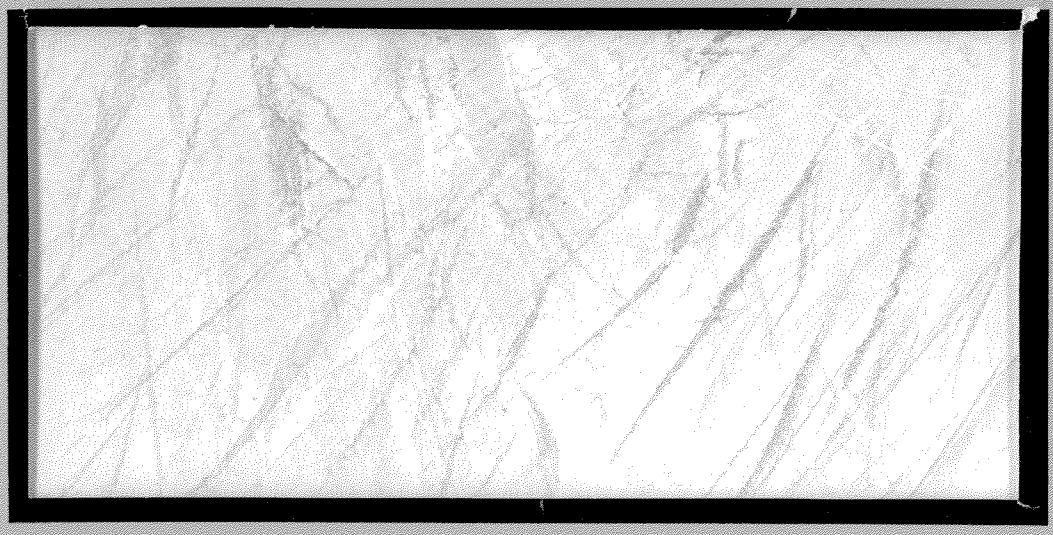
1997-1916

TÜBİTAK

TÜRKİYE BİLİMSEL VE TEKNİK ARAŞTIRMA KURUMU

THE SCIENTIFIC AND TECHNICAL RESEARCH COUNCIL OF TURKEY



Makina, Kimyasal Teknolojiler, Malzeme ve İmalat Sistemleri Araştırma Grubu

Mechanical Engineering, Chemical Technologies, Material Sciences and Manufacturing Systems Research Grant Committee

TEK KAMERA GÖRÜNTÜSÜ İLE CİSİMLERİN 3 BOYUTLU KONUMLARININ BELİRLENMESİ

PROJE NO: MODİSA-8

PROF.DR. BÜLENT E. PLATİN
PROF.DR. ÖMER ANLAĞAN
İ. ALPAR KILINÇ
ERCAN U. ACAR

TEMMUZ 1996
ANKARA

ÖNSÖZ

Bu proje çalışmasında bir cismin 3 boyutlu uzaydaki konumunun bu cisim üzerindeki ikincil bir hedefin tek bir optik kamera görüntüsü ile belirlenmesini sağlayacak bir sistem geliştirilmiştir. Geliştirilen sistemin robot kollarının kontrolu, otomatik sürücülü taşıtlar, uzayda buluşma ve montaj vb. gibi çok değişik mühendislik alanlarında uygulama sahası bulması mümkündür.

Bu proje kapsamında ODTÜ Makina Mühendisliği Bölümü bünyesinde 2 adet yüksek lisans tezi çalışması yürütülmüştür. Bu tezlerden İ. Alpar Kılınç'a ait olanı ODTÜ Prof. Dr. Mustafa N. Parlar Eğitim ve Araştırma Vakfı 1994 yılı Tez Ödülü'nü kazanmıştır.

Proje çalışmaları sırasında ODTÜ Makina Mühendisliği Bölümü'nün tüm altyapı olanaklarından yararlanılmıştır.

Bu projenin maddi desteği tümü ile TÜBİTAK tarafından sağlanmıştır.

İÇİNDEKİLER

ŞEKİL LİSTESİ

ÖZ

Bu çalışmadaki temel amaç, bir cismin 3 boyutlu uzaydaki konumunun bu cisim üzerine monte edilmiş ikincil bir hedefin tek kamera görüntüsü ile belirlenmesini sağlayacak bir sistem geliştirilmesidir. Bu amaca, bu proje kapsamında yürütülen kuramsal ve deneysel çalışmalarda elde edilen sonuçlar ile tümü ile ulaşılmıştır. Geliştirilen sistem yardımı ile, 500 mm - 2000 mm uzaklıktaki bir cismin konumunu ortalama 1 mm ve 0.5° mertebelerinde hassasiyet ile saptamanın mümkün olduğu hem kuramsal çalışmalar ile hem de yürütülen kontrollü deneyler ile gösterilmiştir. Geliştirilen sistemin dayandırıldığı çözüm algoritması, dairesel bir hedefin tek bir kameradaki tek bir görüntüsü yardımı ile bu hedefin dolayısı ile bu hedefin bağlandığı cismin 3 boyutlu uzaydaki konumunun veren 6 konum bilgisini üretmektedir. Bu algoritma daha önce önerilmiş olmasına rağmen, tüm boyutu ile ilk kez bu çalışmada kullanılmış, uygulanabilirlik sınırları kuramsal ve deneysel yöntemlerle test edilmiş ve oluşturulan sistemin performansı saptanmıştır. Bu nedenle, bu çalışma yalnızca alanındaki kuramsal boşlukları kapatmakla kalmayıp, robot kollarının kontrolu, otomatik sürücülü taşıtlar, uzayda buluşma ve montaj vb. gibi çok değişik mühendislik alanlarında uygulamaya dönük somut bir sistem ortaya çıkarmıştır.

ABSTRACT

The main purpose of this study was to develop a system that will determine the three dimensional position of an object by using a single camera image of a secondary target attached on this object. This aim is fully accomplished through the results obtained by theoretical and experimental studies conducted in this project. It is shown by both theoretical studies and controlled experiments that it is possible to determine the position of an object that is about at a distance of 500 mm - 2000 mm, within a precision of 1 mm and 0.5° by means of the developed system. The solution algorithm used in the developed system produces 6 position parameters of an object in three dimensional space by using a single image on a single camera of a secondary target of circular geometry attached on this object. Even though this algorithm is proposed in earlier studies, it is used with all its dimensions first time in this study. The limits of its applicability are tested by theoretical and experimental methods. The performance of the system developed is determined. Therefore, this study does not only fill the gaps in the theoretical sense but also offers a practically implementable system in engineering areas like position control of robotic arms, automatically guided vehicles, rendezvous and assembly in space, etc.

GİRİŞ

İnsan algılaması, ölçmesi, değerlendirmesi, karar vermesi, uygulaması gibi işlemlerden bir kısmını ya da tümünü üstlenen ve otomatikleştirilmiş işlemlerden oluşan sistemlerin istendik bir şekilde çalışabilmesi için bu sistemlerin öncelikle içinde bulundukları ortamın bilgisine gerek vardır. Bu bilginin toplanma işleminin en mükemmel otomatik sistem olan insanda 5 farklı duyu organıyla yapıldığını görürüz. Bunların içinde beyinsel etkinlikler olarak en yoğun olanı ve en çok bilgiyi toplayanı görme organımızdır. Görme duyusu aracılığı ile çevremizdeki cisimlerin şekillerini, yüzey özelliklerini ve konumlarını bu cisimlerin yüzeylerinden yansıyan optik bilgileri kullanarak algılarız. Özellikle geometrisi ve yüzey özelliği önceden bilinen bir cismin tanılanması ve konumunun algılanması sırasında gerçekleştirilen en önemli beyin etkinliği, iki gözümüzden gelen görüntü bilgileri ile eski bilgileri karşılaştırarak bir yargıya varmaktır.

Çevremizdeki cisimlerin 3 boyutlu uzaydaki konumlarına ilişkin bilginin gerekli olduğu ve öncelikle kullanıldığı insan etkinliği ise bu cisimlerin elimizle yakalanmasıdır. Bu yakalama işlemi mühendislik uygulamalarında robot kollarıyla benzetilmeye çalışılmaktadır. Günümüzdeki endüstri tipi uygulamalarda robot kollarına çalışma alanı daha önceden bir teknisyen tarafından öğretilmektedir. Çalışma alanındaki herhangi bir değişikliğin ise robot kola tekrar öğretilme gereği vardır. Ancak akıllı bir robot kolu sisteminin kendini ortamdaki değişikliklere uydurması beklenir. Bu ise ancak ortam bilgisinin sürekli ya da belli aralıklarla toplanıp, işlenerek değerlendirilmesi ile mümkündür. Ortam bilgisinin kullanıldığı mühendislik uygulamaları robotik ile sınırlı olmayıp, otomatik sürücülü taşıtlar, uzayda buluşma ve montaj vb. gibi çok değişik mühendislik alanlarında görülebilmektedir.

Cisimlerin konum bilgilerini türetmek amacı ile çalışma ortamı bilgisini toplamak için kullanılan çok değişik algılayıcılar vardır. Her biri farklı bir fiziksel özelliği kullanan bu algılayıcıların arasında en güçlü olanı, insanın görme duyusuna eşdeğer bir şekilde görüntü üreten algılayıcılar olan optik kameralardır. Çalışma ortamının tek bir kamera görüntüsünden bir çok bilginin türetilmesi mümkündür. Bu bilgilerin

belki de en önemlisi, ortamdaki cisimlerin 3 boyutlu uzaydaki konumlarıdır.

Bu çalışmadaki temel amaç, bir cismin 3 boyutlu uzaydaki konumunun bu cisim üzerindeki ikincil bir hedefin tek kamera görüntüsü ile belirlenmesidir. Bu temel amaca, bu proje kapsamında yürütülen kuramsal ve deneysel çalışmalarda elde edilen sonuçlar ile tümü ile ulaşılmıştır.

Optik temelli algılayıcıların arasında görüntü işleme yöntemleri ile çalışanlarda kullanılan temel ilke, cisimlerin uzayda konum değiştirdikçe bunları izleyen bir kamerada farklı görüntüler oluşturmalarıdır. Bu algılayıcılar ile, cisimlerin geometrik ve yüzeysel (doku, renk vb.) nitelikleri önceden bilinen bir cismin kamera görüntüleri kullanılarak bu cismin gerçek durumu ile ilgili bilgilerinin üretilmektedir.

Bu konuda yapılmış olan araştırmaların çoğunda cisme farklı açılardan bakan iki ayrı kameradan eş zamanlı olarak alınan görüntüler üzerinde stereo görüş ilkesi kullanılmıştır (Barnand, Fischler, 1982; Grimson, 1985; Ayache, Faugeras, 1986; Shoureshi, Mitchell, Cipra, 1987; Weng, Huang, Ahuja, 1987). İki kamera kullanarak insanın görüntü algılama sistemi benzetilmeye çalışılmaktadır. Bu ilkeye dayalı çalışmalarda, eş zamanlı alınan görüntülerin birbirleriyle eşleştirilmesi ile cismin konumu bulunmaya çalışılır. Gerek kamera sayısının iki olması, gerek bu kameralardan alınacak görüntülerin eş zamanlı olma zorunluluğu ve gerekse bu görüntülerin eşleştirilmesi gerektiği, bu tür sistemlerin karmaşıklığını ve maliyetini yükselten faktörlerdir.

Bu projenin stereo görüşe dayalı araştırmalardan en önemli farklılığı, bir cismin stereo görüntüleri yerine bir kameradan alınacak tek görüntüsü ile bu cismin konumunun kestirilmesidir. Üç boyutlu uzayda bir cismin konumunu belirlemek için en az 1 en çok 3 öteleme, geri kalanları da dönme olmak üzere toplam 6 adet konum bilgisine gerek vardır. Tek kameradan alınacak bir görüntüyle uzaklık bilgisini bulmak mümkün değildir. Fakat cismin belirli geometrik özellikleri biliniyorsa, bunlar aracılığı ile uzaklık ile ilgili bilgi de üretilebilir.

Tek kamera görüntüsünden yola çıkılarak yapılmış çalışmalar 1980'li yıllara kadar uzanmaktadır (Kanade, 1981; Strat, Fischler, 1986;

Lowe, 1987; Horaud, 1987; Lamdan, Scwarz, Wolfson, 1988, Sheu, Bond, 1992). Bu çalışmalarda ya gölge oluşumları, ya yüzey üstüne çizilmiş koordinat görüntüleri ya da görüntülenen cismin 3 boyutlu geometrisi ile ilgili bilgiler yoğun olarak kullanılmıştır.

Tek kamera görüntülerinin sağlayamadığı uzaklık bilgisini de türetecek ortamı yaratmak amacı ile cisimlerin üzerine monte edilmiş geometrisi bilinen ikincil hedeflerin kullanımını ancak 1980'li yılların sonundan itibaren görmekteyiz (Olgac, Gan, Platin, 1989; Olgac, Gan, Platin, 1990). Bundan önceki çalışmalardan basitlik ve kolay uygulanabilirlik açısından önemli farklılık gösteren bu çalışmalar, bu projenin yöneticisinin de içinde bulunduğu bir grup tarafından başlatılmış ve yürütülmüştür. Bu proje çalışmasının öncülüğünü de yapmış olan bu çalışmalarda, boyutu bilinen dairesel bir hedefin tek kamera görüntüsü kullanılarak,

a) bu hedefin merkezinin kameraya olan uzaklığını ve
b) hedef düzleminin kameranın optik eksenine dik iki eksene göre dönme açılarını

saptayacak yöntemler geliştirilmiş ve bu yöntemler deneysel çalışmalarla da desteklenmiştir. Bu çalışmaların ilkinde (Olgac, Gan, Platin, 1989) dairenin merkezi kameranın mercek grubunun ekseniyle çakıştırılmış ve dairenin kameranın algılayıcı düzlemindeki ortografik izdüşümü kullanılmıştır.

Literatürde rastlanan benzer içerikli 3 diğer çalışmanın ilkinde (Kite, Magee, 1990) ikincil hedef olarak bir dikdörtgen kullanılmış, konum parametrelerin hepsi bulunmuş, fakat yöntemin hassasiyeti çok düşük kalmış; ikincisinde (Sydow, Cooper, 1992) ise ikincil hedef olarak içinde beş tane benek bulunan bir dikdörtgen kullanılmış; üçüncüsünde ise (Han, Rhee, 1992) içinde 2 benek bulunan bir çember ikincil hedef olarak kamera kalibrasyonu amacı ile kullanılmıştır.

İkincil hedef olarak daire kullanılan ilk çalışmanın (Olgac, Gan, Platin, 1989) daha sonraki geliştirilmiş evrelerinde, (Platin, Gan, Olgac, 1990; Olgac, Craig, Platin, 1991) ortografik izdüşüm yerine perspektif izdüşüm kullanılmıştır. Bu çalışmaların temel handikapları ikincil hedef üzerindeki kısıtlamalar ve konum parametrelerinin bazılarının bilindiği varsayımıdır. Ayrıca, ön çalışmalarda kullanılan konumlama sistemindeki hassaslığın yetersizliği, üretilen deneysel verilerin sağlığını

önemli ölçüde etkilemiştir (Olgac, Craig, Rama, 1991). Bu da, gereğine inanıldığı halde, deneylerdeki parametrik çalışmaların istenilen yoğunlukta yapılmasını engellemiştir.

Bu projenin önerisinde, dairesel hedefli çözümü öngören algılama modeli için çok önemli olan bir deneysel veri bankasının oluşturulması ve bu veri bankası yardımı ile kuramsal olarak oluşturulan algılama modelinin teknolojik olarak yapılabilirliği ve güvenirliği test edilmesi amaçlanmaktaydı. Bu amaca da, bu proje kapsamında yürütülen deneysel çalışmalarda elde edilen sonuçlar ile tümü ile ulaşılmıştır.

Bu projenin önerisinde sunulan önemli amaçlarından bir diğeri ise, önerilen algılama yönteminin gerek sistem parametrelerine ve gerekse algılanması istenen konum değerlerine olan duyarlılığının saptanmasıydı. Bu amaca da, bu proje kapsamında yürütülen kuramsal benzetim çalışmalarında elde edilen sonuçlar ile tümü ile ulaşılmıştır. Oluşturulan sistemin uygulamada karşılaşılacak sınırlarını ve kısıtlamalarını ancak bu tür bilgiler ile sağlıklı bir şekilde saptamak mümkün olmuştur.

Bu proje çalışmasının önerisinde sunulan bir diğer amaç da, dairesel hedefler için yapılmış ve bu proje kapsamında geliştirilmesi ön görülen algılama sisteminin, özellikle hedef düzleminin kendi içindeki açısal konumunu da belirleyebilecek şekilde genişletilmesiydi. Başlangıçta, daha çok kuramsal ağırlıklı bir yapıda çalışmalar gerektirdiği düşünülen bu amaca da, bu proje kapsamında yürütülen kuramsal çalışmaların yanında gerçekleştirilen bir dizi yoğun deneysel çalışmalarda elde edilen sonuçlar ile fazlası ile ulaşılmıştır.

Sonuç olarak, bu proje çalışmasının önerisinde, daha önceden ön çalışmaları yapılmış olan bir optik konum algılama sisteminin ayrıntılı analizi, deneysel doğrulanması, parametrik duyarlılık sınırlarının belirlenmesi, uygulamadaki kısıtlamaların saptanması ve farklı hedefler kullanarak algılama kapasitesinin geliştirilmesi şeklinde özetlenmiş olan amaçlara tümü ile ulaşılmıştır.

Bu proje kapsamında geliştirilen çözüm yöntemi için daha önce yayınlanmış olan iki çalışma (Platin, Gan, Olgac, 1990; Olgac, Craig, Platin, 1991) temel almıştır. Çözüm yöntemi 6 tane konum belirleme parametresini tam ve doğru olarak bulmaktadır. Geliştirilen çözüm

yöntemi ve çözüm yönteminin bilgisayar benzetimleri ile bulunan kuramsal sınırları tüm ayrıntıları ile bir yüksek lisans tezi olarak Ek 1'de (Kılınç, 1994) sunulmuştur. Önerilen çözüm yöntemini gerçek görüntüler kullanarak test edebilmek içinde bir deney düzeneği kurulmuş ve önerilen çözüm test edilmiştir. Bu çalışmalar ise tüm ayrıntıları ile yine bir yüksek lisans tezi olarak Ek 2'de (Acar, 1995) sunulmuştur. Bu projenin önerisinin kapsamını tümü ile içermekte olan bu 2 yüksek lisans tezi, yine proje önerisinde sunulan yöntemlerin tümünü gerekli genişletmeler kullanarak gerçekleştirilmiş ve projenin başarılı bir şekilde sonuçlanmasında belirleyici bir faktör olmuştur.

KURAMSAL ÇALIŞMALAR

Kuramsal çalışmalar, öncelikle bir cismin 3 boyutlu uzaydaki konumunu, bu cismin tek bir kamera görüntüsünden belirleyebilecek bir yapı oluşturulması üzerinde yoğunlaştırılmıştır. 3 boyutlu uzaydaki konum bilgisi, toplam 6 konum bilgisi anlamına gelmektedir. Kuramsal çalışmalar sırasında, bu 6 konum bilgisinin kameraya bağlı eksen takımlarına göre üretilmeğe çalışıldığı varsayılmıştır.

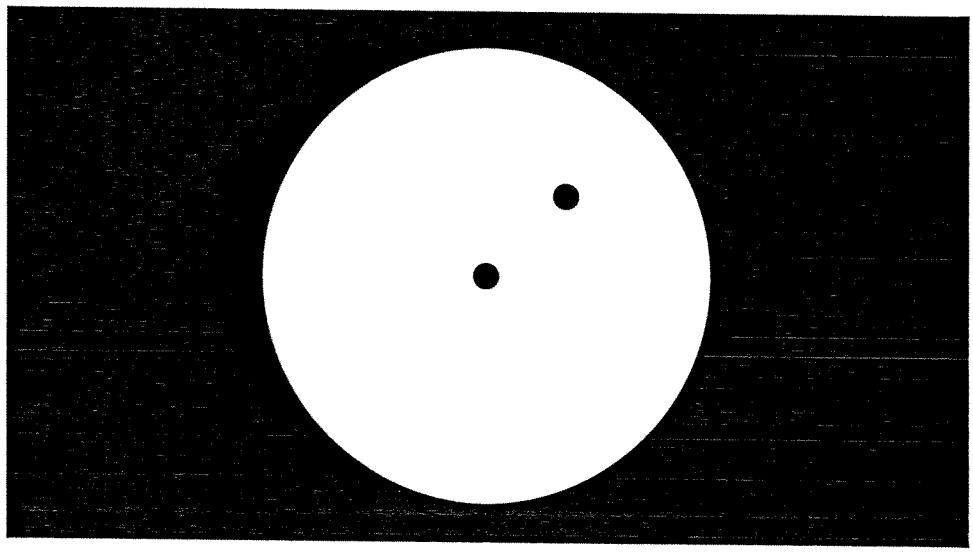
Daha önce yayınlanmış olan iki çalışma (Platin, Gan, Olgac, 1990; Olgac, Craig, Platin, 1991) temel alınarak, tek kameralı ve tek ikincil hedefli alternatif yeni bir çözüm yöntemi geliştirilmiştir. Bu yöntem aracılığı ile, 6 adet konum bilgisi (1 öteleme ve 5 dönme) ikincil hedefin bu kameradaki tek bir görüntüsü kullanılarak bulunabilmektedir.

Bu yöntemin temel avantajları aşağıda sıralanmıştır:

- altı tane konum bilgisinin tümünün üretilebilmesi,
- üretilen çözümlerin tekliği,
- kullanılan çözüm algoritmasının basitliği ve iterasyon gerektirmemesi,
- aynı amaçlı yöntemlere göre daha hassas sonuçlar elde edilmesi,
- monoküler görüntü kullanılması,
- yalnızca gri tonlu görüntülerin kullanılması,
- yapısal aydınlatmaya gerek olmaması,
- hedef üzerindeki noktaların konumlarının bilinmesine gerek olmaması,
- kamera görüntü düzleminin konumunun herhangi bir kamera kalibrasyonunun gerek olmadan bulunabilmesi,
- herhangi bir özel tasarım ya da üretim gerektirmeden, piyasada mevcut donanım kullanılarak uygulanabilmesi,
- Kısmen bloke edilmiş hedef görüntüleri ile bile çalışabilme potansiyeli.

Çözüm yönteminde ikincil hedef olarak daha önceki çalışmaların birinde (Han, Rhee, 1992) kullanılan içinde iki tane benek bulunan

düzlemsel bir daire kullanılmıştır. Hedef geometrisi Şekil 1'de gösterilmiştir. Beneklerden bir tanesi dairenin tam merkezindedir.



Şekil 1. İkincil Hedef Geometrisi

Diğeri ise daire merkezinden yarıçapın yaklaşık yarısı kadar uzakta bir konumdadır. İkincil hedef hakkında bilinmesi gereken parametreler dairenin çapı ve beneklerden bir tanesinin dairenin merkeziyle çakıştığı bilgisidir.

Bu hedefin görüntüsünün elde edildiği kamera ile ilgili bilinmesi gereken tek parametre ise, kamera kalibrasyonu ile doğrudan ilintili olan ve kameranın iç parametresi olarak adlandırılan, kameranın mercek grubunun etkin odak uzaklığıdır.

Çözüm yönteminin geliştirilmesi ve uygulanması ile ilgili tüm ayrıntılar, proje kapsamında gerçekleştirilen ve Ek 1'deki yüksek lisans tezinde (Kılınç, 1994) verilmiştir. Geliştirilen çözüm yöntemi, görüntü elde etme için kullanılan kameraların perspektif görüntü sağladığı temel varsayımına dayanmaktadır. Bu varsayımın geçerli olduğu koşullar ise yine Ek 1'de ayrıntılandırılmıştır.

Çözüm yönteminin ana hatlarını özetlemek gerekirse; dairenin dış çeperinin (yani çemberin) elips şeklindeki görüntü bilgisi 6 adet konum bilgisinin 3'ünün belirlenmesinde kullanılmaktadır. Geri kalan diğer 3 konum bilgisi ise daire içindeki beneklerin merkezlerinin görüntü

bilgileri kullanılarak bulunmaktadır. Çözümü aşağıdaki gibi 3 basmakta toplamak mümkündür:

1. Basamak: Hedef dairesinin merkezindeki beneğin merkezinin görüntü düzlemindeki koordinatları kullanılarak, hedef merkezinin kemeranın optik eksenine göre yatay ve düşey yönlerdeki açısal kaçıklıklar bulunur.

2. Basamak: İlk basmakta bulunan açı bilgileri yardımı ile dairesel hedefin görüntüsündeki çeper bilgileri gerekli koordinat dönüşümünden geçirildikten sonra, dairesel hedefin kamera optik merkezine uzaklığı ile kendi düzlemi dışında yapmış olduğu iki açısal dönme bulunur.

3. Basamak: Son olarak, yukarıdaki bilgilere ek olarak, dairesel hedefin merkezinde olmayan ikinci beneğin merkezinin görüntü bilgileri kullanılarak dairesel hedefin kendi düzlemi içindeki dönme miktarı açısal olarak bulunur.

Yukarıda verilen çözüm yöntemine dayalı olarak Ek 1'deki yüksek lisans tezinde (Kılınç, 1994) geliştirilen bir yan ürün ise, kullanılan kameranın hedef düzleminin kameranın optik merkezine göre konumunun belirlenmesidir. Literatürde kamera kalibrasyonu olarak tanımlanan ve kameranın mercek grubunun her netlik ayarından sonra yeniden saptanması gerekli olan bu bilgi, geliştirilen çözüm yöntemi kullanılarak ve çok az sayıda iterasyon gerektirerek bulunmaktadır.

Yukarıdaki çalışmalara ek olarak, çeşitli kamera-hedef ikilisi konumlarına karşılık gelen yapay görüntüler bilgisayar ortamında hazırlanan benzetimlerle üretilmiş ve geliştirilen çözüm yöntemi bu görüntü bilgileri yardımı test edilmiştir. Ayrıntıları Ek 1'deki yüksek lisans tezinde (Kılınç, 1994) verilen bu çalışma yardımı ile, çeşitli konum parametrelerinin çözüm yöntemine duyarlılığı ve geliştirilen çözüm yönteminin kuramsal sınırları belirlenmiştir.

Bu sınırlar hakkında pratik bir fikir vermek gerekirse;

- 50 mm çapında bir dairesel hedef
- 28 mm odak uzaklıklı bir merceği olan bir kamera
- Boyutları 13 μm x 13 μm olan piksellerden oluşan 400x400 lik bir algılayıcı

dan oluşan bir sistem için, hedef düzlemi

- dönme açıları 10° ile 80°,

- kamera merkezi uzaklığı 500 mm - 1000 mm

arasında kaldığı sürece, geliştirilen yöntem ile elde edilecek ortalama hataların doğrusal konumlamalarda 0.7 mm, açısal konumlamalarda ise 0.25° ile sınırlı kaldığı gösterilmiştir.

Dönme açılarının 10°'den daha küçük değerlerinde, özellikle 0°'ye yaklaştıkça, çözüm yöntemi hala çalışmakta, ancak hedef düzleminin dönme açılarının hesaplanmasındaki hatalar yukarıda verilen hataların 20 katını bulabilmektedir. Bu ise hedef görüntüsünün daireye çok yakın olması beklenen bu özel durum için beklendik bir sonuçtur.

Dönme açılarının 80°'den daha büyük değerlerinde, çözüm yöntemi çalışmakta, ancak bu sefer her türlü doğrusal konum ve hedefin kendi düzlemi içindeki dönme açısının hesaplamasındaki hatalar yukarıda verilen hataların 10-15 katını bulabilmekte, dönme açıları 90°'ye yaklaştığında ise yöntem hiç çalışmamağa başlamaktadır. Bu ise hedef görüntüsünün çok ince ve uzun bir iğ şeklinde olması beklenen bu özel durum için beklendik bir sonuçtur.

DENEYSEL DÜZENEK

Önerilen çözüm yönteminin gerçek bir sistem üzerinde uygulanabilmesi, uygulamadaki güçlüklerin tanılanabilmesi ve kuramsal olarak elde edilen benzetim sonuçlarının test edilebilmesi için bir deney düzeneği oluşturulmuştur. Deneysel düzeneğin ön tasarımı ile ilgili ilk bilgiler Ek 1'deki yüksek lisans tezinde (Kılınç, 1994), bu düzeneğin tasarımı, eleman seçimi, bu elemanların özellikleri, montajı ve kullanımı ile ilgili tüm diğer ayrıntılar, proje kapsamında gerçekleştirilen Ek 2'deki yüksek lisans tezinde (Acar, 1995) verilmiştir.

Bu düzenek aşağıdaki iki alt sistemden oluşturulmuştur:

1. Görüntü algılama sistemi,
2. Hedef konumlama sistemi.

Görüntü algılama sisteminin temel işlevi, 3 boyutlu cisimlerin bir optik kameradaki görüntülerini oluşturmak ve bu görüntüleri bilgisayar ortamında kullanılabilecek bir duruma dönüştürmektir. Bu sistem şu parçalardan oluşmaktadır.

- Kamera
- Mercek grubu
- Görüntü yakalayıcı ve sayısallaştırıcı

Kullanılan kamera CID Technologies Inc. adlı şirketin CID2250D modeli bir CID kameradır. Kameranın görüntü düzlemi 512x512 piksel formatındadır.

Deneylerde özellikleri aşağıda verilen iki ayrı mercek grubu kullanılmıştır:

- Odak uzaklığı 25 mm, açıklık oranı 1:1.4 ve 1 inç görüntü büyüklüğüne sahip Cosmicar CCTV merceği
- Odak uzaklığı 16 mm, açıklık oranı 1:1.4 ve 2/3 inç görüntü büyüklüğüne sahip RS CCTV merceği

Bu çalışmada, görüntü yakalayıcı ve sayısallaştırıcı olarak Data Translation DT3851 görüntü yakalama ve sayısallaştırma kartı kullanılmıştır.

Kontrollu deneylerin yapılabilmesi için hedef konumunun ölçülerek belirlenmesi gerekmektedir. Dolayısı ile, oluşturulan konumlama sisteminin hedefi istenilen herhangi bir konuma istenilen hassaslıkta getirmeye yarayacak 6 serbestlik dereceli bir yapıda olması öngörülmüştür. Tasarımı yapılan ve kısmen yurt dışı satın alma kısmen de yerli üretim ile gereçekleştirilen bu sistemin aşağıda verilen 3 alt bölümü bulunmaktadır:

- Hedef Konumlama Ünitesi
- Kamera Konumlama Ünitesi
- Taban Bloğu

Hedef konumlama ünitesi, ikisi doğrusal ikisi de açısal olmak üzere toplam 4 serbestlik derecesine sahiptir. Bu serbestlik derecelerinden üçü konumlama, dördüncüsü ise hassas merkezleme için kullanılmaktadır.

Kamera Konumlama Ünitesi, üçü doğrusal biri açısal olmak üzere toplam 4 serbestlik derecesine sahiptir. Bu serbestlik derecelerinden biri konumlama, geri kalan üçü ise hassas merkezleme için kullanılmaktadır.

Taban bloğu, yukarıdaki iki konumlama ünitesinin monte edildiği ve koordinat sistemlerinin referans olarak kullandığı tek bir parçadan ibarettir.

DENEYSEL ÇALIŞMALAR

Hedef görüntüsünün yer aldığı karelerin yakalanıp bilgisayar dosyaları şeklinde depolanmasından sonra, ya da isteniyorsa kare yakalanma işleminin hemen ardından, çözüm yönteminde kullanılan piksel adreslerini elde etmek için çeşitli görüntü işleme ve analiz yöntemleri kullanılmıştır. Kullanılan yöntemler Ek. 2'de ayrıntılı bir şekilde anlatılmıştır. Burada, yapılan işlemler sadece özet olarak verilmiştir:

Kontrollu bir çalışma ortamında elde edilen hedef görüntüsünün yer aldığı bir karede, arka planına yeteri kadar gri tonlar hakim olduğundan, uygun bir eşikleme ile hedef görüntüsünü arka plandan ayrıştırmak mümkün olmaktadır. Bu eşikleme işlemin ardından yapılan bir siyah-beyaz ikileme işlemi ile, hedef görüntüsü beyaz, arka plan ise siyah tona dönüştürülmektedir. Bundan sonraki aşama, bu siyah-beyaz görüntü kullanılarak çözüm yöntemi için gerekli olan

- dairenin çeperinin görüntüsü olan elipse ait bilgiler ve
- beneklerin merkezlerinin konum bilgilerinin

elde edilmesidir. Bu amaç ile bu çalışma kapsamında geliştirilen bir kenar çıkarma algoritması kullanılmıştır. Bu algoritmanın ayrıntıları Ek. 2'de verilmiştir.

Benzetim sonuçlarını test etmek amacıyla, çeşitli sistem ve konum parametreleri değiştirilerek bir dizi deneyler yapılmıştır. Deneyler sırasında izlenen ve ayrıntıları Ek. 2'de verilen yöntem aşağıda özetlenmiştir:

1. Kamera görüntü düzleminin taban bloğuna göre yerinin hassas olarak belirlenmesi,
2. Hedefin taban bloğuna göre istenilen bir konuma hassas olarak getirilmesi,
3. Hedefin görüntüsünün yakalanması, sayısallaştırılması ve saklanması,
4. Elde edilen görüntünün işlenmesi ve analizi,
5. Görüntü analizi sonucu elde edilen hedef niteliklerinin çözüm yönteminde girdi olarak kullanılarak hadef konumunun hesaplanması,

6. Hedef konumunun ölçülen ve hesaplanan değerlerinin karşılaştırılması, aradaki hatanın saptanması.

Yapılan oldukça veri yoğun deneysel çalışmalar sonucunda, kuramsal olarak elde edilen uygulama sınırlarına oldukça benzer nitelikte ve mertebede sonuçlar elde edilerek kuramsal çalışmalar doğrulanmıştır. Bu sınırlar hakkında pratik bir fikir vermek gerekirse;

- 60 mm çapında bir dairesel hedef
- 28 mm odak uzaklıklı bir merceği olan bir kamera
- Boyutları 15 μm x 15 μm olan piksellerden oluşan 512x512 lik bir algılayıcı

dan oluşan deneysel sistem için, hedef düzlemi

- dönme açıları 10° ile 80°,
- kamera merkezi uzaklığı 500 mm - 1300 mm

arasında kaldığı sürece, geliştirilen yöntem ile elde edilecek ortalama hataların doğrusal konumlamalarda 1.5 mm, açısal konumlamalarda ise 0.5° ile sınırlı kaldığı gözlemlenmiştir. Bu hata değerleri, kuramsal çalışmalarla kestirilmeğe çalışılan uygulama sınırlarındaki hata değerlerinin yaklaşık iki katına karşı gelmektedir.

İkincil hedefin bir siyah-beyaz görüntüsünden başlayarak bu hedefin konumuna ilişkin 6 adet konum bilgisinin elde Intel486DX66 tabanlı bir kişisel bilgisayarda elde edilmesindeki çözüm hızı 1.3 Hz olarak saptanmıştır. Bu konum bilgilerinin bulunmasında, proje kapsamında geliştirilen çözüm algoritmasının işlenmiş veriler girdi olarak kullanıldığındaki hızı ise aynı tür bir bilgisayar için 18 Hz olmaktadır. Bu da, çözüm algoritmasına işlenmiş veri hazırlayan görüntü işlemle sürecinin hızlandırılması durumunda, sistem performansının oldukça yükselebileceğini göstermektedir.

SONUÇ

Bu çalışmada bir cismin 3 boyutlu uzaydaki konumunun bu cisim üzerindeki ikincil bir hedefin tek bir optik kamera görüntüsü ile belirlenmesini sağlayacak bir sistem geliştirilmesi amaçlanmıştır. Bu amaca, bu proje kapsamında yürütülen kuramsal ve deneysel çalışmalarda elde edilen sonuçlar ile tümü ile ulaşılmıştır.

Önerilen sistem yardımı ile, 500 mm - 2000 mm uzaklıktaki bir cismin konumunu ortalama 1 mm ve 0.5° mertebelerinde hassasiyet ile saptamanın mümkün olduğu hem kuramsal çalışmalar hem de yürütülen kontrollü deneyler ile gösterilmiştir.

Önerilen sistemin dayandırıldığı çözüm algoritması, dairesel bir hedefin tek bir kameradaki tek bir görüntüsü yardımı ile bu hedefin dolayısı ile bu hedefin bağlandığı cismin 3 boyutlu uzaydaki konumunun veren 6 adet konum bilgisini üretmektedir. Bu algoritma daha önce önerilmiş olmasına rağmen, tüm boyutu ile ilk kez bu çalışmada kullanılmış, uygulanabilirlik sınırları kuramsal ve deneysel yöntemlerle test edilmiş ve oluşturulan sistemin performansı saptanmıştır. Bu nedenle, bu çalışma yalnızca alanındaki kuramsal boşlukları kapatmakla kalmayıp, uygulamaya dönük somut bir sistem ortaya çıkarmıştır.

Konuya ilişkin ileriye dönük araştırma önerileri, hedefin otomatik tanınması, görüntü işleme sürecinin hızlandırılması, geliştirilen sistemin kısmen bloke olmuş hedefler için olduğu kadar hareketli hedefler için de çalışmasının sağlanması, çalışma içinde geliştirilen bazı görüntü işleme algoritmalarının piksel-altı koşullarda da çalışmasını sağlıyarak sonuçların hassaslığının daha da arttırılması şekinde özetlenebilir. ODTÜ Makina Mühendisliği Bölümü'nde, ileriye yönelik bu önerilerin bir kısmını kapsayan ve bu proje kapsamında temin edilen techizatı kullanarak yürütülmesi öngörülen bir yüksek lisans tezi başlatılmıştır.

REFERANSLAR

Acar, E.U., "*Experimental Investigation and Implementation of a 3D Configuration Reconstruction Algorithm for an Object Using a Single Camera Image*, (Yüksek Lisans Tezi), Orta Doğu Teknik Üniversitesi, Makina Mühendisliği Bölümü, Eylül, (1995).

Ayache, N., Faugeras, O.D., HYPER: A New Approach for Recognition and Positioning of Two Dimensional Objects, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8, (1986).

Barnard, S.T., Fischler, M.A., Computational Stereo, *Computer Surv.*, 14, 4, (1982).

Grimson, W.E., Computational Experiments with a Feature Based Stereo Algoritm, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-7, (1985).

Han, M.H., Rhee, S., Camera Calibration for Three Dimensional Measurement, *Pattern Recognition*, 25, 2, 155-164, (1992).

Horaud, R., New Methods for Matching 3-D Objects with Single Perspective Views, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9, 3, 401-412, (1987).

Kanade, T., Recovery of the Three-Dimensional Shape of an Object from a Single View, *Artifical Intelligence*, 17, (1981).

Kılınç, İ.A., "*3D Position Reconstruction of Rigid Bodies Using Single Camera Images*, (Yüksek Lisans Tezi), Orta Doğu Teknik Üniversitesi, Makina Mühendisliği Bölümü, Eylül, (1994).

Kite, D.H., Magee, M., Determining the 3D Position and Orientation of a Robot Camera Using 2D Monocular Vision, *Pattern Recognition*, 23, 8, 819-831, (1990).

Lamdan, Y., Scwarz, J.T., Wolfson, H.J., On Recognition of 3-D Objects from 2-D Images, *IEEE Int. Conference on Robotics and Automation*, 3, (1988).

Lowe, D.G., Three Dimensional Object Recognition from Single Two-Dimensional Images, *Artificial Intelligence*, 31, 355-395, (1987).

Olgac. N., Craig, P.D., Platin, B.E., Advancements in Utilizing Monocular Object Configuration Sensor - Part II, *ASME Winter Annual Meeting*, Atlanta, GA, USA, Paper No. 91-WA/DSC-30, ASME Publications DSC-Vol. 30, Advances in Instrumentation, 21-24, (1991).

Olgac. N., Craig, P.D., Rama, P., On the Calibration and Sensitivity Aspects of a Monocular Object Configuration Sensor - Part I, *ASME Winter Annual Meeting*, Atlanta, GA, USA, Paper No. 91-WA/DSC-29, ASME Publications DSC-Vol. 30, Advances in Instrumentation, 13-19, (1991).

Olgac, N., Gan, Z., Platin, B.E., 3D Reconstruction of Object Configurations by Hybrid Projection Analysis Using a Single Camera Image, *1st National Conference on Applied Mechanisms and Robotics*, Cincinnati, OH, USA, Vol. 1, Paper No. 89AMR-4C-6, (1989).

Olgac, N., Gan, Z., Platin, B.E., Cisimlerin 3 Boyutlu Uzaydaki Konumlarının Tek Kamera Görüntüsü Kullanarük Karma İzdüşüm Yardımı İle Belirlenmesi, *4. Ulusal Makina Tasarım ve İmalat Kongresi*, Ankara, Bildiri Kitabı, 203-213, (1990).

Platin, B.E., Gan, Z., Olgac, N., 3D Object Configuration Sensor Utilizing Single Camera Images, *ASME Winter Annual Meeting*, Dallas, TX, USA, Paper No. 90-WA/DSC-22, (1990).

Sheu, D.D., Bond, A.H., A Generalized Method for 3D Object Location from Single 2D Images, *Pattern Recognition*, 25, 8, 771-786, (1992).

Shoureshi, R., Mitchell, R.O., Cipra, R.J., Vision-Based Intelligent Control for Automated Assembly, *ASME Winter Annual Meeting*, DSC-Vol. 5, Intelligent Control, (1987).

Strat, T.M., Fischler, M.A., One-Eyed Stereo: A General Approach to Modelling 3-D Scene Geometry, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8, 6, 730-741, (1986).

Sydow, P.D., Cooper, E.G., Development of a Machine Vision System for Automated Structural Assembly, Report No. NASA TM-4366, NASA Langley Research Center, Hampton, VA 23665-5225, (1992).

Weng, J., Huang, T.S., Ahuja, N., 3-D Motion Estimation, Understanding, and Prediction from Noisy Image Sequences, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9, 3, 370-389, (1987).

EK 1

3D POSITION RECONSTRUCTION OF RIGID BODIES USING SINGLE CAMERA IMAGES

A Master's Thesis
Presented by
İ. Alpar KILINÇ

to
the Graduate School of Natural and Applied Sciences
of Middle East Technical University
in Partial Fulfillment for the Degree of

MASTER OF SCIENCE

in

MECHANICAL ENGINEERING

MIDDLE EAST TECHNICAL UNIVERSITY
ANKARA
September, 1994

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. İsmail TOSUN
for Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Rüknettin OSKAY
Chairman of the Department

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Mechanical Engineering.

Prof. Dr. Bülent E. PLATİN
Supervisor

Examining Committee in Charge:

Prof. Dr. Abdülkadir ERDEN

Prof. Dr. Bülent E. PLATİN

Prof. Dr. Turgut TÜMER

Prof. Dr. Neşe YALABIK

Doç. Dr. Aydan ERKMEN

ABSTRACT

3D POSITION RECONSTRUCTION OF RIGID BODIES USING SINGLE CAMERA IMAGES

KILINÇ, İ. Alpar
M.S. in Mechanical Engineering
Supervisor: Prof. Dr. Bülent E. PLATİN
September 1994, 130 pages.

In this study, an algorithm to determine the 3D configuration of a passive secondary target is presented. The algorithm uses the contour data of a target obtained from a single camera image. The method is based on an earlier study where a circular target was used. The previous method has been expanded by using a composite target and accordingly a new algorithm has been developed. A computer program has been generated to test the proposed algorithm and to determine its accuracy.

It has been proved that the position information of a rigid body can fully be reconstructed in 6 DOF using a secondary passive target of known geometry with a single camera imaging system. The results of the simulation show that the reconstruction error is low enough to compete with other algorithms developed for the same purpose.

This algorithm can be used in automated assembly, mobile robot guidance, robot calibration, camera calibration, and other applications where the use of a passive secondary target is appropriate for configuration sensing.

Keywords: Configuration Sensors, Monocular Vision, Passive Target, Camera Calibration, Machine Vision.

Science Code: 625.01.03

ÖZ

3 BOYUTLU UZAYDA CİSİMLERİN KONUMLARININ TEK KAMERA GÖRÜNTÜSÜ KULLANILARAK BELİRLENMESİ

KILINÇ, İ. Alpar
Yüksek Lisans Tezi, Makina Mühendisliği Ana Bilim Dalı
Tez Yöneticisi: Prof. Dr. Bülent E. PLATİN
Eylül 1994, 130 sayfa.

Bu çalışmada, pasif ikincil bir hedefin 3 boyutlu uzaydaki konumunu belirleyen parametrelerin ölçümü için bir metot sunulmaktadır. Ölçüm için, hedefin tek kamera görüntüsünden elde edilen çeper bilgisi kullanılmaktadır. Bu metot, önceki yıllarda geliştirilmiş, dairesel hedef kullanılan bir çalışmayı esas almaktadır. Kullanılan eski metot, yeni bir hedefin tasarlanması sayesinde değiştirilmiş ve gerekli algoritmalar geliştirilmiştir. Kuramsal olarak yürütülen çalışmaları test etmek ve ölçüm hassasiyetini incelemek amacıyla bir bilgisayar programı yazılmıştır.

Geometrisi bilinen pasif ikincil bir hedefin monte edileceği, 6 serbestlik derecesi olan, herhangi bir cismin konumunun tek kameralı bir görüntü sistemi kullanılarak belirlenebileceği kanıtlanmıştır. Benzetim sonuçları, ölçüm hatalarının aynı amaçla geliştirilmiş diğer programlarla rekabet edebilecek kadar düşük olduğunu göstermektedir.

Bu ölçüm yöntemi, otomatik montaj, gezer robot güdümü, robot kalibrasyonu, kamera kalibrasyonu ve pasif ikincil bir hedefin kullanılmasında engel olmayan diğer uygulamalarda kullanılabilecektir.

Anahtar Kelimeler: Konum Algılayıcıları, Monoküler Görüntü, Pasif Hedef, Kamera Kalibrasyonu, Yapay Görüntü Algılama.

Bilim Kodu: 625.01.03

ACKNOWLEDGEMENTS

I am grateful to my instructor Prof. Dr. Bülent Platin who not only taught me the way of preparing a thesis but also supervised with precious ideas and suggestions.

Many people helped the realization of this thesis by supplying ideas, written materials, softwares or simply with their presence and moral support. Among them I would like to thank to Ercan U. Acar for his valuable efforts in literature survey and for the realization of the optical test bench. I am also grateful to Aytekin Gel who gave me the first lessons on LaTeX. Special thanks go to Cem Kendi and Mutlu D. Cömert with whom we shared the *hard* times within the last three years. I should also mention that I am deeply impressed by the effort Mehmet spent, in order to distract me by supplying advanced Go problems in Japanese!

Finally, I would like to express my gratitude to my family —Güney, Tuncer and Tankut— for the patience they showed and for their support during my studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

d_0	:	image plane distance
d	:	target distance
d_f	:	focal distance
d_{eff}	:	effective target distance
d_x	:	distance of the second virtual image plane
F	:	reference frame
O_f	:	focal point
S_s	:	diagonal sensor size
S_o	:	largest object size
$O_c x_c y_c z_c$	:	the camera coordinate system
$O_o x_o y_o z_o$	:	the optical coordinate system
$O_w x_w y_w z_w$	:	the world coordinate system
$O_i x_i y_i z_i$	:	the image coordinate system
$O_{v1} x_{v1} y_{v1} z_{v1}$	:	the first virtual image coordinate system
$O_{v2} x_{v2} y_{v2} z_{v2}$	:	the second virtual image coordinate system
x_p, y_p	:	coordinates of the principal point
x_s, y_s	:	image coordinates of the outer spot's center
x_0, y_0	:	image coordinates of the target's center
x_{rec}, y_{rec}	:	rectified image coordinates
x_{ref}, y_{ref}	:	refined image coordinates
x_{cor}, y_{cor}	:	corrected image coordinates
b, b_1, b_2, b_3	:	contants for modeling radial distortion
p, q	:	contants for modeling tangential distortion
α_s	:	shearing factor
$\hat{H}_{ij}^{(i,j)}$	:	homogeneous transformation matrix from F_i to F_j
$\hat{R}_x(), \hat{R}_y(), \hat{R}_z()$	:	basic rotation matrices
$\hat{T}_x(), \hat{T}_y(), \hat{T}_z()$	:	basic translation matrices

$\mathcal{H}$	:	perspective projection operator
X, Y, Z	:	coordinates of a point
$\vec{r}_{ij}$	:	vector from point i to point j
α	:	rotation about x axis
β	:	rotation about y axis
γ	:	rotation about z axis
ϕ	:	vertical offset angle
θ	:	horizontal offset angle
Th	:	threshold value
N	:	number of selected pixels
n	:	vertical sensor resolution
m	:	horizontal sensor resolution
k_v	:	vertical pixel size
k_h	:	horizontal pixel size
T_x	:	translation along x axis
T_y	:	translation along y axis
T_z	:	translation along z axis
R_x	:	rotation about x axis
R_y	:	rotation about y axis
R_z	:	rotation about z axis
A, B, C, D, E	:	normalized ellipse coefficients
A_*, C_*, F_*	:	canonical ellipse coefficients
R	:	main target radius
r_0	:	spot radius
r_f	:	spot offset

CHAPTER I

INTRODUCTION

In an automated production environment when a physical task has to be accomplished, the relative position and orientation of the tool and the workpiece should be known. This problem is handled easily in cases where the operational relations of the tool and the workpiece are invariant. For example, in an assembly line, all the operations are known beforehand and moreover the sequence of operations is also defined. Such a system does not bear intelligence and since all the operations are sequentially defined, there is an algorithm type solution based on small portions of the whole task.

On the other hand, more flexible production systems used nowadays are based on task programming. The latter requires an intelligent behavior for the system, since, given a task, the initial state of the environment should first be detected and then an optimal solution leading to the desired configuration has to be found. In systems using artificial intelligence, neither the solution path nor the sequence of operations are defined beforehand but are rather obtained according to the initial and desired states considering the transformation actions available. Of course, this approach requires the knowledge and interpretation of the present status of the environment. The need for gathering real-time information about the environment is the natural consequence of making intelligent actions.

At this point, there comes the importance of sensors used in robotics. Depending on the nature of the required information, a variety of tactile, sonar, visual, infrared and proximity sensors are available. Among them the most powerful information can be obtained using image sensors. This is the reason why so many efforts have been spent in the domain of machine vision.

The recognition of objects using machine vision has attracted many researchers where some have developed methods with the bottom-up reconstruction of the depth information before the recognition. Others spent efforts to first

recognize the objects and then to find the configuration[1] as a by-product of model correspondence. Among them Lowe[1] used perceptual organization approach for model-matching and determined the configuration parameters by correspondence of the 3D model projection and the actual image, similarly Horaud[2] presents a model-based interpretation of single 2D images. A systematic approach to model based object recognition and configuration sensing has been presented in [3] by Sheu and Bond.

The reconstruction of the configuration parameters is of basic importance since in most industrial applications, recognition itself is not the goal but an intermediate step. In recent years a variety of algorithms have been developed for this purpose. The recovery of the depth information from 2D images has been the most challenging research in this domain. Some have solved the depth reconstruction problem with stereo-vision [4]. The purpose being to determine the depth information by using triangulation techniques. The concept of axial stereo is presented in [5] to handle obstacle detection problems. It is a particular case of stereo-vision where the optical axes of the two cameras are coincident.

It has also been proved that the depth reconstruction can be achieved by using single camera images. A unifying approach has been presented by Strat and Fischler[6] based on shading, texture and contour information. The use of a known secondary target[2] is an alternative approach for the reconstruction problem. Olgac et al.[7] used a circular secondary target and orthographic projection model but also made the assumption that the target center lies on the optical axis of the camera. Besides this constraint, the use of a circular target makes the determination of the in-plane rotation, inherently impossible.

The method used by Platin[8] and Olgac[9] is similar except that the orthographic projection model is replaced by a perspective projection model. The constraints on the target configuration and the requirement of a priori knowledge of some configuration variables are the major disadvantages of these algorithms. Olgac et al.[10] have presented an extension of the same study where an experimental method is developed to determine the image plane distance —a characteristic parameter of perspective projection.

[1]The word "configuration" is used for position and orientation
[2]Also called "mark"

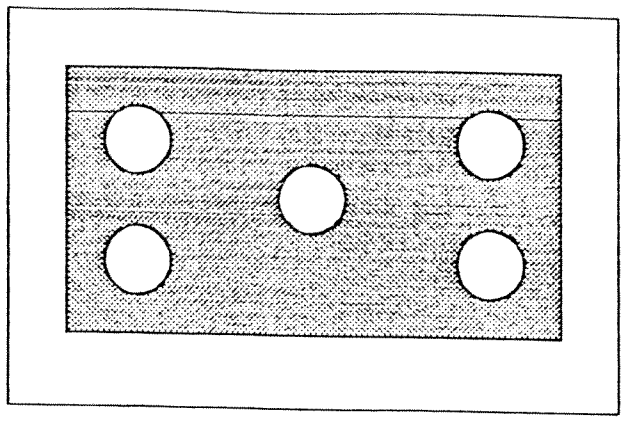


Figure 1.1: The Secondary Target Used by Sydow and Cooper

The technique presented in [11] uses the same secondary target[3] as in this study for camera calibration[4] for a stereo-vision system. However, the purposes and the capabilities of this technique and the proposed new algorithm are quite different. Here it should be stated that the calibration in [11] is incomplete since it makes the same assumption as in [7] requiring a priori the knowledge of two configuration parameters. Another method for 3D reconstruction using single camera has been presented in [12] by Sydow and Cooper where five spots are used as a secondary target. The method has been developed particularly for spatial truss assembly and is able to reconstruct only the position of the target, leaving its orientation arbitrary. The secondary target used in [12] is shown in Figure 1.1.

In [13] the reconstruction of all configuration parameters is made possible by using a rectangle of known size as the secondary target. The method is used to determine the position and orientation of a robot camera. The rectangle having two axes of bilateral symmetry, dual solutions are possible and the test results generated by a simulation program show that the accuracy of the method is poor even at mid-range values of rotation angles.

The method presented in this thesis is inspired from the studies cited above. By trying to find out the weak points of each method, a more powerful algorithm has been developed for the complete reconstruction of configuration parameters. As long as the secondary target is observable, there are no constraints on the configuration variables and a unique solution exists, at least theoretically for all possible configurations of the target. Furthermore, the new algorithm

[3]A circle with two spots inside

[4]Determination of camera configuration with respect to a static target

makes it possible to determine the image plane distance of the camera —a task which was either neglected or accomplished by using a separate calibration process. In fact, in most of the early studies, the image plane distance was named as the "focal length" although these are two different parameters of the camera. Whether this was a conceptual overlooking or not is another question!

1.1 Description of the Problem

It is desired to develop a non-iterative algorithm for the reconstruction of the 3D configuration of a *planar passive secondary target*, using a single 2D image. It is assumed that the image is obtained from a camera connected to a frame grabber hosted by a PC. Accordingly, a target has to be designed such that its image enables the reconstruction of all configuration parameters and shows the necessary contrast with the surrounding objects. Additionally, the algorithm should have the ability to determine the image plane distance which is the key parameter of a perspective projection model. Finally a computer simulation program has to be developed to test the algorithm and to find out its accuracy and its limitations.

The standard image processing operations like segmentation, edge detection, etc. will not be considered as primary problems. Concepts directly related to the formulations of the new algorithm will be studied in detail taking the contour data of the binary target image as the starting point.

This algorithm can be used in automated assembly, mobile robot guidance, robot calibration, camera calibration, and other applications where the use of a passive secondary target is appropriate for configuration sensing.

1.2 Motivations for the Research

Considering the hardware and computation costs, machine vision which is a powerful tool in robotics, should be used within the limits of necessity. High level vision systems are able to recognize objects and produce the necessary information to realize the desired operations. A vision system can use

stereo vision or even color images may be used. However, it is desired to reduce the amount of redundant information and, if possible, to use the simplest system which satisfies the requirements of a problem. Doing so will definitely reduce the flexibility, but on the other hand, will also reduce the hardware and computation costs. This is the main reason why the monocular vision has been presented in this study as an alternative to the stereo vision.

The fact that similar algorithms developed earlier do either require a priori knowledge of one or more parameters or have poor accuracy was another motivation for the research in this area. An effort has been spent to develop new methods to contribute the existing ones and to relax the constraints previously posed.

1.3 Proposed Method for the Solution

The solution process starts with the capture of the image which is digitized by a frame grabber and stored. After a series of standard image processing techniques are applied, the target is recognized, segmented, and its edges are detected. In this study, no additional methods have been developed for the purpose of image processing. This part of the problem may be studied as the continuation of this work and requires an available experimental setup. Hence, a two dimensional contour data is the only input to the reconstruction algorithm. Using this data and curve-fitting techniques, the image of a secondary target is mathematically represented in the image plane coordinates. On the other hand, it is possible to obtain the same representation in terms of the configuration variables using perspective transformations. The problem then is to solve for the unknown configuration variables such that the measured and parametric representations are the same. But, due to the non-linear nature of the perspective projection, the solution process requires an efficient manipulation of the target properties. The capabilities of such an algorithm are mainly determined by the target geometry.

The basic assumptions of this study are as follows:

- the camera is ideal[5]
- lighting condition of the scene provides a clean image
- noise due to data transmission and digitization is negligible
- the secondary target is planar and its geometry is perfect
- the target can be mounted on the rigid body of interest without causing an obstacle for other moving systems in the surroundings
- the values of configuration parameters are kept within a range such that the target is visible
- the image of the target is not masked by other objects in the scene
- the target is located in a range where its image is not blurred

The main advantages of the new algorithm are:

- monocular vision
- simple, non-iterative solution
- capability to reconstruct 6 configuration parameters
- uniqueness of solution
- better accuracy compared to similar algorithms
- possibility to solve for partial target images
- use of gray level images
- ability to determine the image plane distance of the camera without any calibration
- no necessity for structural lighting
- possibility to select different hardware depending on the accuracy requirements

[5]The ideal camera will be described in Section 2.5

- adaptability to commercially available imaging equipment

The parts that should be studied separately in the future are:

- target recognition
- histogram operations
- image binarization and the selection of a proper threshold level
- reflectance characteristics of the target material
- the use of a camera with zoom
- improvement of solution speed[6] for motion tracking

1.4 Experimental Set-up

The project which the study presented here belongs to has another part consisting of the realization of an optical test bench and an imaging system for getting experimental results. For this purpose a CID[7] camera of 512x512 square pixel resolution has been selected. The camera is hosted by a 486DXII PC equipped with an image processing card. The positioning mechanisms are selected from off-the-shelf commercially available optical equipment. Only a few combination plates have been designed for the specific needs of this model. The target configuration is to be set manually using the positioning equipment in a controlled manner. By means of this set up the readouts and the reconstructed values will be compared and a more realistic error analysis will be achieved using experimental data. The optical test bench is pictured in Figure 1.2 where the target and camera bases can be observed.

[6] In robotics applications the required speed is likely to be higher than 15Hz
[7] Charge Injection Device

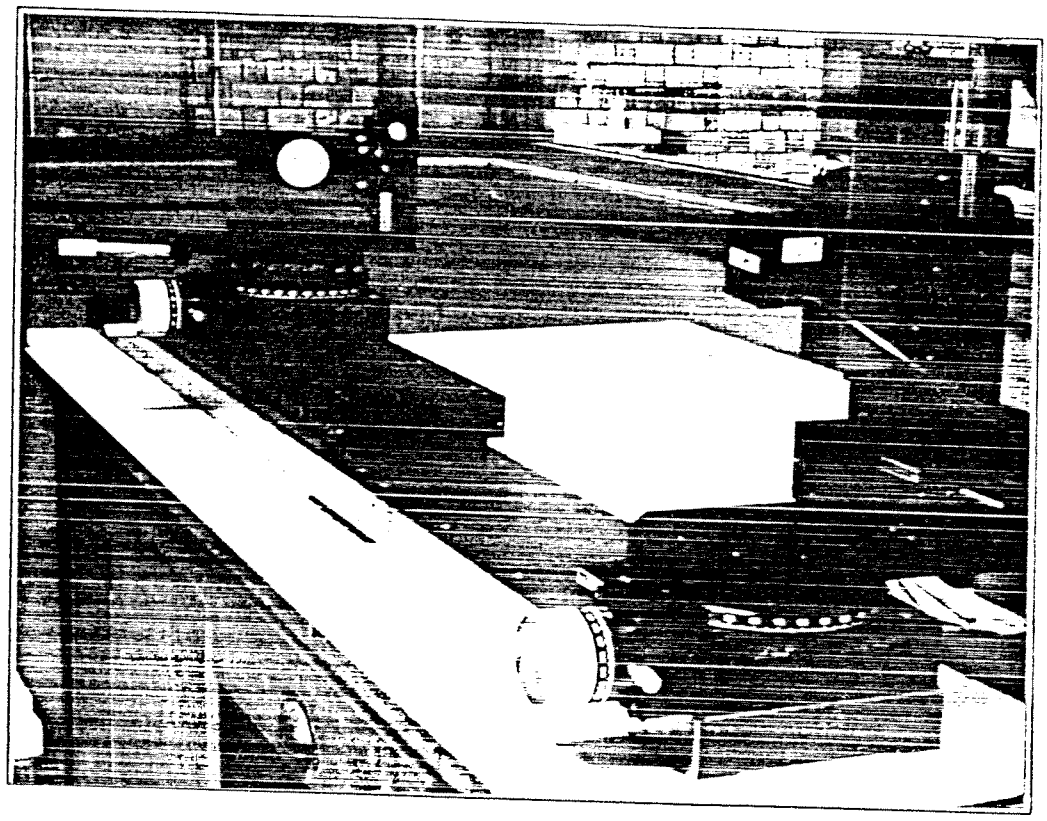


Figure 1.2: The Optical Test Bench

1.5 Outline of the Study

Chapter II describes the concepts related with cameras and images. It starts with the study of simple optical systems and the relevant parameters. It contains the discussion on some confusing points of the available sources. Then the concept of "digital image" is introduced and the operating principles of a camera are explained. The camera modeling using the perspective projection principles is also presented in this Chapter. Finally, some standard image processing techniques are reviewed.

In Chapter III, the so-called target centered solution is reviewed and the new algorithm is presented. This Chapter starts with the definition of coordinate systems and related coordinate transformation matrices. The non-iterative unique solution process to determine the configuration of the secondary target with 6 DOF is explained in every detail.

Chapter IV explains the essential parts of the simulation program developed to support the theory and to test the accuracy of the system. The source file is written in Turbo Pascal language, and is able to generate several

binary images corresponding to different target configurations.

Chapter V investigates the practical limits of the algorithm and compares different solution types. The simulation results concerning each type of solution are given right after the discussions.

The outcomes of the study as well as directions for further improvements are discussed in Chapter VI.

Mathematical subjects directly related with the formulations have been included in the Appendices. Appendix A reviews the homogeneous coordinate transformation matrices. Appendix B shows the analytical steps for the canonical representation of an ellipse equation. Appendix C gives the proof for the invariance of the degree. Appendix D explains the pseudo-inversion technique used for inverting non-square matrices.

CHAPTER II

DIGITAL IMAGING USING A CAMERA

2.1 The Image of a Scene

In order to understand the working principles of a camera one has first to look at its optical properties. In its simple form, an optical system using a single convex lens can be represented as in Figure 2.1, where A' is the image of A.

The "image plane distance"[1] d_0 is directly related to the "object distance"[2] d and the focal distance d_f by the lens equation;

$$\frac{1}{d_0} + \frac{1}{d} = \frac{1}{d_f} \tag{2.1}$$

The focal distance d_f is the characteristic parameter of the lens used, and is a constant value in a single lens system. The object distance d is determined by the location of the observed object with respect to the lens. Therefore, d_0 has to be adjusted to obtain a clear image. Figure 2.1 depicts an optical system for which the lens equation is satisfied. It is observed that the condition to have an unblurred image is that the two rays parting from A should intersect exactly on the image plane. This is the geometric interpretation of Equation 2.1. Its consequence is that the image of a *point* has to be another *point*. On the other hand, if this perfect focusing condition is not met, then the image of a *point* will become a *spot* and spots as images of individual points of an actual object will construct a blurred overall image of this object.

In Figure 2.2b, d_0 is adjusted for the perfect focus when the observed object A is closer to the lens. It should be mentioned that the focal length d_f is the same for Figures 2.2a and 2.2b.

[1]The distance of the image to the center of the lens, also called "principal distance"
[2]The distance of the object to the center of the lens

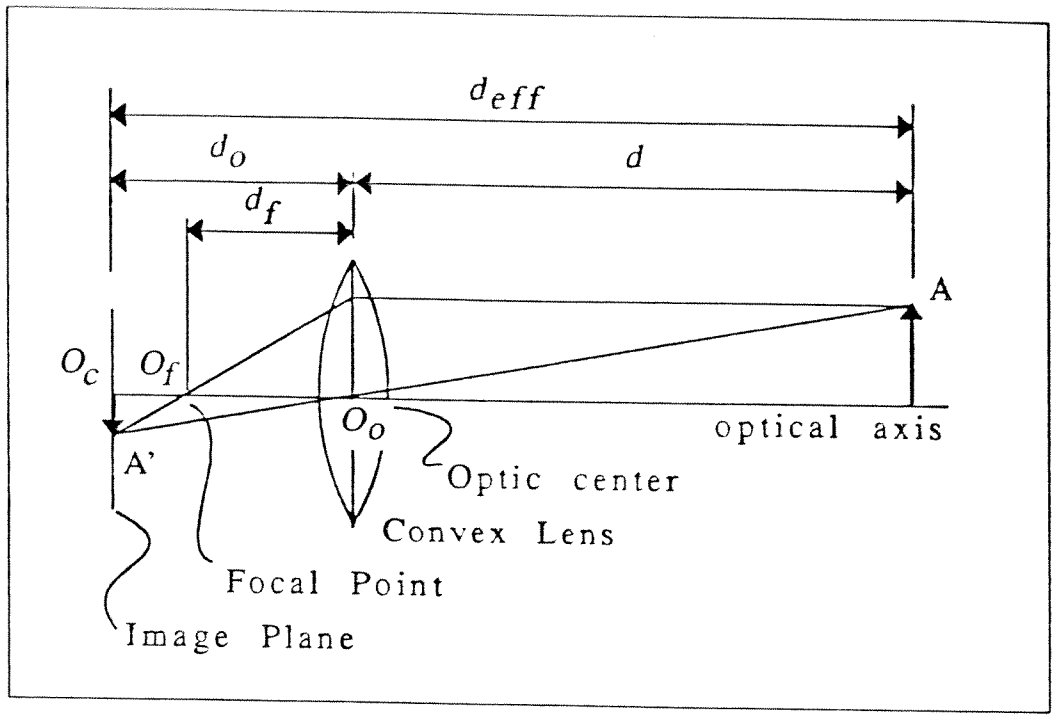


Figure 2.1: A Simple Optical System

In practice, the focusing is done by moving the lens back and forth along the optical axis. Since the image plane remains stationary with respect to the base of the optical system, it makes more practical sense to use the effective distance d_{eff} in the formulations, instead of using the object distance d which depends on the location of the optic center. In fact, when $d = d_{eff} - d_0$ is substituted, the Equation 2.1 becomes;

$$\frac{1}{d_0} + \frac{1}{d_{eff} - d_0} = \frac{1}{d_f} \tag{2.2}$$

and with some manipulations it is possible to obtain a quadratic equation in terms of d_0 as:

$$d_0^2 - d_{eff} d_0 + d_{eff} d_f = 0 \tag{2.3}$$

Solving this equation for d_0 one can obtain:

$$d_0 = \frac{d_{eff} \pm \sqrt{d_{eff}^2 - 4 d_{eff} d_f}}{2}$$

In order to have a real solution for d_0,

$$d_{eff}^2 - 4 d_{eff} d_f \geq 0$$

should also be satisfied. Considering the fact that d_{eff} and d_f should both have positive real values, this inequality reduces to:

$$d_{eff} \geq 4 d_f$$

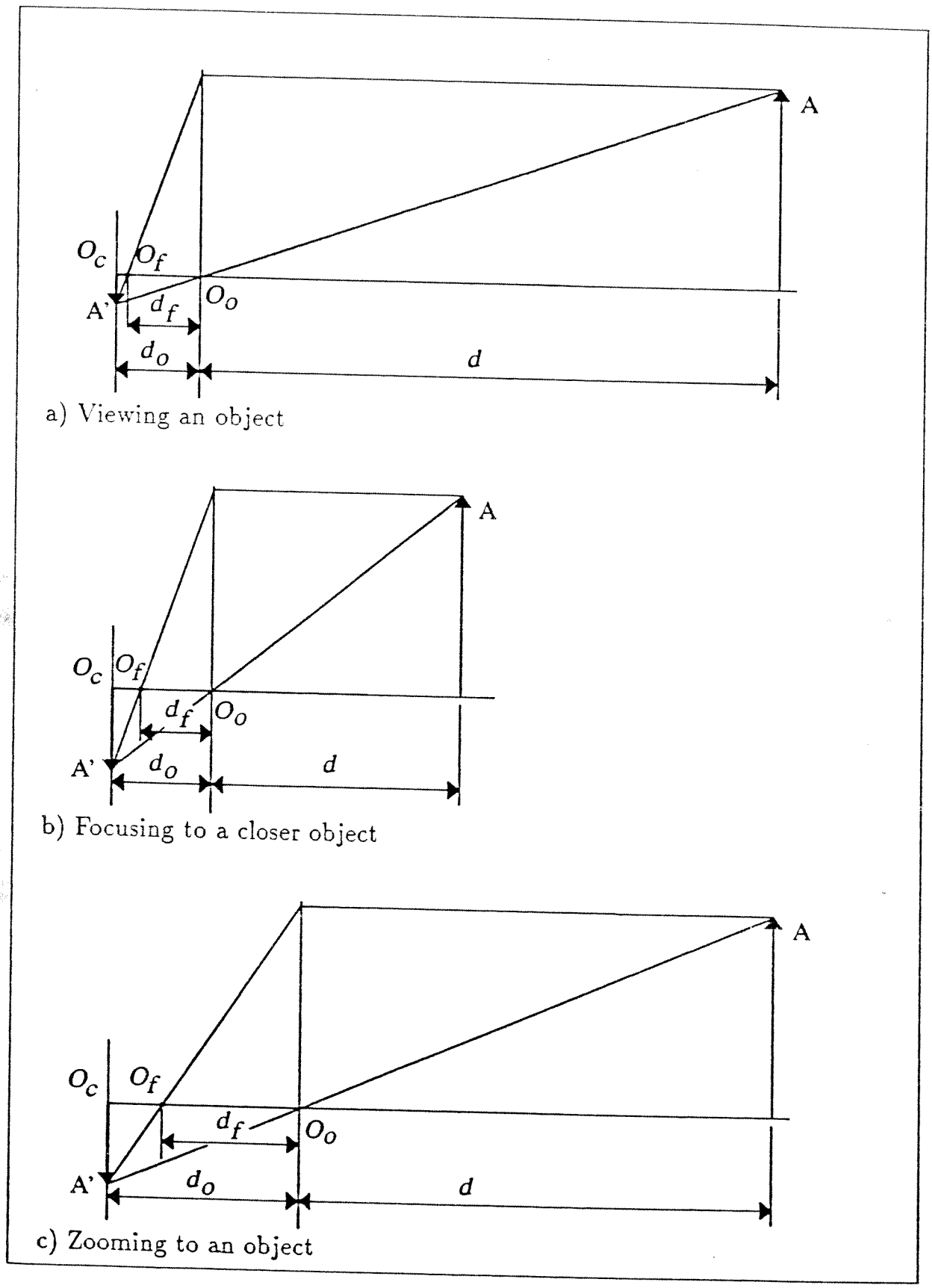


Figure 2.2: Effect of Focus and Zoom

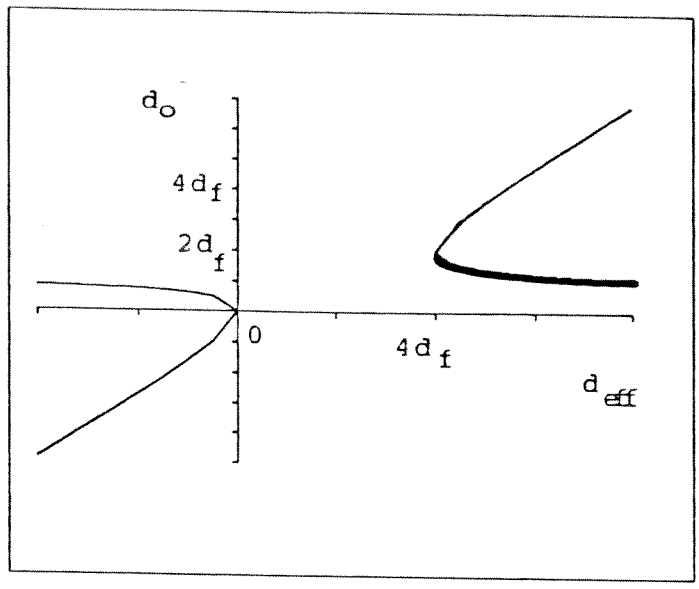


Figure 2.3: Behavior of d_0 vs. d_{eff} Curve

On the other hand, in order to overcome the duality in the solution of Equation 2.3 the following condition can be used:

$$d_{eff} \rightarrow \infty \Rightarrow d_0 \rightarrow d_f$$

which gives the following physically oriented solution:

$$d_0^{(\text{perfect})} = \frac{d_{eff} - \sqrt{d_{eff}^2 - 4d_{eff}d_f}}{2} \qquad d_{eff} \geq 4d_f \qquad (2.4)$$

Figure 2.3 illustrates the d_0 versus d_{eff} graph. The part of the graph corresponding to the physical solution has been indicated by the thick curve.

Despite the fact the Equation 2.4 is not exactly satisfied, objects in the close neighborhood of A will give a clear, tough not perfect, image because when $d_{eff} >> 4d_f$ the variation in $d_0^{(\text{perfect})}$ becomes very small. In most vision problems, this slight blurring is negligible and the image analysis can be carried out. However if the region where the objects of interest are located is wide and close to the optical system, then either an auto-focus mechanism should be used or the image should be unblurred by using image restoration techniques.

By utilizing different lens types, the dimensions of the image can be changed without changing the location of the camera. Figure 2.2c illustrates how the image of the object A can be enlarged by using a different lens. The focal distance of the lens in Figure 2.2c is greater than the focal length of the lens

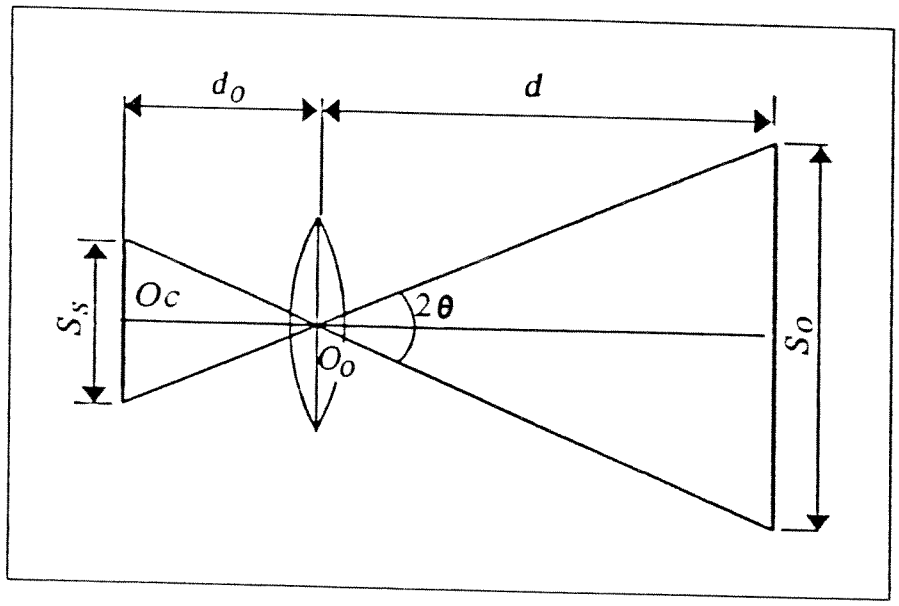


Figure 2.4: Angular Field of View

in Figure 2.2a. Multi lens optical systems which are not going to be described here can be represented by a virtual single lens system, for a constant setting. Actually these systems are able to change the em resultant focal distance without replacement of the objective.

In a single lens system however, the lens should be selected considering the size of objects, the size of the image plane and the observation distance. Using the similar triangles in Figure 2.4 one can write:

$$d_0 = \frac{S_s d}{S_o} \tag{2.5}$$

where S_s is the diagonal size of the sensor, S_o is the largest possible object size and d is the average object distance. Next the desired focal length can be calculated using Equation 2.1 as:

$$d_f^{(\text{desired})} = \frac{d d_0}{d + d_0}$$

Since the commercially available lenses have predefined focal distances like 25, 28 or 50mm, the lens whose focal distance is closer to the calculated value $d_f^{(\text{desired})}$ has to be selected.

The angular field of view of an optical system is also shown in Figure 2.4 as 2θ and can be calculated approximately using;

$$\tan\theta \approx \frac{S_s/2}{d_f} \tag{2.6}$$

The formula being exact for an object at infinity, can be used with an accuracy of 90% for object distances down to $10d_f$.

2.2 Digital Image

In a broad context, "digital image processing" refers to the processing of any two dimensional data by means of a digital computer. A "digital image" therefore is an array of real or complex numbers represented by a finite number of bits. With this generalized definition, the digital image processing is used in a variety of applications. Among these are image transmission and storage for business applications, medical image processing, radar, sonar and acoustic image processing, robotics, automated inspection and security.

A digital image represents the luminosity of objects in a scene in ordinary imaging, the absorption characteristics of the body in X-ray imaging, the temperature profile of a region in infra-red imaging.

In robotics, the machine vision is widely used for gathering visual information in a working environment. In machine vision, a digital image is referred to as the two dimensional data bearing luminosity information of the observed scene. In most applications, "gray level" images are preferred to color images because they contain the necessary geometrical information and are easier to store and process. The visual information is converted to video signals by a CCD[3], a CID[4] or a line-scan camera. Basically, a CCD or a CID camera has a retina composed of "pixels". The image is sampled to an $n \times m$ array where each sample is quantized in its intensity. This array is called a "digital image". A digital image can be represented by a two dimensional function $f(x, y)$. The value of this function at any given point (x, y) in the image plane is the intensity at that point. The resolution of a device can go up to 1024×1024 pixels and even to higher values according to the requirements of the problem. The intensity can be represented by a "gray level" value ranging from absolute black to absolute white. The process of digitization of a given scene at a given time is commonly called "frame grabbing".

[3]Charge Coupled Device
[4]Charge Injection Device

2.3 The Working Principle of a CCD Camera

CCDs are a new family of silicon semi-conductor components capable of performing the general functions of image sensing, analog signal processing, and digital to analog memory. Improved LSI [5] techniques have been developed to realize the full capability of CCD concepts.

CCD linear imaging devices have made possible the new generation of fast facsimile machines. They are also used in high speed mail sorting, rapid quality control measurements and real-time aerial mapping.

CCD area imaging devices are used in small, low power cameras capable of operating in very low light levels. Their application fields are mainly TV broadcasting, automatic production , surveillance and military systems.

The technical details related to the operation of a CCD have been conveyed in a quite simple form. More details and technical data on CCDs can be found in catalogues. The following sections have been summarized from [14].

2.3.1 Charge Coupling

The operating principle of a CCD is the so called *charge-coupling*. Finite amounts of electrical charge called *packets* are created in specific locations in a silicon semi-conductor material. The specific locations are called *storage elements* each created by the field of a pair of gate electrodes very close to the surface of the silicon at that location. By placing the storage elements adjacent to each other, voltages on the adjacent gate electrodes can be alternately raised and lowered. This causes the individual charge packets to be passed from one storage element to the next one. Since each packet can contain different amount of charges, the line of adjacent storage elements becomes a simple analog shift register. Therefore all CCDs are basically shift registers. The transfer from one element to the next one being very efficient, the amount of charge in each packet stays substantially the same even after it has been passed from one to as many as a thousand adjacent elements. The string of charge packets can represent analog

[5] Large Scale Integration

information since the amount of charge in each packet is unique.

The image sensors utilize another basic characteristic of silicon semiconductor devices. This is the photoelectric effect by which free electrons are created in a region of silicon illuminated by photons in the approximate spectral range of 300 nm (UV[6]) to 1100 nm (IR[7]) wavelength. Absorption of such incident radiation in the silicon generates a linearly proportional number of free electrons in the specific area illuminated. If a silicon device structure having a repetitive pattern of small but finite photo sensing sites is created, the number of free electrons created will be directly proportional to the incident radiation on that specific site. If the pattern of incident radiation intensity is a focused light image from an optical system viewing a scene, then the charge packets created in the finite photo-sites array will be a faithful reproduction of the scene projected on its surface. After the appropriate exposure time, the charge packets are simultaneously transferred by charge coupling under an adjacent single long gate-electrode called a transfer gate, to a parallel CCD analog transfer shift register.

Each charge packet corresponds to a picture element also called pixel and when transferred to the adjacent CCD transport shift register, continues to represent the total sensed radiant energy. The transfer gate is immediately returned to the non-transfer clock level so photo-sites can begin integrating the next image information. At the same time, the CCD analog transport register is rapidly clocked to deliver the picture information it contains, in serial format, to the output circuitry.

Area imaging devices have an x-y array of pixels and sense an area image. They are built with both vertical and horizontal transfer gates and transport registers, and deliver an entire field of video information from each integration period in the form of a series of lines of video signal. Such an area imaging device has been shown in Figure 2.5.

[6]Ultra-Violet
[7]Infra-Red

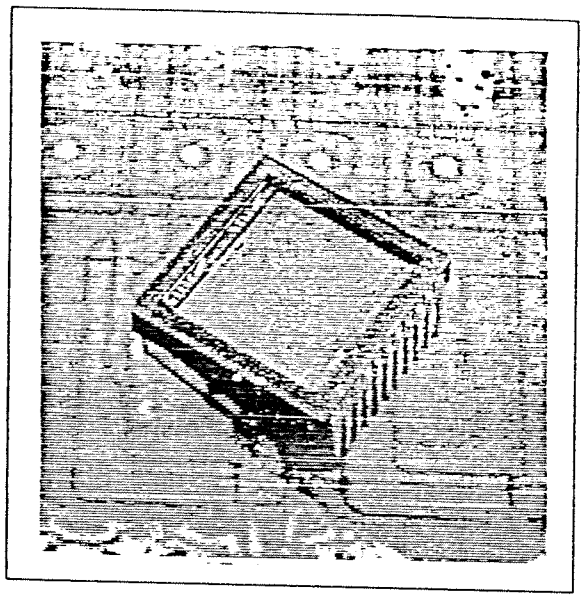


Figure 2.5: A CCD sensor

2.3.2 Light and Integration Time Requirements

Any light source within the visible spectrum is recommended to use with CCDs. Daylight fluorescents or filtered tungsten are the usual choices. Monochromatic lasers are also acceptable in the visible spectrum. The reason why tungsten has to be filtered is that it contains some IR light and the optical cross talk between pixels increases with increasing wavelength (resulting in a poor image). For certain applications where low contrast images are acceptable IR sources up to 1μm may still be used.

The integration time to acquire an image using a CCD depends on the wavelength and light intensity. The exposure is proportional to the integration time by a factor of light intensity. Therefore very short integration times, in the order of nanoseconds are possible with an intense enough light. The exposure adjustment is more important in a normally lighted environment, and the integration time can be kept constant.

2.3.3 The Accuracy of a CCD Camera

The geometric accuracy of a CCD is almost perfect. The angular tolerances for most CCD are given as $\pm 0.0^{\circ}$ and the positional tolerances are in the order of 0.1mm for worst cases. The pixel size tolerances are also very low.

The spacing of two adjacent pixels is accurate within $\pm 0.020\ \mu m$ and the overall cumulative spacing error[8] is $\pm 0.00006 \mu m$/pixel.

The accuracy of the CCD sensor being so good, the overall accuracy of a camera depends more on the quality of the lens used. The so-called lens distortion problem can be handled by using proper image restoration techniques. Its effect is mainly the curved appearance of straight lines on the image plane. Due to its highly radial behavior, the objects that are farther away from the optical axis are most influenced by lens distortion.

2.4 The Advantages of a CID Camera

CID cameras are designed to provide a better performance in certain applications where the use of a CCD does not give satisfactory results. Basically in a CID, the collected charge is not transferred out of the pixel as in CCDs. The readout is non-destructive because the charge remains intact in the pixel after the signal level is determined. To empty the pixel for new frame integration, row and column electrodes are biased to "inject" the charge packet into a substrate collector. The major advantages of a CID can be cited as follows:

- By using the non-destructive readout, camera readout may continue at video rates, but charge injection can be inhibited, so a developing image can be viewed on a monitor as the integration proceeds. This feature is especially useful for low light applications.
- The anti-blooming performance ensures accurate image detail even under extreme lighting conditions, making the CID cameras particularly effective for testing and measurement.
- The contiguous pixel structure further contributes to accuracy for pixel to pixel measurements.
- CID imagers offer a wide spectral response, extending imaging range from 185nm (UV) to 1100nm (IR) without blooming, fringing or loss of resolution.

[8]Cumulative spacing error is equal to the total length block error divided by the number of pixels

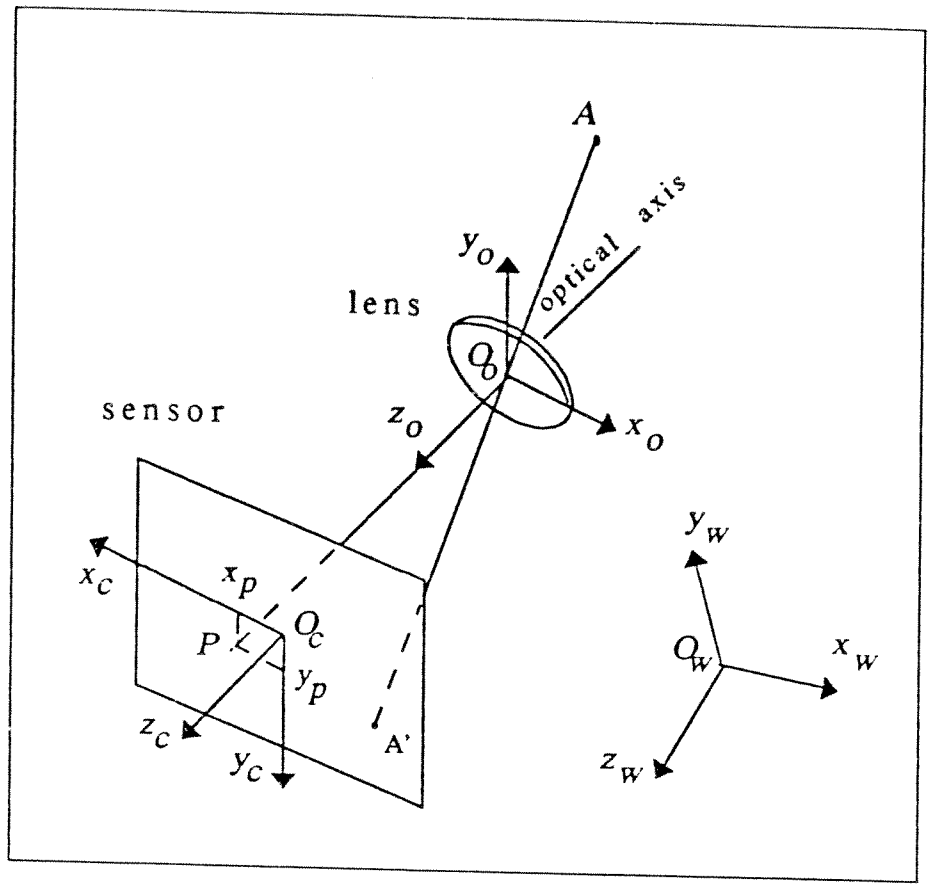


Figure 2.6: A Camera Model

- CID cameras are less vulnerable to disruption in radioactive environments.

2.5 Camera Modeling

With the help of concepts described in the previous sections, the similarity between a perspective projection model and a camera is already obvious. In this section some conventional methods of representing an ideal camera by the perspective projection model will be explained. In the following Chapters this model will be used for the formulations of the reconstruction algorithm.

A camera can be depicted as in Figure 2.6 using a perspective projection model where the perspective center is the optic center O_o and the image plane is the sensor array. The figure also shows the 3D cartesian coordinate systems that have been defined and used in the formulations;

- The world coordinate system $(O_w x_w y_w z_w)$ is positioned at some convenient point, its orientation is also arbitrary.

- The camera coordinate system $(O_c x_c y_c z_c)$ is fixed to the sensor such that O_c is at the geometric center of the sensor, z_c is perpendicular to the sensor plane and x_c is parallel to the horizontal of the sensor.
- The optical coordinate system $(O_o x_o y_o z_o)$ is fixed to the lens such that O_o is at the optic center and z_o is the optical axis.

The parameters related to a camera can be divided into two groups which are the external and internal parameters. For a CCD or CID camera, the external parameters define the position and orientation of the sensor, with respect to the world coordinate system. Therefore the external parameters can be determined if the coordinate transformation matrix $\hat{H}_{wc}^{(w,c)}$ is known. The method presented in this study is mainly intended for the determination of target configurations and necessitates the knowledge of the camera's external parameters. However, by the explicit knowledge of the target configuration, the algorithm can be used in reverse to determine the external parameters of the camera.

The internal parameters of a camera are its image plane distance, coordinates of the principal point, focal distance and image distortion parameters. The idealization of a camera model is rather based upon its internal parameters and they will be examined in detail.

The image plane distance has already been introduced in Section 2.1 and its definition is $d_0 = \overline{O_o P}$. For the ideal camera, it is assumed that d_0 is always adjusted to its perfect value using Equation 2.4. As explained in the preceding sections, it is possible to obtain a clear image even if this assumption is not true. Yet the numerical value of d_0 is required in this study for the solution of the problem. Therefore it is desired to find or to estimate the value of d_0 for a given lens setting. For this purpose an iterative estimation algorithm will be introduced in Section 5.5.

As illustrated by Figure 2.6 the principal point P is the intersection of the optical axis and the image plane. The actual coordinates (x_p, y_p) of the principal point can be determined on an optical test bench. When the principal point and the geometric center of the sensor are not coincident, the corrected image coordinates (x_{cor}, y_{cor}) are found using

$$x_{cor} = x_i - x_p \qquad y_{cor} = y_i - y_p$$

where (x_i, y_i) are the actual image coordinates of a point. The result is simply a coordinate shift. In an ideal camera model, it is assumed that P and O_c are coincident. It is further assumed that the sensor is perpendicular to the optical axis.

The focal length of an ideal camera is taken as the catalogue value supplied for the specific lens. In the simulation part of this study a 28mm lens will be used.

A conventional mathematical model used for image distortion handles the problem in two steps. First the linear distortion is modeled and the so-called *rectified* image coordinates (x_{rec}, y_{rec}) are calculated using

$$x_{rec} = x_{cor}/c_x$$

$$y_{rec} = \alpha_s x_{rec} + y_{cor}/c_y$$

where, c_x, c_y are calibration constants and α_s represents the shearing factor. Then the radial and tangential distortions are modeled and the *refined* coordinates are obtained using

$$x_{ref} = bx_{rec} + p(r^2 + 2x_{rec}^2) + 2qx_{rec}y_{rec}$$

$$y_{ref} = by_{rec} + q(r^2 + 2y_{rec}^2) + 2px_{rec}y_{rec}$$

where,

$$r^2 = x_{rec}^2 + y_{rec}^2 \qquad b = 1 + b_1 r^2 + b_2 r^4 + b_3 r^6$$

The radial distortion is modeled by b_1, b_2, b_3 and the tangential distortion by p and q. In this study, all types of lens distortion are neglected and it will be assumed that the optical system of the camera is perfect.

So far it is observed that the image obtained from an optical system is upside down. To get rid of this effect, an additional coordinate system will be defined.

- The image coordinate system $(O_i x_i y_i z_i)$ is not attached to a physical object, but its position and orientation at any setting is defined by the following coordinate transformation matrix:

$$\hat{H}_{ci}^{(c,i)} = \hat{R}_z(\pi)\hat{T}_z(-2d_0) = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -2d_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

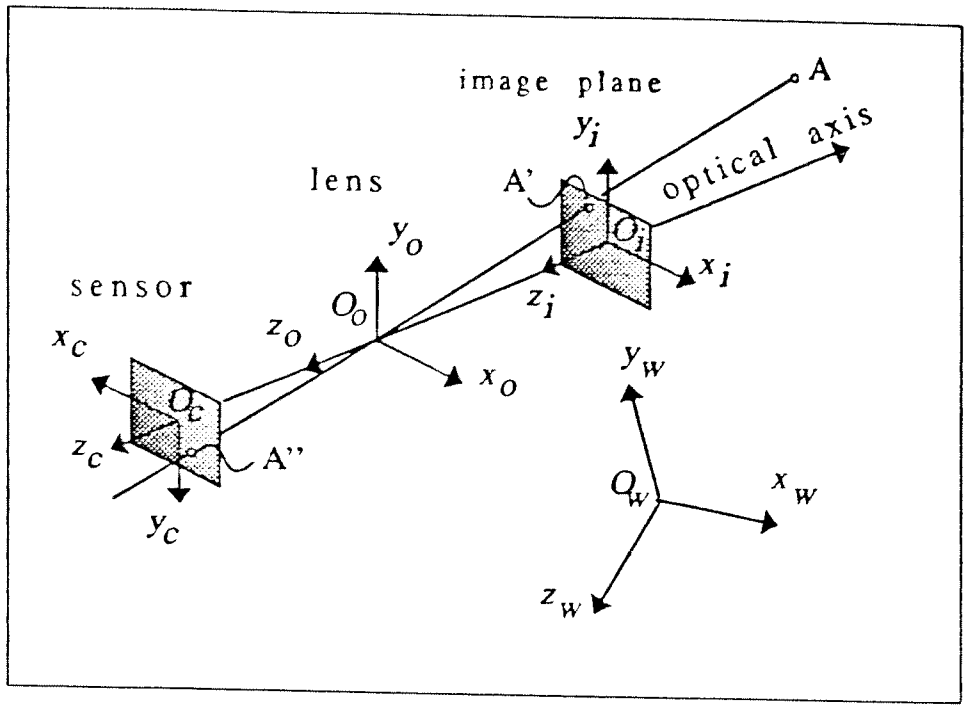


Figure 2.7: Definition of the Artificial Image Plane

The axes x_i and y_i form an imaginary plane, symmetric to the original image plane about O_o. This new configuration is shown in Figure 2.7. Due to the symmetry obtained by the proper selection of $\hat{H}_{ci}^{(c,i)}$ there is no difference in the image coordinates of a point whether it is expressed in the camera coordinate system or in the image coordinate system. Therefore,

$$A''(x_c, y_c) = A'(x_i, y_i)$$

In Figure 2.7 A'' is the image of A, and A' is the conventional representation of A. It is also observed that any other point lying on the line O_oA will give the same image point A''. Therefore it is not possible to determine the range of a point from its image. The ideal camera can be taught as a non-linear transformer from 3D to 2D which uses the following formulation:

$$f(x, y) = \mathcal{H}[g(x, y, z)]$$

where $\mathcal{H}$ denotes the perspective projection operator, $g(x, y, z)$ is any function and $f(x, y)$ is the corresponding image function.

Figure 2.8 gives a more precise geometric description of perspective projection model. The coordinates of A are given by (X, Y, Z) in the image

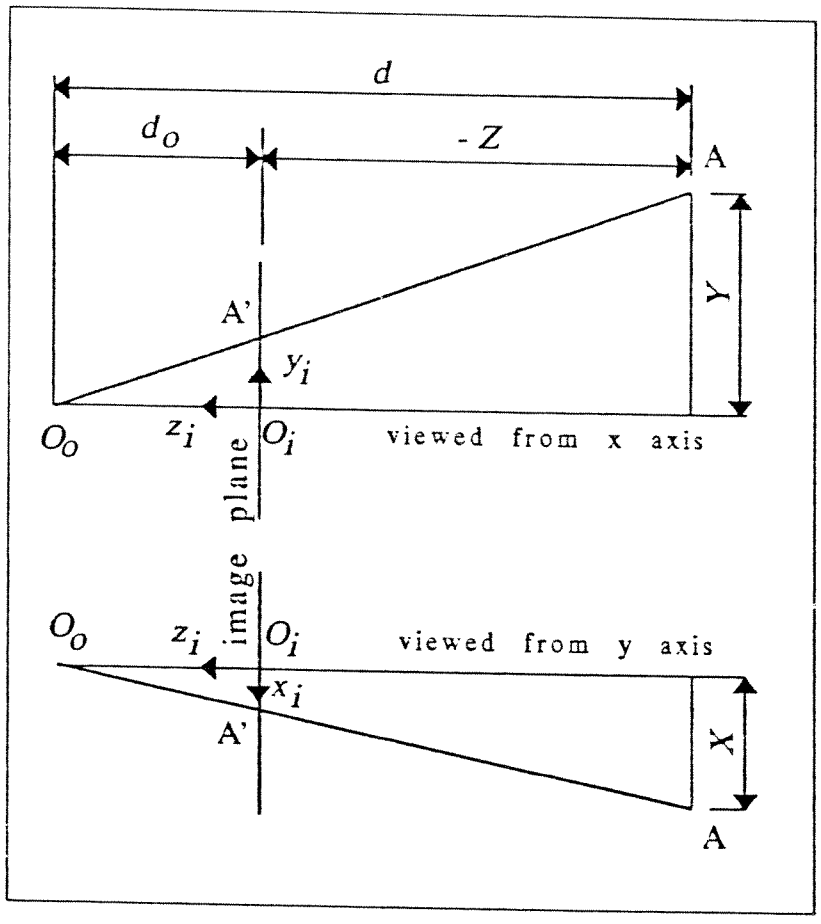


Figure 2.8: Perspective Projection in Detail

coordinate system. By means of similar triangles rule, one can easily write

$$x_i = d_0 \left(\frac{X}{d_0 - Z} \right) \qquad y_i = d_0 \left(\frac{Y}{d_0 - Z} \right) \tag{2.7}$$

This is called the perspective transformation and is used to calculate the image coordinates (x_i, y_i) of a point. Note that the non-linearity of the perspective transformation comes from the fact that in both expressions the Z coordinate appears at the denominator .

An important property of optical systems is the perspective distortion. As opposed to lens distortion, the perspective distortion is inherently present in the idealized model. However, the perspective distortion is only a problem relative to the human eye. It results from the fact that the camera model and the human eye model do not have the same characteristics. Perspective distortion becomes more perceivable when objects are close to the camera but it does not deteriorate the information content of an image.

As a matter of fact, as d increases d_f should also be increased in order to obtain a large enough image of the object. This is known as telephoto in photography and is used to give the details of a small portion of the scene. Viewing a far object this way, the perspective projection will approach an orthographic one and the effect of perspective distortion will vanish. On the other hand, when d gets smaller, d_f should also be decreased in order to have a complete image of the object. This corresponds to a wide-angle lens in photography and is used to obtain an overall image of the scene. The name wide-angle is given due to the increased angular field of view. In the latter case, there will be a high perspective distortion. Figure 2.9 illustrates the image outcomes of these two different cases for a cubic object.

In the telephoto case, the object is far but a large image is obtained whereas in the wide-angle case the same object is nearer and an overall image is obtained. The size of the image is the same for both cases but in the wide-angle case the perspective distortion is higher. An important property of human eye is that its angular field of view is very large[9] but there is almost no distortion when close objects are viewed. This is due to the spherical shaped image sensing retina of the eye.

[9] Around 180°

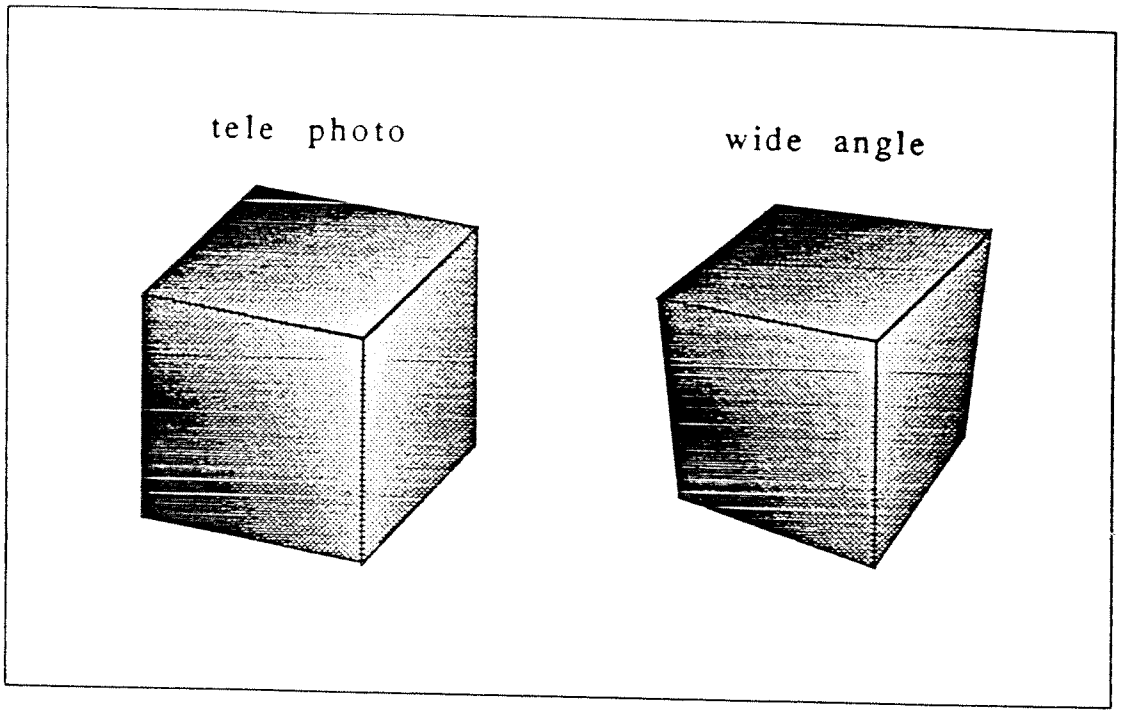


Figure 2.9: Perspective Distortion

The perspective distortion, although not desired in artistic imaging applications, is an important feature of the image and contains a hidden depth information in it.

The ideal camera model can be summarized as follows:

- the center of perspective is located at the optic center
- the principal distance d_0 is adjusted for perfect focusing
- the optical axis is perpendicular to the sensor
- the principal point and the geometric center of the sensor are coincident
- $\overline{O_c O_o} = \overline{O_o O_i} = d_0$
- the focal distance is taken as the supplied catalogue value
- all types of lens distortion are neglected

2.6 Image Processing

In this Section, the standard image processing techniques will be reviewed without supplying any computational details.

2.6.1 Image Enhancement

The image enhancement is to emphasize on some special features of an image for the purpose of analysis or display. Therefore, a rough image is processed using special techniques to obtain a more understandable image. It should be noted that the enhancement process does not increase the information content of the data but makes desired features more perceivable. Some examples are contrast and edge enhancement, noise filtering, sharpening and magnifying.

In the contrast stretching, each gray level is mapped into another gray level by a predefined transformation. It deals with illumination problems such as shadows, hot spots or low contrast images. An example is the histogram equalization where it is desired to obtain a uniform gray level distribution. This operation is especially useful for low contrast images. Other techniques use local neighborhood operations such as convolution masks, transform operations like the DFT[10] etc. Smoothing operations are used to reduce the noise and other undesired effects in images due to sampling, transmission and disturbances. The edge detection is another operation where the local changes in intensity are detected and used in transforming the original image both in x and y directions. The result is another image giving the edge information of the original image.

The lighting condition of the environment plays an important role in the quality of the image. Furthermore, particular lighting techniques called structured lighting can be used to obtain additional information about the shape and location of objects. With an appropriate structural lighting, the image processing operations get simplified. However, the control of the natural illumination of the environment is not an easy task to perform always.

[10]Discrete Fourier Transform

2.6.2 Image Restoration

The image restoration deals with the removal or minimization of known degradations in an image. This includes deblurring of images, noise filtering and correction of geometric distortions due to sensors. For example, the misfocus of a camera will result in a blurred image and due to the properties of the lens group there will be distortions in the image. Using restoration techniques one seeks the best estimate of the image as if the undesired disturbances were absent.

2.6.3 Image Analysis

Probably all the operations cited above have the common goal of supporting the image analysis. Because, in the image analysis, the image is described by making quantitative measurements. This task can be as simple as reading the item number of a product in a supermarket. In industry it can be used for sorting different parts on an assembly line or for inspection of parts. A more advanced task can be to generate the necessary information to control a robot. To proceed with the image analysis it is required to extract some features to identify the objects and then isolate them from the rest of the scene. These operations are called "identification" and "segmentation", respectively.

In robotics applications, generally it is desired to detect the position and orientation of an object subsequent to its recognition. The configuration of any object relative to a fixed frame in 3D space can be expressed by a homogeneous transformation matrix. The elements of the transformation matrix are functions of the six independent parameters which determine the configuration of the object in 3D.

The nature of the information obtained from an image analysis changes with different hardware selections. This implies that different methods for the analysis should be used according to the hardware configuration. In stereo vision, the images obtained from two separate cameras are matched and the range of the object is detected using triangulation techniques. The laser range finding technique uses a beam to detect the range of the objects. The dynamic vision

combines the image flow and the equation of motion of rigid bodies to trace their trajectories. The monocular vision which is the main interest of this study is based on the properties of perspective projection and uses only one camera for image sensing.

CHAPTER III

RECONSTRUCTION ALGORITHM

3.1 General

In this chapter an algorithm for the reconstruction of the 6 configuration parameters of a rigid body will be presented. The algorithm will be developed step by step. First the coordinate systems required for the solution will be defined then the algorithm developed by Platin et al.[8] will be explained. Finally the composite target will be introduced and the formulations of the new algorithm will be given.

In the solution procedure, a passive secondary target of known geometry will be used. Such a passive target can easily be mounted on the object of interest allowing to determine its configuration. The secondary target is designed to be planar, and the plane on which it lies will be called the "target plane". In conventional reconstruction algorithms, a predefined model is matched to the image of the object, subsequent to its recognition. Then this model is deformed such that it fits the image of the object. The configuration is obtained from the appearance of the deformed model. Such an approach requires an iterative solution and a database where the models are stored. The introduction of a secondary target, reduces the problem to the reconstruction of the target configurations in 3D, regardless of the object on which it is mounted. As a matter of fact this approach may also facilitate an object recognition process, since the object is "labeled" by the secondary target.

3.2 Coordinate Systems and Transformation Matrices

Figure 3.1 describes the cartesian coordinate systems that will be used in the formulations. In addition to the previous definitions there is a need

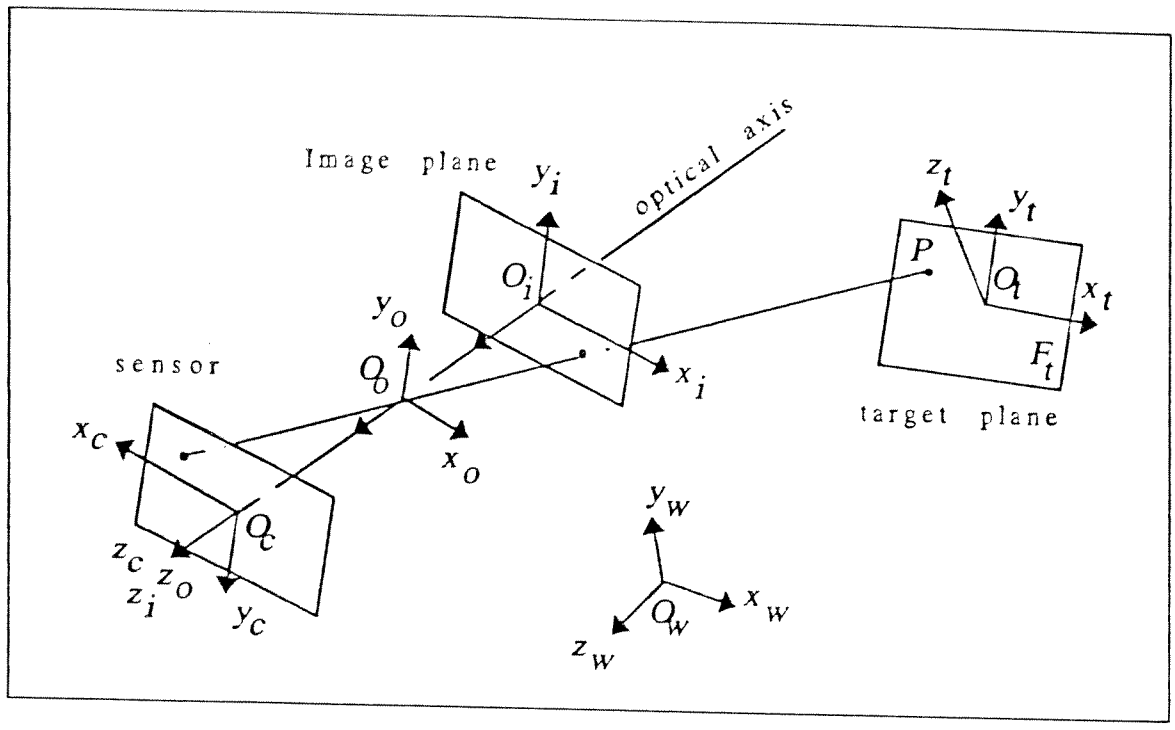


Figure 3.1: Coordinate Systems

for a fifth coordinate system to represent the target configurations. This new coordinate system $(O_t x_t y_t z_t)$ will be called the "target coordinate system" and the plane formed by the axes x_t and y_t will be the "target plane".

As observed in Figure 3.2 The world coordinates of any point P, lying on the target plane can be expressed by the following vector equation:

$$\vec{r}_{wp} = \vec{r}_{wt} + \vec{r}_{tp}$$

This vector equation can be more conveniently expressed using augmented matrices[1] as:

$$\bar{R}_{wp}^{(w)} = \hat{H}_{wt}^{(w,t)} \bar{R}_{tp}^{(t)}$$

It is obvious that the reconstruction will be completed once the matrix $\hat{H}_{wt}^{(w,t)}$ is determined. Using the properties of coordinate transformation, this matrix can be decomposed into three matrices representing consecutive transformations;

$$\hat{H}_{wt}^{(w,t)} = \hat{H}_{wc}^{(w,c)} \hat{H}_{ci}^{(c,i)} \hat{H}_{it}^{(i,t)} \tag{3.1}$$

where $\hat{H}_{wc}^{(w,c)}$ defines the configuration of the camera and is determined by camera calibration, $\hat{H}_{ci}^{(c,i)}$ defines the position of the image plane with respect to

[1]See Appendix A for details

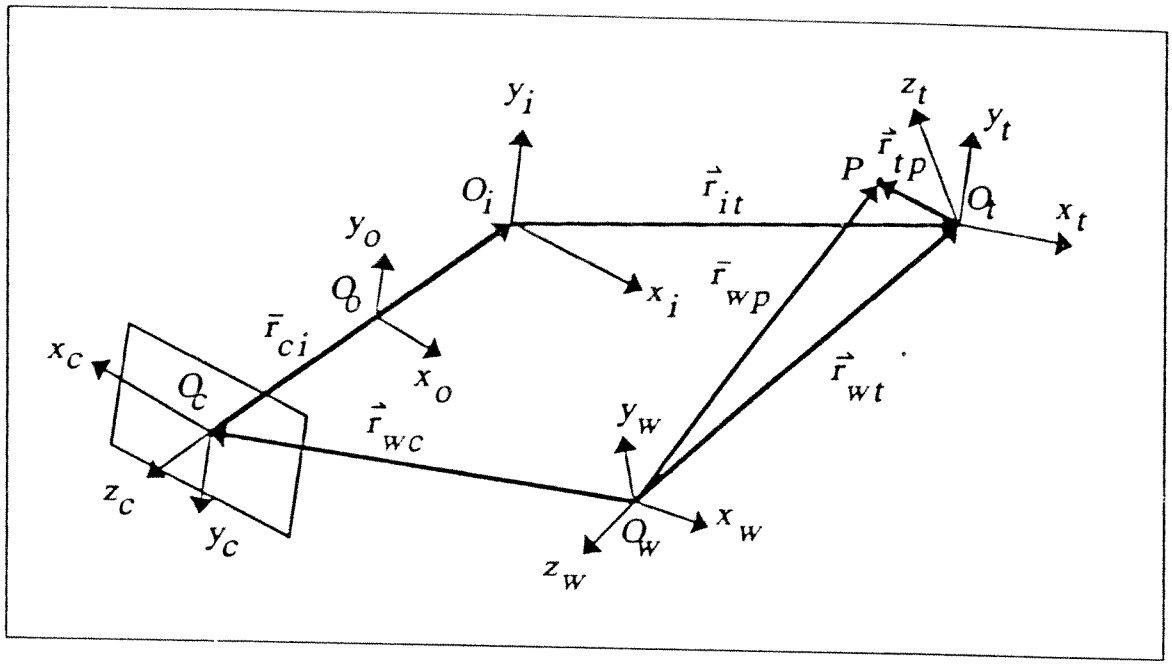


Figure 3.2: Vector Representation

the camera coordinate system and is known for a given lens setting. Therefore the solution will be obtained if the remaining matrix $\hat{H}_{it}^{(i,t)}$, can be determined. The unknown matrix itself can be decomposed to a series of basic rotation and translation matrices as follows:

$$\hat{H}_{it}^{(i,t)} = \hat{T}_z(d_0)\hat{R}_x(-\phi)\hat{R}_y(-\theta)\hat{T}_z(-d)\hat{R}_x(\alpha)\hat{R}_y(\beta)\hat{R}_z(\gamma) \tag{3.2}$$

where each basic transformation is selected properly to suit the reconstruction algorithm. Figure 3.3 illustrates the parameters used in each basic transformation matrix which are given as:

$$\hat{T}_z(d_0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\hat{R}_x(-\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & \sin\phi & 0 \\ 0 & -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\hat{R}_y(-\theta) = \begin{bmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

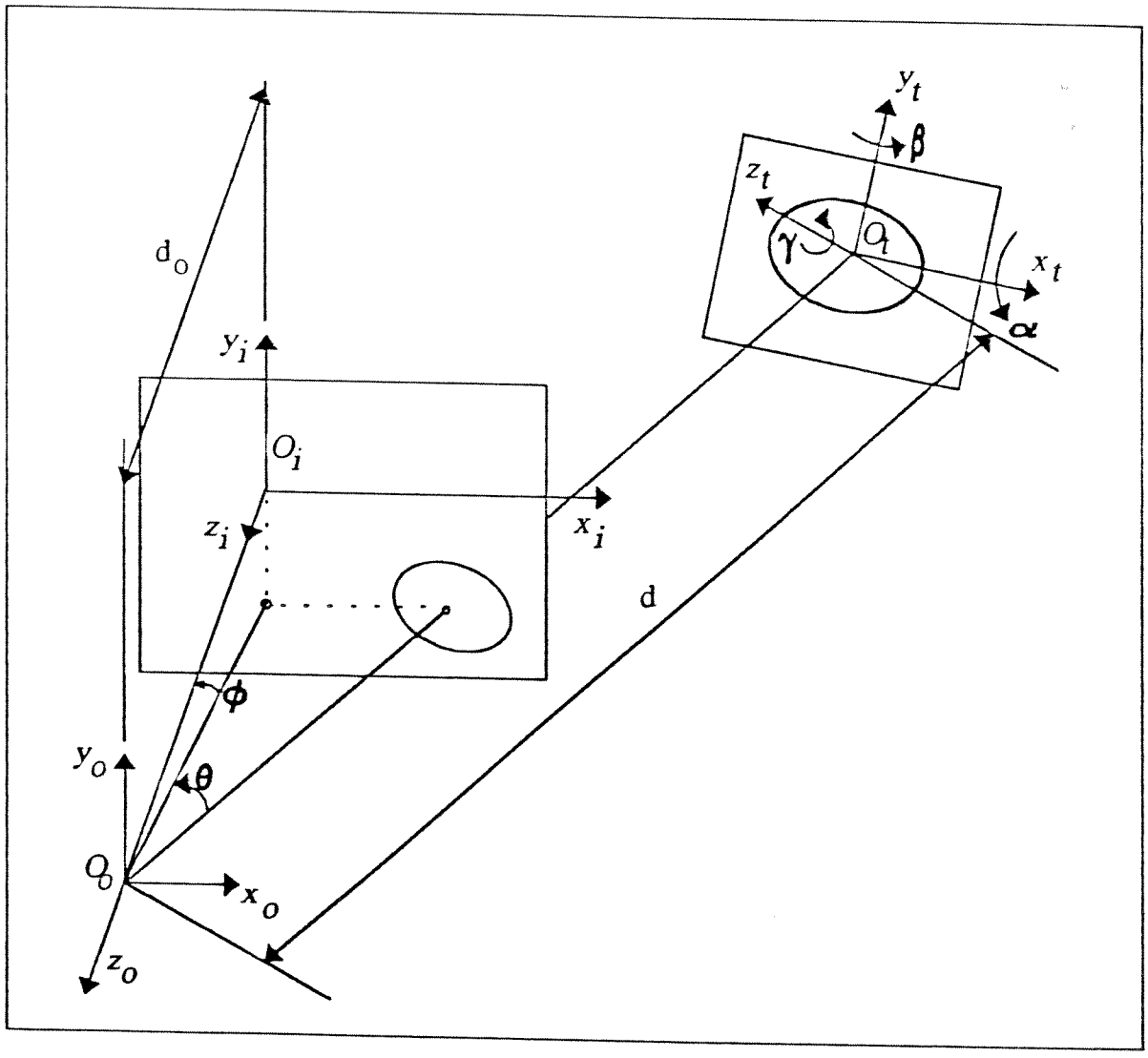


Figure 3.3: Reconstruction Parameters

$$\hat{T}_z(-d) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -d \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\hat{R}_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha & 0 \\ 0 & \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\hat{R}_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\hat{R}_z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 & 0 \\ \sin\gamma & \cos\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The purpose of the reconstruction algorithm is to compute the variables ϕ, θ, d, α, β and γ which determine the target configuration with respect to the image plane.

3.3 Target Centered Solution

In this section, the target centered solution which has been extracted from [8] will be explained and its formulation will be reviewed. The reason behind naming this solution as "target centered" is that it corresponds to the case where the center of the target is located on the optical axis of the camera.

In this early work, a circle of known radius has been used as a passive secondary target. According to Platin et al.[8] this choice is mainly due to two important facts:

> "The circle is one of the better known geometries with smooth analytical characteristics and the perspective projection of a circle offers analytically intriguing properties."

The target circle is positioned on the target coordinate system such that its center

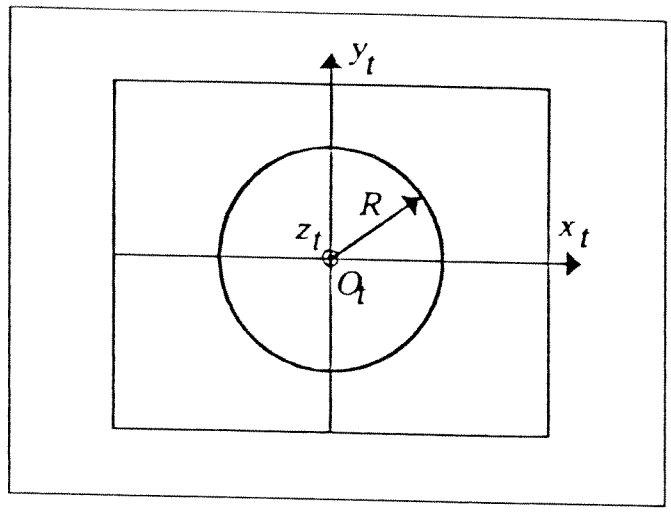


Figure 3.4: A Circular Target

is at O_t. Figure 3.4 depicts the target circle on the target plane where R denotes the radius.

The configuration variables are shown in Figure 3.5. Among the 3 parameters that define the position of O_t in 3D, only one parameter is independent, since the target center is constrained to lie on the optical axis of the camera. On the other hand, the in-plane rotation[2] of the target is left arbitrary because of the selected geometry. Thus the target configuration is defined by 3 parameters, which are the distance d from the perspective center to the target center and two consecutive rotations α and β of the target about axes x_t and y_t, respectively. Although the configuration is constrained and the DOF of the target is 3, the target centered solution will be used later as a tool in the reconstruction of the remaining parameters. Therefore, a good understanding of this model is required before making any attempt to determine the full target configuration.

With the constraints of the model, the transformation matrix $\hat{H}_{it}^{(i,t)}$ becomes:

$$\hat{H}_{it}^{(i,t)} = \hat{T}_z(d_0 - d)\hat{R}_x(\alpha)\hat{R}_y(\beta)$$

When each basic transformation matrix is substituted, the overall transformation matrix can be obtained as:

$$\hat{H}_{it}^{(i,t)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d_0 - d \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha & 0 \\ 0 & \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

[2]The rotation about z_t axis

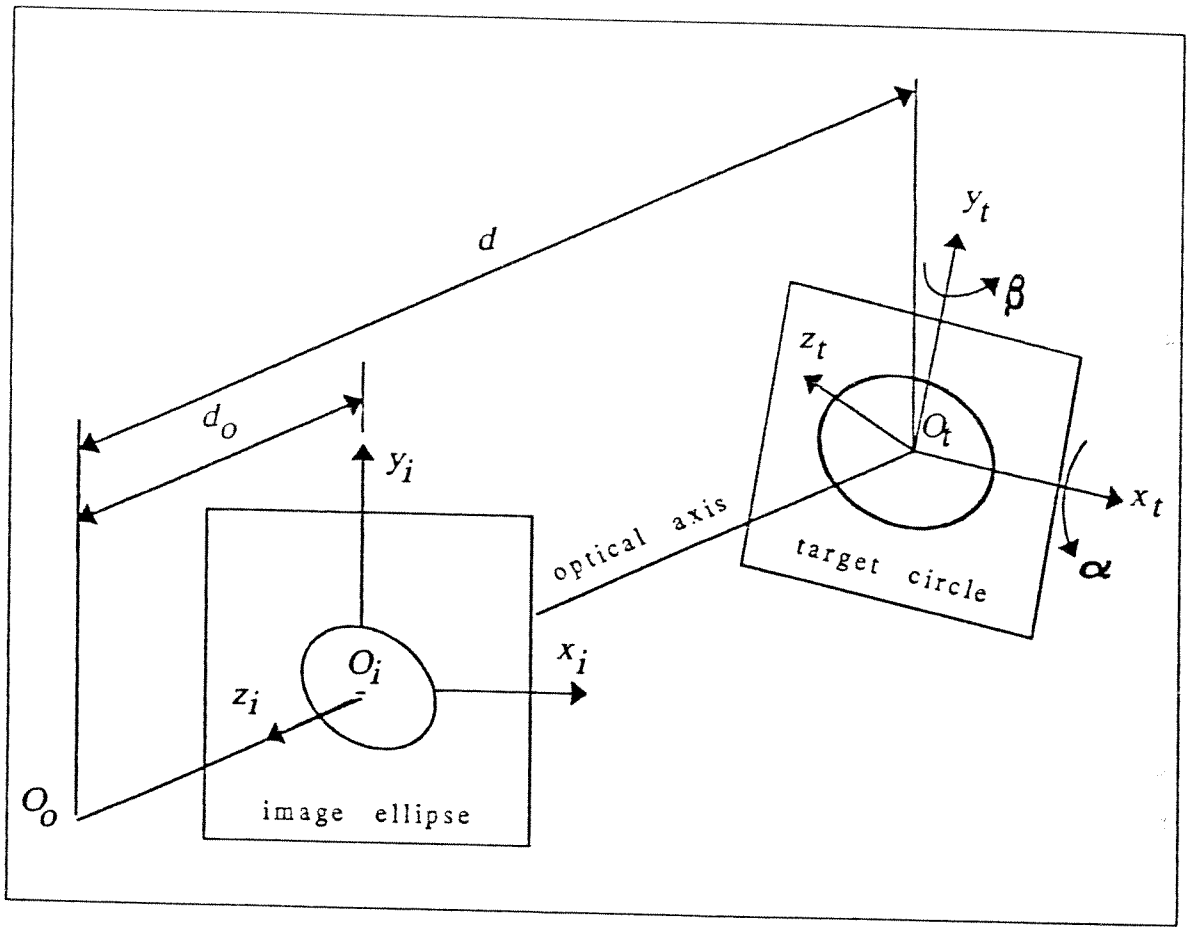


Figure 3.5: Reconstruction Parameters for the Target Centered Case

$$\hat{H}_{it}^{(i,t)} = \begin{bmatrix} \cos\beta & 0 & \sin\beta & 0 \\ \sin\alpha\sin\beta & \cos\alpha & -\sin\alpha\cos\beta & 0 \\ -\cos\alpha\sin\beta & \sin\alpha & \cos\alpha\cos\beta & d_0 - d \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The coordinates of any point lying on the target plane can be expressed in the image coordinate system as follows:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} = \hat{H}_{it}^{(i,t)} \begin{bmatrix} x_t \\ y_t \\ 0 \\ 1 \end{bmatrix}$$

Finally the image coordinates of the point can be calculated using the perspective transformation as:

$$\begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} (X_i d_0)/(d_0 - Z_i) \\ (Y_i d_0)/(d_0 - Z_i) \\ 0 \\ 1 \end{bmatrix}$$

Up to now, using the properties of the transformation matrices and the perspective projection, the image coordinates of a point which is located on the target plane is obtained analytically. However, in the solution process, the starting point is the image, and the target coordinates of an image point can be obtained by solving the equations for x_t and y_t. The solution gives:

$$\begin{aligned} x_t &= \frac{(d\cos\alpha)x_i}{(-\sin\beta)x_i + (\sin\alpha\cos\beta)y_i + (\cos\alpha\cos\beta)d_0} \\ y_t &= \frac{d[(-\sin\alpha\sin\beta)x_i + (\cos\beta)y_i]}{(-\sin\beta)x_i + (\sin\alpha\cos\beta)y_i + (\cos\alpha\cos\beta)d_0} \end{aligned} \tag{3.3}$$

The secondary target being a circle, its mathematical representation on the target coordinate system is given by:

$$x_t^2 + y_t^2 = R^2 \tag{3.4}$$

where R is the radius of the target circle. From the invariance theorem[3] , it is clear that the image of a circle will be another curve of second degree. In the practical limits, the image should be a closed curve, so intuitively it will be an ellipse and will be called the "image ellipse" in the text.

[3]See Appendix C for details

Substituting Equation 3.3 in Equation 3.4, the normalized equation of the image ellipse can be obtained as:

$$Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + 1 = 0 \tag{3.5}$$

where the normalized coefficients are:

$$\begin{aligned} A &= \frac{-[d^2 + (d^2 - R^2)\tan^2\beta/\cos^2\alpha]}{(Rd_0)^2} \\ B &= \frac{2(d^2 - R^2)\tan\alpha\tan\beta}{(Rd_0)^2\cos\alpha} \\ C &= \frac{-[d^2 + (d^2 - R^2)\tan^2\alpha]}{(Rd_0)^2} \\ D &= \frac{-2\tan\beta}{d_0\cos\alpha} \\ E &= \frac{2\tan\alpha}{d_0} \end{aligned} \tag{3.6}$$

The condition for the Equation 3.5 to represent an ellipse is $B^2 - 4AC < 0$. By substituting A, B and C this inequality can be written as:

$$\frac{d^2}{R^2} > 1 - \cos^2\alpha\cos^2\beta$$

To obtain the coefficients of the image ellipse, in the image coordinate system a curve-fitting operation must be performed for N pixel points selected along the contour of the image ellipse. The minimum number should be 5, since there are 5 normalized coefficients to be solved independently. This curve-fitting converts the visual information into mathematical data.

The curve-fitting can be realized by using the pseudo-inversion[4] technique which handles the inversion of non-square matrices. It should be stated that the pseudo-inversion technique is equivalent to the least squares solution when the number of constraint equations is greater than the number of unknowns. Suppose that the ellipse equation given by Equation 3.5 has to be satisfied for N selected pixels. This gives the following set of equations:

$$\begin{aligned} Ax_0^2 + Bx_0y_0 + Cy_0^2 + Dx_0 + Ey_0 + 1 &= 0 \\ Ax_1^2 + Bx_1y_1 + Cy_0^2 + Dx_1 + Ey_1 + 1 &= 0 \\ Ax_2^2 + Bx_2y_2 + Cy_2^2 + Dx_2 + Ey_2 + 1 &= 0 \end{aligned} \tag{3.7}$$

[4]See Appendix D for details

$$\begin{aligned} \ldots &= 0 \\ Ax_N^2 + Bx_Ny_N + Cy_N^2 + Dx_N + Ey_N + 1 &= 0 \\ \text{for } N &\geq 5 \end{aligned}$$

The set given in Equation 3.7 can be expressed in matrix form as follows:

$$\hat{M} \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ . \\ . \\ . \\ -1 \end{bmatrix} \tag{3.8}$$

where $\hat{M}$ is an $N \times 5$ matrix whose elements are function of selected pixel coordinates:

$$\hat{M} = \begin{bmatrix} x_0^2 & x_0y_0 & y_0^2 & x_0 & y_0 \\ x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 \\ x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 \\ . & . & . & . & . \\ . & . & . & . & . \\ x_N^2 & x_Ny_N & y_N^2 & x_N & y_N \end{bmatrix}$$

It is clear that a conventional solution can be obtained only if $[\hat{M}]^{-1}$ exists. However, $\hat{M}$ is not even a square matrix for $N > 5$ and hence its inverse is not defined. At this point, the concept of pseudo-inverse can be used to obtain the ellipse coefficients as:

$$\begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = [\hat{M}]^- \begin{bmatrix} -1 \\ -1 \\ -1 \\ . \\ . \\ . \\ -1 \end{bmatrix}$$

where $[\hat{M}]^-$ is a $5 \times N$ matrix and the '$-$' denotes the pseudo-inversion.

The reconstruction of unknown variables can be achieved if the Equation 3.6 is solved inversely. This means that given the coefficients of the image ellipse, obtained by curve-fitting, Equation 3.6 should be solved for d, α and β. With this statement, it seems that the solution has already been reached. Although this is true in theory, this task is not so easy considering two important practical facts:

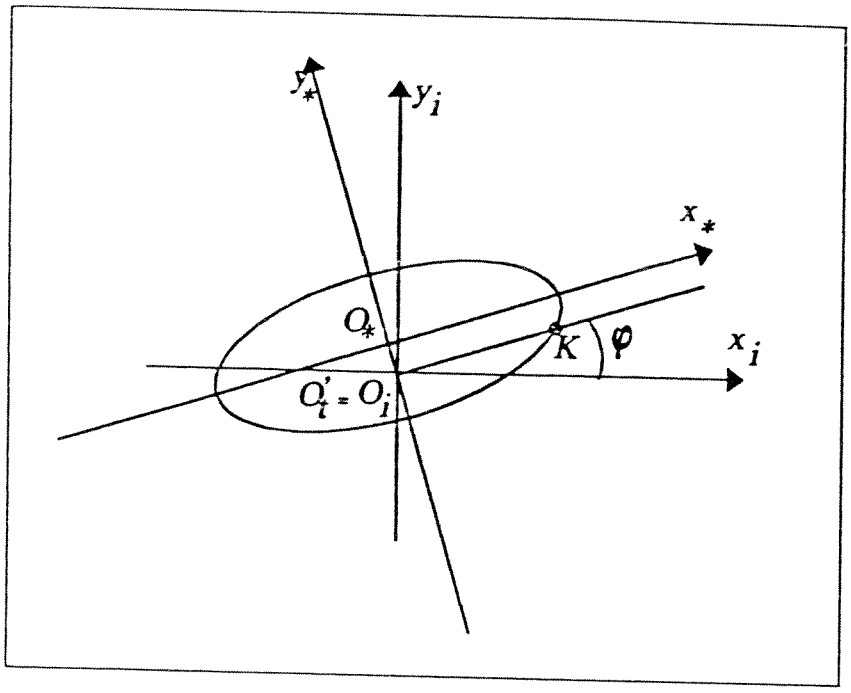


Figure 3.6: Image Ellipse

- the coefficients D and E in Equation 3.6 are very small in magnitude especially for α or β angles close 0^o and can induce numerical errors if any attempt is made to use them directly in the formulations.
- the equations are non-linear and contain complex trigonometric functions whose arguments are the unknown variables.

To get rid of these difficulties, a more elaborated solution based on the canonical representation of the image ellipse, including only the values of the coefficients A, B and C has been devised by Platin et al.[8].

By observing the image ellipse, one can see that the center of the ellipse and the image of the true center of the circle are not coincident. This is an important fact resulting from perspective distortion. The uniqueness of the solution is geometrically justified by this property for the observed center would be shifted in the opposite direction for a "sister plane". The sister plane is the name given to a special target plane whose configuration would give the same image ellipse as the original if the camera model were orthographic. However the non-linear characteristics of perspective projection helps in this case to distinguish the real target plane from its sister plane. The image ellipse is shown in Figure 3.6 as observed on the image plane. The true center of the circle is O_i' and the center of the image ellipse is O_*. In this situation if the length of the chord O_iK can

be found analytically, the distance can be solved easily by using similar triangles rule. The approach to handle this situation by taking full advantage of the circular target requires some mathematical manipulations. First consider the equation of the ellipse given by Equation 3.5, this is a quadratic form and can be written in the following matrix form:

$$\begin{bmatrix} x_i & y_i & 1 \end{bmatrix} \begin{bmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0 \tag{3.9}$$

It is also possible to represent the ellipse in the canonical coordinates (x_*, y_*) as follows:

$$\begin{bmatrix} x_* & y_* & 1 \end{bmatrix} \begin{bmatrix} A_* & 0 & 0 \\ 0 & C_* & 0 \\ 0 & 0 & F_* \end{bmatrix} \begin{bmatrix} x_* \\ y_* \\ 1 \end{bmatrix} = 0 \tag{3.10}$$

Suppose that $\hat{C}^{(i,*)}$ is a coordinate transformation matrix such that:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \hat{C}^{(i,*)} \begin{bmatrix} x_* \\ y_* \\ 1 \end{bmatrix} \tag{3.11}$$

then combining Equation 3.9, 3.10, 3.11, yields to:

$$\begin{bmatrix} A_* & 0 & 0 \\ 0 & C_* & 0 \\ 0 & 0 & F_* \end{bmatrix} = \left[\hat{C}^{(i,*)}\right]^T \begin{bmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & 1 \end{bmatrix} \hat{C}^{(i,*)}$$

The diagonal elements can be obtained[5] as:

$$\begin{aligned} A_* &= \frac{A+C}{2} - \frac{\sqrt{(A-C)^2+B^2}}{2} \\ C_* &= \frac{A+C}{2} + \frac{\sqrt{(A-C)^2+B^2}}{2} \\ F_* &= \frac{Det\left[\hat{M}\right]}{AC-B^2/4} \end{aligned} \tag{3.12}$$

independent of $\hat{C}^{(i,*)}$. A particularly interesting property can be found by combining Equation 3.6 and 3.12 as follows:

$$C_* = -(d/Rd_0)^2$$

which makes more sense when expressed as:

$$\frac{R}{\sqrt{-1/C_*}} = \frac{d}{d_0} \tag{3.13}$$

[5]See Appendix B for details

where it is seen that $\sqrt{-1/C_*}$ is actually the chord O_iK of the ellipse corresponding to the true radius of the circle and the Equation 3.13 is nothing but the application of the similar triangles rule. Actually this fact can also be proved geometrically by intersecting the line $y_i = x_i \tan\psi$ and the ellipse, as shown in Figure 3.6, where $2\psi = \tan^{-1}[B/(A-C)]$.

Now the distance d can be solved independent of α and β as:

$$d = Rd_0\sqrt{-C_*}$$

Finally, using the A and C expressions in Equation 3.6 one can solve for α and β:

$$\alpha = \mathrm{sgn}[E]\tan^{-1}\left[\sqrt{\frac{-(CR^2d_0^2+d^2)}{d^2-R^2}}\right] \tag{3.14}$$

$$\beta = -\mathrm{sgn}[D]\tan^{-1}\left[\cos\alpha\sqrt{\frac{-(AR^2d_0^2+d^2)}{d^2-R^2}}\right] \tag{3.15}$$

where the signs of D and E are used to find a unique solution.

This completes the solution process for a target centered problem, where the target is a circle. The solution is unique for $\alpha, \beta \in [-90°, +90°]$ which covers the practical range.

3.4 The General Solution

Now the question is how the remaining configuration variables can be obtained. Even by adding one more parameter, the matrix expressions become too complicated for the solution procedure. For this reason a different approach to solve the remaining parameters has been presented in [9] where the concept of virtual image plane is introduced. The method is based on [8] and solves the problem by adding one more parameter which corresponds to the angle θ in Figure 3.3, with $\phi = 0$. The circular target is still used, but the solution procedure is modified. The modification does not enable the reconstruction of the additional parameter θ, but makes possible the reconstruction of others for $\theta \neq 0°$. In this case, the solution requires the explicit knowledge of θ which

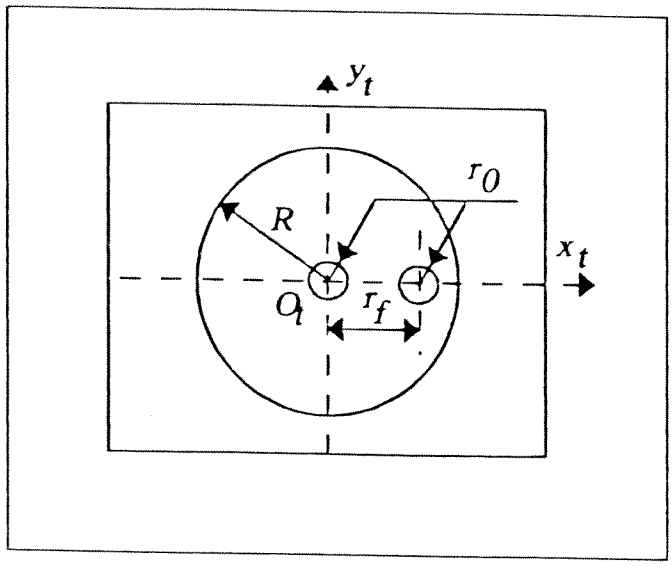


Figure 3.7: The Composite Target

determines the direction where the true center of the circle is located. In [9] no additional method has been devised to determine θ.

The new method that will be presented here is similar but includes the design of a new target, that will allow the determination of the direction where the target center is located. The target centered solution will be modified for the general case where the target has 6DOF. As opposed to [9] the method developed does not need the explicit knowledge of the remaining parameters but makes use of the properties of the new target to reconstruct them. The concept of virtual image plane presented in [9] has been used in a more general manner, and a second virtual image plane has been defined. As opposed to the first one, the second virtual image plane which is used to determine the in-plane rotation γ, cannot make use of the conventional perspective transformations.

3.4.1 Target Geometry

The new target geometry is shown in Figure 3.7. The circular part is preserved and two spots have been added. Both spots have radius r_0 as indicated in the figure. The target consisting actually of three separate geometries, will be called the "composite target".

The spot at the center is used to locate the true center of the circle on the image ellipse. It makes it possible the reconstruction of the variables ϕ and

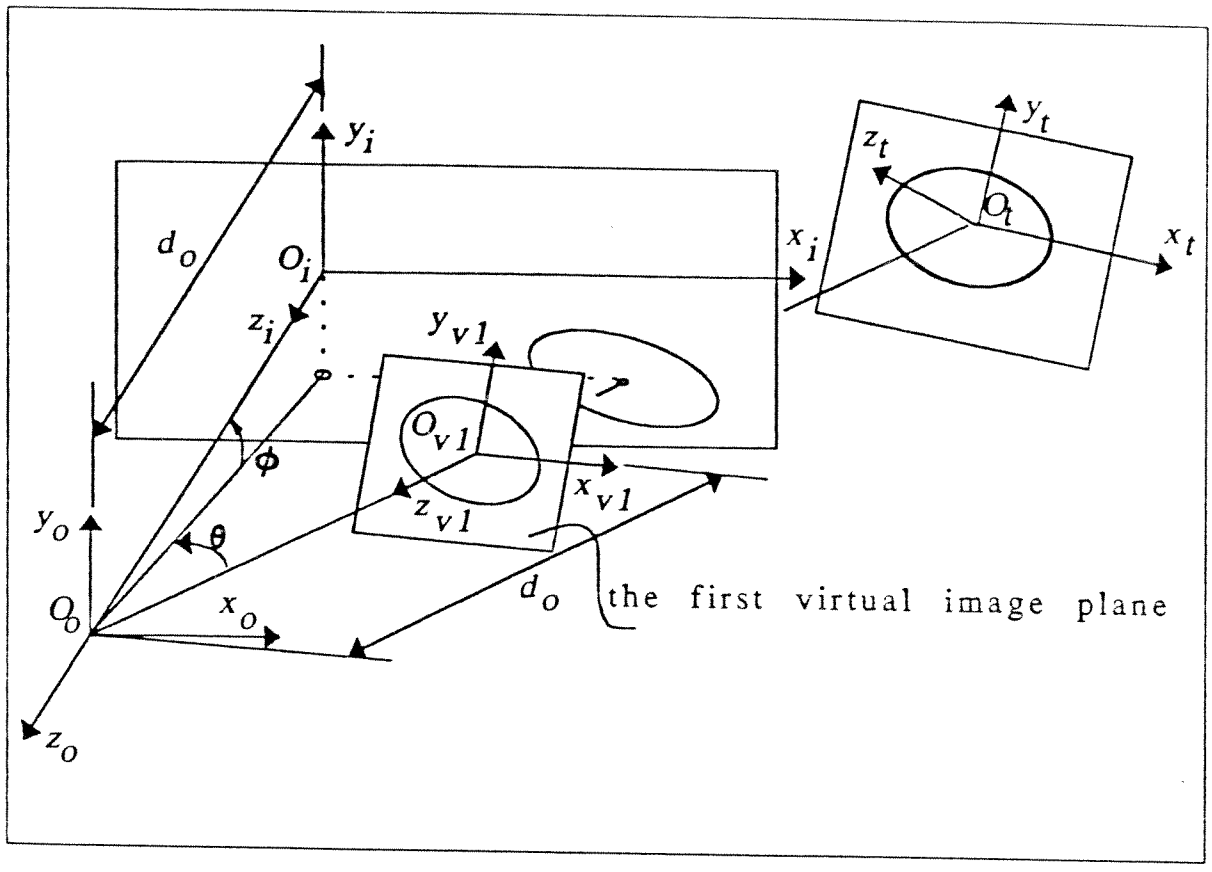


Figure 3.8: The First Virtual Image Plane

θ independent of others. Another spot is placed on the x_t axis, r_f away from the center to be able to determine the in-plane rotation. The asymmetric appearance of the target prevents multiple solutions.

The key assumption is that the center of a spot measured from the image is at the true center of the spot. Accordingly both spots should be small enough to neglect the shift in their true and image centers. It will later be observed in the simulation results that this assumption does not produce large reconstruction errors.

3.4.2 Determination of ϕ and θ

The variables ϕ and θ which determine the direction where the target center is located, and the first virtual image plane are shown in Figure 3.8. A coordinate system is assigned to this virtual image plane and is denoted by $(O_{v1}x_{v1}y_{v1}z_{v1})$. The virtual image plane is positioned such that it represents the image plane of a fictitious camera whose optic axis crosses the target center O_t

as in the target centered solution. The purpose of defining such a plane is to use the outcomes of the target centered solution that has been discussed previously. The virtual image plane makes it possible to handle the problem in two steps, so that the difficulty arising from the complexity of the transformation matrix will not affect the solution.

As a matter of fact, the solution process starts with the determination of the angles ϕ and θ by using the image of the central spot located at O_t. With the two new parameters, the transformation matrix from the image plane to the target plane becomes:

$$\hat{H}_{it}^{(i,t)} = \hat{T}_z(d_0)\hat{R}_x(-\phi)\hat{R}_y(-\theta)\hat{T}_z(-d)\hat{R}_x(\alpha)\hat{R}_y(\beta)$$

When the basic transformation matrices are substituted, the elements of the matrix are given by:

$$\begin{aligned}
\hat{H}_{it}^{(i,t)}[1,1] &= \cos\theta\cos\beta + \sin\theta\cos\alpha\sin\beta \\
\hat{H}_{it}^{(i,t)}[1,2] &= -\sin\theta\sin\alpha \\
\hat{H}_{it}^{(i,t)}[1,3] &= \cos\theta\sin\beta - \sin\theta\cos\alpha\cos\beta \\
\hat{H}_{it}^{(i,t)}[1,4] &= d\sin\theta \\
\hat{H}_{it}^{(i,t)}[2,1] &= \sin\phi\sin\theta\cos\beta + \cos\phi\sin\alpha\sin\beta - \sin\phi\cos\theta\cos\alpha\sin\beta \\
\hat{H}_{it}^{(i,t)}[2,2] &= \cos\phi\cos\alpha + \sin\phi\cos\theta\sin\alpha \\
\hat{H}_{it}^{(i,t)}[2,3] &= \sin\phi\sin\theta\sin\beta - \cos\phi\sin\alpha\cos\beta + \sin\phi\cos\theta\cos\alpha\cos\beta \\
\hat{H}_{it}^{(i,t)}[2,4] &= -d\sin\phi\cos\theta \\
\hat{H}_{it}^{(i,t)}[3,1] &= \cos\phi\sin\theta\cos\beta - \sin\phi\sin\alpha\sin\beta - \cos\phi\cos\theta\cos\alpha\sin\beta \\
\hat{H}_{it}^{(i,t)}[3,2] &= -\sin\phi\cos\alpha + \cos\phi\cos\theta\sin\alpha \\
\hat{H}_{it}^{(i,t)}[3,3] &= \cos\phi\sin\theta\sin\beta + \sin\phi\sin\alpha\cos\beta + \cos\phi\cos\theta\cos\alpha\cos\beta \\
\hat{H}_{it}^{(i,t)}[3,4] &= d_0 - d\cos\phi\cos\theta \\
\hat{H}_{it}^{(i,t)}[4,1] &= 0 \\
\hat{H}_{it}^{(i,t)}[4,2] &= 0 \\
\hat{H}_{it}^{(i,t)}[4,3] &= 0 \\
\hat{H}_{it}^{(i,t)}[4,4] &= 1
\end{aligned}$$

As observed, the elements of the matrix contain complex trigonometric expressions. But there will be considerable simplifications in using these expressions

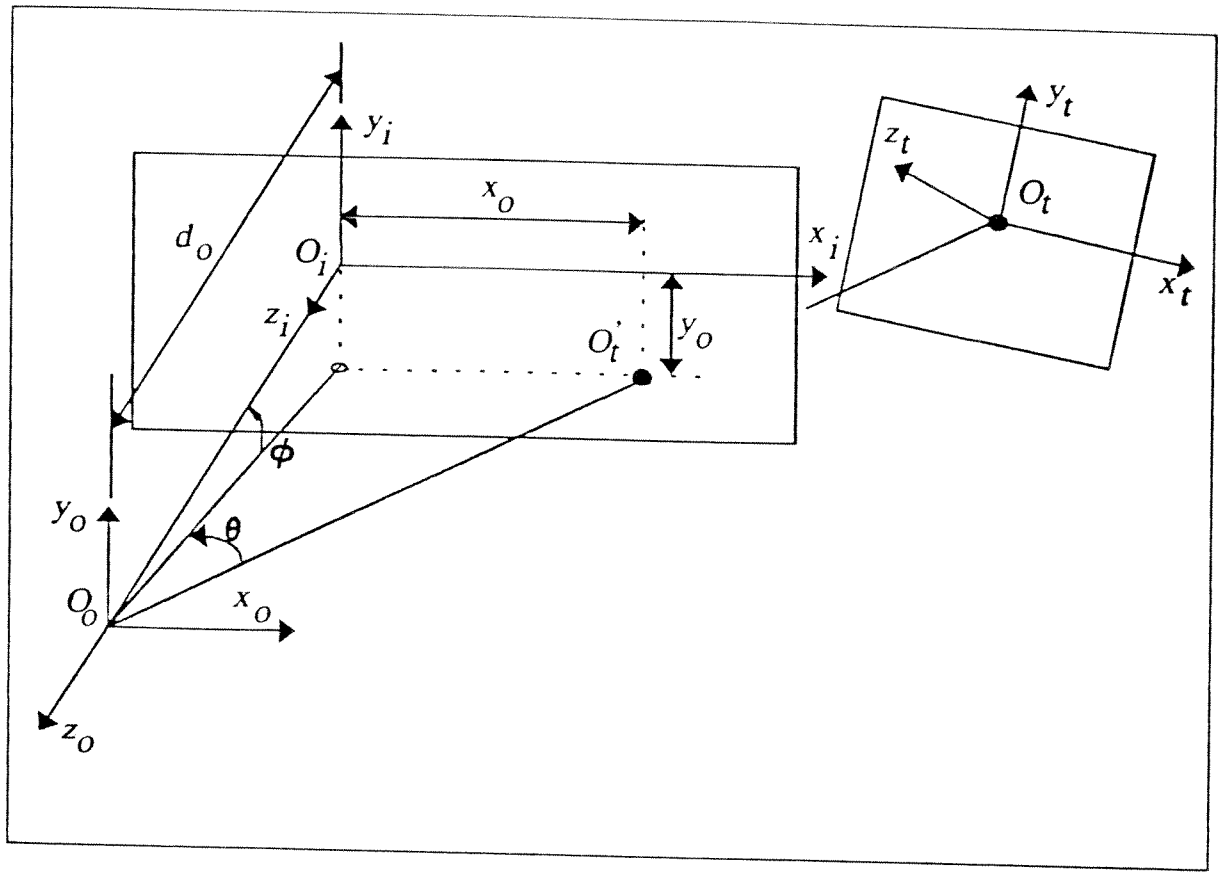


Figure 3.9: Calculation of ϕ and θ

since the purpose is to calculate only the image coordinates of the target center. Using the same techniques described in the target centered formulations, it is possible to relate the coordinates of the image and target planes as:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} = \hat{H}_{it}^{(i,t)} \begin{bmatrix} x_t \\ y_t \\ 0 \\ 1 \end{bmatrix}$$

where (X_i, Y_i, Z_i) are the coordinates of a point on the target plane with respect to the image coordinate system. The perspective transformation gives the coordinates of the image point as:

$$\begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} (X_i d_0)/(d_0 - Z_i) \\ (Y_i d_0)/(d_0 - Z_i) \\ 0 \\ 1 \end{bmatrix}$$

By making the substitution $x_t = 0$, $y_t = 0$ for the target center, the image coordinates (x_0, y_0) of O_t can be obtained easily in terms of the configuration variables. Figure 3.9 shows the geometric interpretation related to the formulations.

As expected, in the expressions

$$\begin{aligned} x_0 &= \frac{d_0 \tan\theta}{\cos\phi} \\ y_0 &= -d_0 \tan\phi \end{aligned}$$

the only unknowns are ϕ and θ since x_0 and y_0 can be measured using the centroid of the image of the central spot. Therefore these equations can be solved for ϕ and θ as:

$$\begin{aligned} \phi &= \tan^{-1}(-y_0/d_0) \\ \theta &= \tan^{-1}(x_0 \cos\phi/d_0) \end{aligned} \tag{3.16}$$

With the help of Equation 3.16 it is possible to determine the configuration of the first virtual image plane. Next, the following procedure is proposed to use the target centered solution:

- the coordinate transformation matrix from the image plane to the virtual image plane is determined
- a number of N points lying on the contour of the original image ellipse are projected onto the virtual image plane
- the curve-fitting is performed in the virtual image plane coordinates
- the resulting ellipse coefficients represent the image ellipse as if it were observed by the fictitious camera
- the target centered solution procedure is applied to the coefficients of the virtual image ellipse and α, β and d are determined with respect to the configuration of the virtual image plane

To project the selected pixel points, the transformation matrix from the image plane to the virtual image plane should be formed. This can be achieved without any difficulty since ϕ and θ have already been determined in Equation 3.16. From Figure 3.8 one can write:

$$\hat{H}_{v1i}^{(v1,i)} = \hat{T}_z(d_0)\hat{R}_y(\theta)\hat{R}_x(\phi)\hat{T}_z(-d_0)$$

When the basic transformation matrices are substituted, the matrix becomes:

$$\hat{H}_{v1i}^{(v1,i)} = \begin{bmatrix} \cos\theta & \sin\theta\sin\phi & \sin\theta\cos\phi & -d_0\sin\theta\cos\phi \\ 0 & \cos\phi & -\sin\phi & d_0\sin\phi \\ -\sin\theta & \cos\theta\sin\phi & \cos\theta\cos\phi & d_0(1-\cos\theta\cos\phi) \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.17}$$

Then the image and virtual image coordinates can be related by:

$$\begin{bmatrix} X_{v1} \\ Y_{v1} \\ Z_{v1} \\ 1 \end{bmatrix} = \hat{H}_{v1i}^{(v1,i)} \begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix}$$

where (X_{v1}, Y_{v1}, Z_{v1}) are the coordinates of an image point expressed in the virtual image coordinate system. Then the projection of the selected points on the virtual image plane can be achieved using perspective transformation:

$$\begin{bmatrix} x_{v1} \\ y_{v1} \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} (X_i d_0)/(d_0 - Z_i) \\ (Y_i d_0)/(d_0 - Z_i) \\ 0 \\ 1 \end{bmatrix}$$

Finally the coordinates of an image point projected onto the virtual image plane are obtained as:

$$\begin{aligned} x_{v1} &= \frac{d_0(x_i\cos\theta + y_i\sin\theta\sin\phi - d_0\sin\theta\cos\phi}{x_i\sin\theta - y_i\sin\phi\cos\theta + d_0\cos\theta\cos\phi} \\ y_{v1} &= \frac{d_0(y_i\cos\phi + d_0\sin\phi)}{x_i\sin\theta - y_i\sin\phi\cos\theta + d_0\cos\theta\cos\phi} \end{aligned} \tag{3.18}$$

3.4.3 Determination of γ, the In-plane Rotation

To determine the in-plane rotation of the target a second virtual image plane will be defined. This time the image of the outer spot that is located on the x_t axis in Figure 3.7 will be utilized.

First of all the transformation matrix from the image plane to the target plane should be formed. This matrix was already defined in Equation 3.2 and represents the general case where the target has 6DOF.

$$\hat{H}_{it}^{(i,t)} = \hat{T}_z(d_0)\hat{R}_x(-\phi)\hat{R}_y(-\theta)\hat{T}_z(-d)\hat{R}_x(\alpha)\hat{R}_y(\beta)\hat{R}_z(\gamma)$$

The elements of the matrix can be obtained after the basic transformation matrices are substituted:

$$
\begin{aligned}
\hat{H}_{it}^{(i,t)}[1,1] &= \cos\theta\cos\beta\cos\gamma + \sin\theta\cos\alpha\sin\beta\cos\gamma \\
&\quad - \sin\theta\sin\alpha\sin\gamma \\
\hat{H}_{it}^{(i,t)}[1,2] &= -\sin\gamma\cos\theta\cos\beta - \sin\gamma\sin\theta\cos\alpha\sin\beta \\
&\quad - \sin\theta\sin\alpha\cos\gamma \\
\hat{H}_{it}^{(i,t)}[1,3] &= \cos\theta\sin\beta - \sin\theta\cos\alpha\cos\beta \\
\hat{H}_{it}^{(i,t)}[1,4] &= d\sin\theta \\
\hat{H}_{it}^{(i,t)}[2,1] &= \sin\phi\sin\theta\cos\beta\cos\gamma + \cos\phi\sin\alpha\sin\beta\cos\gamma \\
&\quad - \sin\beta\cos\gamma\sin\phi\cos\theta\cos\alpha + \cos\phi\cos\alpha\sin\gamma \\
&\quad + \sin\phi\cos\theta\sin\alpha\sin\gamma \\
\hat{H}_{it}^{(i,t)}[2,2] &= -\sin\phi\sin\theta\cos\beta\sin\gamma - \cos\phi\sin\alpha\sin\beta\sin\gamma \\
&\quad + \sin\phi\cos\theta\cos\alpha\sin\beta\sin\gamma + \cos\phi\cos\alpha\cos\gamma \\
&\quad + \sin\phi\cos\theta\sin\alpha\cos\gamma \\
\hat{H}_{it}^{(i,t)}[2,3] &= \sin\phi\sin\theta\sin\beta - \cos\phi\sin\alpha\cos\beta \\
&\quad + \sin\phi\cos\theta\cos\alpha\cos\beta \\
\hat{H}_{it}^{(i,t)}[2,4] &= -d\sin\phi\cos\theta \\
\hat{H}_{it}^{(i,t)}[3,1] &= \cos\phi\sin\theta\cos\beta\cos\gamma - \sin\phi\sin\alpha\sin\beta\cos\gamma \\
&\quad - \cos\phi\cos\theta\cos\alpha\sin\beta\cos\gamma - \sin\phi\cos\alpha\sin\gamma \\
&\quad + \cos\phi\cos\theta\sin\alpha\sin\gamma \\
\hat{H}_{it}^{(i,t)}[3,2] &= -\cos\phi\sin\theta\cos\beta\sin\gamma + \sin\phi\sin\alpha\sin\beta\sin\gamma \\
&\quad + \cos\phi\cos\theta\cos\alpha\sin\beta\sin\gamma - \sin\phi\cos\alpha\cos\gamma \\
&\quad + \cos\phi\cos\theta\sin\alpha\cos\gamma \\
\hat{H}_{it}^{(i,t)}[3,3] &= \cos\phi\sin\theta\sin\beta + \sin\phi\sin\alpha\cos\beta \\
&\quad + \cos\phi\cos\theta\cos\alpha\cos\beta \\
\hat{H}_{it}^{(i,t)}[3,4] &= d_0 - d\cos\phi\cos\theta \\
\hat{H}_{it}^{(i,t)}[4,1] &= 0 \\
\hat{H}_{it}^{(i,t)}[4,2] &= 0 \\
\hat{H}_{it}^{(i,t)}[4,3] &= 0 \\
\hat{H}_{it}^{(i,t)}[4,4] &= 1
\end{aligned}
$$

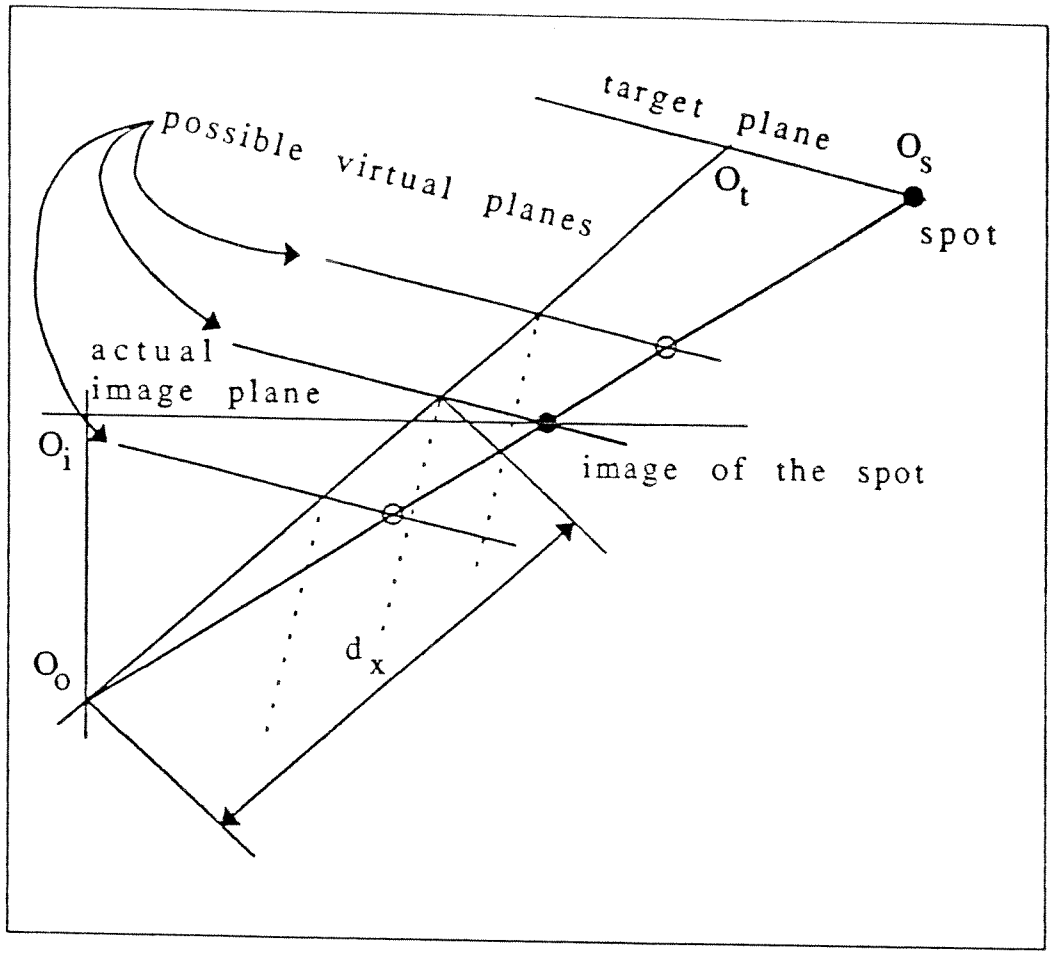


Figure 3.10: Possible Locations for the Second Virtual Plane

As observed, this matrix is too complicated to be directly used in the solution procedure. This is the reason why a second virtual image plane has been introduced. The selection of this virtual image plane is very important. To get the in-plane rotation properly:

- the second virtual image plane should be parallel to the target plane
- its origin should lie on the line O_oO_t

The unknown in-plane rotation can be obtained by a double argument arctangent function. Therefore the magnitude of the coordinates is not important as long as their ratio remains the same. As a consequence, when the two constraints cited above are satisfied, the position of the origin O_{v2} can be arbitrary along O_oO_t. Figure 3.10 depicts this situation in a simplified two dimensional form.

In Figure 3.10, the dashed lines represent the virtual optic axes. A problem that arises at this point is that the virtual optic axis of a virtual image

plane does not necessarily cross the optic center O_o. In a sense the second virtual image plane corresponds to a camera whose optic axis does not cross the optic center! However, the actual image of the spot is obtained using a perspective projection where O_o is the optic center. As a result, it is not possible to obtain the virtual image of a point by using the conventional perspective transformation techniques. In this case the virtual image of the outer spot will be where the line O_sO_o intercepts the virtual plane regardless of the virtual optic axis. To handle this situation, another approach has been devised. It consists of positioning the virtual image plane such that the center of the image spot lies on the intersection line between the actual image plane and the virtual image plane. This can be done because there is the freedom to select the position of the virtual image plane along O_oO_t. The question now is how to determine this particular location.

Suppose that this location is given by a translation along O_tO_o by an amount d_x. Then the transformation matrix can be formed as:

$$\hat{H}_{v2i}^{(v2,i)} = \hat{R}_y(-\beta)\hat{R}_x(-\alpha)\hat{T}_z(d_x)\hat{R}_y(\theta)\hat{R}_x(\phi)\hat{T}_z(-d_0)$$

The newly defined virtual plane is shown in Figure 3.11 and is denoted by $(O_{v2}x_{v2}y_{v2}z_{v2})$. As seen from the figure, the virtual image plane is parallel to the target plane and as a consequence z_t is parallel to z_{v2}.

When the basic transformation matrices are substituted the elements of the matrix can be obtained as:

$$\begin{aligned}
\hat{H}_{v2i}^{(v2,i)}[1,1] &= \cos\beta\cos\theta + \sin\beta\cos\alpha\sin\theta \\
\hat{H}_{v2i}^{(v2,i)}[1,2] &= \sin\beta\sin\alpha\cos\phi + \cos\beta\sin\theta\sin\phi \\
&\quad - \sin\beta\cos\alpha\cos\theta\sin\phi \\
\hat{H}_{v2i}^{(v2,i)}[1,3] &= -\sin\beta\sin\alpha\sin\phi + \cos\beta\sin\theta\cos\phi \\
&\quad - \sin\beta\cos\alpha\cos\theta\cos\phi \\
\hat{H}_{v2i}^{(v2,i)}[1,4] &= d_0\sin\beta\sin\alpha\sin\phi - d_0\cos\beta\sin\theta\cos\phi \\
&\quad + d_0\sin\beta\cos\alpha\cos\theta\cos\phi - d_x\sin\beta\cos\alpha \\
\hat{H}_{v2i}^{(v2,i)}[2,1] &= -\sin\alpha\sin\theta \\
\hat{H}_{v2i}^{(v2,i)}[2,2] &= \cos\alpha\cos\phi\sin\alpha\cos\theta\sin\phi \\
\hat{H}_{v2i}^{(v2,i)}[2,3] &= -\cos\alpha\sin\phi + \sin\alpha\cos\theta\cos\phi \\
\hat{H}_{v2i}^{(v2,i)}[2,4] &= d_0\cos\alpha\sin\phi - d_0\sin\alpha\cos\theta\cos\phi + d_x\sin\alpha
\end{aligned}$$

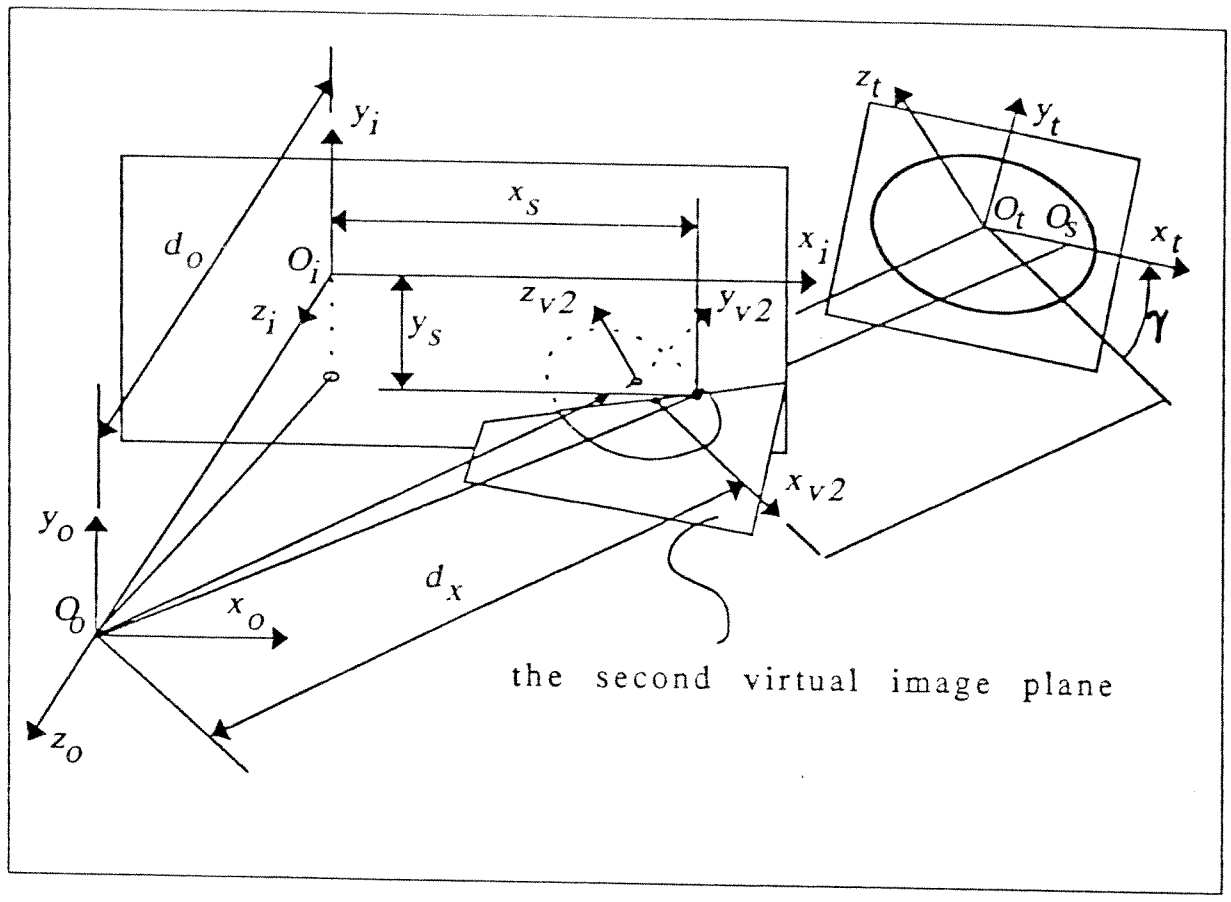


Figure 3.11: The Second Virtual Image Plane

$$
\begin{aligned}
\hat{H}_{v2i}^{(v2,i)}[3,1] &= \sin\beta\cos\theta - \cos\beta\cos\alpha\sin\theta \\
\hat{H}_{v2i}^{(v2,i)}[3,2] &= -\cos\beta\sin\alpha\cos\phi + \sin\beta\sin\theta\sin\phi \\
&\quad + \cos\beta\cos\alpha\cos\theta\sin\phi \\
\hat{H}_{v2i}^{(v2,i)}[3,3] &= \cos\beta\sin\alpha\sin\phi + \sin\beta\sin\theta\cos\phi \\
&\quad + \cos\beta\cos\alpha\cos\theta\cos\phi \\
\hat{H}_{v2i}^{(v2,i)}[3,4] &= -d_0\cos\beta\sin\alpha\sin\phi - d_0\sin\beta\sin\theta\cos\phi \\
&\quad - d_0\cos\beta\cos\alpha\cos\theta\cos\phi + d_x\cos\beta\cos\alpha \\
\hat{H}_{v2i}^{(v2,i)}[4,1] &= 0 \\
\hat{H}_{v2i}^{(v2,i)}[4,2] &= 0 \\
\hat{H}_{v2i}^{(v2,i)}[4,3] &= 0 \\
\hat{H}_{v2i}^{(v2,i)}[4,4] &= 1
\end{aligned}
$$

As usual, one can write the following matrix equation:

$$
\begin{bmatrix} X_{v2} \\ Y_{v2} \\ Z_{v2} \\ 1 \end{bmatrix} = \hat{H}_{v2i}^{(v2,i)} \begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix}
$$

The matrix $\hat{H}_{v2i}^{(v2,i)}$ can be represented by its elements e_{ij} where i denotes the row and j denotes the column:

$$\begin{bmatrix} X_{v2} \\ Y_{v2} \\ Z_{v2} \\ 1 \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & e_{13} & e_{14} \\ e_{21} & e_{22} & e_{23} & e_{24} \\ e_{31} & e_{32} & e_{33} & e_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix}$$

For the image center of the outer spot to lie on the second virtual image plane, the key point is to impose $Z_{v2} = 0$ explicitly and to solve for d_x using $x_i = x_s$ and $y_i = y_s$. This is equivalent to write:

$$[e_{31}\ e_{32}\ e_{33}\ e_{34}] \begin{bmatrix} x_s \\ y_s \\ 0 \\ 1 \end{bmatrix} = 0$$

Observing the elements e_{31}, e_{32} and e_{34}, it is seen that the only unknown is d_x since x_s and y_s can be determined using the centroid of the image of the spot. Therefore, d_x can be calculated as:

$$d_x = \frac{-(e_{31})x_s - (e_{32})y_s + (e_{34}^*)}{\cos\beta\cos\alpha}$$

where

$$\begin{aligned} e_{34}^* &= d_0\cos\beta\sin\alpha\sin\phi + d_0\sin\beta\sin\theta\cos\phi \\ &\quad + d_0\cos\beta\cos\alpha\cos\theta\cos\phi \end{aligned}$$

Now the virtual image coordinates of the center of the outer spot which are going to be denoted by x_s' and y_s' can be determined with the substitution of d_x in the expressions of e_{14} and e_{24};

$$\begin{aligned} x_s' &= (e_{11})x_s + (e_{12})y_s + (e_{14}) \\ y_s' &= (e_{21})x_s + (e_{22})y_s + (e_{24}) \end{aligned}$$

As the second virtual image plane and the target plane are parallel, γ can be calculated using a double argument arctangent function as follows:

$$\gamma = \tan_2^{-1}\left(\frac{y_s'}{x_s'}\right)$$

One important property of this method is that it does not recall the off-set distance of the spot denoted by r_f in Figure 3.7. Thus the in-plane rotation can be reconstructed just with the knowledge that the spot is located on the x_t axis.

3.5 Determination of $\hat{H}_{wt}^{(w,t)}$

Going back to Equation 3.1, $\hat{H}_{wt}^{(w,t)}$ can be written as:

$$\hat{H}_{wt}^{(w,t)} = \hat{H}_{wc}^{(w,c)} \hat{H}_{ci}^{(c,i)} \hat{H}_{it}^{(i,t)}$$

where $\hat{H}_{wc}^{(w,c)}$ is known from camera configuration.

$$\hat{H}_{ci}^{(c,i)} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -2d_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

can be determined for a given lens setting and $\hat{H}_{it}^{(i,t)}$ comes from the reconstruction algorithm.

Now $\hat{H}_{wt}^{(w,t)}$ will be decomposed to extract the configuration variables with respect to the world coordinate system. T_x, T_y and T_z will represent the translation components along x, y and z axes respectively. R_x, R_y and R_z will denote the consecutive rotation angles for an RFB[6]123 sequence.

$$\hat{H}_{wt}^{(w,t)} = \begin{bmatrix} \hat{C}^{(w,t)} & \bar{r}_{wt}^{(w)} \\ \bar{0}^T & 1 \end{bmatrix}$$

can be written using the definition of an homogeneous transformation matrix[7]. The translation components will be obtained easily using

$$\bar{r}_{wt}^{(w)} = \begin{bmatrix} r_1^w \\ r_2^w \\ r_3^w \end{bmatrix}$$

for $T_x = r_1^{(w)}$, $T_y = r_2^{(w)}$ and $T_z = r_3^{(w)}$. The rotation angles R_x, R_y and R_z can be obtained using the direction cosines of $\hat{C}^{(w,t)}$.

For $\cos(R_y) > 0$ this gives

$$\begin{aligned} R_x &= \tan^{-1}\left(\frac{-c_{23}}{c_{33}}\right) \\ R_y &= \sin^{-1}(c_{13}) \\ R_z &= \tan^{-1}\left(\frac{-c_{12}}{c_{11}}\right) \end{aligned}$$

[6]Rotated Frame Based
[7]See Appendix A for details

and for $\cos(R_y) < 0$

$$\begin{aligned} R_x &= \tan^{-1}(\frac{c_{23}}{-c_{33}}) \\ R_y &= \pi - \sin^{-1}(c_{13}) \\ R_z &= \tan^{-1}(\frac{c_{12}}{-c_{11}}) \end{aligned}$$

can be obtained. The case where $\cos(R_y) = 0$ is the singularity of an RFB 123 sequence and should be avoided.

CHAPTER IV

SIMULATION PROGRAM

In order to test the algorithm and to find out its theoretical accuracy limits, a computer program has been developed. The source code has been written in Turbo Pascal. The object oriented format of the program provides a flexible tool for future researches. The simulation consists of three main actions;

- generation of a binary target image for a given target and camera configuration
- reconstruction of configuration variables for a given solution setting
- storage of reconstruction errors for analysis

The error files are exported to Excel environment. Then with the help of a control panel it is possible to obtain hard copies of the error graphs which are automatically placed on a worksheet. A total of 36 graphs are available on 6 different worksheets for a single set of solution. This allows the study of the error on any variable as a function of any other variable. In all the error graphs the y axis is the error axis and the x axis is the variable axis.

One problem is that the error files obtained by running the simulation for an image set contain records in the following format which is not suitable for a clear analysis:

```
ConfigurationAndError = record
                          v:SixDegreesOfFreedom;
                          e:SixDegreesOfFreedom;
                        end;
```

where

```
SixDegreesOfFreedom=array[1..6] of double;
```

is defined beforehand and contains the configuration parameters in the following order for a coordinate system say "AnyCoordinate":

```
var AnyCoordinate:SixDegreesOfFreedom
    ...
begin
 AnyCoordinate[1]:=Tx;
 AnyCoordinate[2]:=Ty;
 AnyCoordinate[3]:=Tz;
 AnyCoordinate[4]:=Rx;
 AnyCoordinate[5]:=Ry;
 AnyCoordinate[6]:=Rz;
end;
```

Therefore, `ConfigurationAndError.v` contains the configuration information and `ConfigurationAndError.e` contains the relative error made during the reconstruction of 6 variables for the given configuration. This kind of data is rather rough especially when the properties of a Nested Set, described later, are considered. For this reason, a data processing algorithm has been developed to decompose the `ConfigurationAndError` format to separate text files for T_x, T_y, T_z, R_x, R_y, and R_z such that data points lying on the same vertical line for a given x axis value, are processed as a set, to obtain the minimum, average and maximum values on that vertical line. This way, the excessive information resulting from the combinatory behaviour of the Nested Sets is converted to a simpler and meaningful format. This time consuming operation is performed by the procedure `MinAveMax`. The `ConfigurationAndError` files are not destroyed and one can always use the "quick graphic" menu to plot all the data points on the screen. The "quick graphic" menu makes an auto-scaling to fit the graph in a predefined screen box.

The object hierarchy for the essential objects in the simulation program are shown in the following figures where the instances are shown inside rectangles.

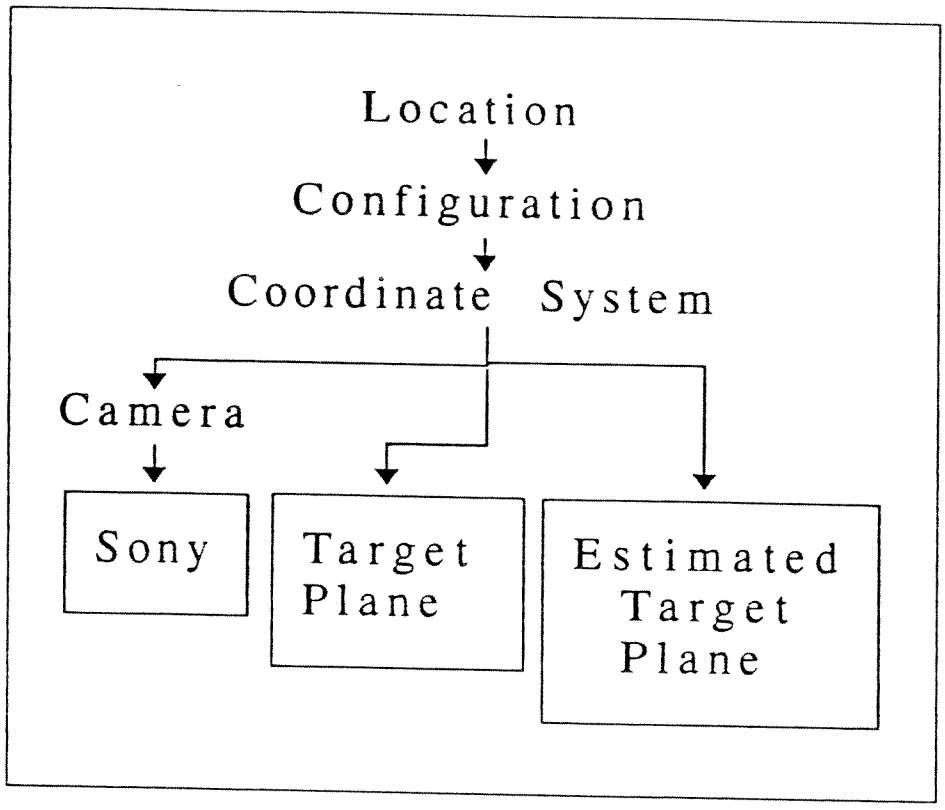


Figure 4.1: Object Hierarchy for Coordinate Systems

Figure 4.1 depicts the hierarchy for the different coordinate systems. Basically a Location contains the coordinates of a point in 3D space and Configuration has 3 additional variables for the rotations. Thus a Coordinate System inherits a location for its origin and an orientation according to an RFB 123 sequence. Its functions and procedures allow the determination of an overall transformation matrix by using the independent configuration variables and vice versa.

The Target Plane and the Estimated Target Plane are the two instances of Coordinate System. On the other hand, the Camera is designed as a descendant of a Coordinate System with some extra features like focusing, zooming etc. The Camera is able to detect the perfect focus value by calculating the target distance using its own functions and procedures.

Figure 4.2 shows the hierarchy for the perspective projection of geometric figures. This tree can be expanded in future to include different target geometries like lines, triangles or ellipses. Every distinct geometry has its own functions and procedures that will allow to represent it analytically on the image plane. All the descendants of Perspective Projection type are able to project

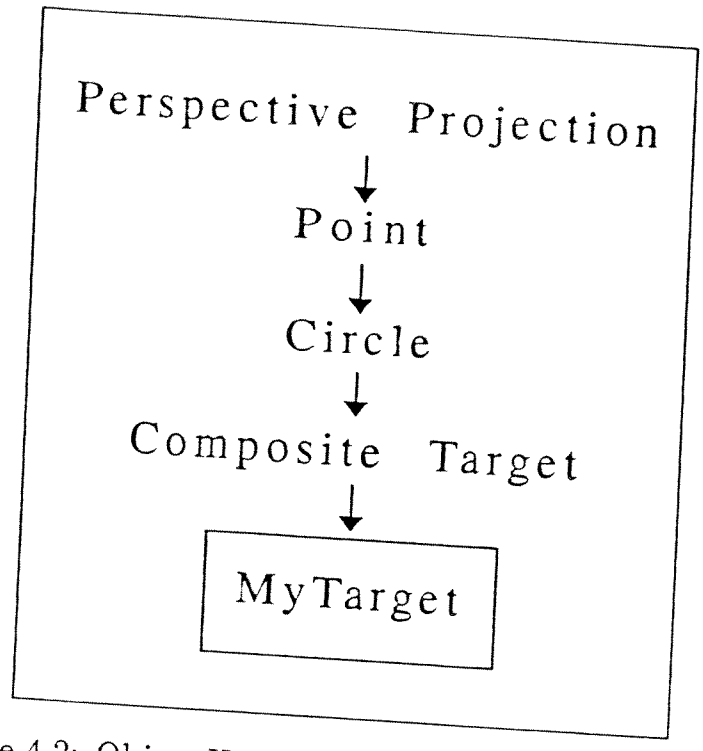


Figure 4.2: Object Hierarchy for Perspective Projection

themselves on a given image plane with a given image plane distance.

In Figure 4.3 it is observed that the image processing and solution depends on the sensor used. The Image Processor contains the algorithms required for the solution like finding the centroid of an area, contour following etc. whereas the Solver has the procedures directly related with the method of solution.

Figure 4.4 shows the smallest family of objects which is used to detect and store the reconstruction errors and later to export them in text format to be used by Excel.

The simulation program contains other procedures and functions for all kinds of mathematical operations related with the reconstruction algorithm. It has the capability to store several settings for the camera, target geometry and different solution types. The image sets can be created in two different ways;

- Nested Sets
- Sequential Sets

The first type consists of creating target images within a desired configuration range. This way all configuration variables are given a starting, ending

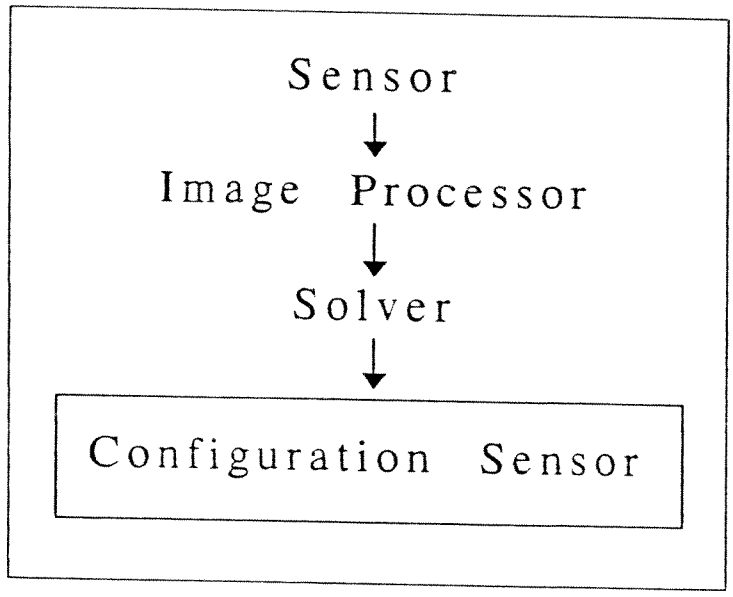


Figure 4.3: Object Hierarchy for the Configuration Sensor

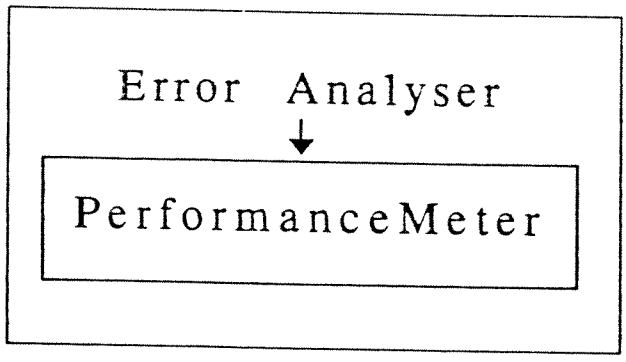


Figure 4.4: Object Hierarchy for PerformanceMeter

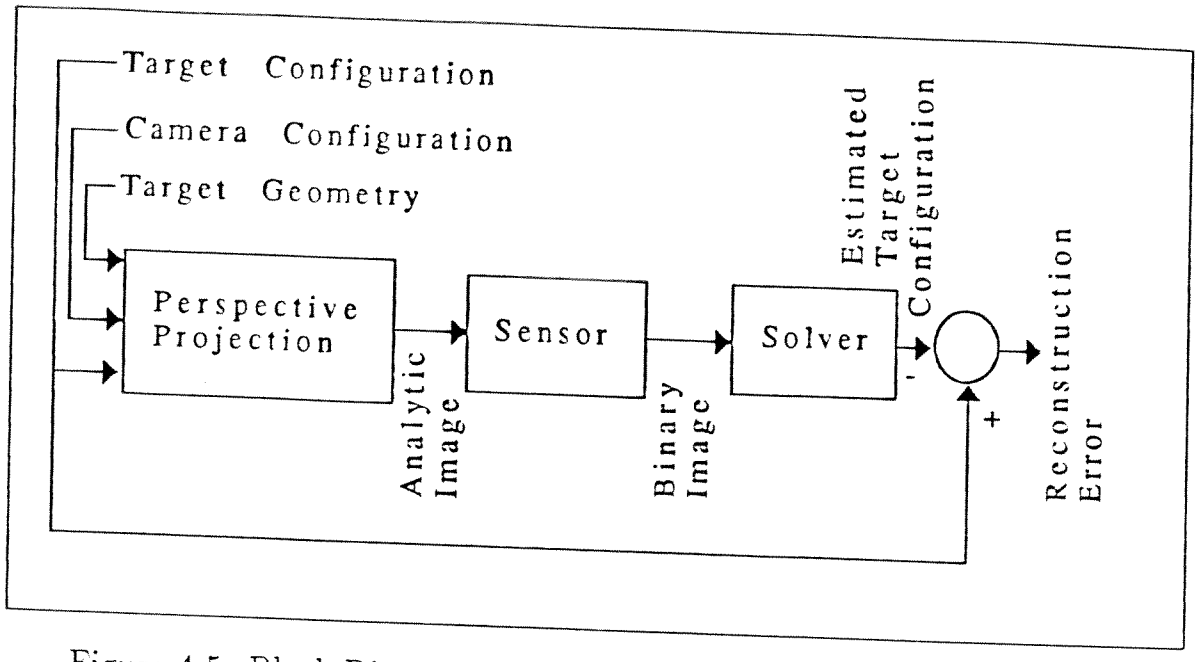


Figure 4.5: Block Diagram Representation of the Simulation Program

and a step value. The units for the position and rotation variables are milimeters and degrees, respectively. The nested loops generate target configurations for all the possible combinations within the starting and ending values with the specified interval. This type of image generation is used to study the performance of the system in a specific configuration range.

In the second type all configuration variables are given a starting value and a velocity. The velocity for the position variables is milimeters per second and for the rotation variables it is degrees per second. The image generation process runs for a specified time interval and during that time as many images as possible are created and solved depending on the speed limits of the hardware. This type of solution is the first step to motion tracking using the proposed algorithm. The number of images that will be generated is not known beforehand, but for reasonable time intervals, it is expected to be less than a Nested Set.

The operation of the simulation program can also be described using the block diagram shown in Figure 4.5.

CHAPTER V

PRACTICAL CONSIDERATIONS AND SIMULATION RESULTS

5.1 General

In order to prevent any confusion between the target configuration parameters and the reconstruction parameters used by the algorithm, the following important points should be kept in mind.

- T_x, T_y, T_z, R_x, R_y and R_z are the variables which determine the configuration of the target plane with respect to the world coordinate system
- ϕ, θ, d, α, β and γ are the reconstruction variables which determine the configuration of the target with respect to the image coordinate system.

For all the simulations that will be presented in this Chapter, the camera configuration is given by

$$\hat{H}_{wc}^{(w,c)} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which means that the camera coordinate system is located at the world coordinate system and rotated 180° about the z axis. This selection provides a better understanding and makes it easier to comment on the specific behaviour of error graphs. Moreover, when a target centered case is simulated such that $T_x = 0$ and $T_y = 0$, the following simplifications will occur:

$$\begin{aligned} \phi &= 0 \\ \theta &= 0 \\ d &= -T_z - d_0 \end{aligned}$$

$$\alpha = R_x$$
$$\beta = R_y$$
$$\gamma = R_z$$

Therefore it is recommended to keep the camera configuration as it is given above for the purpose of error analysis. The reason why the output of the reconstruction algorithm is converted to T_x, T_y, T_z, R_x, R_y, and R_z rather than using its own reconstruction variables is to develop a general approach to the problem. Once the proper working of the simulation program is assured, it can be used to study any specific case where the camera configuration may be different.

5.2 Contour Data from a Gray Level Image

The major difficulty in detecting the contour of the image ellipse is the determination of an appropriate threshold value Th. In cases where the image contrast is low, techniques such as histogram equalization or contrast stretching should be applied prior to binarization. A contour following algorithm is suitable to extract the ellipse contour using the binary image ellipse. If the threshold level is not selected properly, the reconstruction of the target distance will be most influenced by this.

In practice, the threshold level can be determined by making several measurements of the target's configuration using a test bench, under expected lighting conditions. For an 8 bit image, Th should be in the range

$$0 \leq Th \leq 255$$

where 0 and 255 correspond to the absolute black and absolute white, respectively. For a white target, if the perfect threshold is denoted by $Th^{(\text{perfect})}$, then;

- if $Th < Th^{(\text{perfect})}$ the contour will be expanded and the estimated target distance d will be smaller than the actual one
- if $Th > Th^{(\text{perfect})}$ the contour will be contracted and the estimated target distance d will be greater than the actual one

A special thresholding technique has not been developed in this study as the contour data is generated analytically using the exact coefficients of the image ellipse. This kind of practical problems will be studied as the continuation of this project by using the optical test bench to obtain experimental results.

5.3 Selection of Pixels for Curve-fitting

Previously, it has been mentioned that a total of $N \geq 5$ pixels are selected along the contour of the image ellipse to realize the curve-fitting process. However the number of the selected pixels, as well as the method of selection plays an important role in the reconstruction results. The error of reconstruction will be minimized if the selected pixels are close to the actual contour. For this reason a proper thresholding in binarization is necessary such that the contour data will be close to the analytical representation. Different methods have been developed for the purpose of pixel selection, and finally two of them were found to be useful and have been included in the simulation program as optional solution types:

1. MaxMinCross, $N = 8$
2. EqualArcsN, $5 \leq N \leq 700$

These methods do not take care of the errors due to image binarization, they consider the binary image as the exact digital representation of the image ellipse on the image plane. In the following sections, the two methods will be explained separately and simulation results will be compared.

5.3.1 Pixel Selection Using MaxMinCross Type

The MaxMinCross selection type has been designed to select 8 pixels along the contour as illustrated in Figure 5.1. The contour pixels and the theoretical image ellipse are shown and the selected pixels are marked.

The selection is achieved as follows: First the enclosing box is detected using the contour data. This operation consists of finding the maximum and

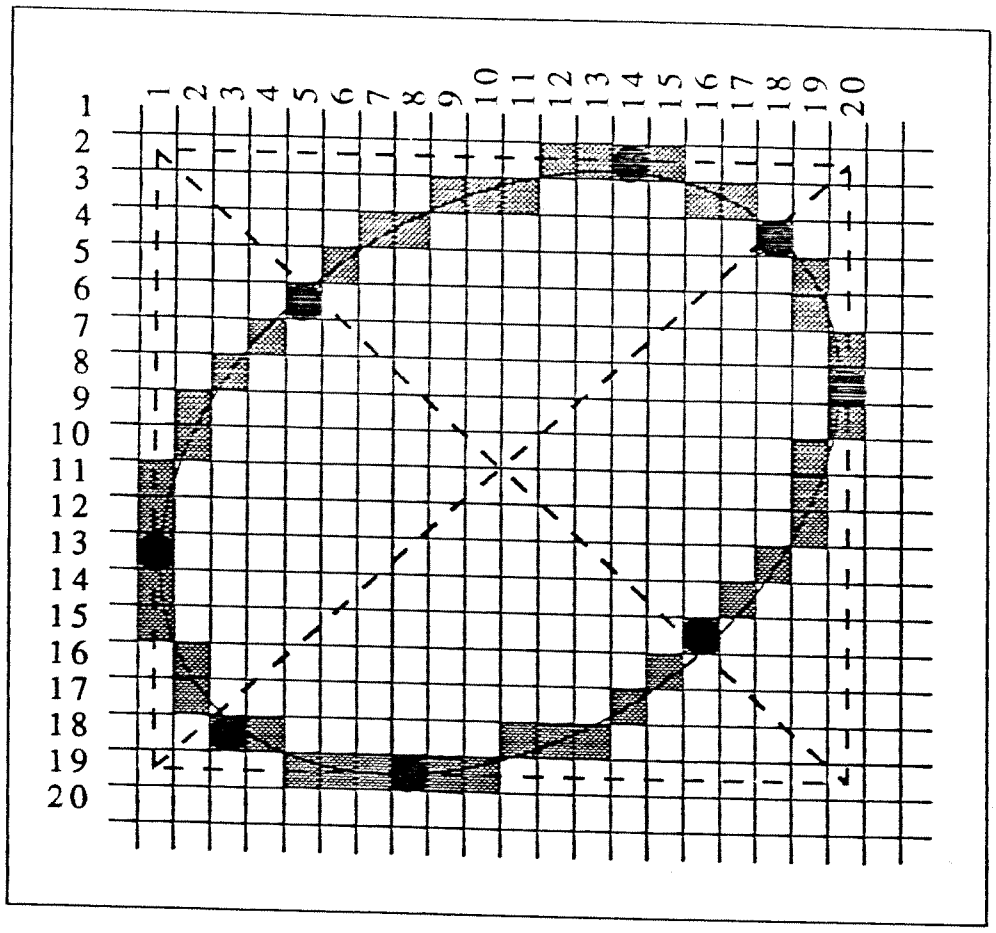


Figure 5.1: MaxMinCross Pixel Selection

minimum x and y values of the contour pixel coordinates. The enclosing box is shown in dashed lines. The first 4 pixels are selected at the regions where the digital curve intersects with the enclosing box. In general the edge of the box and the corresponding extreme portion of the contour have more than one pixels in common. These straight portions of the contour can be observed in the figure. The pixel located at the mid-point of the intersection is selected. In cases where the mid-point value turns out to be a sub-pixel point, its value is rounded and the adjacent pixel is selected. Consider the pixel at location $(14, 2)$, its y coordinate is given by the enclosing box as $y = 2$ and its x coordinate is calculated using $x = \text{round}[(12 + 15)/2]$.

Next the diagonals of the enclosing box are determined and the remaining 4 pixels are selected at the intersection of the diagonals and the digital curve. In here there is no need for averaging, the pixel lying on the contour whose center is closer to the diagonal is selected at each intersection.

The MaxMinCross method, emphasizes inherently the portions of the ellipse where the radius of curvature is small. Besides this, it gives an approximate *hidden* information about the major and minor axes of the image ellipse. The word 'hidden' is used since this information is not carried nor calculated numerically but the pixels selected in this manner represent these features in the curve-fitting operation. In cases where the image ellipse is too thin, sharp corners at the extreme portions of the contour may occur. Due to the property of the MaxMinCross selection which emphasizes the high curvatures, there is a possibility of selecting a pixel more than once in the regions where the curvature is high. The selection can be corrected by adding extra decision steps to the algorithm but then the philosophy underlying the MaxMinCross method would be deteriorated. The multiple selection of a pixel is an undesired fact as it may cause numerical overflows if the total number of distinct pixels selected becomes less than 5. In any case if a minimum number of 5 distinct pixels are available, the problem of multiple selection does not cause any problem for the curve-fitting operation.

Table 5.1: Camera Settings

Camera Setting	1	2
CameraT_x[mm]	0	0
CameraT_y[mm]	0	0
CameraT_z[mm]	0	0
CameraR_x[^{o}]	0	0
CameraR_y[^{o}]	0	0
CameraR_z[^{o}]	180	180
Pixel Size[μm $\times$ μm]	13×13	18×18
Resolution[$n \times m$]	400×400	300×300
d_f[mm]	28	28

5.3.2 Pixel Selection Using EqualArcs

To handle the extreme target configurations, a different approach has been developed instead of modifying the first one. As its name implies, the EqualArcs method consists of selecting N pixels along the contour, such that it is partitioned uniformly. The maximum number of pixels that can be selected is kept at 700 because of hardware capacity. This value is not so critical since in most cases the total number of contour pixels is less than 700. In the EqualArcs method, if the total number of contour pixels is less than N, then all the pixels are selected for curve-fitting. Whenever $N = 700$ is used, one is almost always sure to select all the contour pixels. This way, the maximum available information is gathered from the contour data at the expense of increased computation time. When all the contour pixels are selected, the relative computation speed turns out to be about two times slower than the MaxMinCross selection type, but the reconstruction for α and β around 0° and close to $\pm 80^\circ$ is achieved with less error.

5.3.3 Simulation Results

The Solution Settings and Camera Settings used for the simulations on pixel selection are given in Tables 5.1 and 5.2.

The Nested Set 1, described in Table 5.3 elaborates the region where the rotation angles α and β are close to 0°. The Nested Set 2 given in Table 5.4 is for the study of mid-range rotation angles. Finally Table 5.5 describes Nested

Table 5.2: Solution Settings

Solution Setting	1	2	3	4	5	6	7
Pixel Selection	MMC	EA	EA	EA	EA	EA	EA
N	8	10	100	700	700	700	700
Focus	P	P	P	P	C	I	C
$d_0^{(const)}$[mm]	0	0	0	0	28.5	0	30

MMC: MaxMinCross, EA: EqualArcs
P: Perfect, C: Constant, I: Iterative

Table 5.3: Nested Set 1

	Start	Stop	Step
T_x[mm]	0	0	0
T_y[mm]	0	0	0
T_z[mm]	-600	-600	0
R_x[°]	0	9	1
R_y[°]	0	9	1
R_z[°]	0	0	0

Set 3 intended for the extreme target orientations. The simulation results for the 3 sets have been combined in a single graph for each solution type.

In Figures 5.2 and 5.3 the R_x errors are compared for different pixel selection types. Figure 5.2 shows that both MaxMinCross and EqualArcs10 methods result in high reconstruction errors for highly rotated target configurations. This is due to the sign inversion of the coefficient E. In Figure 5.3 the reconstruction errors become considerably small. The EqualArcs700 method gives slightly better reconstruction results than EqualArcs100 at the expense of reduced computation speed.

Same observations are valid for Figures 5.4 and 5.5 this time for R_y.

Table 5.4: Nested Set 2

	Start	Stop	Step
T_x[mm]	0	0	0
T_y[mm]	0	0	0
T_z[mm]	-600	-600	0
R_x[°]	10	75	5
R_y[°]	10	75	5
R_z[°]	0	0	0

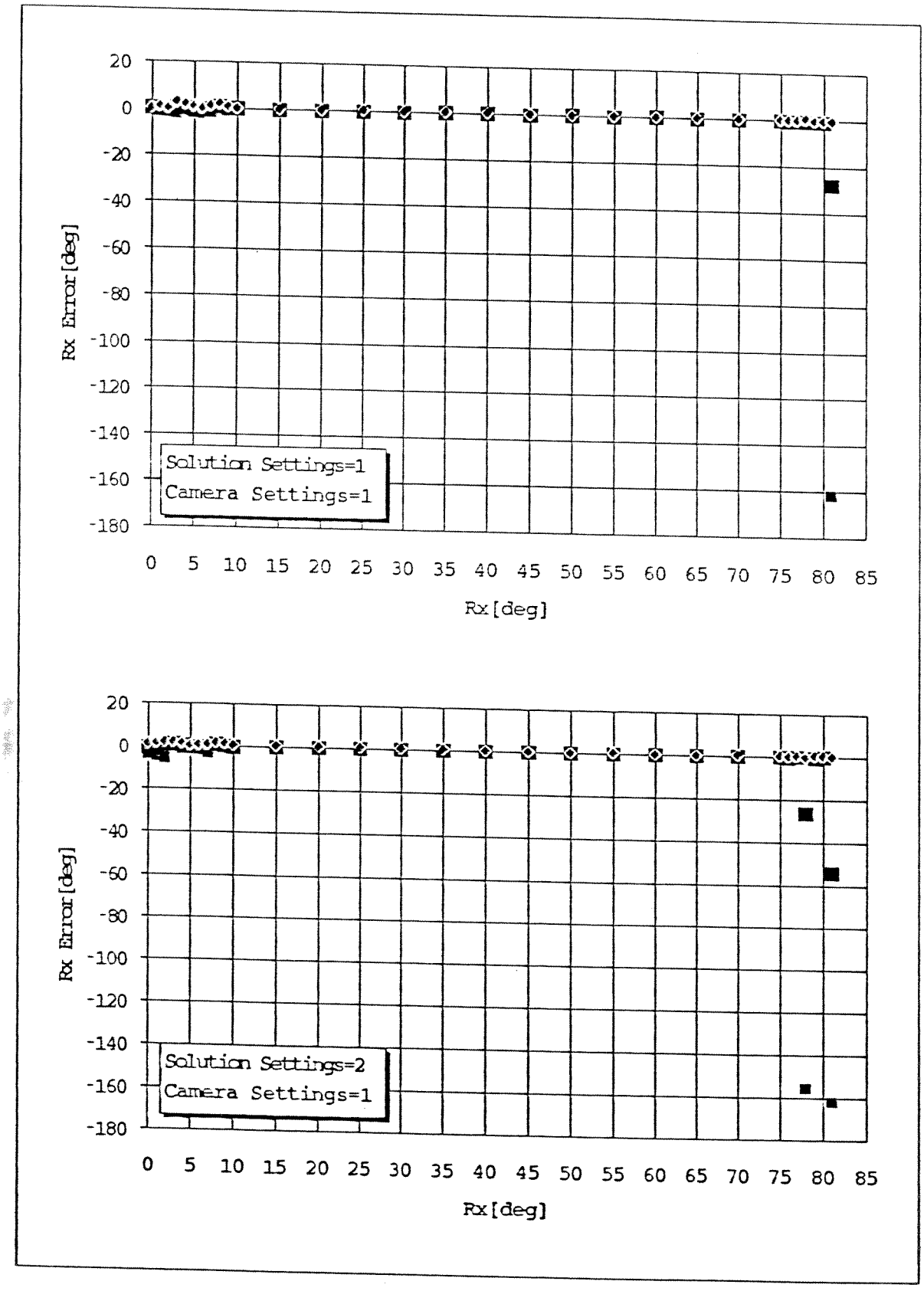


Figure 5.2: R_x Errors for MaxMinCross and EqualArcs10

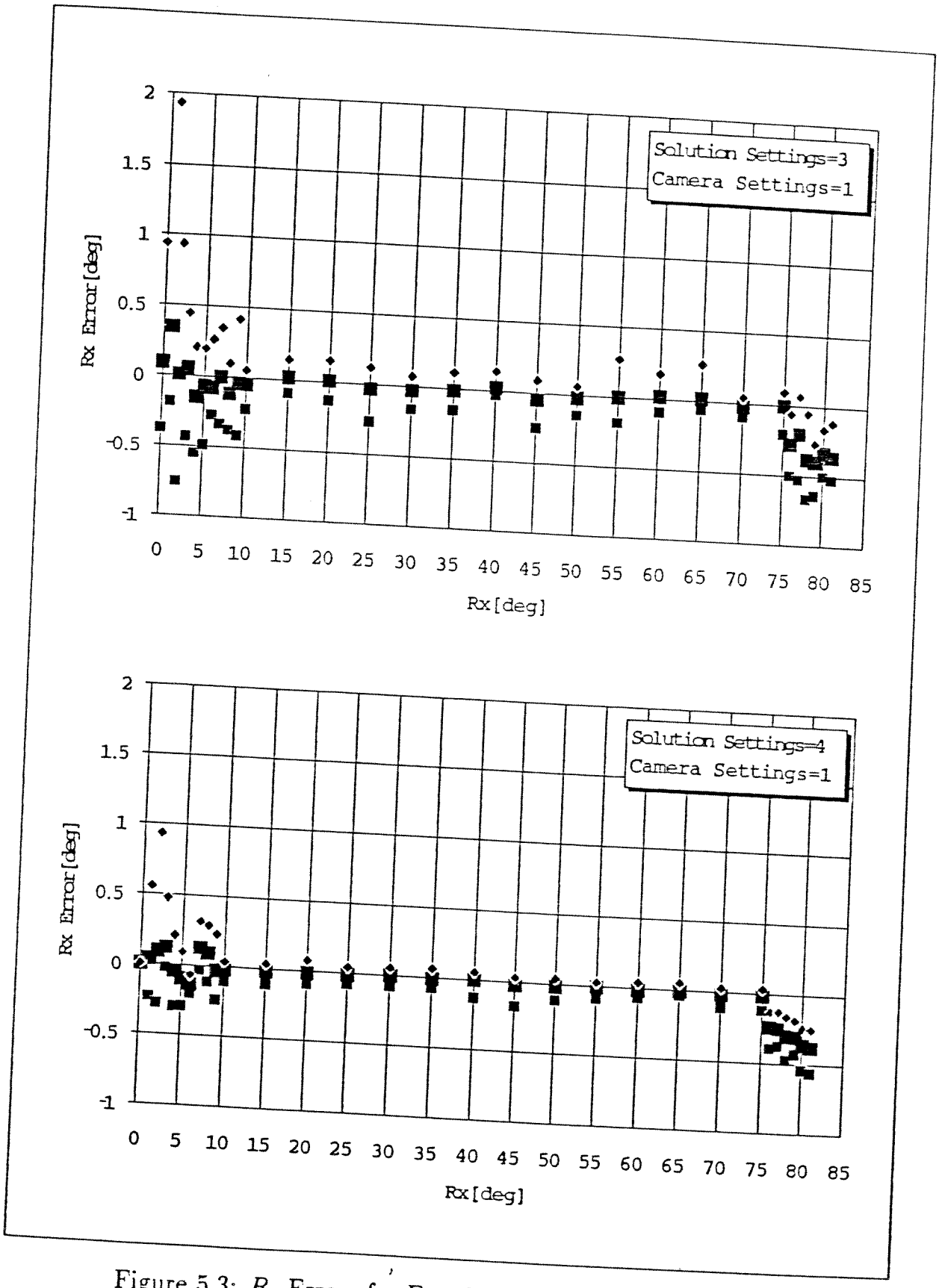


Figure 5.3: R_x Errors for EqualArcs100 and EqualArcs700

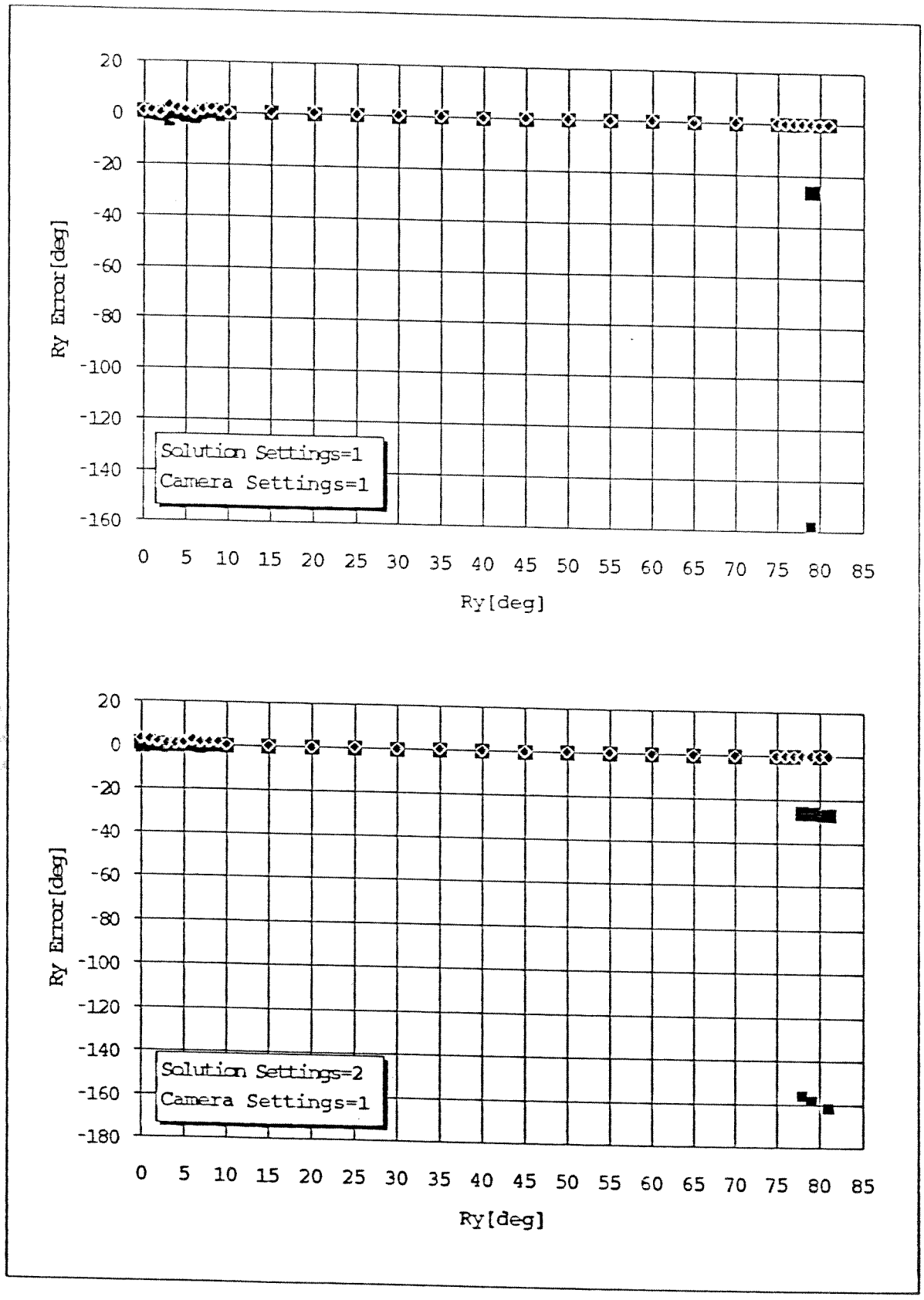


Figure 5.4: R_y Errors for MaxMinCross and EqualArcs10

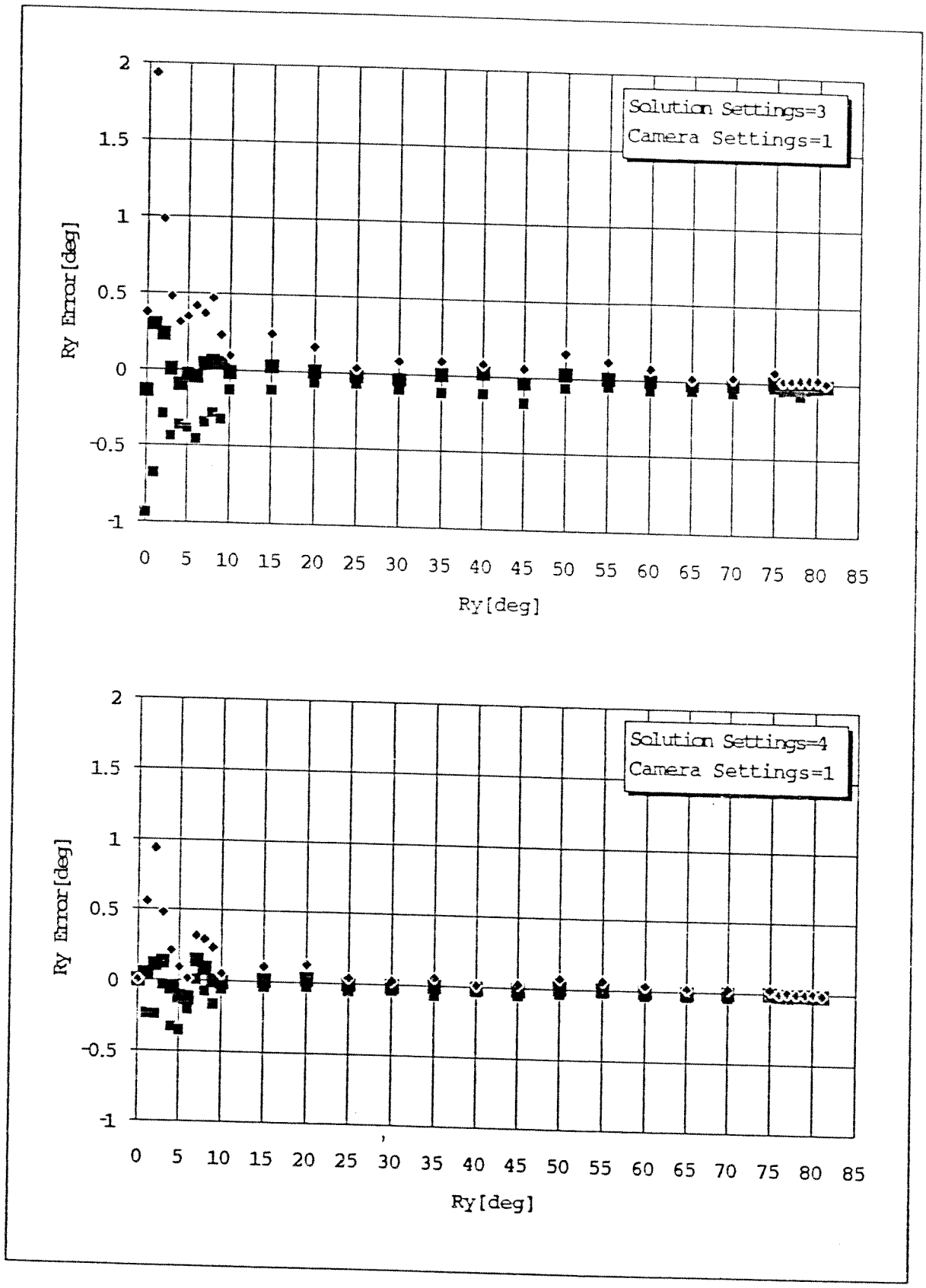


Figure 5.5: R_y Errors for EqualArcs100 and EqualArcs700

Table 5.5: Nested Set 3

	Start	Stop	Step
T_x[mm]	0	0	0
T_y[mm]	0	0	0
T_z[mm]	-600	-600	0
R_x[°]	76	81	1
R_y[°]	76	81	1
R_z[°]	0	0	0

Figure 5.6 however shows the most important property of the MaxMinCross selection method. In reconstructing T_z, the method shows better performance than EqualArcs10 even though the number of pixels selected is less. Actually, the property of MaxMinCross method to select pixels closer where the contour curvature is small, gives a better feeling of the ellipse's major axis and as a result, the target distance is reconstructed with less error.

The EqualArcs100 and EqualArcs700 methods shown in Figure 5.7 are very smooth in the near zero and mid range target orientations. For high rotations however, the MaxMinCross method still gives the best distance estimate.

One can conclude from the observations that the number of pixels selected for curve-fitting and the method of selection have different effects on the reconstruction of different variables. Where T_z is sensitive to the selection type, R_x and R_y are sensitive to the number of pixels used. The image sets used in the simulations for the pixel selection, have all $T_x = 0$, $T_y = 0$ and $R_z = 0$ corresponding to a target centered case. This choice is made because R_x, R_y and T_z are the only parameters directly influenced by the different pixel selection methods, with the given camera configuration.

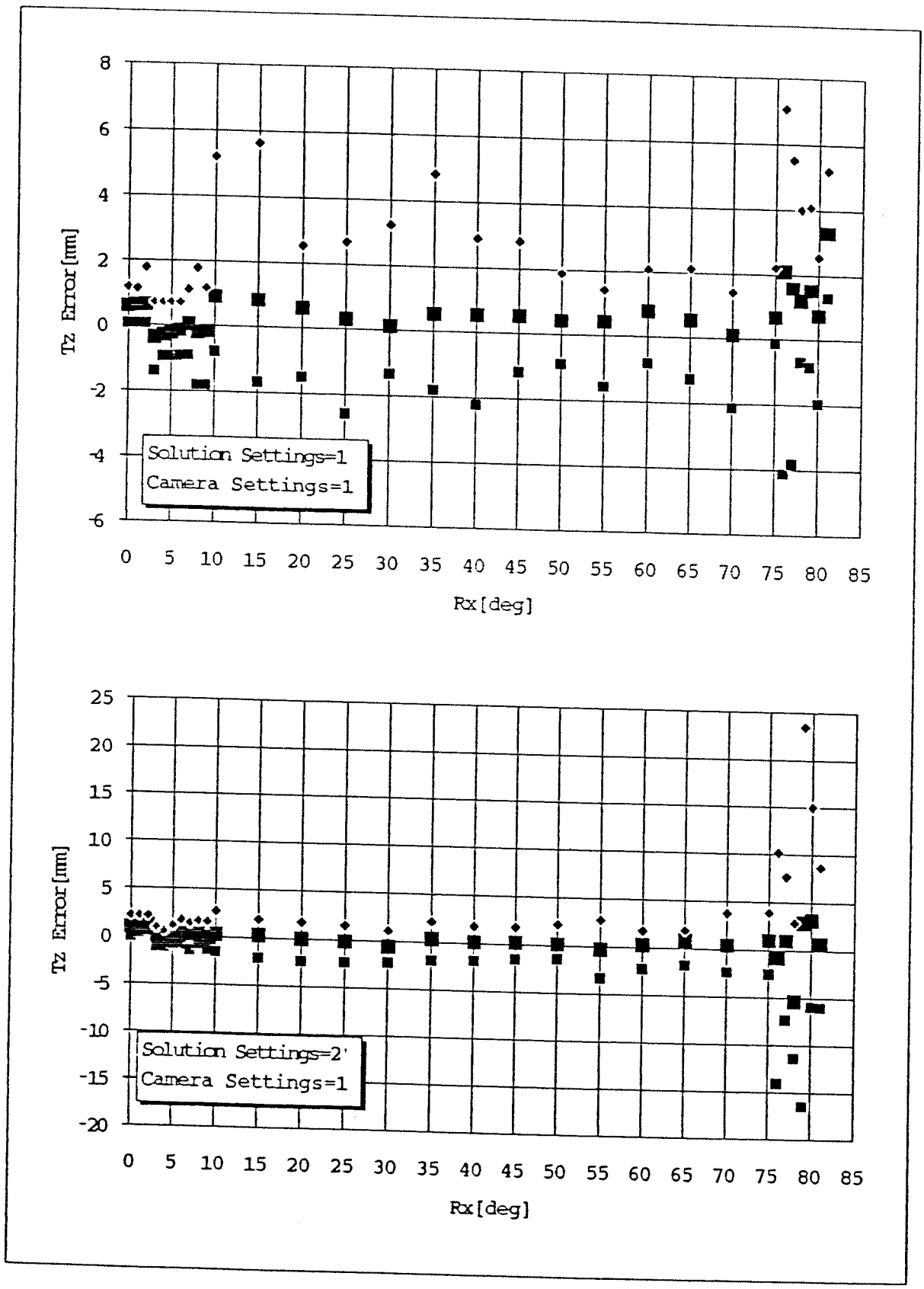


Figure 5.6: T_z Errors for MaxMinCross and EqualArcs10

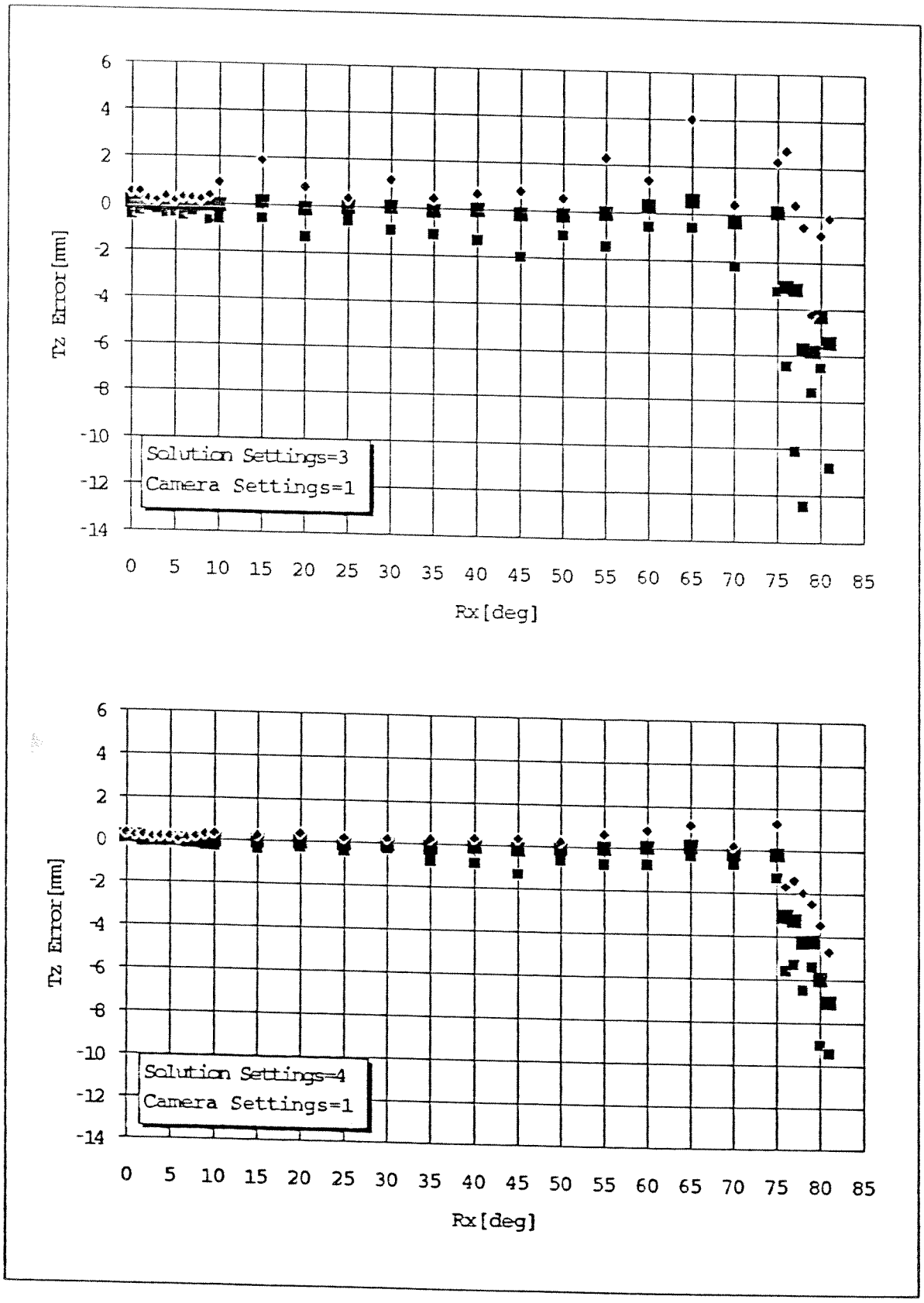


Figure 5.7: T_z Errors for EqualArcs100 and EqualArcs700

In the reconstruction phase, numerical errors occur for the following cases:

- $\alpha \approx 0^\circ$ and/or $\beta \approx 0^\circ$
- $|\alpha| > 80^0$ and/or $|\beta| > 80^\circ$

In the first case, the error is due to the sign inversion of the ellipse coefficients D or E at the curve-fitting. In fact

$$\lim_{\alpha \to 0^-} E = 0^- \qquad \lim_{\beta \to 0^-} D = 0^+$$

$$\lim_{\alpha \to 0^+} E = 0^+ \qquad \lim_{\beta \to 0^+} D = 0^-$$

can be written using the Equation 3.6. The effect of wrong sign reconstruction appears as large errors in α and β since in Equation 3.14 the signs of D and E are used.

To overcome this kind of numerical error a sign correction algorithm has been devised. Its effect is rather unpredictable since it depends on the absolute values of D and E. The simulation program makes use of the sign correction as a default setting.

Going back to the Equation 3.6 it is observed that for the practical range if

$$\begin{aligned} 0^\circ &\leq |\alpha| < 90^\circ \\ 0^\circ &\leq |\beta| < 90^\circ \\ d &> R \end{aligned}$$

are satisfied, the following equations can be written:

$$\begin{aligned} \mathrm{sgn}[B] &= \mathrm{sgn}[\tan\alpha \tan\beta] \\ \mathrm{sgn}[D] &= -\mathrm{sgn}[\beta] \\ \mathrm{sgn}[E] &= \mathrm{sgn}[\alpha] \end{aligned}$$

Combining these equations gives:

$$-\operatorname{sgn}[B] = \operatorname{sgn}[DE] \tag{5.1}$$

The Equation 5.1 can be used to check whether there is a sign conflict in the reconstruction of D or E after the curve-fitting is applied for an ellipse on the image plane[1]. The sign correction algorithm consists of inverting the sign of D or E as follows:

1. Apply the curve-fitting operation to obtain A, B, C, D and E
2. Check Equation 5.1
   - if the check is OK continue the solution process
   - if the check fails then
     - if $|D| < |E|$ then $D = -D$
     - if $|E| < |D|$ then $E = -E$

This algorithm assumes that among D and E the one that has the smaller magnitude has the wrong sign whenever Equation 5.1 is not satisfied. The algorithm is not able to detect sign errors if both D and E have reverse signs because $\operatorname{sgn}[DE] = \operatorname{sgn}[(-D)(-E)]$. The algorithm may also fail if the sign of B is reversed but the latter has never occurred during tests.

The other case where $|\alpha| > 80^0$ and/or $|\beta| > 80^o$ is more critical than the first one due to its degenerate nature. In fact, the ellipse image tends to be a line and it is not possible to apply the reconstruction algorithm in such a situation. The simulation program stops on numerical errors that cause overflows due to an uncompleted contour following operation. The sign error for D and E can also occur for high target rotations as previously demonstrated in Figures 5.2 and 5.4. This problem can be handled easily by increasing the number of selected pixels. The use of all the contour pixels is the radical solution to avoid sign errors on α and β, for highly rotated target configurations.

In any practical target configuration, the matrix operations are carried out without any numerical problems. Actually the only critical matrix operation

[1]This can also be a virtual image plane

Table 5.6: Nested Set 4

	Start	Stop	Step
T_x[mm]	0	0	0
T_y[mm]	0	0	0
T_z[mm]	-600	-600	0
R_x[°]	-10	10	0.1
R_y[°]	0	0	0
R_z[°]	0	0	0

Table 5.7: Nested Set 6

	Start	Stop	Step
T_x[mm]	0	0	0
T_y[mm]	0	0	0
T_z[mm]	-600	-600	0
R_x[°]	-10	10	0.1
R_y[°]	10	10	0
R_z[°]	0	0	0

is the pseudo-inversion of the matrix obtained from Equation 3.7. Since the elements of the matrix are formed using the coordinates of the selected contour pixels which provide at least 5 independent equations, no ill-conditions have been detected during simulations.

The Solution Settings and Camera Settings used for the simulations on numerical errors are given in Tables 5.1 and 5.2.

Figure 5.8 concentrates on the errors for R_x around 0° using Nested Set 4 given in Table 5.6. The excessive errors observed in the range $-4° < R_x < 4°$ are due to sign error of E. As this case demonstrates, the sign check is not able to detect the errors since $R_y = 0$ results in $|D| < |E|$ and every time a sign conflict is detected the sign of D is inverted rather than E.

However, when a similar simulation is done with $R_y = 10°$ using Nested Set 6 given in Table 5.7 then the sign correction algorithm becomes very effective. The result is shown in Figure 5.9 and when it is compared with the previous graph the influence of the sign correction algorithm is clear.

Figure 5.10 shows the simulation for R_y errors around $= 0°$ using Nested Set 5 given in Table 5.8. The fact that the graph is exactly the same

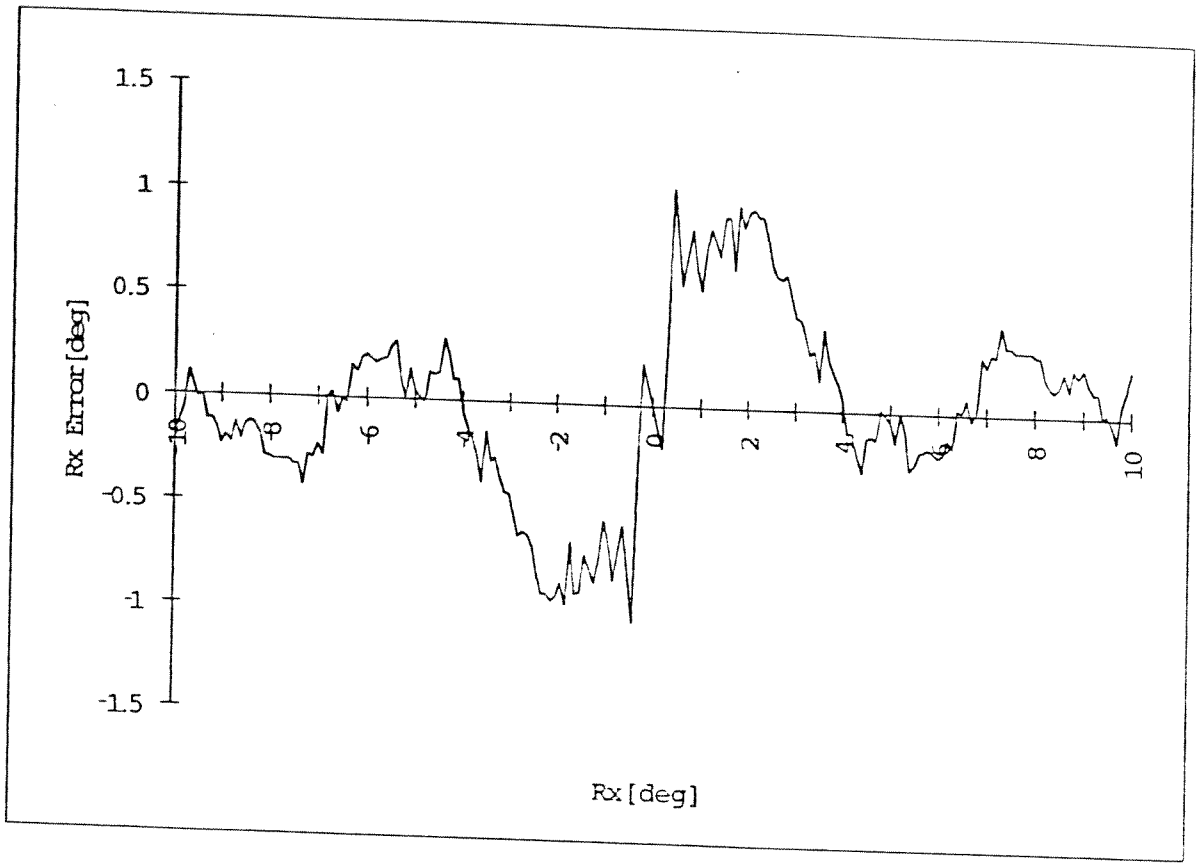


Figure 5.8: R_x Errors Near 0°

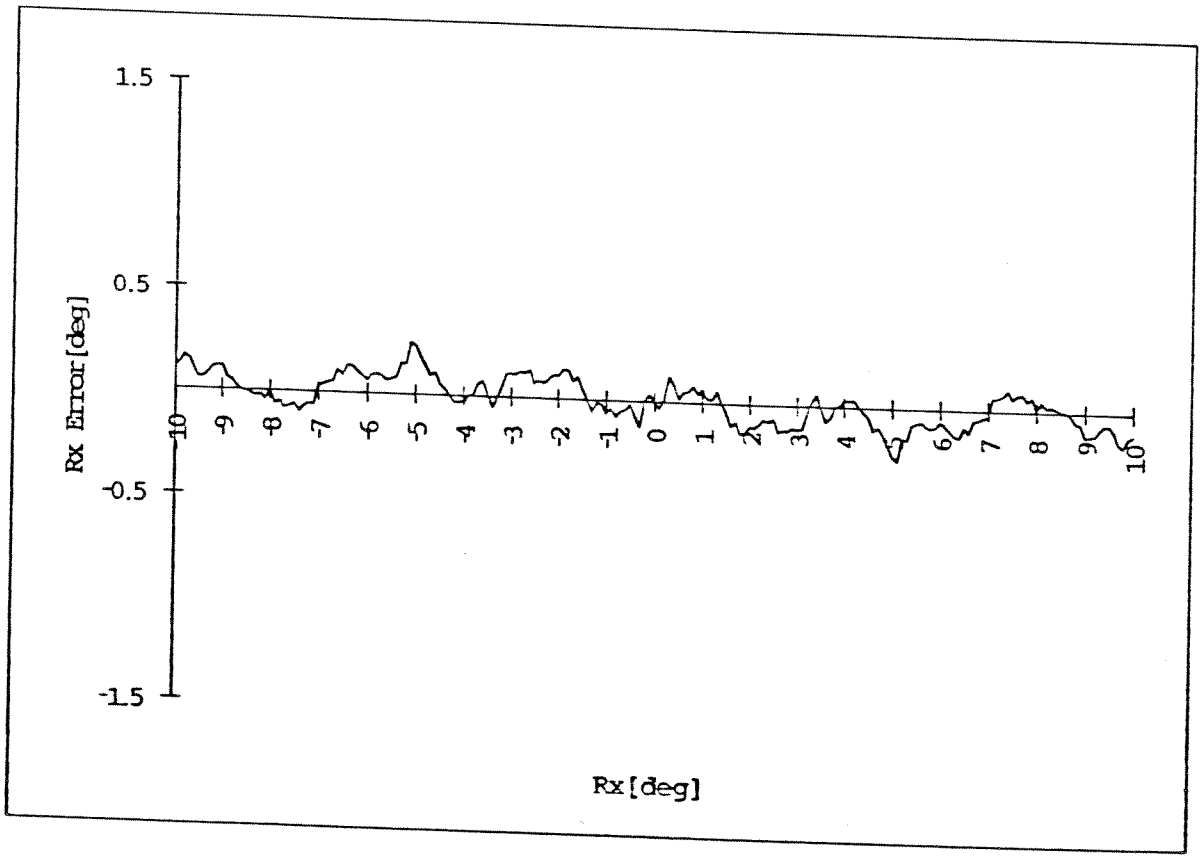


Figure 5.9: R_x Errors Near 0° with $R_y = 10°$

Table 5.8: Nested Set 5

	Start	Stop	Step
T_x[mm]	0	0	0
T_y[mm]	0	0	0
T_z[mm]	-600	-600	0
R_x[°]	0	0	0
R_y[°]	-10	10	0.1
R_z[°]	0	0	0

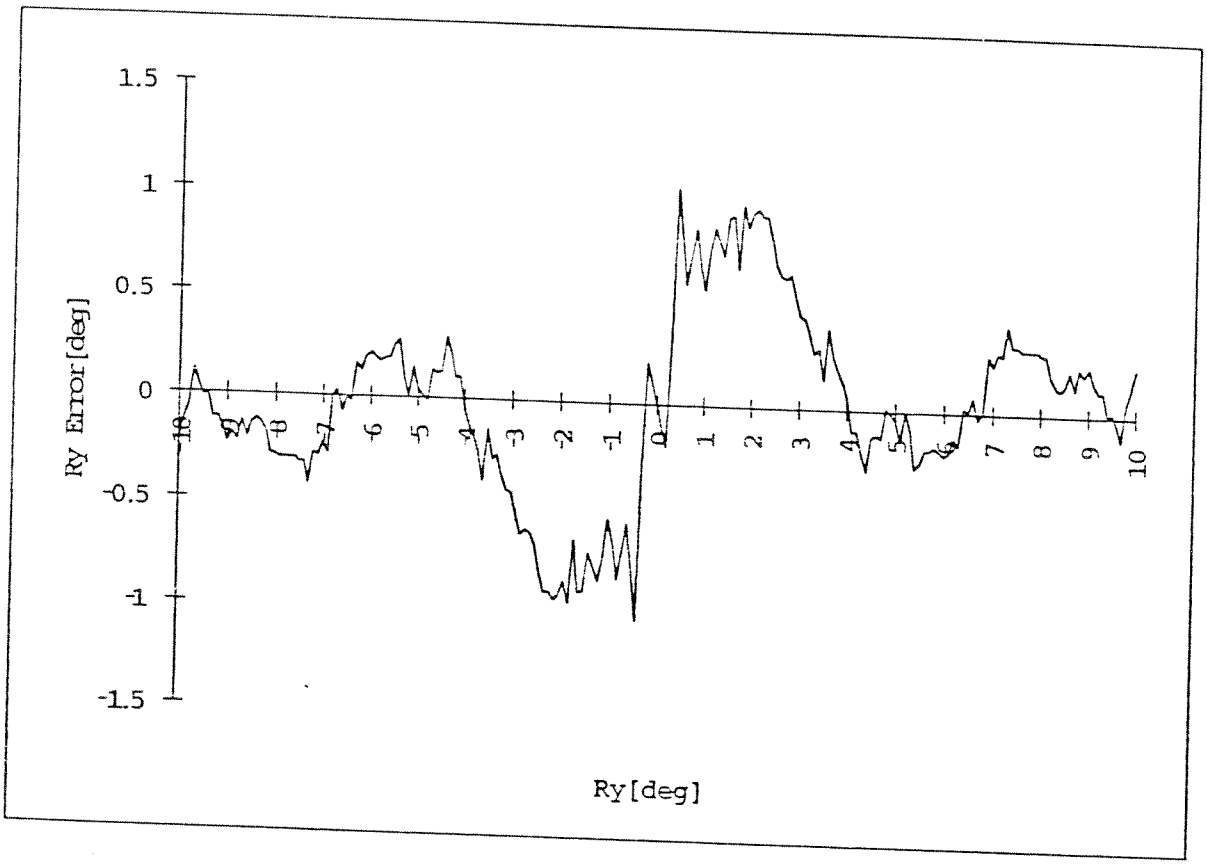


Figure 5.10: R_y Errors Near 0°

as in Figure 5.8 justifies that the symmetric behaviour of the two image sets is preserved during all numerical operations. This conservation is only possible for a sensor having square pixels.

5.5 Determination of d_0

The internal camera parameter d_0 denotes the perpendicular distance between the image plane and the optic center. Its knowledge is necessary for the solution process, however there is not a simple way to obtain d_0 explicitly.

5.5.1 Explicit Determination

The explicit determination of d_0 is only possible by adding an encoder to the camera's focus system. Then the extra system should be calibrated to give the correct d_0 value. However such a system adds to the hardware cost, and is difficult to adapt to a conventional camera lens. The explicit determination of d_0 is the ideal case and corresponds to the "perfect focus" selection in the simulation program, where d_0 is substituted by its calculated value for the given object distance. Although not a practical option, it is used to test the accuracy of the algorithm for the remaining steps as if d_0 was a known parameter.

5.5.2 Use of Constant d_0

A constant value can be used as an estimate for d_0 such that

$$d_f \leq d_0 \leq 2d_f$$

is satisfied for the cases where the target distance does not vary considerably. The constant value has to be found with prior calibration according to the range where the target is supposed to be located. The use of a constant d_0 has been presented in [10] and a calibration technique has been developed. It is useful in cases where the lens setting remains unchanged after a calibration is performed. This means that neither a focusing nor a zoom is applied during the operation of the system because a value that differs significantly from the actual value of d_0 will produce large errors in the reconstruction of the target distance. Whenever there is a need to play with the lens setting to improve the image quality or to enlarge or reduce the image size, the use of a constant d_0 should be avoided.

5.5.3 Iterative Solution to Determine d_0

This method has been invented to give an estimate of d_0 with few iteration steps. It actually uses the target as a calibration mark during the solution process. This means that the system gains a self calibration feature for different target configurations and lens settings when d_f is known. The proposed algorithm consists of first solving the problem using $d_0 = d_f$ assuming that the target is at infinity as a first guess. The reconstruction process is carried until the distance d is obtained, then using the value of d and the formula given by Equation 2.1, d_0 is recalculated and the solution process restarted until d_0 converges. The power of the algorithm is justified by the fast convergence of d_0. In most cases the second iteration leads to the required accuracy. This algorithm can be summarized as:

1. Take $d_0^{(i)} = d_f$ as a first guess for $i = 1$
2. Reconstruct d
3. Use $d_0^{(i+1)} = d\, d_f/(d - d_f)$
4. Increment i by one
5. Repeat steps 2 to 4 until $|d_0^{(i+1)} - d_0^{(i)}| < \epsilon$
   where ϵ is the convergence criteria
6. Process the remaining portion of the reconstruction algorithm

The error made in the estimation of d_0 has not been studied separately but its influence on the reconstruction results of the configuration parameters has been investigated. It should be noted that ϵ is just a convergence criteria and does not limit the error made in the estimation of d_0. However for $d > 10d_f$, satisfactory results are obtained and in general

$$|d_0^{(\text{perfect})} - d_0^{(\text{estimated})}| < \epsilon$$

is satisfied.

Table 5.9: Sequential Set 1

	Start	Vel.[/sec]	Time[sec]
Tx[mm]	-25	0.5	100
Ty[mm]	-25	0.5	
Tz[mm]	-700	-2	
Rx[°]	-80	1.6	
Ry[°]	80	-1.6	
Rz[°]	0	0	

5.5.4 Simulation Results

The Solution Settings and Camera Settings used for the simulations of this part are given in Tables 5.1 and 5.2. The Sequential Set 1 given in Table 5.9, is created to move the target by giving it a combination of translational and rotational motions.

Before analyzing the first simulation result using a Sequential Set, the following properties should be kept in mind:

- Sequential set solutions are not reproducible since the computation time may be slightly different even on the same computer
- The comparison graphs for the same sequential set may not have the same number of data points if the relative computation speed between the solution types is different
- There is not a correspondence between the data points of different simulation results

From Figures 5.11 and 5.12 it is observed that the determination of d_0 has not a considerable effect on the reconstruction of T_x and T_y.

On Figure 5.13 however, a large underestimation of the target distance is observed for the constant $d_0 = 28.5$mm. The constant value for the image plane distance is smaller than the average perfect value which can be calculated as $d_0 \approx 29mm$ using Equation 2.4 and substituting $d_{eff} \approx 800mm$. According to this, the estimated angle of view is larger and the target observed seems to be closer. The results for the iterative estimation of d_0 are surprisingly close to

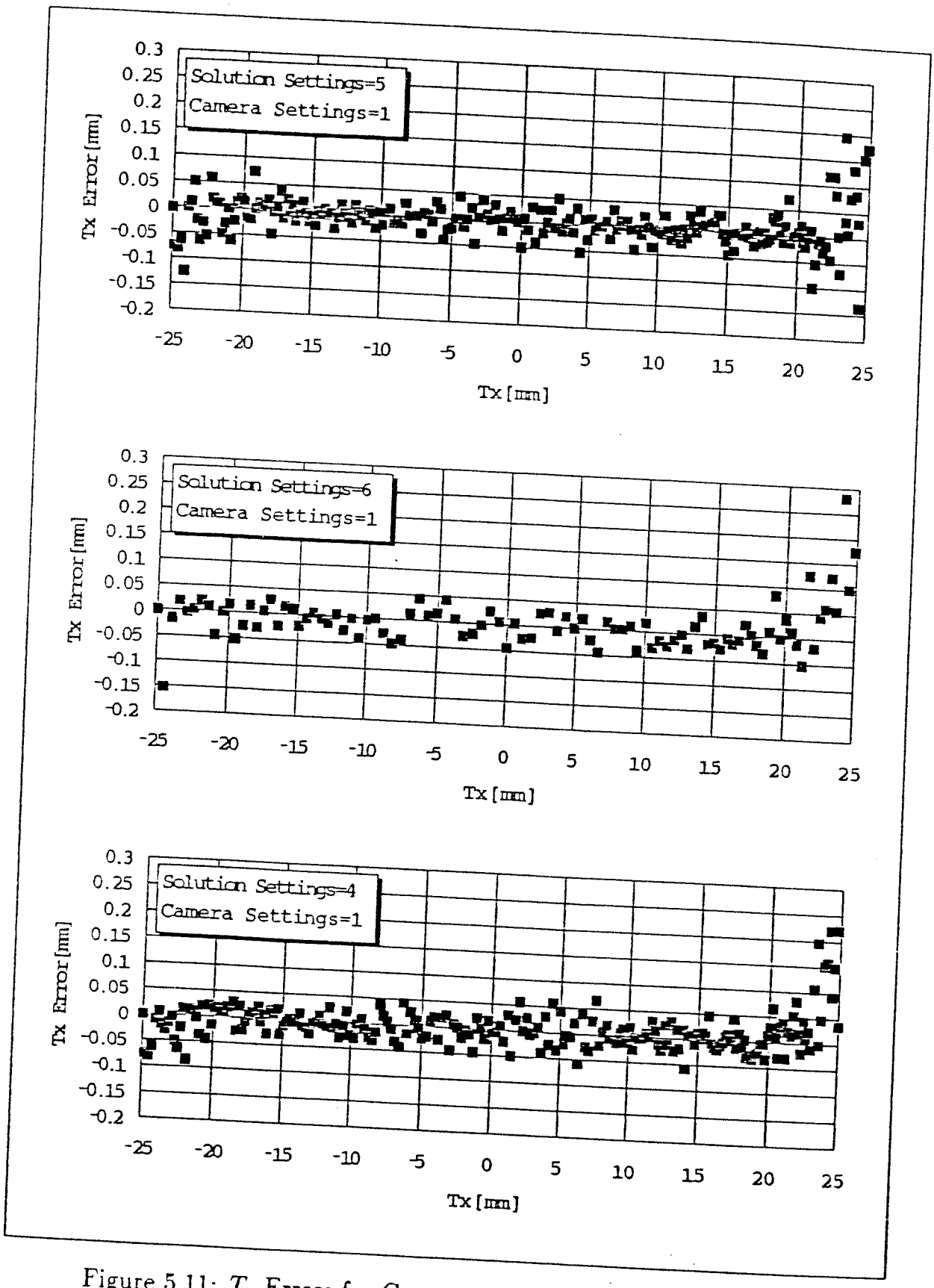


Figure 5.11: T_x Errors for Constant, Iterative and Perfect Focus

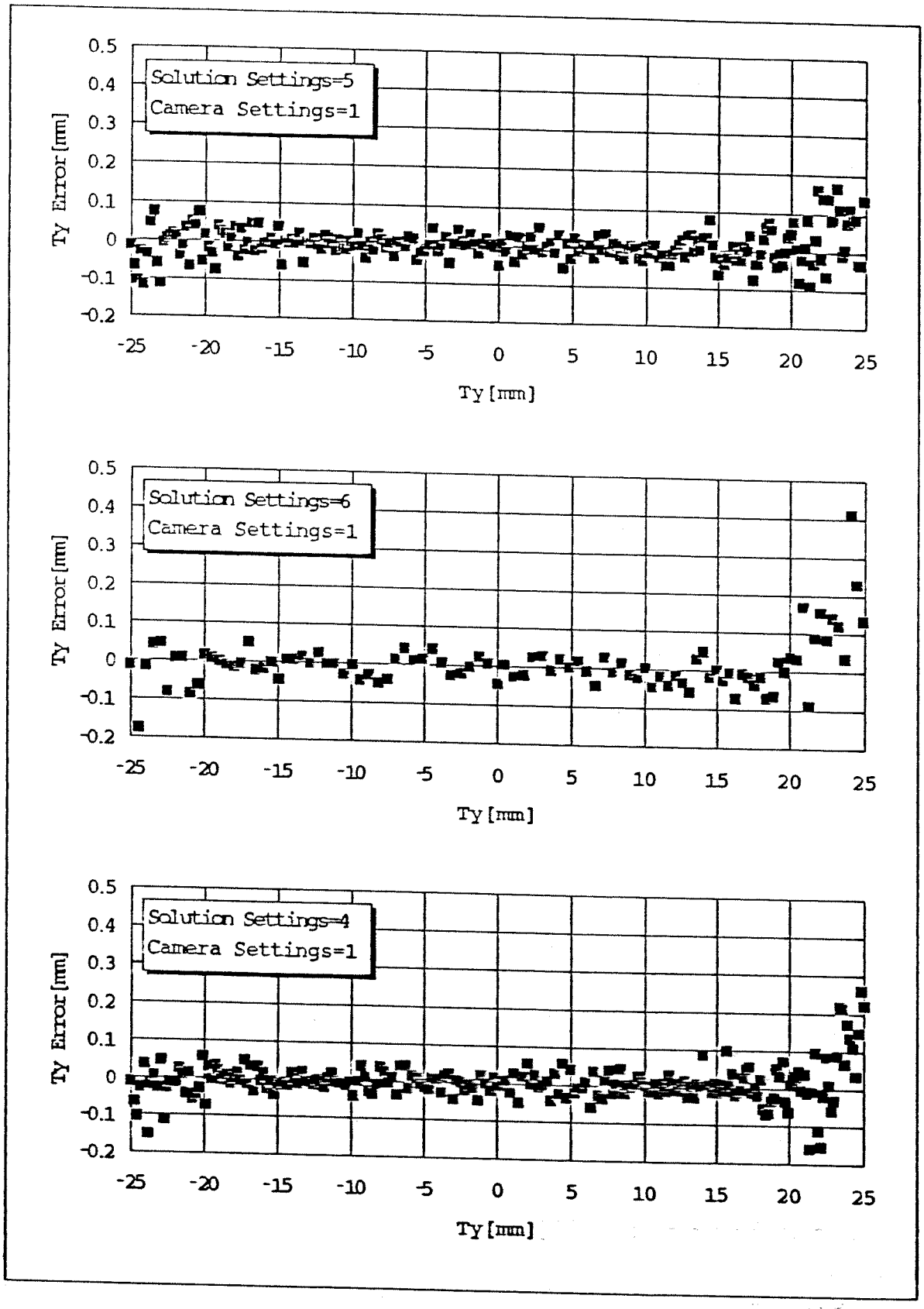


Figure 5.12: T_y Errors for Constant, Iterative and Perfect Focus

the results for perfect d_0 values. The time consuming behaviour of the iterative solution seems to be the only drawback. The number of data points on the graphs gives an idea about the relative computation speeds for the same Sequential Set.

Looking at the Figures 5.14 and 5.15 for R_x and R_y respectively, the behaviour of the 3 graphs are quite similar except that the constant d_0 case gives better results around 0°. This fact is justified by a numerical coincidence when Equation 3.14 is examined. It is observed from the formula that the magnitude of the estimated α and β depends also on d_0 and d. The underestimated d_0 results in an underestimated d and together they make the numerator of the arctangent function smaller. This is particularly contributing the reconstruction of nearly zero rotation angles. However the use of constant d_0 cannot be favored as the overestimated case would just do the reverse contribution.

To justify this fact, another simulation has been carried out using Solution Setting 7 which takes a constant overestimated d_0. The results are shown in Figure 5.16 only for R_x and R_y variables. As expected from the discussion above, the overestimation of d_0 results in higher errors for R_x and R_y around 0° than any other solution setting.

Continuing with the simulation results, Figure 5.17 shows the R_z errors versus T_z. The x axis is selected arbitrarily since $R_z = 0$ is constant for all the images in this set. It is observed that the overall behaviour of the R_z error is almost not affected by the determination of d_0. As the reconstruction of the in-plane rotation is the last step, it is influenced by the errors made on the reconstruction of other variables.

Figures 5.18 and 5.19 give the minimum, average and maximum errors obtained during the whole sequence for translation and rotation variables respectively. This kind of graphs can be obtained by taking one of the configuration parameters constant for all the images in the set. In the Sequential Set 1, R_z was selected as the constant variable.

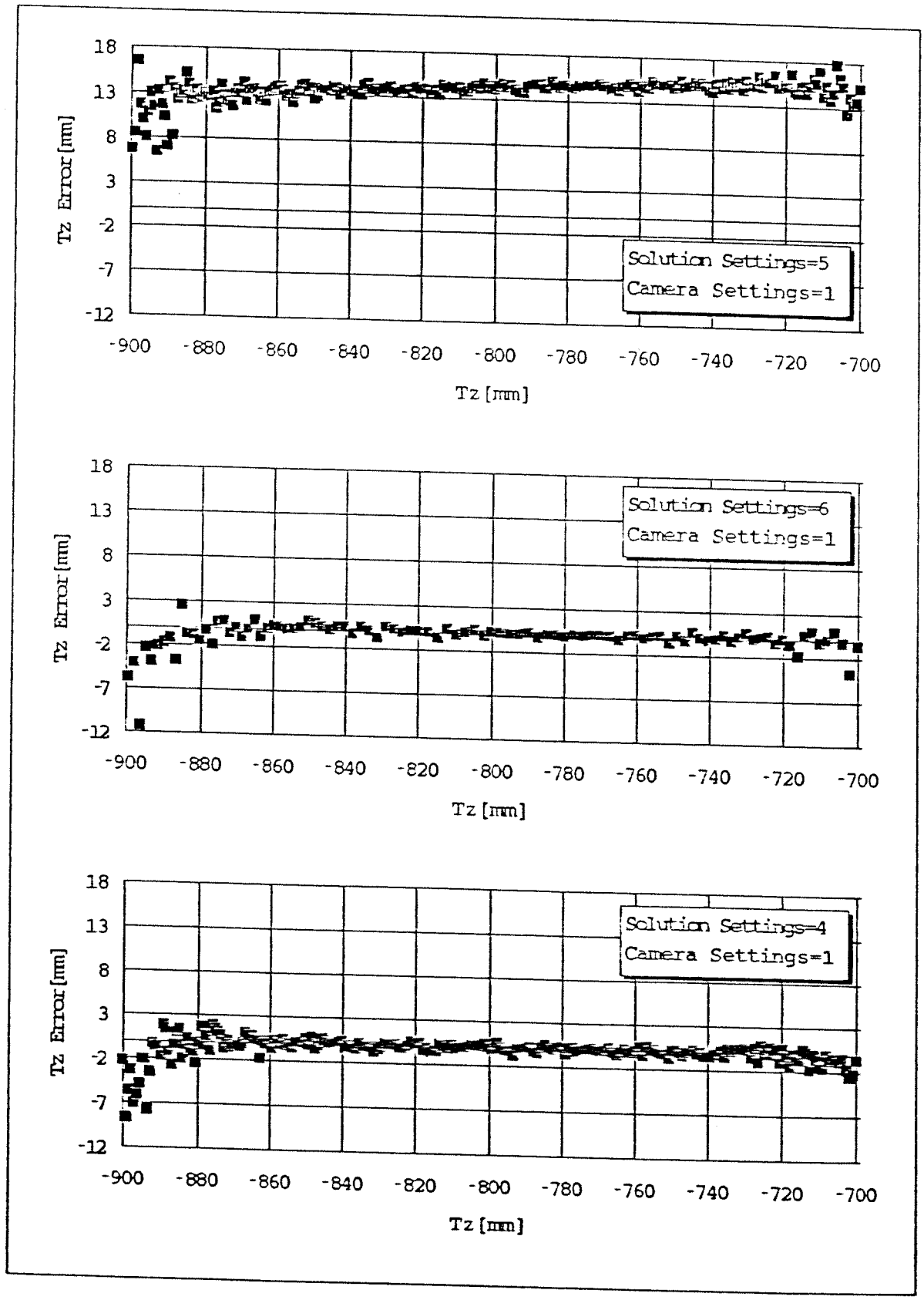


Figure 5.13: T_x Errors for Constant, Iterative and Perfect Focus

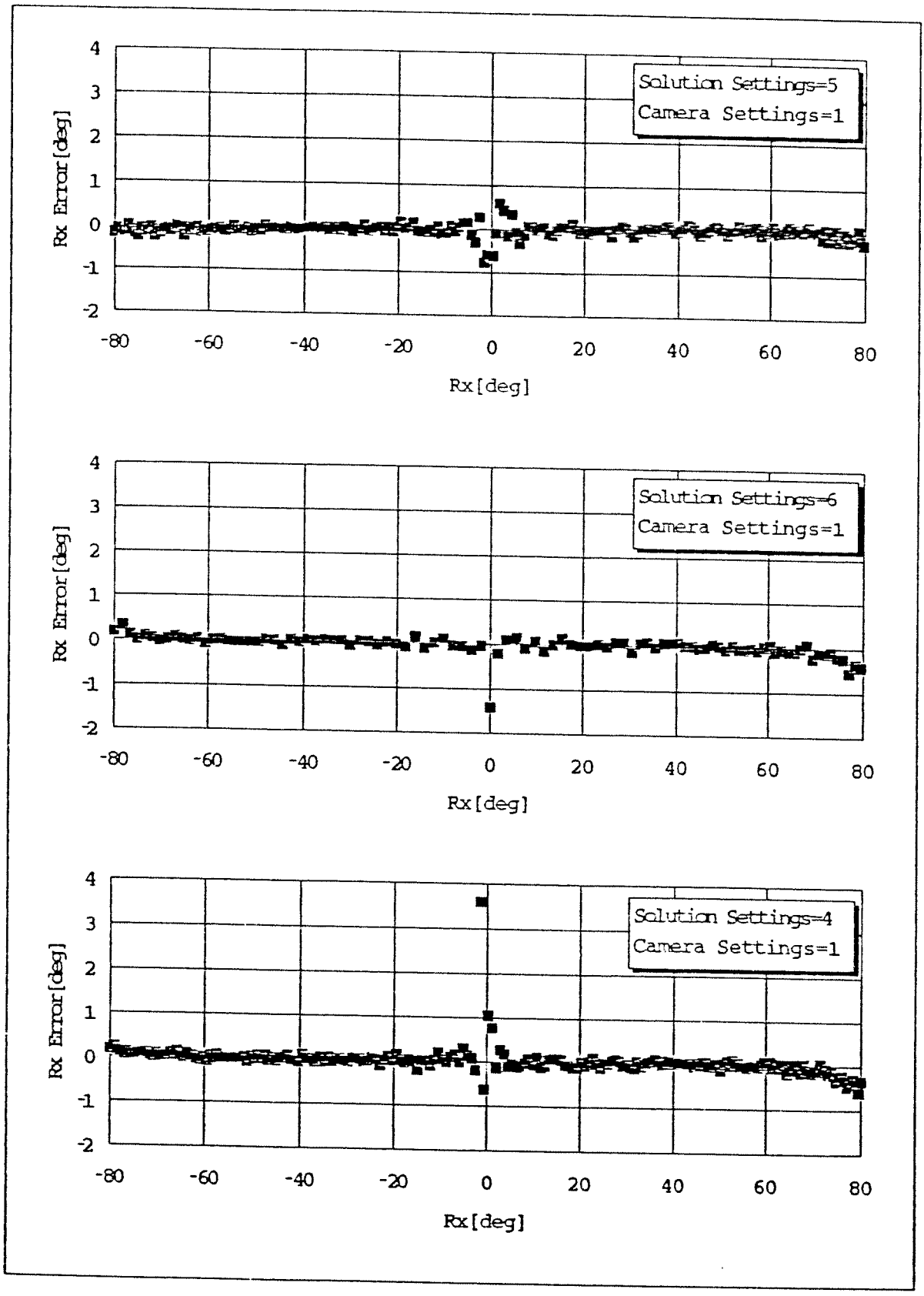


Figure 5.14: R_x Errors for Constant, Iterative and Perfect Focus

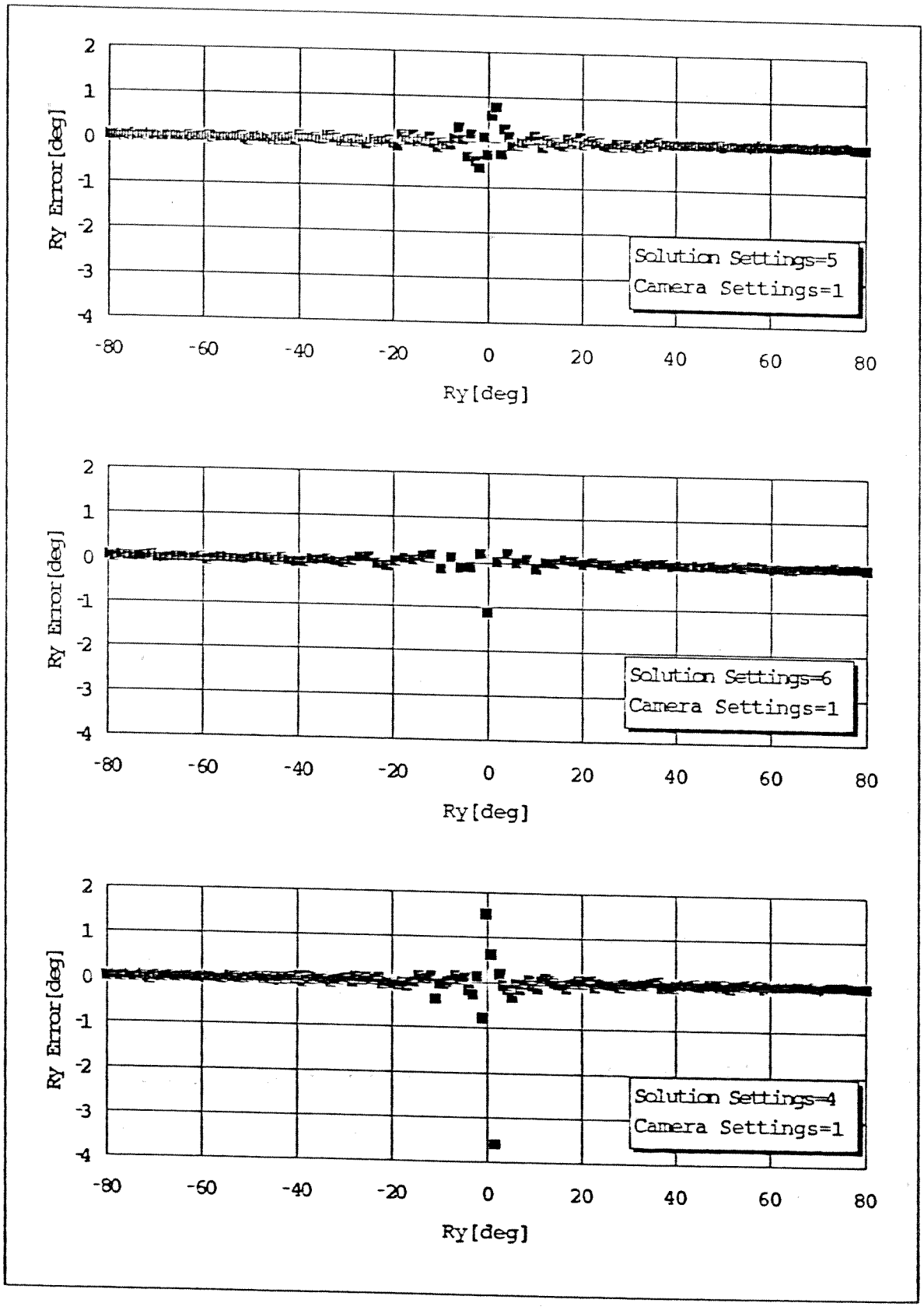


Figure 5.15: R_y Errors for Constant, Iterative and Perfect Focus

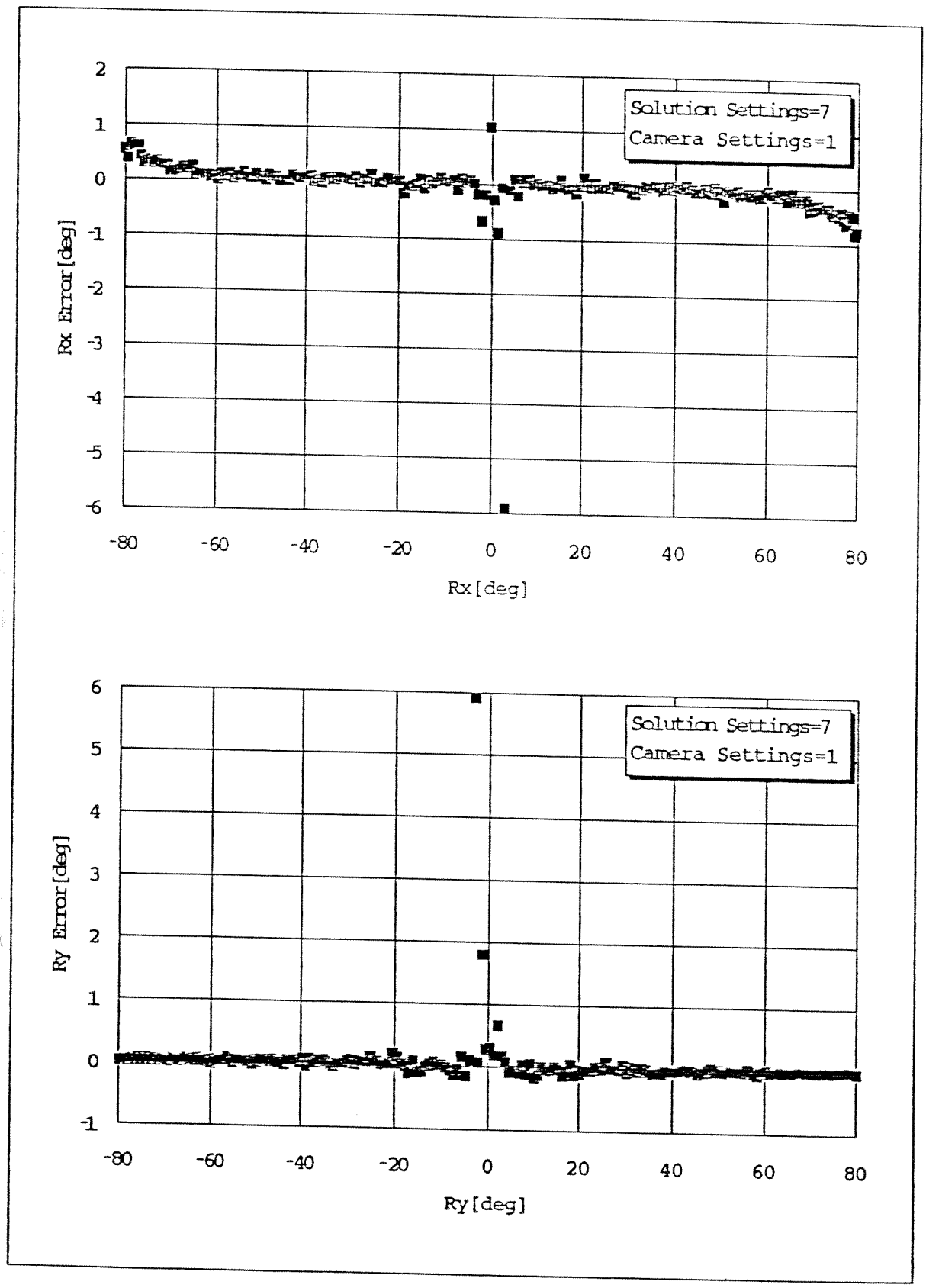


Figure 5.16: R_x and R_y Errors for Over Estimated Constant Focus

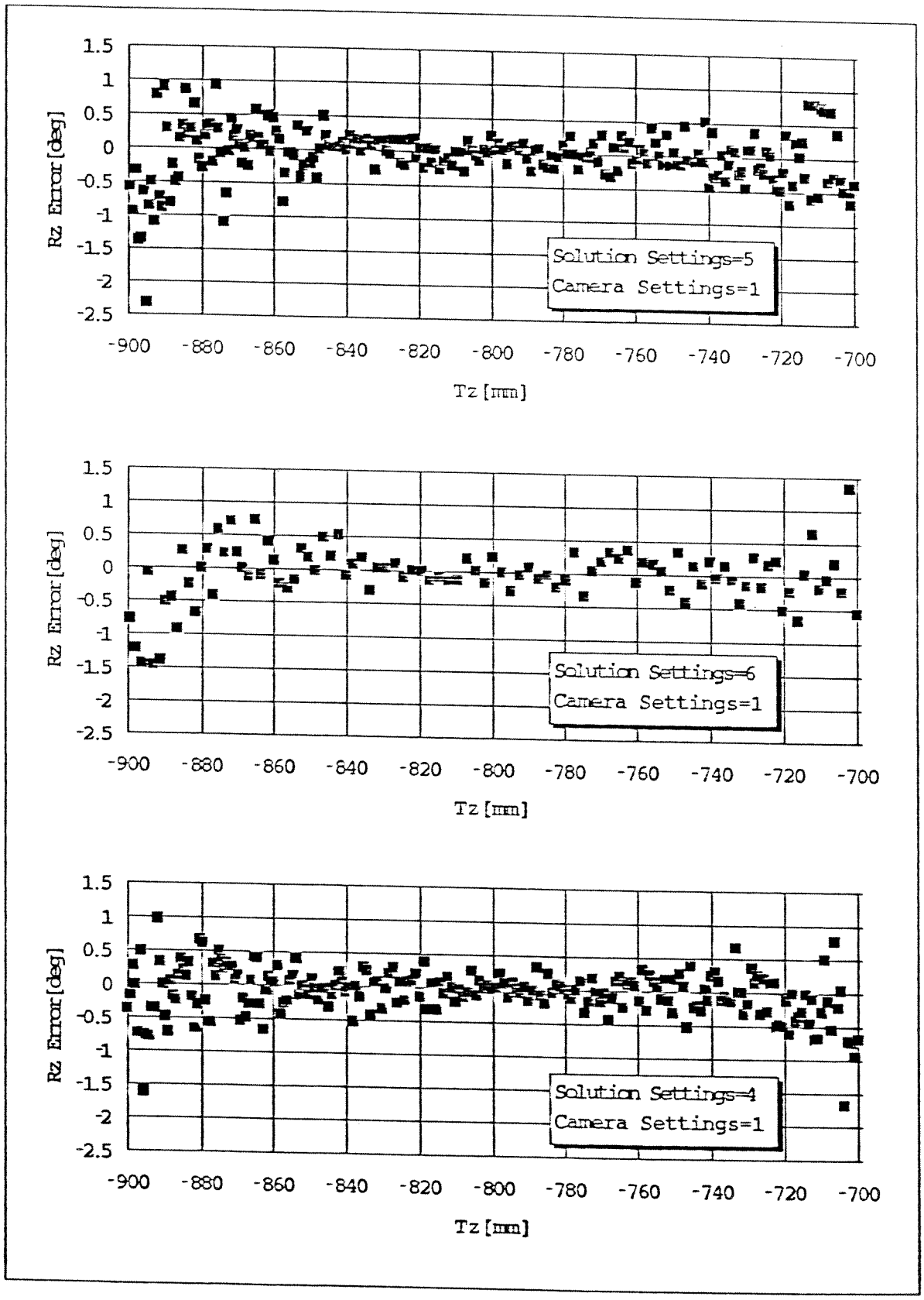


Figure 5.17: R_z Errors for Constant, Iterative and Perfect Focus

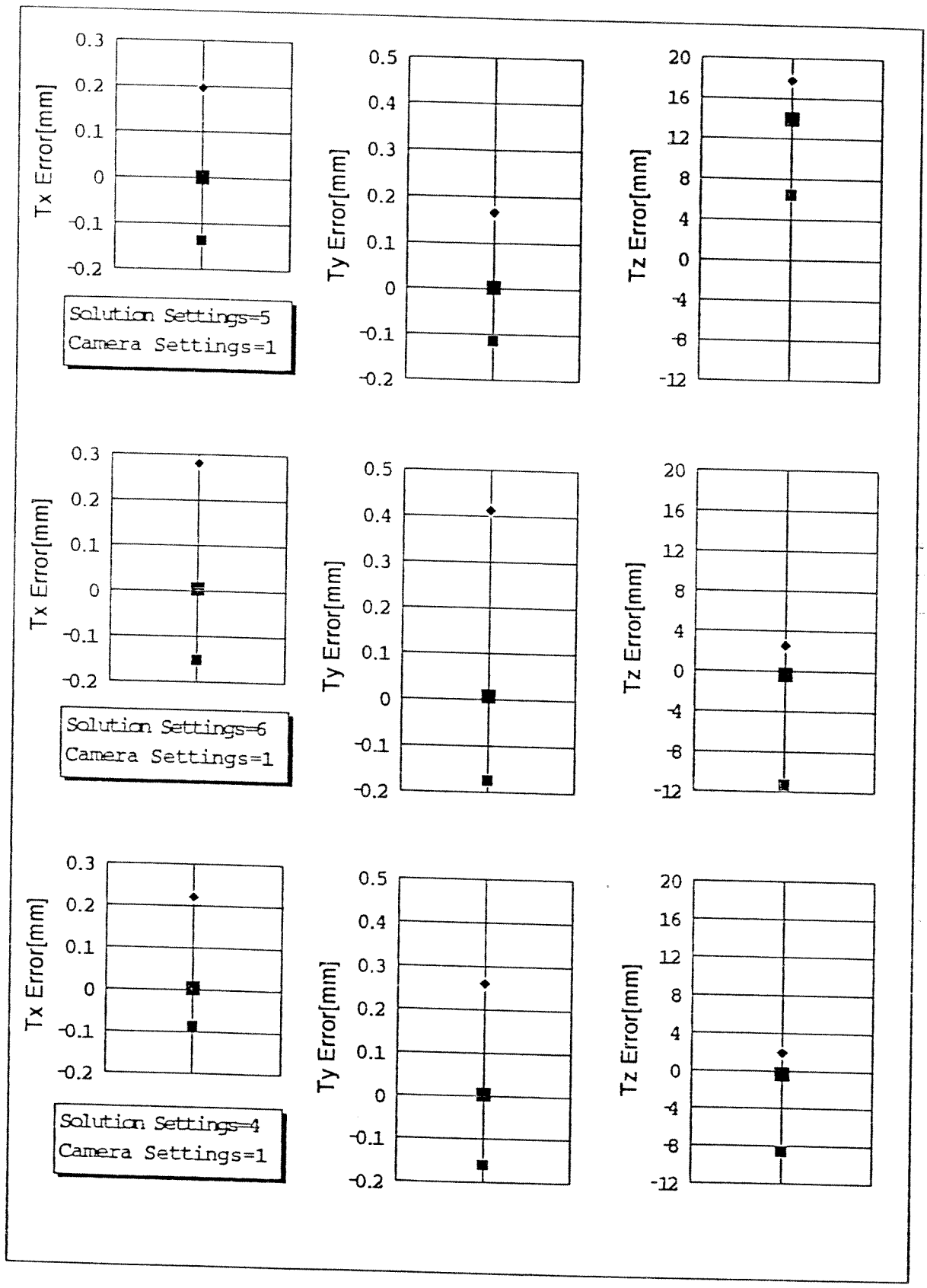


Figure 5.18: Average Translation Errors for Constant, Iterative and Perfect Focus

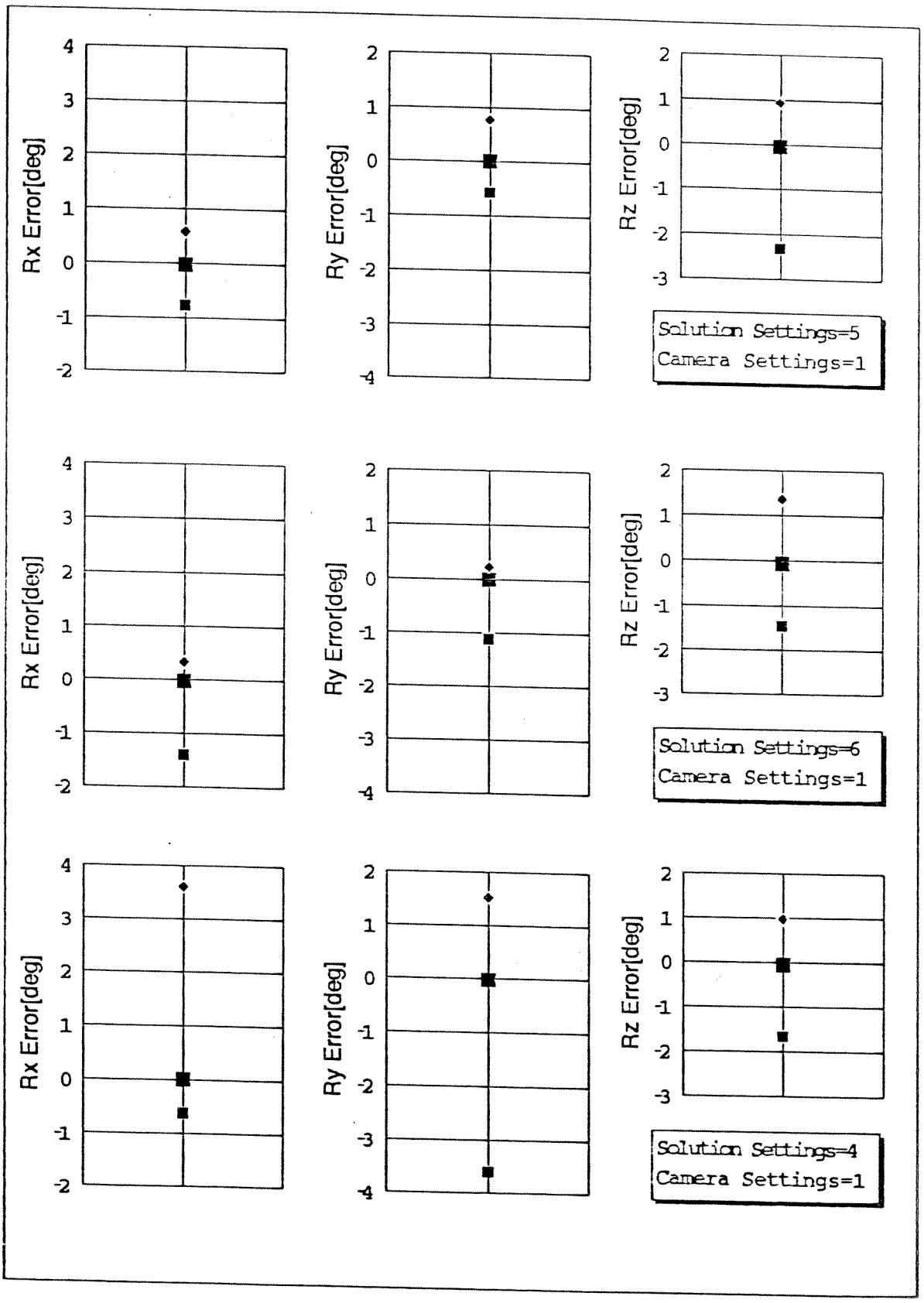


Figure 5.19: Average Rotation Errors for Constant, Iterative and Perfect Focus

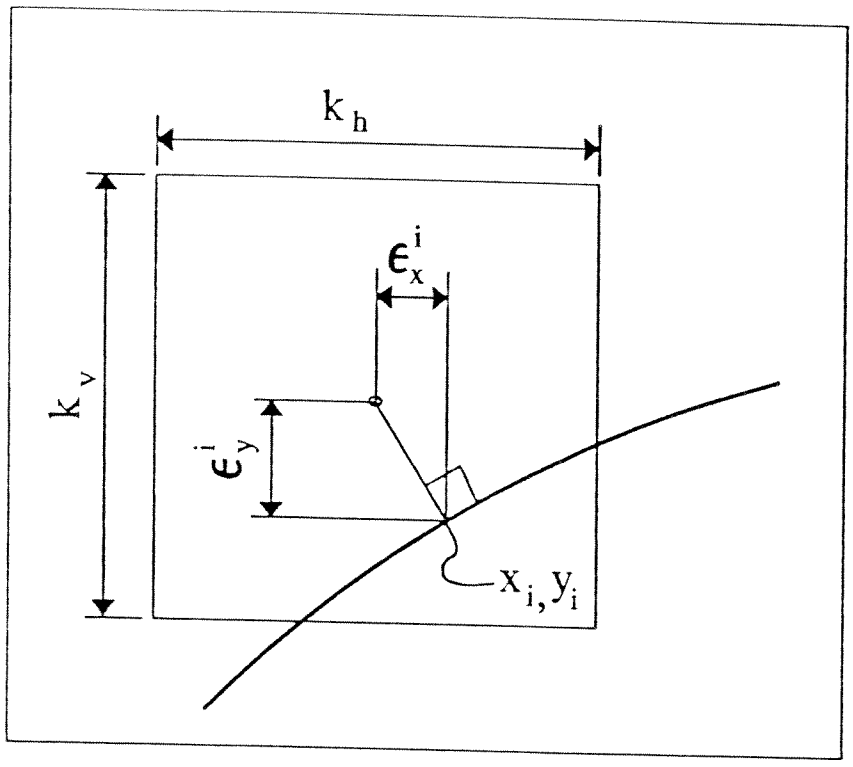


Figure 5.20: Error on the Selection of a Pixel

5.6 Resolution and Pixel Size

The dimension of the sensor array is determined together by the resolution and the pixel size. Where "resolution" determines the number of rows and columns of the sensor array, "pixel size" gives the size of an individual pixel. For the same pixel size, the sensor size denoted by S_s and consequently the angular field of view will be larger if the resolution is increased. Therefore the resolution has no effect on the accuracy of the reconstruction algorithm. The effect of S_s on the angular field of view was given in Equation 2.6.

The pixel size however plays an important role on the accuracy of the algorithm. As the pixel size becomes smaller, the error due to the quantization of the image ellipse vanishes. In fact the reconstruction error is mainly due to the digital nature of the image ellipse. Considering the rectangular geometry of a pixel where k_h is the horizontal pixel size and k_v is the vertical pixel size, on any selected pixel, the cumulative error will be due to the non-zero distance from the center of the pixel perpendicular to the portion of the original curve within the same pixel. Figure 5.20 depicts this situation. This error is expressed separately

in x and y directions as ϵ_x and ϵ_y, respectively. The inequalities

$$\begin{aligned} |\epsilon_x| &\leq k_h/2 \\ |\epsilon_y| &\leq k_v/2 \end{aligned}$$

will be true for all the pixels selected along the contour. In the reconstruction process, the set of ellipse equations previously given by Equation 3.7 will actually be in the following form:

$$\begin{aligned} A(x_0 \pm \epsilon_x^0)^2 + B(x_0 \pm \epsilon_x^0)(y_0 \pm \epsilon_y^0) + C(y_0 \pm \epsilon_y^0)^2 + D(x_0 \pm \epsilon_x^0) + E(y_0 \pm \epsilon_y^0) + 1 &= 0 \\ A(x_1 \pm \epsilon_x^1)^2 + B(x_1 \pm \epsilon_x^1)(y_1 \pm \epsilon_y^1) + C(y_1 \pm \epsilon_y^1)^2 + D(x_1 \pm \epsilon_x^1) + E(y_1 \pm \epsilon_y^1) + 1 &= 0 \\ \ldots &= 0 \\ A(x_N \pm \epsilon_x^N)^2 + B(x_N \pm \epsilon_x^N)(y_N \pm \epsilon_y^N) + C(y_N \pm \epsilon_y^N)^2 + D(x_N \pm \epsilon_x^N) + & \\ E(y_N \pm \epsilon_y^N) + 1 &= 0 \end{aligned}$$

where x_i, y_i denote the coordinates of a point lying on the original curve and $x_i \pm \epsilon_x^i$, $y_i \pm \epsilon_y^i$ takes to account the possible error in the selection of the i^{th} pixel. The errors related to each selected pixel may have a canceling effect on the overall, depending on the sign of ϵ_x and ϵ_y. However, this fact cannot be expressed by a simple mathematical formula.

The set of equations above show that a sensor with small pixels should be preferred to reduce the amount of *maximum possible error* due to the digital nature of the image ellipse. The practical pixel dimensions are given in the order of microns. In the simulation program, the default pixel size is $13\mu\text{m} \times 13\mu\text{m}$ and the resolution is 400×400 pixels. Both values can be changed for comparison between different hardware configurations.

The Solution Settings and Camera Settings used for the simulations concerning the pixel size and resolution are given in Tables 5.1 and 5.2.

The Sequential Set 2, described in Table 5.10, is similar to Sequential Set 1 except that $T_x = 0$ and $T_y = 0$ have been used and R_z is given a constant angular velocity.

In Figures 5.21 and 5.22, it is observed that for the low resolution case which is shown on top, T_x and T_y reconstruction errors are generally higher than

Table 5.10: Sequential Set 2

	Start	Vel.[/sec]	Time[sec]
T_x[mm]	0	0	100
T_y[mm]	0	0	
T_z[mm]	-700	-2	
R_x[°]	-70	1.4	
R_y[°]	70	-1.4	
R_z[°]	0	0.9	

the high resolution case. However when the maximum errors are considered one can be surprised to see that the high resolution camera produces larger errors at some data points. This fact can be justified only by taking to account the shift of the ellipse center and the true center on the image of the center spot. In the calculations, the centroid of the image spot being used, is determined using the contour pixels of the spot. The high resolution camera will of course give a better estimation for the center of the image spot. But, for a moderately rotated target plane, as the calculated point gets closer to the ellipse center, it gets farther from the true center. The center of the image spot being used as an approximation for the true center, the high resolution camera will give slightly larger reconstruction errors at some data points. However when the overall graph is considered it gives better results in terms of the average error. In any case, the error for the reconstruction of T_x and T_y is small enough to justify the approximate determination of the target's true center by using the centroid of the image spot.

There is almost no need to make any comment on Figure 5.23 which clearly shows the advantages of a high resolution camera for the reconstruction of the target distance.

In Figures 5.24, 5.25 and 5.26, once again the advantage of the high resolution camera is obvious for the reconstruction of target rotation angles R_x, R_y and R_z.

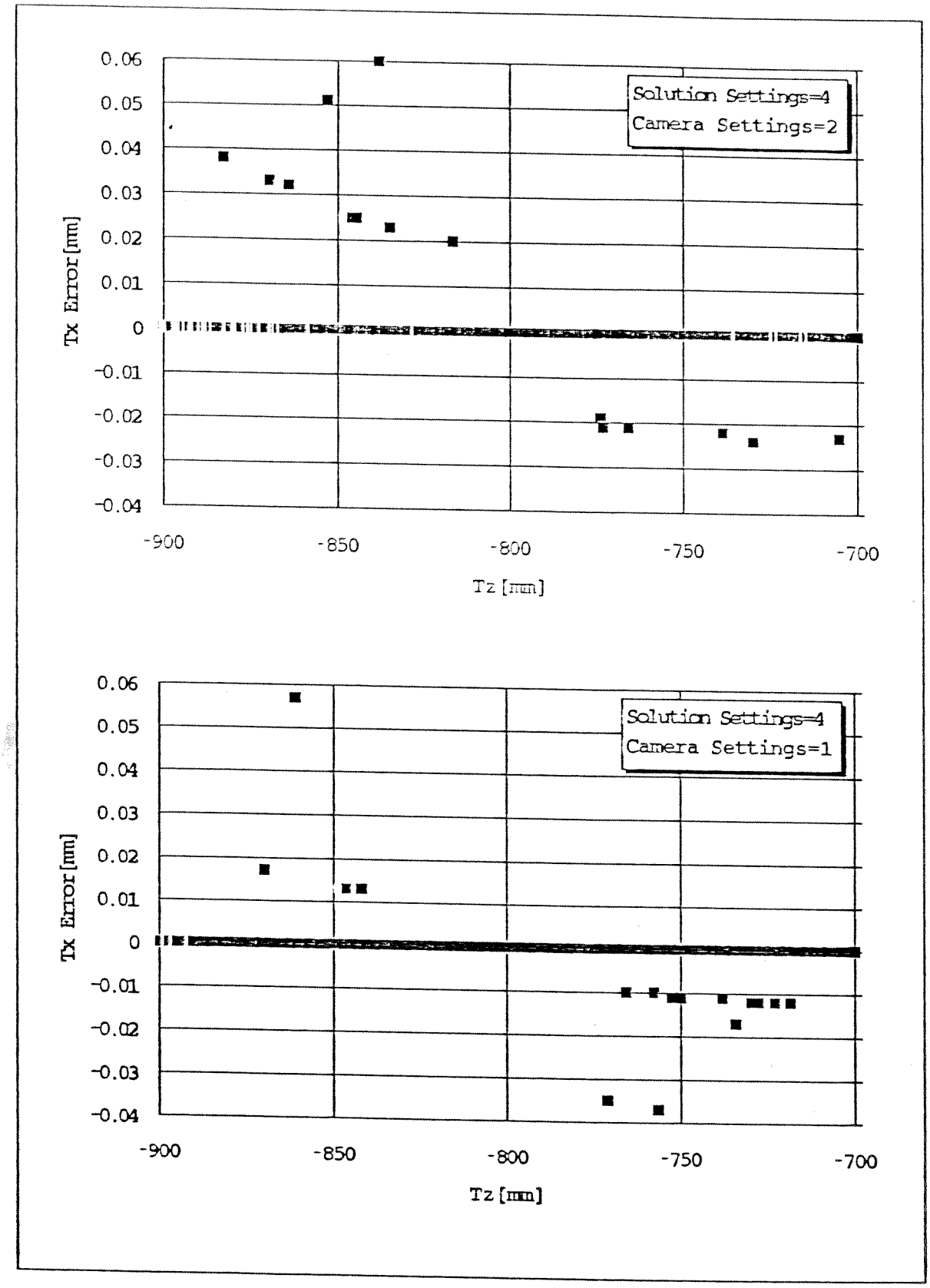


Figure 5.21: T_x Errors for Low and High Resolution Cameras

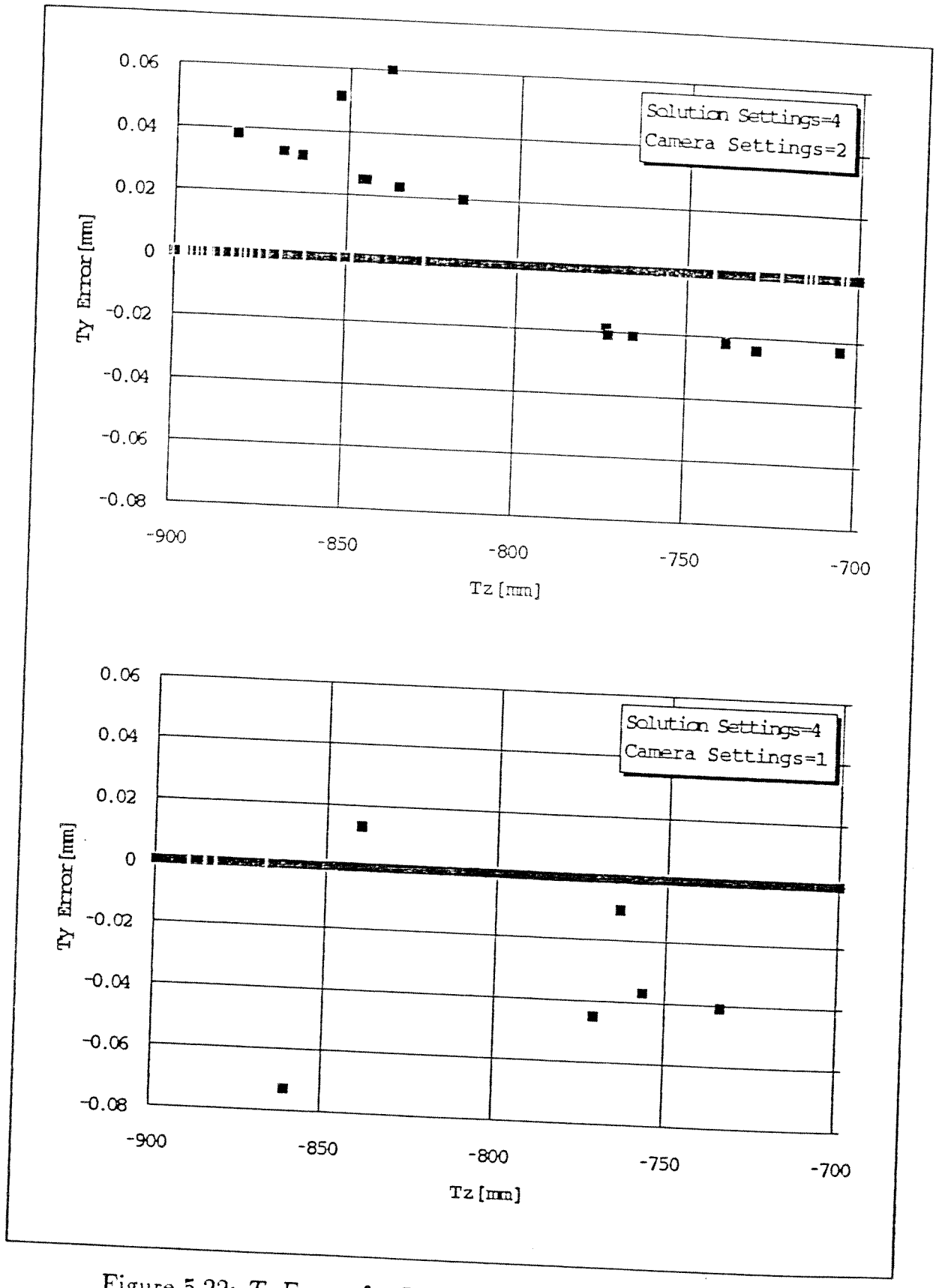


Figure 5.22: T_y Errors for Low and High Resolution Cameras

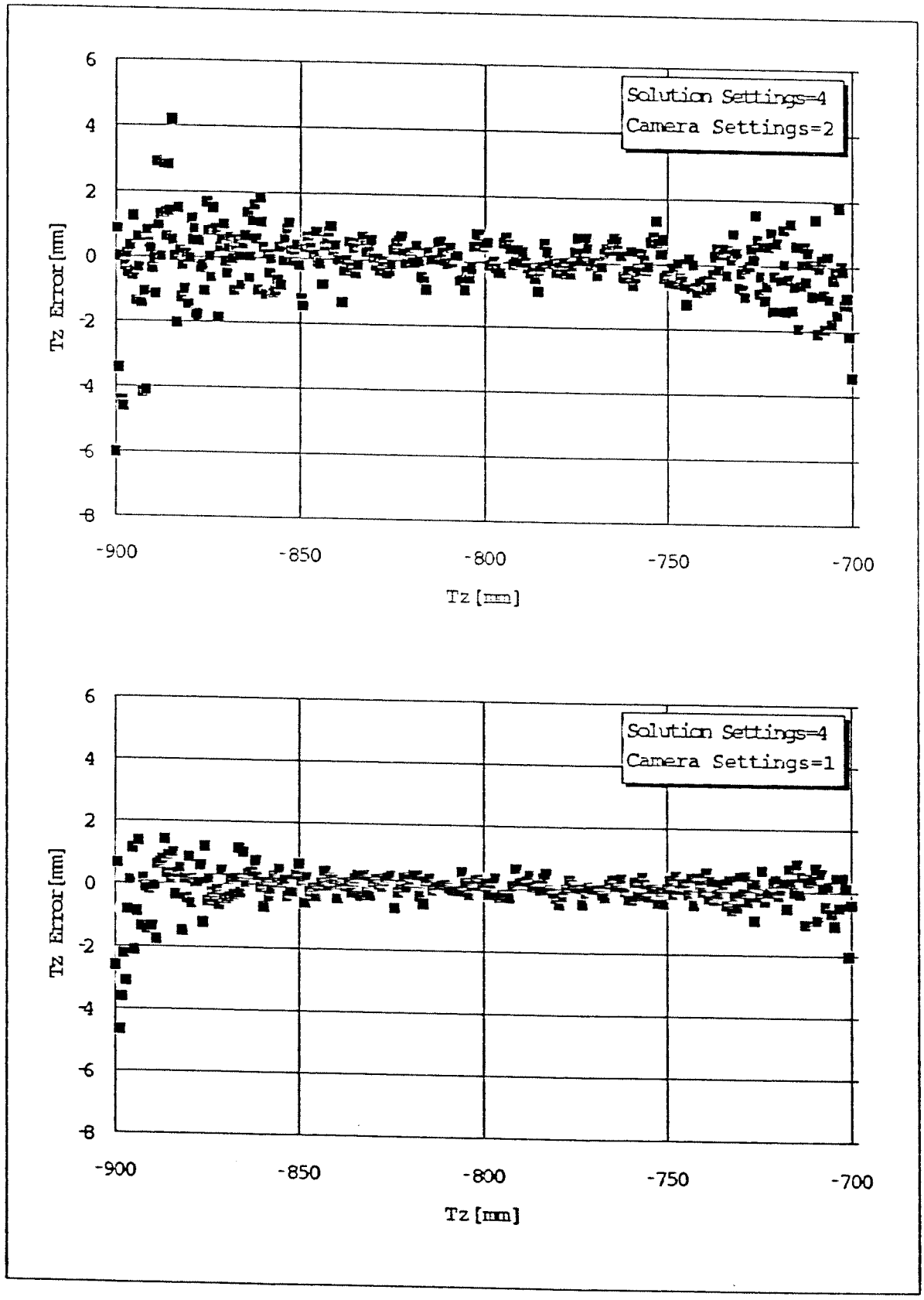


Figure 5.23: T_z Errors for Low and High Resolution Cameras

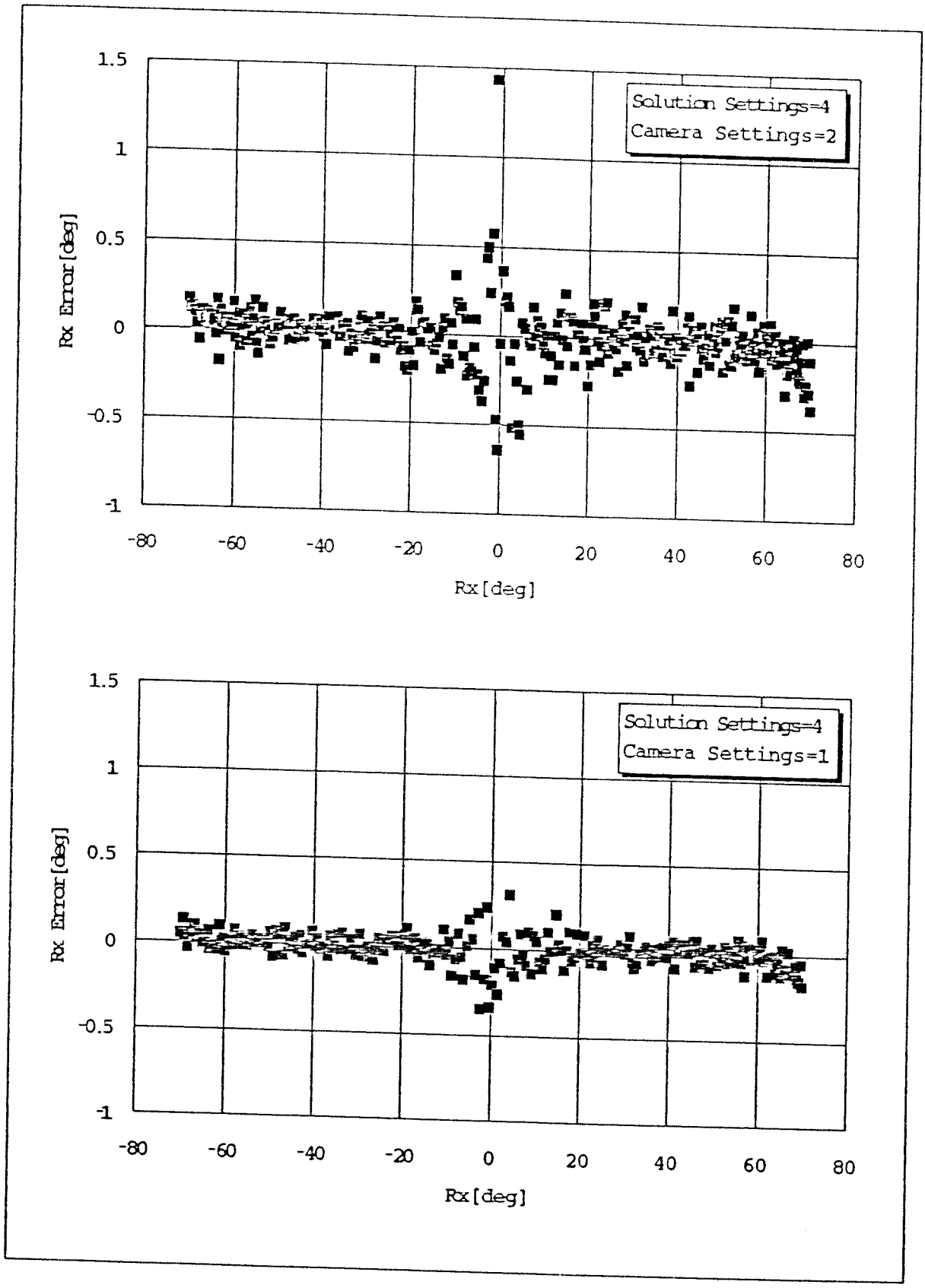


Figure 5.24: R_x Errors for Low and High Resolution Cameras

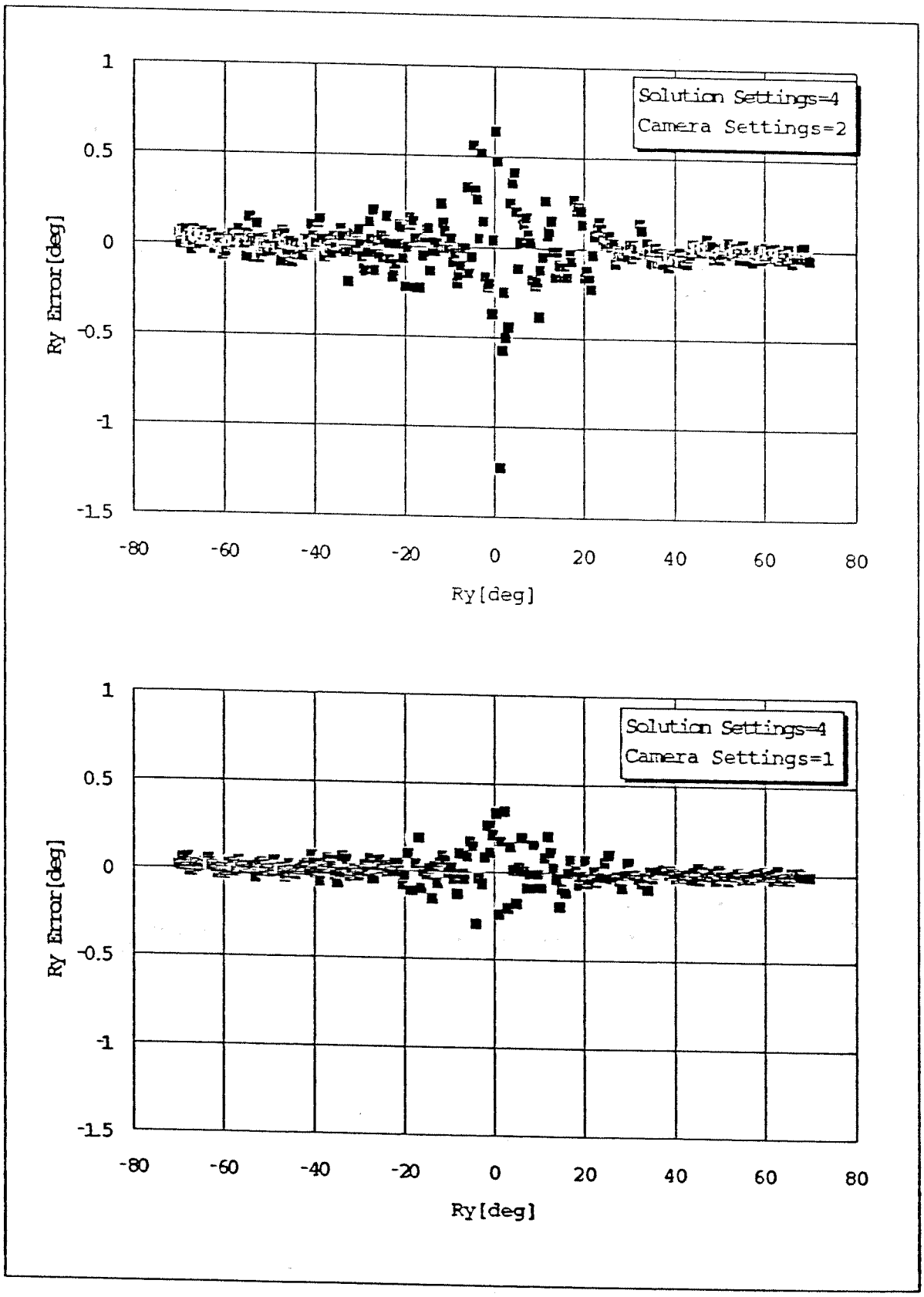


Figure 5.25: R_y Errors for Low and High Resolution Cameras

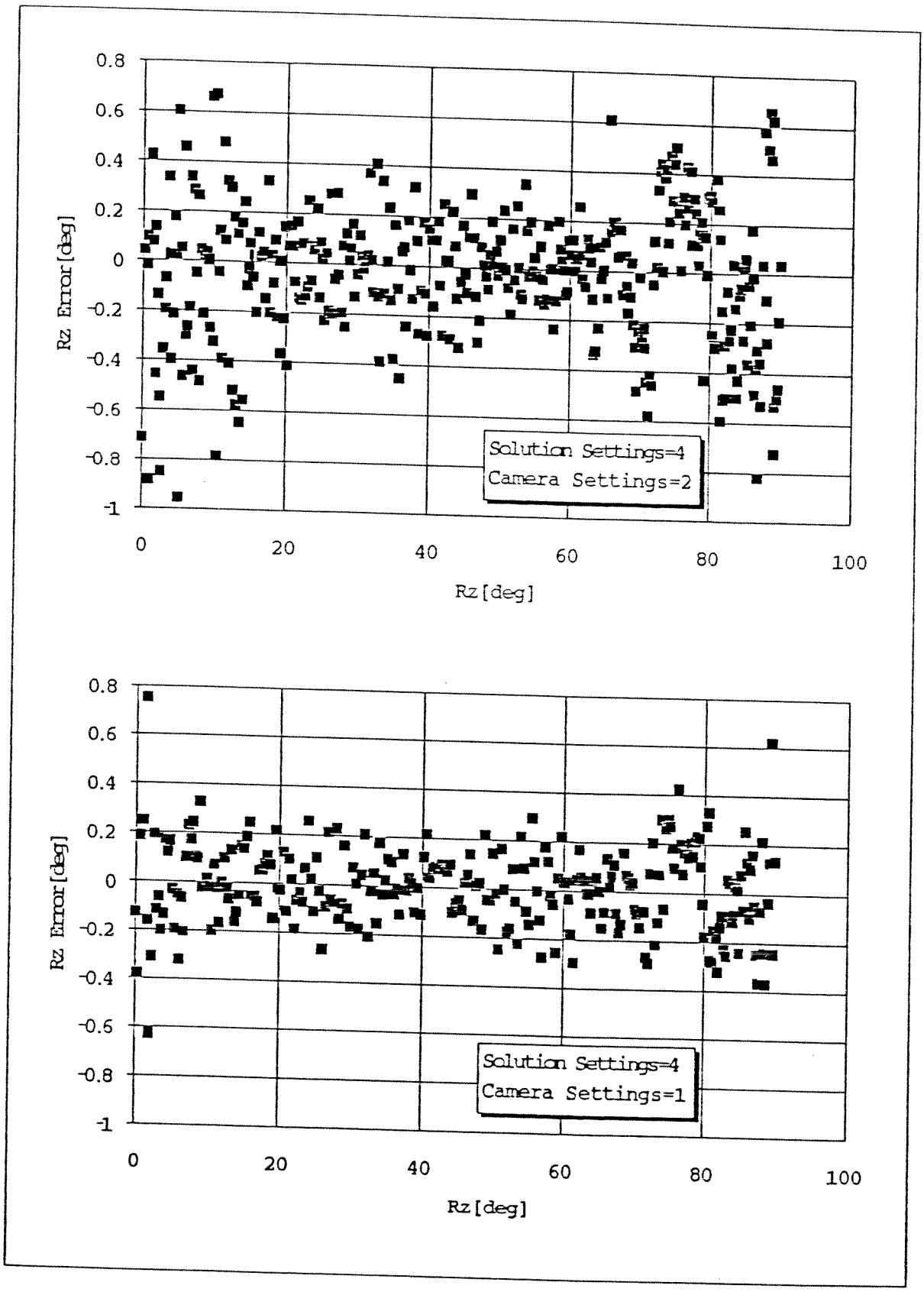


Figure 5.26: R_z Errors for Low and High Resolution Cameras

5.7 Target Parameters R, r_0 and r_f

The three parameters that define the target geometry are R, r_0 and r_f as previously shown in Figure 3.7. The main radius R can be selected according to the target distance. A large target image will result in a better reconstruction and the main radius should be selected accordingly. The following formula may help to the selection of R such that the image will cover approximately 10% of the sensor area:

$$R \approx \frac{d^{(\text{average})} S_s}{d_f} \sqrt{\frac{10\%}{2\pi}}$$

In the simulation, $S_s = 7.3\text{mm}$, $d^{(\text{average})} = 800\text{mm}$, $d_f = 28\text{mm}$ and $R = 25\text{mm}$ have been used.

The radius of the spots given by r_0 should be kept as small as possible, since the centroid coordinates of the image spots are used in the formulations. One should also consider that these spots should be observable in different target orientations. In the simulation program, $r_0 = 2\text{mm}$ have been used as a practical default setting.

The distance of the outer spot's center to the center of the main circle is denoted by r_f. As demonstrated previously, the value of r_f is not used in the reconstruction process. But one can say intuitively that the reconstruction of the in-plane rotation γ will be more accurate if r_f is larger. One thing to consider is that the spot should entirely lie inside the main circle and should not overlap the center spot. So

$$2r_0 < r_f < R - r_0$$

should be satisfied as a geometric constraint. And considering the possible orientations of the target plane, a good practical value for r_f can be given as:

$$r_f \approx (R + r_0)/2$$

In the simulation program, $r_f = 15\text{mm}$ has been used as the default value.

5.8 Problem of Partial Target Images

The methods that use a secondary target for configuration sensing, make actually the assumption that the target is totally visible in the scene. The

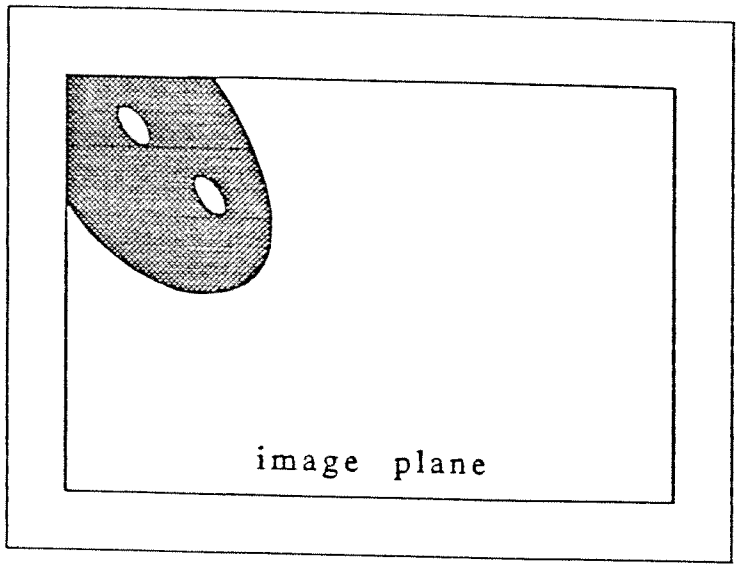


Figure 5.27: A Partial Target Image

methods that use target geometries composed only from spots like in [13] or [12] use this assumption inherently since all the spots must be visible.

The method proposed here uses a composite target where there are only two spots and one main circle. If the two spots are totally visible, the reconstruction process can be completed even if the main ellipse is partial. This is possible because pixel points are selected along the contour of the image ellipse and the curve-fitting can be performed on a partial contour. The only partial contours that have been investigated are due to target configurations leaving the target partially out of the field of view. The reconstruction of partial targets is only possible with the EqualArcs method for pixel selection and the pixels of the contour lying on the edges of the sensor array are not selected. Figure 5.27 depicts a situation where the reconstruction can be completed. The contour of the target image contains actually a part of the row and column that form the upper left corner of the image plane, however pixels should not be selected in that portion of the contour.

Partial target images may also occur due to the presence of other objects in the scene that may overlap the target. In such cases, the contour may be divided into several portions and the reconstruction would require the identification of the overlapping objects such that their contour pixels are not used for the curve-fitting operation. This subject requires a scene interpretation to be completed before the operation of the pixel selection algorithm. Figure 5.28 shows the simulation results for the Sequential Set 3 given in Table 5.11. The

Table 5.11: Sequential Set 3

	Start	Vel.[/sec]	Time[sec]
T_x[mm]	-35	0	50
T_y[mm]	-25	1	
T_z[mm]	-500	0	
R_x[°]	-70	2.8	
R_y[°]	70	-2.8	
R_z[°]	0	0	

target motion is selected such that approximately 25% of the target image falls out of the screen. The results for the curve-fitting on a partial target are satisfactory as long as the two inner spots are totally visible.

5.9 Target Recognition

The composite target can be recognized easily in the scene with the proper selection of target tissue. The target should be considered as a whole disk which has two small holes on it. This geometry can be cut from a high reflective surface material to provide the necessary contrast even for high rotation angles. The important properties of the target that can be used in the recognition process are as follows:

- the two spots lie inside the image ellipse
- the centers of the image ellipse and the central spot are very close to each other

The target image can be distinguished among other candidates using these features once all the ellipses and spots in the image have been segmented.

The lighting condition is also important for the recognition of the target. Although this study does not deal with structured lighting techniques, special lighting can be used to improve the target recognition. The basic requirement for a good lighting is to illuminate the reflective surface such that the rays are perpendicular to it and there are no shadows on the target.

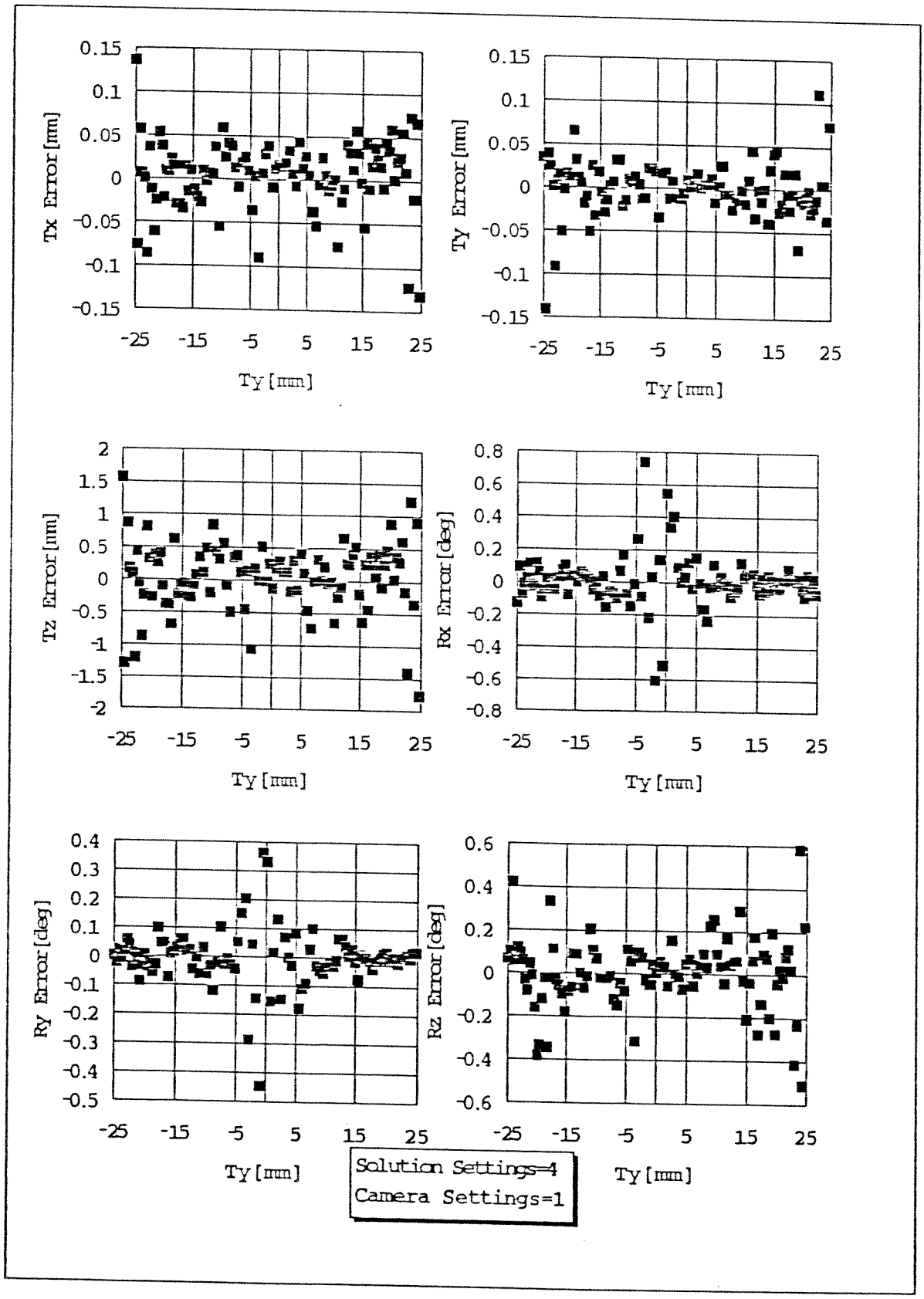


Figure 5.28: Simulation Results for Partial Target Images

The method presented is intended for the configuration sensing of stationary targets. However it can be used for motion tracking if the computation speed is high enough. The simulation program performs an average solution speed of 3 Hz when run on a 486DXII processor on 33 MHz. This value does not reflect the real performance of the algorithm since the program has not been developed to investigate the speed limitations. All kinds of image processing and recognition steps are not considered in the computations but on the other hand additional data operations are carried out for the purpose of error analysis. The results show that the algorithm can be used for motion tracking if the developed software is hardware implemented.

The non-iterative feature of the method is very important for a fast reconstruction. Even as a software, the algorithm can be used for tracking slow moving targets. If this model is used for motion tracking a by product will be the detection of errors due to sign conflicts. In that case, the continuity of motion can be used and the undesired effects of sign errors would be overcome. The target recognition will also be easier because the recognition algorithm will not have to seek all the image. Instead, it will search the target in a specific region by using the knowledge of its past motion.

Table 5.12 gives the solution speeds for different pixel selection types. The computation speed increases from top to bottom as the target contour becomes smaller. One should remember that Nested Set 1 generates target images close to a full circle, Nested Set 2 contains more general ellipse images and in Nested Set 3 the image ellipses are very thin. The EqualArcs10 method seems to be the fastest. In fact it works faster than a MaxMinCross method since it does not need many decision steps for operation. The computation speed decreases from left to right starting from the EqualArcs10 method as the number of selected pixels increases.

Solution speeds for different focus types are tabulated in Table 5.13. It is observed that the perfect or constant focus types work about two times faster than the iterative focus type. Remembering the power of the iterative focus in terms of reconstruction errors, it seems appropriate to use it in practical

Table 5.12: Solution Speeds in Hz for Different Pixel Selection Types

Pixel Selection	MMC	EA10	EA100	EA700
Nested Set 1	3.41	4.11	3.30	1.66
Nested Set 2	4.20	5.08	3.86	2.05
Nested Set 3	4.85	6.01	4.46	2.42

MMC: MaxMinCross, EA: EqualArcs

Table 5.13: Solution Speeds in Hz for Different Focus Types

Focus	P	C	I
Sequential Set 1	2.18	2.17	1.0

P: Perfect, C: Constant, I: Iterative

applications where d_0 cannot be obtained explicitly.

Finally Table 5.14 shows the solution speeds for the high and low resolution cameras corresponding to Camera Settings 1 and 2, respectively. The computation speed increases with increased pixel size and decreasing resolution. Considering the poor accuracy in terms of reconstructed variables, a low resolution camera cannot be suggested as a remedy to improve the speed characteristics.

Table 5.14: Solution Speeds in Hz for Different Sensor Resolutions

Sensor	High Res.	Low Res.
Sequential Set 2	2.63	3.55

CHAPTER VI

SUMMARY AND CONCLUSIONS

In this Chapter, the study will be summarized, the conclusions will be presented and finally directions for future research will be indicated.

6.1 Summary

An alternative approach to the stereo vision has been presented to reconstruct the configuration of rigid bodies by means of the image information of a passive secondary target. The secondary target is designed to be planar and rigidly attached to the object of interest. The system requires a single camera that supplies gray level images and an image processor for operation. The target recognition and subsequent image processing operations have not been the main interests of the research. It is assumed that a binary image is obtained by a proper thresholding and the contour of the target is not deteriorated. Then, the binary target image is used to convert the visual information into numerical data such that the configuration of the secondary target is totally defined in 3D. Therefore, the new algorithm enables the reconstruction of 6 parameters that define the configuration of the target.

In addition to the target configuration, the image plane distance, which is an internal parameter to the camera is also determined with few iteration steps. The iteration is performed by considering the target as a calibration mark during the solution process. The estimated image plane distance is only used in the formulations and is not intended for image restoration techniques such as deblurring. This means that the iterative solution will give satisfactory results only if the actual focus of the camera provides an unblurred image.

The main activity in this study was to develop a method and to test its theoretical accuracy limits. In the theoretical part, the purpose was to combine

existing methods to end up with a simple, non-iterative configuration reconstruction algorithm. For the simulation part, a computer program has been developed in Turbo Pascal. The simulation program has been developed as a tool for future research on the subject. For this reason, its functionality has not been restricted to the special requirements of this thesis. It can be extended easily to study other target geometries and solution algorithms developed accordingly. The data processing and data transfer operations which have been developed separately offer the possibility to make a detailed error analysis in Excel environment.

To realize the full reconstruction of the configuration parameters, a secondary target has been designed. The target is composed of 3 distinct geometrical features which are used in the reconstruction process at different levels.

1. A main circle of radius R, located at (0,0) on the target plane
2. A central spot of radius r_0, located at (0,0) on the target plane
3. An outer spot of radius r_0, located at (r_f,0) on the target plane

The central spot is for the determination of

- ϕ : vertical offset angle from the optical axis
- θ : horizontal offset angle from the optical axis

It is assumed that for a small circle, the shift between the true center and the image center, which is a consequence of perspective projection can be neglected. Thus by finding the centroid of the image of the central spot one can approximately detect the center of the target. Using the image coordinates of the target center and the related formulations, the offset angles ϕ and θ are calculated. The radius of the center spot is selected as $r_0 = 2$mm in order to justify the assumption above.

The main circle is used to determine the parameters

- d : target distance
- α : rotation angle about the x axis
- β : rotation angle about the y axis

A curve-fitting technique is applied by selecting pixels along the image contour of the main circle which is an ellipse. By using ϕ and θ, the selected pixels are projected onto a virtual image plane which corresponds to a fictitious camera whose optical axis points to the center of the target. The ellipse is represented in the virtual image plane coordinates by evaluating the normalized coefficients A, B, C, D and E. The number of selected pixels as well as the method of selection is determined by the user. With some mathematical manipulations, d, α and β are calculated using the coefficients A, B and C. Only the signs of the coefficients D and E have been used in the formulations since their magnitudes are small compared to others especially for an ellipse image which is nearly a full circle. The radius of the main circle which determines the target size is selected as $R = 25$mm. Therefore the secondary target is confined in a square area of 50mm $\times$ 50mm.

Finally,

- γ : the in-plane rotation

is detected using the outer spot. For this purpose, a second virtual image plane has been defined parallel to the target plane. It intersects the original image plane such that the center of the image of the outer spot is on the intersection line. This particular configuration makes easier the determination of γ by using a double argument arctangent function. The offset distance for the spot is taken as $r_f = 15$mm.

There are no theoretical constraints on the freedom of the target plane as long as the target is visible. However, highly rotated target configurations will cause numerical problems due to degeneracy. For this reason,

$$|\alpha| < 82^{\circ}$$

$$|\beta| < 82^{\circ}$$

should be satisfied as practical limits, for a target in the range

$$500\text{mm} < d < 1000\text{mm},$$

using a 28mm lens. One can expect that the practical limits for α and β will be further reduced if the target gets farther.

The variables ϕ, θ, d, α, β and γ are all internal to the reconstruction algorithm as they determine the configuration of the target plane with respect to the image plane whose location is a function of the lens setting. For a given lens setting however, it is possible to obtain a general solution by evaluating the target configuration with respect to the world coordinate system. The following is a list of the parameters that can be measured by using a single camera imaging system and the proposed algorithm:

1. T_x : target translation along x axis
2. T_y : target translation along y axis
3. T_z : target translation along z axis
4. R_x : target rotation about x axis
5. R_y : target rotation about y axis
6. R_z : target rotation about z axis
7. d_0 : image plane distance of the camera

It should be noted that R_x, R_y and R_z are the consecutive rotation angles for an RFB 123 sequence which is used for general applications.

Simulation results concerning the specific system settings have been compared by using the error graphs generated for different image sets. The

simulation results that have been included in the thesis can be summarized as follows:

- comparison for different pixel selection types
- comparison for different focus types
- comparison for low and high sensor resolutions
- an analysis on numerical errors for $\alpha \approx 0$ and/or $\beta \approx 0$
- an analysis for partial target images

The problem of motion tracking is also discussed by considering the relative solution speeds for different system settings. The solution speeds that are tabulated for comparison are not sufficient to estimate the real speed of the algorithm due to two important facts:

- all the operations related to image capture and image processing are not taken into account and the binary target image is generated analytically
- the software is run on a 33MHz IBM 80486 compatible PC

However, they give a relative idea about the solution speeds for different system settings.

6.2 Conclusions

The algorithm has been tested and it is proved that the method works properly within the practical limits. The simulation results are satisfactory in terms of reconstruction errors. For a camera located

- at the world coordinate system, having
- a focal length of 28mm and
- a resolution of 400×400 pixels

- where the pixel size is $13\mu\text{m} \times 13\mu\text{m}$

the following average reconstruction errors are obtained:

- $E_{T_x}^{(\text{average})} \approx 0.03\text{mm}$
- $E_{T_y}^{(\text{average})} \approx 0.03\text{mm}$
- $E_{T_z}^{(\text{average})} \approx 0.69\text{mm}$
- $E_{R_x}^{(\text{average})} \approx 0.10^\circ$
- $E_{R_y}^{(\text{average})} \approx 0.10^\circ$
- $E_{R_z}^{(\text{average})} \approx 0.25^\circ$

Basically there are two regions where the reconstruction errors diverge from the average values.

1. $|R_x|$ and/or $|R_y|$ close to and greater than 80°
2. $|R_x|$ and/or $|R_y|$ around 0°

Table 6.1 shows how the different configuration parameters are affected in these regions.

Other than the critical regions, the algorithm provides satisfactory results and can be used to detect the configuration parameters of a rigid body in an automated environment.

Table 6.1: Approximate Errors for Critical Regions

	$\lvert R_x\rvert, \lvert R_y\rvert \to 80^\circ$	$\lvert R_x\rvert, \lvert R_y\rvert \to 0^\circ$
$E_{T_x}^{(\text{critical})}$	$10 \times E^{(\text{average})}$	$E^{(\text{average})}$
$E_{T_y}^{(\text{critical})}$	$10 \times E^{(\text{average})}$	$E^{(\text{average})}$
$E_{T_z}^{(\text{critical})}$	$15 \times E^{(\text{average})}$	$E^{(\text{average})}$
$E_{R_x}^{(\text{critical})}$	$E^{(\text{average})}$	$20 \times E^{(\text{average})}$
$E_{R_y}^{(\text{critical})}$	$E^{(\text{average})}$	$20 \times E^{(\text{average})}$
$E_{R_z}^{(\text{critical})}$	$10 \times E^{(\text{average})}$	$E^{(\text{average})}$

The study is complete in the sense that all the configuration parameters and the image plane distance can be determined. On the other hand, there are practical points worth to study in the future which have not been within the scope of this thesis. One can cite them as follows:

- problems related to image capture and image processing
- determination of a threshold level for binarization
- segmentation and target recognition
- selection of a reflective target material
- optimization for faster operation

All the image processing operations that will be used until a binary target image is obtained should be studied. The image processing operations and binarization are important steps. More accurate results can be obtained if the digital contour data is close to its analytical representation.

The selection of target tissue is also an important fact as the target should provide the necessary contrast with the environment. This may become a problem especially for cases where the target rotations are high and the image ellipse is very thin.

The source code can be further optimized in terms of computation speed, because the prime interest in developing the simulation program was not to investigate the speed limitations. The method can be used for motion tracking of slow moving objects. For higher performance however, hardware implementation is necessary. As mentioned previously, it is hard to make an estimation of the real solution speed by using the simulation program.

The points that are not considered in this thesis will be studied in the continuation of the project. An optical test bench and a gray level imaging system are available and will be used to obtain experimental results. The theoretical results, and the outcomes of this thesis will be taken as a reference to improve the performance of the real system.

REFERENCES

[1] Lowe, D.G., 1987. "Three-Dimensional Object Recognition from Single Two-Dimensional Images", Artificial Intelligence, Vol. 31, pp. 355-395, 1987

[2] Horaud, R., 1987. "New Methods for Matching 3-D Objects with Single Perspective Views", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAMI-9, No. 3, May 1987

[3] Sheu, D.D., Bond, A.H., 1991. "A Generalized Method for 3D Object Location from Single 2D Images", Pattern Recognition, Vol. 25, No. 8, pp. 771-786, 1992

[4] Barnard, S.T., Fischler, M.A., 1982. "Computational Stereo" , Computer Surv. Vol. 14, No. 4, Dec. 1982

[5] Lee, C.W., Beom, H.R., Cho, H.S. 1993. "An Obstacle DetectionMethod for Mobile Robots Using Axial Stereo", Proceedings of the 12th Triennial World Congress of the International Federation of Automatic Control, Vol. 3, pp. 467-472, Jul. 18-23, Sydney, Australia

[6] Strat, T.M., Fischler, M.A., 1986. "One-Eyed Stereo: A General Approach to Modeling 3-D Scene Geometry", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAMI-8, No. 6, Nov. 1986

[7] Olgac, N., Gan, Z., Platin, B.E., 1989. "3D Reconstruction of Object Configurations by Hybrid Projection Analysis Using a Single Camera Image", From the Proceedings of the 1st National Conference on Applied Mechanisms and Robotics Vol. 1, Paper No. 89AMR-4C-6, Nov. 5-8, Cincinnati, OH, USA

[8] Platin, B.E., Gan, Z., Olgac, N., 1990. "3D Object Configuration Sensor Utilizing Single Camera Images", Presented at the Winter Annual Meeting of ASME, Paper No. 90-WA/DSC-22 Nov. 25-30, Dallas, Texas, USA

[9] Olgac, N., Craig, P.D., Platin, B.E., 1991. "Advancements in Utilizing Monocular Object Configuration Sensor", Paper No. 91-WA/DSC-30, Advances in Instrumentation ASME 1991

[10] Olgac, N., Craig, P.D., Rama, P., 1991. "On the Calibration and Sensitivity Aspects of a Monocular Object Configuration Sensor", Paper No. 91-WA/DSC-30, Advances in Instrumentation ASME 1991

[11] Han, M.H., Rhee, S., 1991. "Camera Calibration for Three-Dimensional Measurement", Pattern Recognition, Vol. 25, No. 2, pp. 155-164, 1992

[12] Sydow, P.D., Cooper, E.G., 1992. "Development of a Machine Vision System for Automated Structural Assembly", NASA Langley Research Center, Hampton, VA 23665-5225, Mar. 1992

[13] Kite, D.H., Magee, M., 1989. "Determining the 3D Position and Orientation of a Robot Camera Using 2D Monocular Vision", Pattern Recognition, Vol. 23, No. 8, pp. 819-831, 1990

[14] Fairchild Weston, 1989. CCD Sensors, Systems and Development Technology pp. 230-291, 1989

[15] Hornbeck, R.W., 1975. Numerical Analysis, Quantum Publishers Inc. 1975.

APPENDICES

APPENDIX A

BASIC TRANSFORMATION MATRICES

A point P in 3D space can be represented by a vector $\vec{r}_{ap}$ using a reference frame say F_a. The vector $\vec{r}_{ap}$ is independent of any reference frame however its components do depend on the reference frame selected. The end of this vector is at the origin O_a of the reference frame and its tip is at the point P. Generally a reference frame is a cartesian coordinate system defined by mutually orthogonal unit vectors $\vec{u}_1$, $\vec{u}_2$ and $\vec{u}_3$. As shown if Figure A.1, $\vec{r}_1$, $\vec{r}_2$ and $\vec{r}_3$ are the components of $\vec{r}_{ap}$ along the axes $\vec{u}_1$, $\vec{u}_2$ and $\vec{u}_3$ respectively. The column representation of $\vec{r}_{ap}$ in F_a can be expressed as

$$\vec{r}_{ap}^{(a)} = \begin{bmatrix} r_1^{(a)} \\ r_2^{(a)} \\ r_3^{(a)} \end{bmatrix}$$

where the elements are the magnitudes of the corresponding component vectors.

By giving an arbitrary rotation to F_a about O_a, a new reference frame F_b can be created. In this case F_b differs from F_a only in its orientation. The transformation from F_a to F_b is referred to as rotational transformation. Figure A.2 illustrates this situation where the angles θ_{ij} are defined as:

$$\theta_{ij} = \angle(\vec{u}_i^a \rightarrow \vec{u}_j^b)$$

In Figure A.3 the vector $\vec{r}_{ap}$ is introduced such that its end is placed at the origin. The vector can be represented in F_a and F_b as

$$\vec{r}_{ap}^{(a)} = \begin{bmatrix} r_1^{(a)} \\ r_2^{(a)} \\ r_3^{(a)} \end{bmatrix}$$

and

$$\vec{r}_{ap}^{(b)} = \begin{bmatrix} r_1^{(b)} \\ r_2^{(b)} \\ r_3^{(b)} \end{bmatrix}$$

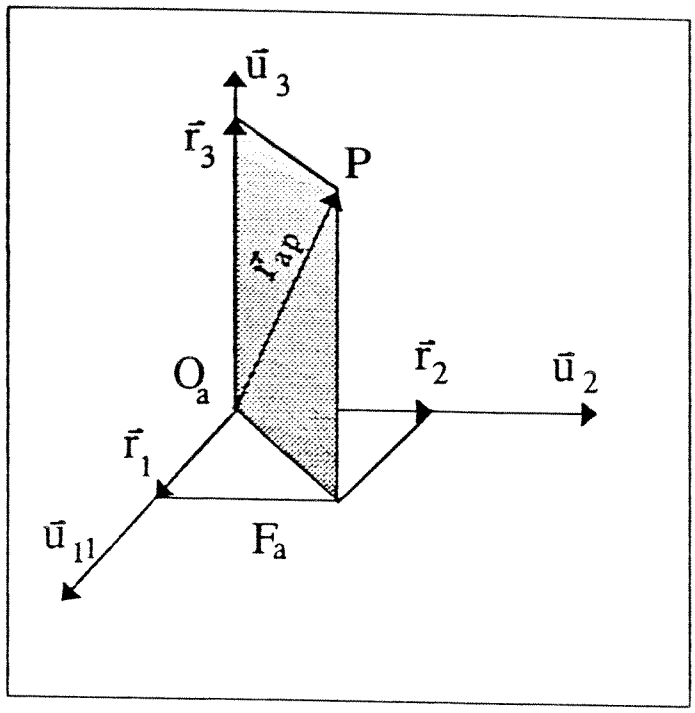


Figure A.1: Components of a Vector

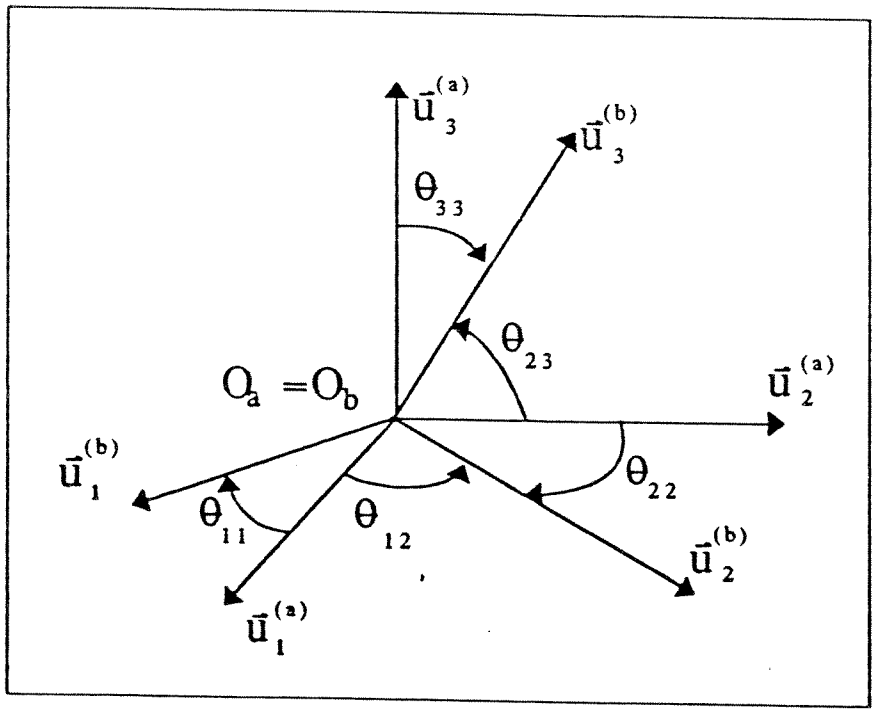


Figure A.2: Rotational Transformation

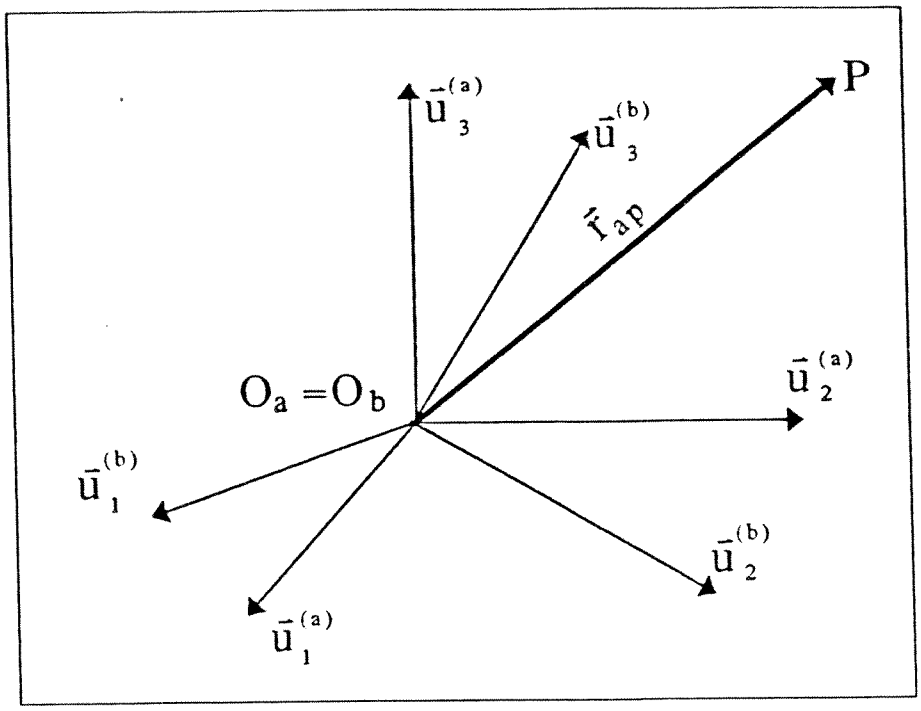


Figure A.3: Representation of a Point

respectively. Note that $\bar{r}^{(a)} \neq \bar{r}^{(b)}$ unless $F_a = F_b$. To relate $\bar{r}^{(a)}$ to $\bar{r}^{(b)}$ a general rotation matrix $\hat{C}^{(a,b)}$ can be defined such that

$$\bar{r}^{(a)} = \hat{C}^{(a,b)}\bar{r}^{(b)}$$

where

$$\hat{C}^{(a,b)} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

The elements of this matrix are called the direction cosines and they can be calculated as:

$$c_{ij} = \cos(\theta_{ij}) \quad i,j = 1 \ldots 3$$

In Figure A.2 only some of the angles were shown not to make graphical confusion. In fact there are a total of 9 direction cosines corresponding to 9 angles. However only 3 direction cosines are independent. Writing the general rotation matrix in terms of the direction cosines does not seem to be a convenient way because one has to determine at least 3 of them and then solve for the remaining 6 to complete the matrix. Instead the general rotation matrix $\hat{C}^{(a,b)}$ can be decomposed into consecutive basic rotations each realized about one of

the axes of the reference frame. Then using any RFB[1] sequence, $\hat{C}^{(a,b)}$ can be formed by the multiplication of three basic rotation matrices.

$$\hat{C}^{(a,b)} = \hat{R}_i(\alpha)\hat{R}_j(\beta)\hat{R}_k(\gamma)$$

The subscripts i, j and k denote sequentially the axis of the reference frame about which the rotation will occur. α, β and γ are corresponding rotation angles. An RFB sequence as its name implies is defined such that the next rotation is realized with respect to the current orientation of the reference frame. The basic rotation matrices about axes $\vec{u}_1$, $\vec{u}_2$, $\vec{u}_3$ are

$$\hat{R}_1(\theta_1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{bmatrix}$$

$$\hat{R}_2(\theta_2) = \begin{bmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix}$$

$$\hat{R}_3(\theta_3) = \begin{bmatrix} \cos\theta_3 & -\sin\theta_3 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Any arbitrary orientation of the reference frame can be composed by an RFB sequence. The selection of the RFB sequence can be done arbitrarily as long as the intermediate steps do not correspond to a physical configuration. The successive rotations can be described as

$$\underbrace{F_a \Longrightarrow}_{\hat{R}_i}\underbrace{F_n \Longrightarrow}_{\hat{R}_j}\underbrace{F_m \Longrightarrow}_{\hat{R}_k} F_b$$

where F_n and F_m are the intermediate frames.

The common RFB sequences are

- RFB 123(roll, pitch, yaw) used in general applications
- RFB 321(yaw, pitch, roll) used mostly in flight mechanics
- RFB 313(precession, nutation, spin) used in gyroscope dynamics and for celestial motions

[1]Rotated Frame Based

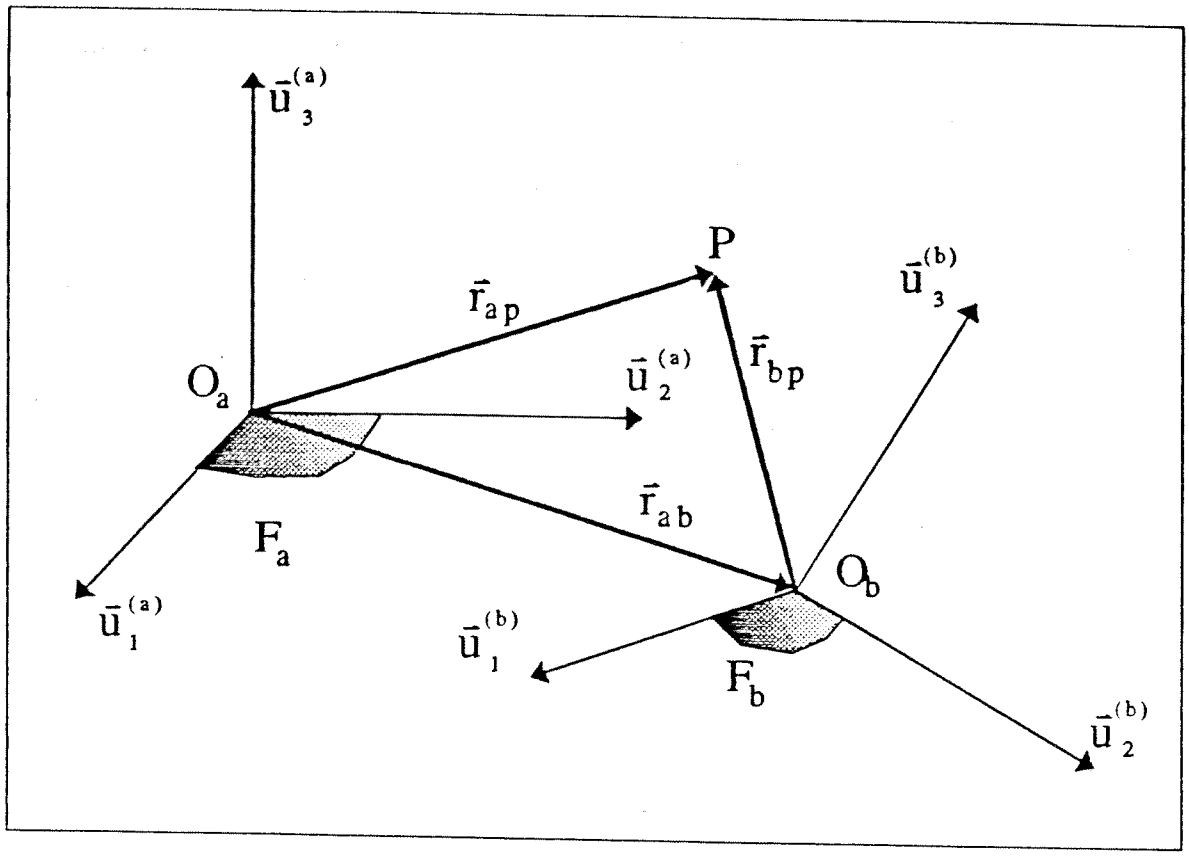


Figure A.4: General Transformation

- RFB 323(yaw, declination, yaw) an alternative to RFB 313 sequence used frequently in robotics

One important property of the general rotation matrix is that $\hat{C}^{(b,a)}$, the inverse transformation, can be obtained very easily as

$$\hat{C}^{(b,a)} = \left[\hat{C}^{(a,b)}\right]^{-1} = \left[\hat{C}^{(a,b)}\right]^{T}$$

Figure A.4 shows a more general case where in addition to the orientation, the origins of F_a and F_b are not coincident. Here again the point P can be represented in F_a and F_b but this time using vectors $\vec{r}_{ap}$ and $\vec{r}_{bp}$. The following vector equation holds for this case;

$$\vec{r}_{ap} = \vec{r}_{ab} + \vec{r}_{bp}$$

This is equivalent to write

$$\bar{r}_{ap}^{(a)} = \hat{C}^{(a,b)}\bar{r}_{bp}^{(b)} + \bar{r}_{ab}^{(a)}$$

in matrix form. Note that this equation is not homogeneous due to the presence of $\bar{r}_{ab}^{(a)}$ in the right hand side. However it can be artificially converted to a homogeneous equation. Adding the trivial equation $1 = 1$ the matrix form becomes

$$\begin{bmatrix} \bar{r}_{ap}^{(a)} \\ 1 \end{bmatrix} = \begin{bmatrix} \hat{C}^{(a,b)} & \bar{r}_{ab}^{(a)} \\ \bar{0}^T & 1 \end{bmatrix} \begin{bmatrix} \bar{r}_{bp}^{(b)} \\ 1 \end{bmatrix}$$

Now one can define the augmented matrices;

$$\bar{R}_{ap}^{(a)} = \begin{bmatrix} \bar{r}_{ap}^{(a)} \\ 1 \end{bmatrix}$$

$$\bar{R}_{bp}^{(b)} = \begin{bmatrix} \bar{r}_{bp}^{(b)} \\ 1 \end{bmatrix}$$

$$\hat{H}_{ab}^{(a,b)} = \begin{bmatrix} \hat{C}^{(a,b)} & \bar{r}_{ab}^{(a)} \\ \bar{0}^T & 1 \end{bmatrix}$$

and the matrix equation becomes homogeneous:

$$\bar{R}_{ap}^{(a)} = \hat{H}_{ab}^{(a,b)} \bar{R}_{bp}^{(b)}$$

$\hat{H}_{ab}^{(a,b)}$ is called the homogeneous transformation matrix. A homogeneous transformation matrix can be

- purely rotational; $\hat{H}_{ab}^{(a,b)} = \begin{bmatrix} \hat{C}^{(a,b)} & \bar{0} \\ \bar{0}^T & 1 \end{bmatrix}$ or
- purely translational; $\hat{H}_{ab}^{(a,b)} = \begin{bmatrix} \hat{I}_3 & \bar{r}_{ab}^{(a)} \\ \bar{0}^T & 1 \end{bmatrix}$

Like the general rotation matrix, a general transformation matrix can be formed by the multiplication of several homogeneous transformation matrices on a rotated-translated frame basis.

$$\hat{H}_{ab}^{(a,b)} = \hat{H}_{an}^{(a,n)} \hat{H}_{nm}^{(n,m)} \dots \hat{H}_{pb}^{(p,b)}$$

$$\underbrace{F_a \Longrightarrow}_{\hat{H}_{an}^{(a,n)}} \underbrace{F_n \Longrightarrow}_{\hat{H}_{nm}^{(n,m)}} \dots \underbrace{F_p \Longrightarrow}_{\hat{H}_{pb}^{(p,b)}} F_b$$

Here also it is convenient that each intermediate step represents a basic rotation or basic translation. For this purpose one can redefine the basic

rotation matrices in augmented form and also define 3 basic translation matrices. The augmented basic rotation matrices are:

$$\hat{R}_1(\theta_1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 & 0 \\ 0 & \sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\hat{R}_2(\theta_2) = \begin{bmatrix} \cos\theta_2 & 0 & \sin\theta_2 & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\hat{R}_3(\theta_3) = \begin{bmatrix} \cos\theta_3 & -\sin\theta_3 & 0 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The basic translation matrices are:

$$\hat{T}_1(d_1) = \begin{bmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\hat{T}_2(d_2) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\hat{T}_3(d_3) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

It should be mentioned that the inverse transformation can be obtained as:

$$\left[\hat{H}_{ab}^{(a,b)}\right]^{-1} = \begin{bmatrix} \left[\hat{C}^{(a,b)}\right]^T & -\left[\hat{C}^{(a,b)}\right]^T \bar{r}_{ab}^{(a)} \\ \bar{0}^T & 1 \end{bmatrix}$$

and

$$\hat{H}_{ba}^{(b,a)} = \left[\hat{H}_{ab}^{(a,b)}\right]^{-1} \neq \left[\hat{H}_{ab}^{(a,b)}\right]^T$$

APPENDIX B

ELLIPSE EQUATION IN CANONICAL COORDINATES

A square matrix which is real and symmetric can be diagonalized by an orthogonal transformation. The diagonalization of a real symmetric matrix is equivalent to reduce a quadratic form to a sum of squares.

The quadratic form can be expressed using a real symmetric matrix $\hat{M}$ as:

$$X^T \hat{M} X = \sum_{i,j=0}^{2} a_{ij} x^i x^j = 0$$

where X is the coordinate vector and

$$\hat{M} = \begin{bmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & 1 \end{bmatrix}$$

is the 3×3 coefficient matrix.

Let the same quadratic form be also represented by

$$\hat{M}_* = \hat{T}^T \hat{M} \hat{T} = \begin{bmatrix} A_* & 0 & 0 \\ 0 & C_* & 0 \\ 0 & 0 & F_* \end{bmatrix}$$

by means of the linear transformation $\hat{T}$, where $\hat{M}_*$ is a diagonal matrix.

If $\hat{U} = \hat{T}^{-1}$, the following statements will be true for the product $\hat{U}^T \hat{M}_* \hat{U}$:

1. $\det \left[\hat{U}^T \hat{M}_* \hat{U} \right] = \det \left[\hat{M} \right] = A_* C_* F_*$
2. $\det_{2\times 2} \left[\hat{U}^T \hat{M}_* \hat{U} \right] = \det_{2\times 2} \left[\hat{M} \right] = A_* C_*$
3. $\mathrm{trace}_{2\times 2} \left[\hat{U}^T \hat{M}_* \hat{U} \right] = \mathrm{trace}_{2\times 2} \left[\hat{M} \right] = A_* + C_*$

therefore the eigenvalues of the upper left 2×2 portion are A_* and C_*. When $\hat{U} = \hat{T}^{-1}$ and $\hat{M}_* = \hat{T}^T \hat{M} \hat{T}$ is substituted the product becomes;

$$\hat{U}^T \hat{M}_* \hat{U} = \left[\hat{T}^{-1}\right]^T \hat{T}^T \hat{M} \hat{T} \hat{T}^{-1}$$

which is nothing but $\hat{M}$ itself. As a result, the eigenvalues of the 2×2 portion of $\hat{M}$ should also be the same. Using this equality,

$$\begin{aligned} A_* &= \frac{A+C}{2} - \frac{\sqrt{(A-C)^2 + B^2}}{2} \\ C_* &= \frac{A+C}{2} + \frac{\sqrt{(A-C)^2 + B^2}}{2} \end{aligned}$$

can be obtained and finally F_* can be found using the first property cited above as:

$$F_* = \frac{\det\left[\hat{M}\right]}{AC - B^2/4}$$

APPENDIX C

INVARIANCE OF THE DEGREE

One important property of perspective projection is the so-called *invariance of the degree.* According to this, the degree of a curve lying on an arbitrary plane is preserved when it is projected to the image plane using the perspective transformation rules. It should be noted that this property does not hold for the cases of degeneracy where the image function of a curve may become either a line or a point.

The invariance of the degree can be proved as follows; Let (x_1, y_1) be the coordinate system of the arbitrary plane, and (x_2, y_2) be the coordinate system of the image plane. The coordinates of a point lying on the arbitrary plane will be expressed in the image coordinate system as (X_2, Y_2, Z_2). And using the perspective transformation, the image coordinates of the same point are obtained as;

$$\begin{aligned} x_2 &= \frac{d_0 X_2}{d_0 - Z_2} \\ y_2 &= \frac{d_0 Y_2}{d_0 - Z_2} \end{aligned}$$

When the expressions for X_2, Y_2 and Z_2 are substituted, the resulting expression is in the following form;

$$\begin{aligned} x_2 &= \frac{a_{11}x_1 + b_{11}y_1 + c_{11}}{x_1 + d_1 y_1 + e_1} \\ y_2 &= \frac{a_{12}x_1 + b_{12}y_1 + c_{12}}{x_1 + d_1 y_1 + e_1} \end{aligned}$$

It is observed that the expressions are bilinear functions of x_1 and y_1 where $a_{11}, b_{11}, \ldots, d_1, e_1$ are constants in terms of the transformation parameters.

The inverse of these equations are also in bilinear form;

$$x_1 = \frac{a_{21}x_2 + b_{21}y_2 + c_{21}}{x_2 + d_2 y_2 + e_2}$$

$$y_1 = \frac{a_{22}x_2 + b_{22}y_2 + c_{22}}{x_2 + d_2y_2 + e_2}$$

Now, let $f_1(x_1, y_1)$ represent a function on the arbitrary plane using

$$f_1(x_1, y_1) = \sum_{\substack{i=0 \\ j=0}}^{n} g_{ij}x_1^i y_1^j = 0$$

where g_{ij}'s are the constant coefficients. Considering the fact $i + j \leq n$ where n is the degree of the original curve, and substituting the expressions of x_1 and y_1, the image function

$$f_2(x_2, y_2) = (x_2 + d_2y_2 + e_2)^n \, f_1[x_1(x_2, y_2), y_1(x_2, y_2)] = 0$$

can be obtained after multiplying both sides by $(x_2 + d_2y_2 + e_2)^n$. In the last expression it is clear that the degree of the image curve is still n in terms of the image plane coordinates x_2 and y_2 which completes the proof of the invariance of the degree.

APPENDIX D

PSEUDO INVERSION

For a general matrix equation such as

$$\bar{v} = \hat{H}\bar{u} \tag{D.1}$$

where $\hat{H}$ is not a square matrix, there is a method to solve the equation for $\bar{u}$. The method requires the inversion of a non-square matrix and is called pseudo-inversion.

Let $\hat{H}$ be an $M \times N$ matrix, then $\bar{v}$ and $\bar{u}$ are columns of $M \times 1$ and $N \times 1$ respectively. If both sides of the Equation D.1 are multiplied by the transpose of $\hat{H}$ the equation becomes:

$$\hat{H}^T\bar{v} = \hat{H}^T\hat{H}\bar{u}$$

At this point it should be noted that $\hat{H}^T\hat{H}$ is an $N \times N$ symmetric matrix. If $\hat{H}^T\hat{H}$ is non-singular then it can be inverted. The condition for a real symmetric matrix to be non-singular is that all the eigenvalues should be different than zero.

If the non-singularity condition is satisfied, then Equation D.1 can be solved for $\bar{u}$ as follows:

$$\bar{u} = (\hat{H}^T\hat{H})^{-1}\hat{H}^T\bar{v}$$

The pseudo-inverse is denoted by $\hat{H}^-$ and is given as:

$$\hat{H}^- = (\hat{H}^T\hat{H})^{-1}\hat{H}^T$$

Some properties of $\hat{H}^-$ are:

$$\begin{aligned} \hat{H}^-\hat{H} &= (\hat{H}^T\hat{H})^{-1}(\hat{H}^T\hat{H}) = I_{(N\times N)} \\ \hat{H}^-\hat{H} &\neq \hat{H}\hat{H}^- \end{aligned}$$

The necessary conditions for $\hat{H}^T\hat{H}$ to be non-singular are:

- $M \geq N$
- $\text{Rank}[\hat{H}] = N$

which means that $\hat{H}$ should have at least N independent rows.

Although a similar pseudo-inverse can be defined for $M < N$, which means $\text{Rank}[\hat{H}] < N$, the solution for $\bar{u}$ will not be unique. In this case the definition of the pseudo-inverse matrix will be modified as follows:

$$\hat{H}^{-} = \hat{H}^{T}(\hat{H}^{T}\hat{H})^{-1}$$

and

$$\hat{H}\hat{H}^{-} = I_{(M\times M)}$$

A generalized inverse can be defined using the SVD[1] of $\hat{H}$. The SVD is given by:

$$\hat{H} = \hat{\Psi}\,\hat{\Lambda}^{1/2}\hat{\Phi}^{T}$$

and the generalized inverse $\hat{H}^{+}$ will be

$$\hat{H}^{+} = \hat{\Phi}\,\hat{\Lambda}^{-1/2}\hat{\Psi}^{T}$$

where $\hat{\Lambda}$ is an $r \times r$ diagonal matrix whose elements are the eigenvalues of $\hat{H}\hat{H}^{T}$. $\hat{\Psi}$ is an $M \times r$ matrix whose columns are the eigenvectors of $\hat{H}\hat{H}^{T}$. $\hat{\Phi}$ is an $N \times r$ matrix whose columns are the eigenvectors of $\hat{H}^{T}\hat{H}$. Finally the minimum norm generalized solution for $\bar{u}$ is given as:

$$\bar{u}^{+} = \hat{H}^{+}\bar{v}$$

[1]Singular Value Decomposition

EK 2

EXPERIMENTAL INVESTIGATION AND IMPLEMENTATION OF A 3D CONFIGURATION RECONSTRUCTION ALGORITHM FOR AN OBJECT USING A SINGLE CAMERA IMAGE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

ERCAN UMUT ACAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF MECHANICAL ENGINEERING

SEPTEMBER 1995

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. İsmail TOSUN
for Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Rüknettin OSKAY
Head of the Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Bülent E. PLATİN
Supervisor

Examining Committee Members:

Prof. Dr. Abdülkadir ERDEN (Chairman)

Prof. Dr. Bülent E. PLATİN

Prof. Dr. Turgut TÜMER

Prof. Dr. Neşe YALABIK

Assoc. Prof. Dr. Aydan ERKMEN

ABSTRACT

EXPERIMENTAL INVESTIGATION AND IMPLEMENTATION OF A 3D CONFIGURATION RECONSTRUCTION ALGORITHM FOR AN OBJECT USING A SINGLE CAMERA IMAGE

Acar, Ercan Umut

M.S., Department of Mechanical Engineering

Supervisor: Prof. Dr. Bülent E. Platin

September 1995, 133 pages.

In this study, an experimental investigation and implementation of a 3D configuration reconstruction algorithm using a passive secondary target are presented. The algorithm used was developed in earlier studies, and its theoretical limits were determined. This algorithm can be used in autonomously guided vehicles, camera calibration, manipulator guidance and other applications where the use of a passive secondary target is appropriate for configuration sensing.

A circle with two internal spots is used as a passive secondary target. An optical bench and a digital imaging system using a camera are constructed to make controlled environment experiments. A computer program is generated to extract the contour data and centroids of the spots from a digital image for the algorithm, and to compute the 6 configuration variables.

In an earlier study, it was proved that the configuration of a rigid body can be reconstructed in 6 degrees of freedom using a secondary target of known geometry with a single camera image. The theoretical limits of the algorithm found in that study are verified by the experimental results. The experimental

investigation shows that the reconstruction error is low enough to compete with other algorithms used for the same purpose.

Keywords: Monocular Vision, Camera Calibration, Machine Vision, Configuration Sensors, Secondary Target.

ÖZ

3 BOYUTLU UZAYDA CİSİMLERİN KONUMLARININ TEK KAMERA GÖRÜNTÜSÜ KULLANILARAK BELİRLENMESİ İÇİN GELİŞTİRİLEN METODUN UYGULAMASI VE DENEYSEL İNCELEMESİ

Acar, Ercan Umut
Yüksek Lisans Tezi, Makina Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Bülent E. Platin

Eylül 1995, 133 sayfa.

Bu çalışmada, 3 boyutlu uzayda cisimlerin konumlarını pasif ikincil hedefler kullanarak belirleyen bir yöntemin uygulaması ve deneysel incelemesi sunulmuştur. Kullanılan yöntem daha önce yapılan çalışmalarda geliştirilmiş ve kuramsal sınırları bulunmuştu. Bu ölçüm yöntemi, kamera kalibrasyonu, otomatik montaj, gezer robot ve manipulatör güdümü ve pasif ikincil bir hedefin kullanılmasında engel olmayan diğer uygulamalarda kullanılabilecektir.

Pasif ikincil hedef olarak, içinde iki tane benek bulunan bir daire kullanılmıştır. Kontrollü deneyler yapabilmek için tek kamera kullanan sayısal görüntüleme ve optik ölçme sistemi kurulmuştur. Yöntemde kullanılan çeper bilgisini ve beneklerin merkezlerini, sayısal görüntüden elde etmek ve 6 tane konum belirleme parametresini hesaplamak için, bir bilgisayar programı yazılmıştır.

Önceki bir çalışmada geometrisi bilinen pasif ikincil bir hedefin monte edileceği, 6 serbestlik derecesi olan, herhangi bir cismin konumunun tek kameralı bir görüntü sistemi kullanılarak belirlenebileceği kanıtlanmıştır. Yöntemin kuramsal sınırları deney sonuçları ile doğrulanmıştır. Deney sonuçları, ölçüm hatalarının aynı amaçlarla geliştirilmiş diğer yöntemlerle rekabet edebilecek kadar düşük olduğunu göstermiştir.

Anahtar Kelimeler: Monoküler Görüntü, Kamera Kalibrasyonu, Yapay Görüntü Algılama, Konum Algılayıcıları, İkincil Hedef.

ACKNOWLEDGEMENTS

I am grateful to my supervisor, Prof. Dr. Bülent E. Platin, the person with whom, working was a pleasure during the preparation of this study.

I would like to express my thanks to Prof. Dr. Ömer Anlağan for his valuable helps about the construction of the optical bench.

Realization of this thesis is supported by many people. Among them I would like to thank to Hüseyin Ütüklerli who gave me the first lessons on LaTeX. For his helps in the measurement of base plate of the optical bench, thanks go to Özgür Ünver. I am grateful to Cenk Güler for his ideas and suggestions. Special thanks go to Evrin B. Erdem, for her moral support and help in typing of the manuscript. I especially thank to Gökmen Mahmutyazıcıoğlu who let me to use his DeskJet printer.

I would like to express my gratitude to Alpar Kılınç whom to work with was a great pleasure.

Finally, the greatest thanks go to my family for their never ending support and reliance throughout my education.

This thesis is supported by TÜBİTAK-MODİSA.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

d_0	:	image plane distance
d	:	target distance
d_x	:	distance of the second virtual image plane
F	:	reference frame
O_f	:	focal point
S_s	:	diagonal sensor size
$O_c x_c y_c z_c$	:	the camera coordinate system
$O_o x_o y_o z_o$	:	the optical coordinate system
$O_w x_w y_w z_w$	:	the world coordinate system
$O_i x_i y_i z_i$	:	the image coordinate system
$O_{v1} x_{v1} y_{v1} z_{v1}$	:	the first virtual image coordinate system
$O_{v2} x_{v2} y_{v2} z_{v2}$	:	the second virtual image coordinate system
x_p, y_p	:	coordinates of the principal point
x_s, y_s	:	image coordinates of the outer spot's center
x_0, y_0	:	image coordinates of the target's center
$\hat{H}_{ij}^{(i,j)}$	:	homogeneous transformation matrix from F_i to F_j
$\hat{R}_x(), \hat{R}_y(), \hat{R}_z()$	:	basic rotation matrices
$\hat{T}_x(), \hat{T}_y(), \hat{T}_z()$	:	basic translation matrices
X, Y, Z	:	coordinates of a point
α	:	rotation about x axis
β	:	rotation about y axis
γ	:	rotation about z axis
ϕ	:	vertical offset angle
θ	:	horizontal offset angle
N	:	number of selected pixels
T_x	:	translation along x axis
T_y	:	translation along y axis
T_z	:	translation along z axis

R_x : rotation about x axis

R_y : rotation about y axis

R_z : rotation about z axis

A, B, C, D, E : normalized ellipse coefficients

A_*, C_*, F_* : canonical ellipse coefficients

R : main target radius

r_0 : spot radius

r_f : spot offset

CHAPTER I

INTRODUCTION

Artificial intelligence techniques are being used extensively in high technology industrial applications. Sensing is the major part of artificial intelligence that is used in industrial applications. In order to control a process or an event by utilizing intelligence, what is going on must be first sensed. Sensing of the environment in which the action will take place is a prerequisite especially in robotic applications. These applications include manipulator guidance, autonomous vehicles, assembly lines, part inspection, gauging, etc. In these applications, information about the environment is gathered either by a human operator or sensory equipment. For example, in industrial manipulators all the working environment is taught to the controller by an operator beforehand. Any change in the environment must therefore be retaught. That is, the system cannot adapt itself to changes without using human intelligence.

Human intelligence is fed by his sensing organs. The most important sensing organ of humans is the vision. Because the majority of information about the human environment is gathered, in a concise form, by vision. We can locate and use the things around us by using our eyes. We can estimate the distances of parts by using stereo vision capabilities of our eyes.

In robotic applications like manipulator and autonomous vehicle guidance, configuration [1] information about the parts in the application area is important. This kind of information can be obtained by using different types of sensors.

[1] Refers to position and orientation

In this study one of the machine vision applications, namely, finding configuration by using a single camera is studied. This study is a continuation of a Master's Thesis "3D Position Reconstruction of Rigid Bodies Using Single Camera Images" which was completed by Alpar Kılınç in the Mechanical Engineering Department of the Middle East Technical University [1].

1.1 Description of the Problem

Suppose a robotic manipulator is guided by computer vision to grasp an object placed on a work bench. In order for the manipulator to determine the configuration of the object from the two dimensional images, calibration of the system which is composed of the manipulator, the work bench and the camera must be made beforehand. The coordinate system in which the manipulator is guided must be defined with respect to that in which the object is placed. The imaging plane of the camera must be positioned with respect to the manipulator as well. The same is true for a vision system for an autonomous vehicle which determines its configuration from an image (Figure 1.1). Consequently, the problem is to find the configuration of an object in the workspace with respect to the camera's sensor plane and also the configuration of the camera's sensor plane with respect to its base frame.

1.2 Overview of the Solutions of the Problem

To solve the above problem mainly two techniques are being used:

- stereo vision
- mono vision

Stereo vision is the simulation of the humans' vision system. Two cameras are used to view a given object. These cameras are oriented such that

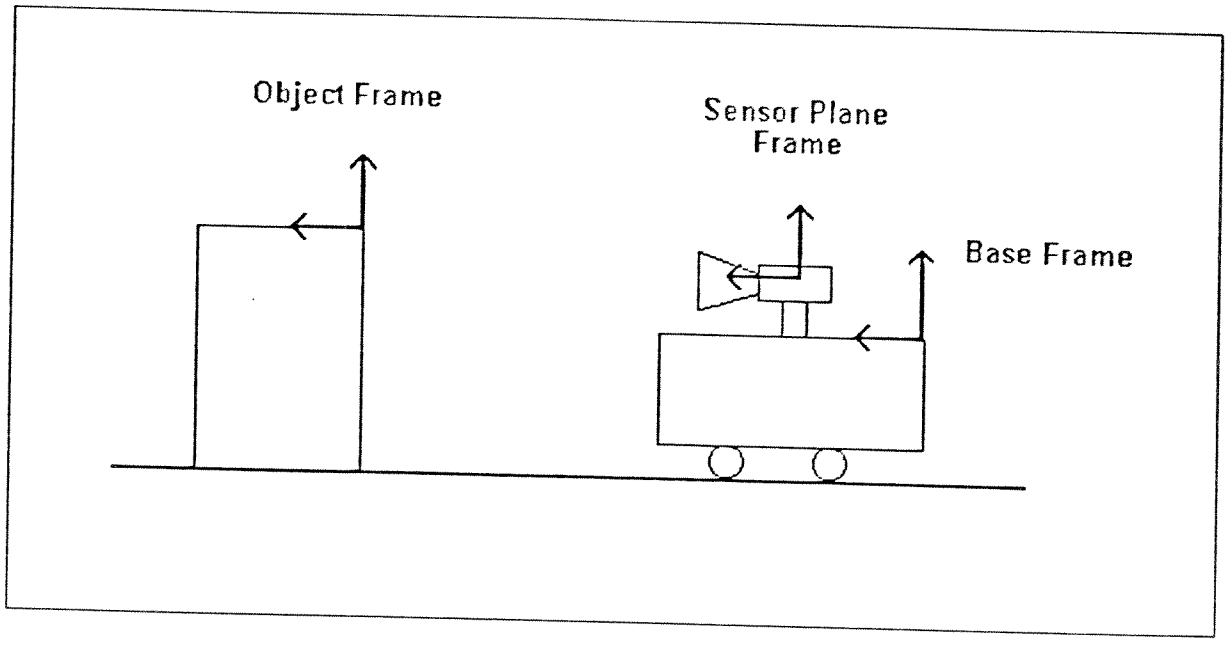


Figure 1.1: Calibration Frames

the object being viewed appears in the image planes of both cameras. The configuration of the object point can be determined by using triangulation [7]. Triangulation can be explained as; given the image points corresponding to a particular object point, the object point can be determined by finding the intersection of the two image point rays as shown in Figure 1.2.

In stereo vision the following points must be taken into account:

1. The allowed or preferred orientations of the cameras.

2. Exact correspondence between points in both image planes. The location of the image point corresponding to the object point being measured must be known exactly in one image plane and constrained in the other image plane. This image point correspondence data is perhaps the most difficult information to autonomously extract.

Since two cameras are used in stereo vision, they must be synchronized. Hence, one frame grabbed from one camera must match to the frame grabbed from the other camera in time. This need of synchronization adds to the cost of hardware.

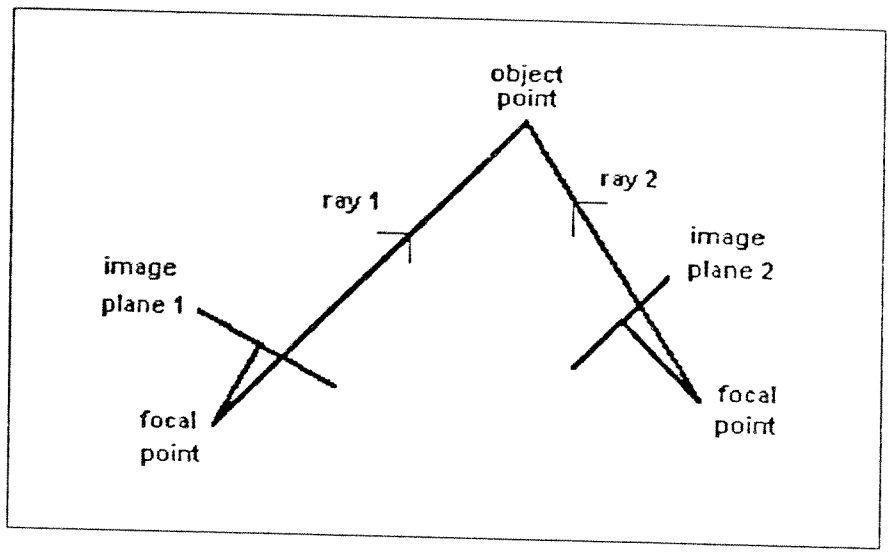


Figure 1.2: Triangulation

Image point correspondence information may be obtained by using special projected light patterns or time modulated projected rays or patterns [7]. This information can be obtained such that the object being imaged contains sufficient structure, which is reflected in the image plane as a set of features such as edges, common regions, or unique shapes.

Mono vision is achieved by using a single image of the object from a single camera. Since the depth information cannot be obtained from a single image due to the nature of the problem, a preknown information about the object is required. This preknown information can be obtained by attaching a secondary target onto the object. The reconstruction of the configuration variables are then achieved by using the image of this secondary target. Sydow and Cooper [11] have used a rectangle having five internal spots to reconstruct only the position of the target. In [10] all configuration parameters are reconstructed by using a rectangle of known size as the secondary target, but the accuracy of the method is poor.

Olgaç et al. [13] used a circular secondary target and orthographic projection model but also made the assumption that the target center lies on the optical axis of the camera. Besides this constraint, the use of a circular target

makes the determination of the in-plane rotation, inherently impossible.

The method used by Platin [14] and Olgaç [15] is similiar except that the orthographic projection model is replaced by a perspective projection model. The constraints on the target configuration and the requirement of a priori knowledge of some configuration variables are the major disadvantages of these algorithms. In [1] this method is further improved by addition of two internal spots to the circular secondary target such that 6 configuration parameters can be reconstructed.

The method developed in [1] has been implemented and tested in this work. A digital imaging system is constructed to implement the method and experimental work is carried out to test the theoretical limits of the algorithm obtained in [1] using an optical bench.

1.3 Camera Calibration

Camera calibration is the process of determining internal geometric and optical characteristics (internal parameters) of a camera and/or 3D position and orientation of a camera relative to a certain world coordinate system (external parameters).

Internal parameters are mainly the focal length, image center, focal point, lens distortion and image scanning parameters. External parameters are the six configuration variables, 3 rotational and 3 translational, that define the location of an object with respect to the camera's sensor plane.

There are different approaches to determine these parameters. In some of them only internal parameters can be found while in other methods external parameters can be determined in addition to some of the internal parameters.

The standard methods of camera calibration for machine vision use sets of points with known position in some world coordinate frame and their

corresponding image coordinates. These points are called feature points. The statement of the calibration problem is the following: given a set of feature points with known world coordinates, and their location in the image, find the external and internal camera parameters that will best map the feature points to their image points.

Tsai [8] presents a method for 3D camera calibration in the presence of radial distortion. First, they assume that the image center is at the center of the image buffer and find the external camera parameters and the focal length and then find the radial distortion parameter. In [12] a two-step calibration technique is proposed. In the first step, the parameters are estimated using a closed form solution based on a distortion free camera model. In the second step, the parameters estimated in the first step are improved iteratively through a nonlinear optimization taking into account lens distortions both radial and decentering. Both methods, published in [8] and [12] result in very accurate 3D measurements.

Some methods use geometrical properties of objects. In [9], in order to find the principal point and the focal length, spheres are used with the fact that the image of a sphere under perspective projection is an ellipse whose axis of least inertia is on a line passing through the principal point. The radial lens distortion parameter is estimated by using straight lines and their distorted images [9]. The correction parameters which would best straighten out the lines in the image are determined.

To find the image center, several methods have been tested in [4]. A taxonomy of 16 different image centers and procedures for measuring them are presented. In this work, among them the numerical center of the frame buffer is taken to be the image center.

1.4 Proposed Method for The Solution

Considering extra hardware requirements and difficulties in image point correspondence, a new algorithm is proposed in [1]. The configuration parameters, in other words, external camera parameters are solved by using a secondary target and a single camera. A planar circular target with two internal spots has been used in the reconstruction algorithm. The internal camera parameters are assumed to be known. The only information required about the shape of the target is radius of the circle and coincidence of one of the spots with the center of the circle. The main advantages of the new algorithm are as follows:

- monocular vision
- simple non-iterative solution
- capability of reconstructing 6 configuration parameters, uniqueness of the solution
- better accuracy compared to similar algorithms
- use of gray level images
- no necessity for structural lighting
- no point to point correspondence or knowledge of the world coordinate points of the target

1.5 Objectives and the Scope of the Study

In this study, the implementation of the algorithm developed in [1] and testing its theoretical limits are aimed. The implementation of the algorithm is targeted by using a digital imaging system and image analysis techniques. To test the validity of theoretical limits of the algorithm developed in [1] a series of experiments are planned. To carry out controlled experiments, the design and

construction of an optical bench are considered to be within the scope of the study.

For the digital imaging system two lenses with different specifications, a CID camera and an image processing card to obtain the digital representation of the image of secondary target have been used. The optical bench is constructed by using available positioning components from the off-the-shelf optical equipment. By means of this set-up the readouts and reconstructed values can be compared.

In this study, the basic assumptions made are :

- the camera is ideal[2]
- lightning condition of the scene provides a clean image
- image center is the numerical coordinates of the frame buffer
- the secondary target is planar and its geometry is perfect
- lens distortion has no considerable effect on the image
- optical axis of the camera is perpendicular to the sensor plane of the camera

and some important features are:

- recognition of the secondary target is performed manually
- simple threshold binarization is performed to obtain binary images
- no other image processing techniques such as convolution is required since the imaging environment is a controlled environment

[2]The ideal camera is described [1]

1.6 Outline of The Study

The components of a digital imaging system are presented in Chapter 2. Detailed definition of an optical system which is used to map 3D scene onto 2D image plane is given first. The sensors which are used to convert 2D scene information into electrical signals are explained in this chapter. Finally, how an image can be represented on a computer is given by the introduction of the frame grabber.

In Chapter 3, the algorithm, its theoretical limits and simulation results obtained in [1] are presented.

In Chapter 4, the experimental set-up used to perform experiments is explained. The description and technical specifications of optical system, camera and frame grabber are given in this chapter.

Image processing techniques and image analysis which are used in experiments are explained in Chapter 5. The algorithms used to extract the feature points are presented also in this chapter.

The calibration of the set up and the procedures followed in performing measurements are given in Chapter 6. The discussion and evaluation of the results are also given in this chapter.

Chapter 7 presents the outcomes of the study and guidelines for future studies.

The outputs of the computer program are given in tabular form in Appendix A. In Appendix B, the technical drawing and measurements of the base plate are given.

CHAPTER II

DIGITAL IMAGING USING A CAMERA

2.1 General

In a digital imaging system which uses a camera as a sensor to sense the scene, basically the followings must be present:

- An Optical System to produce an image of a 3-dimensional scene on a 2-dimensional plane,
- A Sensor to convert the 2-dimensional scene information into electrical signals,
- A frame grabber to convert electrical signals into digital information in order to make computations on the digital image of the scene by using a computer.

In this chapter, the above elements of a digital imaging system will be explained as a compilation of information gathered from various sources [2], [3], [4], [5].

2.2 Optical System

An optical system is used to project an object to the sensor. A typical optical system consist of a single lens or a group of lenses. Lenses are characterized

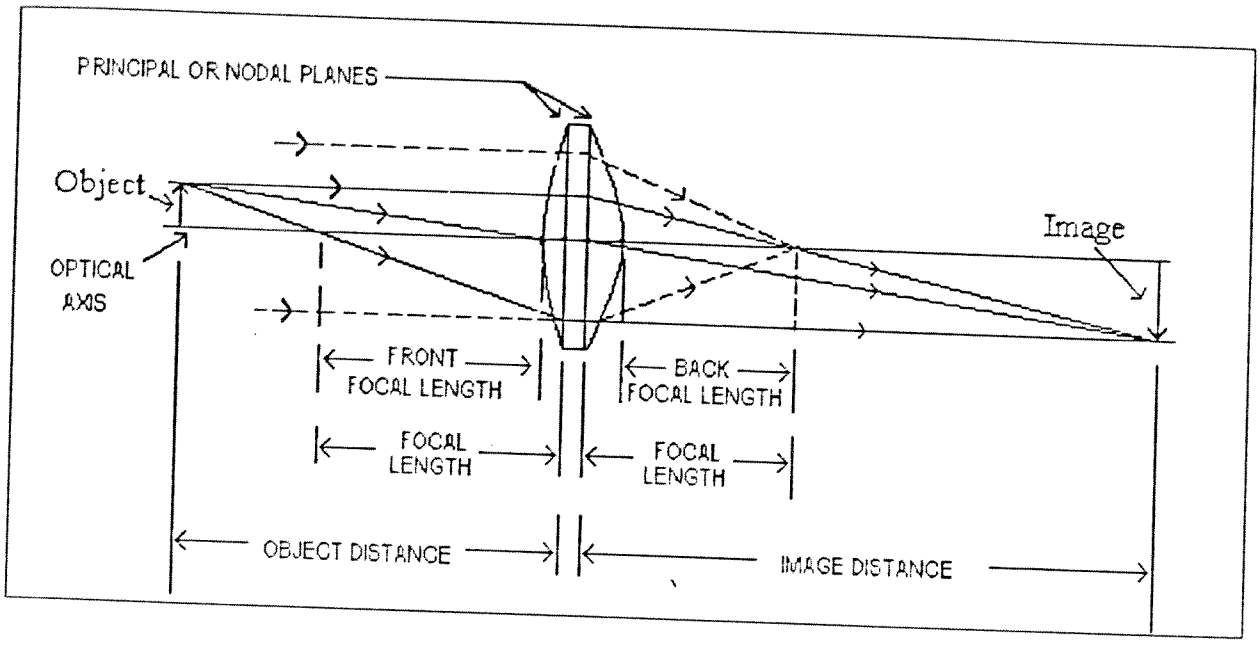


Figure 2.1: Basic Parameters of a Simple Lens

by their focal lengths, f-numbers and depth of fields. In Figure 2.1 an optical system is shown [5].

The focal point of a lens is the point at which the lens will cause a set of parallel rays (rays coming from infinity) converge to a point.

The plane at which the parallel rays appear to bend from being parallel toward the focal point is called the principle plane. A lens which is other than a theoretical "thin lens" will typically have multiple principle planes. The distance from the final principle plane, which is not necessarily located at the back of the lens (nor is it necessarily behind the first principle plane), to the focal point is the focal length of the lens. Often a back focal length of a lens is specified, which is the distance from the rear element of the lens to the focal point.

The f-number of a lens is the focal length divided by the diameter of the aperture (the limiting aperture) of the lens. For example, a lens with a 50 mm focal length and 10 mm aperture diameter will be an f/5 lens. The f-number is useful determining the relative illuminance at the image. The illuminance changes as the area of the aperture, generally indicated on the photographic lenses as f-stops, changes of a factor of two in illuminance. Since the f-number is a function

of the inverse of the lens aperture diameter, the illuminance will change as the inverse of the f-number squared. Therefore, there is a change of about a factor of 2 in illuminance from f/8 (squared is 64) to f/11 (squared is 121 or about 64×2).

The focal length and the f-number of any commercial lens is generally given right on the lens barrel. Given this information, the first step to presenting a usable image is the application of the basic imaging equations.

$$\frac{1}{lens\ focal\ length} = \frac{1}{image\ distance} + \frac{1}{object\ distance}$$

$$lateral\ magnification = \frac{image\ distance}{object\ distance}$$

$$longitudinal\ magnification = (lateral\ magnification)^2$$

By the help of these simple equations, it is possible to follow the Figure 2.1 through the optical system and determine its location and size. By doing this, one can also determine what the field of view of a given system will be by calculating what the image size of the sensor would be at the subject plane.

The depth of field is the distance along the optical axis at the subject through which the object can be located and still be properly imaged. The depth of focus is the distance along the optical axis of the sensor through which the image will be in focus. The distance from the image plane to the minimum focus distance plus the distance from the image plane to the maximum focus distance is the total usable depth of field of the viewing system for the specified feature size. This means that points on the object, which are within this depth range of distances from the camera will be in focus.

In an ideal lens system the optical axis is defined as the straight line passing through all of the radii of curvature of the lens elements [4]. The rotational symmetry of the system naturally leads to imaging properties that are symmetric around the optical axis. In a real lens system things are not so simple. For a

simple lens element like that shown in Figure 2.2 there are actually two axes of symmetry, one optical and one mechanical. The optical axis of the lens is defined as the straight line joining the centers of curvature of the two surfaces of the lens. The mechanical axis of the lens is determined during manufacture by the center line of the machine used to grind the lens' edge. Ideally the optical and mechanical axes would coincide. Practically they do not. The tolerance between them is called decentration.

In compound lens, two or more lens elements are aligned and mounted together to form the complete lens, as illustrated in Figure 2.3. Ideally all elements would be aligned along a common optical axis, but this is not always possible given the decentration in the individual elements. The cumulative effect of the mechanical tolerances for the lens elements is that there is no "ideal" optical axis for the lens. In fact the decentration and misalignment can produce tangential lens distortion and asymmetric radial lens distortion. As a result, the different imaging properties of the lens will not necessarily have a common axis of symmetry.

With adjustable lenses the focus and zoom are changed by varying the positions of the lens elements within the lens body. Moving the lens elements is typically accomplished in one of two ways. In the first method the lens elements are mounted in a threaded section of the lens barrel which can then be rotated around the lens body to move the group along the axis of the lens. In the second method the lens elements are mounted on slides or rails which can then be translated along the axis of the lens body using internal cams. As the spacing of the lens elements is changed so is any misalignment between their mechanical and optical axes. Generally the rotation of a lens group will cause a rotational drift in the position of the lens' optical axis, while sliding of as lens group will cause a translational motion of the lens' optical axis in the image plane. These rotational and translational shifts in the position of the optical axis cause a corresponding rotational and translational shifting of the camera's field of view.

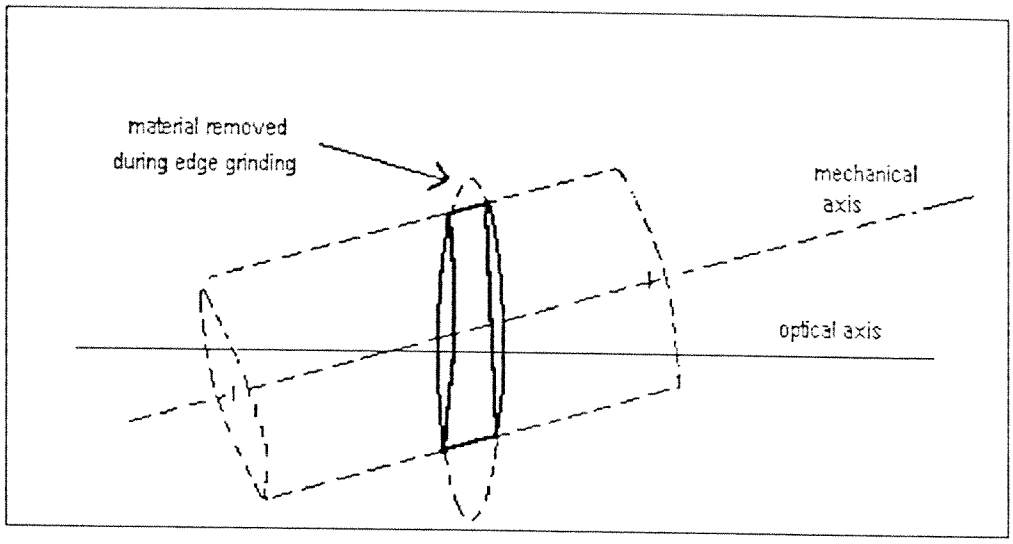


Figure 2.2: Decentration For a Simple Lens

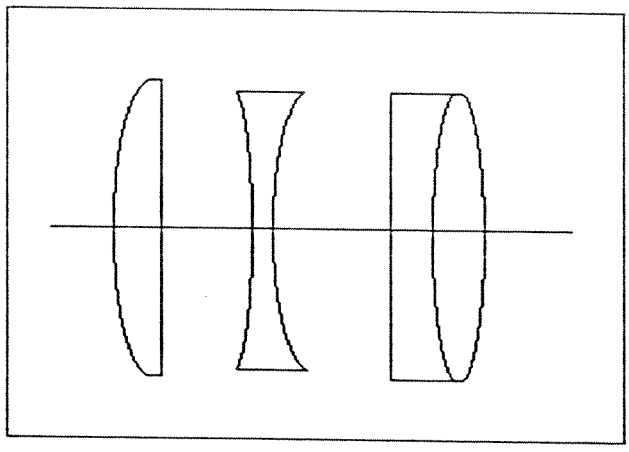


Figure 2.3: Fixed Focal Lenght Lens

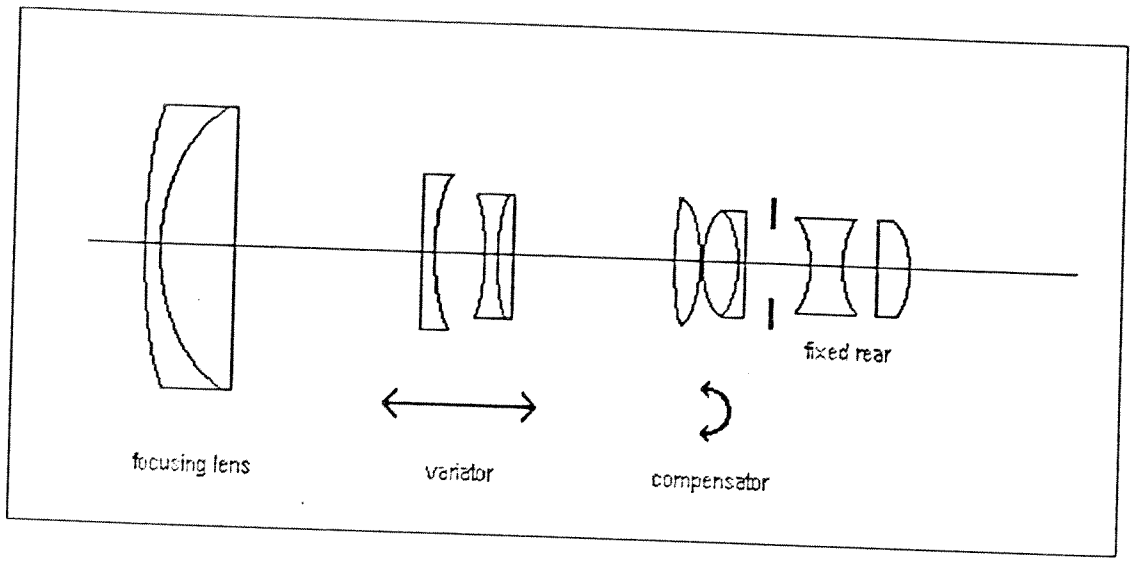


Figure 2.4: Variable Focal Length (zoom) Lens

In variable focus fixed focal length lenses typically all of the lens elements are mounted together in a single fixed assembly. To vary the lens' focus, the separation between the optics and the camera sensor is changed by moving the lens assembly with either a rotational or translational type mechanism.

In variable focal length (zoom) lenses, such as the one illustrated in Figure 2.4, the focal length is changed by moving groups of lens elements relative to one another. Typically, this is done by using a translational type mechanism on one or more of the internal groups. The lens' focus is often varied by using a rotational mechanism on the front lens group.

2.3 Lens Aberrations

Aberrations are changes which degrade the quality of the images. These changes can be caused by the lens, or by the way the lens is used.

In mono color machine vision applications, the following aberrations [5] are important:

- SPHERICAL: points in the image become enlarged in a way that is constant

over the field, but decrease as the cube of the increase in f-number(stopping the lens down greatly reduces the effect).

- DISTORTION: points in the image focus either closer or further from the center that predicted for an undistorted image, changing as the cube of the off-axis rotation, causing features in the image to become noticeably misshaped.

- VIGNETTING: a uniformly illuminated field is imaged to a field which is uneven in illumination typically bright in the middle and falling off in intensity to the sides.

- FIELD CURVATURE: evenly spaced points in the image enlarge and focus closer(or further) from the lens, as the square of increasing distance from the optical axis, decreasing linearly as the lens is stopped down. Astigmatism will cause horizontal features and vertically oriented features to focus at different depths, typically yielding different resolution for horizontal versus vertical features.

2.4 Sensors

Sensors in machine vision systems are usually television cameras. Both vacuum tube and solid state cameras are used in industrial vision systems.

Vacuum tube cameras (Vidicon, etc.) have photo-sensitive sensor planes which are scanned by an electron beam [2]. The beam reads off a charge which is accumulated on the sensor plane by exposure to light. The electron beam scans from left to right and then top to bottom, scanning a rectangle to one of standard RS-170, NTSC, PAL etc. video specifications. Sources of errors for vacuum tube cameras are non-linearity, drift, and non-uniformity of the target coating. Non-linearity appears as a change in the size of the pixel at different points in the image. Drift takes place inside the tube as a result of temperature

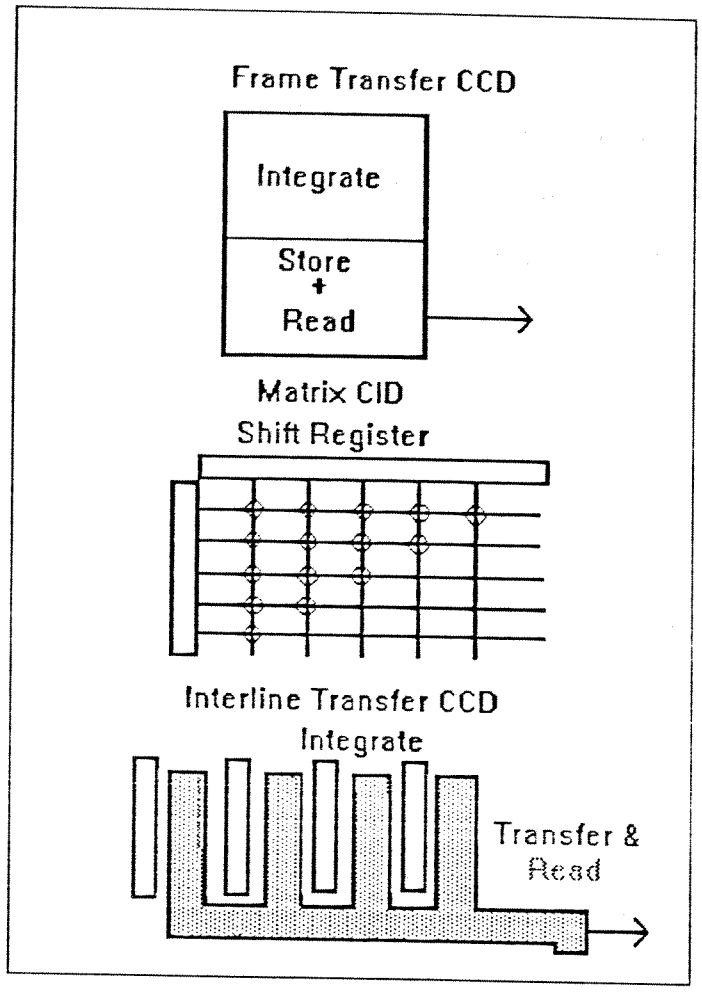


Figure 2.5: Solid State Image Sensor Transfer Concepts

fluctuations (inside the tube is a filament which is a heat source) and from voltage fluctuations. Non-uniformity of target coatings can be as high as 30% as a result of the difficulties tube manufacturers have applying a uniform, thin coating of photo-sensitive material in a bottle. This non-uniformity is not visible to the human eye. But the performance of threshold type binary vision systems can be degraded dramatically by non-uniformity of tube target coatings.

Solid state cameras are constructed utilizing an array of light sensitive semiconductor devices integrated into a single chip. The size and shape of each pixel is determined by the size and shape of the light sensitive semiconductor device in the chip. This type of camera directly samples the image. Even though the output of this type of camera might appear as an analog signal on an oscilloscope, it is the direct result of a two dimensional digitization process which

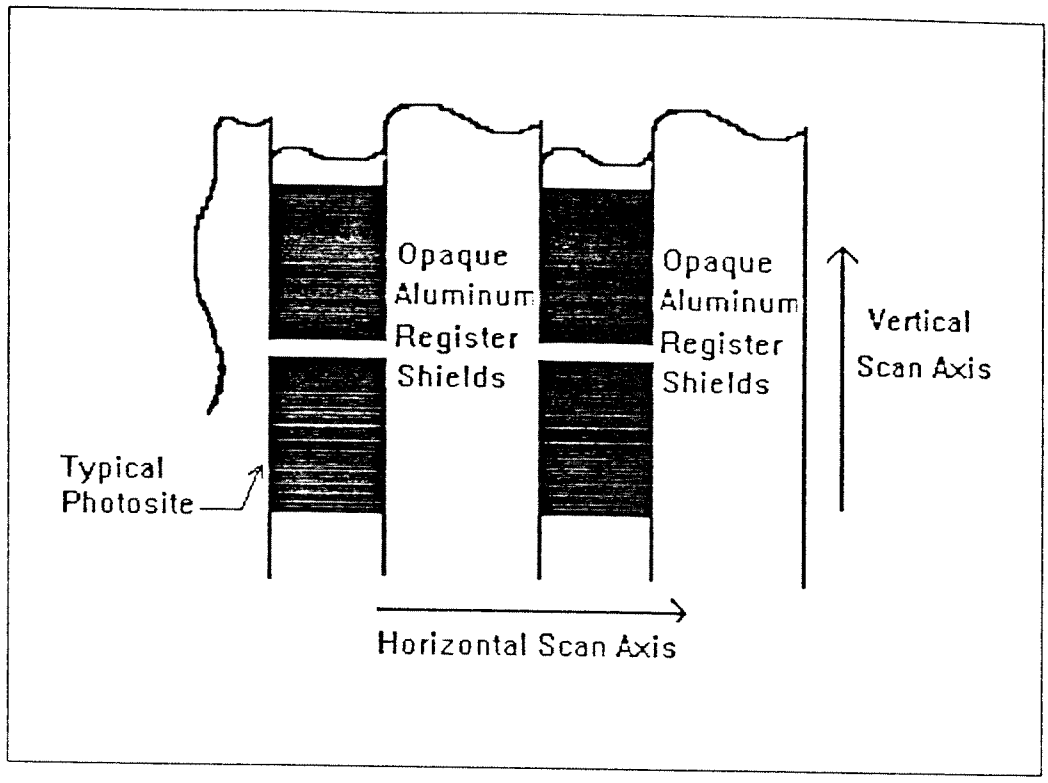


Figure 2.6: The Line Transfer Device Structural Details

converts the continuous image into one which is composed of pixels. Solid state camera sensors, which are fabricated from a single chunk of silicon, are essentially free of drift and non-linearity.

Figure 2.5 shows three popular solid-state image sensor design concepts in transferring and storing video data [3]. The photon charges on the CCD array are transferred to an output storage either directly one line at a time (line transfer) or via a temporary frame storage area (frame transfer).

The frame transfer device has a section for photon integration and another section for storage of a full frame of video data. The frame transfer device tends to present a contiguous surface for collection of photon source data.

The line transfer device interleaves the storage and transfer registers with the registers used for photon integration. One advantage of the interline device is that "smear" is virtually avoided. Because the distance from the photon integration registers to the optically shielded storage and transfer registers is very

short the charge transfer can be very rapid. On the other hand, this approach presents a significant obscureness to the photon image.

Greater than 50% of the horizontal pitch of each pixel is obscured (Figure 2.6). If single or sub pixel accuracy is required for the inspection system, this presents a disadvantage. For the matrix addressed device concept, such as CID, each pixel can be read sequentially or randomly. CID is a contiguous sensor.

In CID arrays, there are no opaque areas like in CCD arrays between pixels where image detail can be lost. This attribute is important for applications where precise dimensional data is critical, for instance the determination of object edges for inspection, measurements, positioning and to achieve sub-pixel resolution for precise imaging.

While the CCD's transfer collected charge out of the pixels during readout (hence erasing the image stored on the sensor). CID's do not transfer charge from site to site in the array. Instead readout is accomplished by transferring or shifting the collected charge packet within an individually addressed pixel. Charge remains undamaged in the pixel after the signal level has been determined, as a consequence readout is non-destructive. Non-destructive readout technique enhance user control of light integration and image readout, useful for low light, stop motion and pulsed capture applications.

Blooming describes the distortion in an image that can occur when solid-state video cameras are exposed to concentrated light intensities. Charge can spill from over saturated elements to adjoining pixels. CCD's are particularly susceptible to blooming if a given pixel receives a light level which is far in excess of the level needed to saturate the pixel. The overcharged pixel will simply spill over to adjacent pixels. CID's are less susceptible to blooming than CCD's because the basic cross section structure of the chip includes a layer which is capable of soaking up a large number of excess charge carriers.

There are two techniques to obtain digital images from a camera's

sensor [4]. They are video output cameras (closed circuit television or CCTV cameras) and non-video digital cameras (scientific, slow scan or pixel clocked cameras).

In video output cameras, each row of the sensor is scanned off of the sensor and converted to a continuous analog signal. The analog continuous signal is then resampled by a digitizer board to obtain a digital representation for the row. In this type of camera, there is a direct relationship between the row numbers on the sensor and the row numbers on the digitizer. However, the relationship between the column numbers on the sensor and the column numbers in the digitizer is not direct, and depends on the synchronization of the digitizer to the start of the analog signal for each row and on the relative sampling rates of the sensor's output clock and the digitizer's sampling clock.

In non-video digital output cameras, the sensor's pixels are digitized directly as they are clocked off the sensor resulting in a one to one correspondence between the sensor's row and column pixel coordinates and the digitizer's coordinates.

2.5 Frame Grabber

The output of a camera is a series of electrical signals corresponding to the series of horizontal scan lines used to acquire the image. The analog variation in each signal's amplitude corresponds to the intensity of the image along the scan line.

Each image's analog signals must be digitized and stored in memory to begin computation on image data. The electronic board that accomplishes this task is called a frame grabber. The frame grabber contains an analog to digital converter that assigns a digital value to each analog signal level. The quantized information content in each pixel corresponds to intensity. This information is defined as a bit and relates to image brightness when digitized into a number of

quantized elements. Acceptable analog signal levels can be one of 8, 64 or 256. The frame grabber then stores the digitized image in random access memory (RAM).

The analog signal coming from the camera must be sampled by the frame grabber such that no spatial distortion of pixel values takes place in memory. In other words, there should be a one to one correspondence between pixels in the sensor plane of the camera and frame memory of the frame grabber.

When the camera and frame grabber sampling rates are the same, then the pixel array format matches the frame memory format. Otherwise over or under sampling will occur. Under sampling takes place when pixel information from the camera is sampled by the frame grabber at a rate lower than the pixel data is clocked off. For example; 755 pixels sampled 640 times results in 640 memory locations filled with data. Each memory locations represents 1.18 pixel. Over sampling takes place when pixel information from the camera is sampled by the frame grabber at a rate higher than the pixel data is clocked out. For example; 512 pixels sampled 640 results in 640 memory locations filled with data. Each memory location represents 0.8 pixel.

CHAPTER III

RECONSTRUCTION ALGORITHM

3.1 Overview

In this Chapter, the algorithm developed in [1] for finding the 6 configuration parameters of a rigid body will be presented, and its theoretical limits will be given. All the details about the development of the algorithm can be found in [1]. The coordinate systems used will be first presented. Then target centered solution will be explained as the basic part of the solution procedure. As a completion of the algorithm, the general solution procedure will be given. Theoretical limits are given at the end of the chapter.

Since a preknown information about a rigid body geometry is required in mono vision, a secondary target mounted on the rigid body can be used. Finding the configuration of the secondary target means finding the configuration of the rigid body on which the secondary target is mounted. Hence a known, well defined and studied target geometry to find configuration parameters will ease the solution procedure.

A circular planar secondary target with two internal spots is used in the solution procedure. Boundary of the circle is used to detect 3 of the configuration parameters. The other 3 parameters are found by using the centers of the spots.

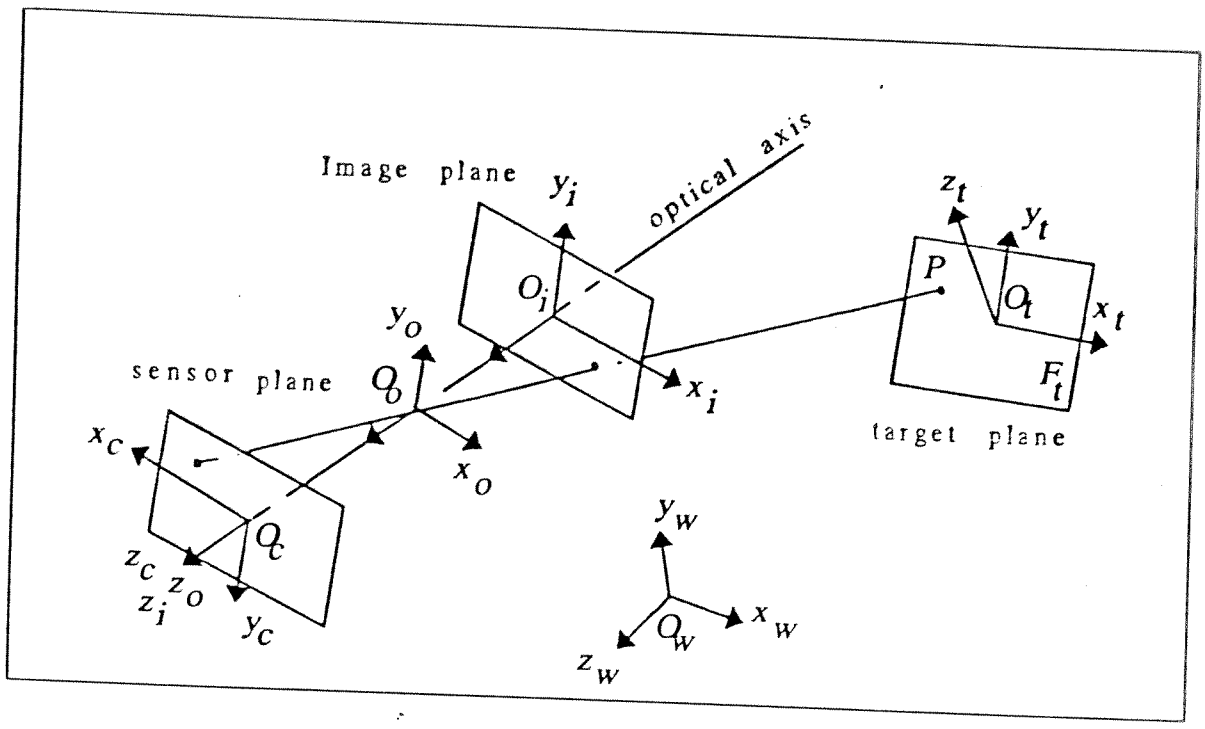


Figure 3.1: Coordinate Systems

3.2 The Model

Five coordinate systems are defined, as shown in Figure 3.1, namely:

- $(O_t x_t y_t z_t)$ The target coordinate system to which the target plane formed by the axes x_t and y_t is attached.
- $(O_w x_w y_w z_w)$ The world coordinate system.
- $(O_c x_c y_c z_c)$ The camera coordinate system to which the sensor plane formed by the axes x_c and y_c is attached.
- $(O_o x_o y_o z_o)$ The optical coordinate system.
- $(O_i x_i y_i z_i)$ The image coordinate system to which the image plane formed by the axes x_i and y_i is attached.

The world coordinates of any point P, lying on the target plane can be expressed by using the coordinate transformation matrices as:

$$\begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \overset{\cdot}{H}_{wt}^{(w,t)} \begin{bmatrix} x_t \\ y_t \\ 0 \\ 1 \end{bmatrix}$$

The coordinate transformation matrix $\overset{\cdot}{H}_{wt}^{(w,t)}$ can be expressed as the multiplication of three consecutive transformations:

$$\overset{\cdot}{H}_{wt}^{(w,t)} = \overset{\cdot}{H}_{wc}^{(w,c)} \; \overset{\cdot}{H}_{ci}^{(c,i)} \; \overset{\cdot}{H}_{it}^{(i,t)}$$

where

$\overset{\cdot}{H}_{wc}^{(w,c)}$ defines the configuration of the camera and is determined by the camera calibration.

$\overset{\cdot}{H}_{ci}^{(c,i)}$ defines the position of the image plane with respect to camera coordinate system and is known for a given lens setting.

$\overset{\cdot}{H}_{it}^{(i,t)}$ defines the configuration of the target coordinate system with respect to the image coordinate system and $\overset{\cdot}{H}_{it}^{(i,t)}$ is determined by using the algorithm.

$\overset{\cdot}{H}_{it}^{(i,t)}$ is decomposed into a series of basic rotation and translation matrices as follows:

$$\overset{\cdot}{H}_{it}^{(i,t)} = \hat{T}_z(d_o)\,\hat{R}_x(-\phi)\,\hat{R}_y(-\theta)\,\hat{T}_z(-d)\,\hat{R}_x(\alpha)\,\hat{R}_y(\beta)\,\hat{R}_z(\gamma)$$

where $\hat{R}_x(..), \hat{R}_y(..), \hat{R}_z(..)$ denotes the basic rotation matrices around x, y and z axes respectively. $\hat{T}_z(..)$ denotes the basic translation matrix along the z axis. Figure 3.3 illustrates the variables ϕ, θ, d, α, β and γ used in the formulations. Once these variables are found, the target configuration with respect to the image plane is determined.

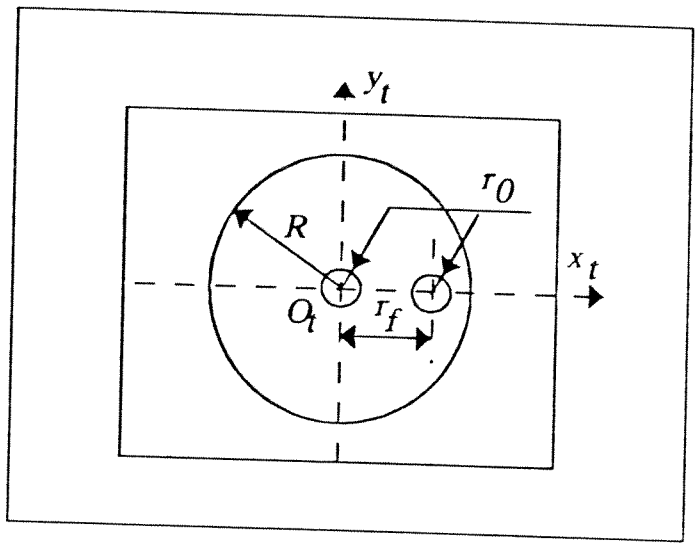


Figure 3.2: The Target

3.3 Target Geometry

The target geometry is shown in Figure 3.2. It consists of a circle with two internal spots.

One of the spots is located at the center of the circle. It is used to locate the true center of the circle. The other spot is placed r_f away from the center, and used to determine the in plane rotation. The asymmetric appearance of the target prevents multiple solutions.

The circle has a radius of R. Spots's radii are equal to each other and it is r_o. Radii of the spots must be chosen small enough that the shift in their true and image centers are almost equal.

3.4 Target Centered Solution

Target centered solution is presented in full detail in [1]. In this section, a general overview and results of the solution will be presented.

The target centered solution corresponds to the case where both ϕ

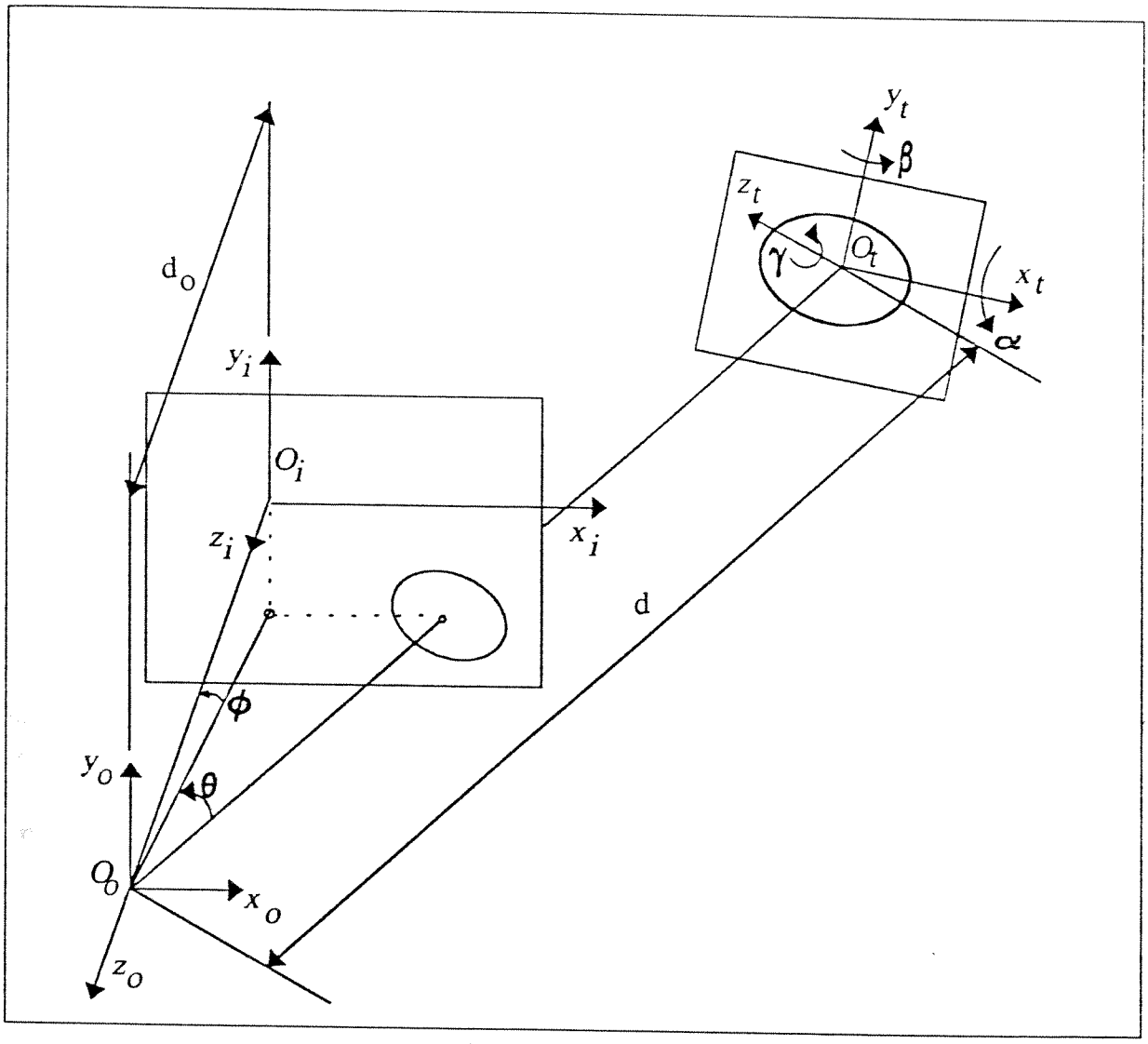


Figure 3.3: Reconstruction Parameters

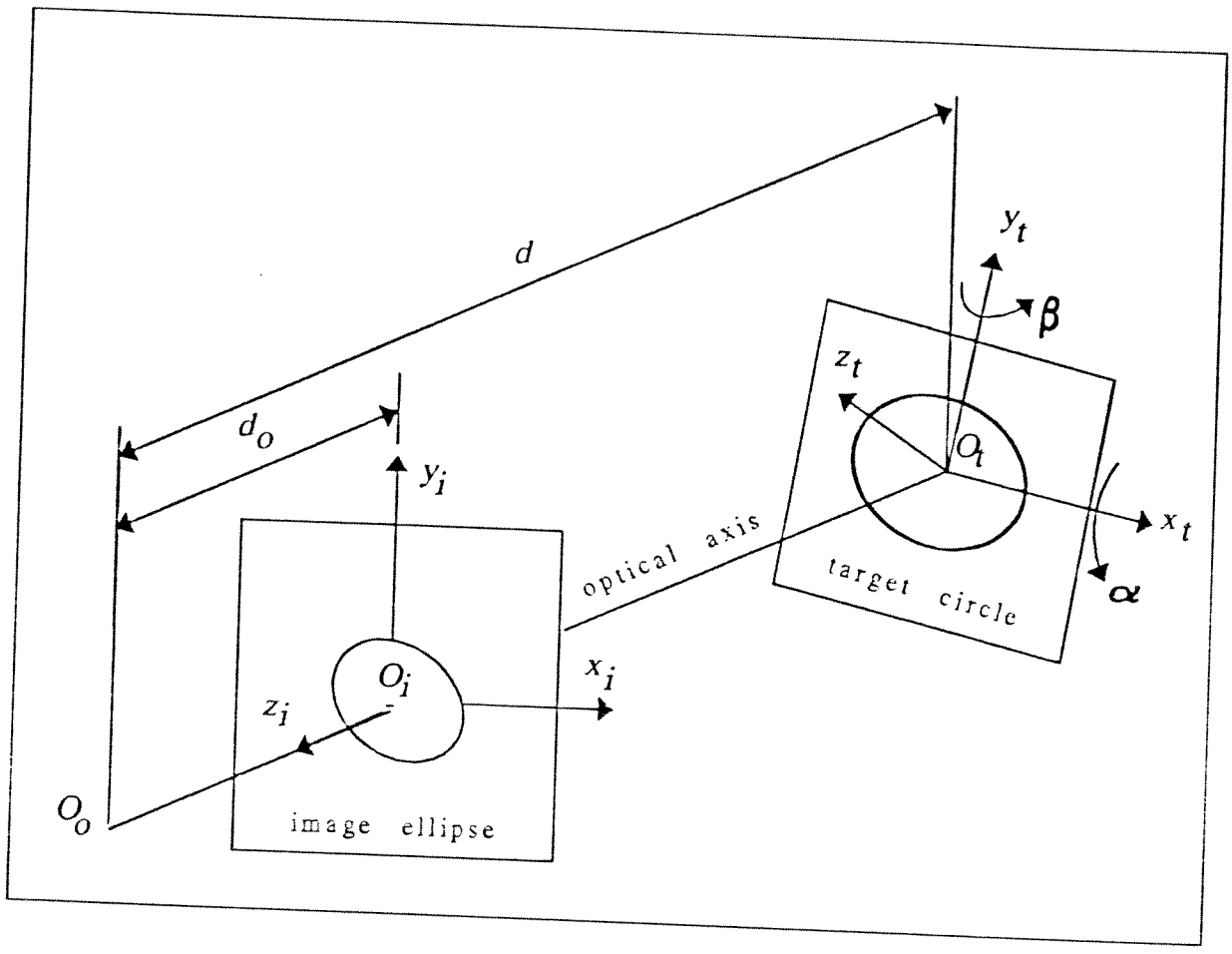


Figure 3.4: Reconstruction Parameters for The Target Centred Case

and θ angles are equal to zero. Which means that target center is on the optical axis. Also in plane rotation is not considered in the solution. So the target configuration is defined by d, α, β where α is the rotation about x_t axis, β is the rotation about y_t axis and d is the distance between O_o (perspective center) and O_t (target center). The configuration variables are shown in Figure 3.4.

The transformation matrix $\overset{\cdot}{H}_{it}^{(i,t)}$ can now be expressed as:

$$\overset{\cdot}{H}_{it}^{(i,t)} = \overset{\cdot}{T}_z (d_o - d)\, \overset{\cdot}{R}_x (\alpha)\, \overset{\cdot}{R}_y (\beta)$$

The coordinates of any point lying on the target plane can be expressed in the image coordinate system as follows:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} = \overset{\cdot}{H}_{it}^{(i,t)} \begin{bmatrix} x_t \\ y_t \\ 0 \\ 1 \end{bmatrix}$$

Finally the coordinates of the point can be calculated using the perspective transformation as:

$$\begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} (X_i d_o)/(d_o - Z_i) \\ (Y_i d_o)/(d_o - Z_i) \\ 0 \\ 1 \end{bmatrix}$$

In the solution process, the image points of the target are available. From the above formulation, the target coordinates of an image can be obtained by solving the above equations for x_t and y_t. The solution gives:

$$x_t = \frac{(d\cos\alpha)x_i}{(-\sin\beta)x_i + (\sin\alpha\cos\beta)y_i + (\cos\alpha\cos\beta d_o)}$$

$$y_t = \frac{d[(-\sin\alpha\sin\beta)x_i + (\cos\beta)y_i]}{(-\sin\beta)x_i + (\sin\alpha\cos\beta)y_i + (\cos\alpha\cos\beta d_o)} \qquad (3.1)$$

The mathematical representation of the circular secondary target, on the target coordinate system is given by:

$$x_t + y_t = R^2 \qquad (3.2)$$

where R is the radius of the target circle.

From the invariance theorem [1], the image of a circle will be another curve of a second degree. It is obvious that it will be a closed curve, by intuition it will be an ellipse, and will be called the image ellipse in the text.

Substituting Equations 3.1 in Equation 3.2 normalized equations of the image ellipse can be obtained as:

$$Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + 1 = 0 \qquad (3.3)$$

where the normalized coefficients are:

$$A = \frac{-[d^2 + (d^2 - R^2)\tan^2\beta/\cos^2\alpha]}{(Rd_o)^2}$$

$$
\begin{aligned}
B &= \frac{2(d^2 - R^2)\tan\alpha\tan\beta}{(Rd_o)^2\cos\alpha} \\
C &= \frac{-[d^2 + (d^2 - R^2)\tan^2\alpha]}{(Rd_o)^2} \\
D &= \frac{-2\tan\beta}{d_o\cos\alpha} \\
E &= \frac{2\tan\alpha}{d_o}
\end{aligned}
\tag{3.4}
$$

To obtain the coefficients of the image ellipse in the image coordinate system, a curve fitting operation must be performed by N pixel points selected along the contour of the image ellipse.

The curve fitting can be realized by using the pseudo-inversion technique [1] which handles the inversion of non square matrices. The pseudo-inversion technique is equivalent to the least square solution when the number of constraint equations is greater than the number of unknowns.

For the N selected pixels, the ellipse Equation in 3.3 will give the following set of equations:

$$
\begin{aligned}
Ax_0^2 + Bx_0y_0 + Cy_0^2 + Dx_0 + Ey_0 + 1 &= 0 \\
Ax_1^2 + Bx_1y_1 + Cy_1^2 + Dx_1 + Ey_1 + 1 &= 0 \\
Ax_2^2 + Bx_2y_2 + Cy_2^2 + Dx_2 + Ey_2 + 1 &= 0 \\
\ldots &= 0 \\
\ldots &= 0 \\
Ax_N^2 + Bx_Ny_N + Cy_N^2 + Dx_N + Ey_N + 1 &= 0
\end{aligned}
\tag{3.5}
$$

The set given in Equation 3.5 can be expressed in matrix form as follows:

$$
M \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ . \\ . \\ . \\ -1 \\ -1 \end{bmatrix}
$$

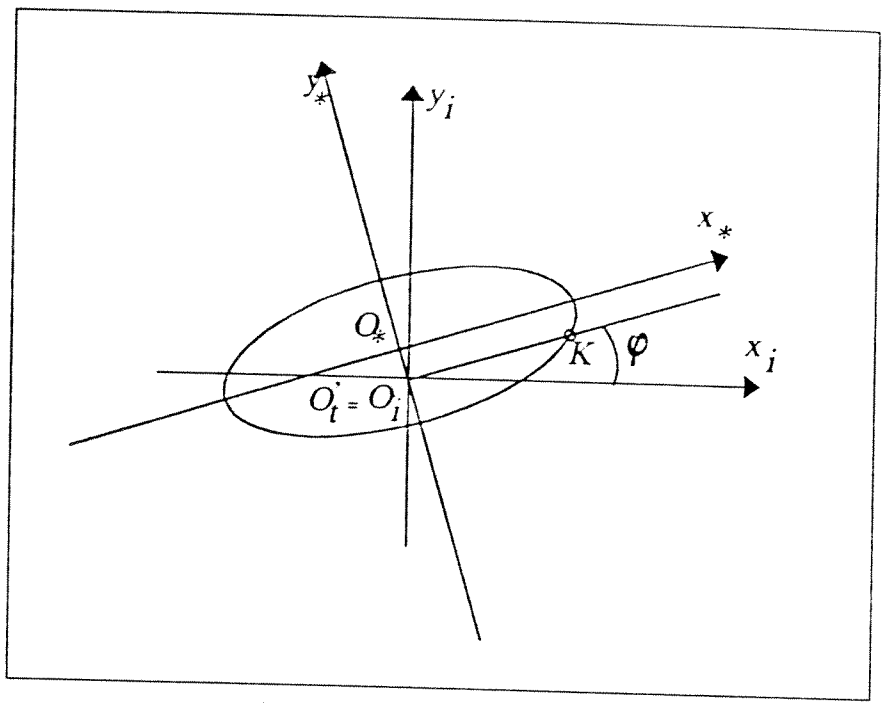


Figure 3.5: Image Ellipse

where

$$M = \begin{bmatrix} x_0^2 & x_0y_0 & y_0^2 & x_0 & y_0 \\ x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 \\ x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_N^2 & x_Ny_N & y_N^2 & x_N & y_N \end{bmatrix} \tag{3.6}$$

by using the pseudo-inversion technique A, B, C, D, E can be found. Although it seems that the solution has been already reached, it is not possible to solve d, α, β from the nonlinear Equations 3.4.

To make the solution process easier, it is possible to define a new coordinate system $(O_*x_*y_*z_*)$ as shown in Figure 3.5. The center of the ellipse and the image of the true center of the circle are not coincident. The true center of the circle is O_*.

The Equation 3.3 of the ellipse can be written in the following matrix form as:

$$\begin{bmatrix} x_i & y_i & 1 \end{bmatrix} \begin{bmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0 \tag{3.7}$$

It is also possible to represent the ellipse in the canonical coordinates (x_*, y_*) as follows:

$$\begin{bmatrix} x_* & y_* & 1 \end{bmatrix} \begin{bmatrix} A_* & 0 & 0 \\ 0 & C_* & 0 \\ 0 & 0 & F_* \end{bmatrix} \begin{bmatrix} x_* \\ y_* \\ 1 \end{bmatrix} = 0 \tag{3.8}$$

Suppose that $\overset{\cdot}{C}^{(i,*)}$ is a coordinate transformation matrix such that:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \overset{\cdot}{C}^{(i,*)} \begin{bmatrix} x_* \\ y_* \\ 1 \end{bmatrix} \tag{3.9}$$

then combining Equations 3.7, 3.8, 3.9, yields to:

$$\begin{bmatrix} A_* & 0 & 0 \\ 0 & C_* & 0 \\ 0 & 0 & F_* \end{bmatrix} = \left[\overset{\cdot}{C}^{(i,*)}\right]^T \begin{bmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & 1 \end{bmatrix} \left[\overset{\cdot}{C}^{(i,*)}\right]$$

The diagonal elements can be obtained [1] as:

$$\begin{aligned} A_* &= \frac{A+C}{2} - \frac{\sqrt{(A-C)^2+B^2}}{2} \\ C_* &= \frac{A+C}{2} + \frac{\sqrt{(A-C)^2+B^2}}{2} \\ F_* &= \frac{Det\left[\hat{M}\right]}{AC - B^2/4} \end{aligned} \tag{3.10}$$

independent of $\overset{\cdot}{C}^{(i,*)}$. By combining Equation 3.4 and 3.10:

$$C_* = (-d/Rd_o)^2$$

which makes more sense when expressed as

$$\frac{R}{\sqrt{-1/C_*}} = \frac{d}{d_o} \tag{3.11}$$

where $\sqrt{-1/C_*}$ is the chord O_iK of the ellipse corresponding to the true radius of the circle and the Equation 3.11 is nothing but the application of the similar triangles rule. Actually this fact can also proved geometrically by intersecting the line $y_i = x_i \tan\varphi$ and the ellipse, as shown in Figure 3.5 where $2\varphi = \tan^{-1}[B/(A-C)]$.

Now the distance d can be solved independent of α and β as:

$$d = Rd_o\sqrt{-C_*}$$

Finally, using the A and C expressions in Equation 3.4 one can solve for α and β:

$$\alpha = sgn[E]\tan^{-1}\left[\sqrt{\frac{-(CR^2d_o^2+d^2)}{d^2-R^2}}\right] \tag{3.12}$$

$$\beta = -sgn[D]\tan^{-1}\left[\cos\alpha\sqrt{\frac{-(AR^2d_o^2+d^2)}{d^2-R^2}}\right] \tag{3.13}$$

where the signs of D and E are used to find a unique solution. The unique solution is for $\alpha, \beta \in [-90^\circ, +90^\circ]$ which covers the practical range.

3.5 The General Solution

The details of the general solution is given in [1]. Here the general solution will be presented as if it can be implemented directly. The 6 configuration parameters ϕ, θ, d, α, β, γ which determine the position and orientation of the target with respect to the image plane will be found by using the contour of the circle and centroids of the spot. First θ and ϕ will be determined from the centroid of the central spot. Then d, α,and β will be found from the contour of the image ellipse. Finally in plane rotation γ will be determined from the centroid of the outer spot.

STEP-1:

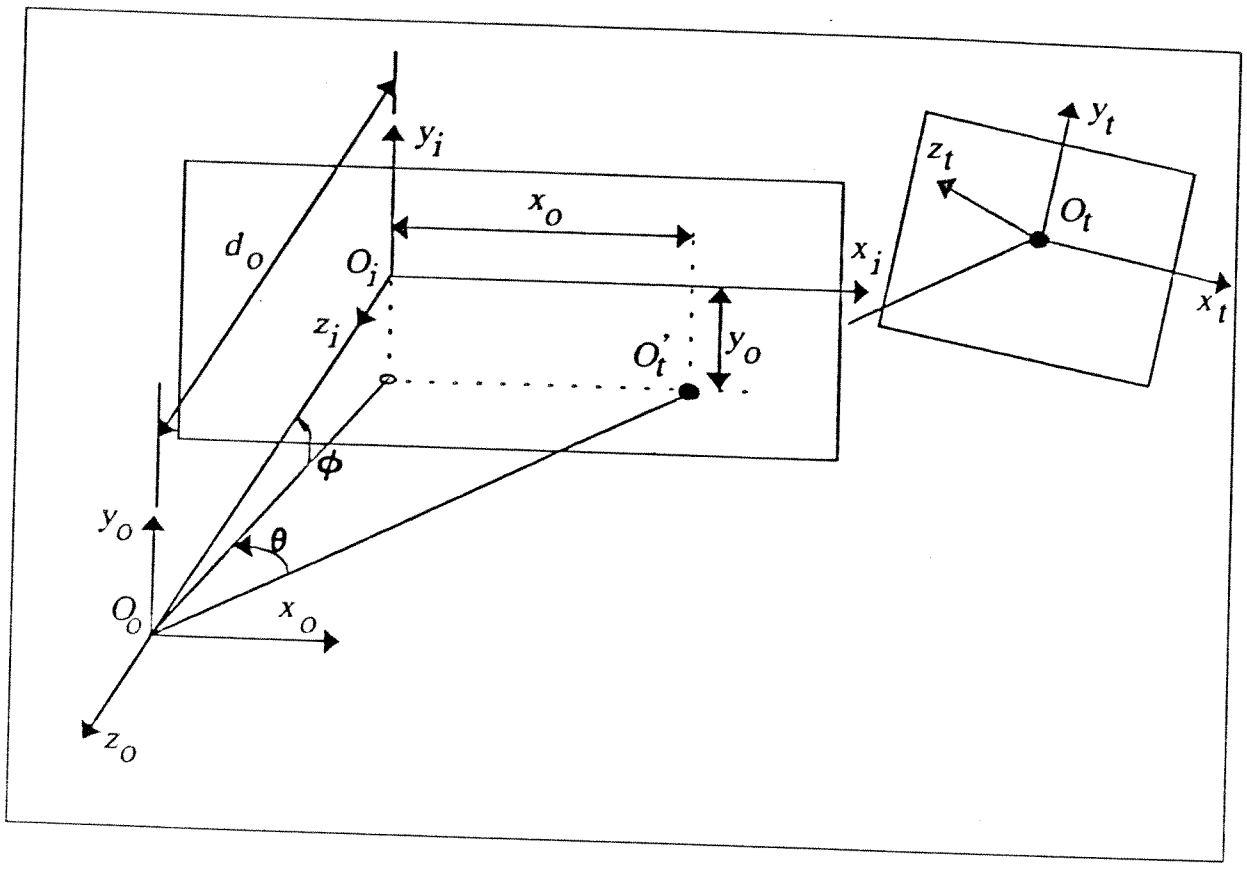


Figure 3.6: Calculation of ϕ and θ

The image of the central spot of the target will represent the true center of the circle on the image plane. Its centroid will be the image of the origin of the image coordinate system. If we can determine θ and ϕ we can locate the direction in which the target center is located.

From Figure 3.6, image points of the true center of the target will be:

$$x_o = \frac{d_0 \tan\theta}{\cos\phi}$$
$$y_o = -d_o \tan\phi$$

Since image points (x_o, y_o) are known, θ and ϕ can be found as:

$$\phi = \tan^{-1}(-y_o/d_o)$$
$$\theta = \tan^{-1}(x_o \cos\phi/d_o)$$

STEP-2:

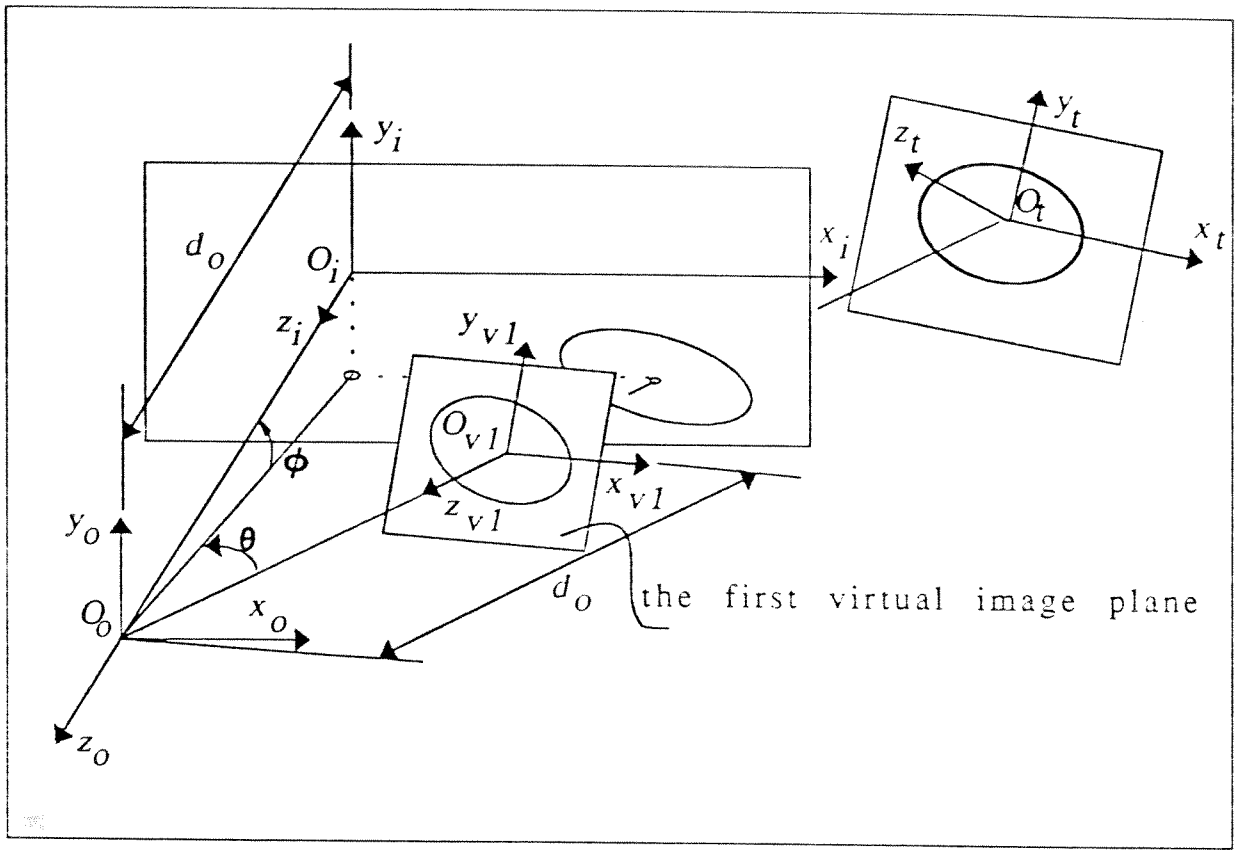


Figure 3.7: The Virtual Image plane

Once the variables θ and ϕ are determined then the direction where the target center is located has been found. It is possible to define a virtual image plane of a fictitious camera whose optic axis crosses the target center O_t, as shown in Figure 3.7. A coordinate system is assigned to this virtual image plane and is denoted by $(O_{v1}x_{v1}y_{v1}z_{v1})$. If we can find the coordinates of the image points in the virtual image plane, then it is possible to apply the target centered solution to find d, α, β with respect to the virtual plane.

Since θ and ϕ are known from STEP-1 it is possible to write the transformation matrix from the virtual image plane to the image plane as:

$$\hat{H}_{vi}^{(v,i)} = \hat{T}_z(d_o)\,\hat{R}_y(\theta)\,\hat{R}_x(\phi)\,\hat{T}_z(-d_o)$$

Then the image and virtual image can be related by:

$$\begin{bmatrix} X_v \\ Y_v \\ Z_v \\ 1 \end{bmatrix} = H_{vi}^{(v,i)} \begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix}$$

where (X_v, Y_v, Z_v) are the coordinates of an image point expressed in the virtual image coordinate system. Then the projection of the selected points on the virtual image plane can be achieved using perspective transformation:

$$\begin{bmatrix} x_v \\ y_v \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} (X_v d_o)/(d_o - Z_v) \\ (Y_v d_o)/(d_o - Z_v) \\ 0 \\ 1 \end{bmatrix}$$

Finally the coordinates of an image point projected onto the virtual image plane are obtained as:

$$\begin{aligned} x_v &= \frac{d_o(x_i \cos\theta + y_i \sin\theta \sin\phi - d_o \sin\theta \cos\phi)}{x_i \sin\theta - y_i \sin\phi \cos\theta + d_o \cos\theta \cos\phi} \\ y_v &= \frac{d_o(y_i \cos\phi + d_o \sin\phi)}{x_i \sin\theta - y_i \sin\phi \cos\theta + d_o \cos\theta \cos\phi} \end{aligned} \tag{3.14}$$

After finding (x_v, y_v) target centered solution can be used to find α, β and d with respect to the virtual plane.

STEP-3:

To get the in plane rotation of the target, a new virtual image plane is defined. It is located such that it is parallel to the target plane and its origin lying on the line $O_c O_t$. Also to ease the solution procedure, one more constraint to the location of the virtual image plane is added. It is positioned such that the center of the image spot lies on the intersection line between the actual image plane and the virtual image plane. A simplified two dimensional form of this situation is shown in Figure 3.8.

Once the location of the virtual image plane is determined, it is possible to write down the transformation matrix from the virtual image plane to the

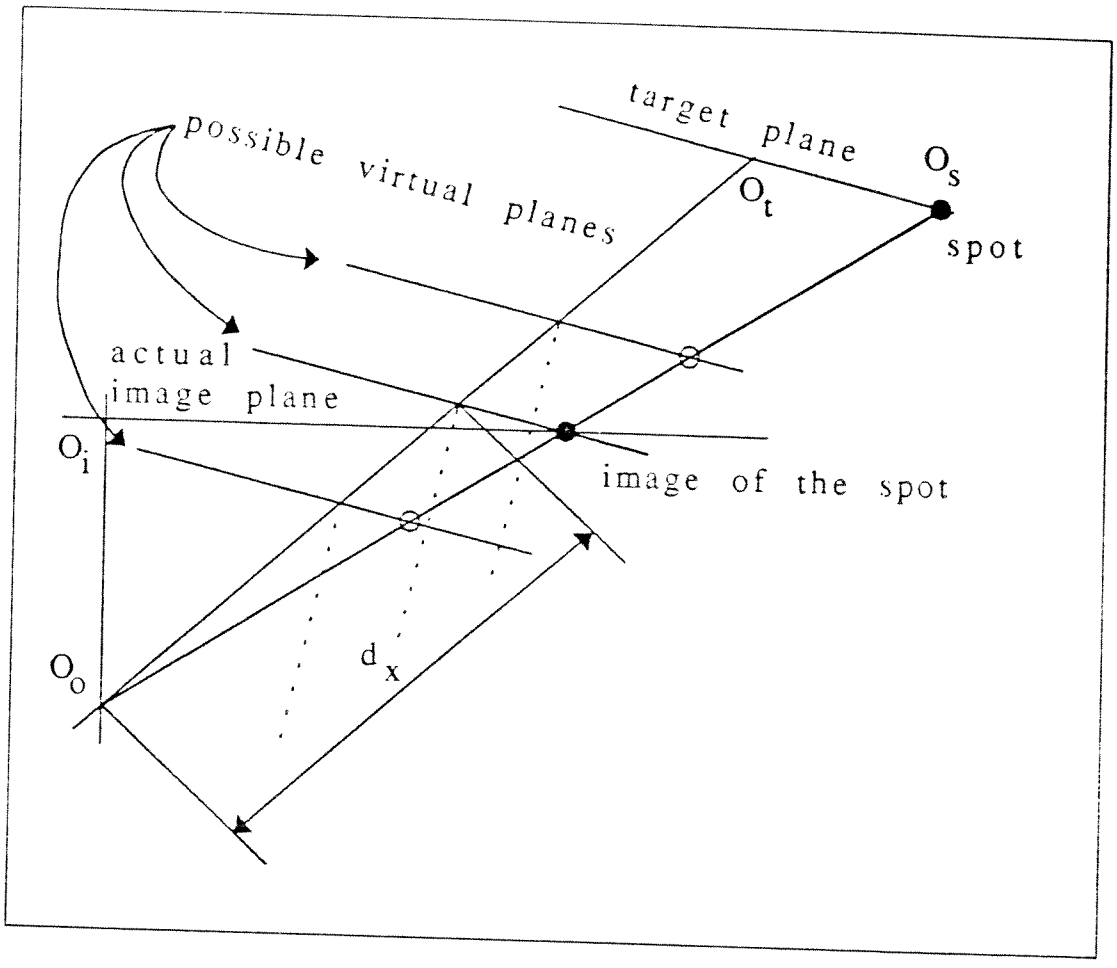


Figure 3.8: Possible Locations for The Second Virtual Plane

image plane as:

$$\overset{\cdot}{H}_{vi}^{(v,i)} = R_y(-\beta)\, R_x(-\alpha)\, T_z(d_x)\, R_y(\theta)\, R_x(\phi)\, T_z(-d_o)$$

The parameter d_x is used to locate the center of the image spot to lie on the intersection of the virtual image plane and the actual image plane. The newly defined virtual plane is shown in Figure 3.9 and is denoted by $(O_{v2}x_{v2}y_{v2}z_{v2})$. As seen from the figure the virtual image plane is parallel to the target plane and as a consequence z_t is parallel to z_v.

When the basic transformation matrices are substituted the elements of the matrix can be obtained as:

$$\overset{\cdot}{H}_{vi}^{(v,i)}[1,1] = \cos\beta\cos\theta + \sin\beta\cos\alpha\sin\theta$$

$$\overset{\cdot}{H}_{vi}^{(v,i)}[1,2] = \sin\beta\sin\alpha\cos\phi + \cos\beta\sin\theta\sin\phi -$$

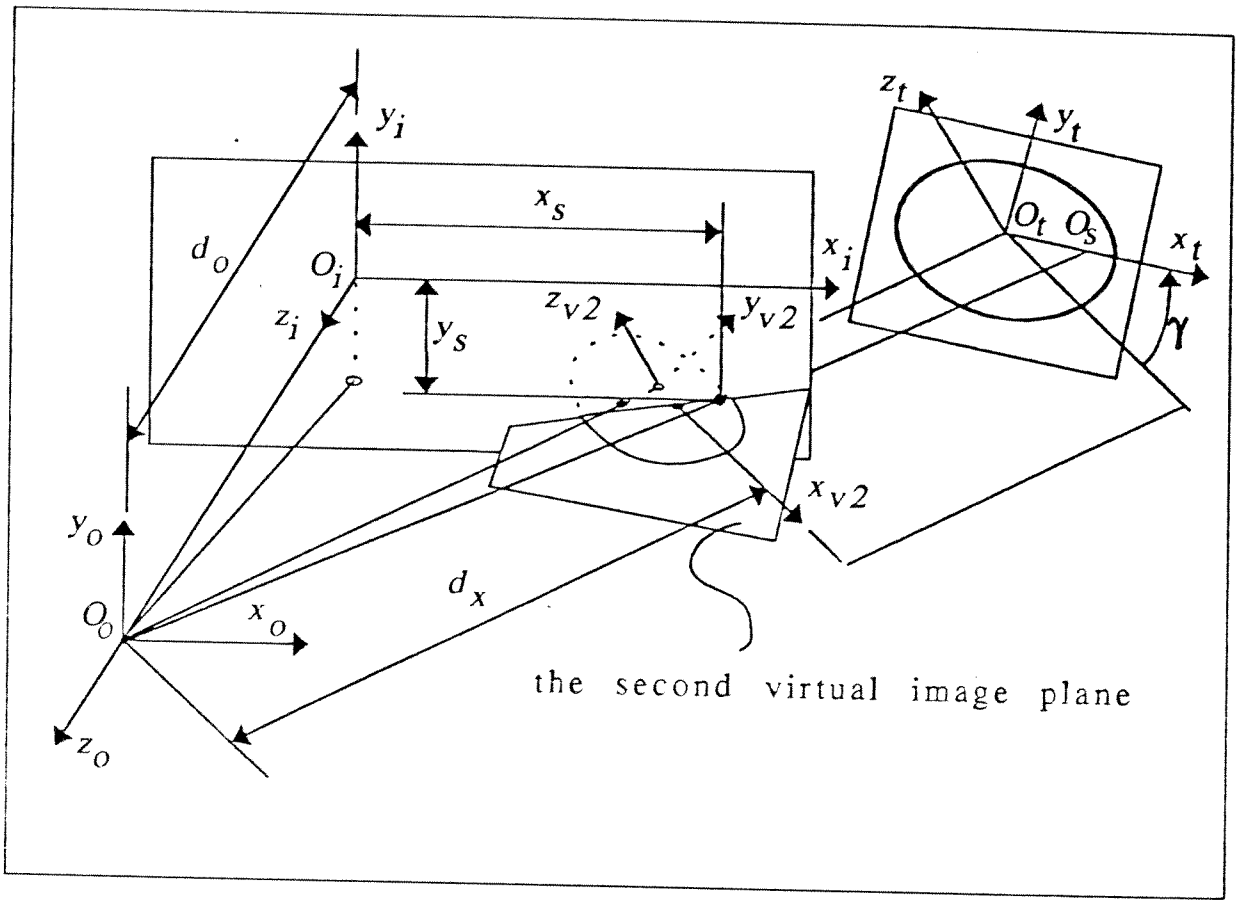


Figure 3.9: The Virtual Plane For STEP-3

$$
\begin{aligned}
\overset{(v,i)}{H_{vi}}\ [1,3] &= \sin\beta\cos\alpha\cos\theta\sin\phi \\
&\quad -\sin\beta\sin\alpha\sin\phi+\cos\beta\sin\theta\cos\phi- \\
&\quad \sin\beta\cos\alpha\cos\theta\cos\phi \\
\overset{(v,i)}{H_{vi}}\ [1,4] &= d_0\sin\beta\sin\alpha\sin\phi-d_0\cos\beta\sin\theta\cos\phi \\
&\quad +d_0\sin\beta\cos\alpha\cos\theta\cos\phi-d_x\sin\beta\cos\alpha \\
\overset{(v,i)}{H_{vi}}\ [2,1] &= -\sin\alpha\sin\theta \\
\overset{(v,i)}{H_{vi}}\ [2,2] &= \cos\alpha\cos\phi+\sin\alpha\cos\theta\sin\phi \\
\overset{(v,i)}{H_{vi}}\ [2,3] &= -\cos\alpha\sin\phi+\sin\alpha\cos\theta\cos\phi \\
\overset{(v,i)}{H_{vi}}\ [2,4] &= d_o\cos\alpha\sin\phi-d_o\sin\alpha\cos\theta\cos\phi+d_x\sin\alpha \\
\overset{(v,i)}{H_{vi}}\ [3,1] &= \sin\beta\cos\theta-\cos\beta\cos\alpha\sin\theta \\
\overset{(v,i)}{H_{vi}}\ [3,2] &= -\cos\beta\sin\alpha\cos\phi+\sin\beta\sin\theta\sin\phi \\
&\quad +\cos\beta\cos\alpha\cos\theta\sin\phi
\end{aligned}
$$

$$
\begin{aligned}
\overset{\cdot (v,i)}{H_{vi}}\ [3,3] &= \cos\beta \sin\alpha \sin\phi + \sin\beta \sin\theta \cos\phi \\
&\quad + \cos\beta \cos\alpha \cos\theta \cos\phi \\
\overset{\cdot (v,i)}{H_{vi}}\ [3,4] &= -d_0 \cos\beta \sin\alpha \sin\phi - d_0 \sin\beta \sin\theta \cos\phi \\
&\quad - d_0 \cos\beta \cos\alpha \cos\theta \cos\phi + d_x \cos\beta \cos\alpha \\
\overset{\cdot (v,i)}{H_{vi}}\ [4,1] &= 0 \\
\overset{\cdot (v,i)}{H_{vi}}\ [4,2] &= 0 \\
\overset{\cdot (v,i)}{H_{vi}}\ [4,3] &= 0 \\
\overset{\cdot (v,i)}{H_{vi}}\ [4,4] &= 1
\end{aligned}
$$

One can write the following matrix equation:

$$
\begin{bmatrix} X_v \\ Y_v \\ Z_v \\ 1 \end{bmatrix} = \overset{\cdot (v,i)}{H_{vi}} \begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix}
$$

The matrix $\overset{\cdot (v,i)}{H_{vi}}$ can be represented by its elements e_{ij} where i denotes the row and j denotes the column:

$$
\begin{bmatrix} X_v \\ Y_v \\ Z_v \\ 1 \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & e_{13} & e_{14} \\ e_{21} & e_{22} & e_{23} & e_{24} \\ e_{31} & e_{32} & e_{33} & e_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 0 \\ 1 \end{bmatrix}
$$

For the image center of the outer spot to lie on the second virtual image plane, the key point is to impose $Z_v = 0$ explicitly and to solve for d_x using $x_i = x_s$ and $y_i = y_s$. This is equivalent to write:

$$
[e_{31}\ e_{32}\ e_{33}\ e_{34}] \begin{bmatrix} x_s \\ y_s \\ 0 \\ 1 \end{bmatrix} = 0
$$

Observing the elements e_{31}, e_{32} and e_{34}, it is seen that the only unknown is d_x, since x_s and y_s can be determined using the centroid of the image of the spot. Therefore, d_x can be calculated as:

$$
d_x = \frac{-(e_{31})x_s - (e_{32})y_s + (e_{34}^*)}{\cos\beta \cos\alpha}
$$

where

$$e_{34}^{*} = d_0 \cos\beta \sin\alpha \sin\phi + d_0 \sin\beta \sin\theta \cos\phi + d_0 \cos\beta \cos\alpha \cos\theta \cos\phi$$

Now the virtual image coordinates of the center of the outer spot which are going to be denoted by x_s' and y_s' can be determined with the substitution of d_x in the expressions of e_{14} and e_{24};

$$x_s' = (e_{11})x_s + (e_{12})y_s + (e_{14})$$

$$y_s' = (e_{21})x_s + (e_{22})y_s + (e_{24})$$

As the virtual image plane and the target plane are parallel, γ can be calculated using a double argument arctangent function as follows:

$$\gamma = \tan_2^{-1}(\frac{y_s'}{x_s'})$$

One important property of this method is that it does not recall the off-set distance of the spot denoted by r_f in Figure 3.2. Thus the in-plane rotation can be reconstructed just with the knowledge that the spot is located on the x_t axis.

3.6 Theoretical Limits of the Algorithm

A computer program is developed in [1] to test the algorithm and to find out its theoretical accuracy limits. The computer program is used to simulate the algorithm. The simulation consists of three main actions:

- generation of a binary target image for a given target and camera configuration
- reconstruction of configuration variables for a given solution setting
- finding out of reconstruction error for analysis

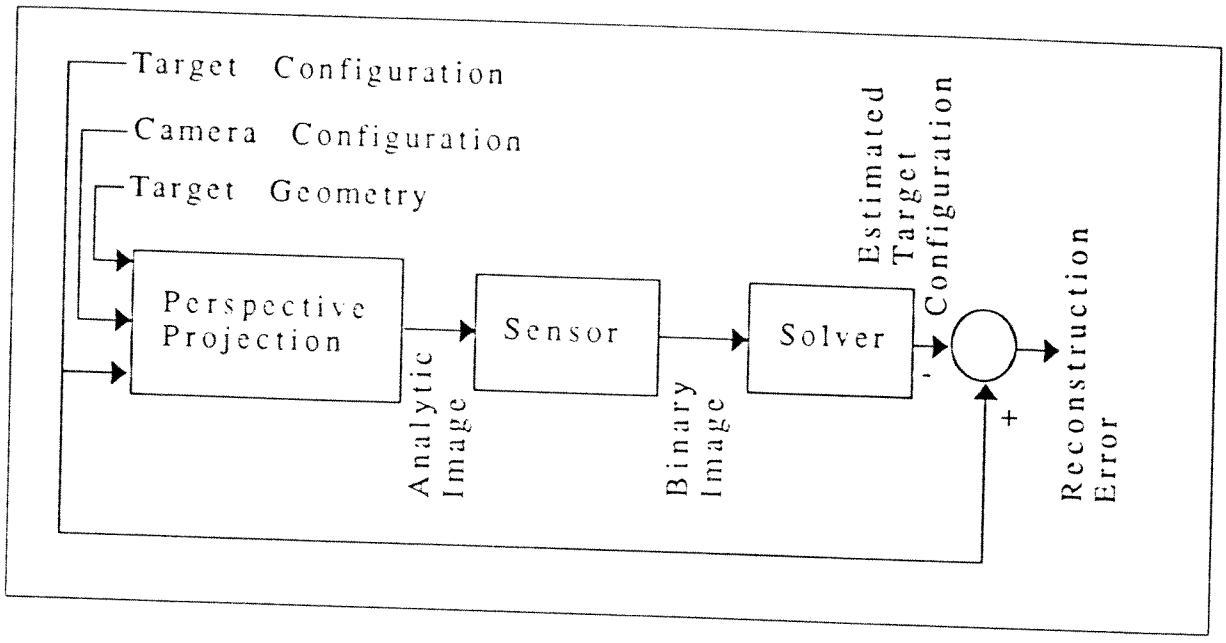


Figure 3.10: Block Diagram Representation of the Simulation Program

The block diagram representation of the simulation program is shown in Figure 3.10. Error analysis is performed by comparing errors for different types of pixel selection from contour data and for different solution settings. The configuration of the target plane with respect to the world coordinate system is determined by the variables T_x, T_y, T_z, R_x, R_y, and R_z. T_x, T_y and T_z represents the three translations along the x, y and z axes respectively. R_x, R_y and R_z represents the rotations around x, y and z axes respectively.

3.7 Selection of Pixels for Curve Fitting

A total of $N \geq 5$ pixels are required along the contour of the image ellipse to realize the curve fitting process.

Two different pixel selection algorithms are developed in [1] and their effects on the reconstruction errors are analyzed.

1. MaxMinCross, N=8
2. EqualArcs, $5 \leq N \leq 700$

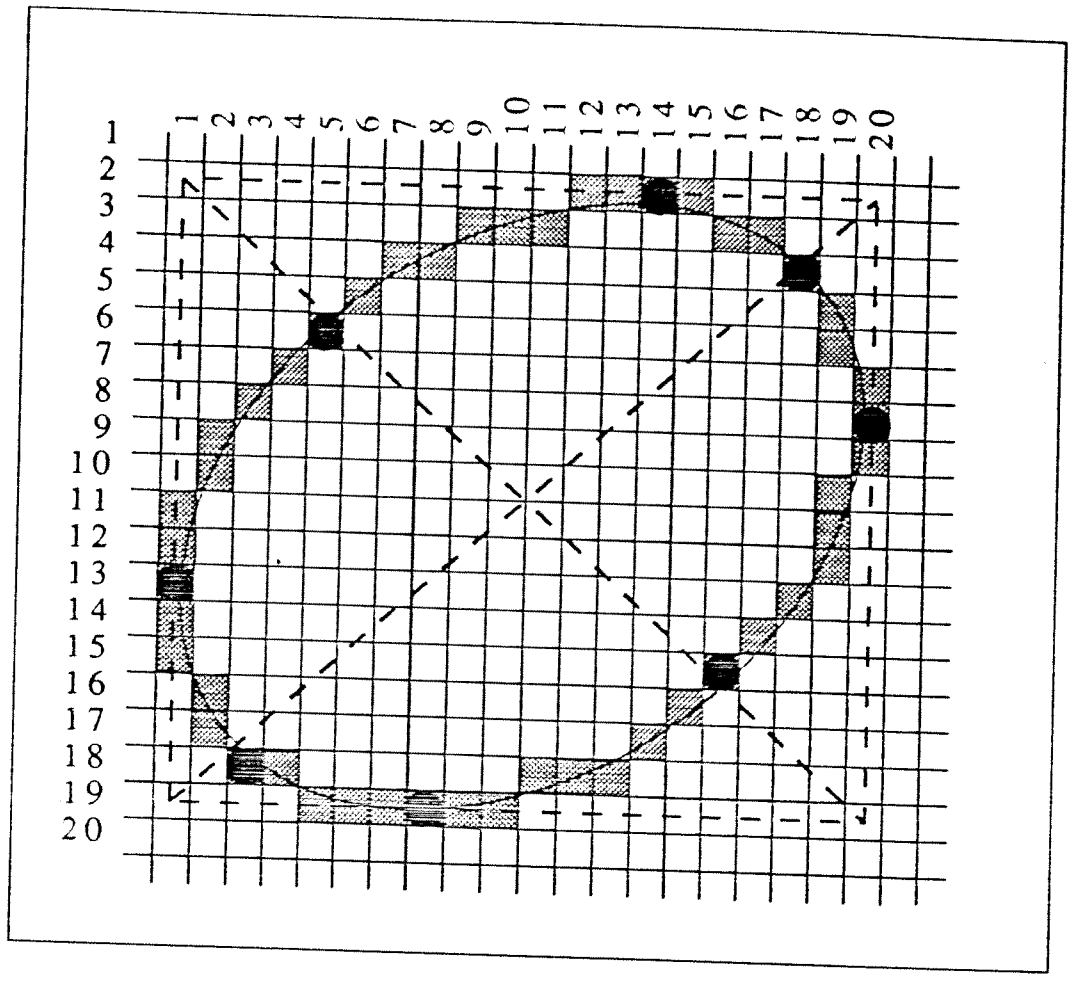


Figure 3.11: MaxMinCross Pixel Selection

3.7.1 Pixel Selection Using MaxMinCross Type

Eight pixels are selected along the contour as shown in Figure 3.11. The theoretical image ellipse and the contour pixels are shown and the selected pixels are marked.

The selection is achieved as follows:

1. The enclosing box is detected using the contour data. The first 4 pixels are selected at the regions where the digital curve is tangent with the enclosing box. If the tangency of the edge of the box and contour have more than one pixel in common, the pixel at the mid point of the tangency is selected. In cases where the mid point value turns out to be a sub pixel point, its value is rounded and the adjacent pixel is selected. This method gives

an approximate information about the major and minor axes of the image ellipse.

2. The diagonals of the enclosing box are determined and the four remaining pixels are selected at the intersection of the diagonals and the contour of the image ellipse.

3.7.2 Pixel Selection Using EqualArcs

The EqualArcs method consists of selecting N pixels along the contour, such that it is partitioned uniformly. If the number of contour pixels is less than N, then all the pixels are selected for curve fitting. When all the contour pixels are selected, the relative computation time was found to be two times greater than the MaxMinCross selection type.

3.8 Simulation Results

The method of selection and the number of pixels have different effects on the reconstruction of different variables. R_x, R_y and T_z are the only parameters directly influenced by the pixel selection methods.

- T_z is sensitive to the selection type. Errors in reconstruction of T_z , by using MaxMinCross is less than the errors obtained using EqualArcs method.
- R_x ,R_y are sensitive to the number of pixels used. Reconstruction errors in R_x and R_y , for mid-range rotations are small for MaxMinCross, but for highly rotated target configurations smaller errors are obtained by selecting all contour pixels.

To correct the numerical errors for the following cases a method was developed in [1]:

- $\alpha \approx 0°$ and/or $\beta \approx 0°$
- $|\alpha| > 80°$ and/or $|\beta| > 80°$

The errors are due to the sign inversion of the ellipse coefficients D or E at the curve fitting. The sign correction method consists of inverting the sign of D or E as follows:

1. Apply the curve fitting operation to obtain A, B, C, D and E
2. Check the following equation

$$-sgn[B] = sgn[DE]$$

- If the check is OK continue the solution process
- If the check fails then
  - if $|D| < |E|$ then D=-D
  - if $|E| < |D|$ then E=-E

The algorithm is not able to detect sign errors if both D and E have reverse signs.

For the case where $|\alpha| > 80°$ and/or $|\alpha| > 80°$, the solution to avoid sign errors is to increase the number of selected pixels.

The image plane distance d_o which denotes the perpendicular distance between the image plane and optic center, should be known for the solution process. In the simulation of the algorithm, two different methods to select d_o is tested.

- Use of constant d_o such that $d_f \leq d_o \leq 2d_f$
- Iterative solution to determine d_o

Iterative method was invented in [1] to estimate d_o. The method for the estimation of d_o can be summarized as:

1. Take $d_o^{(i)} = d_f$ as a first guess for $i = 1$
2. Reconstruct d
3. Use $d_o^{(i+1)} = d\ d_f/(d - d_f)$
4. Increment i by one
5. Repeat steps 2 to 4 until $\left|d_o^{(i+1)} - d_o^{(i)}\right| < \epsilon$ where ϵ is the convergence criteria
6. Process the remaining portion of the reconstruction algortihm

The effect of selecting d_o on configuration parameters are studied in [1]. As a result of study following conclusions are made:

- Determination of d_o does not have a considerable effect on the reconstruction of T_x and T_y
- R_z error is not affected by the determination of d_o
- Over or under estimation for R_x and R_y are observed when constant d_o is selected. Iterative estimation of d_o results in better predictions for R_x and R_y
- Estimation of d_o iteratively gives better results in reconstructing T_z

About the selection of target parameters R, r_o and r_f general guide lines are given in [1].

- A formula for the selection of R such that the image will cover approximately 10% of the sensor area

$$R \approx \frac{d}{d_f} S_s \sqrt{\frac{10\%}{2\pi}}$$

- The radii of the spots given by r_o should be kept as small as possible, considering that these spots should be observable in different target configurations and for various lighting conditions

- A practical value for distance, r_o, of the outer spot's center to the center of the main circle was given as:

$$r_f \approx (R + r_o)/2$$

All of the above results are valid for the target positioned in

$$|\alpha| < 82°$$

$$|\alpha| < 82°$$

and with a distance T_z in the range

$$500mm < d < 1000mm,$$

using a 28mm lens. If the target is placed farther than 1000mm, limits for α and β will be further reduced.

CHAPTER IV

EXPERIMENTAL SET-UP

In order to make controlled environment experiments to test the algorithm presented in Chapter 3, an experimental set-up is constructed.

In the experimental set-up, all the components of a digital imaging system -optical system, sensor, frame grabber- must be present to get images of the target. On top of these, an optical bench is required in order to compare the configuration variables reconstructed by using the algorithm with the actual values of the configuration variables. In this chapter, components of the digital imaging system and the optical bench used in the experiments are explained.

4.1 Frame Grabber

The frame grabber used to grab images is a Data Translation DT3851 Series Board. The DT3851 is a one monitor flexible frame processor for the IBM personal computer AT or compatible computer systems. The Board [1] is fully programmable to acquire and display a wide variety of video formats [6].

The board is built around the onboard graphics processor, a Texas Instruments TMS34020. The graphics processor controls video input, video display, and DT-connect transfers, and performs a limited amount of image processing. Firmware commands manage all board operations. Firmware commands eliminate the extensive register programming required to perform an operation, thereby simplifying programming. The firmware commands are downloaded to

[1] The Board refers to DT3851 Series board in the text

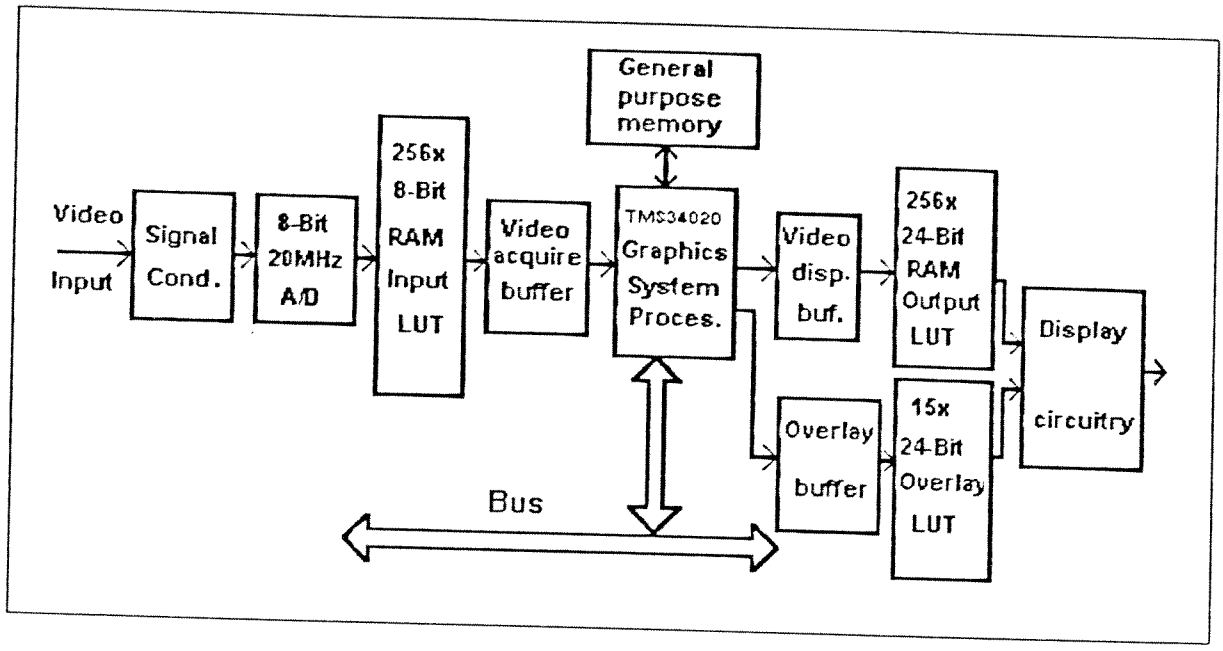


Figure 4.1: Simplified Block Diagram Representation of the DT3851

the board by an application and provide the interface to the onboard graphics processor.

The Board can acquire images up to 2000 pixels by up to 2000 lines with a sampling rate up to 20MHz. The video input circuitry, with its programmable format memory and video clock, accommodates acquired images of variable size.

The Board has three video input modes, namely RS-170 acquire, slow-scan acquire, asynchronous reset-acquire. Each of the three modes can operate using either an internal or external pixel clock. During acquisition, the TMS34020 transfer the digitized data to the video acquire buffer until the end of the frame or window is reached. The TMS34020 can transfer this image to the video display buffer for viewing or to one of the onboard buffers in auxiliary memory for subsequent processing or viewing.

A VGA pass through port allows the Board to drive a single monitor. The VGA passthrough port provides the system monitor on power-up and under DOS. The Board does not display under these conditions. Under windows, the Board provides both system display and image display.

The Board has separate video acquire, video display, overlay, and auxiliary memory buffers. Auxiliary memory contains both firmware commands downloaded by application, and auxiliary buffers of user configurable size. These auxiliary buffers are available to store program instructions, intermediate image processing results and acquired images.

4.1.1 Video Acquisition

The Board acquires images directly from one of four possible software-selectable video input sources. An entire frame is digitized at rates 0 to 20MHz. An image up to 2000 pixels by up to 2000 lines is obtained by digitizing the video signal. There are 8-bits per pixel and hence 256 gray levels can be obtained.

The Board accepts monochrome video input signal from an RS-170 / NTSC(60Hz), CCIR(50Hz), slow-scan, or nonstandard source such as a scanning electron microscope, CT scanner, or magnetic resonance imaging machine.

Signal conditioning can be performed before digitization. For acquisition of NTSC or PAL composite color signals, the Board provides a chrominance notch filter. This filter removes color information from these signals by eliminating a narrow band of frequencies. A programmable gain factor, a multiplier for the incoming video signal, can be used to adjust brightness of the image. A software selectable offset, which determines the video signal level below which all data will be digitized can be used to shift input voltages for optimum resolution.

The frequency of the video input clock determines the video input signal digitization rate. The Board can digitize the incoming signal at a rate up to 20MHz. To program the video input clock, the length of the horizontal line (in time) and the number of pixels must be known.

The portion of each line of video signal to be digitized and stored in memory is determined by the horizontal format memory. This portion is specified

by starting and ending pixel values, where each pixel is defined by a tick of the digitizing clock. The total number of pixels per line is calculated by the following formula:

$$Total\ number\ of\ pixels\ per\ line = \frac{clock\ frequency}{horizontal\ frequency}$$

The starting and ending pixel numbers are determined by the horizontal synchronization pulses available from composite video input signal.

The vertical format memory determines the lines in each video frame to be digitized and stored in memory. The lines are specified by a starting and ending line number. The total number of lines per frame is calculated by the formula:

$$Total\ lines = \frac{horizontal\ frequency}{vertical\ frequency}$$

Vertical synchronization pulses determines the starting and ending line numbers.

4.1.2 Input Look-Up Table

The Board has a 256×8-bit input look-up table (LUT). This LUT can be loaded with different processing setups in order to pass an image unaltered, or to perform pixel point operations such as image multiplication and division, intensity correction, and reverse video.

The digital value of each 8-bit pixel passes from analog to digital converter into the input LUT. The input LUT retrieves the output value for that particular pixel and passes the output value to the video acquire buffer.

4.1.3 Output Look-Up Table

The Board provides a 256×24-bit color LUT for manipulation of false color or intensity values assigned to pixel data. Each 24-bit entry in the table contains an 8-bit red, green, and blue value, one each for the red, green and blue outputs.

4.1.4 Overlay

The Board has a 4-bit non-destructive overlay. This overlay allows to mix stored graphics or text with live video or stored images. A pixel mask allows to select enable/disable bit planes without affecting the data stored in overlay memory. Under Windows in the single monitor environment the system displays in the 4-bit overlay when the VGA passthrough mode is disabled.

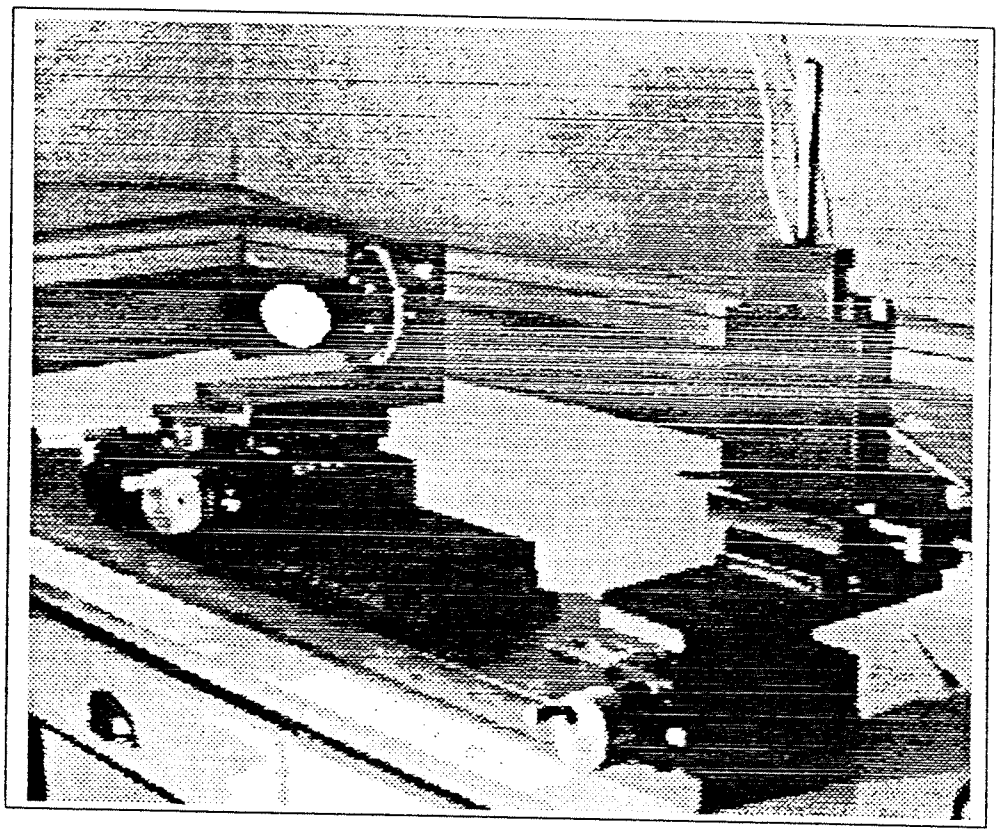


Figure 4.2: The Optical Bench

4.2 Optical Bench

In order to obtain the accurate measurements of the target configuration, an optical bench is designed and constructed. It consists of mainly three separate units:

- Target plane positioning unit

- Camera positioning unit
- Base plate

An image of the optical bench is shown in Figure 4.2.

4.2.1 Target Plane Positioning Unit

Totally 4 degrees of freedom are available to position the target plane. Two rotational and two translational measurements can be taken.

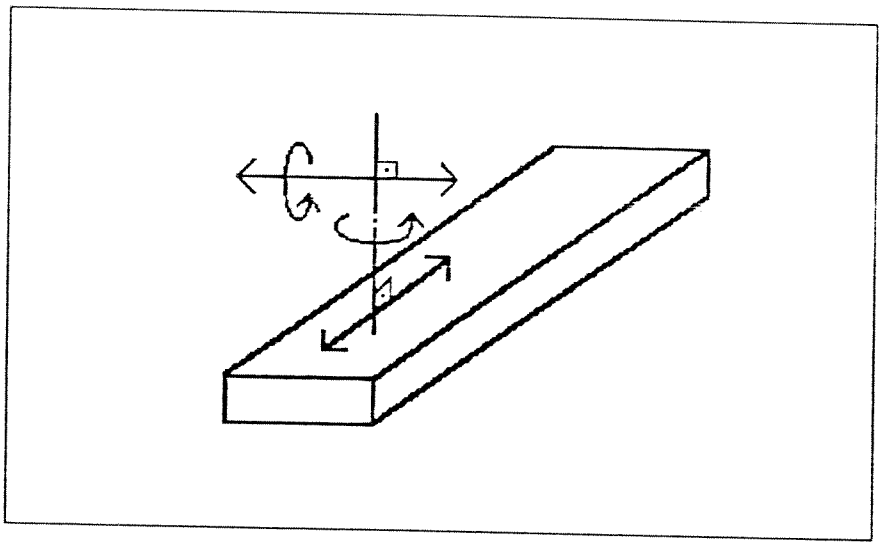


Figure 4.3: Positioning Freedoms For The Target Plane Positioning Unit

The origin of the target plane can be positioned with respect to the camera, in a horizontal plane perpendicular to the camera's sensor plane. Orientation of the target plane can be determined by using the rotational degrees of the positioning unit. The illustration of the target plane positioning unit's positioning freedoms are shown in Figure 4.3. Also its picture is shown in Figure 4.4.

Target plane positioning unit consists of the following Daedal Inc.'s positioning components:

- M4965-12 linear table

- M30006P rotary table
- M4414 linear stage
- M20000 rotary stage
- Two angle brackets

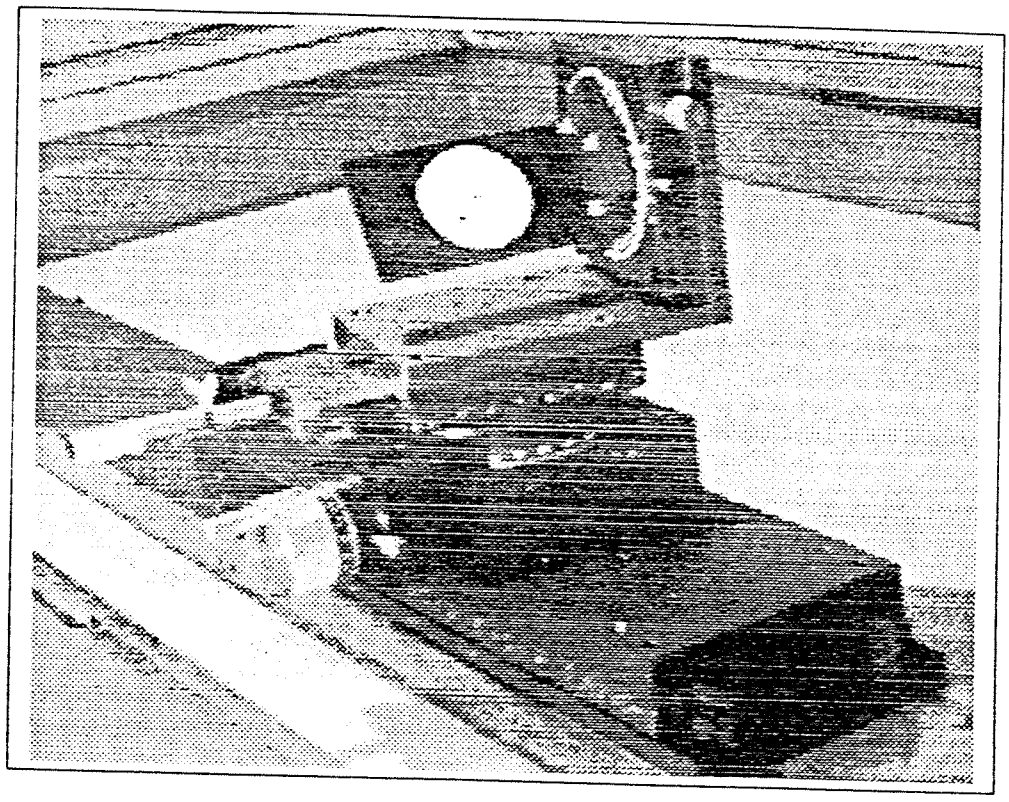


Figure 4.4: The Target Plane Positioning Unit

4.2.2 Camera Positioning Unit

Actually, camera positioning unit is used to locate the sensor plane of the camera, principal point and optical axis of the optical system with respect to the base plate. Three translational and one rotational degrees of freedom are available to position them.

Three linear perpendicular (XYZ) axes are available on a rotary stage. Figure 4.5 depicts the camera positioning unit's positioning freedoms. Figure 4.6 shows a picture of this unit.

Camera positioning unit consists of the following Daedal Inc.'s positioning components.

- M30006P rotary table
- M4434 (XYZ) positioner
- Plate

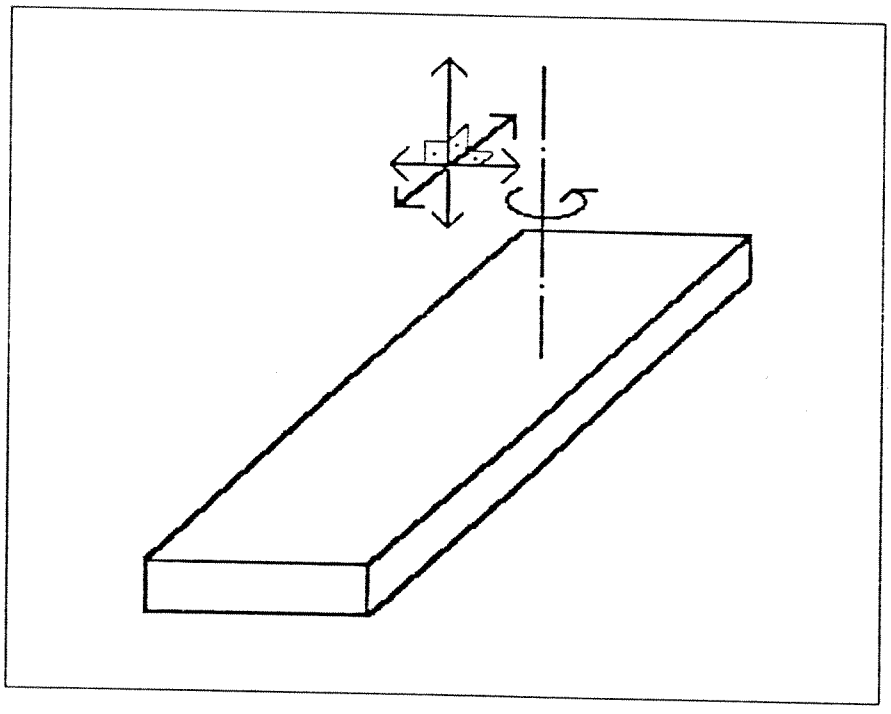


Figure 4.5: Positioning Freedoms For The Camera Positioning Unit

4.2.3 Base Plate

In order not to have deformations in time, base plate was made of cast iron. To release the residual stresses which are the casue of deformations in time, base plate is heat treated. Finally it was grounded to obtain a flat surface. Mounting holes were drilled on this flat surface. Distances between the mounting holes was measured by using a coordinate measurement machine such that the distance between the target plane and camera positioning units can be determined accurately.

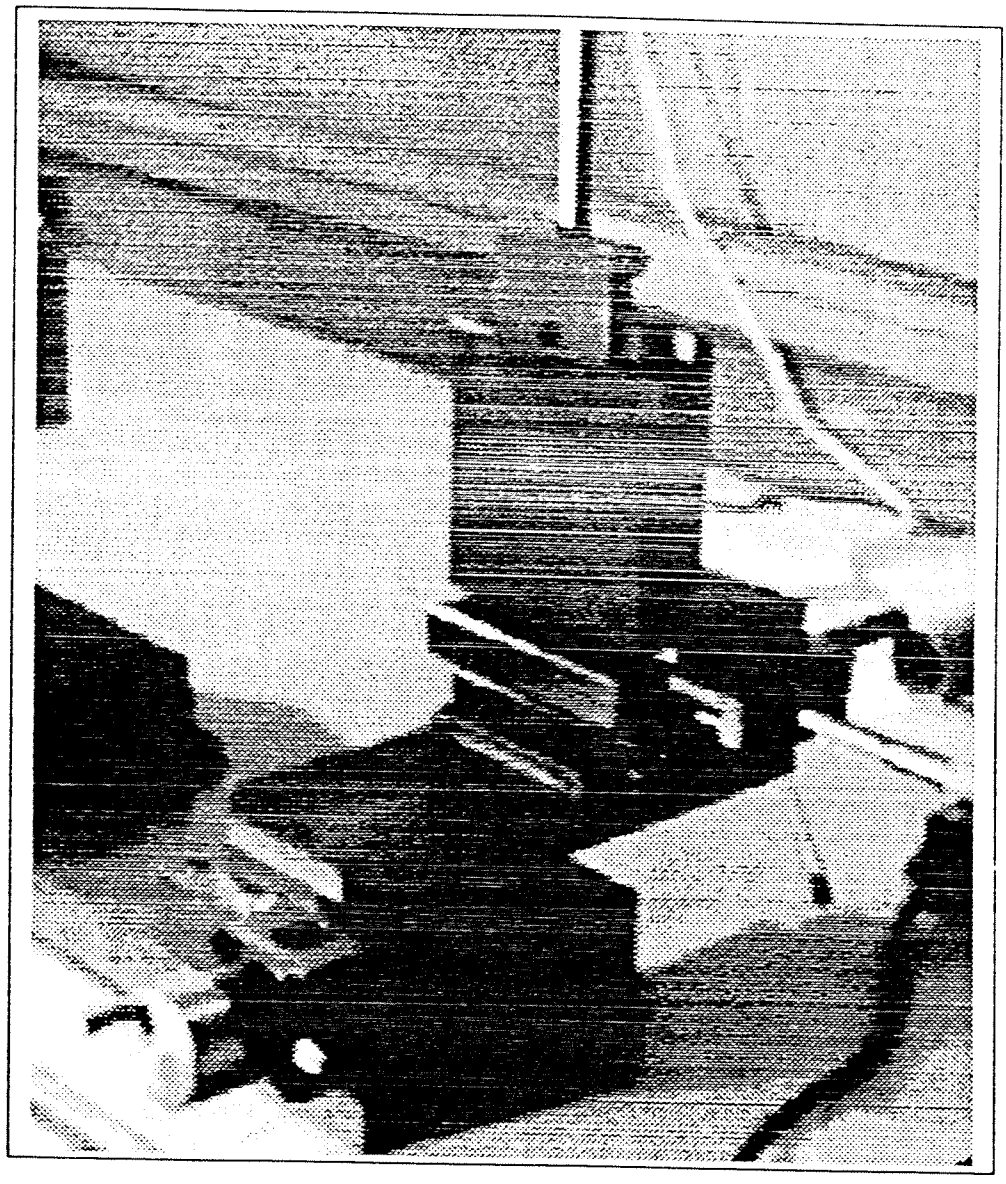


Figure 4.6: The Camera Positioning Unit

4.2.4 Technical Specifications of the Optical Bench

Technical specifications of the individual components the units and the base plate are as follows:

- M4965-12

  A linear positioning stage having a maximum travel of 300mm with a 0.01mm resolution readout. This stage offers extremely smooth linear and precise point to point positioning capability with no backlash or sideplay.

  Travel : 300mm

  Resolution : 0.01mm

  Straight line accuracy: 2μm/25mm of travel

  Positional repeatability: 1.3μm

- M4414

  This stage is a precise linear positioner that offers a 50mm travel with 0.01mm resolution micrometer readout. Two or three of this stages can be combined to obtain 2-axis X-Y or 3-axis X-Y-Z units.

  Travel : 50mm

  Resolution : 0.01mm

  Straight line accuracy: 2μm/25mm of travel

  Positional repeatability: 1.3μm

- M4434

  Three of M4414 linear stages are combined by using an angle bracket to obtain M4434 3-axis X-Y-Z linear positioning unit.

  Travel : 50mm in each axis

  Resolution : 0.01mm

  Straight line accuracy: 2μm/25mm of travel

Positional repeatability: 1.3μm

- M30006

A rotary table that is used to position rotary motion accurately. Its drive mechanism provides precise and accurate rotational positioning with no backlash, assuring bi-directional accuracy and repeatability. An angular readout, graduated in degrees, is provided around the circumference of the table top, while fine position readout dial, found on the control knob, reads directly in 0.01° increments, with the vernier providing 0.002° resolution.

Travel : 360° continuous

Resolution : 0.002°

Runout : 25 μm

Concentricity : 0.3 μm

- M20000

The M20000 rotary positioning stage provides smooth continuous adjustment over a full 360° travel range with a calibrated dial and vernier, readable to 6 arc minutes.

Travel : 360° continuous

Resolution : 6 arc minutes

- Base Plate

Base plate is a flat surface on which mounting holes are available to mount target plane and camera positioning units. It is 1500mm in length and 160mm in width. Technical drawing and measurements for determination of specifications of base plate are given in Appendix B.

Flatness : $\pm$0.1mm in longitudinal direction, $\pm$0.4mm in tranverse direction

Accuracy of distances between mounting holes : $\pm$0.1mm

Finally combined technical specifications of the optical bench can be given as:

- Orientations of the target plane can be determined in three axis as:

  around x-axis, 0-360° , 0.1° resolution

  around y-axis, 0-360° , 0.002° resolution

- Origin of the target plane can be positioned with respect to the camera in three axis as:

  x-axis, 0-50mm, 0.01mm resolution

  y-axis, 0-50mm, 0.01mm resolution

  z-axis, 300-1300mm, 0.01mm resolution

4.3 CID Camera

The CID camera is a CID Technologies Inc.'s CID2250D model. It has a 512H×512V CID35 solid state sensor. The CID35 imager has square pixels 15μm×15μm in size. The square pixel format eases mathematical analysis of the video information. The CID's inherent anti-blooming property and contiguous pixel topography allows highly accurate spatial measurements.

The CID imager is read out in a sequential noninterlaced manner, one field per frame. The camera's video interfaces directly to standard industrial frame grabbers supporting variable scan or slow scan operation. The specifications of CID2250D are as follows:

- Imager Format

  Total pixels :522H×528V

- Displayed pixels : 512H×512V (532 mode)

  Displayed pixels : 512H×505V (525 mode)

  Pixel size : 15.0μm×15.0μm

- Optical Format

  Nominal 11.0mm (2/3") diagonal

- Scanning Format

  Progressive/Sequential

- Lens Mount

  Standard C mount

4.4 Optical System

Two different lenses are used to represent 3-dimensional scene on the 2-dimensional sensor plane.

- Cosmicar CCTV lens having a focal length of 25mm, maximum aperture ratio 1:1.4 and 1 inch image size
- RS lens having a focal length of 16mm, maximum aperture ratio 1:1.4, and 2/3 inch image size

Although the camera's image size (2/3 inch) and Cosmicar lens's image size do not match, it is possible and also better to use a lens having larger image size than the camera's due to reduced level of geometric distortion. By using a lens having larger image size, the area in the image in which geometric distortion is minimum is inherently chosen.

The technical specifications of the lenses are given below:

- Cosmicar CCTV lens

  Focal length 25mm

  Maximum aperture ratio 1:1.4

Angle of view

Diagonal 24°48′

Horizontal 19°58′

Vertical 15°02′

Minimum object distance 60mm

- RS lens

Focal length 16mm

Maximum aperture ratio 1:1.4

Angle of view

Diagonal 38.5°

Horizontal 23.2°

Vertical 38.4°

Minimum object distance 25mm

CHAPTER V

IMAGE PROCESSING AND ANALYSIS

5.1 Overview

In machine vision applications, features which characterize the scene, viewed by a camera, are required to be extracted from an image. Features to be extracted can be edges of parts on an assembly line, geometric centers of objects or specially marked points for camera calibration etc.. The image of the scene, stored in memory by a frame grabber, is usually in raw data form. In order to make computations on the image, features of the image must be identified mathematically and represented as a mathematical data.

The first step in machine vision is usually a preprocessing operation intended to make subsequent processing and analysis easier. This can be accomplished by either removing unnecessary or disturbing details or by transforming the image into a more useful one. Image processing involves the operations of data manipulation, like image enhancement and image restoration.

The image enhancement is to emphasize some special features of an image for the purpose of analysis or display. Therefore, a rough image is processed using special techniques to obtain a more understandable image. The enhancement process does not increase the information content of the data but makes desired features more perceivable. Some examples are contrast and edge enhancement, noise filtering, sharpening and magnifying.

The image restoration deals with the removal or minimization of known degradations in an image. This includes deblurring of images, noise filtering and

correction of geometric distortions due to sensors. Using restoration techniques one seeks the best estimate of the image as if the undesired disturbances were absent.

The image analysis is a post processing operation. It consists of extracting the desired features of the image. In this study the main effort is spent on extracting the desired features from the image. Simple but effective image processing operations are used.

5.2 Image Processing

The first step in machine vision is image processing. Its aim is to obtain an image in which desired features appear distinctly. Several different factors effect the appearance of an image. Among them the most important one is the lightning conditions. It plays an important role on the quality of the image. Special selection of the object being viewed and lighting conditions will diminish the required image processing. In many situations, the objects or features of interest are darker (or lighter) than their background. We can ensure that this is true by controlling the background or the lighting.

Two major preprocessing techniques are point transformations and intensity histograms.

A point transformation replaces the gray level value of each pixel by a new value that is obtained as a nonlinear function of the old gray level. This is typically achieved by comparing the pixel data stored in an image memory with a look up table that provides a new value for each pixel, depending on its current value. The new values then replace the original ones in the image memory. Possible applications for such preprocessing techniques include contrast enhancement, range expansion, single or multiple level thresholding, and gray level inversion.

Intensity histogram lists the number of times each pixel intensity occurs in an image and can be a very useful tool for analysis. For example, a scene including an object will have significantly different histogram than one without that object. The histogram can also be used to establish appropriate threshold levels and help to determine the intensity values to be used in point transformations.

Thresholding is a filtering technique that can be used to create a binary image from one originally containing various shades of gray.

The separation of thresholding replaces each pixel with a 1 or a 0, according to whether it is above or below threshold level. Figure 5.1 shows a thresholding transformation. If the image contains a bright object on a dark background, then the histogram will exhibit two distinct modes as shown in Figure 5.2. Choosing the deepest part of the valley between the modes will often yield a good threshold.

The easiest thresholding is global thresholding. It separates object from the background using a fixed gray level or threshold value. This threshold divides the image pixels into two classes -object pixels (whose gray values lies on one side of the threshold) and background pixels (whose gray value lies on the other side). Global thresholding requires high contrast between objects and background and gray levels of the objects and background on different sides of the threshold value.

5.3 Image Analysis

The ultimate aim in a large number of image processing applications is to extract important features from image data, from which a description, interpretation, or understanding of the scene can be provided by the machine. For example a vision machine system may distinguish parts on an assembly line and list their features such as size and number of holes. In robotic applications 3D

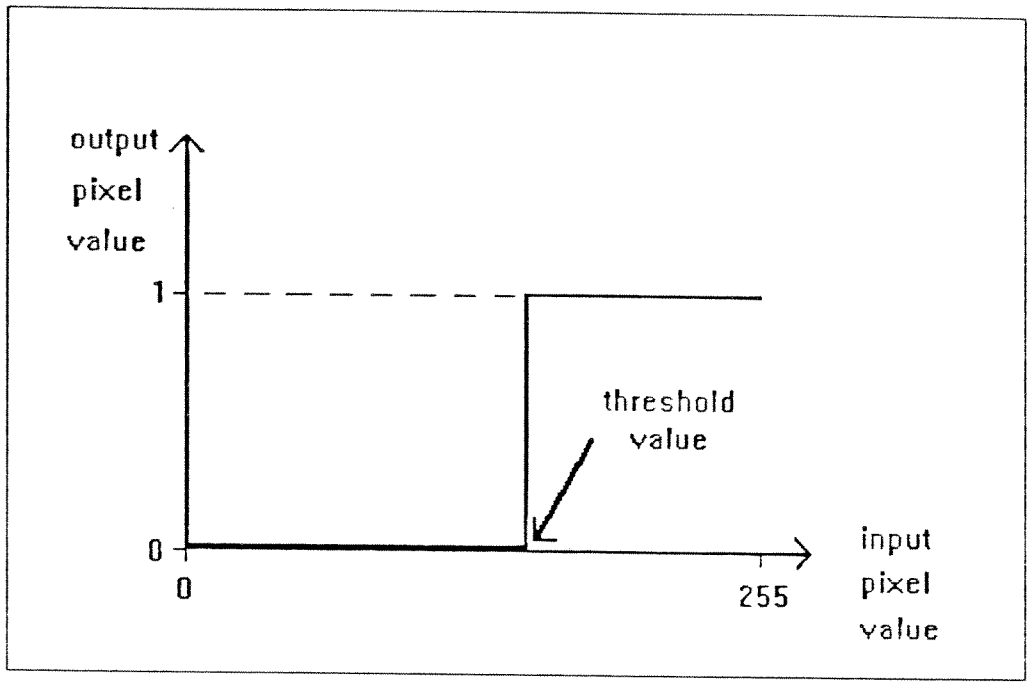


Figure 5.1: Thresholding Transformation

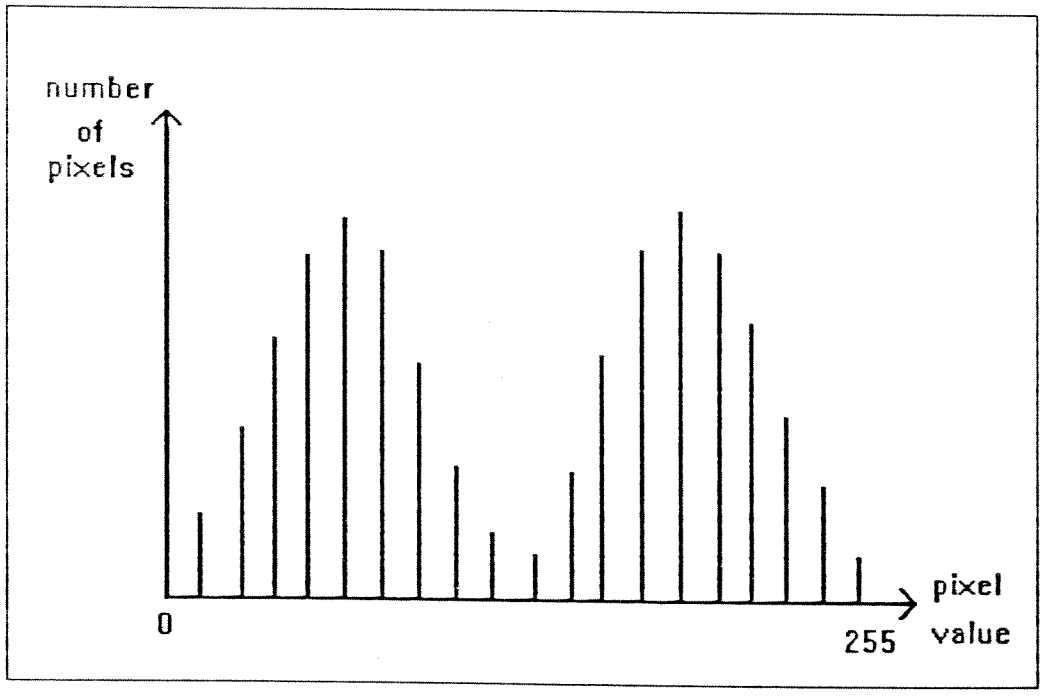


Figure 5.2: A Typical Histogram

configuration information can be the output of the image analysis. More sophisticated machine vision systems are able to interpret the results of analysis and describe the various objects and their relationships in the scene. In this sense image analysis is quite different from other image processing operations, such as restoration, enhancement, where the output is another image. Image analysis basically involves the study of feature extraction, segmentation, and classification techniques.

Segmentation outlines objects in images that require further processing like separation of different objects by extracting their boundaries. Under controlled lightning conditions, the result of global thersholding can be a segmented image. Once the segmented image is obtained, the desired features of the image can be extracted without applying any further image processing.

Feature extraction calculates global features of outlined objects during segmentation. Examples of global 2D features commonly used in machine vision systems, are number of holes, area, perimeter, centroid, length of minimum and maximum radii from the centroid to the perimeter, locations of specially desired points, etc..

Classification matches objects to model descriptions based on their global features. A chosen classifier or decision rule assigns each object to a class of objects with similar properties.

Edge detection and boundary extraction are the two important operations of the image analysis. Edge points in a binary image can be defined as black pixels with at least one white nearest neighbor. Edges characterize object boundaries which is useful in identification of objects in scenes. Convolution operators or simple global thresholding can be used to obtain edge of an object. Under controlled lightning conditions thresholding will directly yield the edge of an object. For some hardware configurations this means that edge of the object is obtained online by applying point transformations. So that operation speed is increased tremendously. Convolution operators can not be used online. Since in

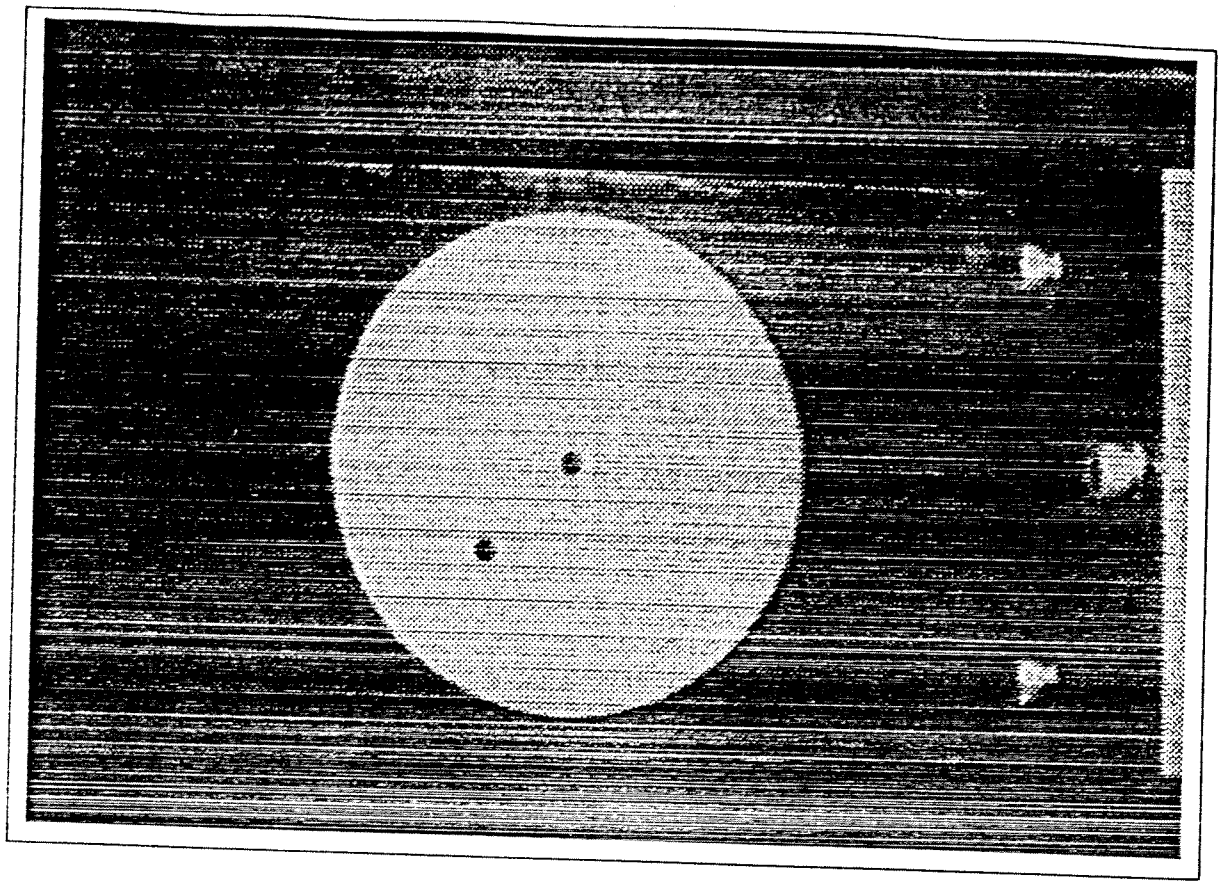


Figure 5.3: Image of The Target

convolution a 3 by 3 matrix operator must visit everywhere of the image which needs considerable amount of time regarding the operation speeds required in industrial applications.

Boundaries are linked edges that characterize the shape of an object. Contour following algorithms trace boundaries by ordering successive edge points. Output of edge detection and contour following is the identified and represented object with manipulable data.

5.4 Image Processing and Analysis Used in This Study

The algorithm presented in Chapter 4 requires extracting some features from the image of the target. These are the contour pixels and centroids of the two spots. To obtain these features with the minimum amount of processing,

the target, its background and illumination is carefully selected.

The target used is a circle with two internal spots. The spots and the outer region of the circle is black. The inside of the circle is white. It is printed with a laser printer on a white page. A picture of the target is shown in Figure 5.3. The background of the target is considerably darker than the target. Therefore a high contrast ratio is obtained for the scene. The target is illuminated by incandescent bulbs. The brightness of the image of the target is controlled by changing the aperture ratio of the lens. For small aperture ratios ligther images and for high aperture ratios darker ones are obtained.

On line point transformations can be performed by using the look up tables of the DT3851 frame grabber. Using this facility of the grabber the processing speed can be increased considerably. An analysis of the histograms for the scene viewed shows that the edge of the circles and spots can be segmented by simple global thresholding. A sample histogram of the scene is given in Figure 5.4.

After thresholding, the image obtained is stored in memory. In Figure 5.5 thresholded image is shown. Since only the target circle and its internal spots are needed for feature extraction, the region of interest is marked manually by a mouse and all subsequent operations are performed on the marked region. To obtain the desired features, first of all, a contour following operation must be performed.

5.4.1 Contour Following

A contour following algorithm, yielding a coarse contour, with some of the boundary pixels appearing twice is given in [16]. This algorithm is modified such that boundary pixels appear twice only in the case of single pixel line extensions. With this modification it is possible to clip external extensions. Once the algorithm starts just outside of a region, it will always follows the boundary

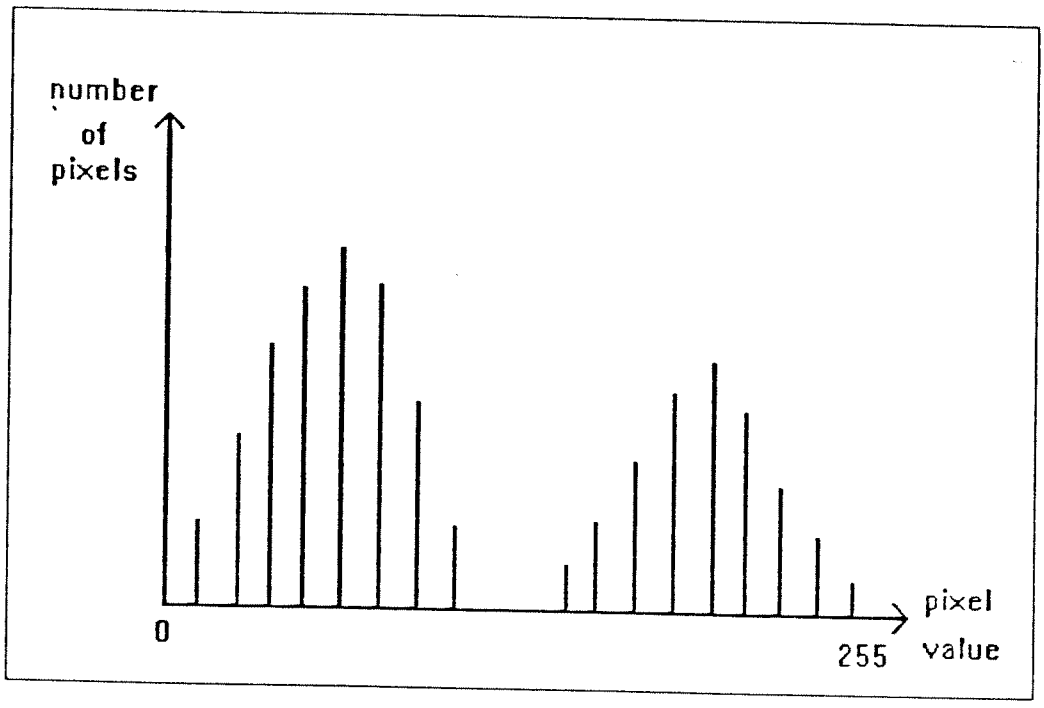


Figure 5.4: Sample Histogram

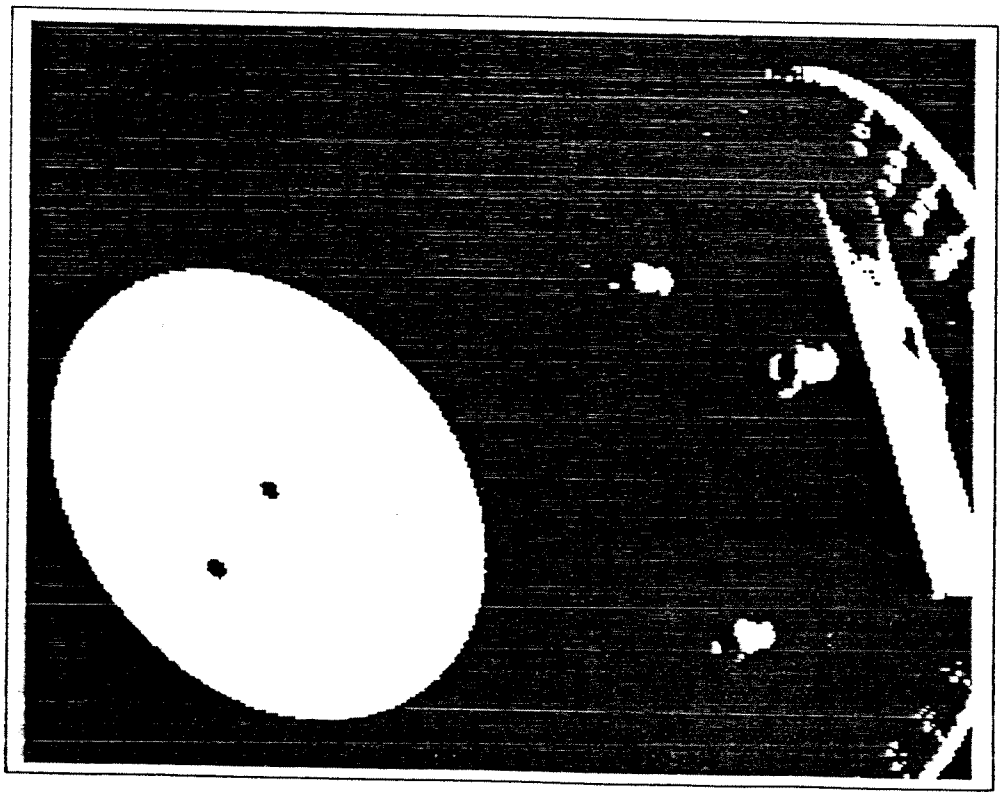


Figure 5.5: Thresholded Image of The Target

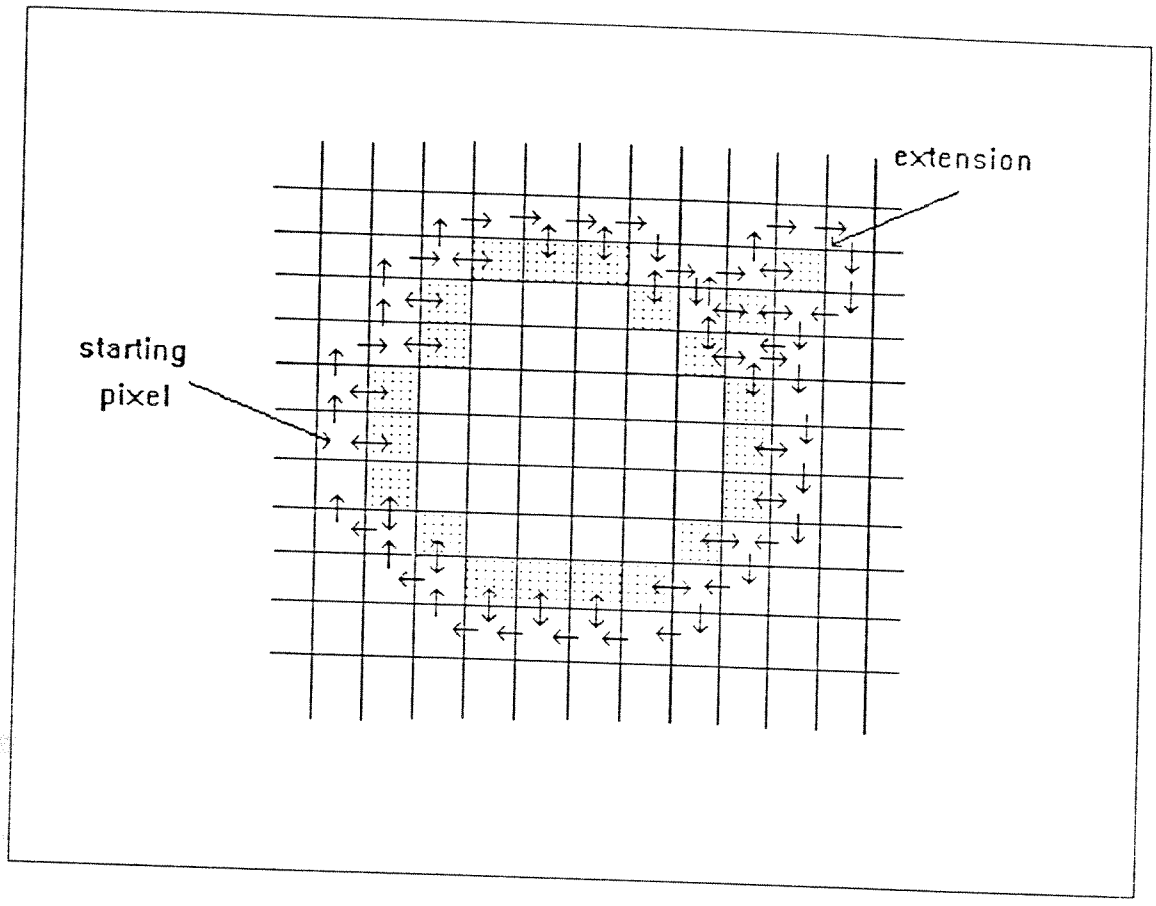


Figure 5.6: An Example Contour Following

until it comes to its starting point. The following outlines the algorithm used in contour following:

1. Start just outside of the region
2. Move forward 1 pixel
3. If the arrived pixel address is equal to the starting pixel adress, then stop
   Else continue
4. If the arrived pixel is white, then turn right
   Else move backward 1 pixel and then turn right
5. If the arrived pixel is white, then turn right
   Else move forward 1 pixel

6. If the arrived pixel is white, then move forward 1 pixel
   Else return to step 3

7. If the arrived pixel is white, then turn right
   Else return to step 3

8. If the arrived pixel is white, then turn right and return to step 3
   Else return to step 3

An example contour following using this algorithm is given in Figure 5.6.

Output of the contour following algorithm is the (x, y) addresses of the boundary pixels of the circle. The same algorithm is also used to detect the boundaries of the internal spots. Once the boundary pixels are available next step is to find the pixel addresses of the 8 special pixels specified by MaxMinCross type pixel selection given in section 3.7.1. The first pixels are easy to determine as specified in the first step of the selection. The pixel addresses will be the pixels having maximum and minimum x and y addresses.

5.4.2 Diagonal Following Algorithm

The second step in which following of a diagonal in a image is necessary requires special interest. Because not a boundary pixel should be missed and the path followed must be as close as possible to the diagonal of the enclosed box. Therefore a new path following algorithm is invented satisfying these two requirements.

The enclosing box of the circle can be determined by using the pixel addresses of the contour of the circle. Then the starting and ending points of the diagonal is determined which are the corners of the box. The following is the algorithm:

1. Determine the aspect ratio

$$aspect\ ratio = \frac{delta\_x}{delta\_y}$$

2. Determine the minimum and maximum step sizes

$$min\_step = minimum\ integer\ closest\ to\ the\ aspect\ ratio$$

$$max\_step = maximum\ integer\ closest\ to\ the\ aspect\ ratio$$

where

$$min\_step : minimum\ step\ size$$

$$max\_step : maximum\ step\ size$$

3. Determine the number of the maximum and minimum steps

If $remainder\ of\ aspect\ ratio > 0.5$ Then

$$n\_max\_steps = maximum\ integer\ closest\ to\ (delta\_y \times remainder\ of\ the\ aspect\ ratio)$$

$$n\_min\_steps = minimum\ integer\ closest\ to\ (delta\_y \times (1 - remainder\ of\ the\ aspect\ ratio))$$

Else

$$n\_max\_steps = minimum\ integer\ closest\ to\ (delta\_y \times remainder\ of\ the\ aspect\ ratio)$$

$$n\_min\_steps = maximum\ integer\ closest\ to\ (delta\_y \times (1 - remainder\ of\ the\ aspect\ ratio))$$

where

$$n\_max\_step : number\ of\ maximum\ steps$$

$$n\_min\_step : number\ of\ miniimum\ steps$$

4. Determine the number of blocks for the maximum and minimum steps

   if $n\_max\_step > n\_min\_step$ Then

$$n\_block\_min = \frac{n\_max\_steps}{n\_min\_steps}$$
$$n\_block\_max = 1$$

   Else

$$n\_block\_min = 1$$
$$n\_block\_max = \frac{n\_max\_steps}{n\_min\_steps}$$

   where

   n_block_min : *number of blocks for the minimum step size*

   n_block_max : *number of blocks for the maximum step size*

5. Determine the total number of steps

$$total\ number\ of\ steps = \frac{delta\_y}{n\_block\_min + n\_block\_max}$$

6. Start at the corner of the enclosing box

7. Move forward min_step pixels

8. Step down 1 pixel

9. Repeat steps 7 and 8 for n_block_min times

10. Move forward max_step pixels

11. Step down 1 pixel

12. Repeat steps 10 and 11 for n_block_max times

13. Repeat steps 6 through to 11 for a total number of steps times

To demonstrate the algorithm, an example is given below and the path followed is shown in Figure 5.7

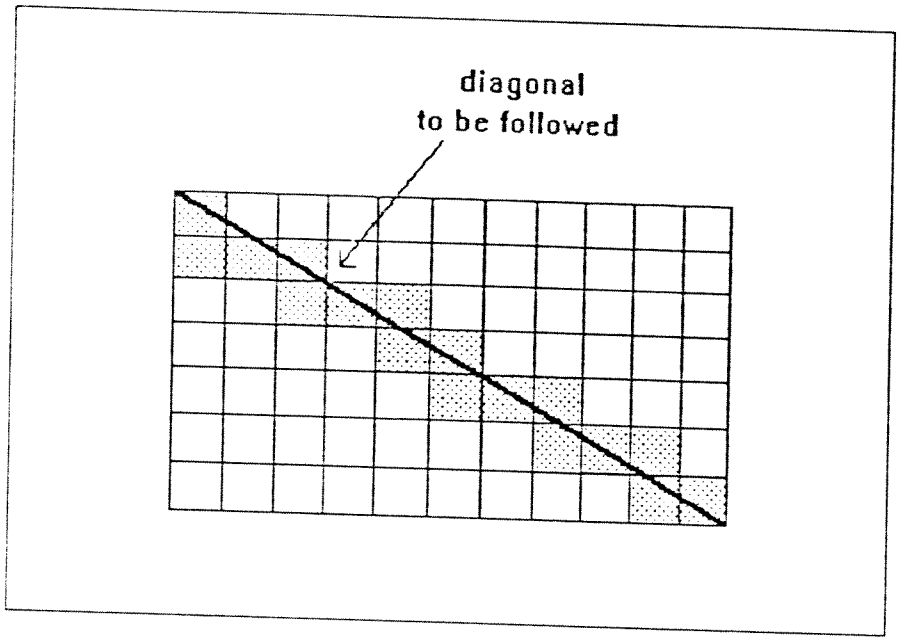


Figure 5.7: An Example For The Diagonal Following Algorithm

1. delta_x = 11

   delta_y = 7

   aspect ratio = 1.57

2. min_step = 1

   max_step = 2

3. n_max_step = 4

   n_min_step = 3

4. n_block_min = 2

   n_block_max = 1

5. total number of steps = 2

5.5 Determination of the Centroids of the Internal Spots

The contour following algorithm is used to determine the boundary pixels of the internal spots. Once the boundary pixels are determined, the cen-

troids of the spots can be calculated by using moments.

The moments of a region R is defined:

$$m_{kj} = \sum_{x,y \in R} x^k y^j \quad k \geq 0,\ j \geq 0$$

Thus the moments of R are sums of products of integer powers of the row and column numbers of pixels in R. The sum of the powers, k+j, is referred as the order of the moment m_{kj}.

Suppose that the area of R is A pixels and the centroid of R is located at $\{x_c, y_c\}$. Then:

$$A = m_{00}$$

$$x_c = \frac{m_{10}}{m_{00}}$$

$$y_c = \frac{m_{01}}{m_{00}}$$

The output of the contour following, diagonal following algorithms and centroids of the spots are the desired features of the target image. The pixel addresses of the centroids and selected pixels are then used by the reconstruction algorithm to compute the configuration variables of the target.

CHAPTER VI

EXPERIMENTS

The features extracted from the image of the target are used in the reconstruction algorithm to find the 6 configuration variables. The output of the algorithm is compared with the measurements made using the optical bench. In the available limits of the optical bench, several different configurations are tested. Also effect of misfocus and different aperture ratios on the reconstructed variables are investigated. The experiments are made according to the guidelines given in [1].

In this chapter, the calibration of the setup for accurate and meaningful measurements is first presented. Then the experimental procedure is given. Finally the results of the experiments for different test conditions are evaluated.

6.1 Calibration of The Set-Up

Before making any measurements the reference frames must be located on the optical bench. Among them extra effort must be spent to locate the sensor plane of the camera. Since the location of the sensor plane with respect to the camera casing is not known and can not be measured, it should be determined by using a characteristic point of the camera and lens. The focal point will be a valid selection to determine the position of the sensor plane with respect to the base frame. The rotation method [9] is used to make this calibration.

The statement of this method is : Given any pair of images produced by a camera which has undergone pure rotation, if the internal parameters and the

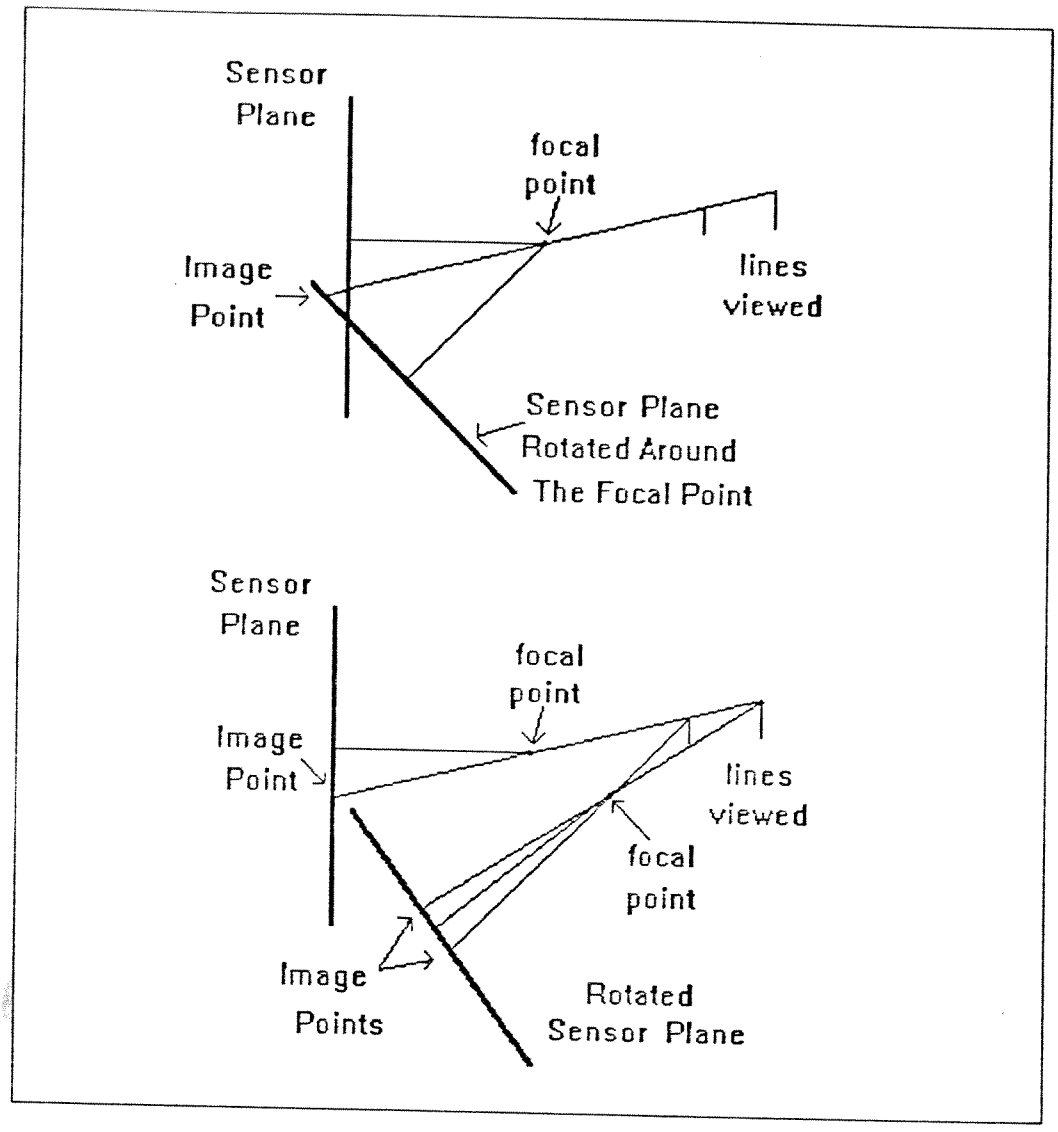


Figure 6.1: The Schematic Representation of The Rotation Method

angle and axis of rotation are known, one can compute where the feature points from one image will appear in the second image of the pair. If there is an error in the internal camera parameters, the features in the second image will not coincide with the feature locations computed from the first image. To use this method to locate the focal point, the camera is mounted first on a X-Y linear positioning stage, then this combination is mounted on a rotary stage. To obtain features a white page with a set of black lines, printed with a laser printer is mounted on a flat board and positioned 1.0 m. away from the camera. Another page with one black line about one third the thickness of the previous ones is mounted on a secondary acrylic sheet. The second page is positioned about 0.3 m. from the

camera in front of the first page so that the line in the second sheet seemed to be perfect continuation of one of the lines in the first sheet. When the camera is rotated, the lines on the first sheet would appear to move either faster or slower than the lines on the more distant sheet. By adjusting the camera position it is possible that no relative motion is detected. When there is no relative motion between lines, the focal point is positioned on the axis of rotation. If the camera does not rotate around the focal point , there will be relative motion between the lines. A schematic representation of the method is given in 2D form in Figure 6.1.

All of the origins of the coordinate frames used in the reconstruction algorithm are referenced with respect to the image center which is the intersection of the optical axis and the sensor plane. Therefore it should be determined. The numerical center of the frame buffer is a valid selection for the experimental set-up being used in this study. Because the CID camera has a contiguous pixel structure unlike the CCD cameras. Also the frame grabber stores the image in memory such that there is one to one correspondence between the CID array and the image stored in memory.

The target is mounted on the target plane positioning unit. To reference this unit with respect to the sensor plane, the center of the target has coincided with the image center. Once this is done all the micrometers used to make measurements are referenced to zero.

After making all of the above calibrations, it is possible to make measurements and compare them with the reconstructed variables.

6.2 Experimental Procedure

The output of the algorithm is a 4 by 4 coordinate transformation matrix which locates the target plane with respect to the world coordinate system. The elements of this matrix determines the following list of parameters:

1. T_x : target translation along x axis

2. T_y : target translation along y axis

3. T_z : target translation along z axis

4. R_x : target rotation about x axis

5. R_y : target rotation about y axis

6. R_z : target rotation about z axis

The parameters R_x, R_y and R_z are the consecutive rotation angles. The experimental set-up being used gives angles for a rotated frame based 213 sequence. Therefore the coordinate transformation matrix is solved for this sequence. It is possible to measure the first two rotations R_x and R_y using the optical bench directly. However the third rotation R_z can not be measured directly. Therefore it is taken to be a constant value. This constant value is compared with the reconstructed value.

Three translations T_x, T_y, T_z can be measured directly from the set-up. When a rotation around R_x or combined rotations around R_x and R_y are present, it is not possible to measure T_x and T_y directly. Hence these two axes are analyzed separately.

6.3 Results of the Experiments

Mainly two sets of experiments have been made using the full range of the optical bench.

- target located at the maximum possible T_z distance of the optical bench
- target located at the minimum possible T_z distance of the optical bench

Table 6.1: Data Set for T_x when T_z=-590mm

	Start	Stop	Step
T_x[mm]	-25	25	1
T_y[mm]	0	0	0
T_z[mm]	-590	-590	0
R_x[°]	0	0	0
R_y[°]	0	0	0
R_z[°]	45	45	0

Table 6.2: Data Set for T_y when T_z=-590mm

	Start	Stop	Step
T_x[mm]	0	0	0
T_y[mm]	-10	10	1
T_z[mm]	-590	-590	0
R_x[°]	0	0	0
R_y[°]	0	0	0
R_z[°]	45	45	0

For both of the sets several subsets are tested. All these sets are tested with a lens having a focal length of 25mm and a target having a 30mm radius outer circle, with 2mm radius internal spots. The limits of the data sets are determined by the practical points. As an example , positive rotations around the y axis can not be made. When the target plane is rotated around y axis in the positive direction, a complete image of the secondary target can not be obtained.

The first four sets that have been tested are given in Tables 6.1, 6.2, 6.3 and 6.4. These sets are used to determine the accuracy of the reconstruction of the configuration variables while changing either T_x or T_y and keeping other variables constant. The reconstruction errors for the rotations R_x, R_y and R_z are given in Figures 6.2 and 6.3. The first two rotations can not be reconstructed successfully. Errors up to ± 6 degrees are present. But the third rotation R_z is found to be approximately equal to its measured value. This is an expected result, since R_z is dependent on the location of the outer spot. It is not completely dependent on the

Table 6.3: Data Set For T_x when T_z=-1305mm

	Start	Stop	Step
T_x[mm]	-25	25	1
T_y[mm]	0	0	1
T_z[mm]	-1305	-1305	0
R_x[^{o}]	0	0	0
R_y[^{o}]	0	0	0
R_z[^{o}]	45	45	0

Table 6.4: Data Set For T_y when T_z=-1305mm

	Start	Stop	Step
T_x[mm]	0	0	1
T_y[mm]	0	48	1
T_z[mm]	-1305	-1305	0
R_x[^{o}]	0	0	0
R_y[^{o}]	0	0	0
R_z[^{o}]	45	45	0

contour of the circle like R_x and R_y. Figures 6.4 and 6.5 show the errors for the translations T_x, T_y and T_z. T_y and T_z are reconstructed successfully, especially T_y. The reason for the chatter in the T_x error is the pixel shift in the image. The target is moved in 1mm steps. This step projected on the sensor plane will cover several pixels. But it can not be fully covered by an integer number of pixels. There will be a difference. In every 1mm step movement this difference will acculumate and finally after some number of steps, it will cause a single pixel jump.

Results for the data sets in Tables 6.3 and 6.4 show that reconstruction errors for R_x, R_y and R_z have the same behaviors for a target located at a farther distance. However T_z errors increase. Since the size of image of the target gets smaller as the target moves away from the camera, amount of information implanted in the image of the target gets smaller. This will result in an increase in T_z errors. The same chattering behavior is observed for T_x and T_y. But this

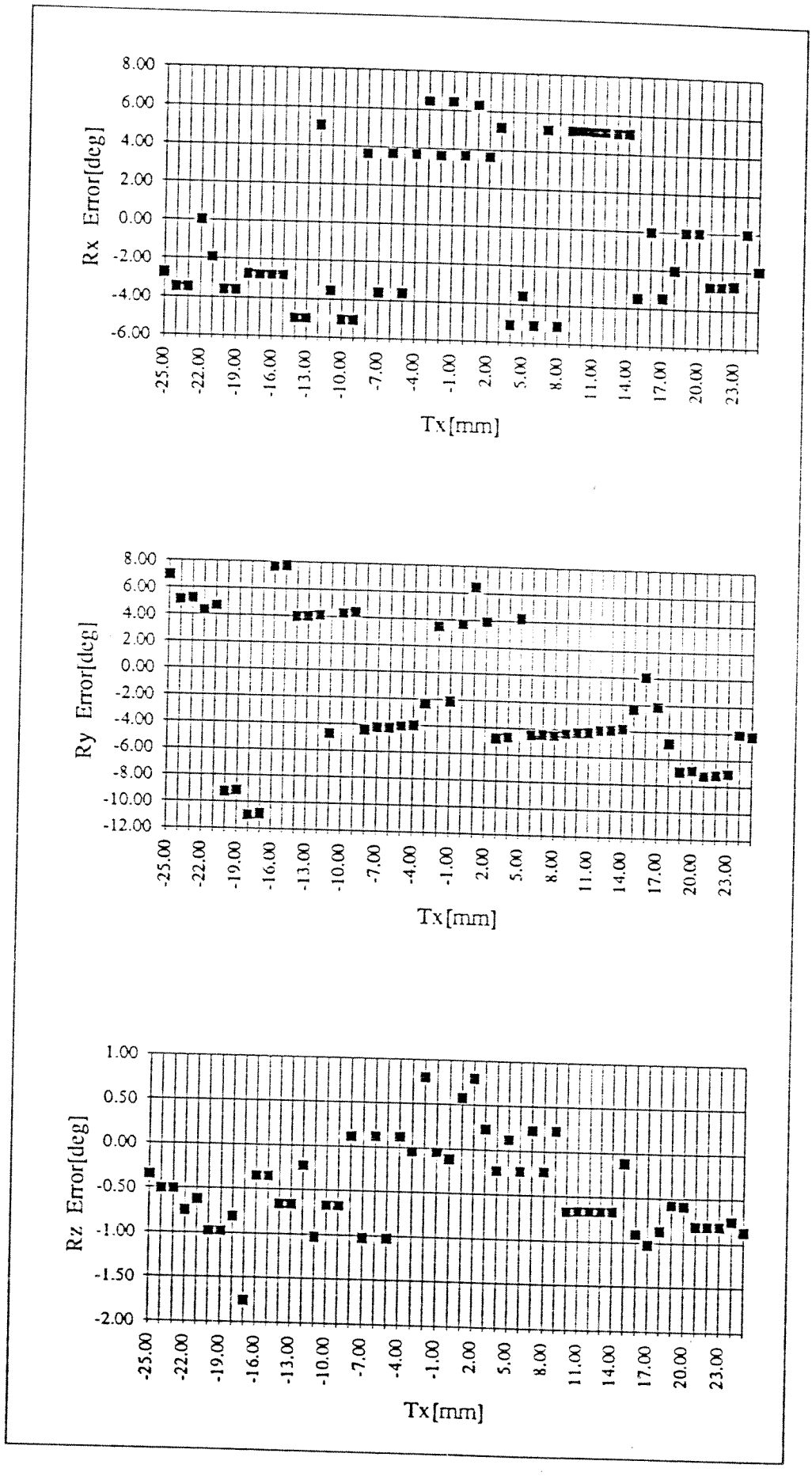


Figure 6.2: Reconstruction Errors For The Data Set in Table 6.1.

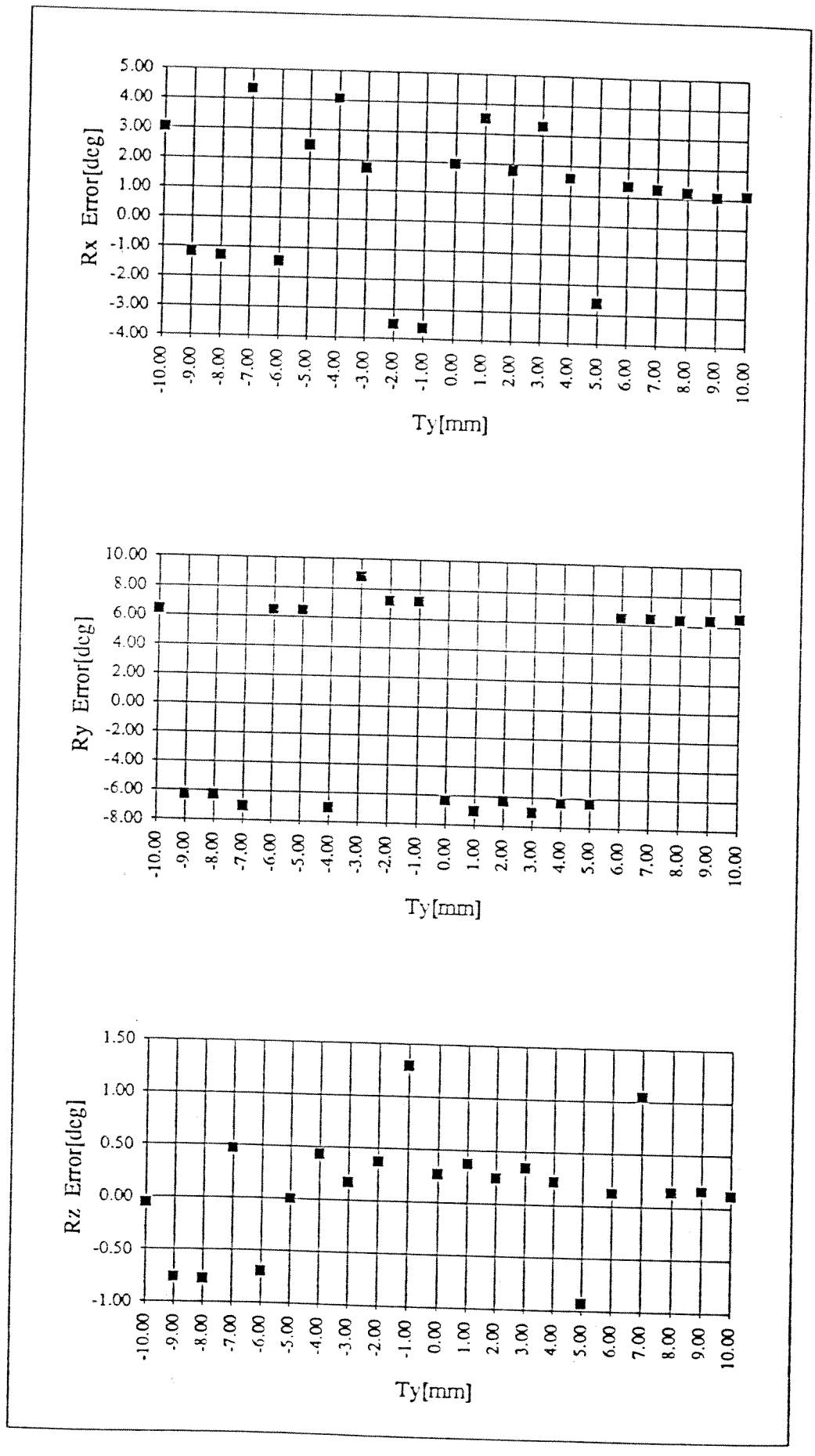


Figure 6.3: Reconstruction Errors For The Data Set in Table 6.2.

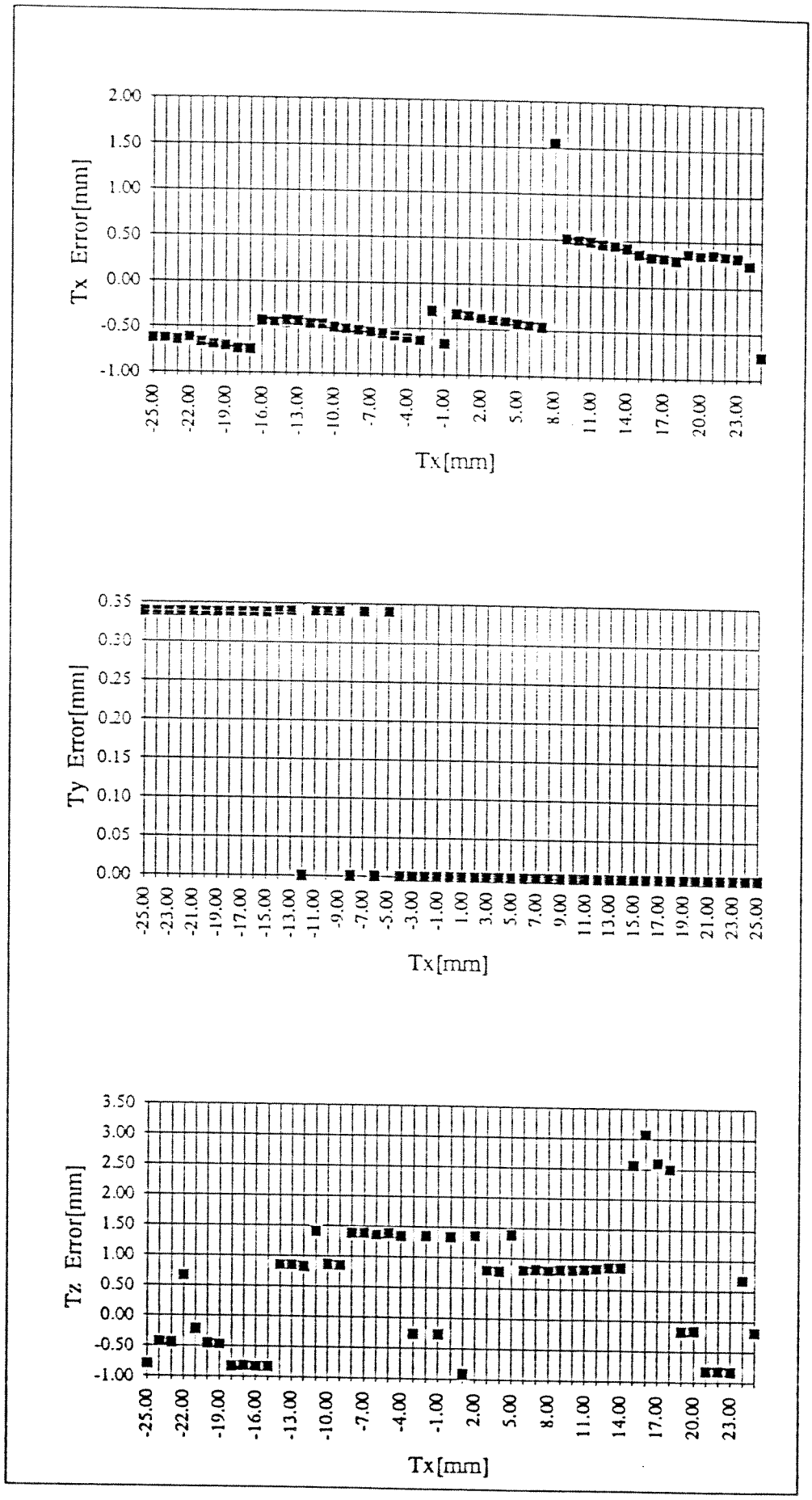


Figure 6.4: Reconstruction Errors For The Data Set in Table 6.1.

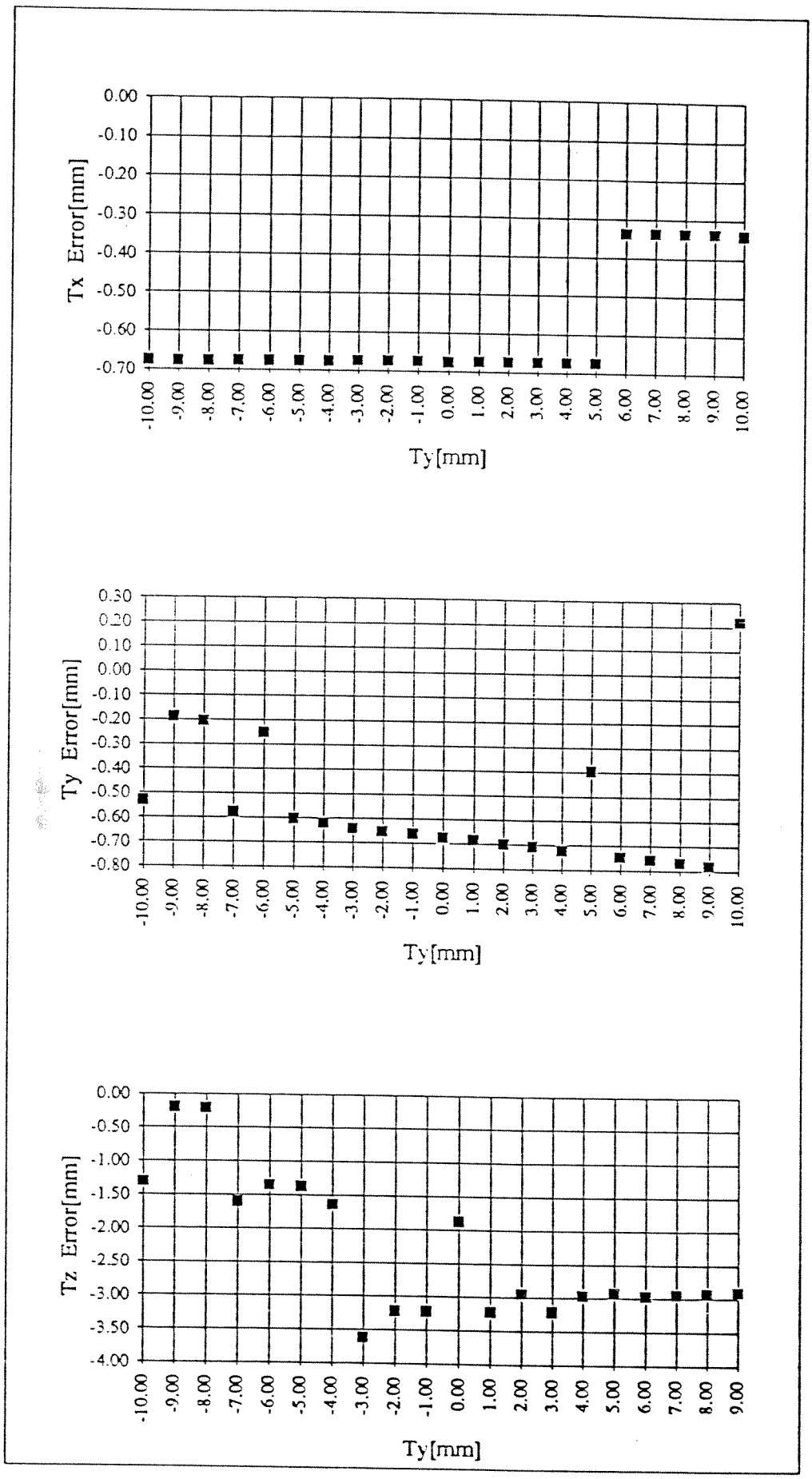


Figure 6.5: Reconstruction Errors For The Data Set in Table 6.2.

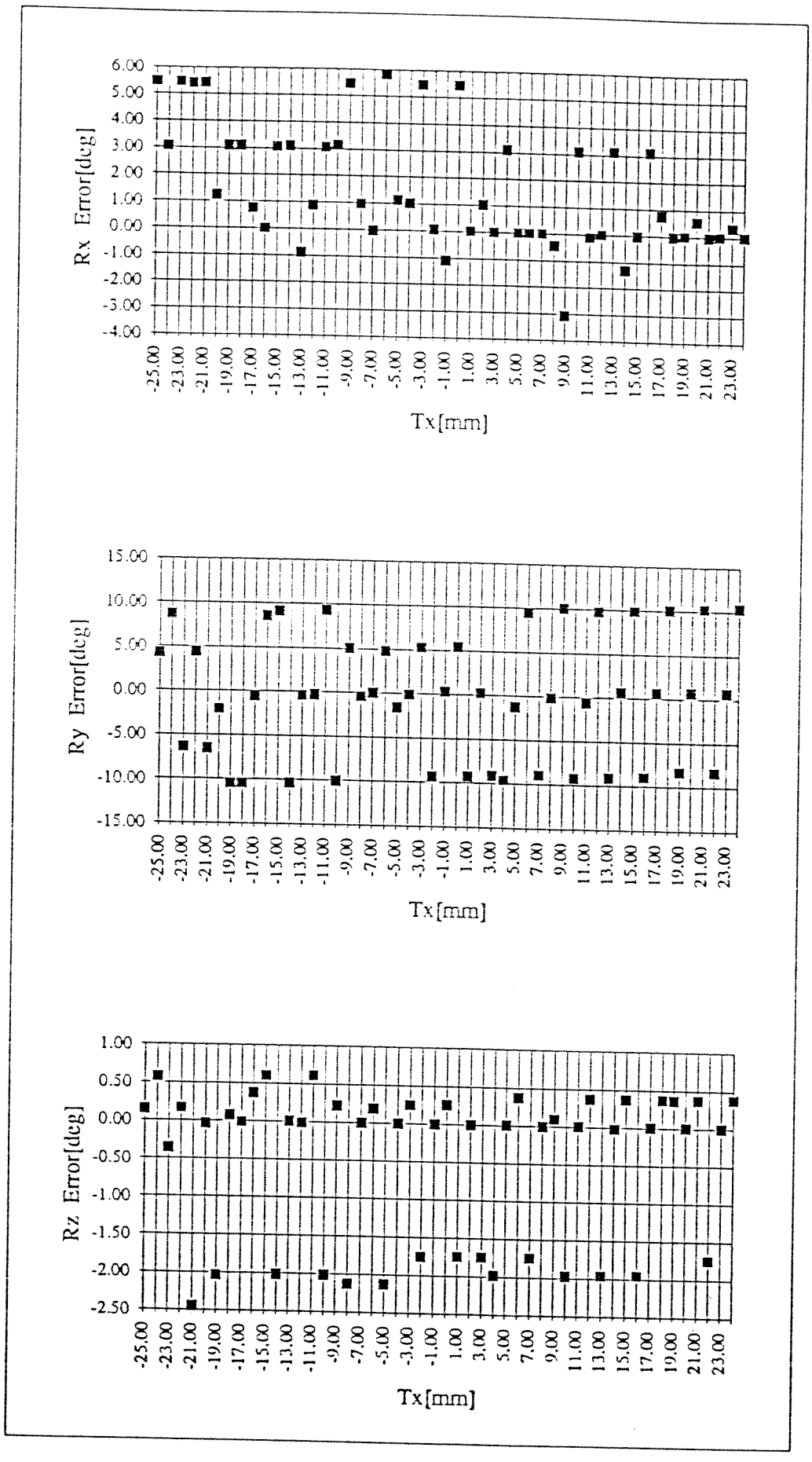


Figure 6.6: Reconstruction Errors For The Data Set in Table 6.3.

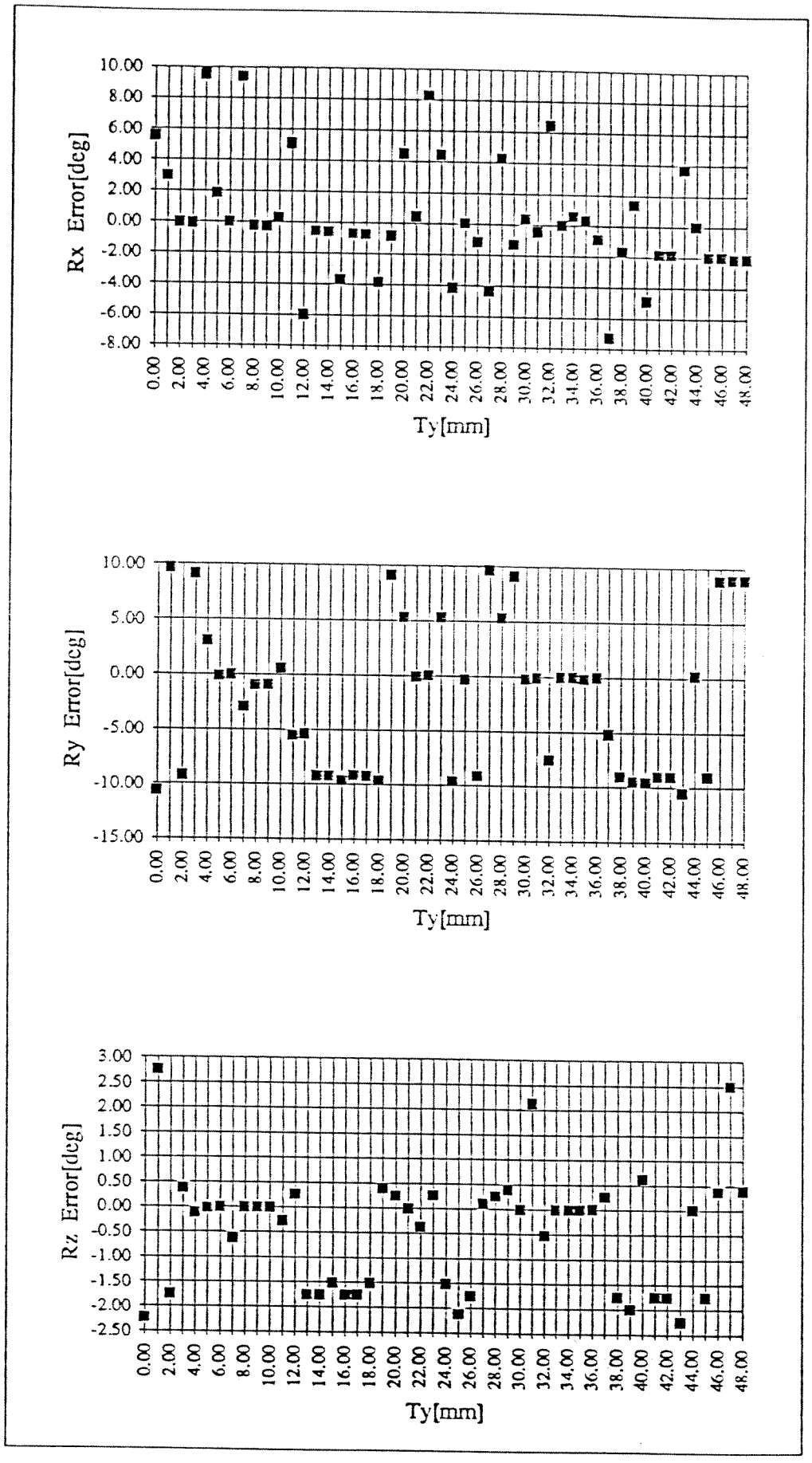


Figure 6.7: Reconstruction Errors For The Data Set in Table 6.4.

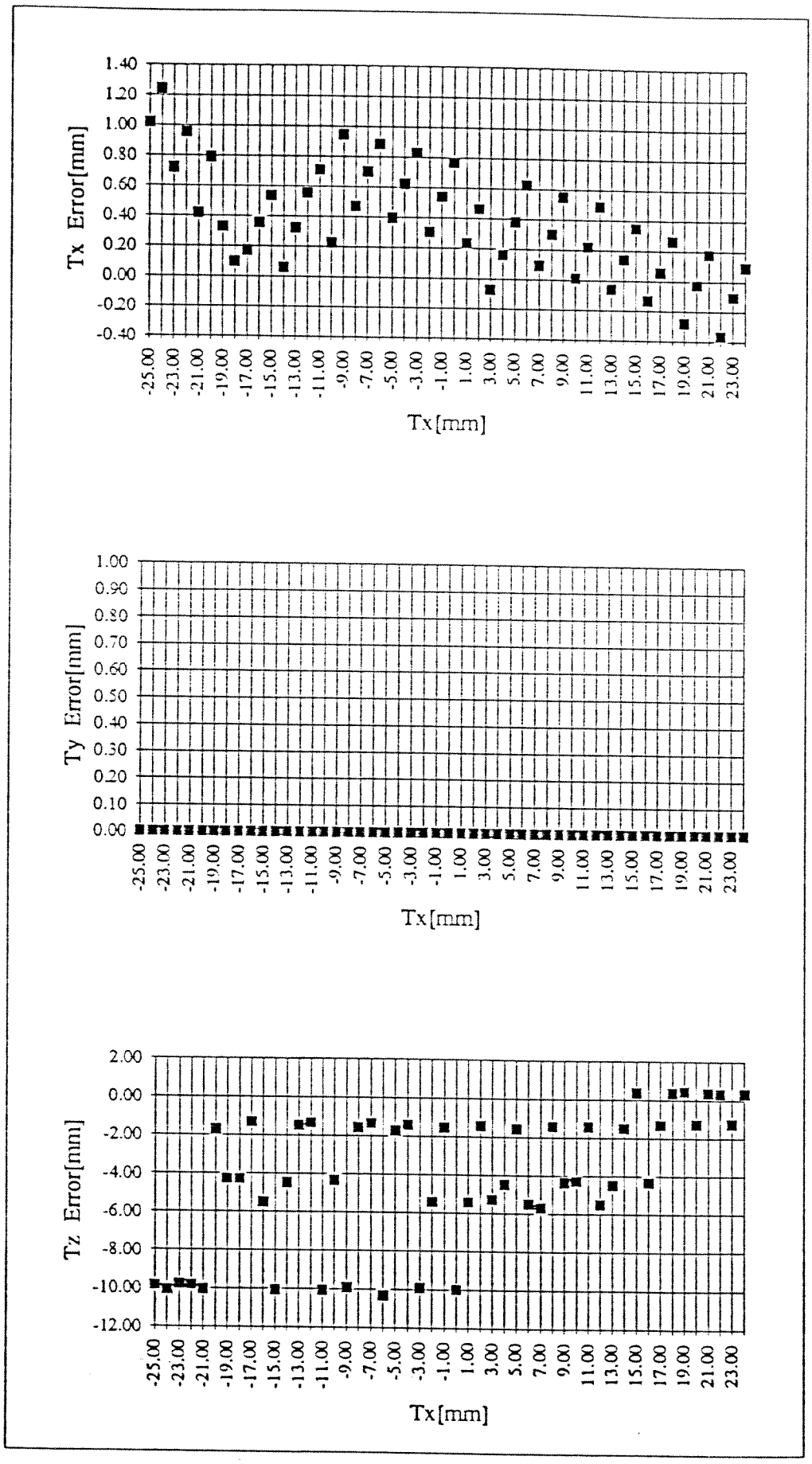


Figure 6.8: Reconstruction Errors For The Data Set in Table 6.3.

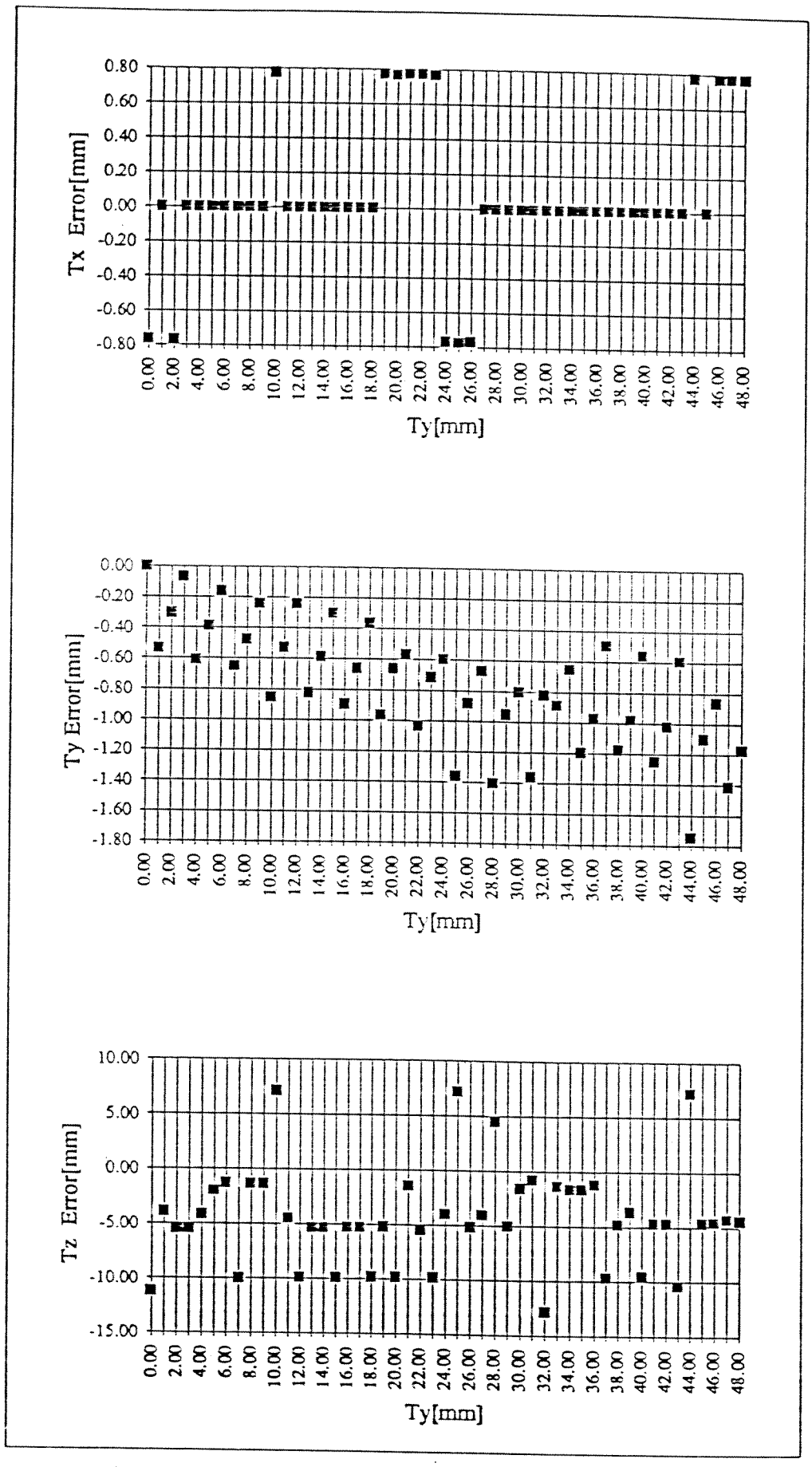


Figure 6.9: Reconstruction Errors For The Data Set in Table 6.4.

Table 6.5: Data Set For R_x when T_z=-590mm

	Start	Stop	Step
T_z[mm]	-590	-590	0
R_x[°]	0	80	5
R_y[°]	0	0	0
R_z[°]	45	45	0

Table 6.6: Data Set For R_y when T_z=-590mm

	Start	Stop	Step
T_z[mm]	-590	-590	0
R_x[°]	0	0	0
R_y[°]	-80	0	5
R_z[°]	45	45	0

time width of the jump decreases as expected.

The second test sets are given in Tables 6.5, 6.6, 6.7 and 6.8. These sets are used to test the accuracy of the reconstruction of R_x and R_y rotations separately. Since the reconstruction of the T_x and T_y depends only on the center of the internal spot and the distance d, they are not analyzed in these sets. Their behavior will be the same as their behavior for the first four sets. Figures 6.10, and 6.11 show the reconstruction errors for R_x, R_y and R_z when T_z has its minimum value. For the maximum T_z, rotation errors are shown in Figures 6.12 and 6.13. The translation T_z errors are shown for the maximum and minimum distances are shown in Figures 6.14, 6.15, 6.16 and 6.17.

Table 6.7: Data Set For R_x when T_z=-1305mm

	Start	Stop	Step
T_z[mm]	-1305	-1305	0
R_x[°]	0	80	5
R_y[°]	0	0	0
R_z[°]	45	45	0

Table 6.8: Data Set For R_y when T_z=-1305mm

	Start	Stop	Step
T_z[mm]	-1305	-1305	0
R_x[°]	0	0	0
R_y[°]	-75	0	5
R_z[°]	45	45	0

Table 6.9: Data Set For R_y when R_x=25°, T_z=-505mm

	Start	Stop	Step
T_z[mm]	-505	-505	0
R_x[°]	25	25	0
R_y[°]	-80	0	5
R_z[°]	45	45	0

From the above results it can be concluded that when R_x or R_y are between 0° and 10°, reconstruction errors are high. But when there is a rotation greater than 10° errors decrease considerably. This is expected, because for small rotations its difficult to detect the rotation of the target. Actually change in the image of the target is not enough to detect the small rotations. For the mid range rotations reconstruction errors gets even smaller. T_z errors are small for a target located closer to the camera, but higher for a target placed farther from the camera. Again this is the result of the worsening of information extracted from the image of the target as the target moves away.

Table 6.10: Data Set For R_x when R_y=25°, T_z=-505mm

	Start	Stop	Step
T_z[mm]	-505	-505	0
R_x[°]	0	80	5
R_y[°]	25	25	0
R_z[°]	45	45	0

Table 6.11: Data Set For R_y when R_x=25°, T_z=-1305mm

	Start	Stop	Step
T_z[mm]	-1305	-1305	0
R_x[°]	25	25	0
R_y[°]	-80	0	5
R_z[°]	45	45	0

Table 6.12: Data Set For R_x when R_y=25°, T_z=-1305mm

	Start	Stop	Step
T_z[mm]	-1305	-1305	0
R_x[°]	0	80	5
R_y[°]	25	25	0
R_z[°]	45	45	0

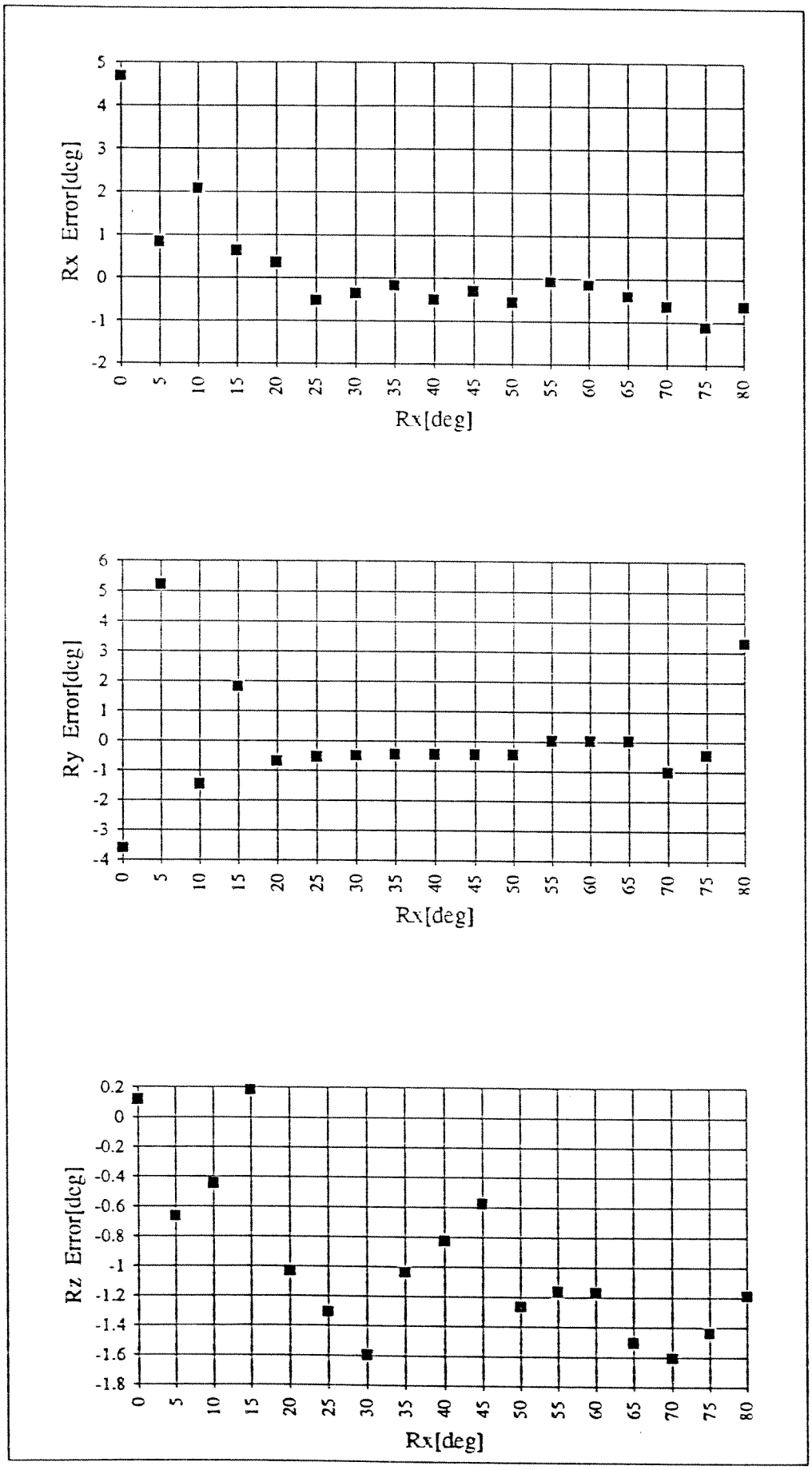


Figure 6.10: Reconstruction Errors For The Data Set in Table 6.5.

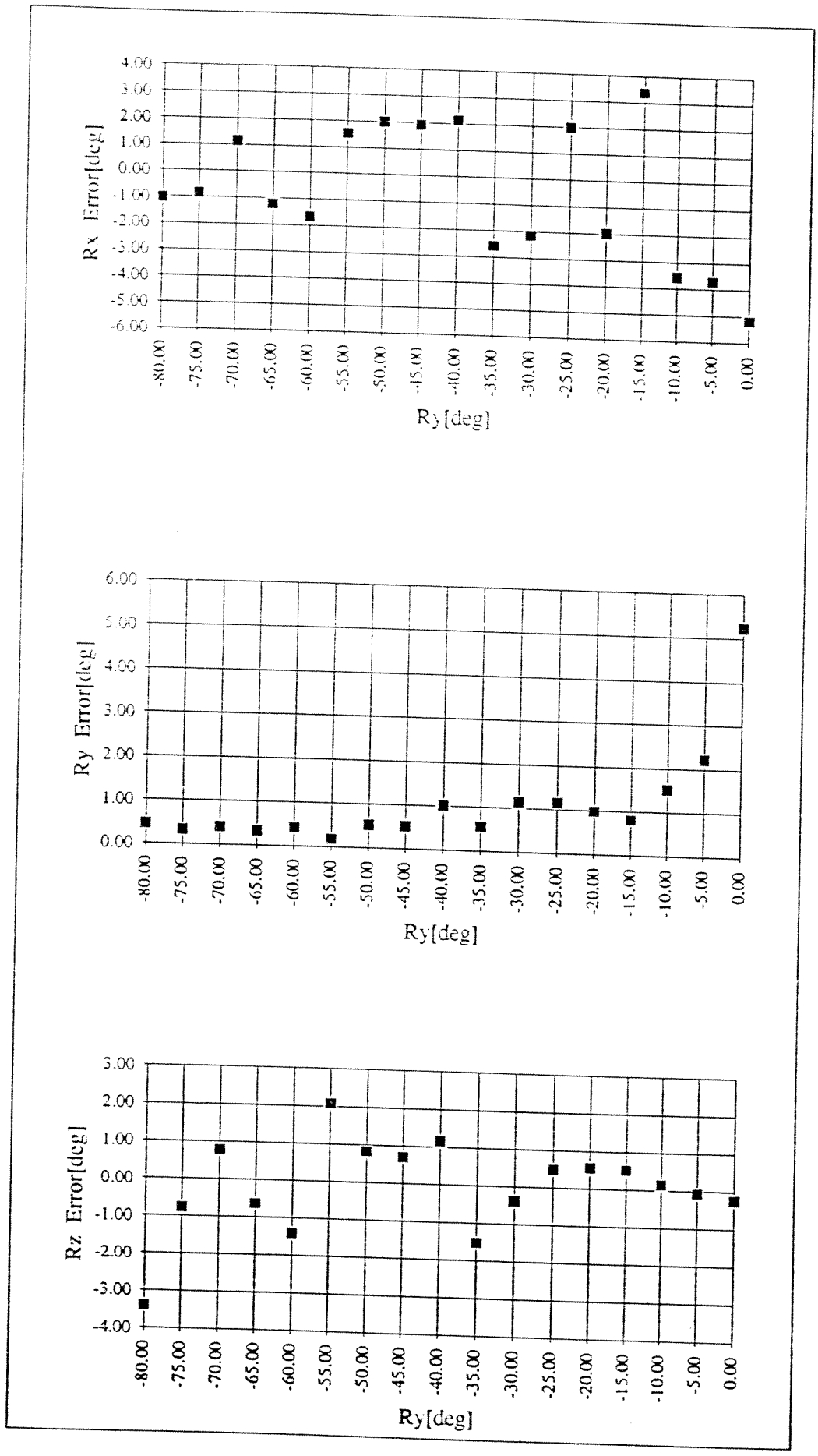


Figure 6.11: Reconstruction Errors For The Data Set in Table 6.6.

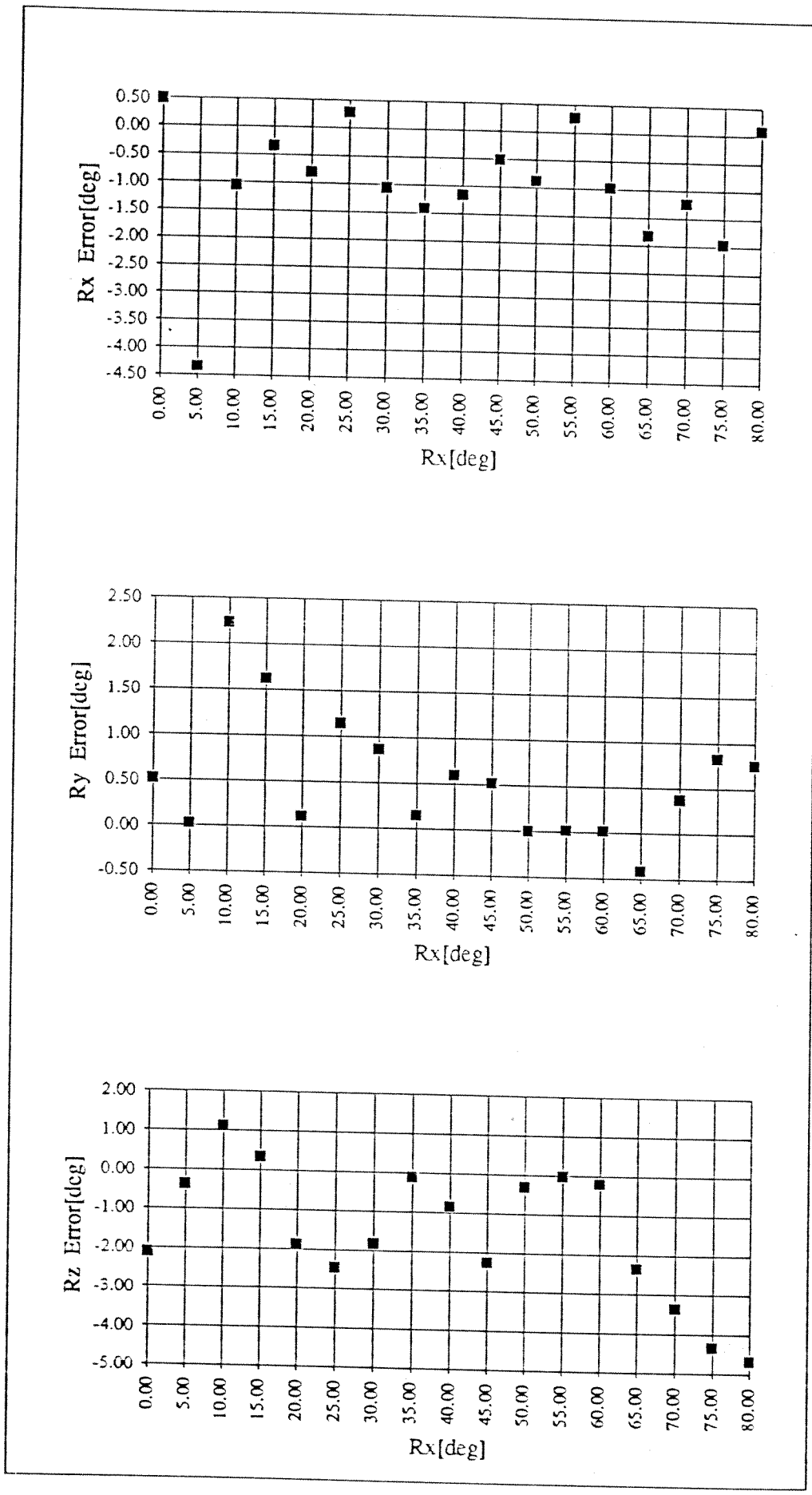


Figure 6.12: Reconstruction Errors For The Data Set in Table 6.7.

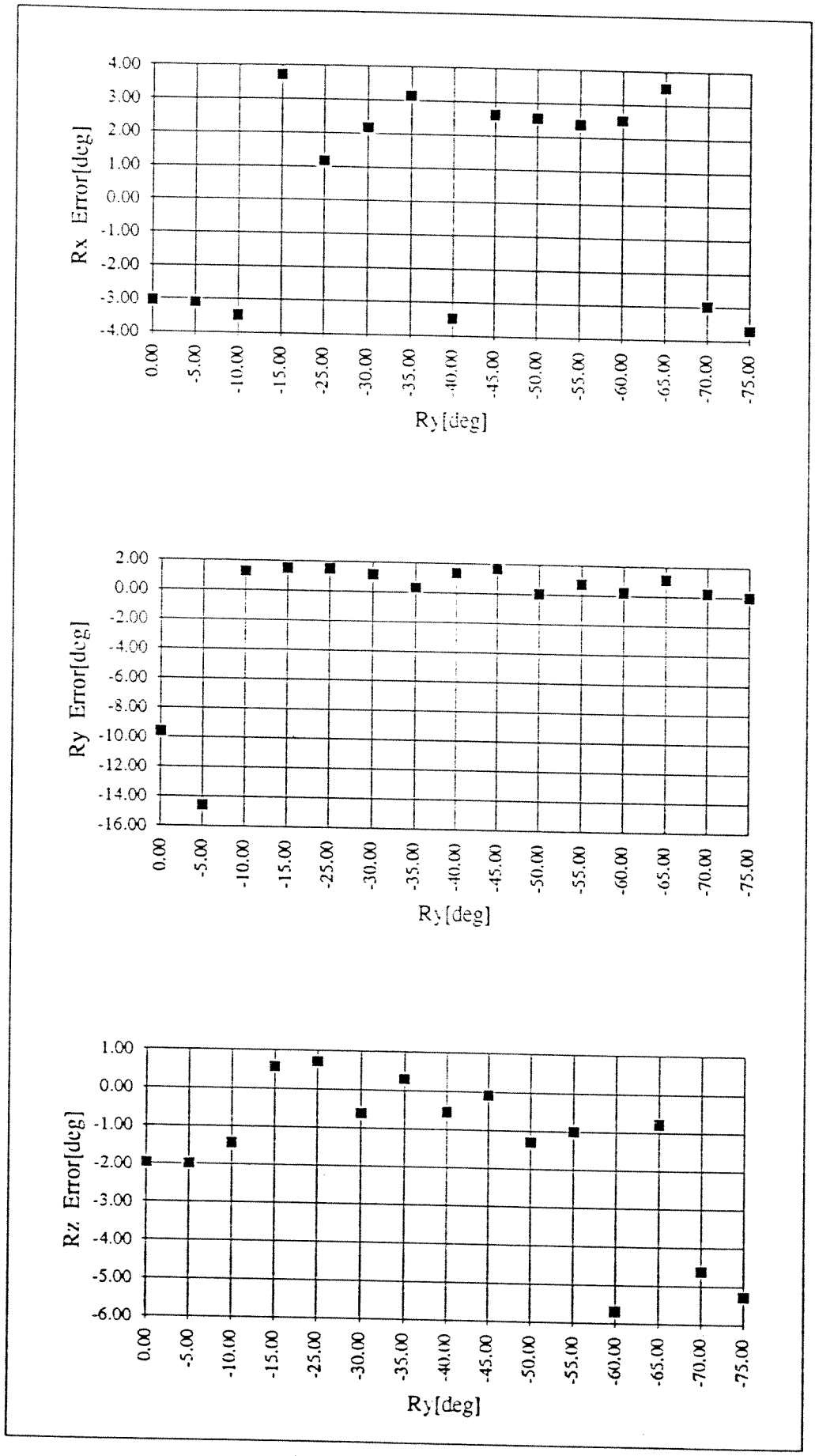


Figure 6.13: Reconstruction Errors For The Data Set in Table 6.8.

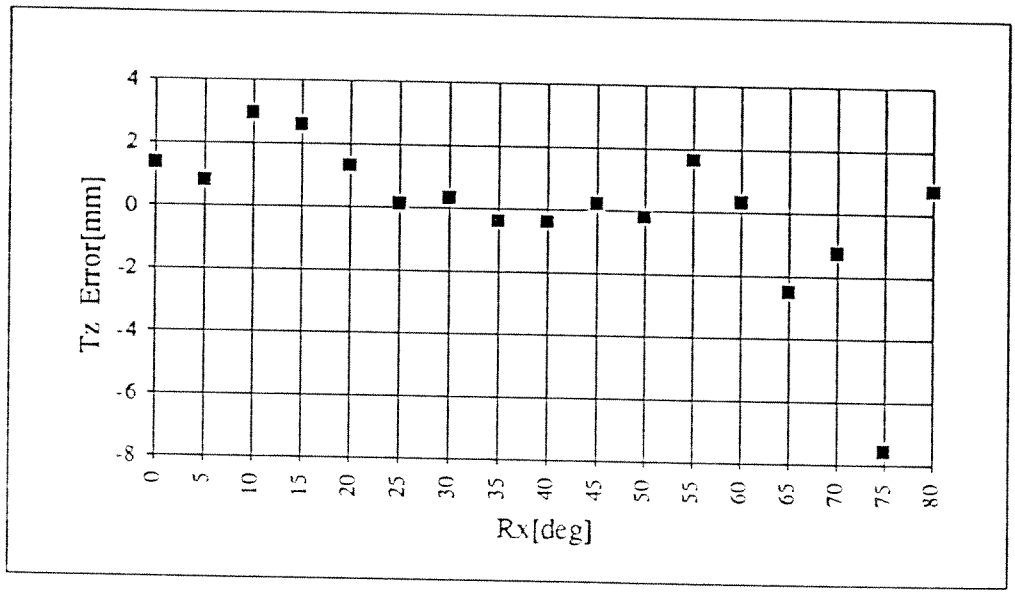


Figure 6.14: Reconstruction Errors For The Data Set in Table 6.5.

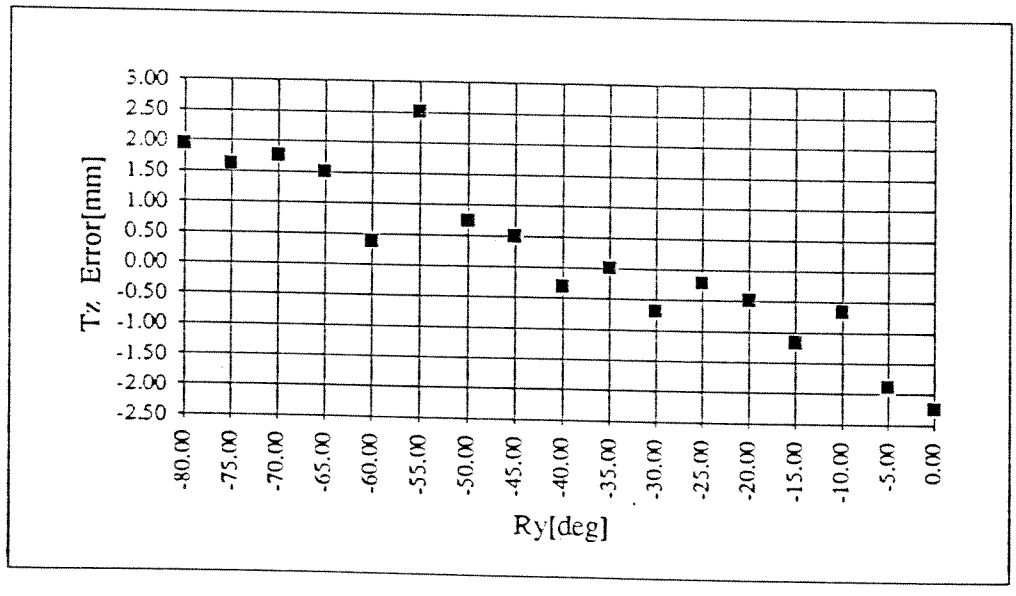


Figure 6.15: Reconstruction Errors For The Data Set in Table 6.6.

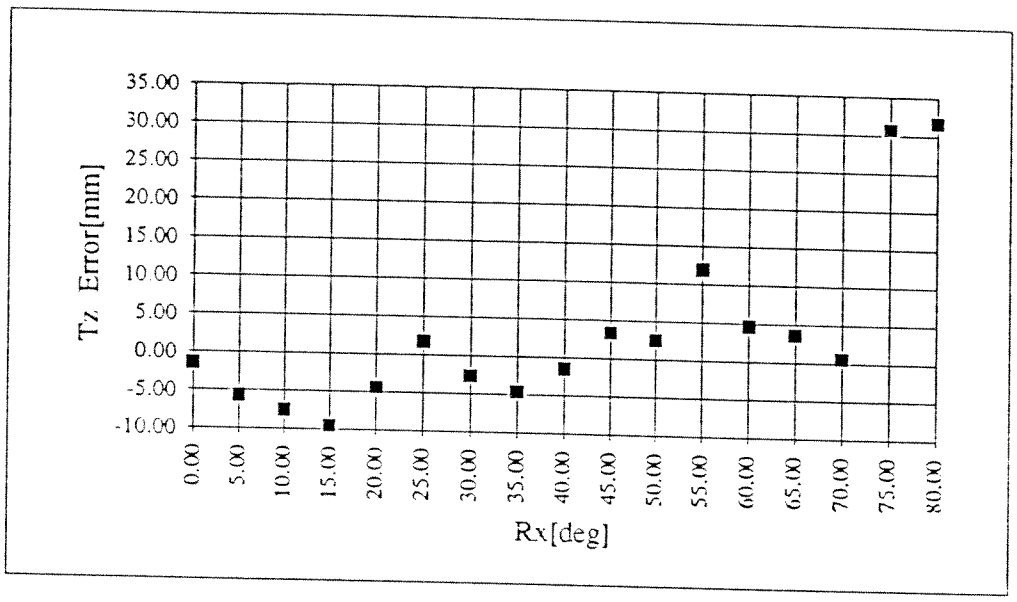


Figure 6.16: Reconstruction Errors For The Data Set in Table 6.7.

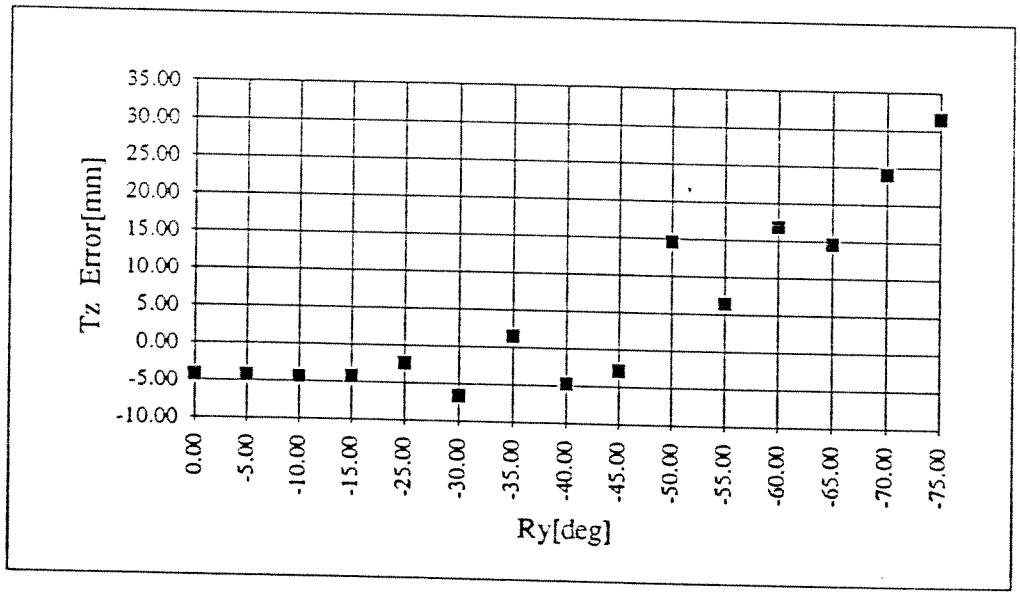


Figure 6.17: Reconstruction Errors For The Data Set in Table 6.8.

Data Sets in Tables 6.9, 6.10, 6.11 and 6.12 are used to investigate the second set of solutions. One of the rotations either R_x or R_y is kept constant at an angle greater than 10° and the other is changed. The results shows that rotations around all the axes are reconstructed successfully. The lower limit for the successful reconstruction of R_x and R_y decreases to 5°. From this result it can be stated that if one of the rotations either R_x or R_y is greater 10°, reconstruction errors decrease adequately for the rotation parameters. The results of these sets are given Figures 6.18, 6.22, 6.20, 6.24, 6.19, 6.23, 6.21 and 6.25.

Tables 6.13 and 6.14 show the data sets and reconstruction errors for combined rotations of R_x and R_y. In these sets R_x and R_y are rotated in equal amounts. From these sets, it is observed that there is a structural error in R_x which is a constant pre rotation of about 1°. This is due to the mounting of the angle bracket on which the target plane is mounted. Considering this structural error, reconstruction errors are found to be low. The results are in agreement with the previous results. The target distance from the camera is again found to be different from the measured value. Since the distance d is reconstructed by using the contour of the circle, any error in the extracted contour due to the setted aperture ratio and the lightning conditions will results in reconstruction errors for T_z.

6.4 Effect of Misfocus and Aperture Ratio

Aperture ratio of the lens controls the amount of light that fall on the sensor plane of the camera. Therefore it directly affects the appearance of the image of the target. For a wrong aperture ratio the image of the target will either be viewed smaller or larger than its actual size. This means that the features extracted from the image of the target will not be true. This will result in over or under reconstructed configurations parameters. However it is expected to have more change in the reconstruction of the translation parameters. This is a result of the reconstruction algorithm being used.

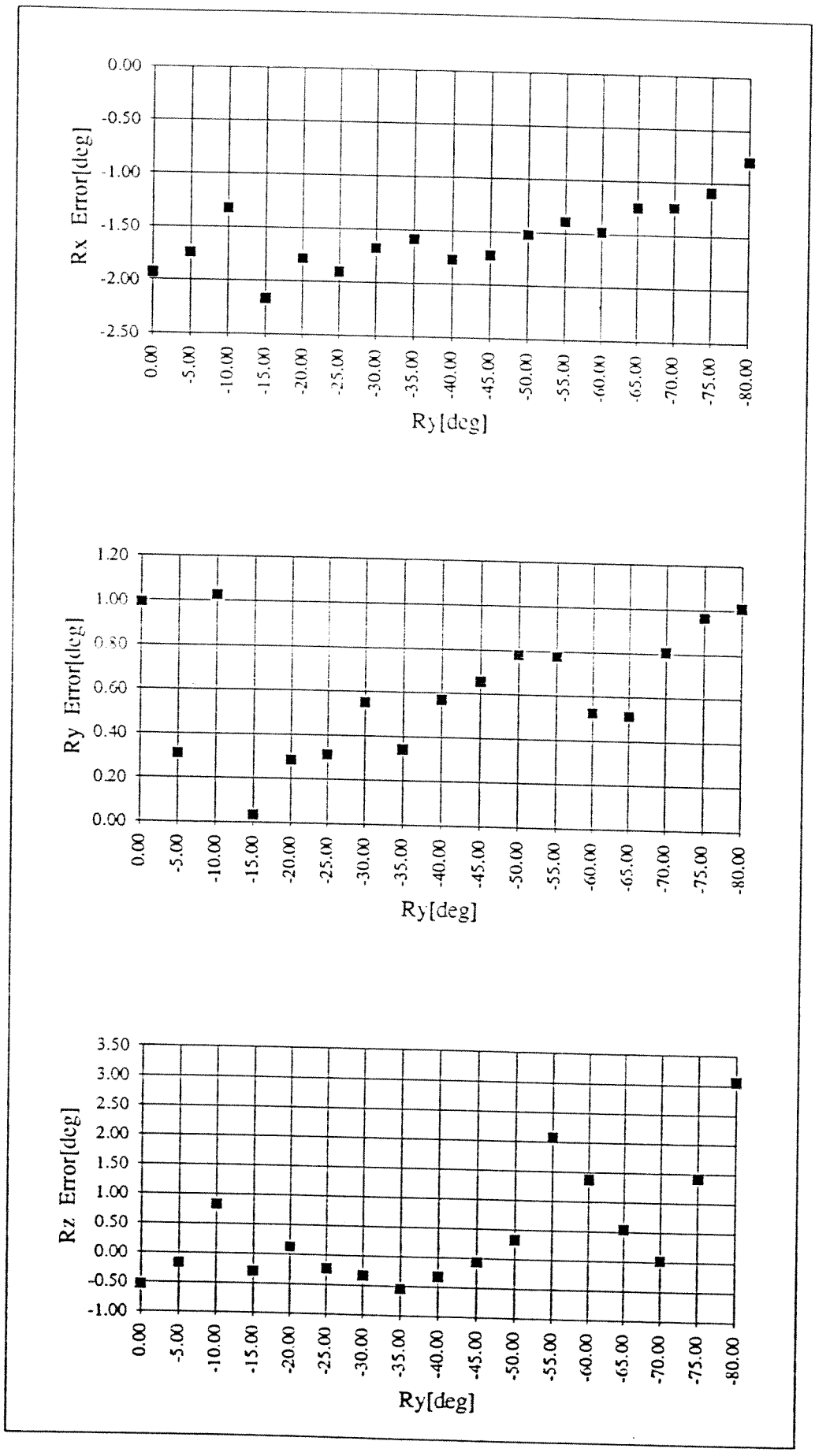


Figure 6.18: Reconstruction Errors For The Data Set in Table 6.9.

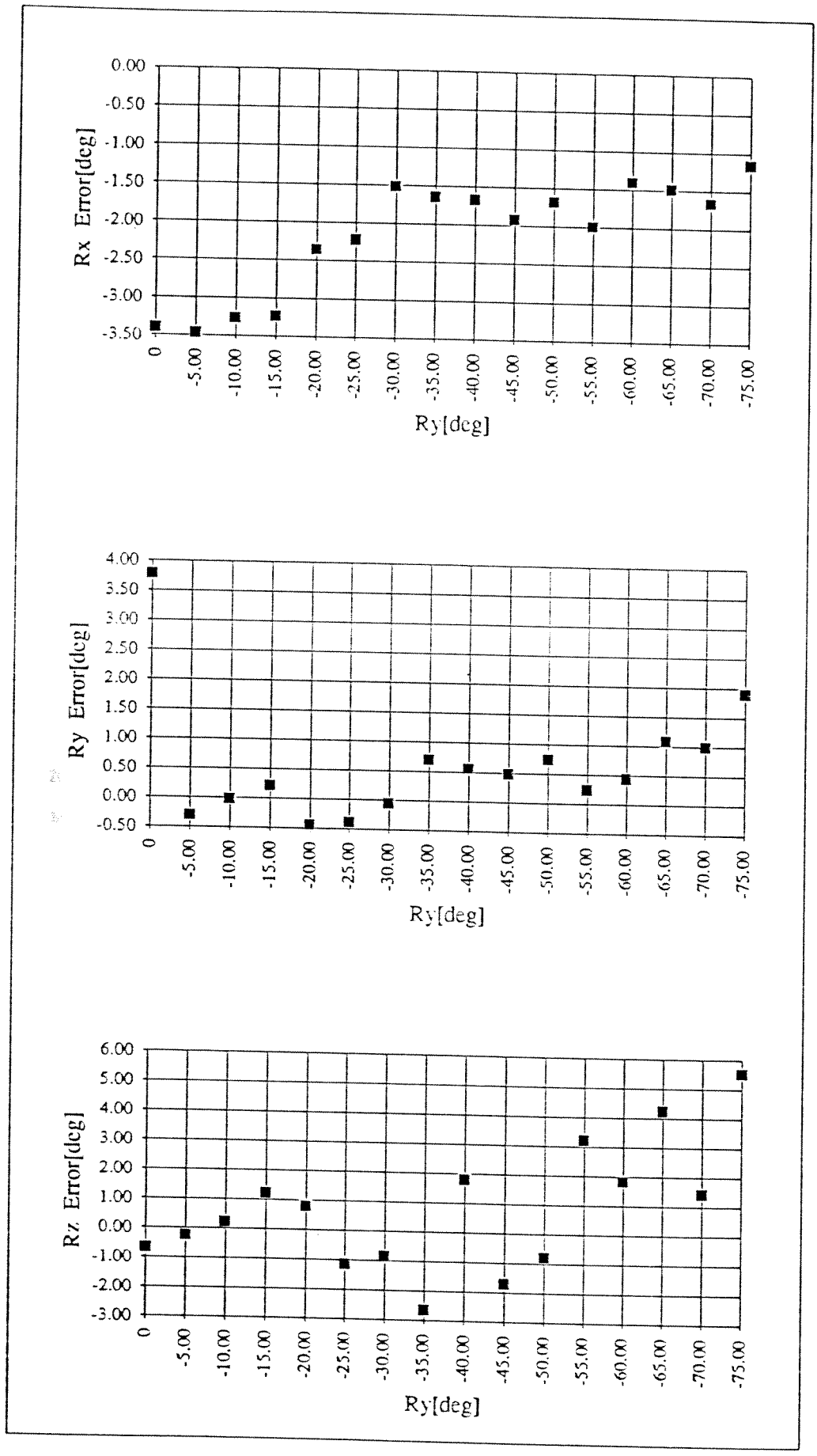


Figure 6.19: Reconstruction Errors For The Data Set in Table 6.11.

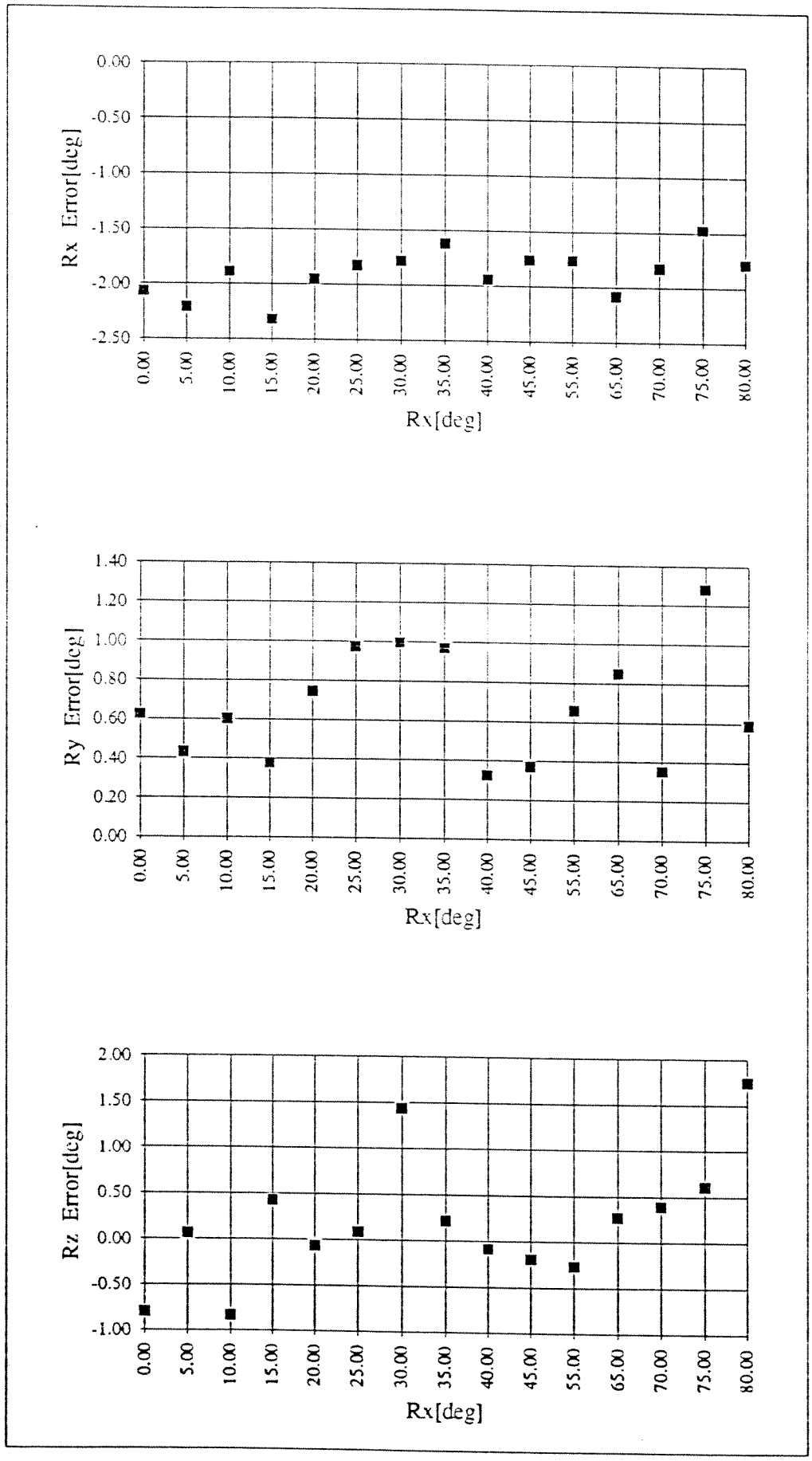


Figure 6.20: Reconstruction Errors For The Data Set in Table 6.10.

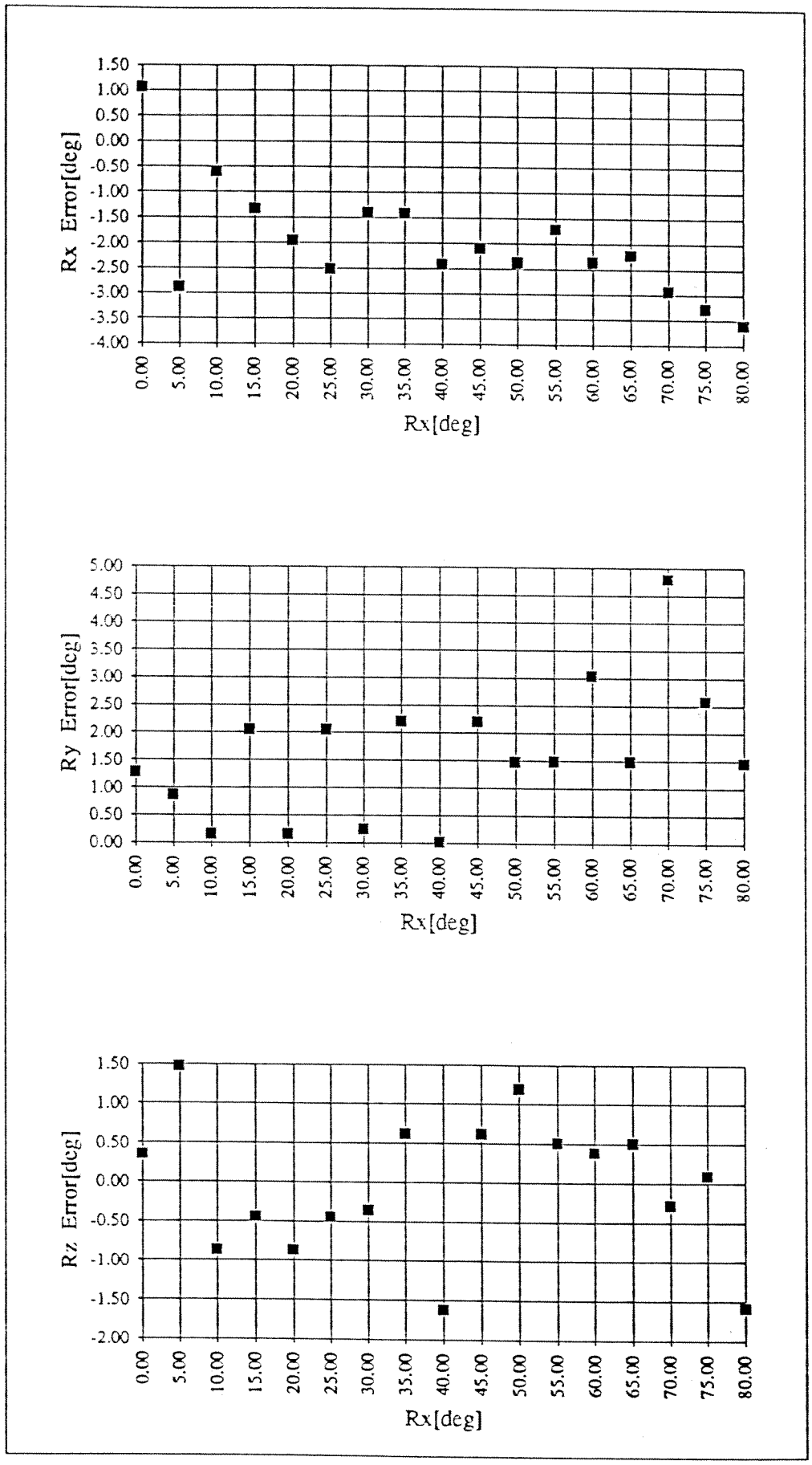


Figure 6.21: Reconstruction Errors For The Data Set in Table 6.12.

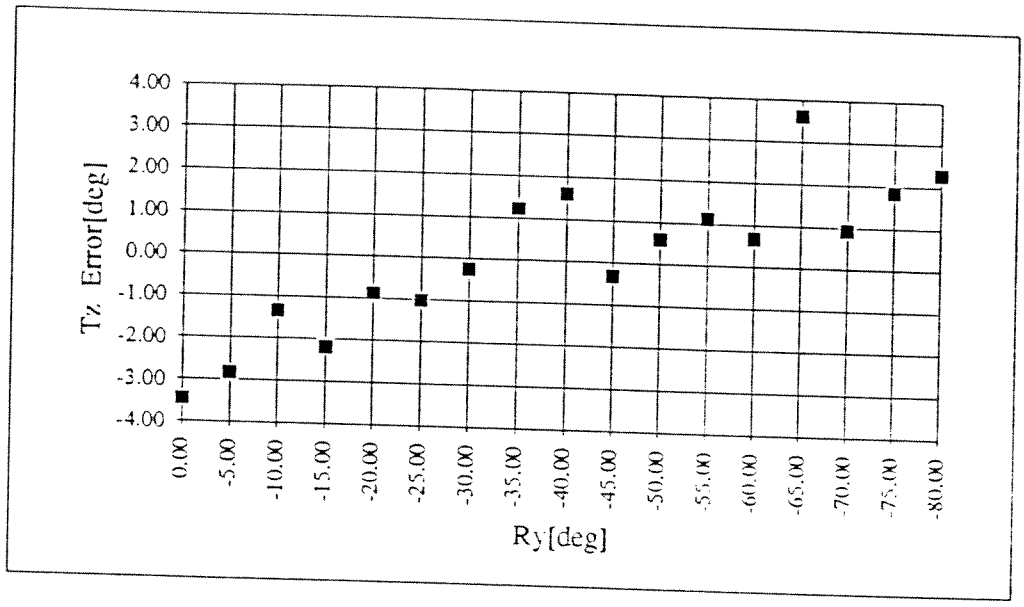


Figure 6.22: Reconstruction Errors For The Data Set in Table 6.9.

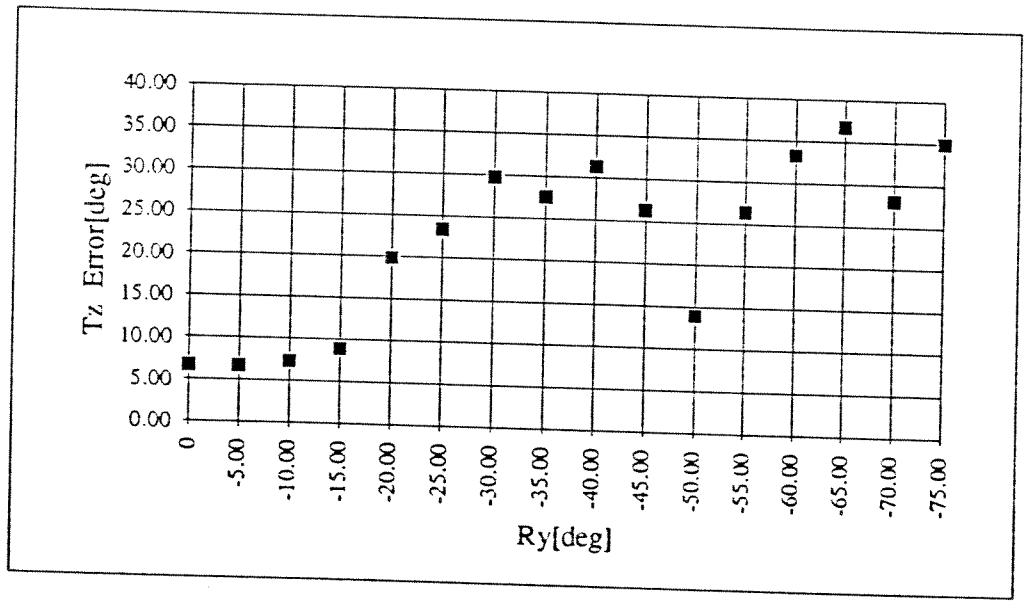


Figure 6.23: Reconstruction Errors For The Data Set in Table 6.11.

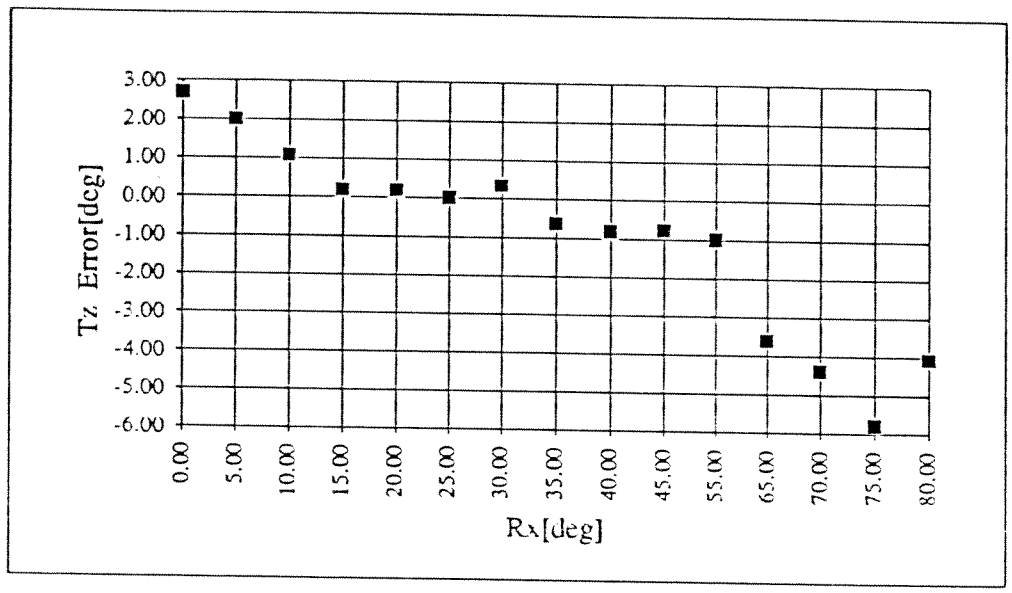


Figure 6.24: Reconstruction Errors For The Data Set in Table 6.10.

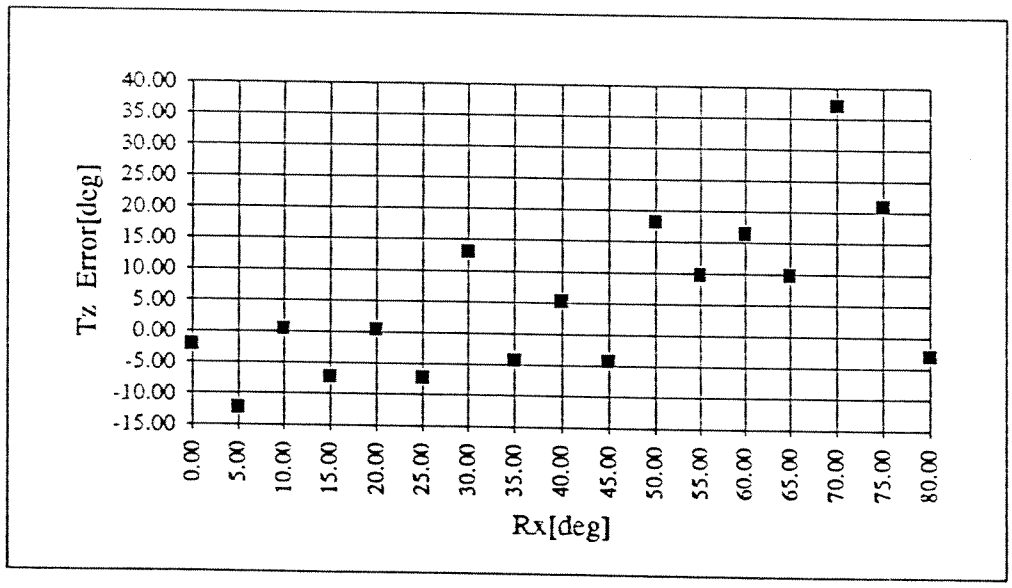


Figure 6.25: Reconstruction Errors For The Data Set in Table 6.12.

Table 6.13: Data Set and Reconstruction Errors For Combined Rotations of R_x and R_y When T_z=-505mm

Measured R_x	Measured R_y	Error in R_x	Error in R_y	Error in R_z	Error in T_z
0	0	6.48	-2.04	-0.05	-0.28
5	5	-1.34	-11.19	-2.22	0.22
10	10	-1.88	0.44	0.01	0.87
15	15	-1.90	1.53	1.41	-1.40
20	20	-2.49	0.40	0.70	-0.99
25	25	-1.82	-0.07	-0.37	2.18
30	30	-1.82	0.73	-0.62	0.29
35	35	-1.78	0.16	-0.60	2.02
40	40	-1.54	0.31	0.02	2.75
45	45	-1.95	0.79	-0.03	0.58
50	50	-1.76	0.97	1.09	-1.06
55	55	-1.36	1.26	1.02	1.35
60	60	-1.55	1.73	1.75	3.29
65	65	-1.16	2.08	0.44	5.88
Average Value		-1.13	-0.21	0.18	1.12
Standard Deviation		2.21	3.31	1.01	1.99

Table 6.14: Data Set and Reconstruction Errors For Combined Rotations of R_x and R_y When T_z=-1305mm

Measured R_x	Measured R_y	Error in R_x	Error in R_y	Error in R_z	Error in T_z
0	0	3.06	5.77	0.12	-4.36
5	5	1.96	4.61	-2.03	10.03
10	10	-0.97	1.18	-0.97	3.75
15	15	-2.02	-0.09	1.87	9.88
20	20	-1.42	0.76	1.60	5.01
25	25	-2.52	0.16	-0.87	0.40
30	30	-1.19	1.60	-0.41	-3.62
35	35	-0.58	0.88	1.57	9.04
40	40	-2.06	1.39	-1.65	-0.35
45	45	-1.70	1.12	1.29	16.59
50	50	-1.41	2.07	3.27	30.79
55	55	-1.24	2.47	1.56	41.45
60	60	-0.84	0.56	0.44	9.88
Average Value		-0.84	1.83	0.44	9.88
Standard Deviation		1.66	1.74	1.64	13.88

Table 6.15: Reconstruction Errors For Focus and Aperture Ratio

	$R_x Error$	R_yError	R_z Error	T_x Error	T_y Error	T_z Error
Correct Aperture	-0.71	0.86	0.02	-0.44	-0.44	-2.75
Wrong Aperture	-1.07	0.68	-0.90	-0.37	-0.39	-9.21
Focus	-0.91	0.31	-0.23	-.0.46	-0.45	-1.04
Misfocus	-0.21	-0.06	1.00	-0.64	0.63	14.45

Table 6.16: Data Set and Reconstruction Errors For Combined Rotations of R_x and R_y For Lens Having a Focal Lenght of 16mm When T_z=-1156mm

Measured R_x	Measured R_y	Error in R_x	Error in R_y	Error in R_z	Error in T_z
0	0	0.73	0.78	-0.01	5
5	5	-3.92	3.98	0.71	-9.04
10	10	-1.17	0.93	-1.86	-13.1
15	15	-0.35	0.88	-3.71	-2.17
20	20	-3.36	-0.89	2.96	-8.45
25	25	-2.06	0.08	2.37	-2.84
30	30	-2.01	-0.6	1.65	-1.05
35	35	-0.98	0.38	0.95	9.91
40	40	-0.83	0.74	3.48	18.1
45	45	-2.65	0.98	0.89	9.12
50	50	-1.86	2.4	4.62	22.84
55	55	-1.79	5.11	1.66	46.9
60	60	-3.16	7.38	9.07	47.45
Average Value		-1.8	1.7	1.75	9.44
Standard Deviation		1.29	2.41	3.1	19.76

Misfocus of the target will have similar effects like the aperture ratio has. Its effect on the T_z is much more than a wrong aperture ratio. This is because of the change of the image plane distance d_o when the focus is changed. Table 6.15 shows the reconstruction errors for a misfocus and wrong aperture ratio. Misfocus refers to an maximum blurred image and wrong aperture ratio refers to an image in which the features are apperant such that they can be extracted.

The parameters T_x and T_y are reconstructed by using the centroid of the central spot. If a smaller spot is used, shift in its centroid on the image will be smaller which will give better results. For the tested wrong aperture the image of the spot gets smaller, and this gives better results. For the misfocus case the opposite is true. Image of the spots get larger. This time error in the computation of the true center of the spots will increase which causes reconstruction errors.

6.5 Results of The Experiments For a Lens Having a Focal Length of 16mm.

As the focal length of a lens decreases, the angle of view will increase. This increase in the angle of view will cause the size of the target image to decrease, which will at the end effect the features extracted from the image. In a sense the amount of information carried by the image of the target will decrease. It will be difficult to identify the features of the image. The feature points will get closer to each other and this will make harder to detect the information obtained from the feature points.

Table 6.16 shows the reconstruction errors for combined rotations of Rx and Ry. As observed from the results, errors increase compared to the errors calculated using a lens having 25mm focal length. This problem can be solved by using a target having a larger radius.

6.6 Summary of the Experimental Results

The theoretical limits of the algorithm obtained in [1] are verified. These limits are as follows:

- $10° < |R_x| < 80°$
- $10° < |R_y| < 80°$
- $500mm < T_z < 1300mm$

Results of the experiments can be summarized as follows:

- T_x and T_y are reconstructed successfully in the whole range of rotations, because they are reconstructed using the central spot.

- Reconstruction errors for T_z are low in the above limits of rotations. But as the target moves away from the camera, error increases.
- R_x and R_y are detremined successfully in the above limits. Reconstruction errors inrease when the rotations get closer to the ultimate limits.
- R_z is reconstructed successfully in the whole range of rotations, because it is reconstructed using the location of the outer spot.

The following average reconstruction errors are obtained in the above limits:

- $E_{T_x}^{(\text{average})} \approx 0.5\text{mm}$
- $E_{T_y}^{(\text{average})} \approx 0.5\text{mm}$
- $E_{T_z}^{(\text{average})} \approx 1.5\text{mm}$
- $E_{R_x}^{(\text{average})} \approx 0.4^\circ$
- $E_{R_y}^{(\text{average})} \approx 0.4^\circ$
- $E_{R_z}^{(\text{average})} \approx 0.5^\circ$

CHAPTER VII

SUMMARY AND CONCLUSIONS

In this chapter, summary of the study and the conclusions will be presented. Finally guidelines for future research will be given.

7.1 Summary

The algorithm used reconstructs 6 configuration variables by using a secondary target which is rigidly attached to the rigid body of interest. The algorithm was developed in [1] as an alternative approach to the stereo vision. Its theoretical limits are available as the outcome of the study [1]. In this study the algorithm has been implemented and its limits have been investigated by using a digital imaging system. The system consists of a lens, a single camera and a frame grabber hosted by a PC. The image of the target has been obtained in 256 gray levels and it has been binarized using a proper threshold. The features required for the algorithm are extracted from the binary image by using image analysis techniques.

The secondary target being used is a planar circle with two internal spots. The geometry of the target can be defined as follows:

1. A main circle of radius R, located at (0,0) on the target plane
2. A central spot of radius r_0, located at (0,0) on the target plane
3. A outer spot of radius r_0, located at (r_f,0) on the target plane

The target has been printed with a laser printer on a white paper. The background of the target has been sufficiently darker than the target such that there exists high contrast ratio in the image. The selected illumination of the target has also guaranteed the high contrast ratio. The high contrast ratio has simplified the binarization process, such that a constant threshold value has been used for all gray level images.

The features extracted from the binary image of this target are:

- contour of the main circle
- centroids of the spots

The binary image of the target has been obtained online by using the input look up table of the frame grabber. A computer program has been developed in C language to extract the features from the binary image and to compute the configuration variables using the algorithm. The target has been recognized and the region of interest has been selected by a human operator. All the subsequent operations like contour following, finding centroids etc. have been done automatically.

The configuration parameters have been compared with the measured values using the set-up. Several different configurations have been tested and their results have been evaluated.

7.2 Conclusions

The algorithm has been tested and it has been proved that the method works properly within the following limits:

$$10^\circ < |\alpha| < 80^\circ$$

$$10^\circ < |\beta| < 80^\circ$$

These limits correspond to highly rotated and unrotated configurations. When the rotations are close to 80°, a very thin image ellipse has been obtained. In the opposite case when the target rotations are close to 10° , the R_x and R_y parameters can not be successfully reconstructed. However satisfactory results are obtained in the reconstruction of the other parameters. The target has been tested in the available range of the experimental set-up which is:

$$500mm < d < 1300mm.$$

In these regions, the algorithm provides satisfactory results and can be used to detect the configuration parameters of a rigid body in an automated environment.

The speed of the solution for an IBM compatible 486DX66 PC is about 18Hz for the algorithm and 1.3Hz for the complete solution. The complete solution includes the extraction of the features from the binary image of the target.

7.3 Guidelines for Future Works

In a controlled environment the study is complete. All the configuration parameters are determined. However the implementation of the algorithm for an uncontrolled environment such as a factory will require the following practical points to study:

- image processing and enhancement
- selection of the target material
- target recognition
- use of sub pixel techniques
- moving targets

- improvement of the algorithm for better results in all configuration ranges
- optimization for faster operation

The image processing and enhancement operations can be used to obtain a digital contour data which is close to its analytical representation. This will increase the accuracy of the results.

The selection of the material is important to obtain a high contrast ratio in the image of the target. Choosing the threshold value for the binarization will be easier for an image having a high contrast ratio.

Sub pixel techniques can be used to obtain a better digital representation of the analytical contour of the ellipse, which will again increase the accuracy of the results.

The algorithm can be used for motion tracking. Using past motions it is possible to estimate future motions.

Two or more circular secondary targets can be used to eliminate the high reconstruction errors for certain target configuration ranges. As an example, two targets can be mounted on two surfaces having 45° in between.

To use the algorithm in a real machine vision system, the solution speed can be optimized for higher speeds. However for higher performance, hardware implementation is necessary.

REFERENCES

[1] Kılınç, İ. A., 1994. "Determination of The Position and Orientation of Rigid Bodies by Using Single Camera Images", M.S. Thesis in Mechanical Engineering, Middle East Technical University, Ankara, 1994.

[2] Lapidus, S. N., 1985. "Understanding How Images are Digitized", in Machine Vision Capabilities For Industry, Zuech N., ed., Machine Vision Association of SME Publications Development Department, 1986.

[3] Chu, C. Y., 1986. "Resolution and Accuracy in Gray level Machine Vision", in Machine Vision Capabilities For Industry, Zuech N., ed., Machine Vision Association of SME Publications Development Department, 1986.

[4] Willson, R.G., 1993. "What is The Center of The Image", School of Computer Science Carnegie Mellon University, CMU-CA-93-122, April 1993.

[5] Harding, K. G., 1985. "Optical Considerations for Machine Vision", in Machine Vision Capabilities For Industry, Zuech N., ed., Machine Vision Association of SME Publications Development Department, 1986.

[6] Data Translation, 1993. DT3851 Series User Manual, 1993.

[7] Schalkoff, R. J., 1989. Digital Image Processing and Computer Vision, John Wiley and Sons, Inc., 1989.

[8] Tsai, R.Y., 1986. "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, pp. 965-980, 1992 .

[9] Stein, G.P., 1993. "Internal Camera Calibraton using Rotation and Geometric Shapes", M.S. Thesis in Electrical Engineering and Computer Science , Massachusetts Institute of Technology, Massachusetts, 1993.

[10] Kite, D.H., Magee, M., 1989. "Determining the 3D Position and Orientation of a Robot Camera Using 2D Monocular Vision", Pattern Recognition, Vol. 23, No. 8, pp. 819-831, 1990.

[11] Sydow, P.D., Cooper, E.G., 1992. "Development of a Machine Vision System for Automated Structural Assembly", Internal Report, No. L-16995, NASA Langley Research Center, Hampton, VA 23665-5225, Mar. 1992.

[12] Weng, J., 1992. "Camera Calibration with distortion Models and Accuracy Evalution", IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 14, No. 10, pp. 965-980, Oct. 1992.

[13] Olgac, N., Gan, Z., Platin, B.E., 1989. "3D Reconstruction of Object Configurations by Hybrid Projection Analysis Using a Single Camera Image", Proceedings of the 1st National Conference on Applied Mechanisms and Robotics, Vol. 1, Paper No. 89AMR-4C-6, Nov. 5-8, Cincinnati, OH, USA.

[14] Platin, B.E., Gan, Z., Olgac, N., 1990. "3D Object Configuration Sensor Utilizing Single Camera Images", ASME Paper No. 90-WA/DSC-22, presented at the Winter Annual Meeting, Nov. 25-30, Dallas, TX, USA.

[15] Olgac, N., Craig, P.D., Platin, B.E., 1991. "Advancements in Utilizing Monocular Object Configuration Sensor, Part II", Advances in Instrumentation, DSC-Vol.30, ASME Book No. H00691, pp. 21-24, also presented at the Winter Annual Meeting, Dec. 1-6, Atlanta, GA, USA.

[16] Jain, A. K., 1989. Fundamentals of Digital Image Processing, Prentice-Hall, Inc., 1989.

APPENDIX A

RESULTS OF EXPERIMENTS

Output of the computer program gives directly calculated configuration parameters. The name of the image file and configuration parameters are recorded in a text file. The followings are the outputs of the program in tabular form.

Table A.1: Results of Experiments for The Data Set in Table 6.1.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
tyn25.img	2.75	-6.92	45.35	-24.37	-0.34	-589.20
tyn24.img	3.52	-5.09	45.51	-23.37	-0.34	-589.56
tyn23.img	3.50	-5.18	45.51	-22.36	-0.34	-589.56
tyn22.img	0.00	-4.30	45.75	-21.38	-0.34	-590.66
tyn21.img	1.93	-4.65	45.63	-20.33	-0.34	-589.78
tyn20.img	3.60	9.29	45.99	-19.31	-0.34	-589.53
tyn19.img	3.61	9.17	45.99	-18.29	-0.34	-589.53
tyn18.img	2.79	11.01	45.81	-17.26	-0.34	-589.17
tyn17.img	2.80	10.88	46.76	-16.25	-0.34	-589.17
tyn16.img	2.80	-7.64	45.35	-15.57	-0.34	-589.17
tyn15.img	2.81	-7.73	45.35	-14.56	-0.34	-589.17
tyn14.img	5.04	-3.91	45.67	-13.58	-0.34	-590.85
tyn13.img	5.03	-3.99	45.67	-12.56	-0.34	-590.85
tyn12.img	-5.07	-4.08	45.23	-11.54	0.00	-590.83
tyn11.img	3.58	4.77	46.03	-10.54	-0.34	-591.40
tyn10.img	5.05	-4.25	45.67	-9.51	-0.34	-590.86
tyn9.img	5.07	-4.33	45.67	-8.49	-0.34	-590.85
tyn8.img	-3.61	4.41	44.89	-7.48	0.00	-591.37
tyn7.img	3.61	4.25	46.03	-6.46	-0.34	-591.39
tyn6.img	-3.64	4.21	44.89	-5.44	0.00	-591.35

Table A.1 (continued)

tyn5.img	3.61	4.10	46.03	-4.42	-0.34	-591.38
tyn4.img	-3.65	3.99	44.89	-3.40	0.00	-591.35
tyn3.img	-6.44	2.32	45.05	-2.37	0.00	-589.73
tyn2.img	-3.62	-3.47	44.20	-1.70	0.00	-591.35
tyn1.img	-6.45	2.12	45.05	-0.34	0.00	-589.72
ty0.img	-3.64	-3.68	45.12	0.34	0.00	-591.34
ty1.img	-6.29	-6.44	44.43	1.35	0.00	-589.09
ty2.img	-3.62	-3.88	44.20	2.38	0.00	-591.35
ty3.img	-5.13	4.82	44.77	3.40	0.00	-590.79
ty4.img	5.13	4.72	45.23	4.41	0.00	-590.78
ty5.img	3.57	-4.21	44.89	5.44	0.00	-591.37
ty6.img	5.12	4.52	45.23	6.45	0.00	-590.80
ty7.img	-5.11	4.43	44.77	7.47	0.00	-590.81
ty6.img	5.12	4.52	45.23	6.45	0.00	-590.80
ty8.img	-5.10	4.34	44.77	8.49	0.00	-590.81
ty9.img	-5.10	4.23	45.67	9.51	0.00	-590.82
ty10.img	-5.08	4.15	45.67	10.52	0.00	-590.83
ty11.img	-5.07	4.06	45.67	11.54	0.00	-590.84
ty12.img	-5.06	3.97	45.67	12.56	0.00	-590.85
ty13.img	-5.04	3.88	45.67	13.58	0.00	-590.86
ty14.img	3.51	2.34	45.12	14.64	0.00	-592.57
ty15.img	0.02	-0.08	45.91	15.68	0.00	-593.06
ty16.img	3.47	2.18	46.03	16.69	0.00	-592.59
ty17.img	1.99	4.87	45.87	17.71	0.00	-592.49
ty18.img	0.01	6.95	45.58	18.64	0.00	-589.83
ty19.img	0.01	6.84	45.59	19.66	0.00	-589.84
ty20.img	2.80	7.23	45.82	20.65	0.00	-589.18
ty21.img	2.80	7.15	45.82	21.66	0.00	-589.18
ty22.img	2.79	7.08	45.82	22.68	0.00	-589.16
ty23.img	0.01	4.10	45.75	23.76	0.00	-590.67
ty25.img	1.97	4.22	45.87	25.76	0.00	-589.81

Table A.2: Results of Experiments for The Data Set in Table 6.2.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
ty10.img	-3.01	-6.40	45.06	0.68	-9.47	-588.69
ty9.img	1.19	6.32	45.77	0.68	-8.81	-589.79
ty8.img	1.31	6.30	45.78	0.68	-7.79	-589.79
ty7.img	-4.34	7.05	44.53	0.68	-6.42	-588.40
ty6.img	1.46	-6.44	45.70	0.68	-5.75	-588.65
ty5.img	-2.49	-6.45	45.00	0.68	-4.40	-588.63
ty4.img	-4.04	7.06	44.57	0.68	-3.38	-588.36
ty3.img	-1.74	-8.79	44.83	0.67	-2.36	-586.38
ty2.img	3.51	-7.20	44.63	0.67	-1.35	-586.77
ty1.img	3.63	-7.19	43.71	0.67	-0.34	-586.78
ty0.img	-1.95	6.36	44.73	0.68	0.68	-588.12
tyn1.img	-3.52	7.04	44.63	0.67	1.69	-586.78
tyn2.img	-1.76	6.30	44.76	0.67	2.70	-587.05
tyn3.img	-3.30	7.04	44.66	0.67	3.71	-586.79
tyn4.img	-1.57	6.36	44.78	0.67	4.72	-587.03
tyn5.img	2.62	6.33	45.93	0.67	5.40	-587.07
tyn6.img	-1.37	-6.44	44.88	0.34	6.74	-587.04
tyn7.img	-1.28	-6.42	43.95	0.34	7.76	-587.06
tyn8.img	-1.20	-6.39	44.86	0.34	8.77	-587.07
tyn9.img	-1.07	-6.38	44.84	0.34	9.78	-587.10
tyn10.img	-1.10	-6.50	44.88	0.34	9.78	-587.1

Table A.3: Results of Experiments for The Data Set in Table 6.3.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
tx0.img	0.00	8.21	44.62	25.45	0.00	-1310.33
tx1.img	0.00	-10.40	44.62	23.91	0.00	-1310.33
tx2.img	-0.34	-0.69	45.00	23.11	0.00	-1308.76
tx3.img	0.00	8.33	46.75	22.37	0.00	-1310.33
tx4.img	0.03	-10.19	44.63	20.82	0.00	-1310.38
tx5.img	-0.55	-0.69	45.00	20.03	0.00	-1308.73
tx6.img	0.00	8.34	44.63	19.28	0.00	-1310.44
tx7.img	0.03	-10.04	44.63	17.74	0.00	-1310.38
tx8.img	-0.73	-0.58	45.00	16.95	0.00	-1308.68
tx9.img	-3.12	8.98	46.96	16.14	0.00	-1305.69
tx10.img	0.03	-9.90	44.63	14.65	0.00	-1310.38
tx11.img	1.33	-0.62	45.02	13.86	0.00	-1308.49
tx12.img	-3.13	9.17	46.97	13.06	0.00	-1305.52
tx13.img	0.00	-9.76	44.63	11.52	0.00	-1304.49
tx14.img	0.11	0.66	45.00	10.78	0.00	-1308.58
tx15.img	-3.09	9.26	46.97	9.99	0.00	-1305.70
tx16.img	3.06	-10.10	44.91	8.45	0.00	-1305.66
tx17.img	0.45	0.21	45.00	7.70	0.00	-1308.59
tx18.img	0.00	8.98	46.74	6.91	0.00	-1304.34
tx19.img	0.00	-9.47	44.63	5.37	0.00	-1304.50
tx20.img	0.00	1.28	44.99	4.62	0.00	-1308.42
tx21.img	-3.10	9.58	46.98	3.84	0.00	-1305.53
tx22.img	0.00	9.09	46.75	3.07	0.00	-1304.71
tx23.img	-1.03	-0.18	45.00	1.54	0.00	-1308.58
tx24.img	-0.02	9.22	46.75	0.77	0.00	-1304.57
tx25.img	-5.45	-5.41	44.74	-0.77	0.00	-1299.99
tx26.img	1.10	-0.27	45.01	-1.54	0.00	-1308.45
tx27.img	-0.06	9.36	46.75	-2.30	0.00	-1304.57
tx28.img	-5.46	-5.23	44.76	-3.83	0.00	-1300.06
tx29.img	-1.02	0.08	45.01	-4.62	0.00	-1308.58
tx30.img	-1.14	1.58	47.14	-5.39	0.00	-1308.25
tx31.img	-5.80	-4.76	44.81	-6.88	0.00	-1299.67
tx32.img	0.02	0.02	45.00	-7.70	0.00	-1308.63
tx33.img	-0.97	0.41	47.13	-8.47	0.00	-1308.44
tx34.img	-5.45	-4.96	44.79	-9.95	0.00	-1300.08

Table A.3 (continued)

tx35.img	-3.13	10.10	47.02	-9.22	0.00	-1305.68
tx36.img	-3.06	-9.23	44.39	-10.71	0.00	-1299.89
tx37.img	-0.90	0.39	45.01	-11.55	0.00	-1308.63
tx38.img	0.91	0.40	45.00	-12.32	0.00	-1308.52
tx39.img	-3.10	10.35	47.03	-13.06	0.00	-1305.54
tx40.img	-3.05	-9.06	44.40	-14.53	0.00	-1299.90
tx41.img	0.00	-8.56	44.63	-15.35	0.00	-1304.50
tx42.img	-0.75	0.57	45.01	-16.17	0.00	-1308.68
tx43.img	-3.09	10.48	44.93	-16.91	0.00	-1305.70
tx44.img	-3.09	10.52	47.04	-17.67	0.00	-1305.70
tx45.img	-3.13	10.54	47.05	-18.44	0.00	-1305.68
tx35.img	1.00	0.11	45.00	-10.78	0.00	-1308.49
tx45.img	-1.23	2.07	45.04	-20.79	0.00	-1308.29
tx46.img	-5.42	6.51	47.45	-21.42	0.00	-1299.96
tx47.img	-5.41	-4.42	44.84	-22.95	0.00	-1300.16
tx48.img	-5.44	6.38	45.36	-23.72	0.00	-1300.20
tx49.img	-3.04	-8.62	44.43	-25.24	0.00	-1299.91
tx50.img	-5.44	-4.21	44.86	-26.01	0.00	-1300.14

Table A.4: Results of Experiments for The Data Set in Table 6.4.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
ty48.img	-5.53	10.66	47.23	0.76	0.00	-1298.80
ty47.img	-2.96	-9.62	42.24	0.00	1.54	-1306.08
ty46.img	0.05	9.22	46.75	0.77	2.30	-1304.57
ty45.img	0.07	-9.17	44.63	0.00	3.07	-1304.56
ty44.img	-9.51	-3.05	45.11	0.00	4.61	-1305.80
ty43.img	-1.85	0.14	45.02	0.00	5.39	-1308.00
ty42.img	0.01	0.00	45.00	0.00	6.16	-1308.74
ty41.img	-9.38	2.96	45.63	0.00	7.65	-1299.98
ty40.img	0.25	0.96	45.00	0.00	8.47	-1308.61
ty39.img	0.28	0.95	45.00	0.00	9.24	-1308.61
ty38.img	-0.26	-0.54	45.00	-0.78	10.85	-1317.16
ty37.img	-5.13	5.53	45.28	0.00	11.52	-1305.47
ty36.img	6.02	5.40	44.74	0.00	12.24	-1300.10
ty35.img	0.56	9.24	46.75	0.00	13.82	-1304.65
ty34.img	0.61	9.21	46.75	0.00	14.59	-1304.68
ty33.img	3.70	9.63	46.52	0.00	15.30	-1300.11
ty32.img	0.72	9.20	46.75	0.00	16.89	-1304.71
ty31.img	0.73	9.22	46.75	0.00	17.66	-1304.70
ty30.img	3.84	9.62	46.52	0.00	18.36	-1300.15
ty29.img	0.85	-9.15	44.62	-0.77	19.96	-1304.76
ty28.img	-4.54	-5.29	44.75	-0.77	20.66	-1300.21
ty27.img	-0.45	0.10	45.01	-0.77	21.56	-1308.58
ty26.img	-8.28	-0.02	45.38	-0.77	23.03	-1304.57
ty25.img	-4.45	-5.31	44.75	-0.77	23.72	-1300.17
ty24.img	4.17	9.62	46.51	0.77	24.60	-1305.98
ty23.img	-0.03	0.30	47.12	0.78	26.36	-1317.20
ty22.img	1.16	9.12	46.75	0.77	26.88	-1304.88
ty21.img	4.32	-9.64	44.89	0.00	27.67	-1305.93
ty20.img	-4.26	-5.31	44.75	0.00	29.40	-1314.51
ty19.img	1.30	-9.13	44.62	0.00	29.95	-1304.95
ty18.img	-0.37	0.22	45.01	0.00	30.80	-1308.46
ty17.img	0.47	0.13	42.88	0.00	32.36	-1309.20
ty16.img	-6.44	7.60	45.53	0.00	32.82	-1297.17
ty15.img	-0.01	0.01	45.00	0.00	33.89	-1308.63

Table A.4 (continued)

ty14.img	-0.58	0.00	45.01	0.00	34.65	-1308.42
ty13.img	-0.32	0.18	45.01	0.00	36.19	-1308.41
ty12.img	0.90	0.01	44.99	0.00	36.98	-1308.84
ty11.img	7.30	5.24	44.75	0.00	37.49	-1300.31
ty10.img	1.68	9.06	46.76	0.00	39.17	-1305.21
ty9.img	-1.35	9.47	47.00	0.00	39.98	-1306.34
ty8.img	4.87	9.56	44.37	0.00	40.56	-1300.46
ty7.img	1.85	9.03	46.75	0.00	42.25	-1305.30
ty6.img	1.87	9.05	46.75	0.00	43.02	-1305.30
ty5.img	-3.66	10.49	47.24	0.00	43.59	-1299.58
ty4.img	0.01	-0.19	45.00	-0.78	45.74	-1317.21
ty3.img	2.01	9.02	46.75	0.00	46.10	-1305.41
ty2.img	2.01	-8.89	44.62	-0.77	46.86	-1305.41
ty1.img	2.10	-8.93	42.50	-0.77	48.41	-1305.71
ty0.img	2.14	-8.96	44.62	-0.77	49.17	-1305.52

Table A.5: Results of Experiments for The Data Set in Table 6.5.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
rx0.img	-3.687	3.603	44.883	0.34	-0.34	-591.341
rx5.img	5.168	-5.197	45.667	0.339	0.339	-590.82
rx10.img	8.926	1.455	45.447	0.341	1.363	-592.971
rx15.img	15.361	-1.807	44.818	0.341	2.384	-592.579
rx20.img	20.648	0.688	46.036	0.34	3.058	-591.311
rx25.img	26.517	0.548	46.307	0.339	4.069	-590.152
rx30.img	31.348	0.488	46.602	0.339	4.749	-590.316
rx35.img	36.18	0.461	46.038	0.339	5.759	-589.628
rx40.img	41.497	0.436	45.822	0.339	6.437	-589.623
rx45.img	46.301	0.452	45.576	0.339	7.122	-590.23
rx50.img	51.547	0.435	46.264	0.339	7.456	-589.834
rx55.img	56.075	-0.035	46.165	0.34	8.16	-591.665
rx60.img	61.134	-0.035	46.17	0.338	8.9	-590.345
rx65.img	66.408	-0.036	46.506	0.338	9.113	-587.51
rx70.img	71.631	1.001	46.609	0.338	9.471	-588.741
rx75.img	77.117	0.448	46.44	0.334	9.7	-582.452
rx80.img	81.625	-3.289	46.191	0	9.844	-590.745

Table A.6: Results of Experiments for The Data Set in Table 6.6.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
ry80.img	-1.04	-80.46	48.40	18.47	-0.34	-594.94
ry75.img	-0.87	-75.31	45.77	17.43	-0.34	-594.63
ry70.img	1.12	-70.39	44.22	16.41	-0.34	-594.77
ry65.img	-1.22	-65.32	45.65	15.38	-0.34	-594.51
ry60.img	-1.66	-60.41	46.41	14.32	-0.34	-593.39
ry55.img	1.53	-55.18	42.94	13.01	-0.34	-595.51
ry50.img	2.00	-50.53	44.18	11.94	-0.34	-593.75
ry45.img	1.91	-45.51	44.30	10.92	-0.34	-593.50
ry40.img	2.14	-41.01	43.84	9.54	-0.34	-592.69
ry35.img	-2.59	-35.55	46.53	8.52	-0.34	-593.01
ry30.img	-2.18	-31.13	45.39	7.15	-0.34	-592.32
ry25.img	2.00	-26.15	44.53	6.13	-0.34	-592.80
ry20.img	-1.98	-20.97	44.45	4.77	-0.34	-592.53
ry15.img	3.37	-15.80	44.49	3.74	-0.34	-591.85
ry10.img	-3.60	-11.50	44.86	2.72	-0.34	-592.36
ry5.img	-3.72	-7.22	45.07	1.70	-0.34	-591.12
ry0.img	-5.18	-5.24	45.24	0.68	-0.34	-590.76

Table A.7: Results of Experiments for The Data Set in Table 6.7.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
rx0.img	-0.50	0.51	47.12	0.00	0.00	-1308.47
rx5.img	9.35	0.03	45.38	0.00	0.77	-1304.30
rx10.img	11.06	2.24	43.89	0.00	2.30	-1302.44
rx15.img	15.33	1.62	44.66	0.00	3.06	-1300.49
rx20.img	20.81	0.12	46.88	0.00	3.84	-1305.55
rx25.img	24.71	1.14	47.46	0.00	4.63	-1311.70
rx30.img	31.06	0.87	46.82	0.00	5.39	-1307.30
rx35.img	36.41	0.15	45.12	0.00	6.15	-1305.36
rx40.img	41.17	0.60	45.84	0.00	6.93	-1308.47
rx45.img	45.51	0.52	47.25	0.00	7.73	-1313.33
rx50.img	50.88	0.01	45.30	0.00	8.50	-1312.54
rx55.img	54.72	0.02	45.02	-0.78	9.34	-1321.83
rx60.img	60.98	0.02	45.20	0.00	8.40	-1314.67
rx65.img	66.83	-0.41	47.35	-0.77	10.05	-1313.50
rx70.img	71.23	0.38	48.35	-0.07	9.78	-1310.56
rx75.img	76.96	0.83	49.34	-0.79	10.26	-1340.77
rx80.img	79.90	0.76	49.67	-0.78	10.36	-1341.67

Table A.8: Results of Experiments for The Data Set in Table 6.8.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
ry0.img	-3.06	9.68	46.99	0.77	0.00	-1305.76
ry5.img	-3.11	9.60	47.00	1.54	0.00	-1305.70
ry10.img	-3.50	-11.23	46.45	2.34	0.00	-1305.56
ry15.img	3.70	-16.45	44.43	3.07	0.00	-1305.76
ry25.img	1.16	-26.46	44.29	5.39	0.00	-1307.59
ry30.img	2.16	-31.17	45.62	6.90	0.00	-1303.37
ry35.img	3.15	-35.31	44.71	7.72	0.00	-1311.49
ry40.img	-3.48	-41.33	45.55	9.22	0.00	-1305.30
ry45.img	2.62	-46.68	45.09	10.00	0.00	-1307.05
ry50.img	2.54	-50.01	46.29	11.70	0.00	-1324.58
ry55.img	2.38	-55.72	46.00	12.40	0.00	-1316.43
ry60.img	2.52	-60.21	50.70	14.06	0.00	-1326.74
ry65.img	3.51	-66.13	45.79	14.81	0.00	-1324.46
ry70.img	-2.99	-70.22	49.62	16.49	0.00	-1334.05
ry75.img	-3.71	-74.98	50.29	17.38	0.00	-1341.59

Table A.9: Results of Experiments for The Data Set in Table 6.9.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
ry250.img	26.93	-0.99	45.54	0.57	3.43	-501.53
ry255.img	26.74	-5.31	45.17	1.43	3.44	-502.17
ry2510.img	26.32	-11.03	44.18	2.30	3.45	-503.65
ry2515.img	27.17	-15.03	45.31	3.44	3.44	-502.83
ry2520.img	26.79	-20.28	44.88	4.31	3.45	-504.11
ry2525.img	26.91	-25.31	45.23	5.46	3.45	-503.97
ry2530.img	26.68	-30.55	45.34	6.62	3.45	-504.75
ry2535.img	26.59	-35.34	45.55	7.80	3.47	-506.20
ry2540.img	26.77	-40.57	45.33	8.96	3.47	-506.60
ry2545.img	26.72	-45.66	45.07	10.07	3.45	-504.69
ry2550.img	26.52	-50.78	44.67	11.25	3.46	-505.57
ry2555.img	26.39	-55.78	42.93	12.12	3.46	-506.11
ry2560.img	26.48	-60.53	43.62	13.27	3.46	-505.68
ry2565.img	26.24	-65.51	44.46	14.51	3.48	-508.59
ry2570.img	26.24	-70.80	44.99	15.58	3.46	-505.93
ry2575.img	26.09	-75.96	43.58	16.48	3.47	-506.83
ry2580.img	25.79	-81.01	41.94	17.36	3.47	-507.28

Table A.10: Results of Experiments for The Data Set in Table 6.10.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
rx025.img	2.07	-25.62	45.80	6.08	-1.16	-507.69
rx525.img	7.21	-25.43	44.94	6.07	0.00	-506.99
rx1025.img	11.89	-25.60	45.84	6.06	0.87	-506.08
rx1525.img	17.32	-25.38	44.58	5.76	1.73	-505.19
rx2025.img	21.95	-25.75	45.07	5.76	2.59	-505.20
rx2525.img	26.82	-25.97	44.92	5.47	3.46	-505.01
rx3025.img	31.79	-25.99	43.56	5.19	4.32	-505.33
rx3525.img	36.63	-25.97	44.78	5.18	4.89	-504.38
rx4025.img	41.94	-25.33	45.09	4.89	5.75	-504.19
rx4525.img	46.77	-25.37	45.19	4.60	6.33	-504.23
rx5525.img	56.77	-25.66	45.27	4.02	7.47	-503.99
rx6525.img	67.09	-25.85	44.73	3.14	8.58	-501.39
rx7025.img	71.83	-25.35	44.62	2.85	8.85	-500.63
rx7525.img	76.48	-26.28	44.39	2.56	9.11	-499.22
rx8025.img	81.79	-25.59	43.25	2.00	9.42	-500.94

Table A.11: Results of Experiments for The Data Set in Table 6.11.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
ry250.img	28.4	-3.78	45.67	0	4.67	-1316.67
ry255.img	28.47	-4.69	45.25	0.00	4.65	-1316.72
ry2510.img	28.27	-9.98	44.78	1.30	4.60	-1317.24
ry2515.img	28.24	-15.22	43.79	2.33	4.66	-1318.82
ry2520.img	27.36	-19.55	44.21	3.13	4.70	-1329.86
ry2525.img	27.22	-24.61	46.13	4.71	4.71	-1333.29
ry2530.img	26.51	-29.95	45.86	5.52	4.73	-1339.73
ry2535.img	26.64	-35.69	47.67	7.09	4.73	-1337.50
ry2540.img	26.67	-40.56	43.20	7.90	4.74	-1341.29
ry2545.img	26.91	-45.49	46.73	9.44	4.72	-1336.26
ry2550.img	26.68	-50.74	45.79	10.13	4.68	-1323.91
ry2555.img	26.99	-55.23	41.77	11.02	4.72	-1336.35
ry2560.img	26.41	-60.44	43.17	12.66	4.75	-1343.29
ry2565.img	26.48	-66.09	40.78	13.48	4.76	-1346.74
ry2570.img	26.66	-71.00	43.57	14.97	4.73	-1338.10
ry2575.img	26.16	-76.90	39.48	15.84	4.75	-1344.90

Table A.12: Results of Experiments for The Data Set in Table 6.12.

Image File	Rx	Ry	Rz	Tx	Ty	Tz
rx025.img	-1.06	-26.47	46.58	5.39	0.00	-1307.08
r525.img	7.88	-26.28	44.65	5.39	0.77	-1307.88
r1025.img	10.60	-27.42	47.19	4.59	2.30	-1300.38
r1525.img	16.33	-24.23	46.35	4.61	3.07	-1305.35
r2025.img	21.95	-25.87	43.53	4.58	3.82	-1297.73
r2525.img	27.52	-25.16	45.87	4.63	4.63	-1310.40
r3025.img	31.40	-27.06	45.45	4.60	5.37	-1302.90
r3525.img	36.41	25.86	66.05	4.61	6.14	-1304.89
r4025.img	42.42	-25.25	45.36	3.89	7.01	-1323.16
r4525.img	47.10	-25.02	46.62	3.87	7.74	-1315.40
r5025.img	52.37	-27.21	44.38	3.07	8.45	-1305.79
r5525.img	56.72	-25.93	45.34	3.10	9.29	-1315.21
r6025.img	62.36	-26.49	43.80	2.35	9.38	-1328.27
r6525.img	67.22	-28.04	44.61	2.34	10.15	-1326.43
r7025.img	72.94	-26.50	44.50	1.55	10.10	-1319.82
r7525.img	78.27	-29.80	45.29	1.59	10.31	-1347.15
r8025.img	83.60	-27.58	44.91	0.78	10.97	-1330.98

APPENDIX B

TECHNICAL DRAWING AND MEASUREMENTS FOR DETERMINATION OF SPECIFICATIONS OF THE BASE PLATE

The distances between mounting holes and flatness of the base plate are measured by using a coordinate measurement machine. This coordinate measurement machine is a 3-axes computer controlled model KM 2956 of KEMCO Co., UK, located in Computer Integrated Manufacturing Laboratory of the Mechanical Engineering Department of METU. Distance measurements performed are tabulated in Table B.1. From the results in this table accuracy of the distances between mounting holes on the base plate is found to be correct within about ± 0.1mm.

Measurements to determine the flatness of the base plate are tabulated in Tables B.2 and B.3. Reference positions for the measurements in Tables B.2 and B.3 are about 120mm and 500mm from the left end of the base plate, respectively. The left end is the end of the base plate to which two mounting holes are 15mm in distance. From the results in these tables flatness of the base plate is found to be within the ranges of approximately ± 0.4mm and ± 0.1mm in X and Y directions, respectively.

Table B.1: Distances Between Mounting Holes On the Base Plate

Hole Numbers	Measured Value[mm]	Designed Value[mm]
1 and 2	124.98	125.00
2 and 3	125.00	125.00
3 and 4	125.10	125.00
4 and 5	125.13	125.00
5 and 6	49.94	50.00
6 and 7	74.91	75.00
7 and 8	50.05	50.00
8 and 9	74.95	75.00
9 and 10	125.06	125.00
10 and 11	49.89	50.00
11 and 12	75.06	75.00
12 and 13	50.03	50.00
13 and 14	75.06	75.00
14 and 15	124.83	125.00
15 and 16	50.07	50.00
16 and 17	75.06	75.00
17 and 18	49.95	50.00
18 and 19	75.05	75

Table B.2: Flatness Measurements of the Base Plate

X[mm]	Y[mm]	Z[mm]	X[mm]	Y[mm]	Z[mm]	X[mm]	Y[mm]	Z[mm]
Reference Position								
0	0	0	0	63.48	-0.01	105.33	275.22	0.40
			0	135.46	0	105.33	155.64	0.27
			0	239.21	0.05	105.33	59.86	0.24
			0	315.70	0.11	105.33	-2.85	0.23
			0	380.50	0.17			
Average Value					0.064			0.285
Standard Deviation					0.076			0.079

Table B.3: Flatness Measurements of the Base Plate

X[mm]	Y[mm]	Z[mm]	X[mm]	Y[mm]	Z[mm]	X[mm]	Y[mm]	Z[mm]
Reference Position								
0	0	0	0	112.66	-0.05	135.21	463.014	0.27
			0	173.73	-0.08	135.21	336.91	0.16
			0	238.21	-0.11	135.21	262.32	0.14
			0	287.43	-0.13	135.21	210.10	0.14
			0	362.73	-0.14	135.21	160.74	0.16
			0	472.01	-0.09	135.21	141.89	0.16
						135.21	90.43	0.18
Average Value					-0.1			0.17
Standard Deviation					0.03			0.05

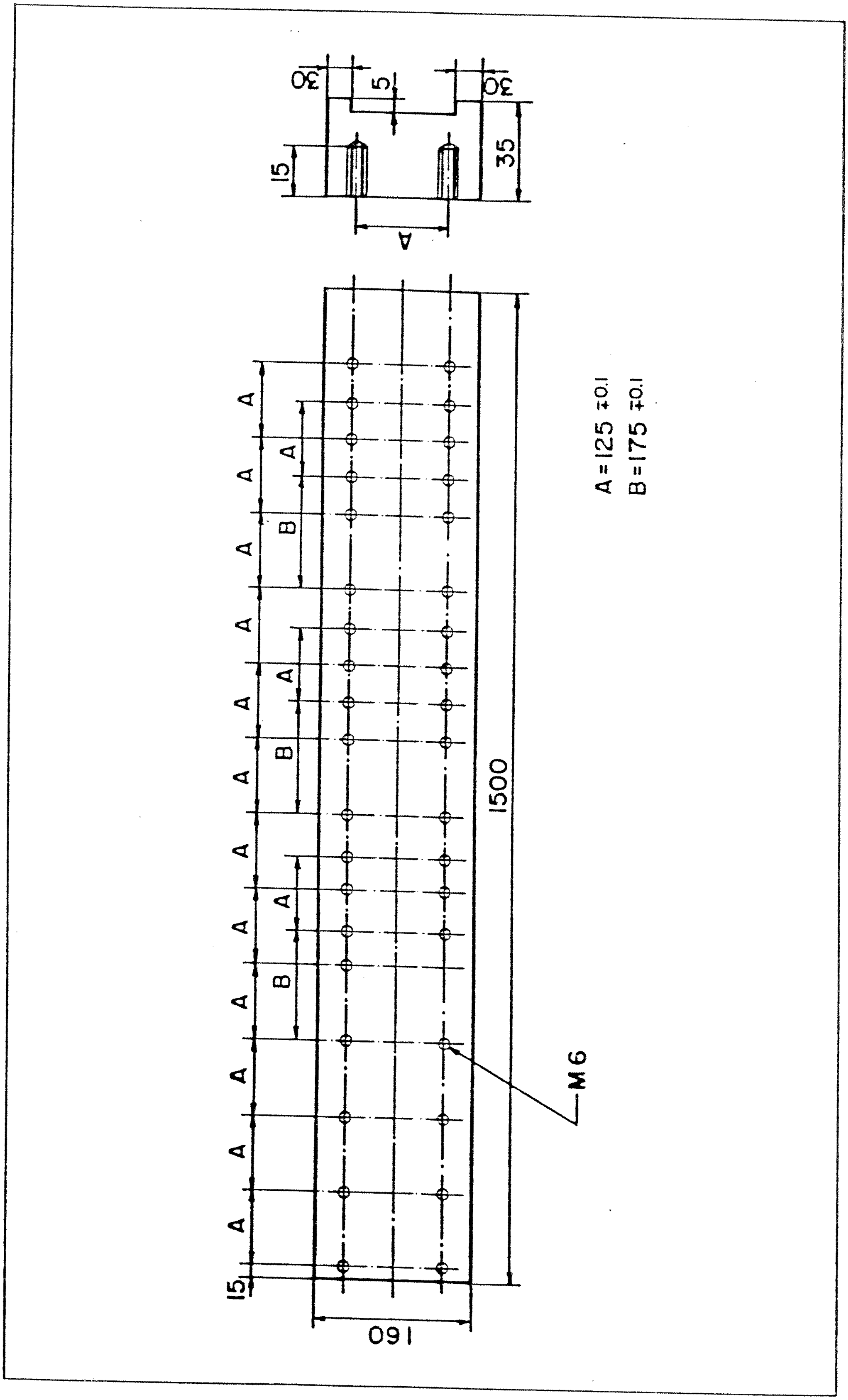


Figure B.1: Technical Drawing of the Base Plate

BİBLİYOGRAFİK BİLGİ FORMU

1- Proje No: MODİSA-8

2- Rapor Tarihi: Temmuz 1996

3- Projenin Başlangıç ve Bitiş Tarihleri: 1 Ağustos 1992 ve 1 Mayıs 1995

4- Projenin Adı:
TEK KAMERA GÖRÜNTÜSÜ İLE CİSİMLERİN 3 BOYUTLU KONUMLARININ BELİRLENMESİ

5- Proje Yürütücüsü ve Yardımcı Araştırıcılar:
Prof. Dr. Bülent E. Platin
Prof. Dr. Ömer Anlağan, İ. Alpar Kılınç, Ercan U. Acar

6- Proje Yürütüldüğü Kuruluş ve Adresi:
Makina Mühendisliği Bölümü
Orta Doğu Teknik Üniversitesi, Ankara 06531

7- Destekleyen Kuruluş(ların) Adı ve Adresi:
TÜBİTAK

8- Öz (Abstract): Bu çalışmadaki temel amaç, bir cismin 3 boyutlu uzaydaki konumunun bu cisim üzerine monte edilmiş ikincil bir hedefin tek kamera görüntüsü ile belirlenmesini sağlayacak bir sistem geliştirilmesidir. Bu amaca, bu proje kapsamında yürütülen kuramsal ve deneysel çalışmalarda elde edilen sonuçlar ile tümü ile ulaşılmıştır. Geliştirilen sistem yardımı ile, 500 mm - 2000 mm uzaklıktaki bir cismin konumunu ortalama 1 mm ve 0.5° mertebelerinde hassasiyet ile saptamanın mümkün olduğu hem kuramsal çalışmalar ile hem de yürütülen kontrollü deneyler ile gösterilmiştir. Geliştirilen sistemin dayandırıldığı çözüm algoritması, dairesel bir hedefin tek bir kameradaki tek bir görüntüsü yardımı ile bu hedefin dolayısı ile bu hedefin bağlandığı cismin 3 boyutlu uzaydaki konumunun veren 6 konum bilgisini üretmektedir. Bu algoritma daha önce önerilmiş olmasına rağmen, tüm boyutu ile ilk kez bu çalışmada kullanılmış, uygulanabilirlik sınırları kuramsal ve deneysel yöntemlerle test edilmiş ve oluşturulan sistemin performansı saptanmıştır. Bu nedenle, bu çalışma yalnızca alanındaki kuramsal boşlukları kapatmakla kalmayıp, robot kollarının kontrolu, otomatik sürücülü taşıtlar, uzayda buluşma ve montaj vb. gibi çok değişik mühendislik alanlarında uygulamaya dönük somut bir sistem ortaya çıkarmıştır.

Anahtar Kelimeler: Konum Algılayıcı, Monoküler Görüntü, Kamera Kalibrasyonu

9- Proje ile İlgili Yayın/Tebliğlerle ilgili Bilgiler:
ODTÜ Makina Mühendisliği Bölümü'nde yürütülmüş 2 adet yüksek lisans tezi

10- Bilim Dalı:
Doçentlik Bilim Dalı Kodu: 625.01.03 **ISIC Kodu:**
Uzmanlık Alanı Kodu:

11- Dağıtım: ☐ Sınırlı ☑ Sınırsız

12- Raporun Gizlilik Durumu: ☐ Gizli ☑ Gizli Değil