

**Videolarda Görüntü, Ses ve Metin Verileri Kullanılarak  
Anlamsal Bilgi Çıkarımı, Depolanması ve Sorgulanması**

**Proje No: 109E014**

Prof.Dr. Adnan YAZICI

Yrd.Doç.Dr. Murat KOYUNCU

Yrd.Doç.Dr. Mustafa SERT

ŞUBAT 2013  
ANKARA

# ÖNSÖZ

Bu proje kapsamında, videoların görüntü, ses ve metin verileri kullanılarak (multi-modal) anlamsal bilgilerin otomatik olarak çıkarıldığı, uygun formatlarda saklandığı ve daha sonra da etkin bir şekilde sorgulanabildiği bir prototip sistem geliştirilmiştir. Geliştirilen sisteme yüklenen bir video öncelikle bazı ön işlemlerden geçirilerek görüntü, ses ve metin verileri elde edilmektedir. Elde edilen görüntü, ses ve metin verileri, her veri tipi için ayrı ayrı geliştirilmiş olan ilgili üç modül tarafından işlenerek videonun içerdiği anlamsal içerikler çıkarılmaya çalışılmaktadır. Daha sonra, bu üç modülden gelen bilgilerin analiz edilmesi, bütünleştirilmesi, eksiklerinin tamamlanması, dublikasyonlarının temizlenmesi ve böylece veritabanına kaydetmeye hazır bilginin elde edilmesini sağlayan bilgi füzyonu işlemi yapılmaktadır. Füzyon sonunda elde edilen video bilgileri, araştırmacıların yine bir TÜBİTAK 1001 Projesi kapsamında daha önce geliştirdikleri Akıllı ve Bulanık Nesneye Dayalı Veritabanı sisteminde saklanmaktadır. Akıllı ve Bulanık Nesneye Dayalı Veritabanı Sistemi temel olarak bir bulanık nesneye dayalı veritabanı ile bir bulanık bilgi tabanından oluşmaktadır. Uygulama alanıyla ilgili büyük boyutlardaki çoklu-ortam verileri nesneye dayalı veritabanında saklanırken, uygulama alanına özgü kuralların tanımlandığı bilgi tabanı vasıtasıyla da veritabanındaki veriler kullanılarak yeni anlamsal bilgiler çıkarılabilmektedir. Ayrıca, yapılacak sorgulara hızlı cevaplar üretebilmek amacıyla hem videolardan çıkarılan anlamsal bilgilerin hem de videoların alt seviye özelliklerinin birlikte kullanıldığı bir dizin yapısı geliştirilmiştir. Geliştirilen sistemde, belirsizlik ve bulanıklık içeren bilgiler de işlenebilmektedir. Böylece bu proje çerçevesinde yapılan araştırmaların sonucunda geliştirmiş olduğumuz prototip sistem, çokluortam veritabanlarının olası kullanıcılarının gereksinim duyacağı bir çok sorgu tipini etkin bir şekilde destekleyebilmektedir.

Bu projenin özgün değeri farklı modlardaki (ses, görüntü ve metin) video verilerinden elde edilen bilgilerin füzyon edilmesi ve böylece daha bütünsel bir anlamsal bilgi seti oluşturularak veritabanında saklanmasında ve etkin olarak sorgulanmasında yatmaktadır. Sisteme yüklenen yeni bir videonun baştan sona otomatik olarak anotasyonunu yapabilen, çıkarılan anlamsal verileri ve alt seviye özellikleri indeksleyerek depolayabilen ve yapılan sorgulamalarla videonun ilgili bölümlerine hızlıca erişimi sağlayan komple bir sistemin gerçekleştirilmiş olması da önemli bir özgün değerdir.

Proje bitiminde, proje öneri dokümanında yer alan başarı kriterleri sağlanarak proje başlangıcında hedeflenen noktaya ulaşılmıştır. Proje kapsamında, 7 adeti SCI-E tarafından tarananlar olmak üzere 8 adet uluslararası dergilerde ve 21 adet konferanslarda (19 adet uluslararası, 2 adet ulusal) olmak üzere toplam 29 adet yayın gerçekleştirilmiştir. Proje kapsamında projenin değişik süreçlerinde görev alan 3 doktora ve 10 lisansüstü öğrencisinin tez çalışmasına imkan verilerek bu öğrencilerin tezlerini başarıyla bitirmeleri sağlanmıştır. Daha bir kaç lisansüstü tez çalışması da devam etmektedir.

Bu proje, BİLİMSEL VE TEKNOLOJİK ARAŞTIRMA PROJELERİNİ DESTEKLEME PROGRAMI kapsamında TÜBİTAK tarafından (109E014 kod numarasıyla) desteklenmiştir.

# İÇİNDEKİLER

ÖNSÖZ .....	i
İÇİNDEKİLER.....	ii
TABLO LİSTESİ .....	iii
ŞEKİL LİSTESİ.....	iv
ÖZET .....	v
ABSTRACT .....	vi
1. GİRİŞ .....	1
2. GENEL BİLGİLER.....	2
3. GEREÇ VE YÖNTEM .....	4
3.1 Genel .....	4
3.2 Görüntüden Anlamsal Bilgi Çıkarımı (2. İş Paketi).....	4
3.2.1 Çekim (Shot) ve Ana Çerçeve (Keyframe) Çıkarımı .....	5
3.2.2 Çekim ve Ana Çerçevelerden Nesne ve Olay Çıkarımı .....	5
3.2.3 Elle Çıkarım .....	5
3.2.4 Otomatik Çıkarım .....	5
3.2.5 Nesne Çıkarımı .....	6
3.2.6 Anlamsal İçerik Modeli (Video Semantic Content Model - VISCOM) .....	8
3.2.7 Uzamsal (Spatial) İlişki Çıkarımı.....	10
3.2.8 Zamansal (Temporal) İlişki Çıkarımı.....	10
3.2.9 Olay Çıkarımı .....	11
3.3 Metinden Anlamsal Bilgi Çıkarımı (3. İş Paketi).....	14
3.4 Sesten Anlamsal Bilgi Çıkarımı (4. İş Paketi) .....	19
3.5 Bilgi füzyonu (5. İş Paketi).....	29
3.6 Video Dizin Yapısı (6. İş Paketi) .....	33
3.7 Veritabanı Entegrasyonu (7. İş Paketi) ve Prototip Sistemde Geline Son Durum..	39
3.8 Farklı Alan Uygulamaları ve Testler (8. İş Paketi).....	42
4. BULGULAR.....	43
5. SONUÇ .....	46
YARARLANILAN KAYNAKLAR .....	49

## TABLO LİSTESİ

Tablo-1: N-Cut bölütleme yöntemi ile bölüt sayısına göre bölütleme süreleri .....	7
Tablo-2: Örnek bir videodan nesne çıkarımı süreleri .....	8
Tablo-3: Nesne çıkarımı doğruluk oranları.....	8
Tablo-4: Ofis güvenliği olay çıkarımı test sonuçları .....	13
Tablo-5: Basketbol olay çıkarımı test sonuçları .....	13
Tablo-6: Futbol olay çıkarımı test sonuçları .....	13
Tablo 7: Veri seti ile ilgili özet bilgi. ....	15
Tablo-8: Kural tabanlı belirleyicinin test sonuçları .....	15
Tablo-9: Melez belirleyicinin test sonuçları .....	16
Tablo 10: Sınıflandırma sonuçları (“Konuşma” – “Konuşma İçermeyen”).....	22
Tablo 11: Sınıflandırma sonuçları (“Saf Konuşma” – “Konuşmayla Karışık”) .....	22
Tablo 12: Sınıflandırma sonuçları (“Konuşma+Müzik” – “Konuşma+Çevresel Ses”).....	22
Tablo 13: Sınıflandırma sonuçları (“Çevresel Ses” – “Müzik”).....	22
Tablo 14: Çevresel sesler için oluşturulan veri kümesi .....	25
Tablo 15: ASF+ASC+ASS+ZCR bileşik özniteliğinin HMM başarımı .....	27
Tablo 16: ASF+ASC+ASS+AH bileşik özniteliğinin HMM başarımı.....	27
Tablo 17: MFCC+ASC+ASS+ZCR bileşik özniteliğinin HMM başarımı .....	28
Tablo 18: MFCC+ASC+ASS+AH bileşik özniteliğinin HMM başarımı .....	28
Tablo 19: ASF+ASC+ASS+AH (ASFCS-H) bileşik özniteliğinin SVM başarımı.....	29
Tablo-20: TRECVID 2007 veriseti üzerinde tek modalite ve füzyon sonuçlarının karşılaştırılması .....	32
Tablo-21: CCV Database veriseti üzerinde tek modalite ve füzyon sonuçlarının karşılaştırılması .....	33

## ŞEKİL LİSTESİ

Şekil-1: Sistem mimarisi .....	3
Şekil-2: Otomatik anlamsal içerik çıkarım sistemi .....	6
Şekil-3: N-Cut bölütleme örneği.....	7
Şekil-4: VISCOM sınıfları ve ilişkileri.....	9
Şekil-5: Örnek bir olay .....	11
Şekil 6: Türkçe için özel anlamlı kelime belirleyicisi tarafından kullanılan kaynakların sınıflandırılması .....	14
Şekil-7: Gürültülü Haber Video Metinlerinden Olay Çıkarım Değerlendirme Sonuçları.....	18
Şekil-8: Temiz Haber Video Tekstlerinden Olay Çıkarım Değerlendirme Sonuçları. ....	18
Şekil-9: Sınıflandırma ağacı .....	20
Şekil-10: Önerilen sistemin blok şeması .....	20
Şekil-11: Ses Sınıflandırma Ağacı .....	23
Şekil-12: Ses özniteliklerinin HMM performansı.....	26
Şekil-13: Ses özniteliklerinin SVM performansı .....	26
Şekil-14: Otomatik Bilgi Çıkarımı .....	30
Şekil-15: FOOD-İndeks yapısı .....	34
Şekil-16: Örnek X-Tree yapısı .....	35
Şekil-17: Mesafe tablosu .....	36
Şekil-18: LMDS sonucunda konumsal yerleştirme.....	36
Şekil-19: Multimedia dizin yapısı .....	37
Şekil-20: Multimedia dizin yapısı .....	38
Şekil-21: İlk video yükleme ekranı .....	40
Şekil-22: Görsel anotasyon ekranı.....	40
Şekil-23: Ses anotasyon ekranı .....	41
Şekil-24: Sorgu ekranı .....	41

## ÖZET

Bu projede, videoların görüntü, ses ve metin verileri kullanılarak (multi-modal) otomatik anlamsal bilgilerin çıkarıldığı, uygun formatlarda saklandığı ve daha sonra da etkin bir şekilde sorgulanabildiği bir prototip sistem geliştirilmiştir. Geliştirilen sistemde, bazı ön işlemlerin yapılmasını takiben, öncelikle görüntü, ses ve metin verilerinin her birinden ayrı ayrı bilgiler çıkarılmaktadır. Daha sonra, bu üç farklı veri tipinden gelen bilgilerin analiz edilmesi, bütünleştirilmesi, eksiklerinin tamamlanması, dublikasyonlarının temizlenmesi ve böylece veritabanına kaydetmeye hazır bilginin elde edilmesi amacıyla bilgi füzyonu yapılmaktadır. Füzyon sonunda elde edilen video bilgileri, çoklu-ortam verileri için projenin araştırmacıları tarafından daha önce yine bir TÜBİTAK 1001 Projesi kapsamında geliştirilmiş olan Akıllı ve Bulanık Nesneye Dayalı Veritabanı sisteminde saklanmaktadır. Ayrıca, sorgulara hızlı cevaplar üretebilmek amacıyla videolardan çıkarılan anlamsal bilgiler ile videoların alt seviye özelliklerinin indekslenemediği etkin bir dizin yapısı geliştirilmiştir. Geliştirilen sistemde, belirsizlik ve bulanıklık içeren bilgiler de işlenebilmektedir. Böylece bu proje çerçevesinde yapılan araştırmaların sonucunda geliştirmiş olduğumuz prototip sistem, çokluortam veritabanlarının olası kullanıcılarının gereksinim duyacağı bir çok sorgu tipini etkin bir şekilde destekleyebilmektedir.

Üç yıl olarak başlatılan proje 4 aylık uzatmayla birlikte 40 ayda tamamlanmıştır. Projede, bulanıklık, nesneye dayalı veritabanı, bilgi tabanlı sistemler ve çoklu ortam uygulamaları konusunda uzmanlaşmış üç araştırmacı görev almıştır. Proje kapsamında projenin değişik süreçlerinde görev alan 3 doktora ve 10 lisansüstü öğrencisinin tez çalışmasına imkan verilerek bu öğrencilerin tezlerini başarıyla bitirmeleri sağlanmıştır. Daha bir kaç lisansüstü tez çalışması da devam etmektedir. Ayrıca, proje sonunda geliştirilen model ve prototip sistemin, akademik dünyada önemli bir boşluğu doldurduğu değerlendirilmektedir. Proje sürecinde 8 adet (7 adet SCI-E tarafından taranan) dergilerde ve 21 adet konferanslarda (19 adet uluslararası, 2 adet ulusal) olmak üzere toplam 29 adet yayın yapılmıştır. Proje sonunda gönderilen ilave makale ve bildirimlerle bu sayının daha da artması beklenmektedir.

**Anahtar kelimeler:** çoklu ortam uygulamaları, video, anlamsal bilgi çıkarımı, nesneye dayalı veritabanı, bilgi tabanlı sistem, füzyon, bulanıklık, belirsizlik

## ABSTRACT

In this project, in order to extract semantic data automatically by using video's visual, audio and textual (multi-modal) data, to store in an appropriate format and then to query in an efficient way, a prototype system has been developed. In the developed system, firstly, semantic contents are extracted from visual, audial and textual data separately after a preprocessing step. Then, the outputs of these three modalities are fused to analyze and unite them, to fill their missing parts, and to clean their duplications. Hence, we get information ready to store in a database. The video data gathered as a result of the information fusion is stored in an Intelligent and Fuzzy Object Oriented Database System which was developed in a TUBITAK 1001 project before by the researchers. Moreover, in order to produce effective answers to the queries, an indexing mechanism that unites extracted video information and video's low level features, has been developed. Information including imprecision and fuzziness is also processed in the developed system.

The project has been terminated in 40 months with a four months extension. In the project, three researchers, who are experts on multimedia applications fuzziness, fuzzy logic, object-oriented database, knowledge-based systems, have been worked. An opportunity is provided for 3 PhD and 10 MS students, who took responsibility during different terms of the project, to work on and finish their thesis successfully. A couple of graduate students is expected to finish their thesis soon. In addition, it is evaluated that the obtained results of the project fill a big gap in the academic literature. During project, 8 journal papers (7 of them in SCI-E) and 21 conference papers (19 international, 2 national), which make 29 in total, are published. With additional papers submitted during final days of the project, we expect increase in the total number of publications.

**Keywords:** Multimedia applications, video, semantic content extraction, object-oriented database, knowledge-based system, fusion, fuzziness, uncertainty

# 1. GİRİŞ

Bilgi ve iletişim teknolojilerindeki gelişmelere bağlı olarak son yıllarda video kullanımı hızla artmış ve günlük hayatta kullanılan birçok uygulamada videoların anlamsal içeriklerinin çıkarılması ve anlamsal içerikler üzerinden etkin sorguların yapılabilmesi ihtiyacı ortaya çıkmıştır. Videoların sadece isim, kayıt tarihi, çözünürlük gibi sınırlı verilerle saklanması ve sorgulanması günümüzde kullanıcı ihtiyaçlarına cevap verecek durumda değildir. Videoların anlamsal içeriklerinin çıkarılması, veritabanlarında saklanması ve ihtiyaç olduğunda bu içeriklerin kullanılarak videonun ilgili bölümlerine hızlı erişimin sağlanması gerekmektedir. Anlamsal içerikten kastedilen videoların içeriğinde yer alan nesnelere ve olaylar ile nesne-nesne, nesne-olay ve olay-olay arasındaki mekânsal ve zamansal ilişkilerdir. Günümüzde videoların anlamsal içerikleri genelde kullanıcılar tarafından manuel olarak çıkarılmaktadır. Oysa manuel çıkarım etkin olmayan, zahmetli, pahalı, çıkarılan anlamların kullanıcılara göre oldukça değişim gösterebileceği bir yöntemdir. Bu nedenle, videolardan otomatik anlamsal bilgi çıkarımı kaçınılmaz hale gelmektedir. Videolardan anlamsal bilgi çıkarımı sıcak araştırma konularından birisidir ve önümüzdeki dönemde bu konudaki araştırmaların artarak devam edeceği tahmin edilmektedir.

Bu proje kapsamında, videoların görüntü, ses ve metin verileri kullanılarak (multi-modal) otomatik anlamsal bilgilerin çıkarıldığı, uygun formatlarda saklandığı ve daha sonra da etkin bir şekilde sorgulanabildiği bir sistem geliştirilmiştir. Proje kapsamında öncelikle görüntü, ses ve metin verilerinin her birinden ayrı ayrı anlamsal bilgiler çıkarılmaktadır. Daha sonra, bu üç modülden gelen bilgilerin analiz edilmesi, bütünleştirilmesi, eksiklerinin tamamlanması, dublikasyonlarının temizlenmesi sonucunda veritabanına kaydetmeye hazır bilgi elde edilmektedir. Kısaca bilgi füzyonu olarak isimlendirdiğimiz bu işlem projenin önemli faaliyetlerinden birisi olmuştur. Füzyon sonunda elde edilen video bilgileri, akıllı ve bulanık nesneye dayalı bir veritabanı sisteminde saklanmaktadır. Akıllı ve Bulanık Nesneye Dayalı Veritabanı Sistemi, nesneye dayalı bir veritabanı ile bir bilgi sisteminin bulanık verileri de işleyecek şekilde geliştirilmesi ve entegrasyonu sonucu elde edilmiş bir sistemdir. Nesneye dayalı veritabanı videolardan elde edilmiş anlamsal verilerin depolanması için kullanılırken, bilgi tabanı video alanıyla ilgili kurallar tanımlamak ve veritabanında saklanmayan anlamsal verileri çıkarmak için kullanılmaktadır. Ayrıca, yapılacak sorgulara hızlı cevaplar üretebilmek amacıyla hem videolardan çıkarılan anlamsal bilgilerin hem de videoların alt seviye özelliklerinin kullanıldığı bir dizin yapısı da proje kapsamında geliştirilmiştir.

Otomatik anlamsal bilgi çıkarımı sonucunda çıkarılan bilginin her zaman %100 doğruluğunu garanti etmek mümkün değildir. Diğer bir deyişle bazen bilgiler belirli bir doğruluk derecesiyle çıkarılabilmekte ve elimizde bulanıklık veya belirsizlik içeren bilgiler oluşmaktadır. Diğer taraftan başka araştırmacıların da belirttiği gibi videodaki bazı bilgi türleri de bulanık (fuzzy) olarak ifade edilebilmektedir. Örneğin, videoda, hava durumuyla ilgili olarak sıcak, ılık, oldukça soğuk, az bulutlu, vs gibi terimler kullanılabilir. Dolayısıyla, geliştirilen sistemde belirsizlik ve bulanıklık içeren bilgiler de işlenebilmektedir.

Proje, birbirlerini takip eden veya birbirlerine paralel yürüyen sekiz iş paketi (listesi aşağıdadır) halinde tanımlanmış ve yürütülmüştür.

- İş paketi 1- Hazırlık
- İş paketi 2- Görüntüden anlamsal bilgi çıkarımı
- İş paketi 3 - Metinden anlamsal bilgi çıkarımı
- İş paketi 4 - Sesten anlamsal bilgi çıkarımı



- İş paketi 5 - Bilgi füzyonu
- İş paketi 6 – Video dizin yapısı
- İş paketi 7 - Veritabanı Entegrasyonu
- İş paketi 8 - Farklı alan uygulamaları ve testler

Proje sonunda, yukarıda verilen iş paketlerinin tamamı bitirilmiştir. Proje boyunca uluslararası dergi ve kongreler için makale ve bildiri hazırlama çalışmaları iş paketlerinden bağımsız sürekli bir faaliyet olarak yürütülmüştür. Gerekli birikim sağlandıkça yeni makale ve bildiriler üretilmiştir.

Proje kapsamında geliştirilen prototip sistemin daha geniş bir özeti Genel Bilgiler başlığı altında bir sonraki bölümde verilmiştir. Proje kapsamında yapılan ayrıntılı Ar-Ge çalışmaları ve sağlanan ilerlemeler Gereç ve Yöntem bölümünde sunulmuştur. Projede elde edilen sonuçlar Bulgular bölümünde özetlenmiştir. Sonuç bölümünde projenin genel değerlendirmesi yapılmış ve olası araştırma konuları irdelenmiştir.

## 2. GENEL BİLGİLER

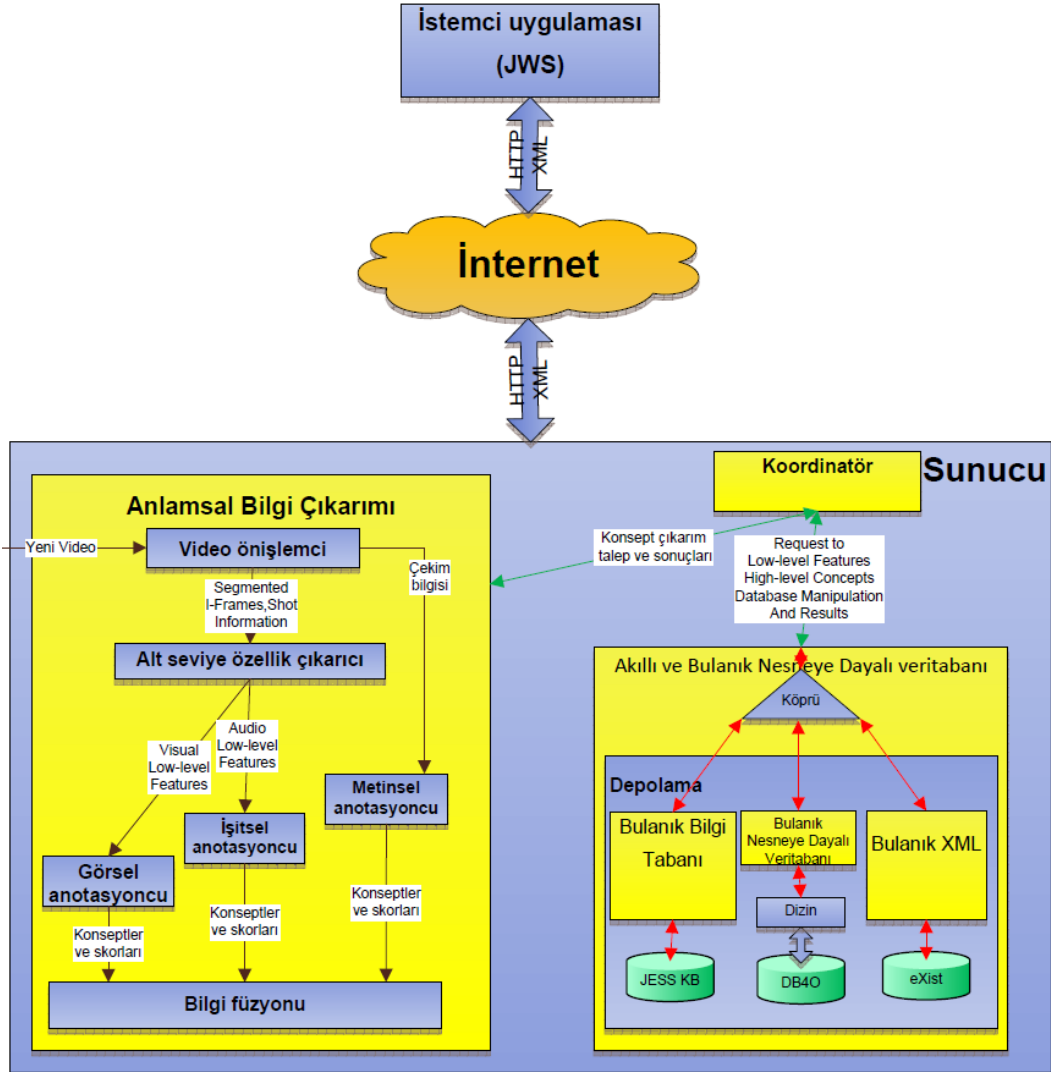
Günümüzde videoların içeriğinde saklı olan anlamsal bilgilerin işlenmesi, önemi her geçen gün artan bir araştırma alanı haline gelmiştir. Literatürde videolardaki anlamsal bilgilerin modellenmesi, saklanması, dizinlenmesi, sorgulanması gibi konularda birçok çalışma bulunmaktadır. Bu çalışmaların her birisi farklı açılardan sorunun çözümüne yönelik öneriler sunmaktadır. Bugüne kadar yapılan çalışmalar, henüz videoların otomatik anotasyondan sorgulanmasına kadar tüm aşamalarda yeterli ve uygulanabilir çözümler üretilmiş değildir ve daha kapsamlı çalışmalara ihtiyaç vardır. Bu çalışma, literatürdeki eksiklikleri bir ölçüde kapatmaya yönelik bir çalışma niteliği taşımaktadır.

Bu proje kapsamında mimari yapısı Şekil-1’de gösterilen sistem bir prototip olarak geliştirilmiştir. Geliştirilen sistem, temel olarak iki alt sistemden oluşmaktadır. Bunlardan birincisi Anlamsal Bilgi Çıkarım Sistemi, ikincisi ise Depolama ve Sorgulama Sistemidir. Bu iki sistem arasındaki koordinasyon ve bilgi akışı Koordinatör olarak isimlendirilen modül tarafından sağlanmaktadır. Koordinatör modülünün bir diğer görevi ise kullanıcı ve sistem arasındaki arayüzü oluşturmaktır.

Anlamsal Bilgi Çıkarım Sisteminin görevi, ham bir videonun içerisinde yer alan anlamsal bilgilerin çıkarılması ve bu bilgilerin Depolama ve Sorgulama Sistemine aktarılmasıdır. Anlamsal bilgi çıkarımı, belli oranda tamamen otomatik olarak, el ile ve bu iki yönetim birlikte kullanıldığı hibrit yöntemle yapılabilmektedir. Bilgi çıkarımı için bir videonun içerdiği görsel, işitsel ve metinsel verilerin her üçü de kullanılmaktadır. Anlamsal bilgi çıkarımı, sisteme verilen bir videoda çekimlerin (shot) tespit edilmesiyle başlamaktadır. Bu noktadan sonra görsel, işitsel ve metinsel veriler ayrı ayrı işlemlere tabi tutulmaktadır ve çekimler çıkarılan bilgilerin hizalanmasında (alignment) temel birim olarak kullanılmaktadır. Görsel bilgi çıkarımında ilk adım olarak ana video çerçevelerindeki video nesnelerinin çıkarılması işlemi yapılmaktadır. Daha sonra nesnelere arasındaki uzamsal ve zamansal ilişkiler tespit edilmektedir. En son olarak da çıkarılan verilerin tamamı kullanılarak video içerisindeki olaylar çıkarılmaya çalışılmaktadır.

Videodan her çekim için çıkarılan ses verisi, öncelikle daha küçük segmentlere ayrılmakta ve öncelikle sessiz segmentler tespit edilmektedir. Daha sonra, ses içeren segmentler belirli akustik sınıflara ayrılmaktadır. Akustik sınıflar olarak başlangıçta ses sinyalleri, sessizlik

(silence), müzik (music), çevresel ses (environmental sound) ve saf konuşma (speech) gibi ana sınıflara ek olarak konuşma üzerine müzik (speech over music) ve konuşma üzerine çevresel ses (speech over environmental sound) gibi birleşik sınıflara ayrılmıştır. Daha sonra, çevresel sesler üzerinde çalışılarak bu seslerin uyarı, savaş, motor ve doğa sesleri gibi daha da ayrıntılı sınıflara ayrılması üzerinde çalışmalar yapılmıştır. Son bir iş olarak da akustik sınıflardan hareketle daha üst seviyede anlamsal bilgiler (semantic classes) çıkarılması üzerinde çalışılmıştır.



Şekil-1: Sistem mimarisi

Video metninden anlamsal bilgi çıkarımı ise temel olarak insan, kuruluş ve yer isimlerini içeren varlık isimlerinin (named entity) çıkarılması işlemi içermektedir. Ayrıca, zaman ve tarih bilgileri ile sayısal bilgiler de çıkarılabilmektedir. Metinden Anlamsal Bilgi Çıkarım Modülü, bilgi çıkarımı için bir kural tabanı (rule-base) ile bir öğrenme mekanizması içermektedir. Böylece sistem başlangıçta bir kurallar listesine sahip iken, zamanla yeni kurallar öğrenerek daha güçlü bir sistem haline gelmektedir. Öğrenme mekanizması olarak "Rote learning" yöntemi kullanılmıştır.

Bilgi Füzyonu Modülü, yukarıda açıklanan farklı üç modaliteden bilgi çıkaran modüllerin çıktılarını girdi olarak kullanan bir modüldür. Dolayısıyla, görsel, işitsel ve metinsel verilerden elde edilen bilgiler, bu modül tarafından analiz edilmekte ve birleştirilerek daha anlamlı ve

dođru verilerin üretilmesine çalışılmaktadır. Füzyon sonucunda elde edilen veri, veritabanında depolanmaya ve daha sonra sorgulanmaya hazır veri anlamına gelmektedir. Buradan çıkan veriler, Koordinatör vasıtasıyla Depolama ve Sorgulama sistemine aktararak kaydedilmektedir.

Sistemin ikinci büyük parçası ise Depolama ve Sorgulama Sistemidir. Bu alt sistem, video ile ilgili verilerin depolanması ve farklı kullanıcı sorgularına cevap verebilmesi için çeşitli modüllerden oluşmaktadır. Temel olarak altsistem içerisinde bir bulanık nesneye yönelik veritabanı ve bir bulanık bilgi sistemi mevcuttur. Bulanık nesneye yönelik veritabanı büyük miktardaki video verisinin depolanmasında kullanılmaktadır. Veritabanı, özellikle video verilerinin saklanmasıyla yönelik olarak modellenmiştir. Bulanık bilgi tabanı ise, video alanına (domain) uygun kuralların tanımlanması ve veritabanında doğrudan depolanmayan, fakat mevcut veriler kullanılarak oraya çıkarılabilen bilgiler için kullanılmaktadır. Hem veritabanı hem de bilgitabanı, belirsizlik ve bulanıklık içeren verileri de işleme yeteneğine sahiptir. Veritabanı ile bilgitabanı arasındaki iletişim Köprü olarak isimlendirilen bir arayüz vasıtasıyla sağlanmaktadır. Veritabanında saklanan verilere daha hızlı erişimi sağlayabilmek için hem anlamsal verileri hem de videonun altseviye özelliklerini birlikte kullanabilen bir dizin yapısı da geliştirilmiştir. Sorgulama esnasında ister anlamsal bilgilerle isterse altseviye özelliklerle sorgulama yapılsın, sistem her iki veri türünü de kullanarak hızlıca ilgili veriye ve oradan da videonun ilgili çekimine (shot) erişim imkanı sağlamaktadır.

Depolama ve Sorgulama Sistemi içerisinde ayrıca bir de XML veritabanı konulmuştur. Bir çok alanda olduğu gibi çoklu-ortam verilerinin de XML olarak tutulması, paylaşılması ve işlenmesi oldukça güncel bir konudur. Proje kapsamında, bir XML veritabanı üzerinde de çalışmalar ve denemeler yapılarak bu konuda da bir bilgi birikimi sağlanmıştır.

## **3. GEREÇ VE YÖNTEM**

### **3.1 Genel**

Bu bölümde, proje kapsamında yapılan çalışmalar, sağlanan ilerlemeler ve projede ulaşılan son durum daha ayrıntılı olarak açıklanmaktadır.

Proje Öneri Dokümanında yer alan proje önerisi çerçevesinde proje ihtiyaçları/gereksinimleri çıkarılmış ve ihtiyaçlar çerçevesinde de sistemin mimari yapısının Şekil-1'de verilen yapıda olması tasarlanmıştır. Mimari yapıdaki farklı altsistem ve modüller farklı iş paketleri kapsamında ele alınarak üzerinde çalışmalar yapılmıştır. Bundan sonraki alt bölümlerde sırasıyla her bir modül daha ayrıntılı olarak tanıtılmaktadır. Tanıtımlar aynı zamanda projedeki iş paketlerine uygun olarak yapılmıştır.

### **3.2 Görüntüden Anlamsal Bilgi Çıkarımı (2. İş Paketi)**

Görüntüden anlamsal bilgi çıkarımındaki amaç, sorgulama açısından ham veri barındıran videoları girdi olarak alıp, bunların içerisinde indekslenebilir ve sorgulanabilir anlamlandırılmış bilginin çıkarılmasıdır. Görüntüden bilgi çıkarımı için video çekimleri (shot) içerisindeki ana çerçeveler otomatik olarak tespit edilmekte, elde edilen ana çevrelerden de elle, otomatik ya da karma bir yöntemle nesnelere ve olaylar çıkarılmaktadır.

Görüntüden anlamsal bilgi çıkarımı konusundaki yaklaşımımız ve her bir alt iş adımıyla yapılanlar ile mevcut durumumuz, daha ayrıntılı olarak aşağıda alt başlıklar halinde verilmiştir.

### 3.2.1 Çekim (Shot) ve Ana Çerçeve (Keyframe) Çıkarımı

Bir videoda çekim (shot), kesintisiz olarak art arda gelen çerçevelerin dizisi olarak tanımlanabilir. Çekim, videonun zamansal boyutunu barındıran temel yapıdır. Video veritabanları video verilerinin organizasyonu için her videonun içeriğine göre ayrılmış çekim bileşenlerine ihtiyaç duyar. Her çekimin ana çerçeveleri de yine bu aşamada çıkarılır. Video çekimleri ve ana çerçeveleri daha sonraki modüllerde anlamsal bilgi çıkarımı için kullanılır.

Projede, Edge Change Ratio [1] yaklaşımı kullanarak videonun çekim sınırlarını belirleyen ve videoyu çekimlere ayıran bir araç geliştirilmiş ve sisteme entegre edilerek kullanılmıştır. Çekimlerin ilk ve son çerçeveleri ile arasında kalan I-Frame'ler ise ana çerçeveler olarak seçilmiş ve nesne çıkarımında kullanılmıştır.

### 3.2.2 Çekim ve Ana Çerçevelerden Nesne ve Olay Çıkarımı

Nesne ve Olay çıkarımında üç alternatif yöntem uygulanabilir: otomatik çıkarım, elle çıkarım, karma çıkarım. Otomatik çıkarımda, videodaki anlamsal içerik herhangi bir insan müdahalesi ve katkısı olmadan otomatik olarak çıkarılır. Elle çıkarımda, video içeriğine bir kullanıcı tarafından karar verilir ve sunulan bir kullanıcı arayüzü yardımıyla içeriğin sisteme girişi sağlanır. Karma yöntemde ise, otomatik ve elle çıkarım yöntemleri aynı anda ya da sırayla birbirlerini tamamlayacak şekilde kullanılır. Genellikle, öncelikle otomatik çıkarım yapılır, daha sonra sonuçlar kullanıcı tarafından kontrol edilir ve gerekli durumlarda düzeltilerek ve ilave bilgiler eklenerek sisteme girişi sağlanır.

### 3.2.3 Elle Çıkarım

Elle çıkarım uzun süren bir işlem olmasına rağmen günümüz teknolojileri ile kullanılması kaçınılmazdır. İnsan müdahalesi olmaksızın bir videonun tüm içeriğinin çıkarılması şu an için mümkün değildir. Bu sebeple, önerilen model bir elle çıkarım modülüne de sahiptir. Bu modül kapsamında video içeriğinin belirlenmesini ve veritabanına girişini kolaylaştırmak için özel bir kullanıcı arayüzü tasarlanmış ve gerçekleştirilmiştir. Bu kullanıcı arayüzü aracılığıyla;

- Kullanıcı, videoyu oynatabilir, çerçeve çerçeve ileri ve geri gidebilir, istediği herhangi bir çerçeve için açıklama girebilir.
- Bir nesnenin bölgesini fare ile seçtiği zaman, koordinat verisi sistem tarafından otomatik olarak alınarak veritabanına kaydedilebilir.
- Nesne ve ilgili video çerçevesi arasındaki bağlantı bu işlem sırasında otomatik olarak kurulur.

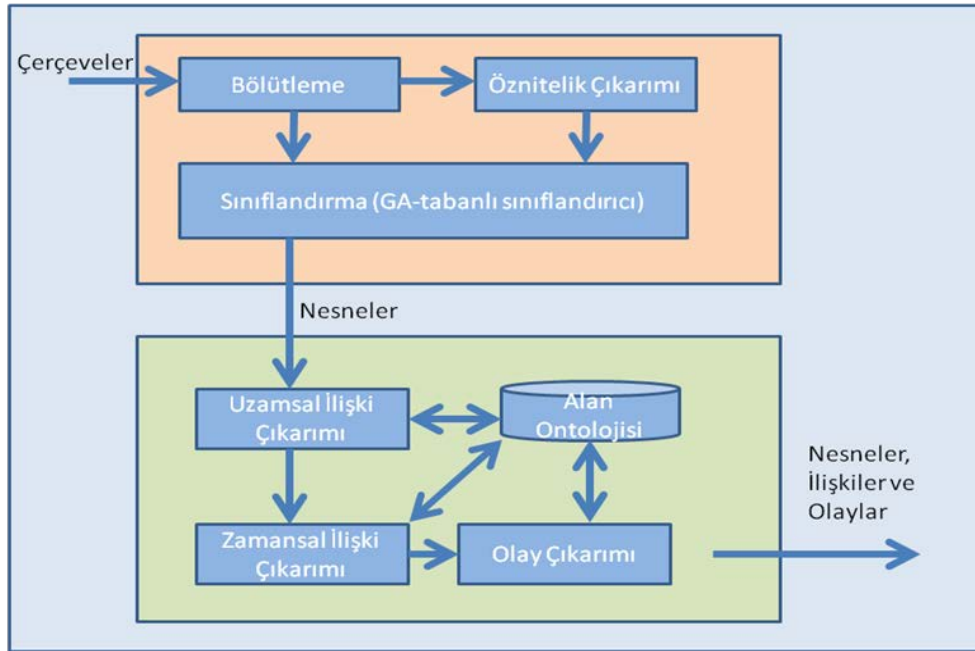
### 3.2.4 Otomatik Çıkarım

Proje kapsamında, videodan otomatik anlamsal bilgi çıkarımı işlemini gerçekleştirmek için bir "Otomatik Anlamsal İçerik Çıkarım Sistemi (Automatic Semantic Content Extraction Framework - ASCEF)" ortaya konmuştur. Şekil-2'de de gösterildiği gibi, ASCEF'te çıkarım işlemi aşağıdaki alt adımlardan oluşmaktadır:

- Nesne çıkarımı
- Uzamsal ilişki çıkarımı
- Zamansal ilişki çıkarımı

- Olay çıkarımı

Otomatik çıkarım işlemi videodan seçilmiş temsili çerçeveleri (ana çerçeveleri) girdi olarak kabul eder. Her çerçeveden öncelikle, o çerçevede bulunan nesnelere çıkarılmaya çalışılır. Nesne çıkarımı için Genetik Algoritma (GA) tabanlı, çoklu sınıflandırma yaparak ve sınıflandırma sonuçlarını bulanık olarak tespit ederek bulanıklılığı destekleyen bir yöntem kullanılmıştır. Nesnelerin arasındaki uzamsal ilişkilerin tespiti için ilgili çerçevedeki nesnelerin konumsal bilgileri kullanılır. Ardışık çerçevelerdeki nesnelerin uzamsal özelliklerindeki değişiklikler tespit edilerek de zamansal ilişkilerin çıkarımı sağlanır. Çıkarılan nesnelere, uzamsal ve zamansal ilişkiler yardımıyla da olay çıkarımı sağlanır. Nesne tanımları GA-tabanlı sınıflandırıcı içinde yapılır. Uzamsal ve zamansal ilişkiler ile olaylara ait tanımlar ise, bir alan ontolojisi içinde tanımlanır.



Şekil-2: Otomatik anlamsal içerik çıkarım sistemi

### 3.2.5 Nesne Çıkarımı

Otomatik bilgi çıkarımının nesne çıkarımına ait kısmı, Şekil-2'nin üst kısmında gösterilmiştir. Çalışmamızın nesne çıkarımı ihtiyacını karşılamak için yarı-otomatik Genetik Algoritma tabanlı bir yöntem kullanılmaktadır. Bu yöntem, öğreticiyle öğrenmeye dayalı (güdümlü) bir yöntemdir ve sistemin çalışıp doğru sonuçlar verebilmesi için, sistem öncelikle ilgi alanımız dahilindeki nesne tiplerinin öznitelikleri kullanılarak eğitilir. Nesne çıkarımı için gerçekleştirilen alt iş adımları aşağıda verilmiştir:

#### Bölütleme

Çerçeve imgelerden aday nesnelerin tespit edilmesi için, öncelikle, literatürde sıkça kullanılan bir bölütleme yöntemi olan "Normalized Cut (N-Cut) Image Segmentation [2]" yöntemi kullanılmıştır. N-Cut yöntemi imgelerden aşırı-bölütlenmiş sonuçlar vermektedir. Bu doğrultuda, çıktı olarak alınan bölütler ve komşu bölütlerle oluşturulan tüm kombinasyonları sınıflandırma sistemine aday nesne girdisi verilir. Uygulaması yapılan N-Cut bölütlemeye ilgili bir örnek Şekil-3'de görülmektedir.



Şekil-3: N-Cut bölütleme örneği

### Öznitelik Çıkarımı

Bölütleme sonrasında çerçeve içerisindeki aday-nesnelere tespit edildikten sonra bunların öznitelikleri çıkarılır ve sınıflandırma için hazır hale gelir. Sistemimizde, imgelerin öznitelikleri olarak MPEG-7 öznitelik tanımlayıcıları kullanılmaktadır. İmgelerin özniteliklerinin çıkarımı için ise, MPEG'in resmi yazılımı olan "eXperimentation Model (MPEG-7 Reference Software) [3]" (XM Yazılımı) kullanılmaktadır. Mevcut durumda, XM yazılımının sistemimizin ihtiyaçlarına göre konfigüre edilmesi ve sistemimiz içerisinde entegre olarak çalışması için gerekli ara bağdaştırıcı yazılımın gerçekleştirimi yapılmıştır ve XM yazılımı sistemimiz içerisinde kullanılabilir durumdadır.

### Sınıflandırma

Öznitelikleri çıkarılan imgelerin sınıflandırılmasında kullanılan Genetik Algoritma tabanlı sınıflandırma yönteminde, her nesne tipi (sınıfı), o sınıfın "En İyi Temsil Eden ve Ayıran Öznitelikler – Best Representative and Discriminative Feature (BRDF)[4]"i kullanılarak tanımlanmaktadır. Diğer bir deyişle, sınıflandırma algoritması, öznitelik seçiminde her sınıfa ait özel olarak tanımlanmış BRDF vektörlerini kullanmakta ve sınıflandırma bu özniteliklerin değerlerindeki farklılıklara göre yapılmaktadır.

### Testler

Gerçekleştirimi yapılan modüller için zaman ölçümü ve doğruluk oranları için çeşitli testler gerçekleştirilmiştir. Elde edilen bazı test sonuçları aşağıdaki tablolarda (Tablo-1-3) özet olarak verilmiştir. Bu sonuçlar elde edilen ilk sonuçlar olup, gerek sistemde yapılacak iyileştirmelerle gerekse sistemin daha iyi eğitilmesiyle bu oranların daha da iyileşeceği değerlendirilmektedir.

Tablo-1: N-Cut bölütleme yöntemi ile bölüt sayısına göre bölütleme süreleri

Bölüt Sayısı	Bölütleme Süresi (s)	Aday Nesne Oluşturma Süresi (s)	Aday Nesne Sayısı
5	15	7	28
8	20	42	209
10	22	143	729
15	41	2986	27.762
20	56	106.578	826.684

Tablo-2: Örnek bir videodan nesne çıkarımı süreleri

İşlem	Toplam Süre (s)	Elde Edilen Parça Sayısı	Parça Başına Ortalama Süre (s)
Ana Çerçeve Çıkarımı	32	1	32
Bölütleme	1.152	48	24
Aday Nesne Oluşturumu	6.816	48	142
Sınıflandırma	864.600	34.584	25
TOPLAM	872.600		

Tablo-3: Nesne çıkarımı doğruluk oranları

	Eşik Değerleri								
	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90
Precision	0,01	0,01	0,02	0,04	0,13	0,37	0,64	0,72	0,74
Recall	1,00	1,00	1,00	1,00	0,97	0,83	0,59	0,37	0,11
F-Measure	0,03	0,03	0,04	0,08	0,23	0,51	0,62	0,49	0,19

### 3.2.6 Anlamsal İçerik Modeli (Video Semantic Content Model - VISCOM)

Olay çıkarım işlemlerini gerçekleştirebilmek için videodaki anlamsal bilginin modellenmesi gerekmektedir. İçerisinde nesnelere, nesnelere birbirleriyle olan uzamsal ve zamansal ilişkilerini ve olayları içeren karmaşık bir yapının modellenmesi ancak gelişmiş bir modelleme aracıyla sağlanabilir. Bu doğrultuda, modelleme işleminin bir ontoloji yapısı içinde gerçekleştirilmesine karar verilmiştir. Aslında oluşturulan ontolojik yapı, bize sistematik bir kural tabanlı sistem sağlamaktadır. Olay çıkarımı için gerekli kurallar, belirlenen ontoloji yapısı içerisinde tanımlanmaktadır. Fakat, belirlenmiş uygulama alanları için bile olsa, belli başlı bazı standartlar ortaya koymadan bir kurallar kümesi ve alan ontolojisi geliştirmek, zorlu, sıkıcı ve ölçülenebilir bir işlem değildir.

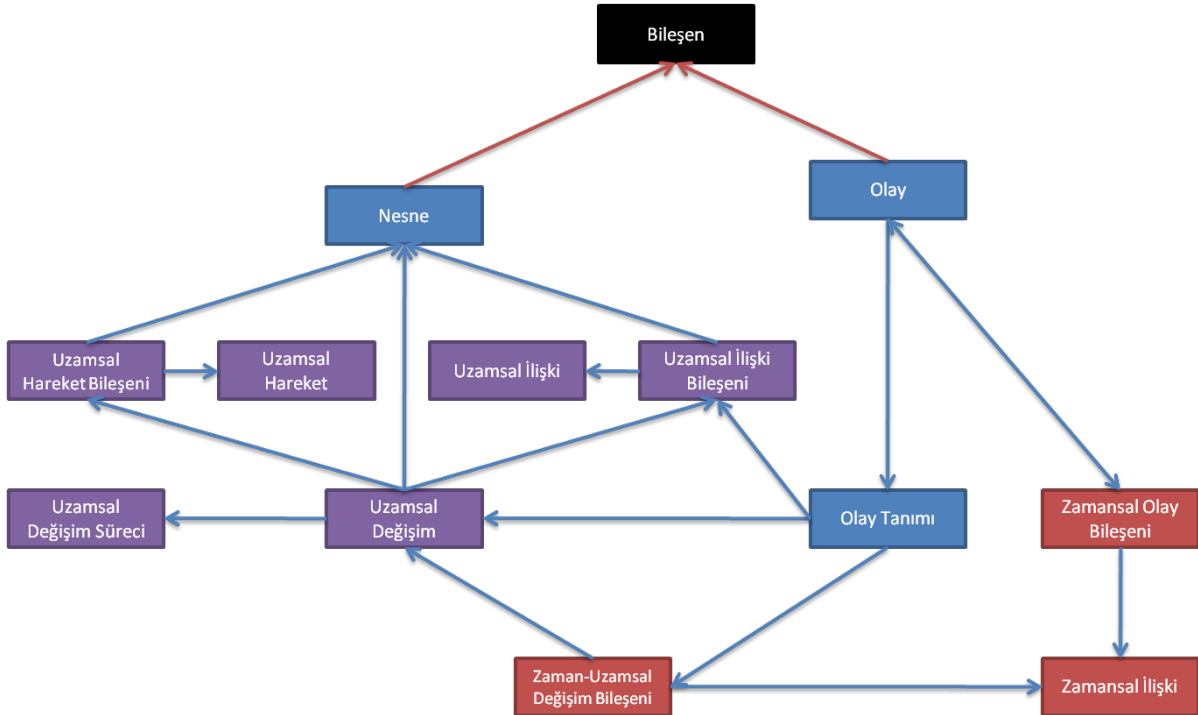
Bu doğrultuda öncelikle bir ontoloji tanımlama standardı (meta-ontoloji) ortaya konmuştur. Anlamsal İçerik Modeli (Video Semantic Content Model - VISCOM)" ismi verilen bu standart uygulama alanlarından bağımsız olarak tanımlanmıştır ve her uygulama alanı için ihtiyaç duyulan alan ontolojileri, bu standart yardımıyla oluşturulmaktadır. ASCEF'de, oluşturulan bu alan ontolojilerini kullanarak olay çıkarımı işlemini gerçekleştirmektedir.

VISCOM, belirli sınıflar ve bu sınıflar arasındaki ilişkilerden oluşmaktadır. Bu sınıflardan bazıları hedef bilgi tiplerimiz olan Nesne, Olay gibi anlamsal içerik tiplerini temsil ederken, diğerleri otomatik anlamsal çıkarım sürecinde kullanılan ara bilgi tipleridir. VISCOM'un barındırdığı sınıflar şunlardır:

- Bileşen: Nesne ve Olay'ın türediği temel sınıftır.
- Nesne: Tüm çıkarım işleminin üzerine kurgulandığı temel sınıf tipidir. Zamansal ve uzamsal bir değişiklik boyutu içermez.
- Olay: Belli bir süreci ifade eden, nesnelere ve nesnelere ilişkilerinden oluşan sınıf tipidir. Olay'lar Olay Tanımı veya Zamansal Olay Bileşeni sınıflarından elemanlar kullanılarak tanımlanır.
- Olay Tanımı: Bir olayın birden fazla oluş şekillerinin her birini ifade eder. Uzamsal veya zamansal-uzamsal değişiklikler ile tanımlanır. Uzamsal İlişki Bileşeni, Uzamsal

Değişim veya Zaman-Uzamsal Değişim Bileşeni sınıflarından elemanlar kullanılarak tanımlanır.

- Zamansal Olay Bileşeni: Olaylar arasında zamansal olarak ilişki tanımlamak için kullanılır. Olay ve Zamansal Olay Bileşeni arasında çift taraflı kullanım ilişkisi bulunur, her ikisini tanımlamak için de bir diğeri kullanılabilir.
- Zamansal İlişki: Zamansal Olay Bileşeni içerisinde Olay elemanlarını veya Zamansal-Uzamsal Değişim Bileşeni içerisindeki Uzamsal Değişim bileşenlerini zamansal açıdan ilişkilendirme (önce, sonra, vs.) için kullanılır.
- Zamansal-Uzamsal Değişim Bileşeni: Uzamsal Değişim elemanlarının zaman içerisindeki değişikliklerini tanımlamak için kullanılır.
- Uzamsal İlişki: Uzamsal İlişki Bileşeni sınıfında elemanlar tanımlarken, Nesnelerin birbirleriyle olan uzamsal ilişkilerini ifade etmek için kullanılır. Aşağıda “Uzamsal (Spatial) İlişki Çıkarımı” bölümünde ayrıntılı olarak verilen ilişki tiplerini içerir.
- Uzamsal Hareket: Uzamsal Hareket Bileşeni sınıfında elemanlar tanımlarken, Nesne’lerin kendilerine göre olan uzamsal değişikliklerini ifade etmek için kullanılır. Yukarı, aşağı, sağa, sola hareket ve hareketsiz gibi tipleri tanımlanmıştır.
- Uzamsal Değişim Süreci: Uzamsal Değişim sınıfında elemanlar tanımlarken, değişimlerin arasında ilişkiyi ifade etmek için kullanılır. Başlangıçtan-başlangıca, başlangıçtan-bitişe, bitişten-başlangıca, bitişten-bitişe değerlerini alabilir.
- Uzamsal Değişim: Uzamsal değişiklikleri tanımlamak için kullanılan ana elemandır. Nesne, Uzamsal Hareket Bileşeni, Uzamsal İlişki Bileşeni ve Uzamsal Değişim Süreci sınıfındaki elemanlar kullanılarak tanımlanır.
- Uzamsal Hareket Bileşeni: Nesnenin kendisine göre olan değişikliklerini tanımlamak için kullanılır. Nesne ve Uzamsal Hareket sınıfındaki elemanlar kullanılarak tanımlanır.
- Uzamsal İlişki Bileşeni: Nesnenin diğer nesnelere göre olan değişikliklerini tanımlamak için kullanılır. Nesne ve Uzamsal İlişki sınıfındaki elemanlar kullanılarak tanımlanır.



Şekil-4: VISCOM sınıfları ve ilişkileri



VISCOM'un sınıfları arasındaki ilişkiler, diğer bir deyişle ontolojik açıdan bağımlılıkları Şekil-4'te gösterilmiştir. Oluşturulan VISCOM meta-ontolojisinde sınıflar, ilişkili diğer sınıfları kullanarak tanımlanmaktadır. Örneğin; Olay sınıfı, Olay Tanımı sınıflarından oluşmaktadır. Her Olay Tanımı sınıfı ise Uzamsal İlişki Bileşeni ve Uzamsal Değişim barındırmaktadır. Bu tanımlar bu şekilde devam etmekte Olay ve Nesne ilişkilendirilmektedir. Oluşturulan meta-ontolojiyi tanımlamak için OWL kullanılmıştır.

Herhangi bir uygulama alanı için alan ontolojisi tanımlanırken, VISCOM içerisindeki tüm elemanların uygun şekilde tanımlanması gerekir. Bu doğrultuda öncelikle Nesne tanımlarından başlanarak Şekil-4'te verilen bağımlılıklara göre adım adım tüm ontoloji oluşturulur. Proje kapsamında oluşturulan sistemi test etmek için, Futbol, Basketbol ve Ofis Güvenliği alanlarında Alan Ontolojileri oluşturulmuştur.

### 3.2.7 Uzamsal (Spatial) İlişki Çıkarımı

Nesne çıkarımı ile çıkarılan nesnelere uzamsal ilişki çıkarımında girdi olarak kullanılır. Uzamsal çıkarım esnasında nesnelere gerçek şekilleriyle ele almak çıkarım işlemi çok karmaşık hale getireceğinden, nesnelere En Küçük Kapsayan Dikdörtgen (Minimum Bounding Rectangle) ile temsil edilmekte ve her nesne dikdörtgen şeklinde modellenmektedir. Uzamsal ilişkiler bulanık olarak ele alınmakta ve her bir ilişki için üyelik değerleri nesnelere birbirlerine göre olan göreceli konumları kullanılarak hesaplanmaktadır. Bu çalışma kapsamında üç tip uzamsal ilişki tanımlanmaktadır:

1. Konumsal İlişki: Üstünde, Altında, Solunda, Sağında
2. Uzaklık İlişkisi: Yakın, Uzak
3. Topolojik İlişki: İçinde, Kısmen içinde, Dokunan, Ayrık

Bulanık olarak ele alınan her ilişki tipi için üyelik değerlerini tanımlayan üyelik fonksiyonları tanımlanmaktadır. Örneğin,  $o_1$  ve  $o_2$  nesnelere arasındaki 'Sağında' konumsal ilişkisinin üyelik değerini hesaplamak için kullanılan üyelik fonksiyonu aşağıdaki gibidir:

$$\mu_{saginda}(o_1, o_2) = \begin{cases} \sin(\Phi + 90), & 0 < \Phi < 90 \vee 270 < \Phi < 360 \\ 0, & diger.durumlarda \end{cases}$$

$\Phi$  :  $o_1$  ve  $o_2$  nesnelere En Küçük Kapsayan Dikdörtgenlerinin merkezlerini birleştiren doğrunun açısı.

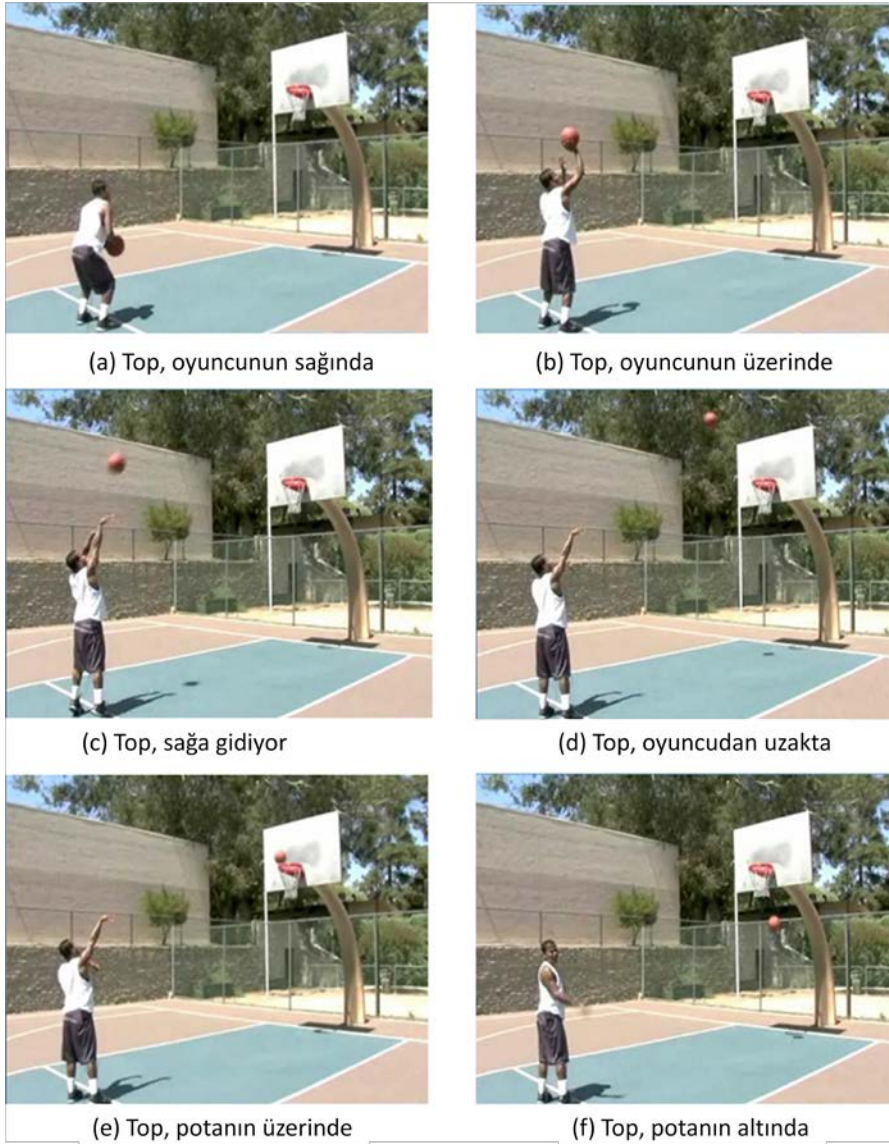
### 3.2.8 Zamansal (Temporal) İlişki Çıkarımı

Zamansal ilişkilerse uzamsal ilişki dizilerine zamansallık eklemek için tanımlanmaktadır. Zamansal ilişkileri tanımlamak için literatürde yaygın olarak kullanılan "Allen's Temporal Interval Algebra [5]" yöntemi tercih edilmiştir. Bu yöntem ile olayların zamansal aralıkları dikkate alınarak, birbirleriyle ilişkileri tanımlanabilmektedir. Allen'in tanımları 13 adet zamansal ilişki barındırmaktadır: Önce (Before), Kapsayan (Overlaps), Boyunca (During), Çakışan (Meets), Başlangıç (Starts), Bitiş (Finishes), Eşit (Equal) ve Eşit haricindekilerin tersleri.

### 3.2.9 Olay Çıkarımı

Olay çıkarımı için yukarıda detayları verilen nesnelere ve nesnelere ait uzamsal ilişkiler kullanılmaktadır. Olay çıkarımı için temel olarak 3 yapı daha mevcuttur:

- Nesne hareketleri: Ardışık çerçevelerde nesnelerin konumları takip edilir ve bir hareket karakteri izleyip izlemedikleri tespit edilir.
- Uzamsal değişiklikler: Ardışık çerçevelerde değişiklik gösteren iki nesne arasındaki uzamsal ilişkiler tespit edilir. Örneğin; bir çerçevede top oyuncunun sağında iken, daha sonraki bir çerçevede top oyuncunun üzerinde ise, top ve oyuncu arasındaki uzamsal ilişkide bir uzamsal değişiklik tespit edilmiş demektir. Uzamsal değişikliklerin çıkarımı için nesne hareketleri ve uzamsal ilişkiler ile uzamsal değişiklik tanımının yer aldığı alan ontolojisi kullanılır.
- Zamansal-uzamsal değişiklikler: Bu değişikliklerse bir olaydaki uzamsal değişikliklerin zamansal sıralamasını tanımlar. Uzamsal değişikliklere benzer şekilde, zamansal-uzamsal değişiklikler, alan ontolojisindeki tanımlar kullanılarak çıkarılır.



Şekil-5: Örnek bir olay

Olay çıkarımı sürecini daha güzel anlayabilmek için, Şekil-5'de verilen bir basketbol videosundan alınmış serbest atış ana çerçevelerini inceleyebiliriz. Olay çıkarımı safhasına geçmeden önce çerçevelerimizde geçen nesnelere (oyuncu, top, serbest atış çizgisi ve pota) çıkarıldığını varsayıyoruz. İlk çerçevede (Şekil 5-a), iki önemli uzamsal ilişki bulunmaktadır: Oyuncu, serbest atış çizgisinin yakınındadır ve top, oyuncunun sağındadır. İkinci çerçevede (Şekil 5-b), topun oyuncunun üzerinde olduğuna dair bir uzamsal ilişki çıkarılmaktadır. Bu durumda, ilk iki çerçeve arasında top ve oyuncu nesnelere için bir uzamsal değişiklik ortaya çıkar. Üçüncü çerçevede (Şekil 5-c), oyuncunun topu attığı ve topun sağa doğru hareket ettiği gözükmektedir. Dördüncü çerçevede (Şekil 5-d), bu durum daha açık bir şekilde gözükmektedir ve top daha sağa hareket etmiş olduğundan bir nesne hareketi (uzamsal hareket) çıkarılmaktadır. Beşinci çerçevede (Şekil 5-e), top potanın üzerindedir ve altıncı çerçevede de (Şekil 5-f) top potanın altındadır. Birbirini izleyen bu iki uzamsal ilişki yeni bir uzamsal değişiklik çıkarılmasını sağlamaktadır. Böylece tüm çerçevelere sırayla baktığımızda, iki uzamsal değişiklik bulunmaktadır. Bu iki uzamsal değişiklik arasındaki zamansal ilişkiyi inceleyerek, buradan bir zamansal-uzamsal ilişki çıkarılmasını sağlayabiliriz. Çıkarılan bu zamansal-uzamsal ilişki de serbest atış olayının tanımı olarak alan ontolojisinde yer alır.

## Test Sonuçları

Otomatik olay çıkarımını test etmek için öncelikle futbol, basketbol ve ofis güvenliği alanlarında alan ontolojileri oluşturulmuştur. Bu alanların her biri için testler gerçekleştirilmiştir.

Olay çıkarımının başarısını ölçmek için Precision, Recall, Precint, Recint ve Boundary Detection Accuracy (BDA) değerleri hesaplanmıştır. Precint, Recint ve BDA değerleri, hem olayın doğru şekilde bulunmasını hem de bulunan olayın başlangıç ve bitiş zamanlarının doğru olarak tespit edilmesini dikkate alan metriklerdir. Bu metrikler aşağıdaki formüllere göre hesaplanmaktadır:

$$Prec_{int} = \frac{\tau_{mb} \cap \tau_{db}}{\tau_{db}}$$

$$Rec_{int} = \frac{\tau_{mb} \cap \tau_{db}}{\tau_{mb}}$$

$$BDA = \frac{\tau_{mb} \cap \tau_{db}}{\max(\tau_{mb}, \tau_{db})}$$

Burada  $T_{db}$  ve  $T_{mb}$  değişkenleri sırasıyla otomatik olarak tespit edilen olay aralığı (başlangıç bitiş zamanı) ve gerçek (ground-truth) olay aralığını ifade etmektedir.

İlk test, ofis güvenlik kamera kayıtlarından oluşan 5 adet video ile yapılmıştır. Videolardan her biri yaklaşık 10 dakika uzunluktadır. Videolardan toplam 1026 ana çerçeve görüntü çıkarılmıştır. Çıkarımı yapılan olaylar ve çıkarım başarımları oranları Tablo-4'de özetlenmiştir.

Tablo-4: Ofis güvenliği olay çıkarımı test sonuçları

Olay	Precision(%)	Recall(%)	Prec <sub>int</sub> (%)	Rec <sub>int</sub> (%)	BDA(%)
Girmek	100.00	100.00	92.31	100.00	92.31
Volta atmak	75.00	100.00	83.97	100.00	83.97
Çıkmak	0.00	0.00	0.00	0.00	0.00
Yazıcıdan Çıktı Almak	0.00	0.00	0.00	0.00	0.00
Oturmak	100.00	83.33	87.44	69.17	86.33
Yazmak (Bilgisayarda)	80.00	80.00	65.80	72.03	65.23
Yürümek	88.46	100.00	77.23	98.91	76.91
Yazı Yazmak	100.00	100.00	100.00	96.77	96.77
Bırakmak	100.00	100.00	95.24	100.00	95.24
<b>TOPLAM</b>	<b>90.00</b>	<b>90.00</b>	<b>80.71</b>	<b>80.13</b>	<b>78.59</b>

İkinci test, her biri yaklaşık 2 dakikalık görüntü içeren 3 adet basketbol videosu ile gerçekleştirilmiştir. Videolardan toplam 207 ana çerçeve görüntü çıkarılmıştır. Çıkarımı yapılan olaylar ve çıkarım başarımları Tablo-5'de özetlenmiştir:

Tablo-5: Basketbol olay çıkarımı test sonuçları

Olay	Precision(%)	Recall(%)	Prec <sub>int</sub> (%)	Rec <sub>int</sub> (%)	BDA(%)
Ribaunt	50.00	50.00	62.50	55.56	62.50
Topa Zıplama	100.00	100.00	94.23	100.00	97.23
Serbest Atış	100.00	100.00	95.00	100.00	95.00
Hücum	100.00	100.00	94.65	100.00	94.65
<b>Total</b>	<b>87.50</b>	<b>87.50</b>	<b>88.42</b>	<b>92.36</b>	<b>89.34</b>

Son test, her biri yaklaşık 2 dakikalık görüntü içeren 5 adet futbol videosu ile gerçekleştirilmiştir. Videolardan toplam 312 ana çerçeve görüntü çıkarılmıştır. Çıkarımı yapılan olaylar ve çıkarım başarımları Tablo-6'da özetlenmiştir:

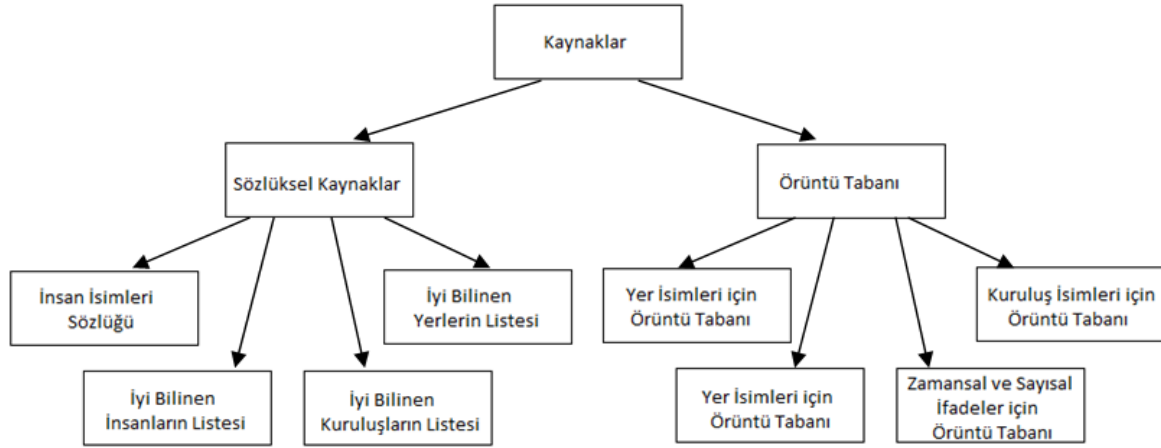
Tablo-6: Futbol olay çıkarımı test sonuçları

Olay	Precision(%)	Recall(%)	Prec <sub>int</sub> (%)	Rec <sub>int</sub> (%)	BDA(%)
Pas	87.50	70.00	82.50	75.52	74.05
Köşe Vuruşu	100.00	50.00	92.00	65.20	65.20
Gol	100.00	100.00	90.00	96.00	96.00
<b>Total</b>	<b>90.00</b>	<b>69.00</b>	<b>86.14</b>	<b>82.15</b>	<b>80.34</b>

### 3.3 Metinden Anlamsal Bilgi Çıkarımı (3. İş Paketi)

Metinden anlamsal bilgi çıkarımı kapsamında iki temel bilgi çıkarım türüne odaklanılmıştır: özel anlamlı kelimeler ve olaylar. Öncelikle, “özel anlamlı kelimelerin belirlenmesi (named entity recognition)” üzerinde çalışılmıştır. Özel anlamlı kelimeleri belirleme (named entity recognition), bilgi çıkarımı (information extraction) konusunun en önemli alt görevlerinden birisidir [6]. Özel anlamlı kelimeleri belirleme, metinlerdeki insan isimleri, yer adları, organizasyon isimleri ve hatta bazı zamansal ve sayısal ifadelerin tanınması şeklinde ifade edilebilir. Bu nedenle çalışmada, varlık ismi tanıma (named entity recognition) yerine “özel anlamlı kelime belirleme” teriminin kullanılması tercih edilmiştir. Proje kapsamında, videolardaki metinlerden özel anlamlı kelimeleri çıkarabilmek adına öncelikle Türkçe için kural tabanlı bir özel anlamlı kelime belirleyicisi geliştirilmiştir.

Türkçe için kural tabanlı Özel Anlamlı Kelime Belirleyicisi (ÖAKB), kural örüntülerini içeren örüntü tabanlarından (pattern bases) ve bazı sözlüksel kaynaklardan (lexical resources) faydalanmaktadır. Geliştirilen sistemin, öncelikli olarak Türkçe haber metinleri üzerinde çalışması düşünülmüş ve bu nedenle kullanılan kaynakların bir kısmı da doğrudan bu Alana uygun olarak seçilmiştir. Ancak, haber metinleri (projede özellikle haber videoları üzerine yoğunlaşmıştır) doğası gereği içinde birçok farklı alanı barındırmaktadır, dolayısıyla bu kaynaklardaki çoğu girdi oldukça geneldir. Kullanılan kaynakların sınıflandırılmış hali Şekil-6'da gösterilmiş ve aşağıda kısaca açıklanmıştır.



Şekil 6: Türkçe için özel anlamlı kelime belirleyicisi tarafından kullanılan kaynakların sınıflandırılması

1. **Sözlüksel Kaynaklar:** Yaklaşık 8300 girdi içeren Türkçe insan isimleri sözlüğü, iyi bilinen siyasetçilerin listesi, Türkiye ve dünyadaki iyi bilinen yerlerin (şehir ve kasaba isimleri) listesi, Türkiye'deki ve dünyadaki iyi bilinen kuruluş isimleri kural tabanlı ÖAKB sistemi tarafından kullanılan sözlüksel kaynakları oluşturmaktadır.
2. **Örüntü Tabanları:** Örüntü tabanları, yer isimlerinin örüntülerini (aşağıda (1) ile örneklendirilmiştir), kuruluş isimlerinin örüntülerini (aşağıda (2) ile örneklendirilmiştir), ve son olarak zaman/gün/para/yüzde ifadelerinin örüntülerini (aşağıda (3) ile örneklendirilmiştir) içermektedir. Haber metinleri üzerinde çalışan özel anlamlı kelime belirleme işlemi sırasında yararlanılan örüntüler, bazı haber yazıları incelenerek ve İngilizce gibi başka diller için yapılmış benzer çalışmalardan yararlanılarak manuel olarak derlenmiştir. Kullanılan örüntülerin boyutu konusunda genel bir fikir vermek gerekirse; yer isimlerinin çıkarımı için kullanılan örüntülerin yaklaşık 90 tanesi (1)'de ve haber metinlerinden çıkarılan kuruluş isimleri için kullanılan örüntülerin 50 tanesi

de (2)'de yer alan örnekteki gibidir. Yine de bu örüntü tabanları, ilgili tabanlara yeni örüntüler ekleyerek ya da var olan örüntüleri değiştirerek kolayca geliştirilebilir.

- (1) X Köprüsü/Sarayı/Mezarlığı: X ya sözlüksel kaynaklar kullanılarak bulunan bir özel anlamlı kelimedir ya da uygun bir özel anlamlı kelime bulunmadığında direkt olarak bir ön belirtkedir (preceding token)
- (2) X Vakfı/Enstitüsü/Federasyonu: X (1) deki gibidir.
- (3) X lira/dolar/avro...: para ifadesi için X bir sayısal ifadedir

Sistem daha da geliştirilerek, ezberden öğrenme (rote learning) tekniği ile daha önce yorumlanmış metinlerden öğrendikleri ile sözlüksel kaynaklarını genişletme yeteneğine sahip hale getirilmiştir. Ezberden öğrenme, bilgi çıkarımı konusunda daha da ilerlemek için üzerinde çalışılmış ve sonuç olarak yüksek doğruluk oranlarına (precision rates) eriştiği görülmüş öğrenme yaklaşımlarından biridir. Freitag'ın çalışmasında [7] açıklandığı gibi; bilgi çıkarımı için kullanılan bir ezberden öğrenme yaklaşımı kısaca daha önce yorumlanmış girdilerin çıkarımını ve gruplanmasını sağlar. Bunu yaparken iki tane istatistiksel öznelikten faydalanır. Bunlar; girdinin görünme sıklığı olan  $n$  ve yorumlanmış girdilerin görünme sıklığı olan  $p$ 'dir. Bunun sonucu olarak her yorumlanmış girdi için  $p/n$ , o girdiye ait güven değeri (confidence value) olarak kullanılabilir.

Kural tabanlı belirleyicinin, ezberden öğrenen bir tamamlayıcı parça ve güven değeri 0.5'in üzerinde olan girdilerle bilgi kaynaklarını genişletme yeteneği ile donatılması sonucunda bir melez belirleyici ortaya çıkmıştır. Verinin türü sağlansa ya da sağlanmasa da yeni oluşturan bu melez belirleyici çalıştırılabilmektedir. Eğer metin girdisinin türü kullanıcı tarafında girilmişse, belirleyici önce o tür için öğrenilmiş sözlüksel kaynakları kullanmaktadır, ardından ilk sözlüksel kaynakları ve örüntü tabanlarını kullanmaktadır. Eğer metin girdisinin türü belirtilmemişse, melez belirleyici önce aynı kural tabanlı belirleyici gibi çalışır, daha sonra sonuçta oluşan metni yorumlamak için tüm öğrenilmiş kaynaklarını kullanır.

Metinden anlamsal bilgi çıkarımı modülünün testi için, birisi TRT ve diğer ikisi NTV haber kanallarından elde edilmiş ve Tablo-7'de özetlenmiş üç adet veri seti kullanılmıştır. Bu veri setleri hem kural tabanlı hem de melez belirleyicinin performans ölçümlerinde kullanılmış ve performanslar genelde bu tür sistemlerin performans ölçümlerinde yaygın olarak kullanılan precision, recall ve f-measure değerleriyle ölçülmüştür.

Tablo 7: Veri seti ile ilgili özet bilgi.

Veri seti	Süresi	Haber hikaye sayısı	Metindeki kelime sayısı	Özel anlamlı kelime sayısı
Video veri seti-1	4 saat	340	20940	2534
Video veri seti-2	1.5 saat	182	12043	1408
Video veri seti-3	1 saat	29	1504	180

Tablo-8: Kural tabanlı belirleyicinin test sonuçları

Veri seti	Precision	Recall	F-Measure
Video veri seti-1	86.24 %	86.77 %	86.51 %
Video veri seti-2 (otomatik çıkarım)	83.92 %	80.20 %	82.02 %
Video veri seti-2 (elle hatasız çıkarım)	85.42 %	87.65 %	86.52 %
Video veri seti-3	86.78 %	86.78 %	86.78 %

Kural tabanlı belirleyicinin performans ölçüm sonuçları Tablo-8’de özetlenmiştir. Melez belirleyicinin performans ölçümleri için ise veri-seti-1 eğitim amacıyla kullanılırken, diğer veri setleri test amacıyla kullanılmıştır. Melez belirleyicinin test sonuçları Tablo-9’da görülmektedir. Kural tabanlı belirleyici ile melez belirleyicinin sonuçları karşılaştırıldığında, melez belirleyicinin kayan yazıdan otomatik çıkarılan veri setinde %1 oranında, elle çıkarılan veri setinde ise %4 civarında daha iyi sonuç verdiği görülmüştür. Otomatik çıkarımda zaten hatalı çıkarımlar olduğu için başarı oranı daha düşük kalmıştır.

Tablo-9:Melez belirleyicinin test sonuçları

Veri seti	Precision	Recall	F-Measure
Video veri seti-2 (otomatik çıkarım)	83.50 %	82.07 %	82.78 %
Video veri seti-2 (elle hatasız çıkarım)	84.98 %	91.10 %	87.93 %

Metinden anlamsal bilgi çıkarımı kapsamında ikinci olarak “olay (event)” çıkarımı üzerinde çalışılmıştır. Haber videolarından olay çıkarımı için, anahtar kelime tabanlı bir yaklaşım kullanılmıştır. Anahtar kelime tabanlı olay çıkarıcı önce, düşünülen olay için sık tekrarlanan anahtar kelimeler kullanılarak eğitilmektedir. Daha sonra olay çıkarıcı, verilen girdi metindeki kelimelere bakarak, tanımlı olayların metin içinde geçip geçmediğine karar vermektedir. Buradan da işlenen videoda geçen olaylar konusunda bilgiye ulaşılabilmektedir. Ya da videonun hangi olayla ilgili olduğu sonucuna ulaşılabilmektedir. Yapılan çalışmada, eğitim veri kümesi olarak, Türkiye Radyo ve Televizyon Kurumu (TRT) tarafından yayınlanmış, toplamda yaklaşık dört saat ve 340 farklı haber hikayesi içeren 35 haber videosu değerlendirilerek genel olay anahtar kelimeleri tespit edilmiştir. Bu haber yazılarında toplamda 463 olay çıkarılmıştır. Aynı olaylar birleştirildiğinde ise 69 farklı olay kalmıştır. Bu işlemden sonra, aynı olayla ilişkilendirilmiş yazı metinleri bir araya getirilerek gruplandırılmıştır (tek bir hikaye metni birden fazla olay taşıyabilir ve birçok gruba dahil olabilir). Olay çıkarıcımız tarafından en çok tekrarlanan 10 olay; “İfade, Ölüm, Duruşma/Soruşturma, Kaza, Hava Durumu, Toplantı, Saldırı, Yaralanma, Seçim ve Operasyon” olarak belirlenmiştir. Bu 10 olayın, toplam tekrarlanma sayısının 303 olduğu ve bu 10 olayın tüm farklı olayların yaklaşık %15 (10/69)’ine, eğitim verileri kümesindeki tüm kodlanmış olayların da %65.4 (303/463)’üne tekabül ettiği tespit edilmiştir.

Sık rastlanan anahtar kelime çıkarma işlemi yukarıdaki olay tiplerine uygun düşen her bir metin grubu için aşağıdaki gibi yapılmaktadır:

- Metin öncelikle “özel anlamlı kelimelerin belirlenmesi” işleminden geçirilerek çıkarılan özel anlamlı kelimeler elenir. Kalan kelimeler arasından yok sayılabilecek kelimeler (“ve”, “ile”, “de”, “ki” gibi), meslek ve ülke adları da elenir.
- Metindeki her bir kelime için, görülme sayısı (kelime frekansı, tf) hesaplanır.
- En çok tekrarlanan 10 kelime (en yüksek 10 tf değeri) incelenen olaya karşılık gelen anahtar kelimeler olarak tutulur. Olay tipinin (e) her bir anahtar kelimesi (k) için bir güven değeri (c (confidence value)) aşağıdaki gibi hesaplanır. Formüldeki  $tf_{i,e}$  e olayının metni içindeki i’nci anahtar kelimenin frekansdır.

$$c_{k,e} = \frac{tf_{k,e}}{\sum_{i=0}^{10} tf_{i,e}}$$

Bu işlemle, yukarıda bahsedilen 10 olay için ilgili güven değerleriyle birlikte olay anahtar kelimelerine karar verilir. Daha sonra otomatik olay çıkarım işlemi, bu anahtar kelimeleri kullanarak haber metninde ilgili olayın olup olmadığına karar vermektedir. Otomatik çıkarım esnasında, eğer o kelimeyle karşılaşırsa, o anahtar kelimenin güven değeri metnin olaya özgü güven skoruna eklenir.

Türkçe haber metinleri için geliştirmiş olduğumuz bu yöntem Zhang ve arkadaşlarının takım spor videoları için yaptığı çalışmaya benzetilmektedir [8]. Zhang ve arkadaşlarının yaptığı çalışmada Webcast metinleri önce Latent Semantic Analysis (LSA) kullanılarak kümelenecek ve her bir kümedeki (basketbolda ayırık olay tiplerine göre) anahtar kelimeler *tf-idf* ağırlıklarına göre sıralanmıştır. *tf-idf* ağırlığı genellikle bilgi çıkarımında girilen kullanıcı sorgusu cevaplarının sıralanması için kullanılır ve şu şekilde hesaplanır:

$$tf-idf_{i,j}=tf_{i,j} * idf_i$$

$tf_{i,j}$ ,  $d_j$  belgesindeki  $t_i$  teriminin tekrarlanma sayısını,  $idf_i=log(N/n_i)$  ve  $N$  dikkate alınan toplam belge sayısını,  $n_i$  ise  $t_i$  teriminin görüldüğü belge sayısını temsil etmektedir [9]. Her bir kümelenebilendeki en üstteki 4 anahtar kelime, basketboldaki olay tipleri için temsili anahtar kelimeler olarak kullanılmaktadır.

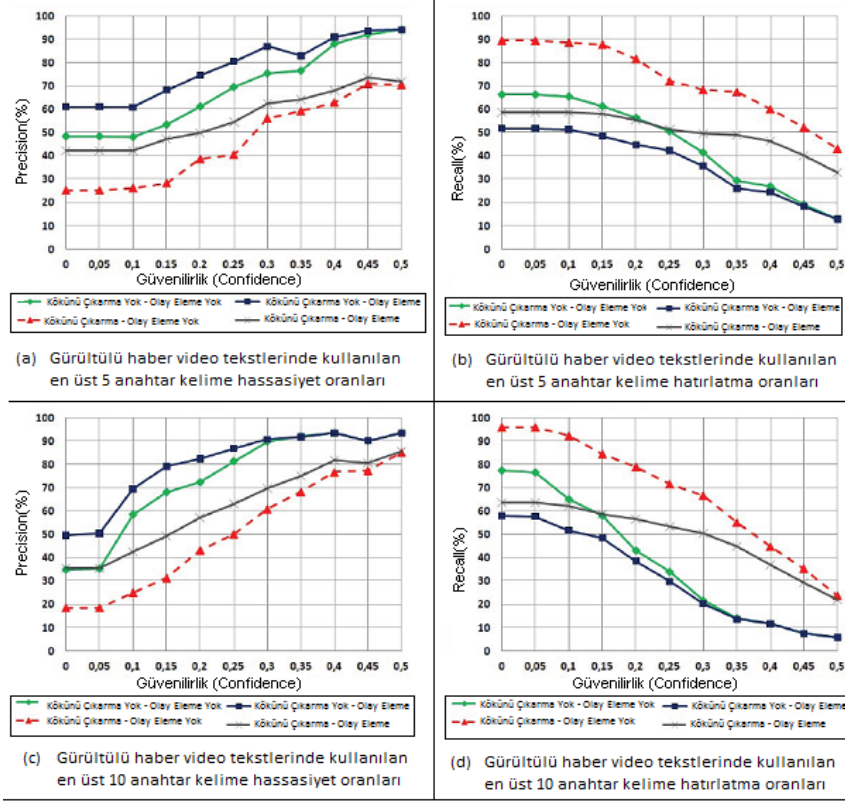
Zhang ve arkadaşlarının yaklaşımını, hedef metin türleri arasındaki farklılıktan dolayı genel video metninden olay çıkarımı için kullanamayız. Çünkü, onların kullanmış olduğu her bir zaman ayarlı Webcast metni tek bir olay göstermesine rağmen, bizim kullanmakta olduğumuz haber video metinleri genellikle birden fazla olay içermektedir. Bundan dolayı, Webcast metinleri LSA kullanarak otomatik kümelenebilmekte ve ortaya çıkan anlamsal olay etiketleri veren metin kümeleri ayırık olmaktadır. Oysa, bizim kullandığımız haber video metinlerinde örtüşen kümeler mevcuttur ve bunlar sistemin eğitilmesi sırasında manuel olarak sisteme öğretilmektedir. Buna ek olarak, ortaya çıkan metin kümelerinin ayırık olmaması bizim *tf-idf* kullanmamızı engellemiş ve onun yerine bizim formülümüzde *tf* kullanılmıştır.

Anahtar kelime tabanlı olay çıkarımı yaklaşımının değerlendirilmesi, TRT tarafından 2007'nin dördüncü, beşinci ve altıncı aylarında yayınlanmış haber videolarından oluşturulmuş veri seti üzerinde yapılmıştır. Test videolarının toplam süresi 1,5 saattir. Veri seti içindeki haber yazısı sayısı 182'dir.

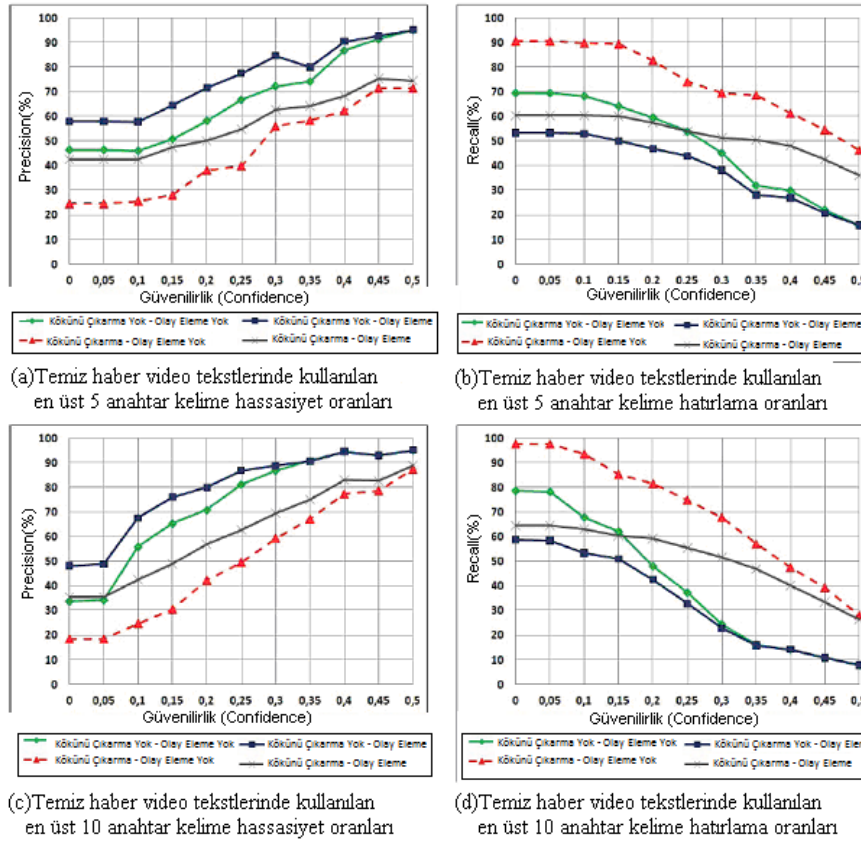
Aşağıdaki parametreler için değerlerin kombinasyonu kullanılarak hazırlanan veri setinde 16 farklı değerlendirme (precision ve recall olarak) yapılmıştır:

- Test Veri Seti: Gürültülü metin (sözcüksel kurallar ve yazım denetleyicisiyle düzeltildikten sonra kayan metin yakalayıcı çıktısı) veya temiz metin (mükemmel manüel transkripsiyon) .
- Kullanılan Anahtar Kelimenin frekansı: En sık görülen 5(üst-5) veya en sık görülen 10 (üst-10) anahtar kelime.
- Kökünü çıkarma (stemming): Anahtar kelimeleri olduğu gibi kullanmak ya da köklerini kullanmak
- Eğitim Veri Seti İçindeki Olayların birliktelik (co-occurrence) kalıplarının kullanımı ya da kullanılmaması: bu parametre haber metninden düşük güvenle çıkarılan olayların, eğitim setinde yüksek güvenle çıkarılan olaylarla birlikte olmayanlarının elenmesi fikrine dayanır. Bu nedenle, son işlem olarak, olayların birliktelik (co-occurrence) kalıp kullanımı dikkate alınarak test veri kümesinden çıkarılan düşük güvenli olayların elenmesi yapılır.





Şekil-7: Gürültülü Haber Video Metinlerinden Olay Çıkarım Değerlendirme Sonuçları.



Şekil-8: Temiz Haber Video Metinlerinden Olay Çıkarım Değerlendirme Sonuçları.

Toplam 4 saat süren ve 35 TRT yayın videosu içeren eğitim veri setinden en sık görülen 10 olay tipi; “İfade, Ölüm, Duruşma/Soruşturma, Kaza, Hava Durumu, Toplantı, Saldırı, Yaralanma, Seçim ve Operasyon” olarak otomatik belirlenmiştir. Bu olay tipleri için en çok görülen anahtar kelimeler de daha önce belirlenmişti. Test veri setinin metni, eğitim seti içindeki videoların haber metni içinden aktarılan olaylarla annotate edilmiştir. Test veri setinin 182 haber metni içinde toplamda 362 olay mevcuttur. Dolayısıyla her bir haber metni birden fazla olay içerebilmektedir.

Değişik parametre ayarları için değerlendirme sonuçları Şekil-7 ve Şekil-8’de güvenilirlik seviyesi 0 dan 0,5’e kadar precision ve recall olarak gösterilmiştir. Sonuçlar, Türkçe haber videolarından otomatik anlamsal olay çıkarımında ilk deneme için umut vericidir.

Şekil-7 ve 8’deki grafikler incelendiğinde, temiz metin üzerinde recall değerlerinin gürültülü metne göre biraz daha yüksek olduğu ve en güvenli seviye için precision değerlerinin biraz daha az olduğu görülmüştür. Şekillerde görüldüğü gibi, en sık görülen 10 anahtar kelimeye karşı en sık görülen 5 anahtar kelime kullanımı, anahtar kelime olarak köklere karşı kelimenin olduğu gibi (token) kullanımı ve eğitim veri kümesinde olayların birliktelik (co-occurrence) kalıplarının kullanımı precision’i artırırken, recall değerlerinde azalmaya yol açmıştır.

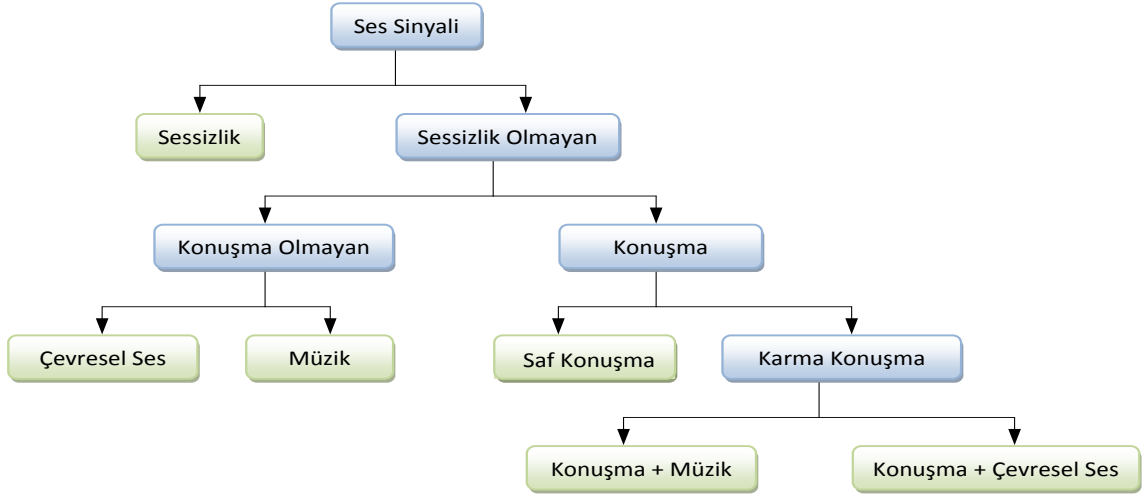
Dikkate alınan 10 olay için alakalı anahtar kelimeler toplamda 4 saat süren eğitim veri seti kullanılarak belirlenmiştir. Bu aslında sınırlı bir settir ve eğer daha büyük boyutta bir eğitim veri setine ulaşılabılırsa, sonuçların daha da iyileşeceği düşünülmektedir.

### **3.4 Sesten Anlamsal Bilgi Çıkarımı (4. İş Paketi)**

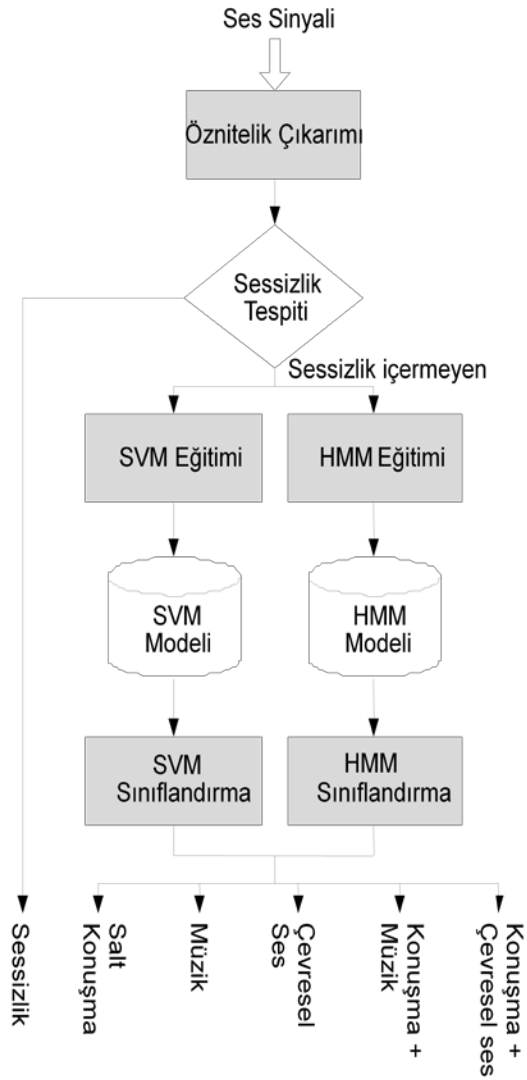
Günümüzde videoların sınıflandırılmasında ses verileri de kullanılmaktadır. Ses verileri oldukça zengin ve tamamlayıcı bir içerik barındırmaktadır. Ses verileri görüntü verilerine yardımcı olarak kullanılabilmesi gibi videoların sınıflandırılması ses verilerinin öncülüğünde de yapılabilir. Sesten Anlamsal Bilgi Çıkarımı iş paketi kapsamında iki aşamalı bir çalışma yürütülmüştür. İlk aşamada, bileşik formattaki ses verileri Destek Vektör Makinesi (Support Vector Machine – SVM) ve Saklı Markov Modeli (Hidden Markov Model – HMM) sınıflandırıcıları kullanılarak altı ana sınıfa (müzik, saf konuşma, çevresel ses, konuşma+müzik, konuşma+çevresel ses ve sessizlik) ayrıştırılmıştır. Oluşturulan sınıflandırma yapısı Şekil-9’da verilmiştir.

Ses sınıflandırılması için önerilen sistem, öznitelik çıkarımı, sessizlik tespiti, SVM eğitimi, HMM eğitimi, SVM sınıflandırma ve HMM sınıflandırma modüllerinden oluşmaktadır. Bu aşamada yapılan sınıflandırma işlemlerinde, MPEG-7 öznitelik kümesinden ASP (Audio Spectrum Projection), ASS (Audio Spectrum Spread), ASF (Audio Spectrum Flatness) ve ASC (Audio Spectrum Centroid) tanımlayıcıları kullanılmıştır. Bileşik formattaki seslerin sınıflandırılması için geliştirilen modülün blok şeması Şekil-10’da verilmiştir.

Geliştirilen sistemde, öznitelikler öge boyutu (granularity) 1 sn olan çerçeveler için hesaplanmakta ve sınıflandırıcı kararları bu çerçeveler için verilmektedir. Ancak, bazı durumlarda sınıflandırıcılar hatalı kararlar verebilmektedir. Örneğin; birer saniyelik ardışık üç çerçeve (s0, s1, s2) ele alındığında, sınıflandırıcının bu çerçeveler için kararı “müzik-saf konuşma-müzik” olabilmektedir. Sürekli bir ses sinyali içerisindeki ses türlerinin bu öge boyutu (1 sn) için bu kadar ani ve/veya hızlı değişimi neredeyse imkansızdır. Bu çerçeve dizisinin tamamının “müzik” olma olasılığı yüksektir. Bu varsayım, sınıflandırıcının vermiş olduğu karara son bir yumuşatma (smoothing) kuralı uygulanmaktadır. Bu kural aşağıdaki biçimde tanımlanmıştır:



Şekil-9: Sınıflandırma ağacı



Şekil-10: Önerilen sistemin blok şeması

$$\text{Yumuşatma Kuralı} = (S_1 \neq S_0 \wedge S_2 = S_0) \rightarrow S_1 = S_0$$

Bu kurala göre, “müzik – saf konuşma – müzik” biçimindeki ardışık çerçevelerinin tamamı “müzik” olarak değiştirilmekte ve bazı sınıflandırma hataları önlenmektedir. Geliştirilen modülde, bu örnekteki yumuşatma kuralı benzeri kurullarla yumuşatma işlemi yapılarak akustik sınıflar elde edilmektedir.

Ses verilerinin güdümlü (supervised) yöntemlerle sınıflandırılmasında kullanılan eğitim kümesinin niteliği ve miktarı büyük önem arz etmektedir. Hazır kullanılabilir bir veri seti bulmak mümkün olmadığından, proje sürecinde veri toplama ve sınıflandırıcı eğitimlerine ağırlık verilmiştir. Bu amaçla, TRECVID 2003, ABC World News Tonight, CNN headline news, internet [10] ve çeşitli müzik disklerinden yaklaşık 4 saatlik veri derlenmiştir. Bu veri kümesinin farklı ses türlerine göre dağılımları şu şekildedir:

- Saf Konuşma: 65 dk
- Müzik: 43 dk
- Çevresel ses: 31 dk
- Müzik üzerine konuşma: 45 dk
- Çevresel ses üzerine konuşma: 26 dk

Toplanan veriler ses sınıflarına göre elle bölütlenmiş ve ses örneklerinin yarısı eğitim; diğer yarısı da test amaçlı kullanılmıştır. Yapılan testlerde, SVM tabanlı ve HMM tabanlı ses sınıflandırma yöntemlerinin performansları karşılaştırılmış ve her bir öznitelik kümesinin etkinliği SVM ve HMM sınıflandırma yöntemleriyle denenmiştir.

Otomatik Konuşma tanıma (Automatic Speech Recognition – ASR) sistemlerinin bir çoğunda HMM sınıflandırma için sol-sağ modeli kullanılmaktadır. Ancak, zamansal verimizin yapısından dolayı bu topoloji tercih edilmemiştir. Bunun yerine, her durumdan bir diğerine geçişi ve çıkılan bir duruma yeniden girişi mümkün kılan beş durumlu bir ergodik model tasarlanmıştır.

SVM sınıflandırıcısında, doğrusal olarak ayıramayan problemlerde iyi sonuçlar veren Radyal Taban Fonksiyonu (RTF) (Radial Basis Function – RBF) gamma ( $\gamma$ ) ve maliyet (C) parametreleri ile birlikte kullanılmıştır. Bu parametrelerin tespiti için eğitim kümesinin küçük bir kısmı sağlama (validation) amacıyla kullanılmıştır. C, SVM tarafından hatalı sınıflandırma yapılması durumunda verilen ceza puanını temsil etmektedir. Bu parametrenin tespiti için k-fold cross validation (k = 5) yöntemi kullanılmıştır. Elde edilen sonuçlara göre, sınıflandırıcı parametreleri aşağıdaki şekilde belirlenmiştir:

- ASP özniteliği için  $\gamma = 6$ ,  $C = 20$
- ASC, ASS ve ASF özniteliklerinin tek bir öznitelikte birleştirildiği ASCSF öznitelikleri için  $\gamma = 2$ ,  $C = 20$

Elde edilen sınıflandırma sonuçları Tablo 10-13 olarak verilmiştir. Doğruluk oranı, doğru biçimde sınıflandırılan örnek sayısının, toplam test verisine oranı olarak tanımlanmıştır. Tablolarda, doğru sınıflandırılmış örnekler “doğru”, hatalı sınıflandırılmış örnekler ise “hatalı” olarak gösterilmiştir.

Tablo 10: Sınıflandırma sonuçları (“Konuşma” – “Konuşma İçermeyen”)

Öznitelik	ASP				ASC + ASS + ASF (ASCSF)			
	SVM		HMM		SVM		HMM	
Sınıflandırıcı	doğru	hatalı	doğru	hatalı	doğru	hatalı	doğru	hatalı
Konuşma İçermeyen	1749	614	2078	285	1791	572	2069	294
Konuşma İçeren	3814	271	3675	410	4042	43	3854	231
Doğruluk Oranı	%86.3		%89.2		%90.5		%91.9	

Tablo 11: Sınıflandırma sonuçları (“Saf Konuşma” – “Konuşmayla Karışık”)

Öznitelik	ASP				ASC + ASS + ASF (ASCSF)			
	SVM		HMM		SVM		HMM	
Sınıflandırıcı	doğru	hatalı	doğru	hatalı	doğru	hatalı	doğru	hatalı
Saf Konuşma	1659	290	1817	132	1773	176	1726	223
Konuşmayla Karışık	1737	399	1558	578	1879	257	1786	350
Doğruluk Oranı	%83.1		%82.6		%89.4		%86.0	

Tablo 12: Sınıflandırma sonuçları (“Konuşma+Müzik” – “Konuşma+Çevresel Ses”)

Öznitelik	ASP				ASC + ASS + ASF (ASCSF)			
	SVM		HMM		SVM		HMM	
Sınıflandırıcı	doğru	hatalı	doğru	hatalı	doğru	hatalı	doğru	hatalı
Konuşma + Çevresel Ses	537	231	639	129	446	322	637	131
Konuşma + Müzik	682	686	735	633	1322	46	1182	186
Doğruluk Oranı	%57.1		%64.3		%82.8		%85.2	

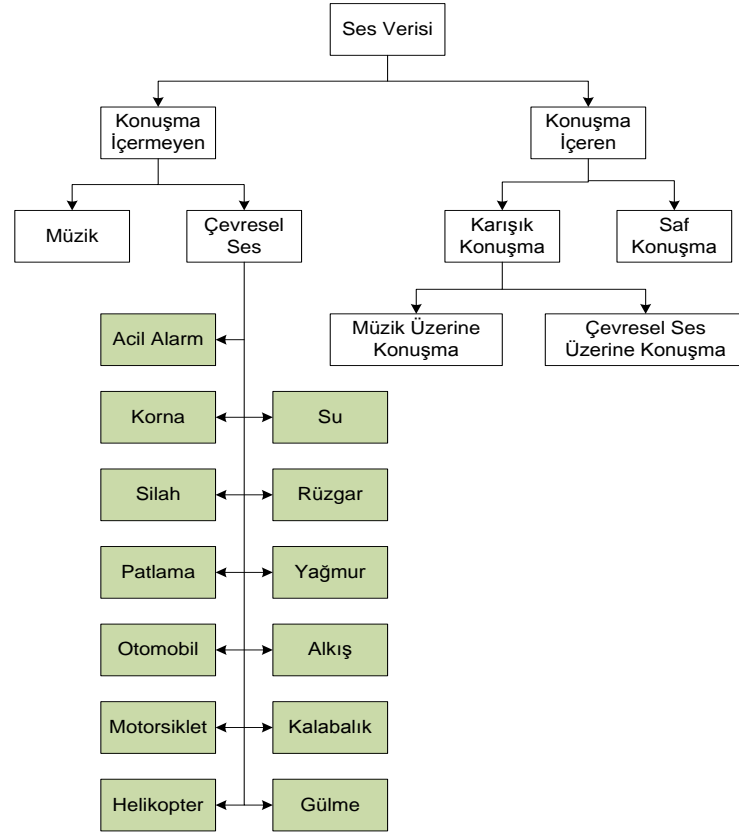
Tablo 13: Sınıflandırma sonuçları (“Çevresel Ses” – “Müzik”)

Öznitelik	ASP				ASC + ASS + ASF (ASCSF)			
	SVM		HMM		SVM		HMM	
Sınıflandırıcı	doğru	hatalı	doğru	hatalı	doğru	hatalı	doğru	hatalı
Çevresel Ses	552	492	929	115	941	103	952	92
Müzik	1101	218	1032	287	1317	2	1305	14
Doğruluk Oranı	%70.0		%83.0		%95.6		%95.5	

Elde edilen deneysel sonuçlar neticesinde, kullanılan iki farklı öznitelik kümesinden ASCSF kümesinin, ASP kümesine kıyasla daha başarılı sonuçlar ürettiği gözlenmiştir. Bu iki küme için SVM ve HMM üzerinde karşılaştırmalar yapılmıştır. ASP niteliklerinin kullanıldığı denemelerde, doğruluk oranının ASCSF niteliklerinin kullanıldığı örneklere göre önemli ölçüde düşük olduğu, yani ASCSF öznitelik kümesinin, “haberler” veri alanında karışık tip sınıflar için en yüksek ayrımı sağladığı gözlenmiştir. Daha açık ifade etmek gerekirse:

- HMM sınıflandırıcı ve ASCSF öznitelik kümesi kullanılarak yapılan “Konuşma İçeren” ve “Konuşma İçermeyen” ayırımında en yüksek doğruluk oranı %91.9 olarak gözlenmiştir.
- SVM sınıflandırıcı ve ASCSF öznitelik kümesi kullanılarak yapılan “Saf Konuşma” ve “Konuşmayla Karışık” sınıflandırmasındaki en yüksek doğruluk oranı %89.4 olarak gözlenmiştir.
- HMM sınıflandırıcı ve ASCSF öznitelik kümesi kullanılarak yapılan “Çevresel ses ve Konuşma” ve “Müzik ve Konuşma” sınıflandırmasındaki en yüksek doğruluk oranı %85.2 olarak gözlenmiştir.
- SVM sınıflandırıcı ve ASCSF öznitelik kümesi kullanılarak yapılan “Çevresel ses” ve “Müzik” sınıflandırmasındaki en yüksek doğruluk oranı % 95.6 olarak gözlenmiştir.

Karmaşıklık açısından, SVM eğitim süresi HMM'ye kıyasla çok fazladır. Öznitelik kümeleri ele alındığında, ASP özniteliğinin çıkarımı, ASC, ASS ve ASF özniteliklerinin çıkarımına kıyasla daha fazla zaman ve fiziksel bellek kullanımına sebep olduğu gözlenmiştir.



Şekil-11: Ses Sınıflandırma Ağacı

Sesten anlamsal bilgi çıkarımı işinin ikinci aşamasında, çevresel ses olduğu önceki aşamada belirlenmiş ses kayıtları on üç (13) farklı akustik sınıfa ayrılmıştır. Bu sınıflar sırasıyla *acil alarm*, *korna*, *silah*, *patlama*, *otomobil*, *motorsiklet*, *helikopter*, *su*, *rüzgar*, *yağmur*, *alkış*, *kalabalık* ve *gülme* olarak belirlenmiştir. Önceki dönem raporlarında “uyarı”, “savaş”, “motor” ve “doğa” olarak belirlenen dört ses sınıfından, daha zengin bir sınıflandırma yapabilmek amacıyla vazgeçilmiştir. Bu sınıfların, diğer ses sınıflarıyla ilişkisi Şekil-11’de verilmiştir.

Yeni ses sınıfları için kullanılan ses özniteliklerinin sayısı artırılmıştır. MPEG-7 ses öznitelik ailesi ve yaygın olarak kullanılan diğer ses öznitelikleri kapsamında toplam on bir (11) adet ses özniteliği düşünülmüştür. Çalışmada ele alınan öznitelikler aşağıda listelenmiştir:

- MPEG-7 Audio Spectrum Projection (ASP)
- MPEG-7 Audio Spectrum Flatness (ASF)
- MPEG-7 Audio Harmonicity (AH)
- MPEG-7 Audio Spectrum Centroid (ASC)
- MPEG-7 Audio Spectrum Spread (ASS)
- Zero Crossing Rate (ZCR)
- Mel Frequency Cepstral Coefficient (MFCC)
- Bileşik ses özniteliği: MPEG-7 ASF, ASC, ASS, ve AH (ASFCS-H).
- Bileşik ses özniteliği: MPEG-7 ASF, ASC, ASS, ve ZCR (ASFCS-ZCR)
- Bileşik ses özniteliği: MFCC, MPEG-7 ASC, ASS, ve AH (MFCC+ASC+ASS+AH)
- Bileşik ses özniteliği: MFCC, MPEG-7 ASC, ASS, ve ZCR (MFCC+ASC+ASS+ZCR)

Bu öznitelikleri elde edebilmek amacıyla Java tabanlı MPEG-7 Audio Encoder aracı kullanılmıştır [11]. Sınıflandırma algoritması olarak yine HMM ve SVM yöntemleri incelenmiştir.

Her bir çevresel ses sınıfı için HMM parametreleri Baum-Welch algoritması kullanılarak hesaplanmıştır. Bu durum HMM’lerde “eğitime problemi” olarak adlandırılmaktadır. Her bir parametrenin ilk değerlerini alarak, Baum-Welch algoritması döngülü (iterative) hesaplamalar için en iyi değerleri bulur. İlk değerlerin bulunmasında rastgele değerler yerine K-Means Clustering algoritması kullanılmıştır. “Değerlendirme problemi” olarak bilenen model ve gözlem dizilerinin bilinip gözlenen dizinin bu modelden üretilme olasılığının hesaplanması İleri-Geri Yön algoritması (Forward-Backward) ile sağlanmıştır.

HMM sınıflandırmada her bir bölüt HMM modeline girdi olarak verilmektedir. Girdi bölütü için en yüksek tanıma olasılığını veren HMM modeli bu bölütü etiketlemektedir. Verilen  $M_1, M_2, \dots, M_n$  modelleri ve 1-sn uzunluğundaki bölüt dizisi S için  $P_i$  model olasılığı aşağıdaki biçimde hesaplanmaktadır:

$$P_i = \text{HmmPredict}(M_i, S); 0 \leq i < 13 \quad (1)$$

Burada *HmmPredict* işlevi HMM’nin çıkarım algoritmasıdır. Daha sonra, modellerden gelen en yüksek olasılıkla girdi bölütü etiketlenmektedir:

$$P_{max} = \max(\text{HmmPredict}(M_1, S), \dots, \text{HmmPredict}(M_n, S)), 0 \leq n < 13 \quad (2)$$

SVM sınıflandırmada, her bir çevresel ses sınıfını öğrenmek için ayrı birer SVM modeli eğitilmiştir. SVM modelleri *bire-karşı-hepsi* (one-versus-all) yaklaşımı ile inşa edilmiştir. Ayrıca, verileri çok boyutlu öznitelik uzayına eşlemlmek için radyal taban fonksiyonu (radial-basis-kernel – RBF) kullanılmıştır. Doğrulama aşamasında, ses bölütleri on üç sınıfa sınıflandırılmıştır. SVM modellerinden gelen olasılık sonuçları hesaplandıktan sonra, ses bölütleri, modellerden gelen en yüksek olasılık ilkesine göre etiketlenmiştir.

SVM test aşamasında, sınıflandırılmak istenen ses bölütleri her bir SVM modeline girdi olarak verilmektedir. SVM ikili bir sınıflandırıcı olduğundan, her modelin çıktısı ikili bir değerdir (+, -). Verilen  $M_1, M_2, \dots, M_{13}$  DVM modelleri ve 1-sn uzunluğundaki bölüt dizisi  $S = \{s_1, s_2, \dots, s_n\}$  için,  $PC$  pozitif bölüt sayısı,  $P_i$  pozitif bölüt sayısının toplam bölüt sayısı içerisindeki yüzdesi olmak üzere,  $V$  model olasılığı aşağıdaki gibi hesaplanmaktadır:

$$V_j = \text{SvmPredict}(M_i; s_j); 0 \leq i < 13; 0 \leq j < n$$

Burada,  $\text{SvmPredict}$  SVM'nin çıkarım algoritmasıdır. Pozitif bölüt sayısı  $PC$  ve olasılıkları aşağıdaki gibi hesaplanmaktadır:

$$PC_i = V_0 + V_1 + \dots + V_j; V_j > 0; 0 \leq i < 13; 0 \leq j < n$$

$$P_i = PC_i / n; 0 \leq i < 13$$

Son olarak, modellerden ( $M_i$ ) gelen en yüksek olasılıkla girdi bölütü etiketlenmektedir:

$$P_{\max} = \max(PC_0, PC_1, \dots, PC_i); 0 \leq i < 13$$

Detaylı testler sonucunda, bazı sınıflandırma hatalarının giderilebileceği sonucuna varılmıştır. Gözlemler sonucunda iki durum kurala bağlanmıştır:

**Kural 1:**  $(s_1 \neq s_0 \ \& \ s_0 = s_2) \Rightarrow s_1 = s_0$ .

**Kural 2:**  $(s_0 = s_3 \ \& \ s_1 \neq s_0 \ \& \ s_1 \neq s_2 \ \& \ (s_1.\text{secondlabel} = s_0 \ \vee \ s_2.\text{secondlabel} = s_0)) \Rightarrow s_1 = s_0, s_2 = s_0$ .

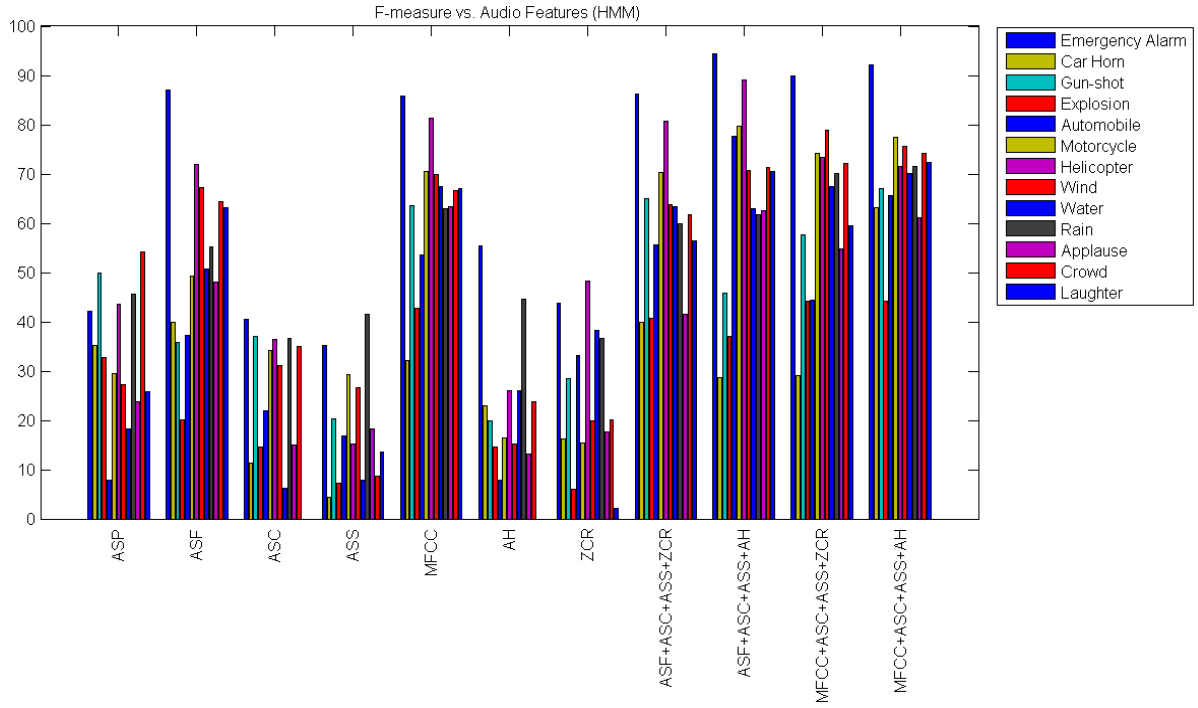
Birinci kuralda belirli bir zamandaki 3 ardışık ses bölütü ( $s_0, s_1, s_2$ ) düşünülmektedir. Kurala göre, ortadaki bölüt kendisinden bir önceki ve bir sonraki bölütlerden farklı etiketlere sahipse ve birinci ve üçüncü bölütler aynı sınıf etiketine (örneğin *alkış* sesi) sahipse, ortadaki bölüt büyük olasılıkla hatalı etiketlenmiştir ve bu bölütün etiketi birinci bölüt ile aynı olur. Bu kuralda, ses sınıflarının 1-sn'lik ardışık bölütlerde ani değişimler gösteremeyeceği ilkesi benimsenmiştir. İkinci kuralda, belirli bir zamandaki 4 ardışık ses bölütü ( $s_0, s_1, s_2, s_3$ ) düşünülmektedir. Bu kural, ortadaki iki bölüt ( $s_1, s_2$ ) farklı etiketlere sahipse ve birinci ve dördüncü bölütler aynı sınıf etiketine sahipse geçerlidir. Ortadaki iki bölütün hatalı sınıflandırma olduğuna karar verebilmek için, SVM modelinin her iki bölüt için vermiş olduğu ikinci en yüksek olasılıklar kontrol edilir. Bu sınıf etiketlerinden herhangi birisi, birinci (ya da dördüncü) bölüt ile aynı sınıf etiketine sahipse, ortadaki iki bölüt ( $s_1, s_2$ ) hatalı sınıflandırma olarak değerlendirilir ve kurala göre tekrar etiketlenir.

Tablo 14: Çevresel sesler için oluşturulan veri kümesi

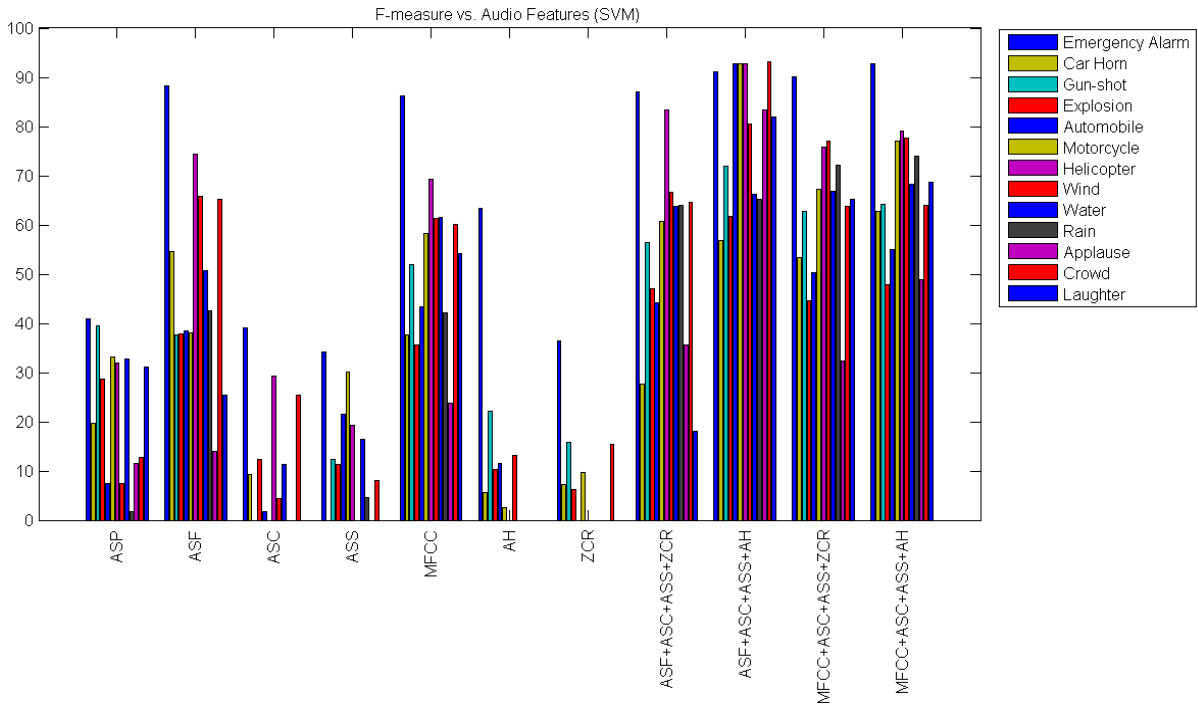
Ses	Süre
Acil Alarm	23 dk 39 sn
Korna	4 dk 44 sn
Silah	9 dk 50 sn
Patlama	19 dk 14 sn
Helikopter	7 dk 24 sn
Motorsiklet	9 dk 57 sn
Otomobil	6 dk 53 sn
Yağmur	13 dk 45 sn
Rüzgar	17 dk 15 sn
Su	27 dk 49 sn
Alkış	10 dk 30 sn
Gülme	13 dk 26 sn
Kalabalık	9 dk 45 sn
<b>Toplam</b>	<b>3 saat 4 dk</b>



Daha önce oluşturulan veri seti, yeni çevresel sesleri içerecek biçimde genişletilmiştir. Deney ve değerlendirmelerde kullanılan veriler internet ortamındaki ses efektlerinden derlenmiştir [12]. Oluşturulan veri kümesinin özeti Tablo 14’de verilmiştir.



Şekil-12: Ses özneliklerinin HMM performansı



Şekil-13: Ses özneliklerinin SVM performansı

Bütün ses klipleri el ile gereksiz bölümlerden arındırılmış ve CD kalitesinde ses olarak kaydedilmiştir. Acil alarm sesleri arasında ambulans, polis, ve yangın alarmları bulunmaktadır. Patlama sesleri bomba seslerinden oluşmaktadır. Silah sesleri tabanca,

*makineli tüfek ve laser silah seslerinden oluşmaktadır. Veri kümesindeki diğer sesler helikopter, motorsiklet, otomobil, trafikten alınan sesler, araç içi/dışı sesler ve su seslerinden (yüzme, dalga, okyanus, damlama, vb.) oluşmaktadır. Bu veri kümesinin %85'lik bir kısmı model eğitimi, geri kalanı ise model testleri aşamasında kullanılmıştır.*

Geliştirilen tanıma sisteminde HMM ve SVM sınıflandırıcıları ASP, ASF, ASC, ASS, AH, ZCR, MFCC, (ASFCS-H), (ASFCS-ZCR), (MFCC + ASC + ASS + AH) ve (MFCC + ASC + ASS + ZCR) öznitelikleri ile test edilmiştir. Yapılan deneysel çalışmalarda, öznitelikler tek başına kullanıldığında, MPEG-7 ASF ve MFCC öznitelikleri dışında oldukça düşük tanıma başarımları gözlenmiştir. Tüm özniteliklerin HMM ve SVM sınıflandırıcı performansları sırasıyla Şekil-12 ve Şekil-13'de sunulmuştur.

ASC, ASS, AH ve ZCR öznitelikleri sırasıyla %27.4, %20.2, %24.7 ve %27.8 ortalama f-measure skorlarına sahiptir. Bu öznitelikler tek boyutlu olduğundan, tek başlarına kullanıldığında tanıma başarımlarının oldukça düşük olduğu gözlenmiştir. Genel ses tanıma uygulamalarında yaygın olarak kullanılan MFCC özniteliğinin %65.3 ortalama f-measure skoru ile, ASF özniteliğine kıyasla (% 55.7) daha iyi sonuç verdiği gözlenmiştir.

Ses karakteristiklerini daha iyi ifade edebilmek amacıyla, ayrı ayrı elde edilen öznitelikler birleştirilmek suretiyle Şekil 12 ve Şekil 13'de görülen yeni öznitelik kümeleri elde edilmiştir. ASF, ASC, ASS ve ZCR öznitelikleri birleştirildiğinde ortalama f-measure skorunun HMM sınıflandırıcısı için %62.1'e yükseldiği gözlenmektedir (Tablo 15).

*Tablo 15: ASF+ASC+ASS+ZCR bileşik özniteliğinin HMM başarımı*

Ses	Recall (%)	Precision (%)	F-Measure (%)
Acil Alarm	86,8	85,9	86,3
Korna	30,0	60,0	40,0
Silah	56,3	76,5	64,9
Patlama	53,1	33,0	40,7
Otomobil	54,5	56,7	55,6
Motorsiklet	62,2	80,8	70,3
Helikopter	85,5	76,5	80,8
Rüzgar	56,0	73,8	63,7
Su	70,5	57,4	63,3
Yağmur	62,6	57,7	60,0
Alkış	35,0	51,2	41,5
Kalabalık	93,5	46,0	61,7
Gülme	52,7	60,9	56,5
<b>Average</b>	<b>61,4</b>	<b>62,8</b>	<b>62,1</b>

*Tablo 16: ASF+ASC+ASS+AH bileşik özniteliğinin HMM başarımı*

Ses	Recall (%)	Precision (%)	F-Measure (%)
Acil Alarm	94,7	94,2	94,4
Korna	16,7	100	28,6
Silah	33,1	74,6	45,8
Patlama	48,4	30,1	37,1
Otomobil	79,2	76,3	77,7
Motorsiklet	75,4	84,4	79,6
Helikopter	85,6	92,7	88,9
Rüzgar	66,1	76,0	70,7
Su	70,1	56,9	63,0
Yağmur	77,6	51,2	61,7
Alkış	67,8	57,9	62,5
Kalabalık	98,4	55,9	71,3
Gülme	74,3	67,1	70,5
<b>Average</b>	<b>68,2</b>	<b>70,5</b>	<b>69,4</b>

ASFCS-H bileşik özniteliği HMM sınıflandırıcısı ile test edildiğinde ortalama %69.4 f-measure skoru elde edilmektedir. Görüldüğü üzere, bu öznitelik kombinasyonunun başarımları ASFCS-ZCR kombinasyonundan daha başarılı sonuç vermektedir (Tablo 16). Ses sınıflarına ayrıntılı bakıldığında, AH özniteliğinin acil alarm, motorsiklet, otomobil, helikopter ve gülme seslerindeki başarımlarına katkısı olumludur.

Öznitelik testlerinde MFCC+ASC+ASS+ZCR ve MFCC+ASC+ASS+AH bileşik öznitelikleri de düşünülmüştür. Tablo 17’de görüldüğü üzere, MFCC+ASC+ASS+ZCR kombinasyonu %63.9’luk ortalama f-measure skoru ile Tablo 18’de verilen MFCC+ASC+ASS+AH kombinasyonundan (%70.6) daha düşük bir tanıma başarımına sahiptir. Deneysel çalışmalar sonucunda, HMM sınıflandırıcısı ile en iyi başarımları MFCC+ASC+ASS+AH öznitelik kombinasyonu ile elde edilmiştir.

*Tablo 17: MFCC+ASC+ASS+ZCR bileşik özniteliğinin HMM başarımları*

Ses	Recall (%)	Precision (%)	F-Measure (%)
Acil Alarm	88,9	90,8	89,8
Korna	26,6	32,0	29,0
Silah	46,9	74,8	57,7
Patlama	70,3	32,1	44,1
Otomobil	41,5	47,7	44,4
Motorsiklet	76,2	72,0	74,1
Helikopter	77,1	70,0	73,3
Rüzgar	68,9	91,9	78,8
Su	75,1	61,1	67,4
Yağmur	69,1	71,1	70,1
Alkış	53,3	56,1	54,7
Kalabalık	85,4	62,3	72,1
Gülme	55,4	64,0	59,4
<b>Average</b>	<b>64,2</b>	<b>63,5</b>	<b>63,9</b>

*Tablo 18: MFCC+ASC+ASS+AH bileşik özniteliğinin HMM başarımları*

Ses	Recall %	Precision %	F-Measure %
Acil Alarm	90,4	93,9	92,1
Korna	60,0	66,6	63,1
Silah	59,2	77,3	67,0
Patlama	67,1	32,8	44,1
Otomobil	59,7	73,0	65,7
Motorsiklet	73,7	81,8	75,7
Helikopter	70,3	72,8	71,5
Rüzgar	70,4	81,8	75,7
Su	70,5	69,8	70,2
Yağmur	77,5	66,4	71,5
Alkış	62,7	59,6	61,1
Kalabalık	90,3	62,9	74,1
Gülme	72,9	72,0	72,4
<b>Average</b>	<b>71,1</b>	<b>70,0</b>	<b>70,6</b>

HMM için gerçekleştirilen testlerin aynısı SVM sınıflandırıcısı için de gerçekleştirilmiştir. Sonuçlar göstermektedir ki, ASFCS-H öznitelik kombinasyonu ortalama %80.6 F-measure skoru ile en iyi sınıflandırma başarımına sahiptir (Tablo 19). MFCC+ASC+ASS+AH öznitelik kombinasyonu, SVM ile kullanıldığında ortalama %69.4 F-measure skoru ile ikinci en iyi başarımı vermektedir.

Tablo 19: ASF+ASC+ASS+AH (ASFCS-H) bileşik özniteliğinin SVM başarımı

Ses	Recall (%)	Precision (%)	F-Measure (%)
Acil Alarm	84,3	99,4	91,1
Korna	74,2	46,0	56,8
Silah	61,2	87,2	71,8
Patlama	79,4	50,4	61,7
Otomobil	92,2	93,5	92,8
Motorsiklet	87,1	99,1	92,7
Helikopter	95,8	89,8	92,7
Rüzgar	75,0	87,0	80,6
Su	81,4	55,9	66,3
Yağmur	53,9	82,7	65,2
Alkış	75,0	93,4	83,3
Kalabalık	96,8	89,7	93,1
Gülme	91,9	73,9	81,9
<b>Average</b>	<b>80,6</b>	<b>80,6</b>	<b>80,6</b>

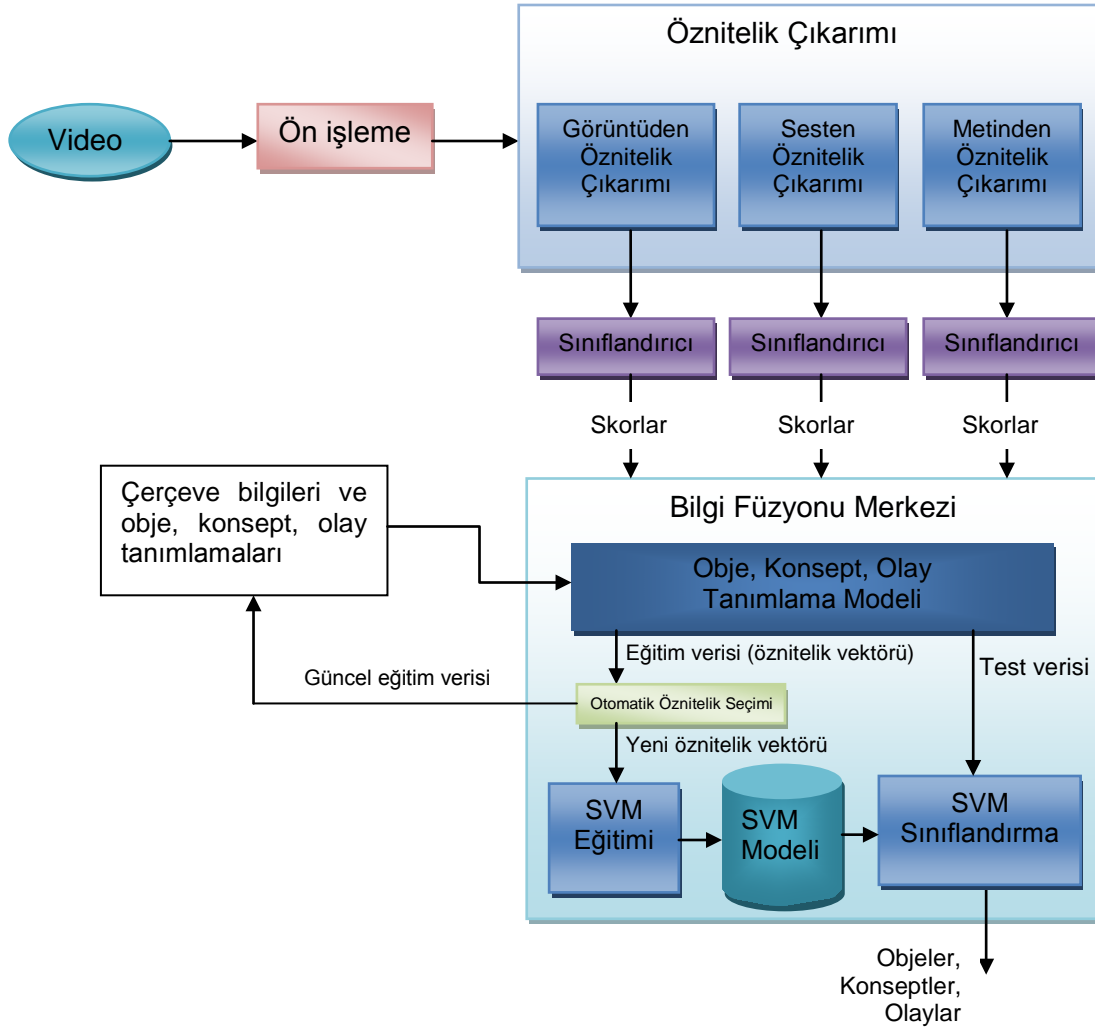
Elde edilen deneysel sonuçlar karşılaştırıldığında, en iyi tanıma başarımının SVM sınıflandırıcı ve ASFCS-H öznitelik kümesi ile elde edildiği görülmektedir.

### 3.5 Bilgi füzyonu (5. İş Paketi)

Bilgi füzyonu erken ve geç füzyon seviyesi olmak üzere iki aşamada gerçekleşebilir. Geç füzyon yaklaşımının daha etkili olduğu bir çok çalışmada kanıtlanmıştır [13], bunun yanı sıra konseptler arası ilişkiden erken füzyon yaklaşımında faydalanmak mümkün değildir. En önemlisi projede yer alan, görüntüden, sestem ve metinden anlamsal bilgi çıkarımı yapan modüller birbirinden bağımsız çalıştığı için projemizdeki füzyon yaklaşımımızın geç füzyon seviyesinde gerçekleşmesi daha uygundur.

Projede, füzyon tekniği olarak da Destek Vektör Makinası (Support Vector Machine – SVM) tercih edilmiştir. Bilgi füzyonu için önerilen çözümün blok şeması Şekil-14'de verilmiştir. Bilgi füzyonu sistemi üç ana alt modülden oluşmaktadır. Bunlar; öğrenilecek ve sınıflandırılacak obje, konsept ve olay bilgilerini dinamik olarak sisteme tanıtan bir ayrıştırma modülü, SVM eğitim modülü ve SVM sınıflandırma modülüdür. İlk modül sayesinde, sistem istenildiğinde dinamik olarak eğitilebilmekte ve sisteme yeni konsept, obje, olay tanımlamaları eklenebilmektedir. Bu modül SVM eğitim ve sınıflandırma modüllerinin girdilerini otomatik olarak oluşturmaktadır. Buna ek olarak sisteme otomatik öznitelik çıkarımı için bir modül daha eklenmiştir. Bir konsept sınıflandırılmasında kullanılan özniteliklerin bir uzman tarafından belirlenmediği durumlarda; bu modül sayesinde özniteliklerin o konseptin sınıflandırılması üzerindeki etkileri, yani bir konseptin öğrenilmesinde kullanılan özniteliklerin kalitesi hesaplanarak, etkisi belirli bir eşik değeri altında kalan öznitelikler elenmektedir. Bunun için, literatürde de sıkça kullanılan ve öznitelikleri ağırlıklandırılan ya da diğer bir

deyişle özniteliklere deęer biçen algoritmaların en başarılılarından biri kabul edilen Relief Algoritması [14] kullanılmaktadır. Bu algoritma özniteliklerin önemini hesaplayıp, her bir öznitelięe uygun bir ağırlık atamaktadır. Atanan bu ağırlıklara göre eşik deęerin altında ağırlıklara sahip öznitelikler elenmektedir. Böylece, hem konsept eğitiminde kullanılan öznitelik vektörünün boyutunun çok büyük olması engellenmekte, hem de ayırdedici etkisi olmayan özniteliklerin kullanılması engellenmektedir. Bunun yanı sıra uygulanan füzyon işleminde aynı konsepte ait farklı modalitelerden gelen tespit skorlarının yanında farklı konseptlerin skorları da kullanılmaktadır. Bu sayede, konseptler arasındaki ilişkiden doğrudan füzyon işlemi içinde faydalanıp, konseptlerin tespit başarısının artması hedeflenmektedir.



Şekil-14: Otomatik Bilgi Çıkarımı

Öznitelik vektörü son halini aldıktan sonra, obje, konsept ve olay bilgilerinde kullanılacak öznitelikler, yeni daraltılmış özniteliklerle güncellenmektedir. Böylece test aşamasında da daraltılmış öznitelik vektörü kullanılması sağlanmaktadır. SVM eğitim modülü, bu öznitelik vektörü üzerinde çalışmakta ve her bir obje, konsept ve olay için birer SVM modeli oluşturulup, test verisindeki hedef obje, konsept ve olayları sınıflandırmaktadır. SVM eğitim modülünün çalışma şeklini kısaca özetlemek gerekirse; önce öznitelik deęerlerinin

ölçeklendirmesi yapılmaktadır. SVM eğitim modülünde, kernel fonksiyonlarından doğrusal olarak ayrılamayan problemlerde iyi sonuçlar veren Radyal Taban Fonksiyonu (Radial Basis Function – RBF) kullanılmaktadır. Parametre seçiminin performans üzerinde önemli rol oynamasından dolayı RBF'in kullandığı gamma ( $\gamma$ ) ve maliyet (C) parametrelerinin değerleri, k-kat (k-fold : eğitim seti k parçaya bölünüp, k-1 parça eğitim, diğer parça doğrulama setini oluşturacak şekilde parametre seçimlerinin k kez tekrar edilmesi) bağımsız geçerlilik sınaması (cross-validation) yöntemi kullanılarak ve  $\gamma$  ve C parametreleri üzerinde aralık tarama (grid-search) yapılarak hesaplanmaktadır. Bağımsız geçerlilik sınaması ile “fazla kapsama” (overfitting) probleminin önüne geçilmekte ve farklı (C,  $\gamma$ ) çiftleri denenerek en iyi bağımsız geçerlilik sınaması doğruluğunu sağlayan çift seçilmektedir. Sınıflandırılacak her obje, konsept, olay için bu işlem ayrı ayrı yapılmış ve her sınıf için en uygun parametre değerleri bulunmuştur. İlk geliştirmede, bu değerler hesaplanırken, sonuç üzerindeki etkileri doğruluk (accuracy) kriterine göre hesaplanmıştır. Ancak, eğitim verisinin dengesiz olduğu durumlarda, örneğin eğitim verisinde negatif örnek pozitif örneğe göre çok fazla ise, bu kriterin doğru bir seçim olmayabileceği görülmüştür. Bu nedenle SVM eğitim modülü, bağımsız geçerlilik sınaması aşamasında kullandığı doğruluk tabanlı kriter dışında precision, recall ve f-ölçütü (f-measure) gibi yeni kriterleri destekleyecek şekilde geliştirilmiştir. Literatürdeki çoğu çalışma, önerdikleri yöntemlerin başarısını doğruluk (accuracy) yerine precision, recall ve f-ölçütü gibi kriterlerle yapmaktadır. Bu bağlamda başarıyı arttıran en iyi parametre değerlerinin hesabı için bu kriterlerin kullanılması daha uygundur.

Füzyon modülünün eğitim ve testleri; NIST (Amerikan Standartlar Enstitüsü) tarafından düzenlenen TRECVID video erişimi değerlendirme yarışmasında paylaşılan ve literatürde en çok kullanılan videolar [15] ile çalışılmıştır (TRECVID 2007 videoları kullanılmıştır). Bunun yanı sıra füzyon modülünün çalışma performansı yeni ve popüler bir dataset olan CCV Database [16] ile yapılan deneylerle de ölçülmüştür.

TRECVID 2007 videolarına ait eğitim verisi 110 videodan oluşmakta, 21,532 referans çekim (shot) içermektedir. Test verisi ise 109 videodan oluşmakta ve 18,142 referans çekim içermektedir. Toplamda 20 konseptin (uçak, hayvan, araba, toplantı, yürüyüş yapan insanlar, hava durumu, vb.) eğitimi ve testi gerçekleştirilmiş ve sonuçları Tablo-20'de verilmiştir. Sistemin başarısı, literatürde TRECVID videoları kullanan çalışmaların başarılarını ölçmede kullandıkları Mean Average Precision (MAP) kriterine göre hesaplanmıştır. Kriteria ait formül aşağıda verilmiştir:

$$AveP = \sum_{k=1}^n P(k) \Delta r(k)$$

$$MAP = \frac{\sum_{q=1}^Q AveP(q)}{Q}$$

MAP başarı değerlendirme kriteri AveP (Average Precision) fonksiyonu ve Q (sorgu sayısı) değerini kullanmaktadır. Literatürdeki çalışmaların birbiriyle olan başarılarını kontrol ederken kullandıkları sorgu sayısı 2000 olduğu için, bu çalışmada da aynı değer kullanılmıştır. AveP hesabında P(k) fonksiyonu precision'a, r(k) fonksiyonu da recall'a denk gelmektedir. n ise konsept testinde kullanılan toplam shot sayısıdır (number of retrieved documents). TRECVID 2007 yarışması katılımcılarının toplamda 20 konsept için genel MAP başarıları %4 ile %13 aralığında değişmektedir [17]. Bizim sistemimizin füzyon neticesinde elde ettiği sonuç %10.76 dır. Füzyon aşaması sayesinde en iyi modalite sonucu olan %9.21 MAP değeri geliştirilerek, genelde % 16.8lik((10.76 - 9.21) / 9.21) bir başarı artımı sağlanmıştır. Konsept ilişkilerinden faydalanılmasının füzyon işlemine katkısının açıkça görülebilmesi için füzyon

sonuçları konsept ilişkilerini içermeyen multimodal füzyon ve konsept ilişkilerinden faydalanan multimodal füzyon şeklinde verilmiştir.

*Tablo-20: TRECVID 2007 veriseti üzerinde tek modalite ve füzyon sonuçlarının karşılaştırılması*

	Tek Modalite Sonuçları						Önerilen Füzyon Sonuçları	
	Görüntü Renk Tabanlı Modalite	Görüntü Doku Tabanlı Modalite	Görüntü Şekil Tabanlı Modalite	Ses Algısal Modalite	Ses Cepstral Modalite	Metin Modalitesi	Multimodal Füzyon	Konsept ilişkilerini Kullanan Multimodal Füzyon
Airplane	6.55%	8.25%	6.35%	7.33%	5.11%	6.09%	<b>9.83%</b>	9.36%
Animal	14.34%	17.18%	5.96%	8.51%	9.36%	7.76%	17.09%	<b>17.65%</b>
Boat_Ship	6.40%	12.79%	8.97%	7.26%	8.77%	8.14%	14.86%	<b>16.80%</b>
Car	15.63%	19.75%	14.25%	14.43%	16.18%	10.42%	21.38%	<b>23.30%</b>
Charts	3.37%	<b>5.95%</b>	2.42%	2.51%	1.34%	1.87%	4.52%	4.98%
Computer_TV-screen	<b>12.74%</b>	9.60%	7.42%	5.38%	8.13%	8.23%	10.83%	11.15%
Desert	1.92%	2.00%	<b>2.86%</b>	0.78%	1.04%	0.57%	2.03%	2.26%
Explosion_Fire	1.96%	2.32%	1.75%	2.47%	1.40%	2.11%	1.66%	<b>2.52%</b>
Flag-US	0.10%	0.42%	0.26%	0.13%	<b>1.61%</b>	0.25%	0.17%	0.31%
Maps	6.36%	10.94%	4.10%	2.61%	4.68%	2.47%	10.72%	<b>11.52%</b>
Meeting	23.69%	24.55%	23.75%	23.76%	29.41%	19.12%	31.16%	<b>31.44%</b>
Military	<b>5.13%</b>	3.14%	1.94%	3.45%	0.96%	0.99%	3.69%	3.87%
Mountain	8.92%	5.90%	3.86%	<b>9.06%</b>	2.70%	2.74%	6.46%	7.59%
Office	8.99%	13.90%	7.09%	5.63%	9.49%	2.66%	13.42%	<b>14.71%</b>
People-Marching	7.00%	7.42%	2.97%	3.37%	3.80%	1.45%	8.34%	<b>9.86%</b>
Police_Security	1.64%	4.87%	2.98%	3.14%	2.89%	2.52%	<b>4.96%</b>	4.94%
Sports	<b>11.08%</b>	5.50%	3.36%	3.67%	4.58%	1.75%	7.58%	9.52%
Truck	7.97%	<b>10.90%</b>	5.47%	5.07%	7.51%	5.77%	10.61%	9.24%
WaterscapeWaterfront	16.75%	17.54%	16.31%	9.44%	10.57%	8.21%	<b>22.48%</b>	22.22%
Weather	1.62%	1.38%	0.35%	1.80%	0.30%	0.27%	1.84%	<b>1.87%</b>
MAP	8.11%	9.21%	6.12%	5.99%	6.49%	4.67%	10.18%	<b>10.76%</b>
MAP Derecesi	4	3	6	7	5	8	2	1

Konsept bazında başarı artımlarına birkaç örnek vermek gerekirse; Boat\_Ship (Bot) konsepti için en iyi modaliteden elde edilen başarı %12.79 iken, Boat\_Ship konsepti için tüm modalitelerin füzyonu sonucu elde edilen başarı ise %14.86 olmuştur. Yani ses ve metin verilerinin yardımıyla başarı daha da artmıştır. Boat\_ship konseptine ait farklı modalitelerden gelen tespit skorlarının yanında diğer konseptlere ait skorların da füzyon işlemi içerisinde kullanılması sonucu daha da artırıp %16.8'e çıkarmıştır. Bunun yanı sıra Car (Araba) konseptinin en iyi modalite performansı da önerilen füzyon yöntemi ile %19.75'den %23.3'e çıkarak %18.3 oranında göreceli olarak iyileştirilmiştir.

CCV Database üzerinde yapılan deneylerde füzyon modülünün başarısını açıkça ortaya koymaktadır. Bu veriseti 9,317 Youtube videosundan oluşmakta ve videoların Basketbol, Futbol, Kedi, Geçit Töreni gibi konseptlerle ilişkilendirilmesini amaçlamaktadır. Jiang ve arkadaşları yaptıkları çalışma [16] kapsamında videolara ait 3 farklı öznitelik (SIFT, STIP, MFCC) paylaşmışlardır. Bunlardan SIFT görsel, STIP hareketli ve MFCC de sese ait özniteliklerdir. Bu proje çalışmasında da aynı modalite ve öznitelikler kullanılmıştır. Konsept ilişkilerini modellemek için video yorumlarından faydalanarak ekstra konsept bilgileri çıkarılmıştır. Yine konsept ilişkilerinin füzyon üzerindeki etkisini detaylı görebilmek için füzyon sonuçları, TRECVID 2007 verisetinde olduğu gibi, iki şekilde incelenmiştir. CCV Database verisetine ait deney sonuçları Tablo-21'de gösterilmiştir.

Tablo-21: CCV Database veriseti üzerinde tek modalite ve füzyon sonuçlarının karşılaştırılması

	Tek Modalite Sonuçları			Önerilen Füzyon Sonuçları	
	SIFT Tabanlı Görsel Modalite	STIP Tabanlı Görsel Modalite	MFCC Tabanlı Ses Modalitesi	Multimodal Füzyon SIFT+STIP +MFCC	Konsept ilişkilerini Kullanan Multimodal Füzyon
Basketball	66.95%	63.37%	44.65%	75.37%	<b>84.82%</b>
Bird	17.40%	14.12%	17.63%	29.46%	<b>47.58%</b>
Graduation	31.58%	22.09%	12.44%	40.88%	<b>57.64%</b>
Birthday	33.32%	15.38%	35.94%	53.33%	<b>70.87%</b>
WeddingReception	18.65%	22.54%	12.41%	26.23%	<b>34.70%</b>
WeddingCeremony	35.20%	32.88%	35.04%	51.78%	<b>66.86%</b>
WeddingDance	56.67%	47.61%	28.00%	63.75%	<b>72.09%</b>
MusicPerformance	48.19%	37.75%	56.71%	67.98%	<b>73.93%</b>
NonMusicPerformanc e	45.21%	53.23%	29.79%	61.95%	<b>66.15%</b>
Parade	48.70%	39.19%	25.62%	62.61%	<b>81.12%</b>
Beach	69.99%	47.50%	37.34%	73.89%	<b>77.49%</b>
Baseball	40.29%	18.38%	9.17%	47.00%	<b>70.27%</b>
Playground	44.59%	30.27%	23.83%	55.40%	<b>68.00%</b>
Soccer	49.27%	39.17%	17.59%	56.59%	<b>69.30%</b>
IceSkating	81.18%	65.82%	16.18%	84.08%	<b>90.67%</b>
Skiing	76.85%	60.26%	29.74%	76.91%	<b>86.87%</b>
Swimming	68.84%	53.80%	15.35%	71.56%	<b>79.13%</b>
Biking	36.84%	23.52%	11.36%	42.14%	<b>62.70%</b>
Cat	34.24%	23.82%	17.40%	42.15%	<b>55.51%</b>
Dog	25.48%	27.64%	22.10%	41.24%	<b>57.31%</b>
MAP	46.47%	36.92%	24.91%	56.21%	<b>68.65%</b>
MAP Derecesi	4	5	6	2	1

Tabloda görüldüğü üzere önerilen füzyon yaklaşımı sonuçları büyük oranda iyileştirmektedir. Genel olarak en iyi modaliteden elde edilen %46.47 değerindeki tespit başarısı füzyon modülü aracılığıyla %68.65'e çıkmıştır. Füzyonun işlemi tüm konseptlerin tespitinde olumlu rol oynamaktadır. Örneğin; ses modalitesinden %35.9 başarı ile çıkarılan Birthday (Doğumgünü) konsepti önerilen füzyon yöntemiyle %70.87 başarı ile tespit edilebilmiştir. Bunun dışında füzyon aşamasında diğer konsept tespit skorlarından faydalanmanın pozitif etkisi de deney sonuçlarında gözlemlenmektedir. Örneğin; Soccer (Futbol) konseptinin multimodal füzyon işlemi ile tespit başarısı %56.59 iken; Gol, Oyuncu, Skor, Top gibi başka konseptlere ait bilgilerin de füzyon işlemine katılması Soccer konseptinin tespit başarısını %69.3'e çıkarmıştır.

### 3.6 Video Dizin Yapısı (6. İş Paketi)

Videolar yapıları gereği oldukça büyük boyutlu dosyalardan oluşmaktadır ve çok miktarda anlamsal içeriğe sahiptir. Bu nedenle yapılacak sorgulamaları hızlandırmak amacıyla bir dizin/indeks yapısının geliştirilmesi büyük önem taşımaktadır. Video veritabanı sorguları bazen doğrudan anlamsal içerikle yapılabileceği gibi bazen de örnek nesnelere (Query By Example/QBE) yapılabilmektedir. Dahası, sorgular bazen her ikisini de içerebilir. Örneğin "Verilen resimdeki aracın hareket halinde olduğu görüntüleri getir" şeklinde verilen bir



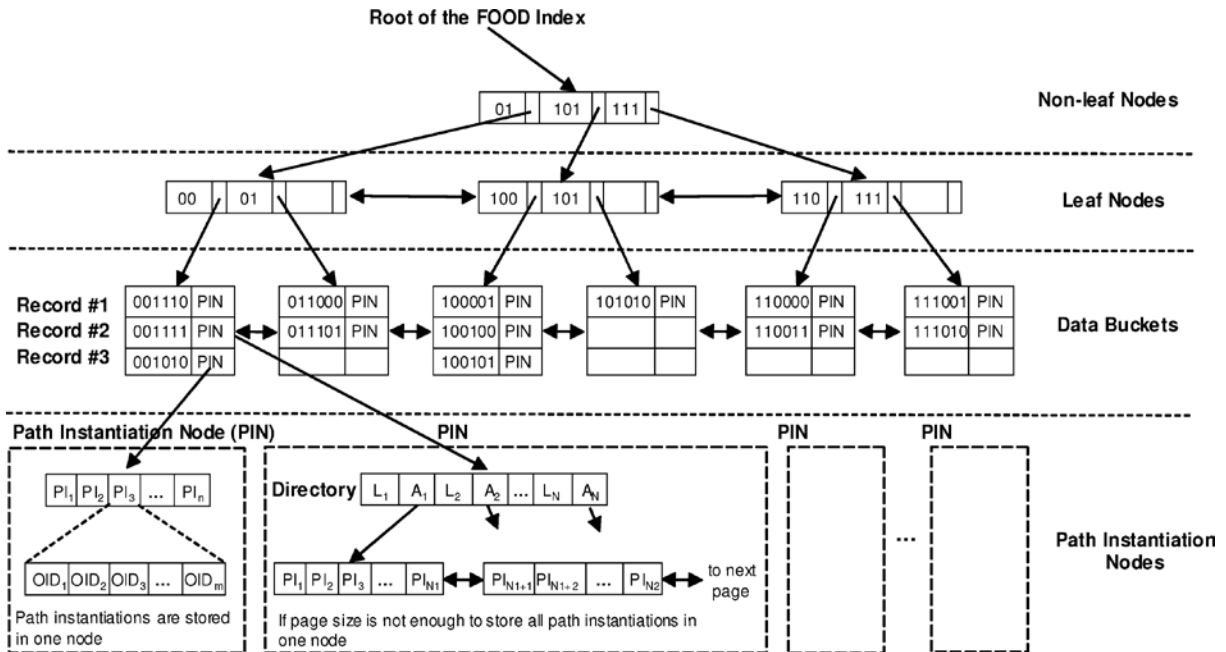
sorguda araç resim olarak sorguya konulmuş durumdadır. Böyle bir sorguda, aracın bulunduğu sahneler aracın alt seviye özellikleri kullanılarak belirlenebilir. Aracın hareket halinde olması ise anlamsal bir içeriktir ve bir olay olarak çıkarılıp veritabanında tutulması gerekir. Dolayısıyla bu tür sorgulara hızlı cevaplar üretebilmek var olan klasik indeksleme yöntemleriyle mümkün değildir. Bu nedenle video veritabanları için alt seviye veriler ile anlamsal verilerin birlikte indekslenebileceği çoklu boyutu destekleyen bir erişim yöntemine ihtiyaç vardır. Proje kapsamında alt seviye içerikler (content) ile anlamsal verileri (concept) birlikte kullanan bir indeks yapısı geliştirilmiştir. Burada içerik, görsel (dominant colors, color layout, edge histogram gibi) ve işitsel (audio spectrum projection, audio spectrum spread, audio spectrum flatness gibi) verilerden çıkarılan alt seviye özellikleri (descriptors) tanımlanmaktadır.

Amaçlanan dizin yapısını gerçekleştirebilmek için iki indeks yapısı birleştirilerek çoklu ortam verilerinin hem anlamsal hem de içerik bağlamında indekslenmesi sağlanmıştır. Anlamsal verileri indekslemek için FOOD-İndeks [18], içerikleri indekslemek içinse X-Tree [19] yapıları entegre bir şekilde kullanılmıştır. Bu iki yapı yaprak düğümleri (leaf nodes) seviyesinde birbirine bağlanmak suretiyle birleştirilmiştir. Bu sayede sorgu türlerine göre indeks yapıları arası erişim mekanizmaları sağlanmıştır. Aşağıda bu yapılar ve yaklaşımlar özetlenmiştir.

### FOOD-İndeks:

FOOD-İndeks yapısı çok boyutlu verileri indekslemek için uygun bir dizin yapısıdır. Şekil-15'de görüldüğü gibi iki farklı kısımdan oluşur:

Yol Dizini (Path Index): İlgili çoklu ortam nesnesine ve o nesneyi içeren diğer tüm nesnelere erişmek için kullanılan kısımdır. Bu kısımda G-Tree altyapısı kullanılmıştır. G-Tree ikili ağaç yapısına sahip olup, FOOD-İndeks kapsamında ikili verilerin dizinlenmesinde ve aranmasında kullanılmaktadır. G-Tree üzerinde arama yapabilmek için, ilgili çoklu ortamın dizinlenmek istenilen bilgisi ya da bilgileri önce ikili karakter dizilerine çevrilir (01101001 gibi). Daha sonra o çoklu ortam nesnesi G-Tree'ye dâhil edilir.



Şekil-15: FOOD-İndeks yapısı

Veri Yığınları (Data Bucket): Çoklu ortam nesnelere ait ikili karakter dizilerinin ve bu nesnelere ait olan yol bilgilerinin tutulduğu veri yapısıdır. Bu veri yapısı sayesinde bir sorgu sonucunda istenen tüm yollar kolayca erişilmektedir.

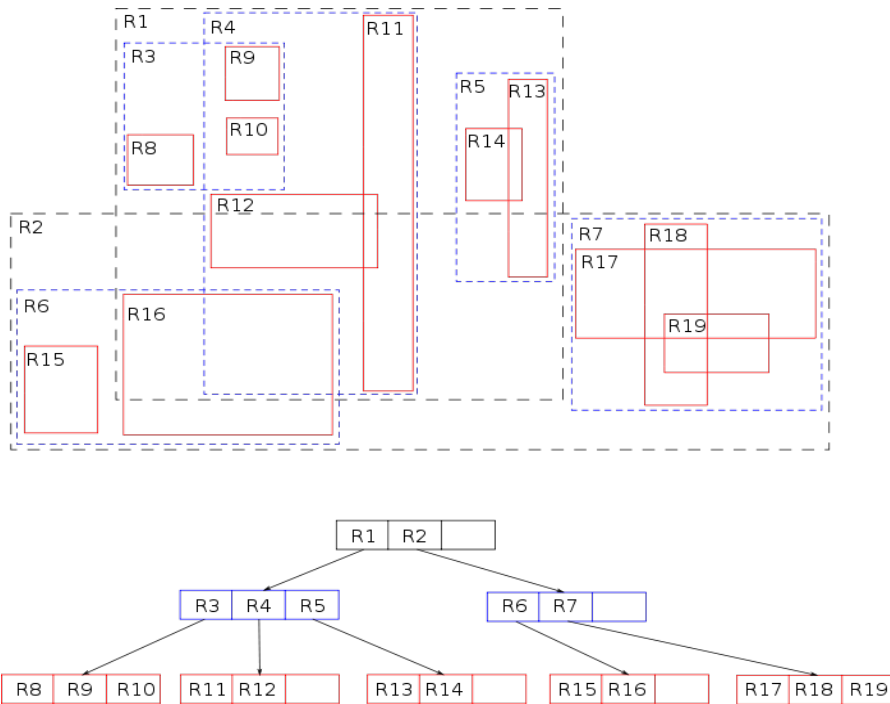
FOOD-İndeks yapısının avantajları şöyle sıralanabilir;

- Çok boyutlu verileri dizinlemekte etkin bir yoldur.
- Nesne tabanlı veritabanlarında kolaylıkla kullanılabilir.
- Belirli mantık çerçevesindeki sorgulara cevap verdiği gibi bulanık mantık (fuzzy logic) sorgularına da etkili bir şekilde yanıt verebilmektedir.
- Çok boyutlu verilerin birden fazla bilgisini dizinlemek amacıyla oluşturulduğu için ölçeklenebilir, yani dizinlenecek değişken sayısı kolaylıkla değiştirilebilir (örneğin sadece ad-soyad üzerinde dizin oluşturulabileceği gibi, ad-soyad-yaş-boy bilgileri hep birlikte dizinlenebilir).
- İkili karakter dizileri kullanıldığı için hızlıdır.

Bu projede FOOD-İndeks yapısı, çoklu ortam nesnelere ait anlamsal bilgilerinin dizinlenmesi için kullanılmaktadır. Çoklu ortam nesnelere ait görüntüsel özellik verilerinin (low level features/contents) dizinlenmesi için X-Tree çok boyutlu konumsal dizinleme yapısı kullanılmaktadır.

### X-Tree:

R-Tree ailesinden gelen X-Tree ağacı uzaysal nesnelere indekslenmesinde kullanılan bir index yapısıdır. X-Tree dizin yapısı, B-Tree altyapısını kullanır. B-Tree’de bilgi, belirli kuralları sağlayan dengeli bir ağaca yerleştirilirken, R-Tree’de nesnelere benzer kuralları sağlayan dengeli bir ağaç oluştururlar.



Şekil-16: Örnek X-Tree yapısı

Ağacın ilk seviyesinde, minimum çevreleyen dikdörtgen (minimum bounding rectangle, MBR) adı verilen ve indeksin kullanımında filtreleme görevi yapan dikdörtgenler oluşturulur.

Böylece düzenli bir şekilde olmayan nesnelere yerine, onları çevreleyen minimum dikdörtgenler üzerinde işlem yaparak, indeks performansı önemli oranda artırılmış olur. Örnek bir X-Tree yapısı Şekil-16'da gösterilmiştir.

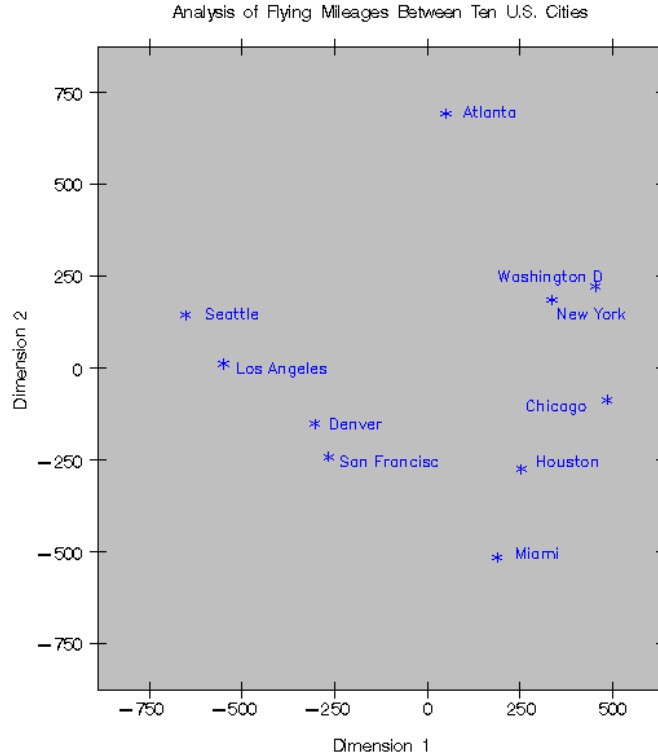
X-Tree yapısı konumsal bilgileri indekslediği için, çoklu ortam verilerinin nasıl X-Tree ile indeksleneceği konusu önemlidir. Bu bağlamda çoklu ortam verilerinden çıkarılan öznelik vektörleri (feature vectors) kullanılır. Bu vektörler sayesinde çoklu ortam verileri matematiksel olarak tanımlanmış olur.

```

data city;
  title 'Analysis of Flying Mileages Between Ten U.S. Cities';
  input (atlanta chicago denver houston losangeles
        miami newyork sanfran seattle washdc) (5.)
        @56 city $15.;
  datalines;
    0
    587 0
    1212 920 0
    701 940 879 0
    1936 1745 831 1374 0
    604 1188 1726 968 2339 0
    748 713 1631 1420 2451 1092 0
    2139 1858 949 1645 347 2594 2571 0
    2182 1737 1021 1891 959 2734 2408 678 0
    543 597 1494 1220 2300 923 205 2442 2329 0
  ;

```

Şekil-17: Mesafe tablosu



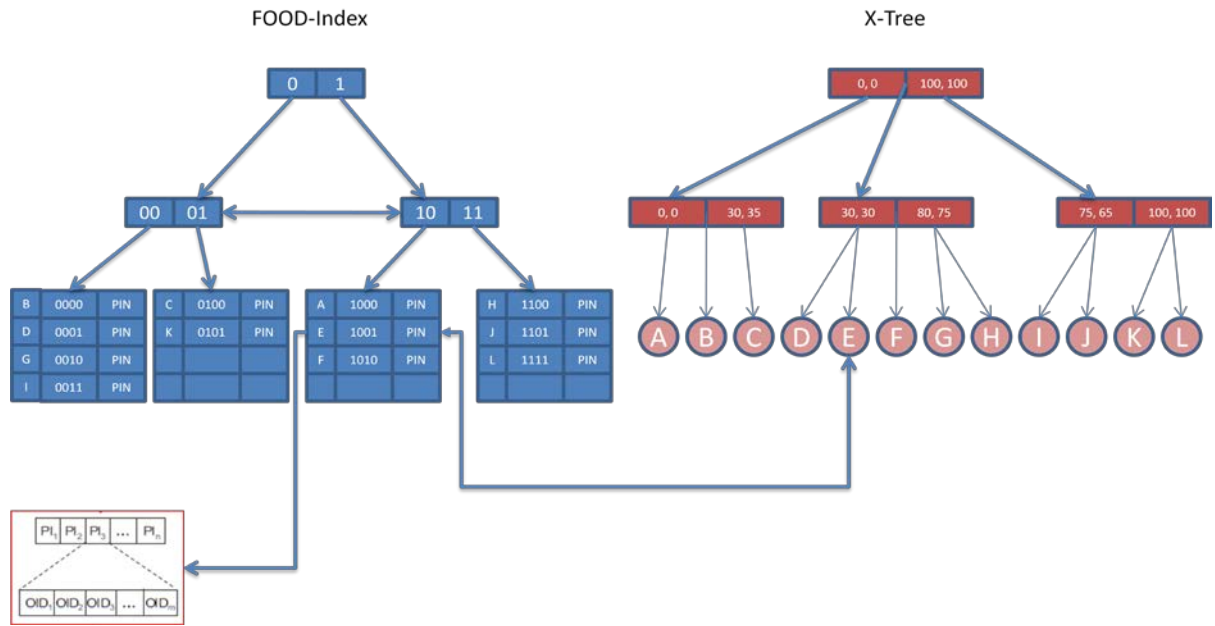
Şekil-18: LMDS sonucunda konumsal yerleştirme

Bu noktada, öznelik vektörlerin boyutları sorun teşkil etmektedir. Çünkü X-Tree yapısı çok boyutlu durumlarda performans sıkıntısı çekmektedir. Bu sorunu aşmak için ise çok boyut ölçeklendirme (Multidimensional Scaling) yaklaşımı kullanılmıştır. Çok boyut ölçeklendirme işleminde esas amaç sayıca çok fazla olan boyutlardan daha az sayıda boyutlara indirgeme

yapmaktır. Bu sayede orijinal veri dağılımı çok bozulmadan etkin erişim yapılması hedeflenmektedir. Bu çalışmada Landmark tabanlı çok boyut ölçeklendirme (LMDS) yaklaşımı kullanılmıştır. Yapılan testler neticesinde, LMDS kullanan bir dizin yapısının çoklu ortam verilerinin dizinlenmesi için etkin bir yöntem olduğu sonucuna varılmıştır.

LMDS yaklaşımında ihtiyaç duyulan, nesnelararası mesafelerin bulunduğu bir matristir. Normal çok boyut ölçeklendirme yaklaşımına göre farkı ise, tüm nesnelararası mesafeler yerine belirli ve az sayıda nesnelararası mesafeler seçilerek (landmarks) çok boyut ölçeklendirme yapılmasıdır. Böylelikle performans kazancı elde edilmiş olur. Örnek olarak, sadece şehirlerarası mesafelerin bulunduğu bir ortamda (Şekil-17), bu şehirlerin birbirine olan konumları LMDS sayesinde gerçeğine çok yakın şekilde tanımlanabilir (Şekil-18).

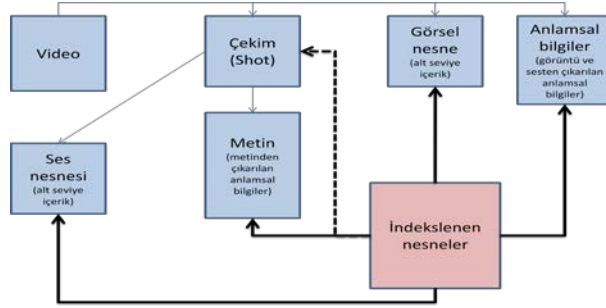
Bu projede ise, X-Tree yapısının ihtiyaç duyduğu koordinat bilgileri, çoklu ortam verilerinden çıkarılan öznelik vektörlerinden elde edilmektedir. Bu süreçte öznelik vektörleri kullanılarak nesnelararası mesafeler hesaplanmış ve bu mesafe tablosu LMDS yaklaşımına girdi olarak verilerek konumsal koordinatlar elde edilmiştir. Son aşamada da X-Tree yapısı bu koordinatlar kullanılarak oluşturulmuştur.



Şekil-19: Multimedia dizin yapısı

Proje kapsamında gerçekleştirilen nihai dizin yapısı Şekil-19'da verilmiştir. Anlamsal verilerin indekslenmesi için yine bu projenin araştırmacıları tarafından geliştirilmiş olan FOOD indeks yapısı, alt seviye içeriklerin indekslenmesi için de X-Tree indeks yapısı kullanılmıştır. Yaprak nodlar seviyesinde iki dizin yapısı birbirleriyle ilişkilendirilerek her iki dizin yapısından da ilgili video çekim bilgilerine erişim mümkün hale getirilmiştir. İndeksleme mantığı Şekil-20'de görsel olarak verilmiştir. Bir video çekimlere ayrılmaktadır. Her çekim üzerindeki aşağıdaki bilgiler çıkarılmakta ve indekslenmektedir.

- Görsel verilerden çıkarılan anlamsal bilgiler (FOOD)
- Ses verilerinden çıkarılan anlamsal bilgiler (FOOD)
- Metin verisinden çıkarılan anlamsal bilgiler (FOOD)
- Görsel nesnelararası mesafelerin alt seviye özellik bilgileri (X-Tree)
- Ses nesnelararası mesafelerinin alt seviye özellik bilgileri (X-Tree)



Şekil-20: Multimedia dizin yapısı

### **Sorgularda indeks yapısının kullanılması:**

Geliştirilen dizin yapısı hem anlamsal veriler hem de içerik verileri üzerinden sorgulama yapma imkanı sağlamaktadır. Sorgulama sonucunda dönen cevaplar ile ilgili video çekimlerine hızlıca ulaşarak çekimler izlenebilmektedir. Temel olarak üç farklı sorgu tipinde dizin yapısının kullanıldığını söylemek mümkündür:

#### **1. İçerik sorgulamaları (query by content)**

Çoklu ortam veritabanlarında, örnek nesnelere (Query By Example/QBE) sorgulama yapılması önemli bir ihtiyaçtır. “örnek olarak verilen resimdeki arabaya benzer arabaların olduğu video çekimlerini getir” türünden sorgulamalar yapılabilmektedir. Geliştirilen dizin yapısı X-Tree aracılığıyla bu tür sorgulamaların etkin olarak cevaplanmasını sağlamaktadır.

Bu tür bir sorgulama aşağıdaki adımlarla gerçekleştirilmektedir:

- Öncelikle, verilen görsel nesnenin (sesli sorgulamalarda ses nesnesinin) alt seviye özellikleri çıkarılır ve LMDS kullanılarak koordinatı hesaplanır.
- Arkasından, aranan nesneye X-Tree ile dizin yapısı üzerinden ulaşılır.
- Sorgu şartlarını sağlayan nesnelere cevap listesine eklenir.
- Şayet cevap nesnelere (sadece nesnenin kendisi değil de) tüm erişim yolu gerekiyorsa, X-Tree ile FOOD indeks arasındaki bağlantılar kullanılarak ilgili nesnelere PI'ları döndürülür.

#### **2. Anlamsal verilerin sorgulanması (query by concept)**

Diğer bir sorgu tipi nesnelere öznelikleri kullanılarak yapılan sorgulardır. Öznelikler normal metin türünde değerlere sahiptir ve FOOD İndeks yapısı üzerinden bu tür nesnelere ulaşmak mümkündür. Bu tür sorgularda izlenen adımlar şöyledir:

- Sorgu nesnesi için bir bit dizini oluşturulur.
- FOOD index üzerinden ilgili nesneye ulaşılır ve nesnenin PI bilgisi cevap olarak döndürülür.

Bu sorgu türünde X-Tree tarafı hiç kullanılmaz.

#### **3. Anlamsal veri ve içeriğin birlikte sorgulanması (Query by concept and content)**

Geliştirilen dizin yapısının en büyük avantajı hem anlamsal verinin hem de alt seviye içeriğin birlikte sorgulanabilmesidir. Şayet sorgu, her iki veri tipinde de şartlar (condition) içeriyorsa, bu durumda hem FOOD indeks hem de X-Tree indeks yapısının kullanılması gerekmektedir. Bu tür sorgularda, öncelikle FOOD indeks kullanılarak aday nesnelere çıkarılmakta, daha

sonra X-Tree üzerinden gidilerek içerikle ilgili şartları da sağlayanlar belirlenmekte ve kesin cevap üretilmektedir.

### **3.7 Veritabanı Entegrasyonu (7. İş Paketi) ve Prototip Sistemde Geline Son Durum**

Proje kapsamındaki ana modüller farklı iş paketleri olarak planlanmış ve ayrı ayrı geliştirilmiş ve test edilmiştir. Bu modüller aşağıda listelenmiştir:

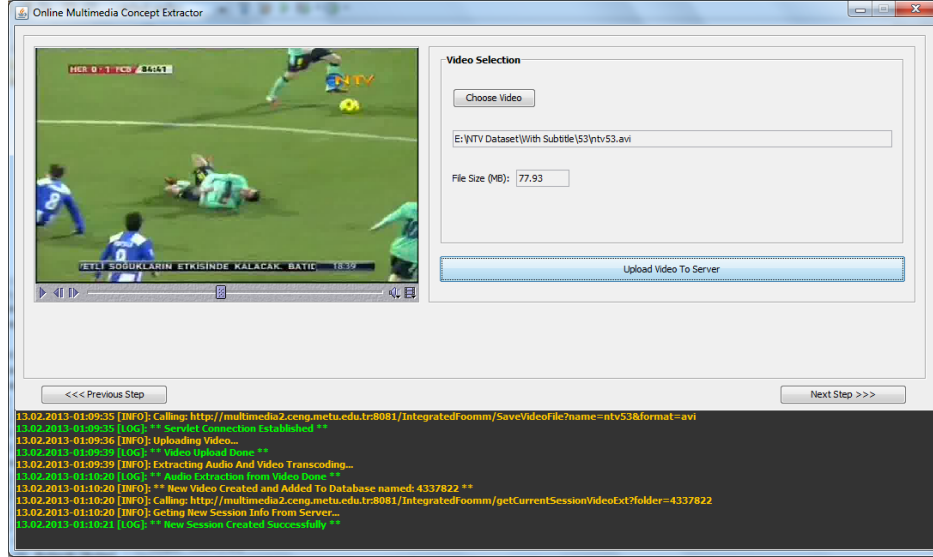
- Görüntüden anlamsal bilgi çıkarma modülü (görsel anotasyoncu) (2.iş paketi)
- Metinden anlamsal bilgi çıkarma modülü (metinsel anotasyoncu) (3.iş paketi)
- Sesten anlamsal bilgi çıkarma modülü (işitsel anotasyoncu) (4.iş paketi)
- Bilgi füzyon modülü (5.iş paketi)
- Veritabanı/Depolama modülü (daha önce geliştirilmişti)
- Dizin modülü (6.iş paketi)

Modüllerle ilgili çalışmaların tamamlanmasından sonra, bu modüllerin birleştirilerek tüm işlemlerin tek bir kullanıcı arayüzünden yönetilebilir bir system haline getirilmesine yönelik çalışmalar yapılmıştır. Bu çalışmalar sonucunda oluşturulan sistemin mimarisi yukarıda Şekil-1'de verilmiş ve oluşturulan mimari yapı Genel Bilgiler başlığı altında zaten açıklanmıştı. Verilen mimari yapı içerisinde, çıkarılan verilerin tutulduğu Akıllı ve Bulanık Nesneye Dayalı Veritabanı, proje araştırmacılarının daha önce TÜBİTAK 1001 - 106E012 Projesi kapsamında geliştirdikleri bir üründür [20]. Akıllı ve Bulanık Nesneye Dayalı Veritabanı modeli, bulanık nesneye dayalı bir veritabanı ile bulanık nesneye dayalı bilgi tabanının entegre edilmesi ve bunların üst seviyede ortak bir dil ile kullanılabilmesi temeline dayanmaktadır. Modelin alt seviyesinde nesneye dayalı veritabanı, büyük boyutlardaki verilerin depolanması ve sorgulanması için, nesneye dayalı bilgi tabanı ise uygulama alanı ile ilgili çıkarım kurallarının tutulması ve işlenmesi için kullanılmaktadır. İki sistem, köprü olarak isimlendirilen bir yapı ile entegre edilmiştir. Köprü, aynı zamanda kullanıcıyı modelin karmaşık yapısından soyutlayarak veri ve bilgi yönetiminin ortak bir dil ile sorgulanmasına imkan sağlamaktadır. Bu proje kapsamında, bir XML veritabanı da eklenerek üzerinde denemeler yapılmıştır.

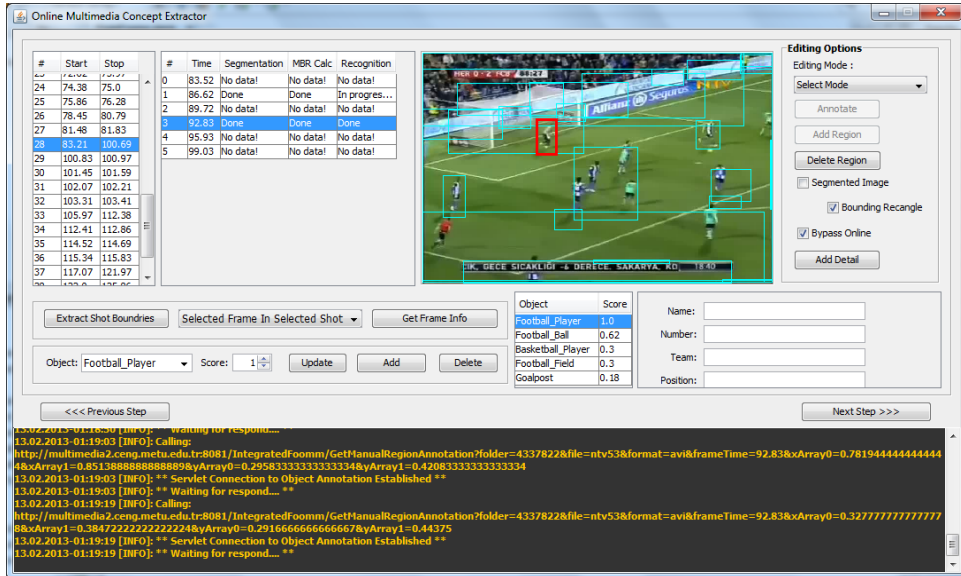
Çıkarılan anlamsal bilgiler veritabanında dört ayrı veri grubu olarak tutulmuştur. Bunlar, görsel, işitsel ve metinsel verilerden elde edilen sonuçlar ile füzyon işlemi sonucunda elde edilen sonuçlardır. Buradaki amaç, örneğin sadece görüntüden anlamsal bilgi çıkarımı sonuçları üzerinde veya sadece sestenden anlamsal bilgi çıkarımı sonuçları üzerinde sorgular yapma imkanı sağlamak ve sonuçları ayrı ayrı görmek ve değerlendirebilmektir. Başlangıçta sadece füzyon işlemi sonucunun veritabanında tutulması planlanmış olmasına rağmen, gelinen aşamada tüm çıkarım sonuçlarının ayrı ayrı tutulmasının sistemin değerlendirilmesi açısından daha iyi olacağı sonucuna varılmıştır. Sonuçta da bu uygulamanın çok daha esnek sorgulama imkanı sağladığı görülmüştür.

Geliştirilen sistemi yönetmek üzere, iki farklı arayüz oluşturulmuştur. Bunlardan birincisi yeni bir videonun sistem yüklenmesi ve anotasyon işlemleri için kullanılmaktadır. İkincisi ise sorgulama işlemleri için geliştirilmiştir.

Şekil-21, Şekil-22 ve Şekil-23'te birinci arayüz ile ilgili örnekler verilmiştir. Şekil-21'de anotasyonu yapılacak bir videonun ilk adım olarak sisteme yüklenmesi yapılmaktadır. Bundan sonraki anotasyon adımları videonun sisteme yüklenmesinden itibaren başlamaktadır.



Şekil-21: İlk video yükleme ekranı

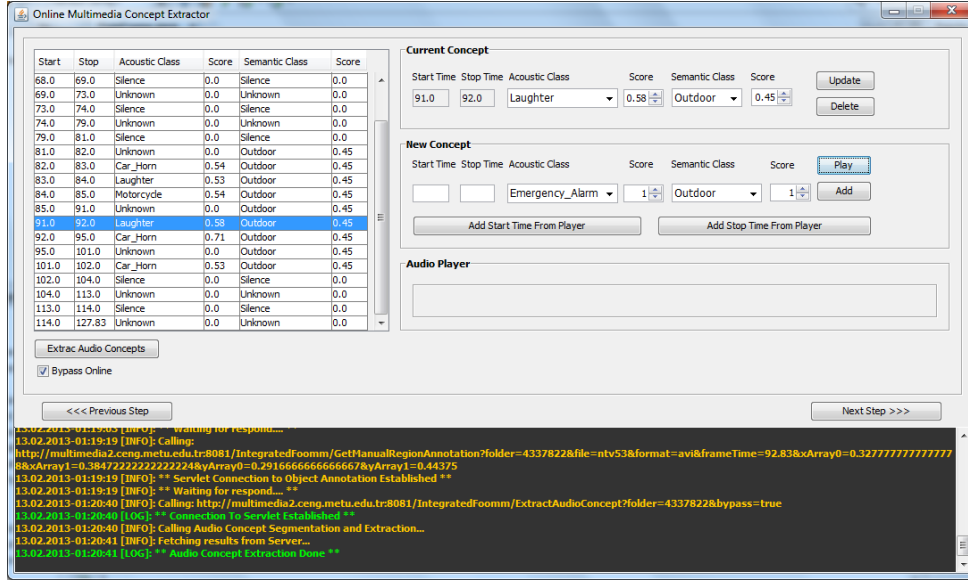


Şekil-22: Görsel anotasyon ekranı

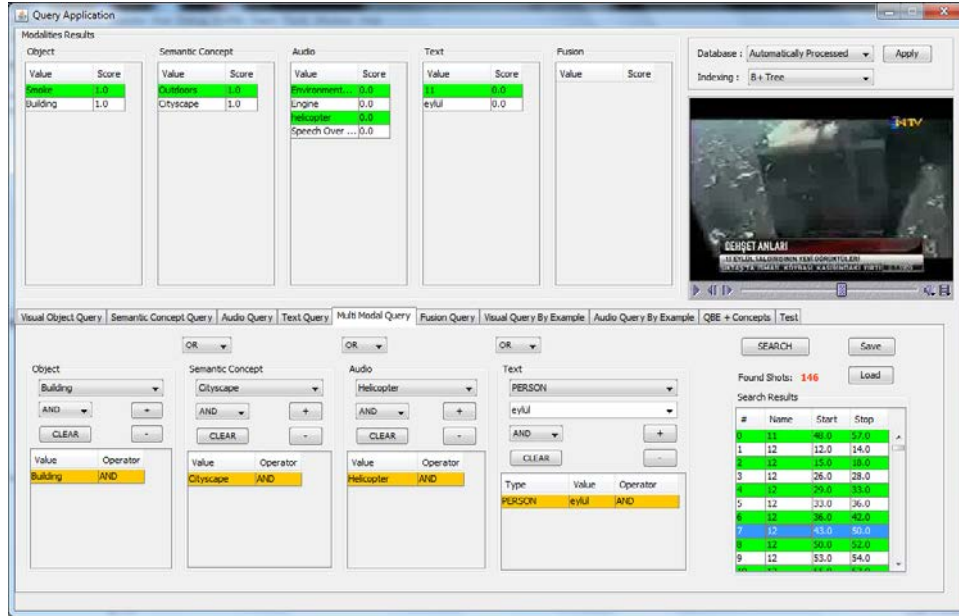
Şekil-22'de görsel anotasyon işlemi görülmektedir. Otomatik anotasyon sonucunda çıkarılan segmentler ve seçilen segmentin nesne olarak nasıl sınıflandırıldığı ekranda görülebilmektedir. Kırmızı çerçeve içerisindeki nesne otomatik olarak 1.0 skoru ile "football\_player" olarak sınıflandırılmıştır (görüntünün hemen altında). İlk beş otomatik sınıflandırma sonucu ekranda görülebilmektedir. Kullanıcı, bu ekran üzerinde isterse otomatik sınıflandırma sonucunu değiştirebilir (hatalı sınıflandırmalarda bu gerekli olabilir), ilave öznitelikler girebilir ve sonucu veritabanına kaydedebilir.

Şekil-23'te ise sestten anlamsal bilgi çıkarımı işlemi görülmektedir. Bir video çekiminin ses verileri daha küçük parçalara bölünerek akustik sınıflara ayrılmaktadır. Bir video çekimine ait ses verisinin segmentleri, bu segmentlerin başlangıç ve bitiş zamanları, segmentin otomatik sınıflandırma sonucunda belirlenen akustik sınıfı, sınıfın belirlenmesindeki skoru ekranda görülebilmektedir. Akustik sınıflara bağlı olarak çıkarılabilen daha üst seviye anlamsal bilgiler varsa (semantic class) onlarda ekranda görülmektedir. Görselde olduğu gibi çıkarılan

sonular, kullanıcı tarafından doğrudan veritabanına kaydedilebilir veya deęişiklikler yapılarak (hatalar arındırılarak) kayıt işlemi tamamlanabilir.



Şekil-23: Ses anotasyon ekranı



Şekil-24: Sorgu ekranı

Sorgu arayüzü ile ilgili bir sorgu ekranı da Şekil-24'te verilmiştir. Arayüzde farklı sorgu tipleri için farklı sekmeler oluşturulmuştur. Verilen örnekte çoklu modalite (multi-modal query) ile sorgu yapma ekranı görülmektedir. Ekranda dört farklı sorgu şartı girilebilmektedir. Ekranda görülen sorgunun doğal dilde yazılmış hali şöyledir:

Görüntüden “bina (building)”, sestten “helikopter”, metinden “eylül” ve konsept olarak (görüntü ve ses verilerinden çıkarılan üst seviye anlamsal bilgi) “şehir görüntüsü (cityscape)” bilgilerini içeren video çekimlerini (shot) getir.



Sorgu sonucu ekranın sağ-alt köşesinde listelenmektedir (yeşil-beyaz satırlar). Buradaki her satır, cevap olarak dönülen bir video çekimidir (shot). Bunlardan istenenin üzerine tıklanarak (mavi satır) görüntü örnekte olduğu gibi izlenebilmektedir.

### **3.8 Farklı Alan Uygulamaları ve Testler (8. İş Paketi)**

Geliştirilen prototip sistem yukarıda anlatıldığı üzere, modüller halinde geliştirilmiş ve her modül bulunabilen uygun veri setleri ile test edilmiştir. Bu testlerle ilgili sonuçlar zaten yukarıda ilgili bölümlerde verilmiştir. Bu iş paketinde ise geliştirilen sistemin entegrasyon sonrasında bir bütün olarak test edilmesi amaçlanmıştır.

Tüm sistemi test edebilecek herkese açık olan uygun hazır bir veri seti (görüntü, ses ve metin açısından ortak kullanılabilir) bulunamamıştır. Bu nedenle, kendi veri setimizi oluşturarak geliştirmiş olduğumuz sistem üzerinde çok sayıda testler yapılmıştır. Test verilerinin oluşturulmasında NTV haber kanalı yayınlarından alınan haber videoları kullanılmıştır. Oluşturulan veri setinin özellikleri şöyledir:

- 76 video
- 1571 çekim (shots)
- 4428 görsel nesne (visual objects)
- 2428 işitsel nesne (audio objects)
- 2796 metin nesnesi (text objects)
- 7719 üst seviye anlamsal konsept (semantic concepts)

Çıkarılan bu veri seti öncelikle sistemin ilgili modüllerinin eğitiminde kullanılmıştır. Örneğin, görsel nesnelere ile Görsel Anotasyoncu Modülünün eğitimi yapılmıştır. Diğer taraftan, bu veriler veritabanına indekslenerek girilmiştir. Dolayısıyla, index ve sorgulama modüllerinin test edilmesinde bu verilerden yararlanılmıştır.

Ayrıca, Anlamsal Bilgi Çıkarım Sisteminin bir bütün olarak testi için ise yine NTV haber videolarından kısa videolar seçilmiştir. Bu videolar siteme yüklenerek baştan sona anotasyon işlemleri otomatik olarak denenmiştir. Otomatik anotasyon sonucunda çıkarılan anlamsal bilgiler veritabanına eklenerek üzerinde değişik sorgular yapılmıştır. Otomatik anotasyon sonucunda oldukça anlamlı sonuçlar çıkmakla birlikte gürültü olarak niteleyebileceğimiz sonuçlar da çıkabilmektedir. Videolarda otomatik anotasyonun zorluğu düşünüldüğünde, bu tür sonuçların çıkması doğal kabul edilmektedir. Sistemin bir bütün olarak performansının ölçülmesi arzu edilmesine karşılık, bu amaçla kullanabileceğimiz nitelikte hazır bir veri setinin olmaması bu konuda sorun oluşturmuştur. Kendi ürettiğimiz veri seti ise, sistemin doğru çalışırılığının gösterilmesi (prove-of-concept) açısından yeterli olmuştur. Ancak, daha ciddi bir performans ölçümü yapılması açısından çok daha büyük bir veri setinin gerekli olduğudeğerlendirilmiştir.

Sonuç olarak, proje sonunda, Şekil-1 de verilen mimari yapı bir prototip sistem olarak başarıyla gerçekleştirilmiş ve doğru çalışırılığı yapılan testlerle gösterilmiştir.

## 4. BULGULAR

1. Geliştirilen prototip sistem ile videoların görüntü, ses ve metin verileri kullanılarak (multi-modal) anlamsal bilgilerin otomatik olarak çıkarılması, uygun formatlarda saklanması ve daha sonra da etkin bir şekilde sorgulanabilmesi konularında bir konsept ispatlama (proof of concept) çalışması yapılmıştır. Yapılan çalışma, bir video ile ilgili bilgi çıkarımı, çıkarılan bilginin depolanması ve en sonunda da sorgulanması süreçlerinin tamamını kapsayan bütün bir sistemi ortaya koyması açısından önemlidir. Proje sonunda ulaşılan sonuçlar, bu konuda doğru adımlar atıldığını göstermektedir.
2. Videoların içindeki nesnelere, bu nesnelere gerçekleştirdiği olaylar, nesnelere birbirleriyle olan mekansal ilişkileri, olayların birbirleriyle olan zamansal ilişkileri gibi anlamsal verilerin çıkarılması günümüzde önemli bir gereksinim haline gelmiştir. Anlamsal verilerin çıkarımında ise videonun sağladığı görsel, işitsel ve metinsel tüm kaynakların kullanılması daha başarılı sonuçlar alınmasını sağlayacaktır. Bu çalışmada elde edilen sonuçlar, videoda mevcut olan üç değişik veri tipinin birlikte kullanılması durumunda bilgi çıkarımındaki başarı oranının artacağını göstermektedir.
3. Videolardan anlamsal bilgi çıkarımında görsel, işitsel ve metinsel verilerin tamamının kullanılması başarı oranını artırmaya karşın, ana veri kaynağının görsel veri olduğu sonucuna varılmıştır. Ses ve metin verileri, görsel veriden elde edilen sonucun başarısını yükseltmek veya görsel veriden elde edilemeyen bilgilerin çıkarılması amacıyla kullanılabilir. Görsel verinin hiç dikkate alınmadığı durumlar ancak bazı özel uygulamalarda ve sorgularda iyi sonuçlar verebilmektedir.
4. Geliştirilen sistemin performans ölçümlerinde uluslararası makale ve bildiri yayınlanma şansını yükseltmek amacıyla var olan ve bilinen veri setlerinin kullanılması tercih edilmiştir. Ancak, her üç modaliteyi içeren ortak kullanılabilir bir veri seti bulunamamıştır. Bu nedenle, mümkün olan yerlerde uluslararası bilinen veri setleri, mümkün olmadığı durumlarda kendi ürettiğimiz video veri setleri ile performans ölçümleri yapılmıştır. Bu nedenle, sistemin her modülünün performansını baştan sona aynı veri setini kullanarak bir bütünlük içerisinde ölçen bir performans ölçüm çalışması yapılamamıştır. Yine de tüm sistemin performansını aynı veri seti ile ölçebilmek için , kendi çalışmalarımızla sınırlı miktarda bir veri seti üretilebilmiş ve bu veri seti hem sistemin doğru çalışırılığını göstermek (proof of concept) ve hem de sınırlı da olsa performans amaçlı kullanılabilmiştir. Her üç modalitedeki anlamsal içerik çıkarımının ve sistemin bütününe performansını daha sağlıklı ölçmek için her üç modaliteyi de içeren ve sayısal olarak da bizim bu çalışmada ürettiğimizden daha büyük boyutlara ulaşan bir veri setine ihtiyaç olduğunu belirtmekde yarar vardır.
5. Projenin "görüntüden anlamsal bilgi çıkarımı" kısmında, bir çok uygulama alanında (örneğin güvenlik/izleme, spor, haber videoları, vs.) kullanılabilir bir otomatik anlamsal bilgi çıkarımı sistemi geliştirilmiştir. Bu doğrultuda, geliştirilen sistemde, uygulama alanından bağımsız olarak bir ontoloji temelli anlamsal içerik meta-ontolojisi ve özel kural tanımları yapılmış; bu meta-ontoloji ve kural tanımları kullanılarak ihtiyaç duyulan uygulama alanları için gerekli ontolojilerin üretilmesi kolaylaştırılmış ve standartlaşma yolunda bir adım atılmıştır.
6. Geliştirilen otomatik anlamsal bilgi çıkarımı uygulaması, mevcut anlamsal video modelleme ve anlamsal içerik çıkarımı çalışma alanlarına çeşitli açılardan katkı sağlamıştır. Bunlardan ilki, anlamsal bilgi çıkarımı işleminin tamamen otomatik olarak yapılabilmesidir. Ek olarak, videolar için genel bir ontoloji temelli anlamsal meta-ontoloji modeli (VISCUM) geliştirilmiştir. Ayrıca, sınıf, ilişki ve kural tanımlarına bulanıklılık (fuzziness) katılarak, anlamsal içerik temsil kabiliyeti ve çıkarım başarısının artırılması

sağlanmıştır. Mantıksal içeriğin temel taşı sayılan nesnelere tespit edilmesi için otomatik Genetik Algoritma temelli bir nesne çıkarımı yöntemi sisteme entegre edilmiştir.

7. Geliştirilen sistem ile yapılan nesne ve olay çıkarımı testlerinde başarılı sonuçlar elde edilmiştir. Yapılan testlerde, olay çıkarımının başarısının, doğrudan nesne çıkarımına bağlı olduğu, dolayısıyla nesne çıkarımındaki başarının çok önemli olduğu tespit edilmiştir. Bu doğrultuda, gerçekleştirilen uygulamada, genetik algoritma temelli bir yöntem kullanılmıştır. Ancak, genetik algoritma temelli yaklaşımın bazen yetersiz kaldığı gözlemlenmiştir. Bu gibi durumlarda sınıflandırma için en başarılı yöntemlerden biri olarak bilinen Destek Vektör Makineleri (SVM) yönteminden yararlanılması gerektiği sonucuna varılmıştır.
8. Genel ses sınıflandırma sistemlerinde ses verisi konuşma, müzik ve çevresel ses olarak adlandırılan genel ses kategorilerine ayrıştırılmaktadır. Konuşma ve müzik sesleri bazı tonal ve harmonik karakteristiklere sahiptir ve bu tür karakteristikler Zero Crossing Rate (ZCR), Mel Frequency Cepstral Coefficient (MFCC) ve Spectrum Flux (SF) gibi yaygın olarak kullanılan ses öznitelikleri ile uygun biçimde karşılanabilmektedir. Ancak, bu verilerin birlikte bulunduğu (müzik üzerine konuşma, çevresel ses üzerine konuşma vb.) ses kayıtlarının sınıflandırılması, farklı ses karakteristiklerini ifade eden alt düzeyli öznitelikler ile çalışılmasını gerektirmektedir. Çevresel sesler söz konusu olduğunda, bu seslerin yapısal olmayan doğası ve genellikle gürültü benzeri ve düz spektrum karakteristikleri nedeniyle, tanıma işlemleri daha karmaşık olmaktadır. Bu tür karakteristikler, çevresel seslerin tanınmasını insanlar için bile zor kılabilir. Örneğin, alkış ve yağmur sesleri farklı çevresel sesler olmasına rağmen, bu sesler insan algısı için bile benzerlikler içerdiğinden ayırt edilmesi zor olabilmektedir. Bu duruma, çevresel seslerin dinamik yapısı da etki etmektedir. Örneğin, farklı ortam ve/veya zamanlarda farklı sesler benzer olarak algılanabilmektedir. Çevresel ses sınıflandırmada diğer bir problem de bu seslerin çeşitliliğidir. Bu çeşitliliğin artması, otomatik tanıma sistemlerinin performansını önemli ölçüde düşürmektedir.
9. Bu çalışmada, oldukça zengin bir içeriğe sahip olan ses verilerinin hiyerarşik olarak sınıflandırılması hedeflenmiştir. Çok zengin ve farklı karakteristiklere sahip ses verilerinin tek bir öznitelik ile yeterince iyi ifade edilemediği gözlemlenmiştir. Bu nedenle, sesin farklı özelliklerini ortaya çıkaran farklı öznitelik kombinasyonlarının tanıma başarımlarını artırdığı görülmektedir. Deneysel sonuçlara göre, önerilen ASFCS-H özniteliği, hiyerarşik sınıflandırmada SVM sınıflandırıcısı ile oldukça iyi sonuçlar vermektedir. MPEG-7 ASF özniteliği, ses verisinin kısa-süreliliği güç spektrumunu oldukça iyi ifade edebildiğinden, genel ses tanıma uygulamalarında MFCC öznitelğine alternatif olabileceği gözlemlenmiştir. Diğer yandan, bazı çevresel seslerin (alarm, yağmur ve su) harmonik özellikler taşıdığı ve bu nedenle, harmonik ses özniteliğinin gürbüz spektral özniteliklerle (MPEG-7 ASF, MFCC) birleştirilmesiyle daha yüksek tanıma doğruluklarının elde edilebileceği düşünülmektedir.
10. TRECVID 2007 ve CCV Database gibi oldukça popüler datasetler üzerinde yapılan deneylerin gösterdiği üzere farklı kaynaklardan elde edilen verilerin bu çalışmada önerilen füzyon yaklaşımı ile birleştirilmesi, tüm modalite sonuçlarından (görüntü, ses ve metinden anlamsal bilgi çıkarımı yapan sistemlerin sonuçları) daha iyi sonuç vermektedir. Füzyon metodu içinde farklı konseptlere ait bilgilerden faydalanmak, diğer bir deyişle konseptler arasındaki ilişkiyi doğrudan füzyon işlemi içinde kullanmak da sonuçları önemli derecede iyileştirmiştir. Önerilen füzyon metodu sayesinde, TRECVID 2007 datasetinde en iyi sonucu veren modalitenin başarısı genel olarak %16.8 oranında artmıştır. Bunun yanı sıra CCV Database datasetinde füzyon işlemi aracılığıyla genel olarak %33.8 göreceli performance artışı (en iyi sonuç veren modaliteye göre -görüntü modalitesi-) gözlemlenmiştir. Yapılan deneylerde,

modalitelerin sonuçlarının birbirini tamamlayacak şekilde birleştirildiği sonucu çıkarılabilir. Örneğin; CCV Database datasetinde yer alan WeddingCeremony (düğün seramonisi) konsepti sadece görüntüden anlamsal bilgi çıkarımı yapılarak %35.2 başarı ile tespit ediliyorken, sestem elde edilen bilgilerle bu sonuç %51.78'e çıkmaktadır. Görüntü ve sese ek olarak metinsel bilgilerin kullanımı ile sonuç daha da artarak %66.86'ya çıkmaktadır. Her iki data setinde de, WeddingCeremony konseptinde olduğu gibi, konseptlerin büyük çoğunluğu için füzyon işleminin sonuçları önemli derecede arttırdığı ya da en iyi modalite sonucuna yakın sonuçlar verdiği gözlemlenmiştir.

11. Füzyon sistemin başarısı, başka geleneksel füzyon yöntemleri ile karşılaştırılarak da kanıtlanmıştır. Önerilen füzyon yöntemi, karşılaştırılan Ortalama füzyon (Averaging), MAX seçimi (MAX selection), MIN seçimi (MIN selection), lineer ağırlıklı ortalama füzyonu (linear weighted fusion) gibi yöntemlerden % 5'ten daha fazla oranda iyi sonuç vermiştir. Tüm bu sonuçların gösterdiği üzere multimodal bilgi füzyonu otomatik anlamsal bilgi çıkarımında bilgilerin daha yüksek performanslarda elde edilmesini sağlamaktadır.
12. Çoklu ortam uygulamaları içerik olarak çok zengindir ve veritabanlarında oldukça çok verinin/bilginin tutulmasını gerektirmektedir. Çok miktardaki bilgiye hızlı erişimi sağlamak üzere dizin yapılarının geliştirilmesi bir zorunluluktur. Bu projede videonun hem anlamsal içerikleri hem de görsel alt seviye özellikleri (MPEG-7 features) kullanılarak bir dizin yapısı geliştirilmiştir. Dizin yapısının erişim hızına önemli katkı sağladığı gözlemlenmiştir.
13. Genelde literatürde var olan çalışmalar tek veri tipi üzerinde yoğunlaşmıştır. Görüntü, ses ve metin verilerinin her üçünü de kullanan çalışma sayısı çok sınırlıdır. Bu açıdan çalışmanın literatüre önemli katkı sağlayacağı değerlendirilmektedir.
14. Çoklu-ortam uygulamalarının anlamsal içeriklerinin otomatik olarak çıkarılması, depolanması ve sorgulanması konusundaki bilimsel çalışmaların sonuçları henüz yeterince tatmin edici değildir. Dolayısıyla çoklu-ortam üzerinde daha çok bilimsel araştırmaların yapılmasına ihtiyaç vardır. Önümüzdeki dönemde de konu sıcak ve önemli bir araştırma konusu olmaya devam edecektir. Gerçekleştirilmiş olan proje, ileriye yönelik olarak da önemli bir bilgi birikimi sağlamıştır.

## 5. SONUÇ

Projemizin öneri dokümanında projenin başarısının aşağıda verilen iki ölçütle değerlendirileceği yer almıştır:

- Birinci ölçüt, yapılan akademik çalışmaların ve ortaya çıkan özgün çalışmaların uygun yerlerde yayınlanması sonucu ortaya çıkacak bilimsel yayın performansdır. Bu proje kapsamında en az sekiz (8) uluslararası konferans bildirisi ve en az beş (5) adet de SCI (Expanded) yer alan dergilerde makale yayınlanmasıdır.
- İkinci ölçüt ise amaçlanan modelin gerçekleşme oranı ve performansdır.

Proje sonunda geliştirilen model ve prototip sistemin, akademik dünyada önemli bir boşluğu doldurduğu değerlendirilmektedir. Proje sürecinde 8 adet uluslararası dergide (7 adeti SCI-E tarafından taranan dergilerdir) ve 21 adet konferanslarda (19 adet uluslararası, 2 adet ulusal) olmak üzere toplam 29 adet yayın yapılmıştır (yayınlanmak üzere kabul alanlar bu rakama dahildir). Görüldüğü üzere, hem dergi yayınında hem de konferans yayınında hedeflenen sayının üzerine çıkılarak önemli bir başarı kaydedilmiştir. Ayrıca, henüz sonuçları bilinmediğinden bu sayılara ilave edilmeyen projenin son döneminde sunulan makale ve bildiriler de mevcuttur. Bu nedenle, projeden üretilen yayın sayısının bu raporda verilen rakamların üzerine çıkması beklenmektedir.

İkinci ölçüt açısından bakıldığında ise, videoların görüntü, ses ve metin verileri kullanılarak (multi-modal) anlamsal bilgilerin otomatik olarak çıkarılması, uygun formatlarda saklanması ve daha sonra da etkin bir şekilde sorgulanması işlemlerini yapabilen bir prototip sistem geliştirilmiştir. Gerçekleştirilen sistem, proje başlangıcında hedeflenen tüm işlevleri içermektedir. Dolayısıyla, proje başlangıcında planlanan tüm konular üzerinde çalışılmış ve başarılı sonuçlar alınmıştır. Dolayısıyla kapsam açısından proje başarılı olarak sonuçlandırılmıştır.

İkinci ölçüt açısından bakılması gereken bir diğer husus ise ulaşılan performans ölçümleridir. Yapılan çalışmaların performansını kıyaslamak amacıyla mümkün olan yerlerde uluslararası veri setleri kullanılarak performans ölçümleri yapılmıştır. Ancak, yukarıda Bulgular bölümünde de açıklandığı üzere, görüntü, ses ve metin verilerinden anlamsal bilgi çıkarımı ve arkasından da bilgi füzyonu işlemlerinin tamamında ortak kullanılacak yeterli bir video veri seti bulunamamıştır. Bu nedenle, sistemin modülleri şayet mümkünse literatürde mevcut araştırmacıların kullandığı veri setleri ile, mümkün değilse kendi ürettiğimiz veri setleri ile test edilerek performansları ölçülmüştür. Bunun sonucu olarak, başka çalışmalarla karşılaştırmalar modüller bazında yapılabilmektedir. Gerek bu konuda çalışan araştırmacıların kullandığı açık bir veri setinin olmaması nedeniyle gerekse videodaki görsel, işitsel ve metin verilerini birlikte kullanan çalışmaların çok sınırlı olması nedeniyle sistemin bir bütün olarak performansı başka çalışmalarla kıyaslanamamıştır. Sistemin modüllerinin performans sonuçları genel olarak literatürde yayınlanan sonuçlarla uyumludur. Gerçekleştirilen yayınlar, önerilen modelin akademik dünyada kabul gördüğünün bir göstergesidir.

Gerçekleştirilen prototip sistem, önerilen modelin uygulanabilirliğini göstermesi açısından önemlidir. Böylece ortaya konulan konseptlerin pratikte uygulanabileceğini göstermiştir. Bu çalışma, bu alanda geliştirilecek ürünlere de ışık tutacaktır.

Bu proje videoların görüntü, ses ve metin verileri kullanılarak (multi-modal) anlamsal bilgilerin otomatik olarak çıkarılması, uygun formatlarda saklanması ve daha sonra da etkin bir şekilde sorgulanabilmesi konularıyla ilgili olarak belirli bir bakış açısı getirmektedir. Ancak, oldukça kapsamlı bir konu olması nedeniyle bu alanda çözülmesi gereken daha çok sorun mevcuttur. Bu proje kapsamında edinilen tecrübe ve bilgi birikimine dayanarak bu alanda üzerinde çalışma yapılması gerektiğini düşündüğümüz konular aşağıda verilmiştir:

- Günümüzde veritabanlarına video içerisinde yer alan anlamsal veriler genelde manüel olarak girilmektedir. Oldukça zahmetli olan bu işi otomatikleştirecek sistemler geliştirmek önemli bir çalışma alanı oluşturmaktadır (videodan otomatik anlamsal içerik çıkararak veritabanına kaydetmek). Yapmış olduğumuz bu çalışma, konuyla ilgili önemli sonuçlar içermekle birlikte, konunun zorluğundan dolayı bu konudaki çalışmaların artarak devam etmesi gerekmektedir.
- Görsel veri açısından, kamera açılarının ve 3. boyuttaki hareketlerin de dikkate alınarak konumsal ilişki çıkarımı işleminin iyileştirilmesi yapılabilir.
- Bu çalışmada, genetik algoritma tabanlı bir nesne çıkarımı yöntemi kullanılmıştır. Görsel nesne çıkarım yönteminin, anlamsal bilgi çıkarım sürecindeki en önemli mekanizma olduğu düşünülmektedir. Dolayısıyla, bu konuda daha etkin yöntemlerin geliştirilmesine ihtiyaç vardır.
- Görsel veriden bilgi çıkarımında, nesne, uzamsal ve zamansal ilişki çıkarımı ve en sonunda da olay çıkarımı adımları uygulanmıştır. İlişkilerin ve olayların çıkarılmasında sadece görsel nesne verilerine dayalı bir meta-ontoloji modeli kullanılmıştır. Bu yapının görsel, işitsel ve metinsel tüm kipleri içerecek şekilde genişletilmesi de bu konuda yapılabilecek bir çalışmadır.
- Metinden anlamsal bilgi çıkarımı ile ilgili olarak; hem varlık ismi tanıma hem de olay çıkarımı için ümit verici değerlendirme sonuçları elde edilmiştir. Ancak, bu sonuçları iyileştirmek için her iki konuda da farklı yöntemlerin denenmesi, daha fazla işaretlenmiş veri (annotated data) kümesi oluşturularak ve kullanarak farklı istatistiksel ve makine öğrenimi yöntemlerinin denenmesi ve olay çıkarımı için göz önünde bulundurulmuş olay tiplerinin artırılması, bu konuyla ilgili yapılabilecek ilave çalışmalardır.
- Sesten anlamsal bilgi çıkarımı aşamasında, ses verileri altı (sessizlik dahil) ana kategoriye göre sınıflandırılmıştır. Bu sınıflardan “çevresel ses” sınıfı kendi içerisinde 13 alt sınıfa ayrıştırılmıştır. Elde edilen ses sınıfları, genel ses tanıma uygulamalarının yanısıra video içerik analizinde de önemli bir rol oynamaktadır. Ses verisinin çok zengin bir içeriğe sahip olması nedeniyle, bu konudaki çalışmaların devam etmesi gerekmektedir.
- Füzyon yöntemi her ne kadar iyi sonuç verse de başarımın daha da artması için bazı iyileştirmeler yapılabilir. Çalışmanın genişletilebileceği bu alanlar aşağıda sıralanmıştır:
  - Her konsept için oluşturulan SVM modelinde sabit bir kernel fonksiyonu kullanmak yerine her konsept için en uygun kernel fonksiyonunu otomatik olarak bulan bir yöntem geliştirilebilir.
  - Modalitelerin zamansal eşlenmesinde kullanılan yöntem geliştirilebilir. Örneğin; füzyon sırasında sadece görsel çekime (shot) denk gelen metinsel verileri kullanmak yerine, komşu çekimlerdeki metin verileri de belirli oranlarda füzyon işlemine katılabilir.

- Bazı konseptler doğrudan tek bir modaliteye bağılı olabilir ve dięer modalitelerden gelen bilgiler füzyon işlemleri sırasında faydalı olmayıp gürültü (noise) görevi görebilir. Bu durumda uygun modalitelerin seçilmesi için kullanılan Relief tabanlı yaklaşım, dięer yaklaşımların da analiz edilmesi ile deęiştirilebilir ya da geliştirilebilir.

## YARARLANILAN KAYNAKLAR

- [1] R. Zabih, J. Miller, and K. Mai. A Feature-Based Algorithm for Detecting and Classifying Scene Breaks. Proc. ACM Multimedia 95, San Francisco, CA, pp. 189-200, Nov. 1995.
- [2] J. Shi and J. Malik, "Normalized cuts and image segmentation". IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, pp. 888-905, 2000.
- [3] Moving Picture Experts Group (MPEG), "MPEG-7 reference software experimentation model (XM Software)", 2003.
- [4] Y. Yildirim, T. Yilmaz and A. Yazici, "Ontology-supported object and event extraction with a genetic algorithms approach for object classification", In Proc. of the 6th ACM int. conf. on Image and video retrieval (CIVR '07), New York, USA, 2007, pp. 202-209.
- [5] J. F. Allen, "Maintaining knowledge about temporal intervals", Communications of the ACM, Vol. 26, pp. 832-843, 1983.
- [6] D. Nadeau ve S. Sekine. "A survey of named entity recognition and classification", Linguistica Investigiones, 2007, vol. 30, no. 1, pp.3-26.
- [7] D. Freitag. "Machine learning for information extraction in informal domains", Ph.D. thesis, Computer Science Department, Carnegie Mellon University, 1998.
- [8] Y. Zhang, C. Xu, Y. Rui, J. Wang, H. Lu, Semantic event extraction from basketball games using multimodal analysis, in: Proceedings of the IEEE Conference on Multimedia and Expo (ICME), 2007.
- [9] P. Jackson, I. Moulinier, Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization, 2nd Edition, John Benjamins, 2007.
- [10] Royalty Free Music and Sound Effects, Partners In Rhyme, Inc, <http://www.partnersinrhyme.com>. Son erişim tarihi: 15.02.2011
- [11] MPEG-7 Audio Encoder, <http://mpeg7audioenc.sourceforge.net/index.html>. Son erişim tarihi: 06.01.2012.
- [12] "The bbc sound effects library," [Online]. Available: <http://www.sound-ideas.com/>.
- [13] Cees G. M. Snoek, Marcel Worring, and Arnold W. M. Smeulders, "Early versus Late Fusion in Semantic Video Analysis", in Proceedings of the ACM International Conference on Multimedia, Singapore, 2005, pp. 399-402.
- [14] Sun, Y. and Wu, D., "Feature Extraction Through Local Learning", in Journal of Statistical Analysis and Data Mining, 2009, pp. 34-47.
- [15] Digital Video Retrieval at NIST, <http://trecvid.nist.gov/>
- [16] Jiang, Y.-G., Ye, G., Chang, S.-F., Ellis, D., & Loui, A. C. (2011). Consumer Video Understanding: A Benchmark Database and An Evaluation of Human and Machine Performance. Proceedings of ACM International Conference on Multimedia Retrieval (ICMR), oral session (pp. 29:1--29:8). New York, NY, USA: ACM.



- [17] P. Over, G. Awad, W. Kraaij, and A. F. Smeaton. Trecvid 2007- Overview. In TRECVID'07, 2007.
- [18] A.Yazıcı, Ç.İnce and M. Koyuncu. "FOOD Index: A Multidimensional Index Structure for Similarity-Based Fuzzy Object-Oriented Database Models". In IEEE Transaction on Fuzzy Systems, Vol.16, No.4, pp.942-957, 2008.
- [19] Berchtold, Stefan; Keim, Daniel A.; Kriegel, Hans-Peter. "The X-Tree: An Index Structure for High-Dimensional Data". Proceedings of the 22nd VLDB Conference (Mumbai, India): 28–39, 1996.
- [20] N. B. Ozgur, M. Koyuncu and A. Yazici, "An intelligent fuzzy object-oriented database framework for video database applications", Fuzzy Sets and Systems, Vol.160, pp. 2253-2274, 2009.

**TÜBİTAK**  
**PROJE ÖZET BİLGİ FORMU**

Proje Yürütücüsü:	Prof. Dr. ADNAN YAZICI
Proje No:	109E014
Proje Başlığı:	Videolarda Görüntü, Ses ve Metin Verileri Kullanılarak Anlamsal Bilgi Çıkarımı, Depolanması ve Sorgulanması
Proje Türü:	Araştırma
Proje Süresi:	36
Araştırmacılar:	MUSTAFA SERT, MURAT KOYUNCU
Danışmanlar:	
Projenin Yürütüldüğü Kuruluş ve Adresi:	ORTA DOĞU TEKNİK Ü. MÜHENDİSLİK F. BİLGİSAYAR MÜHENDİSLİĞİ B.
Projenin Başlangıç ve Bitiş Tarihleri:	01/09/2009 - 01/01/2013
Onaylanan Bütçe:	193370.0
Harcanan Bütçe:	160114.03
Öz:	<p>Bu projede, videoların görüntü, ses ve metin verileri kullanılarak (multi-modal) otomatik anlamsal bilgilerin çıkarıldığı, uygun formatlarda saklandığı ve daha sonra da etkin bir şekilde sorgulanabildiği bir prototip sistem geliştirilmiştir. Geliştirilen sistemde, bazı ön işlemlerin yapılmasını takiben, öncelikle görüntü, ses ve metin verilerinin her birinden ayrı ayrı bilgiler çıkarılmaktadır. Daha sonra, bu üç farklı veri tipinden gelen bilgilerin analiz edilmesi, bütünleştirilmesi, eksiklerinin tamamlanması, dublikasyonlarının temizlenmesi ve böylece veritabanına kaydetmeye hazır bilginin elde edilmesi amacıyla bilgi füzyonu yapılmaktadır. Füzyon sonunda elde edilen video bilgileri, çoklu-ortam verileri için projenin araştırmacıları tarafından daha önce yine bir TÜBİTAK 1001 Projesi kapsamında geliştirilmiş olan Akıllı ve Bulanık Nesneye Dayalı Veritabanı sisteminde saklanmaktadır. Ayrıca, sorgulara hızlı cevaplar üretebilmek amacıyla videolardan çıkarılan anlamsal bilgiler ile videoların alt seviye özelliklerinin indekslenemediği etkin bir dizin yapısı geliştirilmiştir. Geliştirilen sistemde, belirsizlik ve bulanıklık içeren bilgiler de işlenebilmektedir. Böylece bu proje çerçevesinde yapılan araştırmaların sonucunda geliştirmiş olduğumuz prototip sistem, çokluortam veritabanlarının olası kullanıcılarının gereksinim duyacağı bir çok sorgu tipini etkin bir şekilde destekleyebilmektedir.</p> <p>Üç yıl olarak başlatılan proje 4 aylık uzatmayla birlikte 40 ayda tamamlanmıştır. Projede, bulanıklık, nesneye dayalı veritabanı, bilgi tabanlı sistemler ve çoklu ortam uygulamaları konusunda uzmanlaşmış üç araştırmacı görev almıştır. Proje kapsamında projenin değişik süreçlerinde görev alan 3 doktora ve 10 lisansüstü öğrencisinin tez çalışmasına imkan verilerek bu öğrencilerin tezlerini başarıyla bitirmeleri sağlanmıştır. Daha bir kaç lisansüstü tez çalışması da devam etmektedir. Ayrıca, proje sonunda geliştirilen model ve prototip sistemin, akademik dünyada önemli bir boşluğu doldurduğu değerlendirilmektedir. Proje sürecinde 8 adet (7 adet SCI-E tarafından taranan) dergilerde ve 21 adet konferanslarda (19 adet uluslararası, 2 adet ulusal) olmak üzere toplam 29 adet yayın yapılmıştır. Proje sonunda gönderilen ilave makale ve bildirimlerle bu sayının daha da artması beklenmektedir.</p>
Anahtar Kelimeler:	çoklu ortam uygulamaları, video, anlamsal bilgi çıkarımı, füzyon, bulanıklık, belirsizlik
Fikri Ürün Bildirim Formu Sunuldu Mu?:	Hayır

<p>Projeden Yapılan Yayınlar:</p>	<p>1- Named Entity Recognition Experiments on Turkish Texts (Bildiri),  2- Content-Based Retrieval of Audio in News Broadcasts (Bildiri),  3- Fuzzy Querying in Intelligent Information Systems (Bildiri),  4- Lightweight Object Localization with a Single Camera in Wireless Multimedia Sensor Networks (Bildiri),  5- A Text-Based Fully Automated Architecture for the Semantic Annotation and Retrieval of Turkish News Videos (Bildiri),  6- A Framework for Fuzzy Video Content Extraction, Storage and Retrieval (Bildiri),  7- A Hybrid Named Entity Recognizer for Turkish with Applications to Different Text Genres (Bildiri),  8- Comparison of Multidimensional Data Access Methods for Feature-Based Image Retrieval (Bildiri),  9- Improving Automatic Semantic Annotations of News Videos in Turkish through Web Alignment and Event Extraction (Bildiri),  10- A Genetic Algorithms Based Classifier For Object Classification In Images (Bildiri),  11- Providing Automated Actions in Wireless Multimedia Sensor Networks via Active Rules (Bildiri),  12- Flexible Content Extraction and Querying for Videos (Bildiri),  13- Multilingual Video Indexing and Retrieval Employing an Information Extraction Tool for Turkish News Texts: A Case Study (Bildiri),  14- Exploiting Class-Specific Features in Multi-Feature Dissimilarity Space for Efficient Querying of Images. Flexible Query Answering Systems (Bildiri),  15- Turkish Large Vocabulary Continuous Speech Recognition By Using Limited Audio Corpus (Bildiri),  16- A Novel Fuzzy Visual Object Classification Approach (Bildiri),  17- A RELIEF-Based Modality Weighting Approach for Multimodal Information Retrieval (Bildiri),  18- A Multimodal Fusion Approach by Exploiting Concept Interactions for Efficient Multimedia Analysis (Bildiri),  19- Efficient and Accurate Object Classification in Wireless Multimedia Sensor Networks (Bildiri),  20- Semi-Automatic Semantic Video Annotation Tool (Bildiri),  21- Bir Çokluortam Veri Yönetim Sistemi Yazılım Mimarisinin ?Views and Beyond? Yaklaşımıyla Belgelenmesi: Durum Raporu (Bildiri),  1- Exploiting information extraction techniques for automatic semantic video indexing with an application to Turkish news videos (Makale/Kitap/Kitapta Bölüm),  2- A Flexible and Scalable Audio Information Retrieval System for Mixed-Type Audio Signals (Makale/Kitap/Kitapta Bölüm),  3- A Hybrid Named Entity Recognizer for Turkish (Makale/Kitap/Kitapta Bölüm),  4- Multimodal Information Fusion for Semantic Video Analysis (Makale/Kitap/Kitapta Bölüm),  5- Automatic Semantic Content Extraction in Videos using a Fuzzy Ontology and Rule-based Model (Makale/Kitap/Kitapta Bölüm),  6- A Semi-Automatic Text-Based Semantic Video Annotation System for Turkish Facilitating Multilingual Retrieval (Makale/Kitap/Kitapta Bölüm),  7- Efficient Localization and Tracking of Multiple Objects in Wireless Multimedia Sensor Networks (Makale/Kitap/Kitapta Bölüm),  8- Comparison of Feature-based and Image Registration-based Retrieval of Image Data Using Multidimensional Data Access Methods (Makale/Kitap/Kitapta Bölüm)</p>
-----------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------