**T.C.**
**GEBZE TEKNİK ÜNİVERSİTESİ**
**FEN BİLİMLERİ ENSTİTÜSÜ**

# FİZİK TERAPİ EGZERSİZ TİPLERİNİN TANINMASI VE TAKİP EDİLMESİ İÇİN BİR BİLGİSAYARLA GÖRME YÖNTEMİ

**İLKTAN AR**
**DOKTORA TEZİ**
**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**GEBZE**
**2015**

**T.C.**
**GEBZE TEKNİK ÜNİVERSİTESİ**
**FEN BİLİMLERİ ENSTİTÜSÜ**

# FİZİK TERAPİ EGZERSİZ TİPLERİNİN TANINMASI VE TAKİP EDİLMESİ İÇİN BİR BİLGİSAYARLA GÖRME YÖNTEMİ

**İLKTAN AR**
**DOKTORA TEZİ**
**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

DANIŞMANI
DOÇ. DR. YUSUF SİNAN AKGÜL

**GEBZE**
**2015**

**T.R.**

**GEBZE TECHNICAL UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# A COMPUTER VISION APPROACH TO RECOGNIZE AND TO TRACK THE TYPES OF PHYSIOTHERAPY EXERCISES

**İLKTAN AR**

**A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPY**

**DEPARTMENT OF COMPUTER ENGINEERING**

THESIS SUPERVISOR

ASSOC. PROF. DR. YUSUF SINAN AKGUL

**GEBZE**

**2015**

| | |
|---|---|
| GEBZE TEKNİK ÜNİVERSİTESİ | **DOKTORA JÜRİ ONAY FORMU** |

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 12/06/2015 tarih ve 2015/37 sayılı kararıyla oluşturulan jüri tarafından 23/06/2015 tarihinde tez savunma sınavı yapılan İlktan Ar'ın tez çalışması Bilgisayar Mühendisliği Anabilim Dalında DOKTORA tezi olarak kabul edilmiştir.

**JÜRİ**

| | |
|---|---|
| ÜYE (TEZ DANIŞMANI) | : Doç. Dr. Y. Sinan Akgül |
| ÜYE | : Prof. Dr. Şebnem Baydere |
| ÜYE | : Doç. Dr. M. Elif Karslıgil |
| ÜYE | : Doç. Dr. Mehmet Göktürk |
| ÜYE | : Doç. Dr. Yakup Genç |

**ONAY**

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

...../...../..... tarih ve ...../..... sayılı kararı.

İMZA/MÜHÜR

# ÖZET

Eylem tanıma problemi sabit veya hareketli görüntülerde insanların hareketlerin bilgisayarla tanımlanmasıdır. Bu problem günümüzde en dikkat çeken bilgisayarla görme konularından biridir. Literatürde eylem tanıma problemi: spor müsabakalarının hareketli görüntülerinden sporcuların yaptığı eylemlerinin neler olduğunun tanınması, dans ve bale kareografilerinin tanınması, ortopedik hastaların klinik incelemeleri, yaşlı insanların bakımı gibi konularla örneklenebilir.

Bu tezde, hareketli görüntülerdeki insan eylemlerini tanımak için bilgisayarla görme tabanlı çeşitli sistemler oluşturulmuştur. Geliştirilen bu sistemlerde kullanılmak üzere hareketli görüntülerdeki, hareket tabanlı öznitelikleri tanımlayabilmek için ve duruş tabanlı öznitelikleri tanımlayabilmek için özgün öznitelik çıkarma yöntemleri oluşturulmuştur.

Tez kapsamında insan eylemlerinin tanınmasına örnek konu olarak evde yapılan fizik tedavi egzersizlerinin analizi seçilmiştir. İlgili sağlık kuruluşlarından evde yapılan fizik tedavi seanslarında en sık kullanılan egzersiz tipleri hakkında bilgi alındıktan sonra bu egzersizlerin hareketli görüntülerini barındıran bir veri kümesi oluşturulmuştur. Ev Ortamındaki Fizik Tedavi Egzersizleri (EOFTE) veri kümesinde MS Kinect algılayıcı kullanılarak kaydedilen 240'ar tane renk ve derinlik hareketli görüntüleri bulunmaktadır. Daha sonra, fizik tedavi egzersizlerinin analizi için temel ve gelişmiş olmak üzere iki farklı sistemler tasarlanmıştır. Bu sistemler hareket, duruş ve nesne bilgisi olmak üzere üç ana fizyoterapi bilgi türünün, üretken bir Bayes ağı yapısında değerlendirerek egzersizleri yüksek başarı oranı ile tanımaktadır. Sistemlerde renk ve derinlik hareketli görüntüleri birlikte kullanılmıştır. Sadece renk hareketli görüntülerini kullanarak da yüksek tanıma oranları elde edilmiştir. Son olarak, egzersiz seanslarının takibi için egzersiz tekrar sayısının tahmin eden özgün bir yöntem geliştirilmiştir.

Geliştirilen sistemlerin ve yöntemlerin etkinliği birçok eylem tanıma veri kümesinde test edilmiş ve elde edilen sonuçların başarılı olduğu literatürdeki benzer çalışmaların sonuçları ile karşılaştırılarak gösterilmiştir.

**Anahtar Kelimeler: Eylem Tanıma, Egzersiz Tanıma, Öznitelik Çıkarımı, Bayes Ağı, Haar Benzeri Hareket Öznitelikleri, Izgarasal Duruş Öznitelikleri.**

# SUMMARY

Action recognition problem is defined as the computerized recognition of human actions from still images or videos. This problem is one of the most interesting topics in computer vision. Examples of action recognition problem in literature are: recognition of sport specific actions (e.g. style of tennis serve) from sport videos, recognition of dance's and ballet's choreography, clinical examination of orthopedic patients, examination of human physical status, and care of elderly people.

In this thesis, computer vision based systems are constructed to recognize human actions from videos. In order to use within these systems, novel feature extraction methods are developed which obtain motion-based features and stance-based features.

The analysis of home-based physiotherapy exercises is selected as a subtopic in action recognition. After receiving information about the most common types of home-based physiotherapy exercises from related health care institutes, a dataset (EOFTE) that contains videos of these exercises was created. The EOFTE dataset contains 240 color and 240 depth videos which are captured by a Microsoft Kinect sensor. Then, two different systems have been proposed, as basic and enhanced version, to analyze physiotherapy exercises. These two systems use the main sources of information in physiotherapy (motion patterns, stance information, exercise object) within a Bayesian network and recognize the type of exercise in the given videos accurately. Although the recognition rate obtained by using both color and depth videos is higher than using only color videos, the proposed system can robustly use only color videos. Finally a novel approach is presented to estimate the exercise repetition counts.

The effectiveness of the developed systems and methods is validated on several action recognition datasets. The results obtained in the experiments proved that the developed systems and methods are successful.

**Key Words: Action Recognition, Exercise Recognition, Feature Extraction, Bayesian Network, Haar-like Motion Features, Grid-based Stance Features.**

# TEŞEKKÜR

Bu tez çalışmasının her aşamasında desteğini ve yardımlarını hiçbir zaman esirgemeyip bilgisi ile bu çalışmanın oluşmasının yolunu açan danışmanım Doç. Dr. Yusuf Sinan AKGÜL'e tavsiyeleri, destekleri, yönlendirmeleri ve en önemlisi bana karşı olan inancı ve sabrı için,

Tez izleme komitemde bulunan Doç. Dr. Mehmet GÖKTÜRK'e ve Doç. Dr. Mine Elif KARSLIGİL'e tezimin ilerleme sürecindeki görüşleri ve önerileri için,

Prof. Dr. Şebnem BAYDERE'ye ve Yrd. Doç. Dr. Yakup GENÇ'e tez jürimde yer aldığı için,

Tüm çalışma arkadaşlarıma fikirleri, dostlukları ve huzurlu bir çalışma ortamı sağladıkları için,

Eşim Esin AR'a, oğluma ve tüm aileme manevi destekleri ve sabırları için,

Türkiye Bilimsel ve Teknolojik Araştırma Kurumuna (TÜBİTAK) doktora eğitimim boyunca yurtiçi doktora burs programıyla beni desteklediği için,

içten teşekkürlerimi sunarım.

# İÇİNDEKİLER


| | Sayfa |
|---|---|
| ÖZET | v |
| SUMMARY | vi |
| TEŞEKKÜR | vii |
| İÇİNDEKİLER | viii |
| SİMGELER ve KISALTMALAR DİZİNİ | x |
| ŞEKİLLER DİZİNİ | xii |
| TABLOLAR DİZİNİ | xiv |
| | |
| 1. GİRİŞ | 1 |
| 2. ÖNCEKİ ÇALIŞMALAR | 4 |
| 2.1. Öznitelik Çıkarımı | 4 |
| 2.1.1. Hareket Tabanlı Öznitelik Çıkarma Yöntemleri | 4 |
| 2.1.2. Duruş Tabanlı Öznitelik Çıkarma Yöntemleri | 10 |
| 2.1.2.1. Cheema Arkadaşlarının Duruş Tabanlı Öznitelik Çıkarma Yöntemi | 12 |
| 2.1.3. Nesne Tabanlı Öznitelik Çıkarma ve Nesne Tespit Yöntemleri | 17 |
| 2.1.3.1. Fei-Fei'nin Nesne Tespit Yöntemi | 18 |
| 2.1.3.2. Viola ve Jones'un Nesne Tespit Yöntemi | 19 |
| 2.1.3.3. Gall ve Lempitsky'nin Nesne Tespit Yöntemi | 21 |
| 2.2. Eylem Tanıma Konusunda Yapılmış Çalışmalar | 22 |
| 2.3. Fizik Terapi Egzersizlerini Konu Alan Yapılmış Çalışmalar | 26 |
| 3. VERİ KÜMELERİ | 29 |
| 3.1. Gall Veri Kümesi | 29 |
| 3.2. KTH (Kraliyet Teknoloji Enstitüsü) Veri Kümesi | 31 |
| 3.3. Weizmann Veri Kümesi | 32 |
| 3.4. Ev Ortamındaki Fizik Tedavi Egzersizleri (EOFTE) Veri Kümesi | 33 |
| 4. HAAR-BENZERİ ÖZELLİKLERLE HAREKET TABANLI ÖZNİTELİK ÇIKARIMI | 40 |
| 5. RASTGELE ORMANLAR YAPISIYLA GELİŞTİRİLMİŞ HAREKET TABANLI ÖZNİTELİK ÇIKARIMI (ROY_HÖ) | 43 |

6. IZGARASAL YAPIYLA DURUŞ TABANLI ÖZNİTELİK ÇIKARIMI 46
7. HAREKET VE NESNE ÖZNİTELİKLERLE EYLEM TANIMA 50
7.1. Tasarlanılan Sistemin Modeli ve Detayları 50
7.2. Deneysel Sonuçlar 54
8. HAREKET VE DURUŞ ÖZNİTELİKLERİYLE EYLEM TANIMA 56
8.1. Tasarlanılan Sistemin Modeli ve Detayları 56
8.2. Deneysel Sonuçlar 59
9. FİZİK TEDAVİ EGZERSİZLERİNİ ANALİZ EDEN TEMEL VE GELİŞTİRİLMİŞ SİSTEM 64
9.1. Fizik Tedavi Egzersizlerini Analiz Eden Temel Sistem 65
9.2. Fizik Tedavi Egzersizlerini Analiz Eden Geliştirilmiş Sistem 67
9.3. Fizik Tedavi Egzersiz Seanslarında Tekrar Sayısını Bulma Yöntemi 71
9.4. Deneysel Sonuçlar 74
10. SONUÇLAR ve YORUMLAR 83

KAYNAKLAR 86
ÖZGEÇMİŞ 96
EKLER 97

# SİMGELER ve KISALTMALAR DİZİNİ

| **Simgeler ve Kısaltmalar** | | **Açıklamalar** |
|---|---|---|
| $\overline{D}$ | : | Duruş bilgisi öznitelik vektörünün simgesi |
| D | : | Kontur nokta kümesi elemanlarının merkeze uzaklık vektörü |
| d | : | Kontur noktasının ağırlık merkezine uzaklığı |
| dk | : | Duruş kanıtları simgesi |
| et | : | Egzersiz tipi simgesi |
| hk | : | Hareket kanıtları simgesi |
| k | : | Filtre tipi simgesi |
| L | : | Egzersiz tipini tanıma fonksiyonu |
| m | : | Ortalama simgesi |
| M | : | X ve Y eksenindeki ağırlık merkezi |
| n | : | Maksimum filtre sayısı |
| nk | : | Nesne kanıtları simgesi |
| P | : | Olasılık fonksiyonu |
| Q | : | Potansiyel fonksiyonu |
| s | : | Varyans simgesi |
| t | : | Zaman ekseni simgesi |
| v | : | Hareketli görüntü simgesi |
| x | : | 2-boyutlu uzayda yatay eksen |
| y | : | 2-boyutlu uzayda dikey eksen |
| Z | : | Potansiyel fonksiyon sabiti |
| 3DH | : | 3-boyutlu Haar filtresi |
| 3DHF | : | 3-boyutlu Haar-benzeri özellikler |
| BKDHÖ | : | Birleşik Kısa Dönem Hareket Örüntüleri |
| DT | : | Duruş Tipi |
| DTKG | : | Dilim Tabanlı Küresel Gösterim |
| DUB | : | Duruş Bilgisi |
| DVM | : | Destek Vektör Makinesi |
| EOFTE | : | Ev Ortamındaki Fizik Tedavi Egzersizleri |
| ET | : | Egzersiz Tip kümesi |

| | | |
|---|---|---|
| GD | : | Güvenirlik Değeri |
| GKHG | : | Geliştirilmiş Küresel Hareket Gösterim |
| G-KYM | : | Gri tonlarda-Kırmızı Yeşil Mavi |
| HAB | : | Hareket Bilgisi vektörü |
| HB | : | Hareket Bilgisi |
| HEG | : | Hareket Enerji Görüntüsü |
| HGG | : | Hareket Geçmiş Görüntüsü |
| HÖDÖD | : | Hareket-tabanlı Ölçekle Değişmeyen Özellik Dönüşümü |
| HT | : | Hareket Tipi |
| IY_DÖ | : | Izgarasal Yapıyla Duruş Öznitelikleri |
| KDHÖ | : | Kısa Dönem Hareket Örüntüleri |
| KHG | : | Küresel Hareket Gösterimi |
| KK | : | Kod Kelime |
| KT | : | Kompakt tanımlayıcı |
| KTH | : | Kraliyet Teknoloji Enstitüsü |
| KYM | : | Kırmızı Yeşil Mavi |
| KYMD | : | Kırmızı Yeşil Mavi Derinlik |
| NB | : | Nesne Bilgisi |
| NT | : | Nesne Tipi |
| ÖDÖD | : | Ölçekle Değişmeyen Özellik Dönüşümü |
| ÖK | : | Öznitelik Kümesi |
| ÖT | : | Öznitelik Tanımlayıcısı |
| ROY_HÖ | : | Rastgele Ormanlar Yapısıyla Hareket Öznitelikleri |
| YÖA | : | Yavaş Özellik Analizi |

# ŞEKİLLER DİZİNİ


| Şekil No: | | Sayfa |
|---|---|---|
| 1.1: | Tez çalışması kapsamında geliştirilen yöntemlerin, sistemlerin ve veri kümelerinin arasındaki bağlantılar. | 3 |
| 2.1: | Hareket Enerji Görüntüsü (HEG) ve Hareket Geçmiş Görüntüsü (HGG) örneği. | 6 |
| 2.2: | Örnek bir köpek görüntüsü ve bu görüntünün Radon dönüşümü. | 7 |
| 2.3: | Oturma eylemi hareketli görüntüsünün ana çerçeveleri. | 7 |
| 2.4: | Şekil 2.3'de verilen oturma eyleminin hareket özniteliklerinin Radon dönüşüm yüzeyi. | 7 |
| 2.5: | Izgarasal gösterimde ÖDÖD ve HÖDÖD özellik tanımlayıcıları. | 9 |
| 2.6: | a) Zıplama, b) yürüme, c) koşma eylemlerinin uzam zaman şekilleri. | 12 |
| 2.7: | Dikdörtgen alanlarla vücut parçalarının kapsanması. | 12 |
| 2.8: | a) Örnek bir silüet görüntüsü, b) ilgili kontur görüntüsü. | 14 |
| 2.9: | Kontur noktalarının, ağırlık merkezine göre sıralanması. | 15 |
| 2.10: | Sağa tekme ve yumruk eylemlerinin duruş ağırlık değerleri. | 16 |
| 3.1: | Gall veri kümesinden alınan örnek çerçeveler. | 30 |
| 3.2: | Gall veri kümesinde bulunan nesnelerin örnek görüntüleri. | 30 |
| 3.3: | Weizmann veri kümesinden alınan örnek çerçeveler. | 32 |
| 3.4: | Sopa egzersizinden örnek bir görüntü. | 35 |
| 3.5: | Çapraz sopa egzersizinden örnek bir görüntü. | 36 |
| 3.6: | Sırtüstü bacak egzersizinden örnek bir görüntü. | 36 |
| 3.7: | Havlu egzersizinden örnek bir görüntü. | 37 |
| 3.8: | Düz sarkaç egzersizinden örnek bir görüntü. | 37 |
| 3.9: | Dairesel sarkaç egzersizinden örnek bir görüntü. | 38 |
| 3.10: | Sandalye egzersizinden örnek bir görüntü. | 38 |
| 3.11: | Topuk egzersizinden örnek bir görüntü. | 39 |
| 4.1: | Hareket tabanlı öznitelik çıkarmada kullanılan Haar-benzeri filtreler. | 41 |
| 5.1: | Rastgele orman yapısıyla geliştirilmiş özellik çıkarma örneği. | 44 |
| 6.1: | Sopa kaldırma egzersizinden örnek bir görüntünün arka planı ve aslı. | 47 |
| 6.2: | En büyük ön plan nesnesini barındıran örnek bir görüntü. | 48 |
| 6.3: | Duruş bilgisi (DUB) öznitelik vektörünün oluşturulması. | 48 |

7.1: Hareket ve nesne bilgisi kullanarak eylem/nesne tanıma modeli. 52
8.1: Hareket ve duruş bilgisini kullanarak eylem tanıma veri akış diyagramı. 57
8.2: Hareket ve duruş bilgisiyle eylem tanımada kullanılan filtre tipleri. 58
9.1: Fizik tedavi egzersizlerini analiz eden temel sistemin modüler yapısı. 65
9.2: Geliştirilmiş sistemin veri akış diyagramı. 68
9.3: Geliştirilmiş sistemdeki üretken Bayes Ağının grafiksel modeli. 69
9.4: Örnek bir egzersiz seansı için *CV(t)-t* eğrisi. 73
9.5: Derinlik hareketli görüntüleri kullanılmadan başarılı tanınan dairesel sarkaç egzersizinden sıralı çerçeveler. 76
9.6: Hem renk hemde derinlik hareketli görüntüleri kullanılarak başarılı tanınabilen düz sarkaç egzersizinden sıralı çerçeveler. 77
9.7: Tekrar sayısı üç olan bir sopa egzersizinde tahmin edilen tekrar başlangıç/bitiş çerçeveleri. 82

# TABLOLAR DİZİNİ


| Tablo No: | | Sayfa |
|---|---|---|
| 3.1: | Egzersiz tiplerinin nesne ve duruş özellikleriyle bağlantısı. | 39 |
| 7.1: | Gall veri kümesindeki eylem nesne eşleşmesi. | 53 |
| 7.2: | Gall veri kümesinde çeşitli senaryolarda eylem/nesne tanıma başarısı. | 55 |
| 7.3: | Gall veri kümesinde eylem/nesne tanıyan sistemin karışıklık matrisi. | 55 |
| 8.1: | Weizmann veri kümesinde eylem tanıma karışıklık matrisi. | 61 |
| 8.2: | KTH veri kümesinde eylem tanıma karışıklık matrisi. | 62 |
| 8.3: | Duruş ve/veya hareket bilgisiyle eylem tanıyan sistemlerin doğruluk oranları. | 62 |
| 9.1: | Sadece renk hareketli görüntüleriyle egzersiz tanıma sonuçları. | 75 |
| 9.2: | Hem renk hem de derinlik hareketli görüntüleriyle egzersiz tanıma sonuçları. | 76 |
| 9.3: | Geliştirilmiş sistemin çeşitli bilgi türleriyle elde ettiği tanıma oranları. | 78 |
| 9.4: | Geliştirilmiş sistemin farklı bilgi ve veri türleriyle tanıma oranları. | 80 |
| 9.5: | Geliştirilmiş sistemin eylem tanıma karışıklık matrisi (sadece renk). | 81 |
| 9.6: | Geliştirilmiş sistemin eylem tanıma karışıklık matrisi (renk ve derinlik). | 81 |
| 9.7: | Egzersiz tekrar sayısının tahmin edilmesiyle ilgili sonuçlar. | 82 |

# 1. GİRİŞ

Hareketli görüntülerdeki (ing. video) insan eylemlerinin (ing. action) tanınması günümüzdeki en dikkat çeken bilgisayarla görme araştırmalarından biridir [1]-[4]. Çok çeşitli kullanım alanları olan bu konuda farklı çalışmalar yapılmaktadır. Örneğin spor müsabakalarının hareketli görüntülerinden sporcuların yaptığı eylemlerinin neler olduğunun tanınması (teniste temel vuruş stili v.b.) [5], dans ve bale kareografilerinin tanınması [6], ortopedik hastaların klinik incelemeleri [3], insanın fiziksel durumunun incelenmesi [1], yaşlı insanların bakımı [7] gibi konular bu tip çalışmaların uygulama alanını oluşturmaktadır. Ayrıca konu, dudak okuma, el ifadeleri anlama, yüz hareketleri anlama gibi alt dallara da bölünebilmektedir.

Fizik tedavi ya da yaygın adıyla Fizyoterapi, insanların sinir, kas, ve iskelet sistemlerinde ağrı ve hareket kısıtlılığı gibi bozuklukların tanısının yapılması ve tedavi edilmesi süreçlerini kapsayan çalışmalar bütünüdür. Bu çalışmalarda ana amaç hastanın kaybolan fonksiyonlarını ona tekrar kazandırmaktır. Fizik tedavi kullanılan yöntemler: sıcak-soğuk uygulama, elektroterapi, hidroterapi masaj ve egzersizler (ing. exercise) şeklinde sıralanabilir [8].

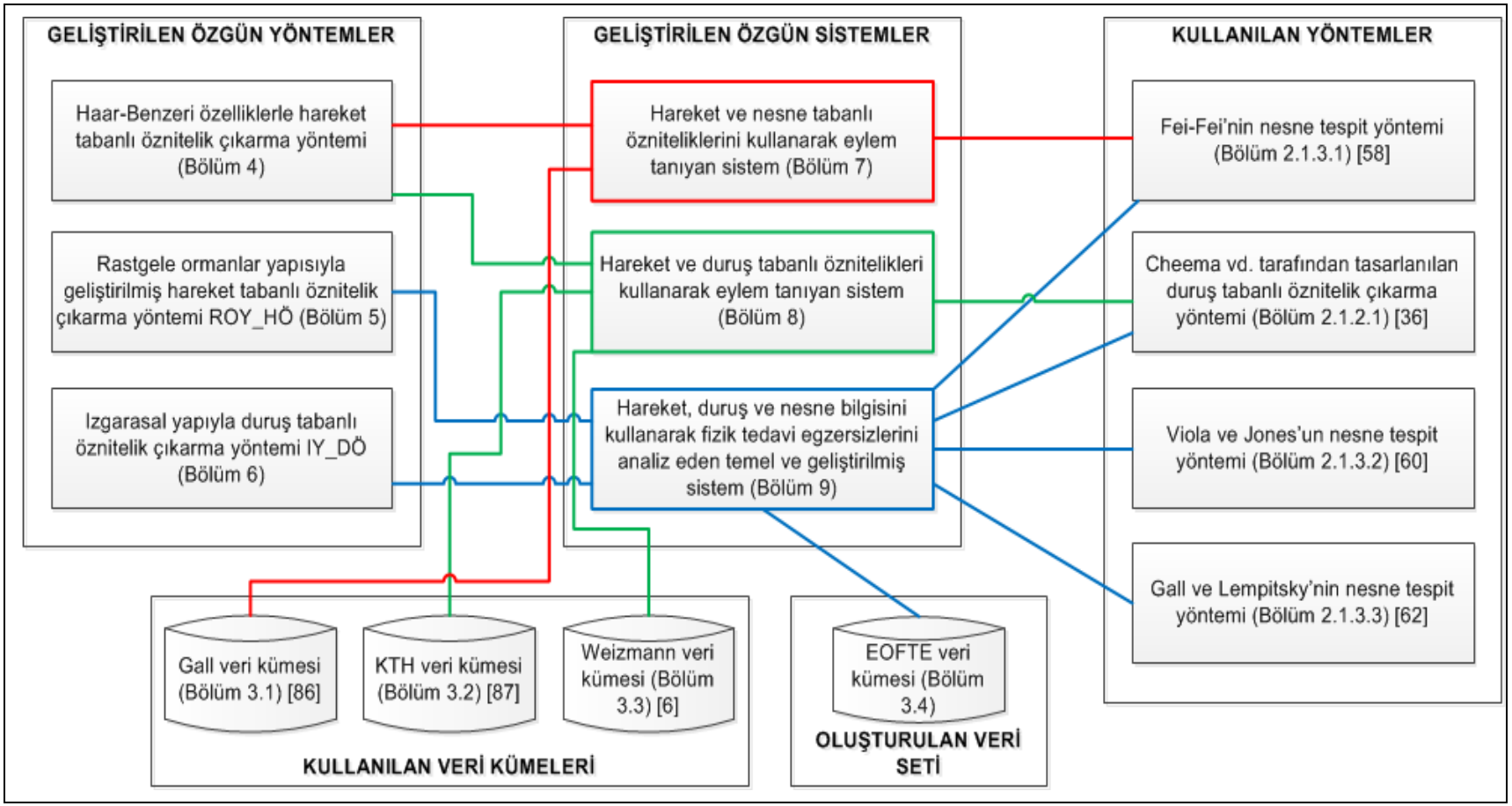
Yapılan tez çalışmasında hareketli görüntülerdeki insan eylemlerinin tanınması ana hedeftir. Bu hedef doğrultusunda uygulama alanı olarak evde yapılan fizik tedavi hareketlerinin analizi seçilmiştir. Fizik tedavi (ing. physiotherapy) hastalarından, evlerinde yapılması istenilen omuz ve diz egzersizlerini inceleyerek bu egzersizleri tanıyan ve egzersiz adedini tahmin eden bir sistem geliştirilmiştir. Özgün bir problem olan bu konu üzerinde yapılan çalışmalar genellikle maliyetli sistemlerle sonuçlanmaktadır [9]. Ayrıca bu sistemler, ev ortamı yerine özel rehabitilasyon merkezlerinde kullanılmaktadır. Bu noktada düşük maliyet ile geliştirilebilecek bir sistemin zamansal ve ekonomik bazda getirisi olacaktır.

İncelenilen fizik tedavi hareketleri omuz, diz, bel gibi belirli bölgelerdeki eklemleri, kasları vs. eski çalışma düzenine kavuşturmak amacıyla içinde tekrar içeren ve belirli kurallar dahilinde icra edilen egzersiz serilerinden oluşmaktadır. Bu egzersizlerden örnek hareketli görüntüler sabit bir KYMD (Kırmızı Yeşil Mavi Derinlik, ing. RGBD) kamera (MS Kinect) vasıtasıyla alınarak egzersiz seanslarından oluşan bir hareketli görüntü kümesi oluşturulmuştur. Oluşturulan veri kümesinde farklı aktörler tek başlarına seçilen fizik tedavi egzersizlerini çeşitli sayılarda tekrar etmektedir.

Geliştirdiğimiz sistem fizyoterapi egzersizlerinde bulunan üç ana bileşeni: hareket örüntülerini [10], [11], duruş bilgilerini [12], egzersiz nesnelerini ayrı ayrı birer tanıma problemi olarak ele alınmıştır. Bu öznitelik kaynakları üretken bir Bayes ağında birleştirerek, hareketli görüntülerden oluşan egzersiz seansında tekrarlanan egzersizin tipini yüksek başarı oranıyla tanımıştır [12], [13]. Egzersizin tipi tanındıktan sonra hasta tarafından egzersizin seans içerisinde kaç defa tekrar edildiği bilgisine ulaşmak için ayrıca bir yöntem oluşturulmuştur [12], [13]. Bu yöntem aracılığıyla egzersiz tekrar adedi yüksek başarı oranıyla tahmin edilmiştir.

Tez çalışması boyunca geliştirilen sistemler bu doküman kapsamında modüler bir biçimde ele alınmıştır. Şekil 1.1'de verilen görsel ile, tez çalışması kapsamında oluşturulan veri kümesi ve kullanılan diğer veri kümeleri, geliştirilen özellik çıkarma yöntemleri ve kullanılan özellik çıkarma yöntemleri ile tasarlanılan özgün sistemler arasındaki bağlantılar rahatlıkla gözlemlenebilmektedir. Geliştirilen sistemler eylem tanıma konusunda yoğunlaşmıştır. Diğer çalışmalardan alınarak kullanılan yöntemlerin detayları ve kullanılan hazır veri kümeleri hakkında detaylı bilgi Şekil 1.1'de gösterildiği üzere ilgili bölümlerde anlatılmıştır.

Bu tez kapsamında, ikinci bölümde hareketli görüntülerdeki insan eylemlerinin tanınması ana problemi için literatürde varolan önceki çalışmalar incelenmiştir. Üçüncü bölümde kullanılan veri kümeleri hakkında bilgi verildikten sonra oluşturulan fizik tedavi egzersizlerini barındıran Ev Ortamındaki Fizik Tedavi Egzersizleri (EOFTE) adlı veri kümesi tanıtılmıştır. Dördüncü, beşinci, altıncı bölümlerde tez kapsamında geliştirilen özgün hareket ve duruş tabanlı öznitelik çıkarma yöntemleri hakkında bilgi verilmiştir. Yedinci bölümde hareket ve nesne bilgilerini kullanarak eylem tanıyan sistemin, sekizinci bölümde ise hareket ve duruş bilgilerini kullanarak eylem tanıyan sistemin detayları ve deneysel sonuçları verilmiştir. Dokuzuncu bölümde hareket, duruş ve nesne bilgisini kullanarak fizik tedavi egzersizlerini analiz eden temel ve geliştirilmiş sistemlerin detayları ve elde edilen deneysel sonuçlar aktarılmıştır. Son kısım olan onuncu bölümde ise tez kapsamında yapılan özgün çalışmalar genel olarak değerlendirilmiş ve gelecekte yapılabilecek çalışmalar hakkında öngörülerde bulunulmuştur.




Şekil 1.1: Tez çalışması kapsamında geliştirilen yöntemlerin, sistemlerin ve veri kümelerinin arasındaki bağlantılar.

# 2. ÖNCEKİ ÇALIŞMALAR

Literatürde var olan çalışmalar konularına ve kapsamlarına göre bölümlendirilerek tez çalışmasına uygun konu başlıkları altında derlenmiştir. Öncelikle tez kapsamında geliştirilen öznitelik çıkarma yöntemlerinin diğer yöntemlerden farklarını anlatabilmek amacıyla öznitelik çıkarımı başlığı oluşturulmuş ve bu başlık altında literatürdeki ilgili yöntemler hakkında bilgiler verilmiştir. Daha sonra tez kapsamında geliştirilen eylem tanıyan sistemlerin karşılaştırılabilmesi için literatürde varolan eylem tanıyan sistemler hakkında bilgi verilmiştir. Son olarak ise fizik terapi egzersizlerini konu alan çalışmalar hakkında bilgi verilmiş ve bunların başarılı ve başarısız yönleri hakkında bilgi verilmiştir.

## 2.1. Öznitelik Çıkarımı

Tez kapsamında sabit ve hareketli görüntülerden çıkartılabilecek öznitelikler üç bölümde toplanmıştır. Bunlar sırasıyla hareket tabanlı öznitelikler, duruş tabanlı öznitelikler ve nesne tabanlı özniteliklerdir. Özniteliklerin tipleri için bu üç ana bölüm daha da genişletilebilir ama geliştirilen öznitelik çıkarma yöntemlerinin karşılaştırılabilmesi açısından üç kategori ile sınırlandırılmıştır.

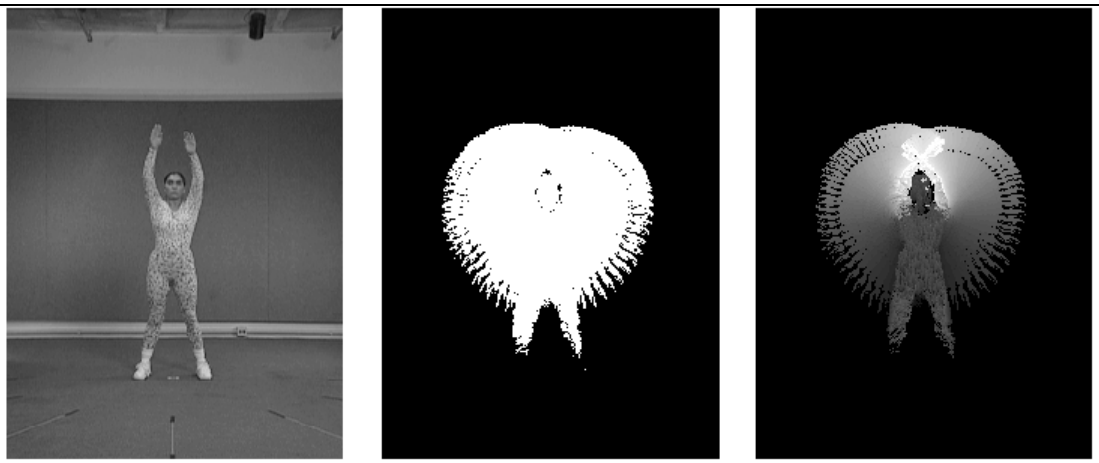
### 2.1.1. Hareket Tabanlı Öznitelik Çıkarma Yöntemleri

Hareket bilgisi bir hareketli görüntüde, nesnelerin, kameranın ve ışık kaynağının yer değiştirmesi gibi nedenler ile ortaya çıkan bir bilgi türüdür. Hareket bilgisi, eylem tanıyan sistemlerin öncelikle üzerinde durduğu bir öznitelik türüdür.

Hareket bilgisini sayısal dünya da tanımlamak için çeşitli yöntemler kullanılmaktadır. Bu yöntemler hareketli görüntüyü giriş verisi kabul ettikten sonra çıkış verisi olarak hareket bilgisini tanımlamak için kullanılan öznitelikleri vermektedir. Bu yöntemlerden sıkça kullanılanları ise Hareket Enerji Görüntüsü (HEG, ing. Motion Energy Image (MEI)) ve Hareket Geçmiş Görüntüsü (HGG) (ing. Motion History Image (MHI)) [14], Radon Dönüşümü [15], [16], Optik Akış (ing. Optical Flow) [17], [18], Gabor filtreleme [19], [20], Zamansal ızgaralar [21], Gradyanlar histogramı (ing. Histogram of Gradients (HoG)) filtreleme [22] ve Akış histogramı (ing. Histogram of Flow (HoF)) filtreme [23] şeklinde sıralanabilir.

Bobick ve Davis [14] geliştirdikleri yöntemi kullanabilmek için öncelikle bilinen arka plan (ing. background) görüntüsünü kullanarak verilen test görüntüsünde arka plan çıkarımı (ing. background subtraction) uygulamışlar ve ön plan (ing. foreground) görüntüsüne ulaşmışlardır. Ön plan görüntüsünde sahnede daha önceden sabit olarak bulunmayan nesnelerin bölgeleri bulunmuştur. Bu bölgelerde aktörün olduğu varsayımından yola çıkarak arka plan pikselleri siyah renk ile ön plan pikselleri beyaz renk ile gösterilerek görüntü siyah-beyaz hale gelmiştir. Bu gösterim şekline silüet (ing. silhouette) denilmektedir. Takip eden silüet çerçeveleri arasında piksel bazında çıkarma işlemi yapılarak hareketin olduğu pikseller tespit edilmiştir. İkili renkten oluşan bu fark görüntüsüne HEG adı verilmiştir. HEG incelendiğinde hareketin olduğu piksellerin dolayısıyla hareketin nerede olduğu sorusunun yanıtına ulaşılmıştır. Nasıl bir hareket yapılmıştır sorusuna yanıt arayan araştırmacılar HEG tabanlı yeni bir hareket özniteliği tanımlamışlardır ve adını HGG koymuşlardır. HGG görüntüsü belirli bir zaman aralığında HEG'lerin değişimini incelemektedir. Örneğin 256 renk gri seviyesi (ing. gray-level) ile tanımlanan HGG'de 0 siyah, 255 beyaz rengi tasvir ettiği kabul edildiğinde seçilen çerçeveye yakın zamanda yapılan hareketi barındıran piksellerin daha beyaza yakın, daha geçmişte yapılan hareketi barındıran piksellerin daha siyaha yakın gri tonlarla gösterilmektedir. HGG ile amaçlanan, hareketli görüntüde yapılan eylem boyunca hareketin hangi iki boyutlu konumda, zaman içerisinde nasıl değiştiğinin gözlemlenmesidir. Şekil 2.1 a)'da iki elini başının üstünde yere dikey olarak tutan bir aktörün örnek görüntüsü, Şekil 2.1 b)'de ellerini aşağı doğru hareket ettirmesiyle oluşan HEG örneği, Şekil 2.1 c)'de ise aynı hareket sonucunda oluşan HGG örneği verilmiştir [14]. Araştırmacılar geliştirdikleri hareket tabanlı öznitelik vektörlerini olan HGG'ları tüm eğitim kümesi (ing. training set) ve test kümesinde (ing. testing set) çıkarmışlar ve bu HGG'ları kullanarak Hu momentlerini [24] oluşturmuşlardır. En sonunda elde ettikleri Hu momentlerini kullanarak basitçe eğitim ve test örnekleri arasında şablon eşleştirme (ing. template matching) tekniğini kullanarak eylem tanıma yapmışlardır.

Davis ve Bobick [25] geliştirdikleri hareket özniteliklerini [14] kullanan sanal bir aerobik egzersiz yönlendirme sistemi tasarlamışlardır. Bu sistemde kameralar tarafından yakalanan hareketli görüntü ve egzersizi yapan şahsın hareketleri arasında karşılaştırma yapılmakta ve egzersizi yönlendiren eğitimci uyarılmaktadır. Egzersizi yapan şahısın yapması gereken eylemler ise klipler (ing. clips) şeklinde görsel olarak şahsın görebileceği bir ekrana yansıtılmaktadır.



a) b) c)

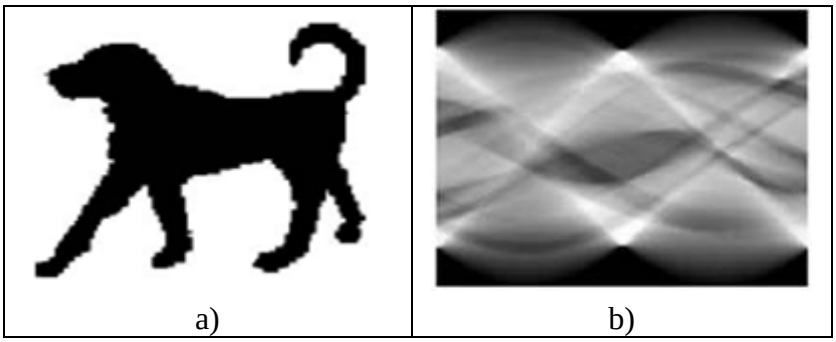
Şekil 2.1: Hareket Enerji Görüntüsü (HEG) ve Hareket Geçmiş Görüntüsü (HGG) örneği. a) Gri tonlarda egzersiz görüntüsü, b) egzersizin Hareket Enerji görüntüsü, c) egzersizin Hareket Geçmiş görüntüsü.

Araştırmacılar, Bobick ve Davis'in [14] tanıttığı hareket tabanlı öznitelikleri temel alarak bir çok yeni yaklaşım geliştirmişlerdir. Bu yeni yaklaşımlar ise Ahad arkadaşları [26] yaptığı çalışmada detaylıca incelenmiştir.

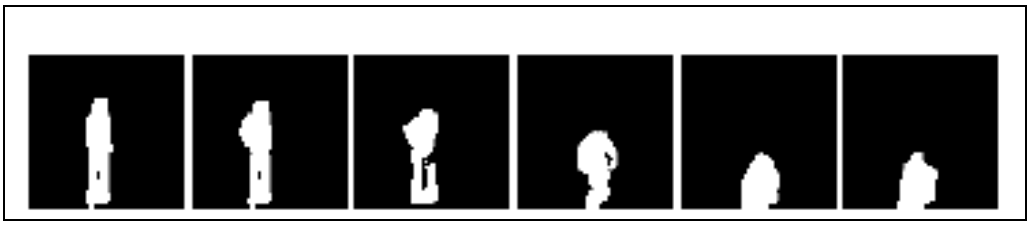
Radon dönüşümü, Deans tarafından [27] tanımlanmıştır. Radon dönüşümü sabit görüntüler için hesaplanılır. Bu hesaplamada görüntü üzerinde farklı açılarda doğrular olduğu varsayılır ve bu doğrulara denk düşen projeksiyonlardan kümeler oluşturulur. Görüntü işlemede (ing. image processing), Hough dönüşümü (ing. Hough transform) gibi görüntülerde doğruların bulunmasında kullanılmaktadır. Şekil 2.2'de siyah-beyaz bir köpek görüntüsü ve bu görüntünün Radon dönüşümü görsel olarak verilmiştir.

Tabbone arkadaşları [28] Radon dönüşümünü geliştirerek $\mathfrak{R}$ dönüşümünü tanımlamışlardır. $\mathfrak{R}$ dönüşümü basitçe Radon dönüşümünün ölçeklemeye (ing. scaling) ve yer değiştirmeye karşı bağımsız hale getirmek olarak ifade edilebilir. Wang arkadaşları [15] $\mathfrak{R}$ dönüşümünü hareketli görüntünden elde edilen silüet görüntüler üzerine uygulayarak yerel öznitelikleri oluşturmuştur. $\mathfrak{R}$ dönüşüm yüzeyi ise zaman ekseninin (ing. axis) de eklenmesiyle her silüetten elde edilen $\mathfrak{R}$ dönüşümlerinden hareket tabanlı öznitelik çıkarma amacıyla kullanılmıştır [16]. Şekil 2.3'de IXMAS (ing. INRIA Xmas Motion Acquisition Sequences) veri kümesinden [28] alınan, içerisinde oturma eylemi bulunan, bir hareketli görüntünün

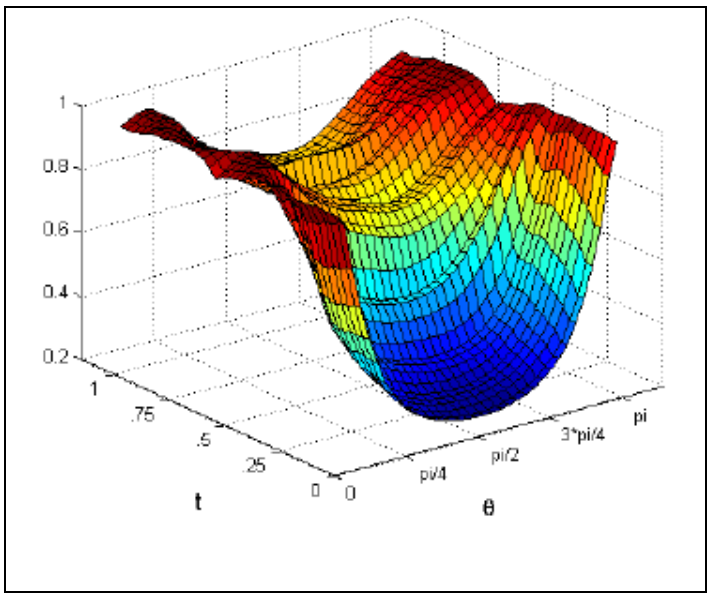
ana çerçevelerini gösteren silüet görüntü serisi verilmiştir. Hareketli görüntüden elde edilen tüm silüet görüntülerinin kullanılmasıyla oluşturulan $\Re$ dönüşüm yüzeyi ise Şekil 2.4'de verilmiştir.




Şekil 2.2: a) Örnek bir köpek görüntüsü, b) görüntünün Radon dönüşümü.



Şekil 2.3: Oturma eylemi hareketli görüntüsünün ana çerçeveleri.




Şekil 2.4: Şekil 2.3'de verilen oturma eyleminin hareket özniteliklerinin Radon dönüşüm yüzeyi.

Optik akış kişinin/kameranın görüş alanındaki nesnelerin hareketi olarak basitçe tanımlanabilir. Optik akış alanı (ing. optic flow field) ise görüntü düzleminde kolayca anlaşılabilen parlaklık örüntülerinin hızlarının (ing. velocity) kümesi olarak tanımlanır [29]. Görüntüdeki her piksel için verilen zaman diliminde yatay ve dikey olarak hızın hesaplanılması mantığıyla çalışmaktadır [17], [18]. Optik akış hareket tabanlı özniteliklerin çıkarımında kullanılmakta olan bir yöntemdir.

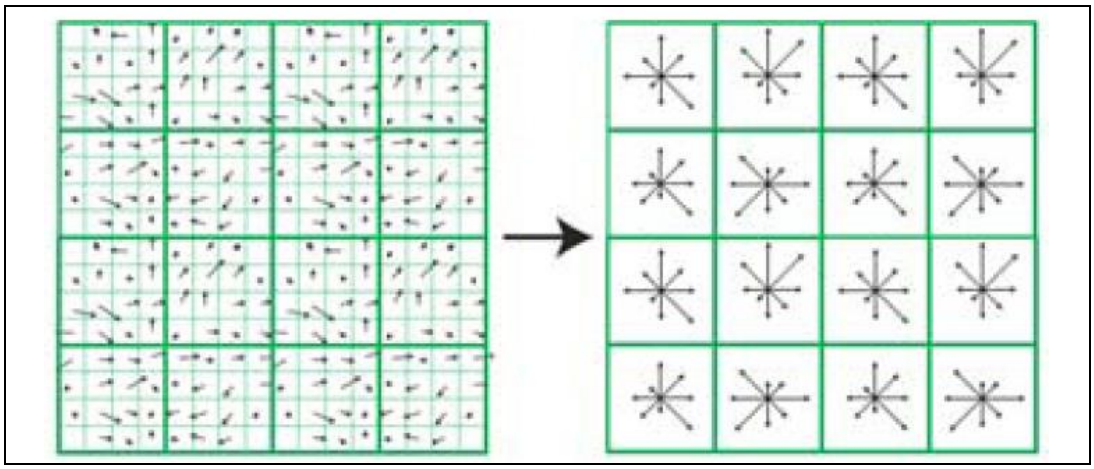
Efros arkadaşları [17], hareketli görüntünün tüm çerçevelerinde aktörü bularak pencere içine aldıktan sonra bu penceredeki görüntülerin boyutlarını aynı seviyeye getirmiştir. Daha sonra takip eden pencerelerdeki optik akışı her çerçeve için hesaplamıştır. Optik akış vektör alanını yatay ve dikey bileşenleri gözeterek önce iki parçaya bölümlemiş daha sonra ise bu bölümlenmiş parçalarını yarım dalga doğrultma (ing. half wave rectification) yaklaşımıyla tekrar iki bileşene ayırmıştır. Elde ettiği dört farklı görüntüye Gauss filtresi uygulayarak bulanıklaştırmış (ing. blurring) ve bunların tümünü hareket tabanlı öznitelikler olarak tanımlamış ve eylem tanıma probleminde bir özellik kümesi olarak kullanmıştır.

Danafar ve Gheissari [18] hareketli görüntülerdeki her çerçevede insanı tespit edip pencere içerisinde aldıktan sonra pencere içerisinde insanın dikey pozisyonda olduğu varsaymıştır. Bu varsayımdan yola çıkarak (örn. koşma, yürüme, el sallama eylemleri) pencere içerisinde yukarıdan aşağıya 1/5, 2/5, 2/5 oranında sabit bölümleme yaparak insan bölgelerini (baş-boyun, gövde-eller, bacaklar) elde etmişlerdir. Her bölgedeki hareketi ayrı ayrı yorumlamak için optik akış yöntemini kullanarak her bölge için dört farklı kanaldan oluşan hareket tabanlı öznitelikleri elde etmiş ve bunları eylem tanımada özellik olarak kullanmıştır.

Laptev arkadaşları [23] hareket tabanlı öznitelikleri elde etmek için öncelikle hareketli görüntülerdeki ilgi noktalarını (ing. interest points) tespit etmiş ve üç boyutlu uzayda komşu olan noktaları küboidler (ing. cuboids) olarak gruplamıştır. Bu küboidler içerisinde hareketi tanımlamak için ise optik akışları elde etmiştir. Optik akışlar piksel bazında elde edildiği için bunların histogramını alarak hareket tabanlı öznitelikleri tanımlamış ve bunları eylem tanımada kullanmıştır.

Chen ve Hauptmann [30] gözetim hareketli görüntülerinden insan eylemlerini tanımak üzere Hareket-tabanlı Ölçekle Değişmeyen Özellik Dönüşümü (HÖDÖD) (ing. Motion Scale-Invariant Feature Transform, MoSIFT) adı ile bir algoritma oluşturmuşlardır. Bu algoritma ilgi noktalarının bulunmasıyla sadece yerel görünüşün değil yerel hareketin de kodlanmasıdır. Önerdikleri yaklaşımı üç ana

bölümde incelemişlerdir. Bunlar ilgi noktalarının bulunması, öznitelik tanımlayıcısının (ing. feature descriptor) oluşturulması ve sınıflandırıcının inşa edilmesidir. Basit olarak, ilgi noktalarının bulunması kısmında Ölçekle Değişmeyen Özellik Dönüşümü (ÖDÖD) (ing. Scale-Invariant Feature Transform, SIFT) özelliklerine ek olarak hareket bilgisini de kullanmak istemişler ve aynı piramitsel yapı içerisinde, komşu çerçevelerle incelenen çerçeve arasındaki farkları tespit ederek hareket bilgisine ulaşmışlardır. ÖDÖD yapısındaki gradyanlara benzer olarak HÖDÖD yapısında optik akışlar rol almaktadır. Kullanılan veri kümesi gözetim (ing. surveillance) hareketli görüntülerinden oluştuğundan optik akış yönteminin eksik olduğu: oryantasyondan (ing. orientation) bağımsız olma durumu, göz ardı edilmiştir. Hem görünüş hem de hareket özelliklerini barındıran histogramlar birleştirilerek 256 boyutlu (128+128) bir HÖDÖD özellik tanımlayıcısı oluşturmuşlardır. HÖDÖD ve ÖDÖD özellik tanımlayıcılarının ızgarasal gösterimindeki bir örnek Şekil 2.5'te verilmiştir [30].



Şekil 2.5: Izgarasal gösterimde ÖDÖD ve HÖDÖD özellik tanımlayıcıları.

Chen ve Hauptmann, öznitelik tanımlayılarını oluşturduktan sonra bunların yorumlanması aşamasına geçmiştirler. Bu aşamada sık kullanılan Kelimeler Torbası (ing. Bag-of-Words, BoW) yapısı yerine doğal dil işleme ve metin sınıflandırmada sıkça kullanılan Bi-gram modelleri [31] incelenmiştir. Görünüş ve hareket özelliklerini K-Ortalamalar kümelemesi (ing. K-Means Clustering) algoritması kullanılarak kümelendikten sonra her bir kümeye ait elemana farklı kelimeler atanmıştır [30]. Bu farklı kelimeler ayrıştırıp bi-gram modelleri oluşturulmaya çalışıldığında çok fazla işlenecek veri ortaya çıkmıştır. Bunu engellemek için en sık

tekrar eden n tane kelimenin bi-gram modelleri oluşturulmuştur. Son aşama olan sınıflandırma aşamasında ise polinom çekirdekli (ing. kernel) destek vektör makineleri (DVM) (ing. Support Vector Machines, SVM) kullanılmıştır. 2009 yılında yapılan bu çalışmada elde edilen sonuçlar dikkat çekicidir. KTH (Kungliga Tekniska Högskolan) [87] veri kümesinde %95,83 başarı oranıyla eylemler tanınmıştır. Bu sonuç 2009 yılına kadar yapılan tüm çalışmalardan haylice yüksek olup hareket özniteliklerinin görünüş öznitelikleriyle birlikte nasıl başarılı bir şekilde kullanılabileceğinin en güzel örneklerinden biri olarak günümüzde halen bir çok araştırmacıya yol göstermektedir. Elde edilen bu sonuç görünüş özellikleriyle aktörün duruş bilgisinden faydalanıldığını göstermektedir. Ayrıca hareket ve duruş özniteliklerinin beraber kullanıldığında daha faydalı olabileceği gözlemlenmiştir.

### 2.1.2. Duruş Tabanlı Öznitelik Çıkarma Yöntemleri

İnsan eylemlerini tanımayı hedefleyen sistemler insanın duruş (ing. pose) bilgisinden faydalanabilmektedir. Her eylemin farklı bir duruşunun ve/veya duruşlardan oluşan bir kümesinin varolduğu varsayımından yola çıkan bu yöntemleri kullanan çeşitli çalışmalar bulunmaktadır. Bu varsayıma örnek olarak, futbol oynayan bir insanın duruşunun görüntüsü, kitap okuyan bir insanın duruşunun görüntüsü, yatan bir insanın duruşunun görüntüsü ve ayakta duran bir insanın duruşunun görüntüsü verilebilir. Örneklerdeki duruş görüntüleri eylem hakkında ayırt edici bilgiye sahiptir

Duruş tabanlı özniteliklerin eylem tanıma sistemlerinde kullanılmasıyla elde edilen sınıflandırma başarısının, insanın görünüşü tabanlı eylem tanıma sistemlerine oranla daha yüksek olduğu vurgulanmıştır [32]. Bu öznitelikler bir görüntüden ve/veya hareketli görüntülerden çıkarılabilmektedir.

Literatürde duruş tabanlı öznitelikler statik ve dinamik öznitelikler olarak iki ayrı kategoriye ayrılmıştır [1], [33]. Dinamik öznitelikler zaman içerisinde duruş bilgisinin değişimini incelemekte, statik öznitelikler ise zamandan bağımsız olarak tek görüntü üzerinde duruş bilgisini incelemektedir.

Duruş özniteliklerini çıkartan ve bu öznitelikleri insan eylemlerini tanımada kullanan çalışmalarda görüntüde insan vücudunun tespit edilmesi esas alınmaktadır. İnsan vücudunun görüntüdeki yeri tespit edildikten sonra tüm vücuttan elde edilen duruş öznitelikleri oluşturulabildiği gibi [34]-[36] eklem yerlerinin ve benzeri özel

noktaların birbirlerine göre konumları incelenerek duruş öznitelikleri de oluşturulabilmektedir [32], [37].

Literatürde görüntü ve/veya hareketli görüntüler için duruş özniteliğinin elde edilebilmesi için çok çeşitli yöntemler oluşturulmuştur. Bunlardan bazıları: Temel Bileşen Analizi (ing. Principal Component Analysis) kullanarak özşekiller (ing. eigenshape) ile [34], Gradyanlar Histogramı yöntemiyle Temel Bileşen Analizini birleştirerek [35], 3-boyutlu uzam zaman şekillerini elde ederek [38], uzamsal (ing. spatial) alanda vücut parçalarını kapsayacak dikdörtgen alanlar tespit edilip bu alanlardaki çeşitli dağılımların histogramı alınarak bilgileriyle ifade edilmesiyle [39], silüet görüntüleri kullanarak insanın konturlarının bulunması ve bunların merkeze bağlı uzaklığının hesaplanılmasıyla [36], eklem noktaları tespit edilip zaman içerisinde hareketlerinin incelenmesiyle [40], 5 farklı eklem noktasının tahmin edilip uzamsal incelenmesiyle [37], şeklinde sıralanabilir.

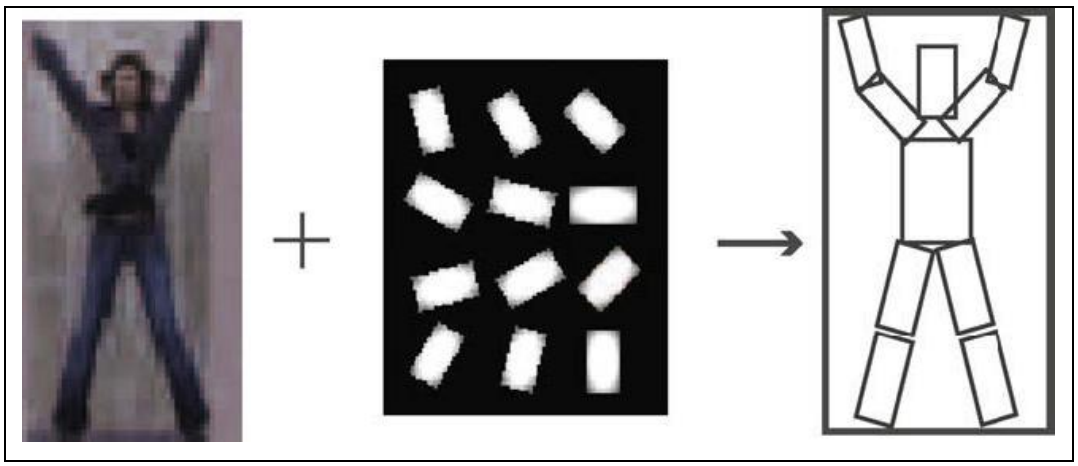
Şekil 2.6'da kollarını açarak zıplama, yürüme ve koşma eylemlerinin zaman içerisinde değişen 3-boyutlu uzam zaman şekilleri gösterilmiştir [38]. Şekil 2.7'de ise çeşitli büyüklüklerde ve çeşitli yönlerde oluşturulan dikdörtgen alanların görüntüdeki vücut parçalarını kapsayacak bir biçimde dağıtılması gösterilmiştir [39].

Duruş özniteliğini çıkararak insan eylemlerini tanıma çalışmalarında yakın zamanda ise hareketli ve/veya sabit görüntülerde eklemlerin 2 veya 3 boyutlu koordinatlarına göre iskelet modelleme çalışmaları ön plana çıkmaktadır [41]. Bu çalışmalarda eklem yerlerinin tespit edilmesi için genellikle derinlik bilgisini kullanılmaktadır. Derinlik bilgisini almak için çeşitli algılayıcılar kullanılmaktadır. Ayrıca eklem noktalarını tespit etmek için de çeşitli yazılımlar ve donanımlar barındıran sistemler kullanılmaktadır.

Bu tez çalışması kapsamında silüet görüntülerde duruş tabanlı öznitelik çıkarma üzerinde yoğunlaşılmıştır ve Cheema arkadaşları [36] tarafından geliştirilen duruş tabanlı öznitelik çıkarma yöntemi, tez kapsamında geliştirilen bazı sistemlerde kullanılmıştır.

a) b) c)

Şekil 2.6: a) Zıplama, b) yürüme, c) koşma eylemlerinin uzam zaman şekilleri.



Şekil 2.7: Dikdörtgen alanlarla vücut parçalarının kapsanması.

### 2.1.2.1. Cheema Arkadaşlarının Duruş Tabanlı Öznitelik Çıkarma Yöntemi

Cheema arkadaşları duruş bilgisinin, insan eylemlerini tanımada önemli bir ayırt edici öznitelik kümesi olduğunu belirtmişlerdir. Bazı basit ve/veya temel eylemlerin, hareket bilgisine gerek kalmadan, sadece bir görüntüden tespit edilebileceği öne sürmüşlerdir. Hareket bilgisinin yerine duruş bilgisinin incelenerek aktörün pozundan yaptığı eylemin tanınabileceğini savunmuşlardır [36].

Duruş bilgisini kullanarak eylem tanıyan sistemler genellikle hareketli görüntülerdeki anahtar pozları (ing. key poses) çıkararak bunları tanımlamak için çeşitli metotlar kullandıklarını belirtmişlerdir. Anahtar pozlar ile ifade edilmek

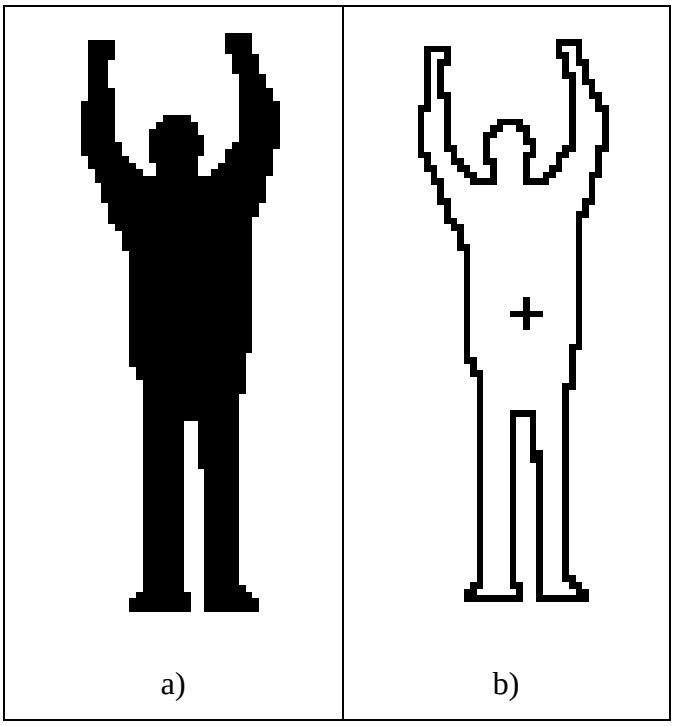
isteneni bir örnek senaryoda düşünürsek: voleybolda smaç servis atan oyuncu için poz bilgi “ayakta durma”, “topu havalandırma”, “sıçrama” ve “havada topa vurma” duruşları bu eylemin anahtar pozları olarak tanımlanabildiğini görürüz.

Duruş bilgisini çıkaran metotlar aracılığıyla ilgili görüntüdeki duruş özniteliklerinin çıkarıldığı açıklanmıştır. Duruş özniteliklerini elde etmek için bazı yöntemler: insan vücudunun düşük seviye görsel öznitelikler ile (örneğin renk bilgisi) modellenmesi [42], gradyanlar histogramı tanımlayıcısının kullanılması [33], insan görüntüsünün çevresinin çift çizgiler ile tanınması [43], şeklinde sıralanabilir.

Cheema arkadaşları [36] geliştirdiği eylem tanıma sistemiyle verilen silüet görüntülerinden öncelikle görüntüde aktör boyutundan bağlı değişmeyen kontur tabanlı duruş öznitelikleri çıkarılmıştır. Bu öznitelikler daha sonra kümelenerek daha genel bir tanımlayıcı elde edilmiştir. Anahtar pozların eylemlerdeki ağırlıklarına göre bu tanımlayıcının değerleri hesaplanılarak eylemler kendi aralarında değersel olarak farklılaştırılmış ve gelen örneklerdeki eylemler bu sistem aracılığıyla sınıflandırılmıştır.

Cheema arkadaşları [36] geliştirdiği eylem tanıma sistemde bir modül olarak görev alan duruş tabanlı öznitelik çıkarma yöntemi sabit görüntülerde ve hareketli görüntülerde çalışabilmesi, görüntüdeki aktör boyutundan bağımsız olması, hızlı ve başarılı sonuçlar vermesinden dolayı tez kapsamında geliştirilen bazı sistemlerde duruş özniteliklerini elde etmek için kullanılmıştır. Ayrıca tez çalışmasında geliştirilen duruş tabanlı öznitelik çıkarma yöntemini karşılaştırmak içinde rakip yöntem olarakda kullanılmıştır.

Duruş tabanlı özniteliklerin elde edilmesi işlemi için silüet görüntüler kullanılmıştır. Bu silüet görüntüler ilgili veri kümelerinde hazır olarak bulunabileceği gibi basit bir arka plan çıkarması ile de elde edilebilir. Silüet görüntülerde ön plan aktörü temsil eden siyah piksellerden arka plan ise beyaz piksellerden veya tam tersi olmak üzere iki renkten oluşur. Şekil 2.8’de a) bendinde, ayakta durup iki elini sallayan bir insanın silüet görüntüsü örnek olarak verilmiştir. Şekil 2.8’de b) bendinde aynı şeklin a) bendinde verilen silüet görüntüsünün ağırlık merkezi + simgesi ile gösterilmiştir.

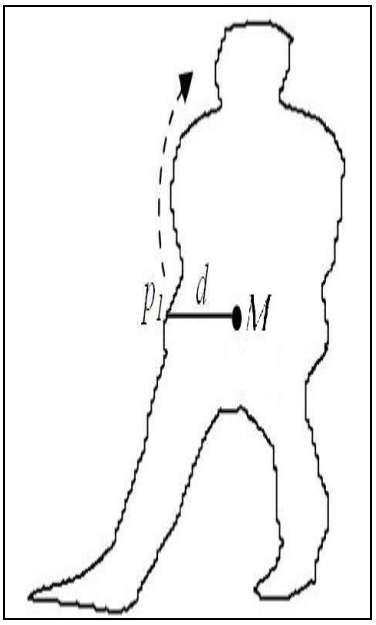



Şekil 2.8: a) Örnek bir silüet görüntüsü, b) ilgili kontur görüntüsü.

Silüet görüntünün x ve y ekseninde ağırlık merkezi $M = (x_m, y_m)$, $n$ değeri ön plandaki piksellerin toplam sayısını belirtmek üzere Eşitlik (2.1) ile hesaplanılmıştır:

$$x_m = 1/n \sum_{i=1}^{n} x_i, \quad y_m = 1/n \sum_{i=1}^{n} y_i . \tag{2.1}$$

$M$ noktası hesaplanıldıktan sonra silüet görüntüsünde konturlar hesaplanılmıştır. Toplam adedi $s$ olan tüm kontur noktaları $P$, $M$ noktasının tam solundaki kontur noktasından $p_1$ başlanılarak saat yönünde sıralanarak $P = [p_1, p_2, ..., p_s]$ elde edilmiştir. Bu tanımlayıcı değerler örnek bir görüntü üzerinde Şekil 2.9'da verilmiştir [36].




Şekil 2.9: Kontur noktalarının, ağırlık merkezine göre sıralanması.

P noktaları ile ağırlık merkezi arasındaki uzaklıklar (d), öklid mesafesi baz alınarak $j = 1,2,...,s$ olmak üzere Eşitlik (2.2)'deki gibi hesaplanıldıktan sonra değişken görüntü boyutlarında elde edilen farklı kontur sayılarından ($s$) düzgün dağılımda $r$ tane değer seçilebilmesi için Eşitlik (2.3) kullanılmıştır. Duruş tabanlı öznitelikleri aktör ve/veya görüntü boyutuna bağımlı olmaktan kurtarmak için tasarlanılan bu seçim işleminde $\lceil \cdot \rceil$ tavan fonksiyonu olarak kullanılmıştır.

$$d_j = \| p_j - M \| , \tag{2.2}$$

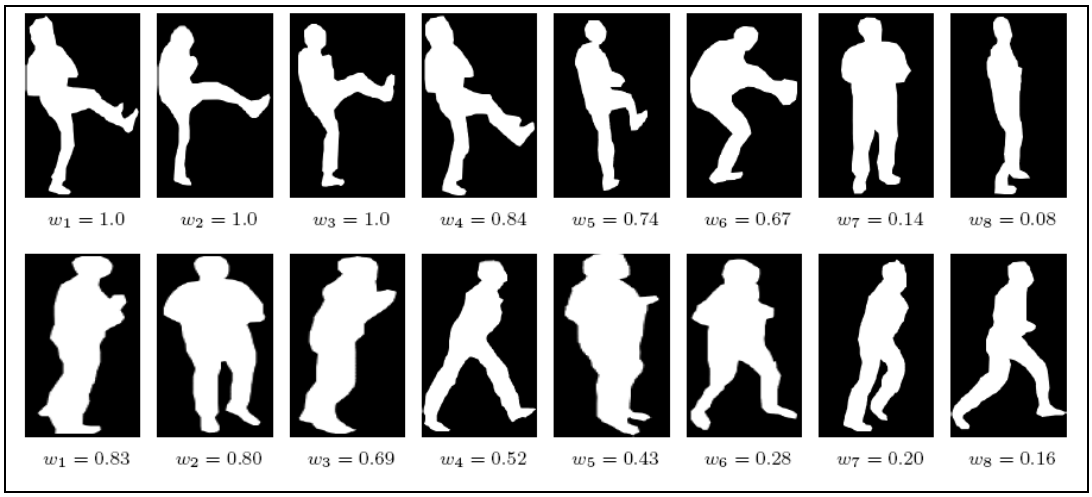
$$D[j]=d\left\lceil \frac{j*s}{r} \right\rceil, \tag{2.3}$$

Son olarak seçilen sabit sayıda kontur noktasının merkeze bağlı öklid mesafesini barından D vektörü normalize edilerek duruş tabanlı öznitelikleri temsil eden $\overline{D}$ vektörü haline Eşitlik (2.4) kullanılarak getirilmiştir:

$$\overline{D[j]} = \frac{D[j]}{\Sigma_1^r D[j]}. \tag{2.4}$$

Tez çalışması dahilinde Cheema arkadaşları [36] tarafından tasarlanılan duruş tabanlı öznitelik çıkarma yöntemini kullanılmıştır. Her bir hareketli görüntü çerçevesi için elde edilen $\overline{D}$ vektörü, başka bir açıklama yapılmadığı takdirde, duruş bilgisi temsil etmek için kullanılmıştır.

Cheema arkadaşları [36] duruş tabanlı özniteliklerini verilen tüm hareketli görüntülerin her bir görüntüsünde hesapladıktan sonra k-ortalamalar kümelemesi yöntemini kullanarak *k* tane küme ve *k* tane küme merkezi elde etmişlerdir. Küme merkezleri, özetle, birbirlerine yakın duruş bilgilerinin en baskın, en tanımlayıcı noktalarıdır. Her bir eylemde birden fazla duruş bilgisi bulunmaktadır. Zaman içerisinde duruş bilgileri değişmektedir. Eylemlerden bazıları zaman içerisinde aynı duruş bilgisine sahip olabilmektedir. Dolayısıyla bazı duruş bilgileri eylemlere özgü bazı duruş bilgileri ise daha genel bir yapıya sahiptir. Bu noktada bir ayrıma gidilebilmesi için duruş bilgisi küme merkezleri için basit bir ağırlıklandırma çalışma yapılmıştır. Bu ağırlıklandırma işlemiyle eylemde daha az rastlanılan duruş bilgileri daha yüksek bir değere, eylemde daha sık rastlanılan duruş bilgileri daha düşük bir değer ile ilişkilendirilmiştir. Örnek bir eylemdeki ağırlıklandırma işlemiyle elde edilen değerler çerçeve bazında Şekil 2.10'da [36] verilmiştir. Bu değerler aracılığıyla gelen test örneklerinde bulunan eylemin sınıflandırılması ise ağırlık oylaması yöntemiyle yapılmıştır.




Şekil 2.10: Sağa tekme ve yumruk eylemlerinin duruş ağırlık değerleri.

### 2.1.3. Nesne Tabanlı Öznitelik Çıkarma ve Nesne Tespit Yöntemleri

Görüntülerde nesnelerin tespit edilmesi, takip edilmesi, sınıflandırılması ve benzeri problemlerin çözülmesi konusunda çok çeşitli çalışmalar yapılmıştır. Bu çalışmaların çoğunluğunda varolan nesne örnekleri ve/veya modelleri test görüntülerinde bulunmasını takiben bu bulunan nesnenin türünün ne olduğu sorusuna yanıt aranmıştır [44].

Günümüzde halen Bilgisayarla Görme alanında önemli bir araştırma konusu olan nesnelerin tanınması, eylem tanıma problemi ile ilişkilidir. Bu ilişkiyi tanımlamak için öncelikle nesnelerin fonksiyonellikleri üzerine çeşitli çalışmalar yapılmıştır [45]. Nesnelerin fonksiyonelliklerine göre kategorize edilmesi robotik başta olmak üzere çeşitli uygulama alanları için çok önemlidir. Bir sürahide bulunan sıvının dökülebilir olması, bir camın kırılabilir olması, bir sandalyenin üzerine oturulabilir olması ve benzeri örneklerle anlatılabilecek nesne fonksiyonelliğinin belirlenmesi konusunda çeşitli çalışmalar mevcuttur [46], [47].

İnsan eylemleri ve nesneler arasındaki ilişkinin birlikte incelenmesiyle hem insan eylemlerinin tanınması probleminde hem de nesnelerin tespit edilip sınıflandırması probleminde, elde edilen başarıların arttığı gözlemlenmiştir [48], [49]. Örneğin; insanın bir bardağı tutup ağzına doğru getirmesi eylemini gördüğümüzde içme eyleminin yapıldığına kanaat getirilmiştir [50]. Verilen bu örnekte nesne bilgisi, bize yapılan eylem hakkında önemli bir ipucu sağlamıştır.

Eylem ve nesne arasındaki ilişki iki farklı yaklaşımla kullanılmıştır. Yapılan eylemin ne olduğu bilgisinin kullanılmasıyla nesnenin sınıfı bulunmuş [51] veya nesnenin ne olduğu bilgisinin kullanılmasıyla yapılan eylemin sınıfı bulunmuştur [52].

Eylem ve nesne arasındaki ilişki basit olarak “çatala dokunma”, “kaşığı tutma”, “oyuncak arabayı itme” ve benzeri eylemlerde incelendiği gibi [53] daha karmaşık olan çeşitli spor hareketlerinin (voleybolda smaç, teniste servis v.b.) [54] tanınmasında yarar sağlamıştır.

Fizik tedavi egzersizlerinde ise nesnelerin kullanılmasının bir çok yararı bulunmaktadır [55]. Egzersizlerin nesneler kullanılarak desteklenmesiyle hastaların daha hızlı bir iyileşme sürecine girdiği gözlemlenmiştir [56]. Hastalar, egzersiz nesneleri ile etkileşim içerisinde bulunduğu egzersizlerde (örneğin paralel yürüme

çubuklarını kullanarak düzgün bir biçimde yürüme egzersizi) nesnelerden destek aldığını ve kendilerini daha güvende hissettiklerini belirtmişlerdir [8].

Fizik tedavi merkezlerindeki jimnazyumlarda hastalar doktorların gözetiminde verilen egzersizleri yaparken kullandıkları nesnelere örnek olarak egzersiz istasyonları, egzersiz topları, elastik bantlar, paralel yürüme çubukları, pedal egzersiz setleri verilebilir. Evde yapılan fizik tedavi egzersizlerinde ise günlük kullanımda her evde bulunabilen nesnelerden (örneğin havlu, sandalye v.b.) yararlanılmaktadır [57]. Ayrıca, oluşturulan EOFTE veri kümesinde (Bölüm 3.4), ev ortamında rahatlıkla bulunabilecek nesnelerin kullanıldığı egzersiz tipleri mevcuttur. Bu veri kümesinde kullanılan üç ana nesne tipi: sandalye, sopa, ve havludur.

Tez çalışması kapsamında fizik tedavi egzersizlerinin tanınmasında üç ana bileşenden biri olarak karşımıza çıkan nesne bilgisi ile egzersiz tipi arasında ilişki kurulması hedeflenmiştir. Nesnelerin tespit edilip sınıflandırılması bu ilişkinin tanımlanması için ilk hedeftir. Bu hedef doğrultusunda literatürde var olan üç farklı yöntem incelenmiş ve kullanılmıştır. Bu yöntemler ilerleyen alt bölümlerde detaylı bir şekilde aktarılmıştır. Birden fazla yaklaşımının kullanılmasıyla bu farklı yaklaşımların birbirleriyle karşılaştırılması ve nesne tespit/sınıflandırma işleminin daha başarılı bir şekilde tamamlanması amaçlanmıştır.

#### 2.1.3.1. Fei-Fei'nin Nesne Tespit Yöntemi

Fei-Fei geliştirdiği sistem ile görüntülerde, verilen nesne kategorisine ait olan nesnelerin nerede olduğunu ve bulunan nesnenin sınıfının ne olduğunu yüksek başarı oranı ile tespit etmiştir [58]. Bu sistemde kullanılan yaklaşım, nesnelerin birden çok parçadan oluştuğu ve bu parçaların tespit edilmesiyle nesnenin sınıfı ve konumu hakkında bilgi edinilebileceği varsayımı üzerine kurulmuştur.

Fei-Fei'nin sistemi [58] sadece sabit görüntüler üzerinde çalışmaktadır. Sistem önce görüntü üzerinde ÖDÖD [59] noktalarını tespit eder. Daha sonra bu noktalardaki çeşitli öznitelikleri Kelimeler Torbası (ing. Bag of Words, BoW) yapısında kullanarak nesnelerin sınıflarını ve konumlarını tahmin etmektedir.

ÖDÖD noktalarının tespiti görüntülerin farklı ölçeklerde (ing. scale) küçültülmüş/büyütülmüş hallerinden elde edilmesiyle başlar, bu yeni görüntülerin alçak geçirgen filtrelerden (ing. low-pass filters) biri olan Gauss filtrelerinden geçirilerek gürültüden arındırılması ile devam etmektedir. Gürültüsü azaltılmış farklı

ölçeklerdeki görüntülerin arasındaki fark Gaussların Farkı (ing. Difference of Gaussians) yöntemi ile hesaplanır. Daha sonra bu fark değerleri kullanılarak uç noktalar bulunur. Uç noktalar bir noktanın 26 farklı komşu noktasına (yönteme özgü bir komşuluk sayısı) bakılarak belirlenir.

Aday ilgi noktası olarak da tanımlanabilecek uç noktalar belirlendikten sonra bunlardan kararlı olmayanların elenmesi gerekir. Kararlı olmayan noktalar düşük kontrasta (ing. contrast) sahip gürültüden daha çok etkilenebilen noktalardır. Bunlar çeşitli filtreler ve eşik seviyeler kullanılarak elenir. En sonunda kalan ilgi noktalarındaki öznitelikleri tanımlanması gerekmektedir. Bunun için dönmeden (ing. rotation) bağımsız bir kümede gradyanların büyüklüğü (ing. magnitude) ve oryantasyonları hesaplanılır. Bu oryantasyon ve büyüklük değerleri her ilgi noktası için hesaplanıldıktan sonra noktalar 4x4'luk komşuluk esasına göre kümelenir. Bu yeni alan içerisinde noktaların aynı açı değerlerine sahip büyüklük değerleri birbiriyle toplanılarak histogram bazlı bir gösterim elde edilir.

Yaptığımız tez çalışmasında Fei-Fei'nin sistemi [58] EOFTE veri kümesinde bulunan farklı nesnelerin tespiti için kullanılmıştır. Bu kullanım için her bir nesnenin 30 farklı görüntüsü (farklı oryantasyon ve farklı büyüklükteki) oluşturulmuştur. Sistem her bir görüntüde bu nesneleri aramakta ve bulduğunda nesnenin sınıf bilgisini ve konumunu geri dönmektedir.

### 2.1.3.2. Viola ve Jones'un Nesne Tespit Yöntemi

Viola ve Jones'un geliştirdiği nesne tespit yöntemi [60] çok sayıda görüntüde hızlıca ve yüksek başarı oranıyla aranan nesneleri bulan bir yöntemdir. Bu yöntem özetle art arda sıralı (ing. cascade) Adaboost sınıflandırıcıları [61] ile görüntüde, aranan nesnenin bir örneğini barındıran bölgeyi bulmak için çalışır. Yaptıkları çalışmada verilen görüntülerdeki insan yüzlerini yüksek başarı oranıyla tespit etmişlerdir.

Viola ve Jones'un sistemi [60] giriş olarak pozitif ve negatif örnek görüntüler istemektedir. Pozitif görüntülerden anlatılmak istenen aranılacak sadece ilgili nesneyi tam olarak gösteren örnek görüntüler olmakla beraber negatif görüntülerden anlatılmak istenilen ise ilgili nesne olmamasına rağmen onunla karıştırılabilecek görüntülerdir. Yaptıkları çalışmada pozitif görüntüleri günümüz biyometrik fotoğraflara benzeyen sadece yüz bölgesini kapsayan görüntülerden oluşturmuşlardır.

Yaklaşık olarak 5000 pozitif örnek görüntü, 10000 negatif örnek görüntü ile 130 test görüntüsü üzerinde yüzlerin konumlarını tespit etmeye çalışmışlardır. Ayrıca burada örnek görüntü kümeleri ufak kesitler olarak alınmıştır (örneğin 28x28 piksel).

Sistem eğitim görüntülerinde 3-boyutlu Haar benzeri öznitelikleri (Bölüm 4) elde ettikten sonra bu öznitelikleri Adaboost algoritmasına vermektedir. Adaboost algoritmasının temel amacı eğitim örneklerini kullanarak bir dağılım ortaya çıkarmak ve bu dağılıma bağlı düşük hatalı bir varsayım oluşturmaktır. Bu algoritmada özniteliklere bağlı olarak zayıf sınıflayıcılar oluşturulur. Bu zayıf sınıflayıcılardan oluşan küme içerisinden hata oranı yüksek olanlar elenir. Geriye kalan zayıf sınıflayıcılar kullanılarak daha güçlü bir sınıflayıcı oluşturulur. Daha dikkatli bir analiz yapıldığında seçici olmayan özniteliklere bağlı zayıf sınıflayıcıların hata oranı yüksek olduğundan elendiği gözlemlenmektedir.

Güçlü sınıflandırıcı oluşturma yapısı art arda sıralı bir şekilde devam etmiştir. En nihayetinde istenilen hata oranına yaklaşmış güçlü sınıflandırıcılardan oluşan ana sınıflandırma yapısı ortaya çıkar. Ayrıca, her bir zayıf sınıflandırıcı başlangıç olarak aynı dağılım ile yola çıkmaktadır.

Test görüntülerde pencereler içerisindeki alanlarda yüz tespiti için tekrar 3-boyutlu Haar benzeri özellikler çıkartılarak Adaboost sınıflandırıcısına verilir ve belirli bir eşik seviyesi (ing. threshold) üstünde olan pencerelerde yüz vardır şeklinde işlem tamamlanır.

Yapılan bu tez çalışmasında nesnelerin tespiti için Viola ve Jones'un sistemi [60] kullanılmıştır. Bu sistemin kullanılabilmesi için EOFTE veri kümesinde var olan nesnelerden bir eğitim kümesi oluşturulmuştur. Pozitif eğitim kümesinde 500 görüntü, negatif eğitim kümesinde ise 1000 görüntü olacak şekilde, EOFTE veri kümesinde var olan hareketli görüntüler dışında, tekrar görüntü alınmıştır. Bu noktada arka plan bilindiğinden basit bir arka plan çıkarımı yapılarak ön planda bulunan nesnelerin farklı büyüklük, oryantasyon ve örtüş durumlarına göre görüntülerinden pozitif eğitim kümesi oluşturulmuştur. Negatif eğitim kümesi ise %80'i arka plandan %20'si ise aktörden alınan çeşitli kesitlerden oluşan görüntülerden oluşturulmuştur. Çalışma yapısı olarak test kümesinde bulunan hareketli görüntü veya görüntülerin her bir çerçevesinde çalışabilecek bir düzeyde nesne tespit sistemi olarak hazırlanmıştır.

### 2.1.3.3. Gall ve Lempitsky'nin Nesne Tespit Yöntemi

Gall ve Lempitsky [62] görüntülerde ilgili nesne sınıflarından (örneğin arabalar, yayalar) örnekler tespit etmeyi amaçlayan bir yöntem oluşturmuşlardır. Bu yöntem, özetle, görüntüde bulunan nesnenin farklı kesitlerini bularak nesnenin merkezini ve dolayısıyla nesnenin konumunu tespit etmeyi hedeflemektedir. Örneğin araba nesnesi ele alındığında bu nesnenin farklı kesitleri olarak lastikleri, camları, bagajı, kaputu v.b. kısımları verilebilir. Her bir kesit nesnenin merkezi için olasılık tabanlı bir oy vermektedir. Bu oyların hepsi ele alındığında ise nesnenin merkezi ve dolayısıyla nesnenin konumu ve sınıfı bulunmaktadır.

Temel olarak Hough Dönüşümü (ing. Hough Transform) ile doğru, daire, elips ve benzeri şekiller görüntü üzerinde tespit edilebilmektedir [63]. Hough dönüşümü daha sonra Ballard [64] tarafından Genelleşmiş Hough Dönüşümü (ing. Generalised Hough Transform) haline geliştirilerek sadece analitik fonksiyonlar ile tanımlanan nesnelerin değil gelişigüzel nesnelerin de bir modele bağlı olarak tespit edilmesini sağlamıştır. Genelleşmiş hough dönüşümü ölçekten ve dönmeden bağımsız olarak çalışabilmektedir. Gall ve Lempitsky [62] genelleşmiş Hough dönüşümünden yola çıkarak nesneyi oluşturan bağımsız kesit görüntülerini kullanan nesne sınıflarına özel bir sınıflayıcı tanımlamıştır. Bu sınıflayıcı Rastgele Orman (ing. Random Forest) yapısını kullanarak nesne sınıfına özgü bir biçimde çalışmaktadır.

Yöntemde ilk önce, ön planda bulunduğu varsayılan nesneden alınan kesitler ile pozitif eğitim kümesi, arka planda bulunan ve nesne ile alakası bulunmayan kısımlarda alınan negatif eğitim kümesi oluşturulur. Bu kesitlerdeki görüntü özellikleri kullanılarak bir rastgele orman sınıflandırıcısı oluşturulur. Her bir rastgele ağaçtaki düğüm (ing. node) test görüntüsündeki seçili penceredeki alan ile eğitim kesitleri arasındaki benzerliği karşılaştırır. Bu karşılaştırmalar sonucunda pencere içerisindeki alan nesnenin bir parçası olarak çıkarsa nesnenin merkezi hakkında oy kullanılır. Bu işlem test görüntüsündeki tüm pencereler için devam eder. Nihayetinde eşik seviyesinden daha fazla oy alan bir nokta tespit edilir ise bu nokta nesnenin merkezi olarak tayin edilir.

Yapılan tez çalışmasında, Gall ve Lempitsky [62] yöntemi kullanılarak her nesne sınıfı için bir sınıflandırıcı oluşturulmuştur. Hareketli görüntünün seçilen çerçevelerinde tüm sınıflandırıcılar art arda çalıştırılarak hangi sınıfa ait nesnenin görüntüde nerede olduğu tespit edilmiştir. EOFTE veri kümesindeki her bir hareketli

görüntüde ön planda aktörün etkileşimde bulunduğu bir nesne olduğu kısıtlaması mevcuttur. Bu kısıtlama (ing. constraint) ile eğer ki birden fazla tip nesnenin merkezi bulunursa hangisinin daha yüksek oy alan tercih edilmiştir. Yüksek oy alan nesnenin sınıf bilgisi Rastgele Ormanlar yapısı kullanılarak aynen Gall ve Lempitsky [62]'in çalışmasında olduğu gibi elde edilmiştir. Nesnenin sınıf bilgisi tez kapsamında kullanılan sistemde nesne bilgisi olarak ele alınmıştır.

## 2.2. Eylem Tanıma Konusunda Yapılmış Çalışmalar

Yapılan bir çalışmada, yakın zaman içerisinde geliştirilmiş insan eylemlerinin tanınması sağlayan sistemler hakkında bilgi verilmiştir [1]. Bu çalışmada görüntü dizilerinden çıkarılan çeşitli öznitelikler incelenmiş ve bunlar iki ana kategoride toplanmıştır. Bu kategorilere küresel gösterimler ve yerel gösterimler olarak ad verilmiştir. Küresel gösterimi kullanan yaklaşımlarda, öncelikle hareketli görüntülerdeki kişi (aktör), ilgili görüntüler üzerinde arka plandan ayrıştırılması veya takip etme (ing. tracking) metotlarıyla tespit edildikten sonra ilgi alanının/noktalarının bulunması, ilk işlemler olarak sıralanmıştır. Bu işlemlere takiben de ilgi alanını temsil eden genel öznitelik tanımlayıcısı oluşturulduğu belirtilmiştir. Yerel gösterimi kullanan yaklaşımlarda ise birbirinde bağımsız uzam-zaman ilgi noktaları bulunur. Daha sonra bu ilgi noktaları arasındaki korelasyon kullanılarak sonuca gidildiği belirtilmiştir. Küresel gösterimlerin, yerel gösterimlere oranla gürültüye karşı daha hassas olduğu fakat yerel gösterimlerden çok daha hızlı çalıştığının altı çizilmiştir. Ayrıca bu çalışmada, çerçeve (hareketli görüntüdeki her bir görüntü) veya çerçeve dizileri için çeşitli öznitelikler elde edildikten sonraki aşama olan eylemlerin sınıflandırılması, üç ana kategoride toplanılmıştır: doğrudan sınıflandırma, zamansal durum-uzam modelleri ve eylem tespiti. Doğrudan sınıflandırma yaklaşımını benimseyen çalışmalarda öznitelikler zamana göre değişim modellerine bakılmaksızın sınıflandırılma işleminin yapıldığı, zamansal durum-uzam modelleri yaklaşımını benimseyen çalışmalarda ise zaman-uzam modellerinin oluşturulduğu, eylem tespit yaklaşımlarda ise hareketli görüntüdeki eylemin hiçbir eylem modellemesi yapılmadan tespit edildiği belirtilmiştir.

Kovashka ve Grauman [65], uzam-zamansal (ing. spatio-temporal) ilgi noktalarının genellikle çok yerel olduğu ve bunların tam olarak uzam ve zaman komşuluklarını tanımlayamadığı önerisinden yola çıkarak sınıfa özel değişken şekilli

uzam-zaman komşuluklarını içeren sözcük hazinesi oluşturarak hareketlerin tanınması konusunda çalışmışlardır. Öncelikle 3 boyutlu gradyanlar histogramı metoduyla ilgi noktaları tespit edilip bu noktalardaki yerel hareket ve görünüş öznitelikleri, gradyanlar histogramı ve Frekans Histogramı (ing. Histogram of Frequency) metotları aracılığıyla elde edilmiştir. Harflerden kelimeler elde edilmesi mantığıyla en yakın özellikleri barındıran ilgi noktaları kümelenmiştir. Çoklu çekirdek öğrenmesi (ing. multi-kernel learning) ile bu sözcüklere ağırlıklar atanarak hareketli görüntüler içerisinde ayırt edici sözcükler tespit edilmiştir. Son olarak aynı işlemlere tabii tutulan tüm test örnekleri tek tek, testteki örneğin dışındakileri eğitim kümesi kabul ederek, Destek Vektör Makineleriyle sınıflandırılarak KTH veri kümesinde %94,53, UCF Spor veri kümesinde %87,27 tanıma başarısıyla eylemler sınıflandırılmıştır.

Bregonzio arkadaşları [66], geliştirdikleri sistem ile hareketli görüntülerde zaman bilgisinin önemini vurgulayarak eylem tanıma yapmışlardır. Tek görüntü üzerinde tespit edilen ilgi noktalarının çok yerel kaldığı ve eylemi icra eden aktör üzerinde yoğunlaşmadığı vurguladıkları çalışmalarında ardışık çerçeveler arasındaki farklar hesaplanıldıktan sonra bu fark görüntülerinde 2 boyutlu Gabor filtreleri kullanılarak ilgi noktalarının kümesi oluşturulmuştur. Bu kümede hangi ilgi noktalarının daha ayırt edici olduğunu bulmak için yerel tepe noktaları hesaplanılmıştır. Bu noktaları kapsayan pencere ile ön plandaki hareketli nesne tespit edilmiş ve bu nesne (aktör) ile ilgili öznitelikler çıkartılmıştır. Öznitelik olarak nesne penceresinin en-boy oranı, yatay ve dikeydeki hızı gibi temel özellikler kullanılmıştır. Farklı zaman ve uzam ölçeklerinde elde edilen bu hareket tabanlı özniteliklere ek olarak görünüş tabanlı özniteliklerde sistem içerisinde harmanlanmıştır. Çoklu çekirdek öğrenmesi algoritması ile eğitim kümesi kullanılarak eğitim aşaması tamamlanmış ve test örnekleri üzerinde sistemin başarısı ölçülmüştür. Weizmann veri kümesinde %96,66, KTH veri kümesinde ise %94,33 doğruluk oranı ile eylem tanıma yapılmıştır. Test aşamasında, test hareketli görüntüsündeki aktörün ilgili hareketli görüntüleri eğitim kümesine dahil edilmeden eğitim yapılmıştır.

Klaser arkadaşları [22], yaptıkları çalışmada uzam-zaman eksenindeki değişimleri yakalamaya çalışan bir tanımlayıcı geliştirmişlerdir. 3 boyutlu değişimlerin oryantasyonlarını inceleyen araştırmacılar geliştirdikleri gradyanlar histogramı benzeri üç boyutlu tanımlayıcı vektörler aracılığıyla insan eylemlerinin

tanınması üzerine yoğunlaşmışlardır. Weizmann veri kümesi üzerinde test ettikleri sistemleri %84,3 oranında başarı elde etmiştir.

Zhang ve Tao [67], Yavaş Özellik Analizi (YÖA) (ing. Slow Feature Analysis) yaklaşımını temel alan eylem tanıyan bir sistem geliştirmişlerdir. YÖA hızlıca değişen bir giriş sinyali üzerinde yavaşça değişen özellikleri elde etmek için kullanılan bir metottur [20]. Hareketli görüntülerde yer değiştirmenin olduğu piksellerin sınırlarından çıkarılan çok sayıda rastgele seçilen küboidlerde (uzam-zaman küpleri) YÖA özellikleri çıkartılır. Bu YÖA özellikleri kabaca komşu çerçeveler arasındaki hareketi tanımlayan öznitelikler olarak açıklanmaktadır [68]. YÖA özellikleri ayrıca küboidlerin ızgarasal bir yapı ile aktörün vücut bölgelerine göre farklılaştırılması ile zenginleştirildiği bu çalışmada özellikler çeşitli makine öğrenmesi metotları ile (gözetimli, gözetimsiz, ayırt edici) de farklılaştırılmıştır. Her bir hareketli görüntü için elde edilen bu öznitelikler daha sonra DVM algoritması ile eğitim kümesi kullanılarak test örneklerinden eylemin sınıflandırılması için kullanılmıştır. Araştırmacılar küboidleri kendi içerisinde parçalayarak aktörün görünüş özelliğinden de faydalanmışlardır.

Ikizler ve Sclaroff, insan eylemlerinin tanınmasının sadece vücut hareketlerinden elde edilecek özniteliklerin kullanılmasıyla yeteri kadar başarılı olmayacağı varsayımından hareketle eylem sırasında kullanılan nesnelerin ve eylemin icra edildiği sahnenin önemli bir bilgi kaynağı olarak kullanılmasının gerekliliğini vurgulamışlardır [69]. Bu eksiklikleri giderecek ve de hareketli görüntülerde insan eylemlerini tanıyacak bir sistem oluşturmuşlardır. Kullanılan hareketli görüntüler öncelikle gürültülerden arındırılıp ön planda hareket eden insan ve/veya cisimler kutulanarak tespit edilmiştir. Daha sonra bu kutular içerisinde kalan alanda gradyanlar histogramı ve Optik Akış (ing. Optical Flow) metotları aracılığıyla çeşitli öznitelikler elde edilmiştir. Hareketli görüntülerdeki sahneden ise seçilen üç rastgele çerçeveden renk bilgisi temelli bir öznitelik vektörü ve de seçilen beş rastgele çerçeveden öznitelik vektörü elde edilmiştir. Çıkarılan tüm öznitelikleri giriş bilgisi kabul eden Çoklu Örnek Öğrenmesi (ing. Multiple Instance Learning) sınıflandırma algoritmasıyla sistem sonuçlandırılmıştır. Sahne ve nesneler hakkında öznitelikleri kullanan bu sistem 1168 tane Youtube hareketli görüntüsünden oluşturulan bir veri kümesi üzerinde test edilmiştir. Bu test sonucunda tasarlanılan sistem hareketli görüntülerin %75,21'indeki eylemleri başarıyla tanımıştır.

Ballan arkadaşları [70], hem görünüş hemde hareket tabanlı tanımlayıcıları birleştirerek geliştirdikleri sistem aracılığıyla hareketli görüntülerde eylem tanıma yapmışlardır. Sistem çalışmasına, öncelikle ilgi noktaların tespiti için farklı uzam-zaman ölçeklerinde çalışabilir şekle getirilmiş [71] dedektörünün zenginleştirerek başlar. Dedektör aracılığıyla tespit edilen ilgi noktalarını, merkez alan küboitlerden 3 boyutlu gradyan özellikleri ve optik akış özellikleri elde edilerek, yeni bir öznitelik tanımlayıcısı oluşturulur. Bu öznitelik tanımlayıcıları kümeleme algoritmaları kullanılarak kümelenir ve küme merkezlerini birer görsel kelime olarak tanımlanarak verilen hareketli görüntü bazında kelimelerin histogramları oluşturulur. Bu histogramların her eylem türüne göre farklı olacağı varsayımından yola çıkılarak elde edilen histogramlar ile DVM eğitim kümesi örnekleri ile eğitilir ve test örnekleri kullanılarak test edilir. Benzeri çalışmalarda olduğu gibi bu çalışmada da test örneğindeki aktörün hareketli görüntüleri eğitim kümesinden hariç tutularak yapılan testte elde edilen sonuçlar KTH veri kümesi için %92,10, Weizmann veri kümesi için %92,41 ortalama doğruluk oranını elde etmiştir.

Dhillon arkadaşları [72] geliştirdikleri görünüş ve hareket bilgilerini birleşik bir biçimde yorumlayan sistem ile hareketli görüntülerde eylem tanıma yapmışlardır. Sistem, hareketli görüntüde eylemi icra eden aktörün çeşitli takip yöntemleri [73], [74] ile izlenmesi sonucunda elde edilen görünüş ve hareket özellikleri değerlendirmiştir. Aktörün takip edilmesi ile elde edilen vücut bölgelerindeki değişimden hareket bilgisi, bu bölgelerin ortak görümününden ise istatiktiksel yöntemler kullanılarak görünüş bilgisi çıkartılmıştır. Kelimeler Torbası yaklaşımı ile hareket ve görünüş bilgileri her bir hareketli görüntü için ayrı ayrı histogramlardan birleştirilerek hareketli görüntüler için tanımlayıcılar elde edilmiştir. Bu histogramlar ise en sonunda DVM sınıflandırıcı kullanılarak eylem tanıma işleminde kullanılmıştır. Diğer çalışmalardan farklı olarak takip etme yöntemlerinin başarısına bağımlı bir sistemdir.

Thurau [75], sabit görüntü ve hareketli görüntülerde eylem tanıyan bir sistem geliştirmiştir. Sistem öncelikle görüntü üzerinde bir arama penceresi döndürerek aktörü tespit etmekte ve bu arama penceresi içerisindeki gradyanları hesaplamaktadır. Bu arama penceresi daha sonra ızgarasal bir yapı biçiminde bölümlenerek kabaca vücut bölgeleri arasındaki ilişkiyide kullanan bir yapı haline gelmektedir. Bu ızgarasal yapıda bulunan her hücre (ing. cell) gradyanlar histogramı özniteliği çıkartılmış ve bunlar belli bir sıra ile birleştirilerek görüntüyü tanımlayan

genel tanımlayıcı elde edilmiştir. Araştırmacı ayrıca karmaşık hareketlerin temel hareketlerin birleşimden meydana geldiği varsayımından yola çıkarak temel tanımlayıcıları kendi içerisinde kümelemiştir. Bu kümeleme sonucunda gradyanlar histogramı öznitelikleri yakın kümelere atanmıştır. Bu atama sonucunda elde edilen etiketlerin zaman içerisindeki sıralamasına göre eylemler arasında benzerlik kurulmuştur. Bu benzerlik n-gram [76] yapısında eylem sınıflandırma için kullanılmıştır.

Microsoft XBOX Kinect platformu gibi algılayıcılar aracılığıyla renk bilgisine ek olarak derinlik bilgisi de sunan donanımların yaygınlaşmasıyla bu benzeri donanımları kullanarak hareket/eylem analizi yapan sistemler geliştirilmiştir. Shotton arkadaşları [77], insanların duruş bilgisini çıkaran çalışmaları MS Kinect algılayıcısını kullanan örneklerden biridir. Bu çalışmada zamansal bilgi kullanmadan, sadece derinlik bilgisi kullanarak, insan eylemlerini bölgesel bazda bulan ve bunun için duruş özniteliklerini çıkarmayı hedefleyen bir sistem geliştirilmiştir. Bu sistem iskelet çıkarmak yerine derinlik bilgisiyle elde edilen insan şeklinin, çeşitli bölgelere bölümlenmesinin eylem tanıma çalışmalarında daha verimli kullanılabileceğini önermiştir.

Sung arkadaşları [41], MS Kinect ortamının günümüzdeki düşük maliyetiyle bir çok işlemde sistemde rahatlıkla kullanılabileceğini bildirmekle beraber Kırmızı-Yeşil-Mavi-Derinlik (KYMD) (ing. RGBD) görüntülerinden insan eylemlerini tespit eden bir sistem geliştirmişlerdir. Sistem hareketli görüntüdeki aktörün iskelet bilgisini oluşturduktan sonra eklem yerlerinin pozisyon bilgisini, ellerin konum bilgisini ve de hareket bilgisini tek bir çatı altında birleştirmek için bir Markov model oluşturmuştur. Bu model aracılığıyla araştırmacılar, insanların evde sıklıkla yaptıkları eylemleri, hareketli görüntüler üzerinden tanımaya çalışmışlardır.

## 2.3. Fizik Terapi Egzersizlerini Konu Alan Yapılmış Çalışmalar

Fizik terapi uzmanları hastalarını eski sağlıklarına kavuşturabilmek için terapi seansları uygularlar. Bu terapi seanslarında işlevini yitirmiş veya işlevleri azalmış bölgelerin muayenesi yapıldıktan sonra tedavi sürecine geçilmektedir. Bu tedavi sürecinde fizik terapi uzmanının yönlendirmesiyle hasta genellikle farklı egzersiz tiplerinden oluşan bir küme egzersizi tekrar etmektedir. Egzersizler tekrar edilmeden

önce fizik terapi uzmanı egzersizin nasıl yapılması gerektiği sözel veya fiziksel olarak tasvir etmektedir. Özellikle felç geçirmiş hastaların eski sağlıklı hallerine dönebilmeleri için hasta tarafından tekrar edilen egzersizlerden analiz edilmesi önemlidir. Bu analiz esnasında sözel, görsel, hareketli görüntü, kinematik ve benzeri geri dönüşlerin gözlemlenmesini sağlayacak sistemlerin kullanılması yarar sağlamıştır [78]. İnsan sinir sisteminin çalışmasını esas alan yapay sinir ağları, hataların tespit edilmesi ve düzeltilmesi için geri dönüş değerleri barındırmaktadır. Bu geri dönüş değerleriyle, tasarlanılan sistemlerin hata oranları azaltılmaya çalışılmaktadır. Yapılan bir analiz sonucunda, egzersiz seanslarının rehabilitasyon sürecinin başarıya ulaşmasındaki ana etkenlerden biri olduğu ve bu seansların motor hareket öğrenme sürecini hızlandırdığı tespit edilmiştir [79].

Fizik terapi uzmanlarının hastaya fiziksel destek sağlayarak yaptırdığı egzersizler, hasta tarafından daha başarılı bir şekilde tamamlanabilmektedir. Örneğin yürüme platformunda yapılan bir deneyde insanların fizik terapi uzmanının desteğiyle daha az hata yaparak yürüdüğü tespit edilmiştir. Fakat bu deneyde destek almayan insanların kendi hatalarını görerek sonraki tekrarlarda daha az hata yaptığı da tespit edilmiştir. Yürüme egzersizinde yapılan hatanın tespit edilebilmesi için hastanın kendi yaptığı egzersizi izleyebilmesi ve değerlendirmesinin önemi ortaya çıkmıştır. Bu değerlendirmenin fizik terapi uzmanı tarafından yapılması ise hastayı izleme ve terapiyi güncelleme açısından ön plana çıkmıştır [80].

Diz rehabilitasyonu, diz incinmesi veya diz ameliyatından sonra tekrar eski fonksiyonlarını kazanabilmesi için çok önemli bir işlemdir. Bu tedavinin başarılı olabilmesi için fizik terapi uzmanının hastaya verdiği ödev egzersizlerin düzenli olarak tekrar edilmesi gerekmektedir. Egzersiz tekrarlarının düzenli olarak yapıldığının kontrolünü sağlayacak sistemlerin, hastanın iyileşme sürecine katkı sağlayacağının önemle altı çizilmiştir [81].

Nater arkadaşları [82], yaşlı ve/veya engelli kişilerin evlerindeki yaşamlarını desteklemek amacıyla kameralar aracılığı ile normal ve anormal eylemlerin tespiti üzerine çalışmıştır. Çalışmada iki hiyerarşik yapı kullanılmıştır. İlk hiyerarşik yapı insan görünüşü üzerine odaklanıp yukarıdan-aşağı bir yaklaşımla (ing. top-down) hareketli görüntülerden silüet görüntüleri elde edip bunları kümeleyerek sembollerle eşleştirmiştir. İkinci hiyerarşik yapı ise aşağıdan-yukarı bir yaklaşımla (ing. bottom-up) hareket ve duruş tabanlı sembollerin birbirlerini takip etmesini kurallar dahilinde sınırlandırarak eylem/davranış örüntülerini elde etmiştir. Bu çalışma dahilinde

normal davranış/eylemler kümesine uzakta kalan örüntüler her iki hiyerarşi içerisinde tespit edilip anormal davranış olarak etiketlenmiştir. Konu üzerinde hazır bir veri kümesi bulunmadığından araştırmacılar yeni bir veri kümesi oluşturmuşlardır. Bu veri kümesi birbirlerinden farklı hareketleri ardı ardına icra eden aktörlerin hareketli görüntülerini içermektedir. Önerdikleri sistem bu veri kümesi üzerinde test edilmiş ve normalden uzak davranışların tespit edilmesinde yüksek başarı oranı elde edilmiştir. Ayrıca çalışmada anormal davranışa en güzel örnek olarak aktörün yürürken bir anda düşüp kalması örneği verilmiştir.

Zhou ve Hu yaptıkları çalışmada 1980-2007 yılları arasındaki rehabitilasyon için insan hareketlerinin takibi üzerine yapılan çalışmaları incelemiştir. Bu araştırmada temel olarak insan vücudunun ve eklem yerlerinin takibi üzerine yapılan çalışmaların ilerleyen teknoloji ile birlikte daha başarılı sonuçlar elde ettiğini vurgulanmıştır. Ayrıca fizik tedavi alanında artan hasta sayısının ve bu hastaların devamlı takibinin getirdiği zorlukların da altı çizilmiştir [9].

Soutschek arkadaşları [83], geliştirdikleri sistem ile fizik tedavi egzersizlerini yönetmeyi ve desteklemeyi amaçlamışlardır. Kullandıkları Uçuş Zamanı (ing. Time of Flight) kamera ile hem renk hemde derinlik hareketli görüntülerini elde ederek kalça, diz ve ayak bileğinde bulunan beyaz etiketleri takip etmişlerdir. Bu takip sonucunda elde edilen eklemler arasındaki açıyı zaman içerisinde yorumlayarak egzersizi takip etmişlerdir.

Fizik tedavi hareketlerinin gözlemlenmesi ve incelenmesi konusunda geliştirilen sistemler genellikle hareket takibi üzerine yoğunlaşmaktadır [9]. Literatürde var olan sistemler incelendiği genel olarak egzersiz hakkında çok detaylı bilgiler elde edilmesi üzerine inşa edilmiş ve kliniklere özel tasarlanmış olarak karşımızı çıkmaktadır. Özel algılayıcılar kullanarak diz ve eklemler arasındaki açıların zamansal değişimini izlemek [83], üzerinde özel algılayıcılar barındıran bir elbise ile eklemlerin ve egzersizin detaylı incelemesini yapmak [84], üst vücut fonksiyonlarını gözlemlemek üzere tasarlanmış algılayıcılar kümesi tasarlamak [85] şeklinde ön plana çıkan çalışmalardaki sistemlerin donanım tabanlı maliyetlerinin yüksek olduğu gözden kaçmamalıdır.

# 3. VERİ KÜMELERİ

Tez kapsamında geliştirilen sistemleri test etmek için çeşitli veri kümeleri kullanılmıştır. Bu veri kümelerinden; Gall veri kümesi [86], KTH veri kümesi [87], ve Weizmann veri kümesi [6] yapılan çalışmalarda yöntemlerin sonuçlarının karşılaştırılmasında sıkça kullanılan açık veri kümeleridir. Açık veri kümesi internet aracılığıyla ücretsiz olarak indirilip araştırma amaçlı kullanılabilen veri kümeleri olarak tanımlanmıştır.

Tez kapsamında geliştirilen fizik tedavi egzersizlerini inceleyen ve egzersiz tipi ve ilgili egzersizin tekrar sayısını bulan sistemlerin tam olarak test edilebilmesi için ev ortamında çekilmiş fizik tedavi egzersiz seanslarının hareketli görüntülerinden oluşan bir veri kümesi gereklidir. Bu istek doğrultusunda yapılan araştırma neticesinde bu tip eylemleri barındıran açık bir veri kümesi bulunamamıştır. Bundan dolayı tez çalışması kapsamında evde yapılan fizik tedavi egzersizlerini barındıran bir veri kümesi oluşturulmuş ve araştırmacıların erişimi için bu veri kümesi açık hale getirilmiştir. Oluşturulan veri kümesi ile ilgili detaylar Bölüm 3.4'de verilmiştir.

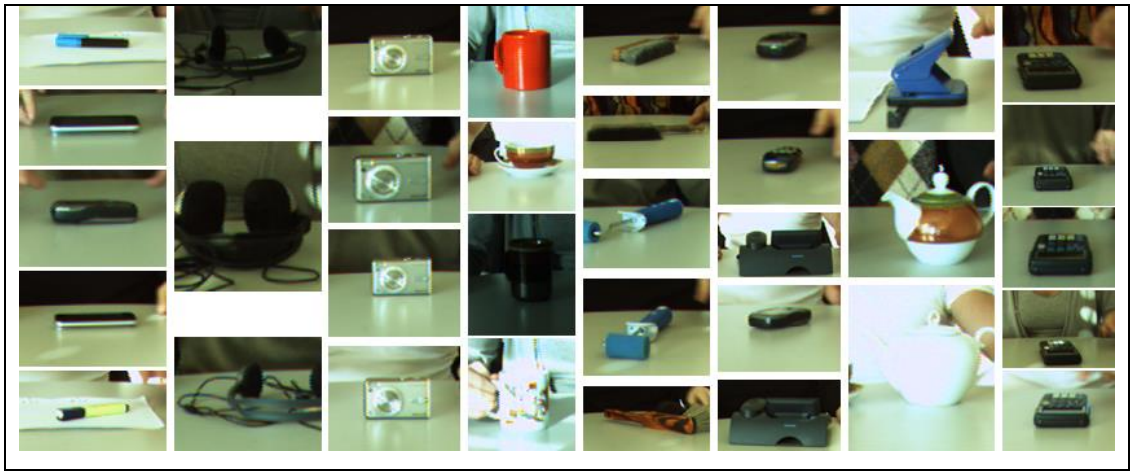
## 3.1. Gall Veri Kümesi

Gall arkadaşları [86] 2011 yılında, hareketli görüntülerde şahısların kullandığı nesnelerin fonksiyonelliklerine kategori edilmesi konusunda bir çalışma yapmışlardır. Bu çalışmada oluşturulan sistemin başarısını ölçmek için bir veri kümesi oluşturmuşlardır. Bu veri kümesinde altı farklı kişiden masanın üstünde bulunan nesneleri kullanarak ardı ardına otuz eylem yapmaları istenmiştir. Bu farklı eylemler: kupa içerisine sıvı dökme, sol elini kullanarak içme, sağ elini kullanarak içme, fırça kullanma, uzaktan kumanda kullanma, ufak boya rulosu kullanma, delgeç kullanma, cep telefonunu kullanarak bir arama yapma, müzik dinlemek için kulaklık takma, bir resim çekme, kalem kullanma, oyun konsoluyla bilgisayar oyunu oynama, şeklinde sıralanmıştır. Şekil 3.1a'da bu veri kümesinden (Gall veri kümesi) alınan uzaktan kumanda kullanma eyleminden alınan bir çerçeve, Şekil 3.1b'de ise gene aynı veri kümesinden alınan ufak boya rulosu kullanma eyleminden alınan bir çerçeve örnek olarak verilmiştir.

a) b)

Şekil 3.1: Gall veri kümesinden alınan örnek çerçeveler. a) uzaktan kumanda kullanma, b) ufak boya rulosu kullanma.

Gall veri kümesinde bulunan nesneler kupa, fincan, fırça, uzaktan kumanda, rulo, delgeç, cep telefonu, kulaklık, fotoğraf makinesi, fosforlu kalem, şeklinde sıralanmıştır. Bu belirtilen nesneler kategorileri içerisinde çeşitli sayılarda nesneler bulunmaktadır. Veri kümesindeki nesnelerin görüntülerden bir kesit Şekil 3.2'de [86] gösterilmiştir.



Şekil 3.2: Gall veri kümesinde bulunan nesnelerin örnek görüntüleri.

Gall veri kümesinde hareketli görüntülerin her çerçevesi için derinlik bilgisi de (görüntüsü) bulunmaktadır. Bu bilgi tez çalışması kapsamında önerilen sistemlerin test edilmesi aşamasında çok yararlı olmuştur.

## 3.2. KTH (Kraliyet Teknoloji Enstitüsü) Veri Kümesi

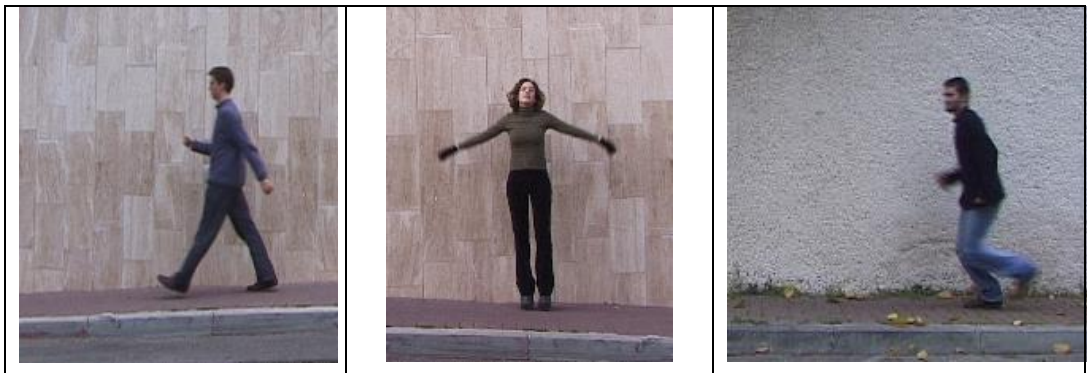
Schült arkadaşları [87] 2004 yılında hareket tanıma üzerine yapmış oldukları çalışmada yerel uzam-zamanda ilgi çeken noktaları tespit etmişlerdir. Bu noktalarda ki öznitelikleri çıkararak yerel hareket hakkında bilgi edinmeyi hedeflemişlerdir. Örneğin ayaklar, eller, kollar ve benzeri uzuvlara denk gelen noktaların hareket özniteliklerini hareketli görüntünün tümünde analiz ederek hareketin dokusal yapısına ulaşmaya çalışmışlardır. Bu çalışmada her çerçevede tespit edilen ilgi çeken noktaların birbirleriyle eşleştirilmesinden sonra literatürde var olan öznitelik çıkarma yöntemleriyle bu noktaları tasvir eden öznitelik vektörleri oluşturulmuştur. Bu işlem sonunda ise hareket tanıma aşamasına gelinmiş ve destek vektör makinesine çıkarılan öznitelik vektörleri verilerek sınıflandırma yapılmıştır. Ayrıca geliştirdikleri yöntemi değerlendirmek amacıyla hareketli görüntülerden oluşan ve KTH veri kümesi olarak adlandırılan bir veri kümesi hazırlamışlardır [87].

İnsan eylemlerini tanıyan sistemlerin başarılarını karşılaştırmakta sıkça kullanılan KTH veri kümesi temel olarak 599 tane düşük çözünürlüklü hareketli görüntüden oluşmaktadır. Yürüme, koşma, rahat adımlarla yürüyüş ve koşma (ing. jogging), boks, alkış ve el sallama olmak üzere altı farklı hareketi yapan 25 farklı şahsın hareketli görüntülerden oluşan bu kümedeki hareketli görüntüler 160 x 120 piksel çözünürlüktedir. Temel olarak kullanılan küme, KTH veri kümesindeki dış ortam sahnesinde çekilmiş olan insan eylemlerinin hareketli görüntüleridir. Daha sonra eklenen iç ortamda, farklı yakınlaştırma seviyeleriyle dış ortamda, aktörlerin çeşitli kıyafetler giymesi ile dış ortamda çekilen hareketli görüntüler ile veri kümesindeki hareketli görüntü sayısı 2391'e ulaşmıştır. Bu haliyle genişletilmiş ve/veya geniş KTH veri kümesi olarak tanımlanmaktadır. Genişletilmiş KTH veri kümesi temel kümeye oranla daha karmaşık durumlar içermektedir. Bu durumlar sahne içerisinde bulunan elemanların birbirlerini örtmesi (ing. occlusion), daha fazla gürültü, ışık farklılığı, farklı yakınlaştırma seviyelerinin birbiri ardına aynı hareketli görüntüde uygulanması, şeklinde sıralanabilir.

## 3.3. Weizmann Veri Kümesi

Blank arkadaşları [6] tarafından 2005 yılında yapılmış bir araştırmada hareketli görüntülerde insan gövdesinin ve uzuvlarının, silüet görüntülerden oluşan hareketli görüntülerde, 3 boyutlu olarak incelenmesiyle hareketin tanınması üzerine yoğunlaşmışlardır. Bunun için yeni yerel ve genel öznitelikler tanımlamışlardır. Oluşturdukları sistemin başarısını değerlendirmek amacıyla hareketli görüntülerden oluşan ve Weizmann veri kümesi olarak adlandırılan bir veri kümesi hazırlamışlardır.

Hareketli görüntülerde insan eylemlerinin tanınması alanında, önerilen farklı sistemlerin karşılaştırılmasında sıkça kullanılan bir veri kümesi olarak karşımıza çıkan Weizmann veri kümesinde 90 tane 180x144 çözünürlükte hareketli görüntü bulunmaktadır. Bu hareketli görüntülerde dokuz farklı şahıs, on farklı eylemi gerçekleştirmektedir. 10 farklı eylem: koşma, yürüme, tek ayak üzerinde sekerek ilerleme, eğilme, tek elini sallama, iki elini sallama, bir uçtan bir uca yengeç yürüyüşü, zıplayarak ilerlemek, yerinde sabit kalarak zıplamak, kolları ve bacaklarını açıp kapayarak yerinde zıplama şeklinde sıralanmıştır. Şekil 3.3'de bu veri kümesinden alınan birkaç örnek görüntü verilmiştir [6].



a) b) c)

Şekil 3.3: Weizmann veri kümesinden alınan örnek çerçeveler. a) yürüme, b) zıplama, c) koşma eylemleri.

## 3.4. Ev Ortamındaki Fizik Tedavi Egzersizleri (EOFTE) Veri Kümesi

Tez çalışması kapsamında evde yapılan fizik tedavi seanslarının hareketli görüntülerinin incelenmesi ve oluşturulan sistemin bu hareketli görüntüler ile test edilmesi gerekmiştir. Yapılan araştırma sonucunda böyle bir açık veri kümesi bulunamamıştır. Bu gereklilik sonucunda Microsoft Kinect algılayıcısı kullanılarak ev tabanlı fiziksel terapi egzersizlerinin hem KYM (Kırmızı Yeşil Mavi) hem de derinlik hareketli görüntülerinden oluşan EOFTE (Ev Ortamındaki Fizik Tedavi Egzersizleri) veri kümesi oluşturulmuştur.

EOFTE veri kümesi oluşturulmadan önce Marmara Üniversitesi tıp fakültesinde bulunan Fizyoterapi ve Rehabilitasyon bölümüne üç defa gidilmiştir. Bu ziyaretler esnasında bölüm hocalarıyla birebir diyalog kurularak en sık karşılaşılan fizik tedavi hasta tipleri hakkında bilgi edinilmiştir. Hastaların en çok omuz, diz ve bel bölgelerindeki şikayetler ile ilgili olarak fizik tedavi bölümüne başvurulduğu bilgisine ulaşılmıştır. Bu bilgi çeşitli araştırmacılar tarafından da gözlemlenmiştir [8]. Ayrıca bölüm, hastaların faydalanması için evde yapılabilecek omuz, dil ve bel egzersizlerini gösteren bir resimli kılavuz hazırlamıştır.

Hastaların hastane ortamından ev ortamına geçtiklerinde, ilgili fizik tedavi uzmanı tarafından evde yapılmak üzere ödev egzersizler verilmektedir. Bu ödev egzersizlerin düzenli olarak tekrar edilmesi gerekmektedir. İyileşme süreci açından çok önemli olan bu durum ne yazıksa ki hastalar tarafından yeterince önemsenmediği doktorlar ile yapılan görüşmelerde açığa çıkmıştır. Hastane içerisinde yapılan egzersizlerin uzmanlar tarafından takip edilmesi sonucunda iyileşme kaydeden hasta evde bu ev egzersizlerine devam etmemekte veya bu evde yapılan egzersizler hakkında ilgili doktora (çoğunlukla) geri dönüş sağlamamaktadır.

Marmara Üniversitesi Tıp fakültesi, fizik tedavi ve rehabilitasyon bölümünde görevli doktorlar tarafından hastalarına uygulamaları için verdikleri ev egzersizlerinden yola çıkılarak hazırlanılan veri kümesinde kullanılan üç ana nesne tipi: sandalye, sopa, ve havludur. Esneme, destek, kuvvet uygulama egzersizlerinde sıkça kullanılan bu üç nesne bulunma kolaylığından dolayı seçilmiştir. Ayrıca, her evde aynı nesne kullanılmadığından ana nesneler mekanlara göre farklılık göstermiştir. Böylece ana nesne sınıflarının kendi içerisinde varyasyonları sağlanmıştır.

Evde yapılan fizik tedavi egzersizlerini yansıtmak amacıyla oluşturulan EOFTE veri kümesinde beş gönüllü aktör görev almıştır. Bu aktörler sekiz farklı egzersizi uygulamışlardır. Her bir egzersiz seansı 30 saniye ve 60 saniye arasında değişmektedir. Bu süre zarfında aktör aynı tip egzersizi istediği hızda istediği tekrar sayısı kadar egzersizin orijinal yapısına uymak kaydıyla tekrarlamaktadır. Her bir seans için altı farklı çekim yapılmıştır. Toplamda 240 tane hareketli görüntü bulunmaktadır. Bu 240 seans için hem renk hareketli görüntüleri hem de derinlik hareketli görüntüleri elde edilerek 480 tane hareketli görüntüden oluşan bir veri kümesi inşa edilmiştir.

EOFTE veri kümesinin oluşturulması esnasında aktörden bir seans içerisinde aynı tip egzersizi en az iki defa tekrar etmeleri istenmiştir. Her bir tekrarda egzersizin tam olarak yapılması ön şart konulmuştur. Egzersizde kullanılan nesne sahnede hazır olarak bulunmasına veya nesnenin aktöründe elinde bulunmasına dikkat edilmiştir. Egzersizle ilgili nesne dışında diğer egzersizlerle ilgili nesneler sahnede tam olarak yer almamıştır.

Microsoft Kinect algılayıcısı kullanılarak elde edilen renk tabanlı hareketli görüntüler (KYM) 640x480 piksel çözünürlüğünde çerçevelerden oluşmaktadır. Hareketli görüntüler saniyede 30 çerçeve ile (ing. fps) yakalanmıştır. Derinlik tabanlı hareketli görüntüler ise 320x240 piksel çözünürlüğünde olup saniyede 30 çerçeve ile yakalanmıştır. Hareketli görüntüleri dijital ortamda kaydetmek için “The Kinect for Windows SDK version 1.5” platformu kullanılmıştır [88].

Microsoft Kinect algılayıcısının her bir piksel için 11 bitlik derinlik bilgisi sağlamaktadır. Fakat bu derinlik bilgisi, zeminden veya duvardan yansıma durumlarında, gölge oluşması durumunda ve benzeri durumlarda net ölçülememektedir. Bu sorunlar sıkça karşılaşılan bir durum olduğundan araştırmacılar bu konu üzerinde çeşitli çözümler aramışlardır. Bu tez çalışmasındaki EOFTE veri kümesi oluşturulduğunda aynı problemlerle karşılaşılmıştır. Derinlik bilgisinin ölçülemediği durumlar için Xia arkadaşları [89] izlediği önişlemler serisi takip edilmiştir. Yöntem en yakın komşuluk ara değerlemesi (ing. nearest neighbor interpolation) ise ölçülemeyen derinlik değerlerine atama yaptıktan sonra tüm derinlik görüntüsünde medyan filtreleme kullanarak resmi daha pürüzsüz, daha az gürültü içeren bir görüntüye çevirmektedir. Bu yöntem için tez kapsamında medyan filtrenin boyutu 4x4 olarak kullanılmıştır. Bu işlem tüm derinlik görüntüleri için EOFTE veri kümesi oluşturulurken tekrarlanmıştır.

EOFTE veri kümesinde bulunan sekiz farklı egzersiz tipinin örnek görüntüleri ve açıklamaları aşağıda maddeler halinde verilmiştir. Egzersiz tipleri ile nesne ve duruş özellikleri arasındaki ilişki ise Tablo 3.1'de verilmiştir. Tüm egzersiz tiplerinin detayları ise maddeler halinde aşağıdaki gibi açıklanmıştır:

- Sopa egzersizi: Bu egzersizde aktör ayakta durmakta ve elinde bir sopa tutmaktadır. Dirsekleri sabit ve yukarı bakan bir konumda (dik olacak) başlanılan egzersiz, sopanın yavaşça baş üzerine doğru yukarı kaldırılıp tekrar aşağı indirilmesi ile devam etmektedir. Sopa ile germe olarak tanımlanabilecek bu egzersiz bir omuz egzersizidir. Egzersiz tipi ile ilgili örnek görüntü Şekil 3.4'te verilmiştir.

Şekil 3.4: Sopa egzersizinden örnek bir görüntü.

- Çapraz sopa egzersizi: Bu egzersizde bir önceki egzersiz tipi gibi bir omuz germe egzersizidir. Aktör bel hizasında yatay bir biçimde sağ ve sol uçlarına yakın noktalardan tuttuğu sopayı Şekil 3.5'te görüntü gibi çapraz pozisyona getirir ve kısa bir süre (birkaç saniye) bu pozisyondan sopayı tutar. Aktör daha sonra sopayı tekrar başlangıç noktasına geri getirerek egzersizi tekrar eder. Bu egzersizde dikkat edilmesi gereken husus gövdenin olabildiğince az hareket edip esas hareketin omuzlardan kaynaklanmasını sağlamaktır. Egzersiz tipi ile ilgili örnek görüntü Şekil 3.5'te verilmiştir.

Şekil 3.5: Çapraz sopa egzersizinden örnek bir görüntü.

- Sırtüstü bacak egzersizi: Bu egzersiz bir diz ve bel egzersizidir. Sırtüstü pozisyonda yatar durumda olan aktör dizini sabit tutar ve egzersize başlar. Sabit tutulan diz ile bacağını kaldırabildiği kadar yukarıya kaldırılması istenen aktör daha sonra bacağını tekrar başlangıç konumuna geri getirir. Diğer bacağı için aynı işlemi tekrar eder. Bu egzersizde dikkat edilmesi gereken husus bele yük bindirilmemesi ve hastanın bu egzersizi bir germe-açma egzersizi olarak görmesinin sağlanmasıdır. Egzersiz tipi ile ilgili örnek görüntü Şekil 3.6'da verilmiştir.

Şekil 3.6: Sırtüstü bacak egzersizinden örnek bir görüntü.

- Havlu egzersizi: Havlu ile germe bir omuz egzersizidir. Yeterince uzun bir havlunun bir el omuz üstünde diğer el bel arkasında olacak şekilde havlunun iki ucunda da tutulması ile egzersiz başlar. Bel arkasındaki el ile havlunun çekilip üst kolun ve omuzun gerilmesi ile tekrarlanır. Bu egzersizde aktör kameraya karşı sırtı dönük olarak havlunun görülmesini sağlamaktadır. Aktör isterse tekrar edilen seanslar içerisinde havluyu sağ veya sol eliyle tutabilir yani çeken ve gerilen

kollarını değiştirebilir. Bu egzersizde dikkat edilmesi gereken husus aktörün bu egzersizi tekrarlamaya başladığında zamanla yorulması ve yaptığı germenin azalmasıdır. Egzersiz tipi ile ilgili örnek görüntü Şekil 3.7'de verilmiştir.

Şekil 3.7: Havlu egzersizinden örnek bir görüntü.

- Düz sarkaç egzersizi: Yapılan egzersiz sarkaç hareketinin anımsattığından adı sarkaç egzersizi olarak bilinmektedir. Bu egzersiz bir omuz egzersizidir. Sandalyeden bir el ile destek alan aktör bel ve dizini hafifçe kırarak öne eğilir. Öne eğildikten sonra kameraya karşı durarak destek aldığı kol sabit olmak üzere diğer kolunu öne ve arakaya gidecek bir şekilde sallandırır. Egzersizi bitirme aşamasına gelindiğinde aktör ilk konumuna yani kolunu sallandırmaya başladığı ana geri döner ve egzersizini tam o noktada tamamlar. Bu egzersizde dikkat edilmesi gereken husus sabit kolun oynatılmaması ve ağrıyan omuzun mümkün olduğunca çok çalıştırılarak açılmasının sağlanmasıdır. İlgili örnek görüntü Şekil 3.8'de verilmiştir.

Şekil 3.8: Düz sarkaç egzersizinden örnek bir görüntü.

- Dairesel sarkaç egzersizi: Düz sarkaç egzersizine benzer bir yapıda olup serbest kolun kameraya karşı öne ve arkaya sallandırılmasına yerine serbest durumdaki kol ile sanal bir çember çizilmeye çalışılınır. Hastanın bu sarkaç egzersizlerinde dikkat etmesi gereken bir başka husus ise öne doğru eğilirken dengesini kaybetmemesi ve belini zorlamamasıdır. Egzersiz tipi ile ilgili örnek görüntü Şekil 3.9'da verilmiştir.

Şekil 3.9: Dairesel sarkaç egzersizinden örnek bir görüntü.

- Sandalye egzersizi: Sandalye üzerine oturulmuş halde başlanılan bu egzersiz de bir diz egzersizidir. Bu egzersizde sandalyede oturan aktörün dizi bükülü durumda egzersize başlar. Dizini sabit tutarak bacağını kameraya doğru uzatır ve eski konumuna geri getirerek egzersizi tekrar eder. Aktör bacağını kameraya doğru uzattığı zaman kısa bir süre (1-2 saniye) bacağını sabit bir şekilde havada tutar. Egzersiz diğer bacakla devam eder ve bu şekilde tekrar eder. Egzersiz tipi ile ilgili örnek görüntü Şekil 3.10'da verilmiştir.

Şekil 3.10: Sandalye egzersizinden örnek bir görüntü.

- Topuk egzersizi: Hem bel hem de diz egzersizi olarak sayılabilecek bir egzersizdir. Aktör sandalye üzerinde dik oturur durumdayken ayağını sandalye gerisine doğru çeker. Bu çekme işleminde topuğu yerden kesilerek yukarı bakan konuma gelir. Egzersiz ayağın başlangıç konumuna dönmesi ile diğer ayağın aynı egzersizi yapması şeklinde tekrar eder. Bu egzersizde ayağın sandalyenin gerisine doğru gitmesine dikkat edilmelidir. Egzersiz tipi ile ilgili örnek görüntü Şekil 3.11'de verilmiştir.

Şekil 3.11: Topuk egzersizinden örnek bir görüntü.

Tablo 3.1: Egzersiz tiplerinin nesne ve duruş özellikleriyle bağlantısı.

| *Egzersiz Tipi* | *Kullanılan Nesne* | *Duruş* |
|---|---|---|
| Sopa egzersizi | Sopa | Ayakta durma |
| Çapraz sopa egzersizi | Sopa | Ayakta durma |
| Sırtüstü bacak egzersizi | Yok | Sırtüstü yatış |
| Havlu egzersizi | Havlu | Ayakta durma |
| Düz sarkaç egzersizi | Sandalye | Eğilme |
| Dairesel sarkaç egzersizi | Sandalye | Eğilme |
| Sandalye bacak egzersizi | Sandalye | Oturma |
| Topuk egzersizi | Sandalye | Oturma |

# 4. HAAR-BENZERİ ÖZELLİKLERLE HAREKET TABANLI ÖZNİTELİK ÇIKARIMI

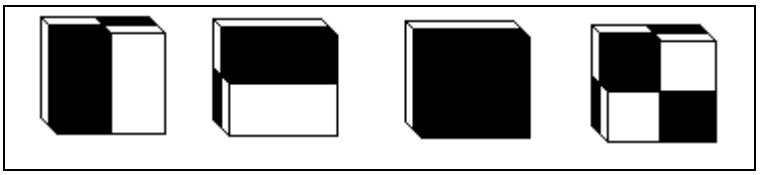
Yapılan tez çalışmasının omurgasını oluşturan hareket tabanlı öznitelikleri çıkarma için özgün bir yöntem geliştirilmiştir. Geliştirilen yöntem, hareketli görüntüde var olan kısa dönem hareket örüntülerinin yakalanması ve bunun tüm hareketli görüntünün geneli hakkında bilgi sağlayacak küresel bir gösterim elde edilmesi prensibine dayanarak çalışmaktadır. Bu küresel gösterim ile hareket tabanlı öznitelikler tanımlanmıştır.

Her bir eylemin hareket bilgisinin birbirinden farklı olacağı ve bu hareket bilgilerinin örüntüler olarak ifade edilebileceği üzerine çeşitli çalışmalar yapılmıştır [90]-[92]. Hareket örüntüsünü oluşturmak için örüntü tanımada alanında sıkça kullanılan Haar-benzeri özellikler ilk defa Papageorgiou arkadaşları [90] yaptıkları çalışma ile tanıtılmıştır. Görüntü üzerine dikdörtgen şeklinde konumlandırılmış pencerelerde çalıştırılan yöntem, bu alan içerisindeki piksellerin renk değeri toplamlarının farkı olarak tanımlanmıştır. 2-boyutlu görüntülerde çalışan bu yöntemde pencereler içerisindeki alanın içerisindeki piksellerin bir kısmı +1, bir kısmı -1 değeri çarpılmasıyla elde edilen değerler toplanılmıştır. Elde edilen toplam o bölge içerisinde anlamlı bir örüntü olup olmadığını yorumlamak için kullanılmıştır.

Haar-benzeri özellikler ile 2-boyutlu görüntülerde yüz tanımayı amaçlayan Viola ve Jones, Haar-benzeri özellikleri kolay bir şekilde hesaplamak için integral görüntü adlı bir metot tanımlamışladır [60]. İntegral görüntü yöntemini daha sonra 2-boyutlu görüntülerde yayaların tespiti için kullanmışlardır [93].

Ciu arkadaşları 2-boyutlu Haar-benzeri özellikleri 3-boyutlu uzaya taşımıştır. Yaptıkları çalışmada yayaları tespit etmeyi hedefleyen araştırmacılar, kübik filtreler kullanarak uzam-zaman hacminde örüntüler çıkartmışlardır. Ayrıca yaptıları bu çalışmada integral görüntü yöntemini 3-boyutta çalışacak bir şekle getirerek integral hacim yaklaşımını oluşturmuşlardır [91].

Haar-benzeri özellikler nesne bulma ve tanımada kullanılan görüntü temelli özelliklerdir. Haar-benzeri filtreler sadece 2 boyutlu eksende çalıştıkları için zamana karşı duyarsızdırlar. 3 boyutlu Haar-benzeri filtrelerde ise zaman ekseni bulunmaktadır. Geliştirdiğimiz hareket tabanlı öznitelik çıkarma yönteminde Haar-benzeri özellikler temel alınmıştır.



Şekil 4.1: Hareket tabanlı öznitelik çıkarmada kullanılan Haar-benzeri filtreler.

Geliştirdiğimiz yöntemde onaltı farklı 3-boyutlu Haar-benzeri filtre kullanılmıştır. Kullanılan filtrelerin ana gösterimleri Şekil 4.1'de verilmiştir. Filtre tipleri ve filtrelerin x-y-z eksenlerindeki boyutları ilgilenilen probleme göre değişebilir. Tez kapsamında elde edilen deneysel sonuçlar ile uzamsal boyutta filtrelerin büyüklükleri 4x4 ve 8x8 piksel olarak alınmıştır. Zamansal boyutta ise 4 ve 8'lik çerçevelerle varyasyon sağlanılmıştır. Her bir filtre insanın belirli bölgelerine denk gelerek o bölgelerdeki hareket hakkında öznitelikleri oluşturması hedeflenilmiştir.

Haar-benzeri filtrelerin $3DH_k$, $k = 1,2,...,16$, seçimi belirlendikten sonra bu filtreler verilen hareketli görüntünün $v$, tüm çerçevelerine $t$ Eşitlik (4.1)'deki gibi uygulanmıştır ve sonucunda 3-boyutlu Haar-benzeri özellikler $3DHF$ elde edilmiştir:

$$3DHF_{\mathrm{v}}(\mathrm{x,y,t,k}) = v(\mathrm{x,y},t)*3DH_{\mathrm{k}}, \tag{4.1}$$

Burada $3DHF$ uzam eksenine ($x$, $y$), zaman eksenine ($t$), ve filtre tipine ($k$) bağımlı bir özelliktir. İlerleyen aşamalarda bu bağımlılıklardan kurtarılarak yerellikten genelliğe geçiş sağlanmıştır. Ama bundan önce tüm hareketli görüntüler için ortak bir değerler kümesi sağlanması için $3DHF$ değerleri 0-255 aralığına normalize edilmiştir.

Kısa dönem hareket örüntülerini ardışık çerçevelerdeki uzamdan bağımsız yerel hareket öznitelikleri olarak tanımlanmıştır. Kısa Dönem Hareket Örüntülerinin *KDHÖ* çıkarılmasıyla hareket özniteliklerine uzamdan bağımsızlık kazandırılmıştır. Bu işlem için histogram oluşturma Eşitlik (4.2)'deki gibi yapılmıştır:

$$KDH\ddot{O}_{\mathrm{v}}(t,\mathrm{k}) = \mathrm{Histogram}[3DHF_{\mathrm{v}}(\mathrm{x,y},t,\mathrm{k})]. \tag{4.2}$$

İşlem sonucunda her bir filtrenin çıktıları tek bir histogramda birleştirilerek çerçeve bazında kısa dönem hareket örüntüleri elde edilmiştir. Bu kısa dönem hareket

örüntüleri temel olarak $k$ ile tanımlanan 3 boyutlu Haar-benzeri filtrenin tipine bağlı olarak çeşitli hareket örüntüleri yakalamaktadır.

*KDHÖ* verilen hareketli görüntünün çerçeve sayısına ve kullanılan Haar-benzeri filtrelerin her biri için 256 bölümlük histogramlar olarak tanımlanmıştır. Kısa zamanda hareket bilgisinde önemli değişiklerin olduğu eylem tiplerinde (örneğin hızlıca kolun, elin sallanması eylemi) *KDHÖ* başarılı bir biçimde hareket örüntülerini yakalayabilmektedir. Zıttı durumlarda ise eylem hakkında önemli bilgileri kaçırabildiğinden başarılı olamamaktadır.

Tüm hareketli görüntü tanımlayabilecek bir gösterim elde edebilmek için *KDHÖ,* Küresel Hareket Gösterimine *KHG* (ing. global motion representation) dönüştürülmüştür. Bu dönüşüm için sık kullanılan istatistiksel yöntemler olan ortalama (ing. mean) ve varyans (ing. variance) başvurulmuştur. Verilen *v* hareketli görüntüsü için $KHG_1$ *KDHÖ* ortalamalarını (μ), $KHG_2$ ise *KDHÖ*'lerin varyansını (σ) hesaplamak için Eşitlik (4.3) ve Eşitlik (4.4) kullanılmıştır:

$$KHG_1(\mathrm{k}) = \mu\,(KDH\ddot{O}(t, \mathrm{k})), \tag{4.3}$$

$$KHG_2(\mathrm{k}) = \sigma\,(KDH\ddot{O}(t, \mathrm{k})). \tag{4.4}$$

*KHG*'nin filtre tipi bağımlılığı çözmek ve nihai olarak hareket örüntülerini tanımlamak için her bir hareketli görüntü için tek tanımlanabilecek olacak HAB (HAreket Bilgisi) tanımlanmıştır. Verilen hareketli görüntüden elde edilen *KHG*'ler birbirleri arkasına birleştirilmesi ile *HAB* Eşitlik (4.5)'teki gibi elde edilmiştir:

$$HAB_{\mathrm{v}}(t) = KHG_{1\mathrm{v}}(t, \mathrm{k}_1) \,||\, KHG_{2\mathrm{v}}(t, \mathrm{k}_1) \,||\, KHG_{2\mathrm{v}}(t, \mathrm{k}_n). \tag{4.5}$$

n değeri maksimum filtre sayısını ( tez kapsamında 16) temsil etmektedir.

# 5. RASTGELE ORMANLAR YAPISIYLA GELİŞTİRİLMİŞ HAREKET TABANLI ÖZNİTELİK ÇIKARIMI (ROY_HÖ)

Rastgele ormanlar Breiman tarafından 2001 yılında tanımlanmış bir sınıflandırma algoritmasıdır [94]. Rastgele ormanlar içerisinde karar ağaçlarından oluşan bir küme barındıran bir yapıdır. Orman ismi de ağaçların topluluğu şeklinde yorumlanmasından gelmektedir. Bu yapıda her bir karar ağacı kendi içerisinde düğümlerden oluşur ve diğer ağaçlardan bağımsız olarak kendi sınıflandırmasını sağlar. Eğitim aşamasında gelen girdiler (özellikler) en üst düğümden (ki kök olarak adlandırılır) süzülerek/bölümlenerek alt düğümlere iner. Eğitim aşamasındaki tüm veriler kullanılmaz bu verilerden sadece belirli bir kısmı kullanılır. Eğitim aşaması tamamlandıktan sonra test aşamasına geçildiğinde test örneğinin özelliklerini daha önce eğitilen düğümlere göre bölümlenerek en aşağıda bir yaprağa denk gelir. Bu yaprağın sahip olduğu sınıf etiketi veya oranı test örneğine bu karar ağacı tarafından atanır. Tüm karar ağaçları bu atamayı yaptıktan sonra oylar toplanarak gelen test örneğinin sınıfı belirlenir [94].

Yapılan tez çalışması kapsamında rastgele ormanlar, gelişmiş hareket tabanlı öznitelikleri çıkarma amacıyla kullanılmıştır. Bu amaç doğrultusunda rastgele ormanların sonuç çıktısı kod şekline çevrilmiştir. Bu yaklaşım çeşitli araştırmalarda da [95], [96] kullanılmış bir yöntemdir. Detaylara indiliğinde her bir karar ağacı için yaprak sayısı uzunluğunda bir kod üretilmekte ve bu kod basitçe bir tane 1 ve geri kalanlar yapraklar için 0 olarak alınmaktadır. Bu kodlar birleştirilerek kod-kelimeye dönüştürülerek bir anlamda kümeleme yapılmaktadır. Kod-kelimeler her bir test örneği için tüm karar ağaçlarından gelen kodların ardı ardına eklenmesiyle oluşan bir tanımlayıcıdır.

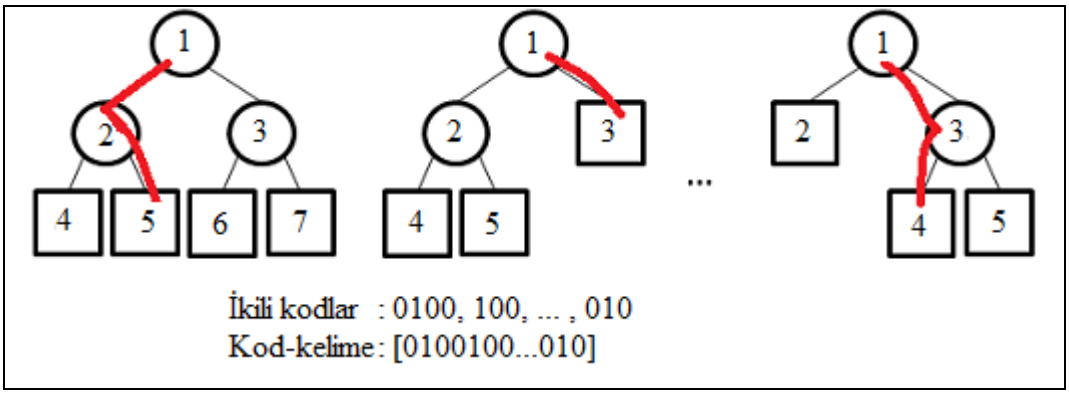
Hareket tabanlı öznitelikler basitçe K-Ortalamalar algoritması kullanarak kümeleneceği gibi bu tez çalışmasında olduğu gibi rastgele ormanlar yapısı kullanarak da icra edilebilir. Hatta bu konuda yakın zamanda yapılmış çalışmalarda ağaç yapısı kullanılarak daha başarılı kümeleme sonuçları elde edilmiştir [97], [98]. Bu başarılar bize ağaç yapısı kullanılması konusunda yol göstermiştir.

Geliştirilen sistem öncelikle eğitim aşamasında ormanların oluşturulmasıyla başlamıştır. Sisteme girdi olarak kısa dönem hareket tabanlı öznitelik çıkarma için tasarlanılmış *KDHÖ* (Eşitlik (4.2)) filtrelerinin sonuçlarını yani hareketli görüntüler

için her çerçeve ve filtre bazında elde edilen öznitelikler bütünleşik bir yapıya kavuşturularak verilmiştir. Bütünleşik yapıdan kasıt edilen öznitelik tanımlayıcısı *KDHÖ*'lerin ardı ardına eklenmesi sonucuyla Eşitlik (5.1)'deki gibi Birleşik Kısa Dönem Hareket Örüntüleri *BKDHÖ*'nin (ing. Concatenated Short-Term Motion Patterns) oluşturulmasıdır:

$$BKDH\ddot{O}_v(t) = KDH\ddot{O}_v(t, k_1) \,||...||\, KDH\ddot{O}_v(t, k_n). \qquad (5.1)$$

*BKDHÖ* verilen bir hareketli görüntünün her bir çerçevesi için 256 x n boyutunda (n tez kapsamında 16 seçilmiştir) bir vektör olarak oluşmaktadır. Bu vektör ile öznitelikler elde edildiğinde her bir çerçeve için aynı boyutta farklı veriler (genellikle) ihtiva eden matrisler oluşmaktadır. Her bir çerçevede yapılan eylem hareketli görüntüde ki eylemin türünden olacağı varsayımdan yola çıkılmıştır. Rastgele orman yapısı bu öznitelik vektörlerinden gelen veri ile eğitilmiştir. Fakat çıktı olarak sadece bir yaprağın seçilmesi yerine tüm karar ağaçları, Şekil 5.1'deki gibi [12], izlediği yola göre değil sonuçlandığı yaprağa göre kodlanmıştır.




Şekil 5.1: Rastgele orman yapısıyla geliştirilmiş özellik çıkarma örneği.

Yapraklara göre kodlama işlemi basitçe karar ağacının bütün düğümlerinin sol üstten, sağ altta giden bir şekilde numaralandırılması ile başlar. Bu numaralandırma sonucunda gelen örnek hangi yaprağa denk geliyor ise o yaprak ikili tabanda 1 diğer yapraklar ise 0 olacak şekilde numaralarına göre ard arda dizilerek o ağaç için bir ikili kod (ing. binary code) oluşturulur. Bu işlem tüm ağaçlar için tekrarlandığından elde edilecek birden fazla ikili kod birleştirilerek bir kod-kelimesi (ing. code-word)

tanımlanılır. Bu kod-kelimesinin uzunluğu her bir örnek için aynı olacağından herhangi bir boyut normalizasyonuna gerek yoktur.

Rastgele orman yapısıyla geliştirilmiş hareket tabanlı özellik çıkarma yönteminde elde edilen kod-kelimeler daha öncede anlatıldığı üzere her bir çerçeve için bir tanedir. Dolayısıyla hareketli görüntünün tümü hakkında genel bir bilgi ihtiva etmemektedir. Bu problemin üstesinden gelmek için daha öncede de başvurulan (Eşitlik (4.3) ve Eşitlik (4.4)) istatistiksel yöntemler olan ortalama ve varyans kullanılmıştır. Bu işlemdeki amaç kod-kelimelerinden en baskın olanın ve kod kelimeler arasındaki değişiklikleri gözlemleyerek hareketli görüntündeki eylemin geneli hakkında bilgi edinebilmektedir. Verilen bir hareketli görüntü *v*'nün her bir çerçevesi için kod-kelimenin KK (ing. codeword) Eşitlik (5.2) ile elde edildiğini varsayarak:

$$\mathrm{KK}_{\mathrm{v}}(t) = \mathrm{KodKelime}(BKDH\ddot{O}_{\mathrm{v}}(t)), \tag{5.2}$$

küresel gösterimde tüm hareketli görüntüyü kapsayacak Geliştirilmiş Küresel Hareket Gösterimi olan *GKHG* (ing. enhanced global motion representation) istatistiksel yöntemler kullanılarak Eşitlik (5.3) aracılığıyla oluşturulmaktadır:

$$GKHG(\mathrm{v}) = \mu(\mathrm{KK}_{\mathrm{v}}(t),\ t=1...T \parallel \sigma(\mathrm{KK}_{\mathrm{v}}(t),\ t=1...T). \tag{5.3}$$

*T* değeri hareketli görüntü *v*'deki çerçeve sayısını ifade etmektedir.

Rastgele Ormanlar yapısıyla, daha ayırt edici hareket tabanlı öznitelikler çıkartılması amaçlanmıştır. Farklı bir yapıda işlenen hareket öznitelikler kullanılarak elde edilen başarının karşılaştırması gerekmektedir. Karşılaştırma açısından Küresel Hareket Gösterimi ile tanımlanan öznitelikler Temel Bileşen Analizine (TBA) verilmiştir. Weizmann veri kümesinde TBA ile seçilen hareket öznitelikleri ve Rastgele Ormanlar yapısıyla oluşturulan hareket öznitelikleri, Destek Vektör Makineleri kullanarak eylem tanıma probleminde çalıştırılmışlardır. Seçilen öznitelik sayıları eşitlendiğinde Rastgele Ormanlar yapısıyla oluşturulan öznitelikler kullanılarak elde edilen doğruluk oranı %2,76 daha başarılıdır. TBA için en ideal değerler seçildiğinde ise doğruluk oran farkı %1,27'ye düşmektedir.

# 6. IZGARASAL YAPIYLA DURUŞ TABANLI ÖZNİTELİK ÇIKARIMI (IY_DÖ)

Fizyoterapi uzmanları hastalarının postürünü yani vücut kısımlarının diziliş ve düzenlerini inceler. Bu diziliş ve düzen, iskeleti ve vücut destek yapılarını zedelenmelere ve çeşitli deformasyonlara koruyacak şekilde dengeli ve düzenli olmalıdır. Her eylemin (örneğin ayakta durma, yatış, oturma, v.b.) ideal bir pozisyonu mevcuttur [99], [100].

Fizyoterapi uzmanları, görsel olarak hastalarını izlediklerinde veya detaylı ölçümlerle hastalarının duruşlarını analiz ettiklerinde hastada varolan problemler hakkında bilgi edinirler. Ayrıca daha önceki bölümde aktarıldığı üzere duruş bilgisi bilgisayarla görme alanında eylemleri sınıflandırma konusunda ayırt edici bir öznitelik olarak karşımıza çıkmıştır. Örnek bir senaryo olarak aynı aktörün bir sopayı tek eliyle ve iki eliyle havaya kaldırıp indirdiği eylem dizisini verilebilir. Bu senaryo kapsamında yapılan iki farklı eylem, nesne bilgisi bakımından incelendiğinde aynı sopanın kullanılmasından dolayı, hareket bilgisi bakımından incelendiğinde ise her iki eylemde de yukarı aşağı ve aşağı yukarı yönlü baskın hareketler olduğundan bu iki bilgiyi kullanarak yapılan eylemlerin farklılaştırılması zor olmuştur. Fakat duruş bilgisi düşünüldüğünde iki eyleminde duruşlarının birbirlerinden kolayca ayrıştırılabildiği fark edilmiştir. Buna ek olarak yüksek gürültü oranının olduğu, nesnelerin birbirlerini kapattığı, hareket bazında çok yavaş değişimlerin olduğu hareketli görüntülerin analizinde duruş tabanlı öznitelikler diğer özniteliklere karşın daha başarılı sonuçlar vermiştir [32].

Tez çalışması kapsamında duruş tabanlı özniteliklerin çıkarılması için bir yöntem önerilmiştir. Bu yöntem ile sistemin genel başarısını arttıracak bir modül oluşturulması amaçlanmıştır. Geliştirilen yöntem sabit arka plana sahip hareketli görüntülerde (ev ortamının modellenmesi) çalışmak üzere tasarlanılmıştır.

Tasarlanılan duruş tabanlı öznitelik çıkarma yöntemiyle öncelikle hareketli görüntülerden arka plan çıkarımı yaparak silüet görüntüler elde edilmiştir. Arka plan görüntülerden kasıt sadece sahnenin bulunduğu görüntülerdir ve bu sabit görüntülerde sahne dışında ne bir aktör ne de etkileşimde bulunan bir nesne bulunmaktadır. Şekil 6.1-a'da örnek bir arka plan görüntüsü ve Şekil 6.1-b'de aynı sahne üzerinde ön planda sadece bir aktörün bulunduğu örnek görüntüler verilmiştir.

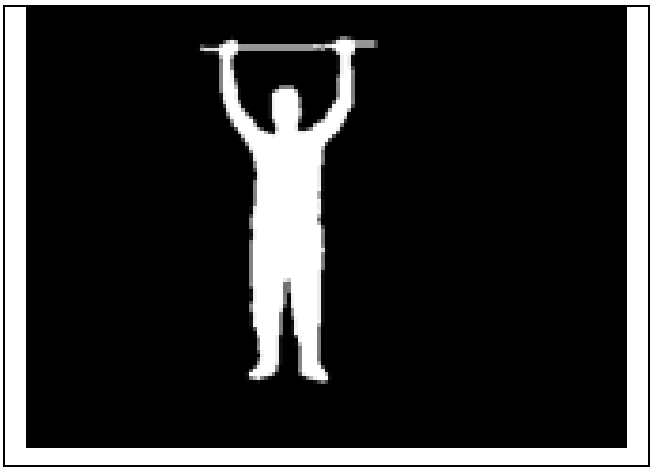
Bu örnek görüntüler çalışma kapsamında oluşturulan fizik tedavi egzersiz veri kümesindeki hareketli görüntülerden alınmıştır.

a) b)

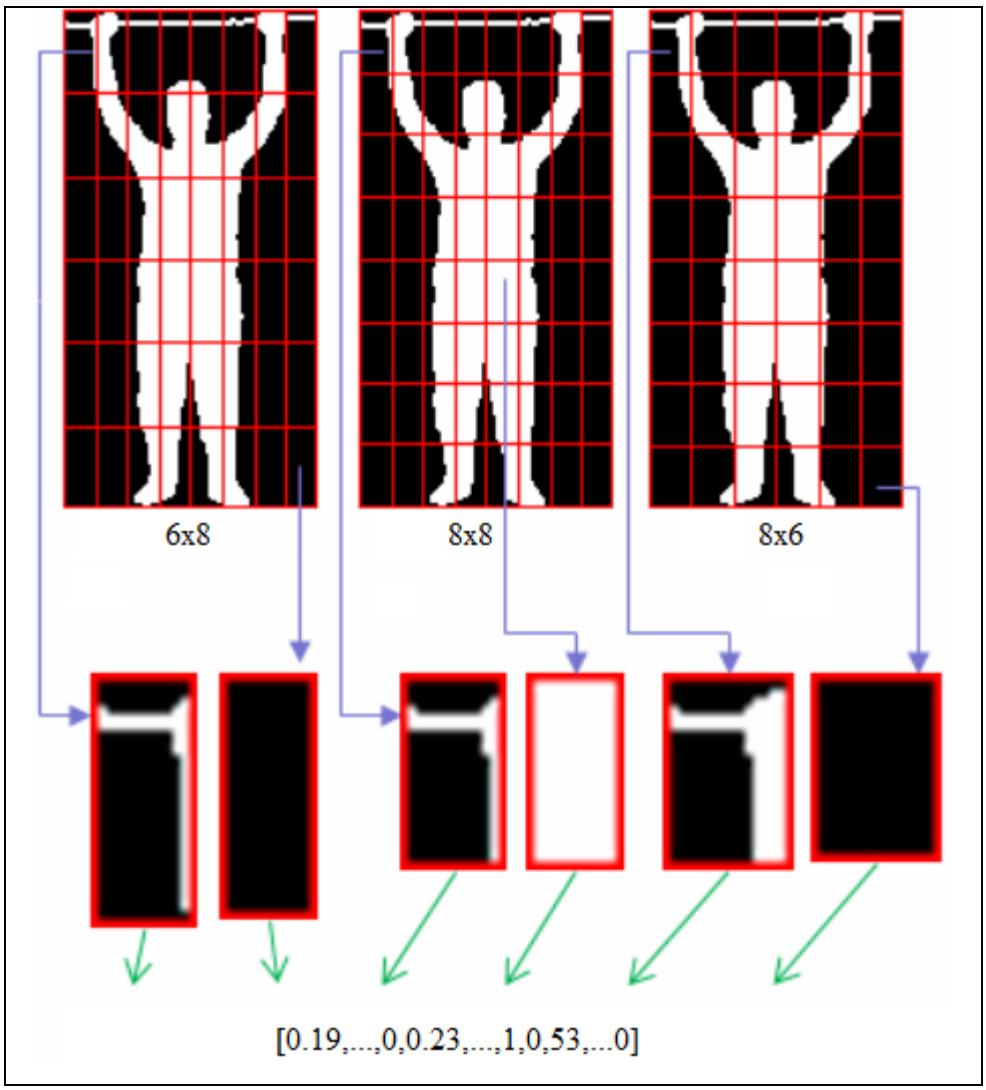
Şekil 6.1: Sopa kaldırma egzersizinden örnek bir görüntünün arka planı ve aslı. a) arka plan görüntüsü, b) egzersiz görüntüsü.

Arka plan çıkarımı işleminden sonra elde edilen siyah-beyaz renklerdeki silüet görüntülerde en büyük nesne tespit edilmiştir. En büyük nesneyi tespit etmek için sekiz komşuluk esasına dayalı olarak birleşik bileşen analizi (ing. connected component analysis) [101], [102] kullanılmıştır. Bu işlem renk veya derinlik görüntülerinde aynı şekilde yapılmıştır. Hem renk hem de derinlik görüntülerinin ortak kullanıldığı durumlarda ise her ikisinde de ortak bulunan ön plan pikselleri esas alınarak görüntüdeki en büyük nesne tespit edilmiştir. Şekil 6.1-b'de verilen örnek egzersiz görüntüsünde en büyük ön plan nesnesi tespit edildiğinde elde edilen görüntü Şekil 6.2'de verilmiştir.

En büyük ön plan nesnesinin konumları belirlendikten sonra bu nesne çerçeve içine alınmıştır. Farklı çözünürlükteki görüntülerde, farklı aktör boyutlarında ölçekten bağımsız olarak elde edilmek istenilen duruş özniteliklerinin çıkarılması için bu pencere içerisinde ızgara biçiminde hücreler oluşturulmuştur. Bu bölümlerin boyutları farklı problemler için değişebileceği gibi bu tez çalışmasında 6x8, 8x8, 8x6 olmak üzere üç farklı bölümle kullanılmıştır. Şekil 6.3'te [12] verilen *f* çerçevesi için en üst satırda üç farklı boyutta ızgaralar kullanıldığı elde edilen hücreler görülebilmektedir.



Şekil 6.2: En büyük ön plan nesnesini barındıran örnek bir görüntü.




Şekil 6.3: Duruş bilgisi (DUB) öznitelik vektörünün oluşturulması.

Her ızgara hücresinin içerisine vücudun belli bir bölgesinin düştüğü varsayımdan yola çıkılarak görüntüde bulunan Duruş Bilgisine (*DUB*) ulaşılmak hedeflenilmiştir. Bu hedef doğrultusunda hücrelerden istatistiksel bilgi çıkartılmıştır. Verilen hareketli görüntünün çerçevesinde en büyük ön plan nesnesini barındıran alan pencere içine alınarak $f$ görüntüsü elde edilmiştir. Bu silüet görüntü (siyah-beyaz renkli) ( $f$ ) ızgarasal yapı içerisine alındıktan sonra ızgara hücrelerindeki $cl = 1,2,..., p$ önplan ve tüm piksel sayıları hesaplanarak Eşitlik (6.1) ile verilen hücre bazında duruş bilgisi $DUB_p$ elde edilmiştir. Izgara türleri, boyutları değiştirilerek ızgara hücre sayısı arttırabileceği gibi probleme özgü ızgara türleri de tanımlanabilir.

$$DUB_p = \frac{\text{ön_plan_piksel_adedi}_p}{\text{Tüm_piksel_adedi}_p} \tag{6.1}$$

Bütün $DUB_p$ değerleri tek bir vektörde birleştirilerek Şekil 6.3'te görüldüğü üzere hareketli görüntünün ilgili çerçevesi için $DUB_f$, duruş tabanlı öznitelikleri barındıran tanımlayıcı vektör oluşturulmuştur.

Duruş tabanlı öznitelikler hareketli görüntünün seçilen bir çerçevesi ve arkaplan görüntüsü kullanılarak elde edilmektedir. Hareketli görüntülerde birden çok çerçeve olmaktadır. Çerçeve sayısı her bir hareketli görüntü için aynı olmayabilmektedir. Duruş tabanlı öznitelikleri yerellikten kurtarıp hareketli görüntünün tümünü tanımlayabilecek hale getirmek için çeşitli yaklaşımlar kullanılabilir, örneğin anahtar pozun seçilmesi [36]. Geliştirilen yöntemde ise düzgün örnekleme (ing. uniform sampling) kullanılarak verilen hareketli görüntünden $n$ tane örnek seçilmiştir. Yapılan deneysel çalışmalar sonucunda $n$=20 olarak alınmıştır.

Verilen hareketli görüntüden seçilen yirmi örnek çerçeve için oluşturulan $DUB_f$ vektörlerinden, $f = 1...n$ olmak üzere, genel bir öznitelik tanımlayıcısı oluşturmak için istatistiksel metotlara başvurulmuştur. $\mu$ istatistiksel ortalama olmak üzere verilen $v$ hareketli görüntüsü için duruş tabanlı öznitelik tanımlayıcısı $DUB_v$, $n$ tane örneğin istatiksel ortalaması olarak Eşitlik (6.2)'deki gibi oluşturulmuştur:

$$DUB_v = \mu(DUB_n).. \tag{6.2}$$

# 7. HAREKET VE NESNE ÖZNİTELİKLERLE EYLEM TANIMA

İnsanların eylemleri ile bu eylemlerinde kullandığı nesneler arasında bir ilişki mevcuttur. Bu ilişki basit olarak boya yapan ustanın boya fırçası kullanması örneğinden de anlaşılacağı üzere genelde birebir ilişkidir.

Hareket ve nesne bilgisi arasındaki ilişki üzerine çeşitli çalışmalar yapılmıştır [50]-[54]. Bu çalışmalar ile hareket bilgisini kullanarak nesnenin sınıfı tanınmış veya nesnenin tipi belirlenerek hareketin sınıfı tanınmıştır.

Tez çalışması kapsamında geliştirdiğimiz hareket tabanlı öznitelikleri çıkartan yöntemimizi (Bölüm 5) değerlendirmek ve nesne-hareket bilgisi arasındaki ilişkiyi kullanarak karşılıklı tanıma yapabilmek için geliştirdiğimiz özgün sistem bir alt bölümde detaylarıyla aktarılmıştır. Bayes ağları kullanılarak tasarlanılan bu hibrit sistem aracılığıyla görüntülerdeki nesnelerin sınıfı ve bu görüntülerde yapılan eylemin sınıfı tanınmıştır. Hareketli görüntülerin renk tabanlı olanları ve derinlik tabanlı olanları hem birlikte hem de ayrı ayrı ele alınarak sistemin bunları tanıma başarısı deneysel olarak test edilmiştir [10].

## 7.1. Tasarlanılan Sistemin Modeli ve Detayları

Geliştirilen sistem, test hareketli görüntülerinde bulunan tek nesnenin sınıf bilgisini ve bu görüntülerde bulunan aktörün yaptığı eylemin sınıf bilgisini tanımak için inşa edilmiştir. Sistem ana hatlarıyla üç farklı basamaktan oluşmaktadır. Bu basamaklar sırasıyla önişlemler, öznitelik çıkarımı, ve nesne/eylem tanımadır.

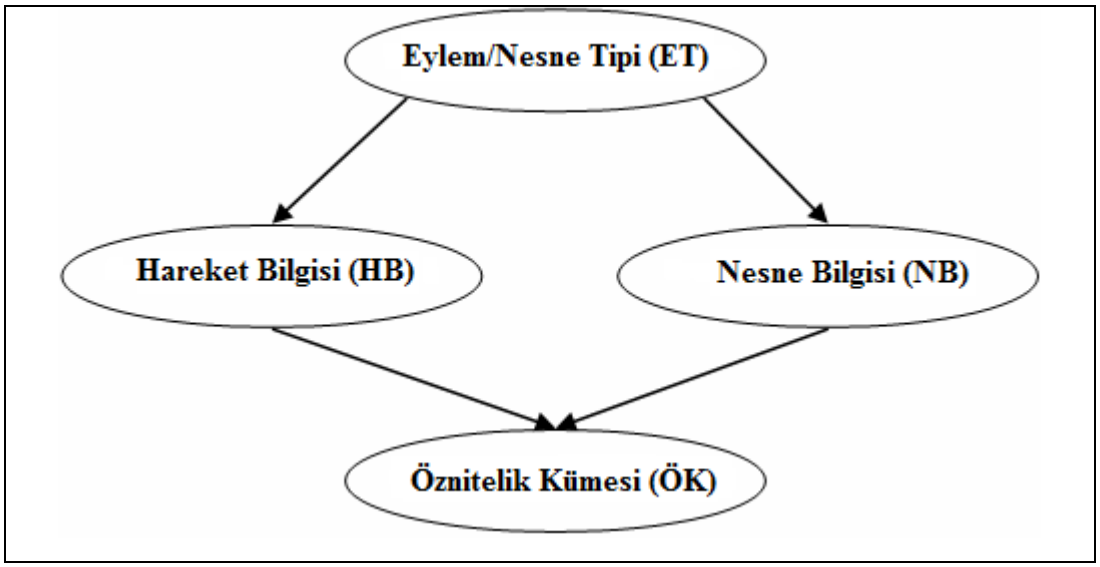
Ön işlemler basamağında hareketli görüntüleri oluşturan KYM resimlerde renk azatlaması yapılarak 256 gri-seviyeli hale getirilmiştir. Bu gri seviyeli resimler G-KYM (Gri-Kırmızı Yeşil Mavi) olarak adlandırılmıştır. Daha sonra gri renkli görüntülerdeki gürültünün azaltılması için medyan filtreleme [103] kullanılmıştır. Medyan (ing. Median) filtrenin boyutu kullanılarak 3x3 olarak alınmıştır. Bu boyut Gall veri kümesi (Bölüm 3.1) için deneysel çalışmalar sonucunda elde edilen en başarılı sonuçtur.

Öznitelik çıkarımı basamağında hareket tabanlı öznitelikleri elde etmek için geliştirdiğimiz Haar-benzeri hareket örüntüleri kullanılmıştır (Bölüm 4). Bu noktada dikkat edilmesi gereken husus derinlik hareketli görüntüleri kullanıldığında da aynı

öznitelik çıkarma yönteminin kullanılmasıdır. Bu işlem bize derinlik görüntüleri için farklı bir yöntem tanımlamak zorunda bırakmamıştır. Hem derinlik hem de renk görüntülerinden elde edilecek hareket örüntüleri kullanılmak istenildiğinde ise elde edilen farklı *KHG*'ler Eşitlik (2.5)'te olduğu gibi art arda eklenilerek öznitelik vektörü tanımlanmıştır.

Nesne tabanlı bilgilere ulaşmak için ise Fei-Fei'nin [58] nesne tespit yöntemi kullanılmıştır (Bölüm 2.1.3.1). Hatırlanacağı üzere bu yöntem verilen görüntüdeki nesnenin konumunu tespit eden bir yaklaşımdır. Hareketli görüntülerde çok sayıda çerçeve olduğundan nesne tespitinin her bir görüntü için tekrar edilmesi gereklidir. Ayrıca nesne kategori sayısı doğrultusunda her bir kategori için bu işlemin tekrar etmesi gerekmektedir. Bu işlemlerden kaynaklanacak zaman kaybını azaltabilmek için hareketli görüntüdeki tüm çerçevelerin yerine bunlardan seçilecek bir küme görüntü üzerinde çalışılması daha yararlı bulunmuştur. Bunun için hareketli görüntüde sabit sayıda çerçeveden bir tanesi alınarak zaman ekseninde örnekleme yapılmıştır. Bu örnekleme her 20 çerçeveden bir tanesi seçilerek eşit aralıkla düzgün bir örnekleme yoluna gidilmiştir.

En son basamak olan nesne/eylem tanıma kısmında verilen hareketli görüntüdeki eylemi/nesneyi tanıyan bir sınıflandırıcı kullanılması gereklidir. Bu sınıflandırma işlemi için ayrımcı (ing. discriminative) veya üretken (ing. generative) yaklaşımlar kullanılabilir. Ayrımcı sınıflandırma yöntemlerine örnek olarak Yapay Sinir Ağları (ing. Neural Networks), Mantıksal Regresyon (ing. Logical Regression), Eklemeli Modeller (ing. Additive Models) ve diğerleri verilebilir. Üretken sınıflandırma yöntemlerine ise Gizli Markov Modelleri (ing. Hidden Markov Models), Fisher Ayırma Analizi (ing. Fisher Discriminant Analysis), Bayes Ağları (ing. Bayesian Networks) ve benzerleri verilebilir [104]. Bu bilimsel çalışmada birleşik dağılımların gösteriminde ve de koşullu bağımsızlıkların (ing. conditional independence) çözümlenmesinde sağladı yararlar bakımında Bayes Ağları sınıflandırıcı olarak seçilmiştir [105], [106].




Şekil 7.1: Hareket ve nesne bilgisi kullanarak eylem/nesne tanıma modeli.

Bu çalışmada tasarlanılan Bayes Ağının grafiksel modeli Şekil 7.1'de verilmiştir. Rastgele değişkenler: Eylem/nesne tipi *ET*, hareket bilgisi *HB*, nesne bilgisi *NB*, öznitelik kümesi *ÖK* olmak üzere aralarındaki koşullu bağımsızlık bağlantıları grafiksel modelden gözlemlenebilmektedir. Bir hareketli görüntünün *v*, öznitelik kümesi *ök* olmak üzere bu hareketli görüntünün eylem/nesne sınıfı SINIF fonksiyonuyla Eşitlik (7.1) ile hesaplanılır:

$$\text{SINIF}(\ddot{o}k) = \underset{\ddot{o}k \in \ddot{O}K}{\arg\max} \sum_{HB,NB} P(ET,HB,NB,\ddot{O}K). \tag{7.1}$$

Eşitlik (7.1)'de *P(ET,HB,NB,ÖK)* ile tanımlanan ifade birleşik olasılık dağılım tablosudur (ing. joint probability distribution table) ve bu ifade açık olarak:

$$P(ET,HB,NB,\ddot{O}K) \propto P(ET)\, P(HB|ET)\, P(NB|ET)\, P(\ddot{O}K|HB,NB) \tag{7.2}$$

Eşitlik (7.2) ile ifade edilir. Eşitlik (7.2)'de verilen ifadelerden *P(ET)* her hareketli görüntü için sabit bir değer olarak alınır. Nesne ve hareket bilgisi arasındaki ilişkiyi temsil eden ifade *P(NB|ET)* ve *P(HB|ET)* için kullanılan veri kümesine göre hazırlanmış tablodan yararlanılır. Bu çalışma için kullandığımız nesne-eylem ilişkisi Tablo 7.1'de verilmiştir.

Tablo 7.1: Gall veri kümesindeki eylem nesne eşleşmesi.

| **Eylem** | **Nesne** |
|---|---|
| Bardağa su dökme | Sürahi |
| İçme | Kupa |
| Uzaktan kumanda kullanma | Uzaktan kumanda |
| Oklava kullanma | Oklava |
| Hesap makinesi kullanma | Hesap makinesi |
| Arama yapma | Cep telefonu |
| Kulaklıkla müzik dinleme | Kulaklık |
| Oyun oynama | Gameboy |
| Resim çekme | Fotoğraf makinesi |
| Kalem kullanma | Kalem |
| Boya yapma | Boya fırçası |

Tablo 7.1 incelendiğinde *P(NB|ET)* ifadesiyle nesne bilgisi ve eylem/nesne sınıfı arasındaki koşullu olasılığa örnek olarak *P(NB=kulaklık|ET=kulaklıkla müzik dinleme)=1* rahatça verilebilir. Müzik dinleme eyleminin sadece kulaklıkla yapılabilecek olmasından dolayı nesne bilgisi ve eylem/nesne sınıfı arasında birebir ilişki mevcuttur. *P(ÖK|HB,NB)* ifadesi ise aksiyomları kullanarak Eşitlik (7.3) ile genişletilir:

$$P(ÖK|HB,NB) = \frac{P(HB,NB|ÖK)\ P(ÖK)}{P(HB,NB)}\ . \tag{7.3}$$

Eşitlik (7.3)'te verilen *P(ÖK)* ve *P(HB,NB)* ifadeleri verilen her *v* için aynı olacağından dolayı ihmal edilerek *P(HB,NB|ÖK)* ifadesi Eşitlik (7.4) ile

$$P(HB,NB|ÖK) = P(HB|ÖK)\ P(NB|ÖK), \tag{7.4}$$

dönüştürülür. Bu dönüşüm sonunda elde edilmesi gereken iki koşullu olasılıktan ilki *P(NB|ÖK)* değeri ise daha önce anlatıldığı üzere Fei-Fei'nin [58] nesne tespit yöntemi ile bulunmuştur. Basitçe yöntem seçilen çerçeveler üzerinde nesne kategorisine göre arama yapmakta ve sonuç olarak seçilen çerçevelerde bulunan nesnelerin genele göre oranları *P(NB|ÖK)* değerini oluşturmaktadır. *P(HB|ÖK)*

değerinin hesaplanılması için ise öncelikle hareket tabanlı öznitelikler kullanılarak eğitilen Destek Vektör Makinesinin her bir hareket bilgisine skor atamasına takiben bu skorların tahmine çevrilmesi için Gibbs dağılımından [107] Eşitlik (7.5) ile yararlanılır:

$$P(HB|ÖK) = 1/Z\ e^{(-Q_m(HB,OK))}. \tag{7.5}$$

Eşitlik (7.5)'te verilen $Q_m(HB,ÖK)$ potansiyel fonksiyonu hareket örüntüleri ile ilgili bilgiyi ifade etmekte $hb|(hb \in HB)$ olup Z ise sabit değer olarak 0,5 alınmıştır.

## 7.2. Deneysel Sonuçlar

Geliştirilen hareket ve nesne bilgisini kullanarak eylem tanıyan sistem [10] Gall veri kümesinde (Bölüm 3.1.1.) test edilerek değerlendirme sonuçlarına ulaşılmıştır. Bu veri kümesinde 60 farklı hareketli görüntü hem renk hem derinlik bilgisi olarak mevcuttur. 10 farklı eylemin/nesnenin 6 farklı aktör ile ilişkisinin test edildiği senaryoda, tüm sistemin başarısı, bir hareketli görüntünün dışarıda test öğesi olarak bırakılması kaydıyla kalan öğelerle eğitilmesi sonucunda elde edilen tanıma bilgilerinin ortalaması verilmiştir. Ortalama tanıma başarısı farklı koşullar altında değerlendirildiğinden bu koşullar ayrı birer deney olarak gözlemlenebilmektedir. Farklı koşullar sırasıyla: sadece nesne bilgisinin kullanılması, nesne ve hareket bilgisinin sadece renk hareketli görüntülerinde kullanılması, hareket bilgisinin hem renk hem de derinlik hareketli görüntülerini kullanarak elde edilmesi ve nesne bilgisinin eklenmesi şeklinde sıralanarak Tablo 7.2'de verilmiştir.

Tablo 7.2 incelendiğinde sadece nesne bilgisinin kullanılmasıyla eylem/nesne tanıma başarısının düşük bir seviyede kaldığı %61,7, fakat hareket bilgisi işin içine katıldığında bu başarı oranının %71,7'ye, hatta derinlik ve renk hareketli görüntülerinden hareket bilgisinin çıkarımı ve nesne bilgisi ile birleştirilmesiyle bu başarı oranının %83,3'te yükseldiği gözlemlenmiştir. Hareket bilgisi derinlik hareketli görüntülerini analiz etmesiyle 60 hareketli görüntünün 50'sinin başarılı bir şekilde tanındığı ortaya çıkmıştır. İlgili karışıklık matrisi ise Tablo 7.3'te verilmiştir.

Tablo 7.2: Gall veri kümesinde çeşitli senaryolarda eylem/nesne tanıma başarısı.

| **Koşul** | **Eylem/Nesne Tanıma Başarısı** |
|---|---|
| Sadece nesne bilgisi kullanıldığında | % 61,7 |
| Nesne ve hareket bilgisi sadece G-KYM (Gri-Kırmızı Yeşil Mavi) hareketli görüntülerden elde edildiğinde | % 71,7 |
| Hareket bilgisi hem derinlik hemde G-KYM hareketli görüntülerinden elde edilmesi ve nesne bilgisi ile birleştirilmesiyle | % 83,3 |

Tablo 7.3: Gall veri kümesinde eylem/nesne tanıyan sistemin karışıklık matrisi. Derinlik bilgisi olmadan-derinlik bilgisi kullanarak başarılı tanıma sonuçları.

| | Su d. | Su içme | Boya y. | Kumanda kullanma | H.M. kul. | Telefon kul. | Kulaklık kul. | G. oynama | Fotograf çekme | Yazı y. |
|---|---|---|---|---|---|---|---|---|---|---|
| Su dökme | 4-5 | 1-1 | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 | 1-0 |
| Su içme | 0-0 | 6-6 | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 |
| Boya yapma | 0-0 | 0-0 | 4-6 | 0-0 | 1-0 | 0-0 | 0-0 | 1-0 | 0-0 | 0-0 |
| Kumanda kullanma | 0-0 | 0-0 | 0-0 | 3-5 | 0-0 | 1-0 | 0-0 | 1-1 | 1-0 | 0-0 |
| Hesap M. kullanma | 0-0 | 0-0 | 0-0 | 0-0 | 4-5 | 0-0 | 0-0 | 1-0 | 0-0 | 1-1 |
| Telefon kullanma | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 | 5-5 | 0-0 | 1-1 | 0-0 | 0-0 |
| Kulaklık kullanma | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 | 0-0 | 6-6 | 0-0 | 0-0 | 0-0 |
| Gameboy oynama | 0-0 | 0-0 | 0-0 | 2-1 | 0-0 | 1-1 | 0-0 | 3-4 | 0-0 | 0-0 |
| Fotograf çekme | 0-0 | 0-0 | 0-0 | 1-1 | 0-0 | 1-1 | 0-0 | 0-0 | 4-4 | 0-0 |
| Yazı yazma | 0-0 | 0-0 | 0-0 | 0-0 | 2-2 | 0-0 | 0-0 | 0-0 | 0-0 | 4-4 |

# 8. HAREKET VE DURUŞ ÖZNİTELİKLERİYLE EYLEM TANIMA

İnsanların eylemlerindeki hareket örüntüleri ve eylemler yapılırken insanların duruş şekilleri arasında bir ilişki mevcuttur. Bu ilişki için çeşitli örnekler verilebilir: yatakta uyuyan bir insanın duruşu yataya yatay bir şekilde uzanmış olması, sıra bekleyen bir kişinin yere dikey olarak ayakta duruyor olması v.b.

Hareket ve duruş bilgisi arasındaki ilişkiyi kullanarak eylem tanıyan çeşitli sistemler mevcuttur [108]-[112]. Bu çalışmalar ile hem hareket bilgisini hem de duruş bilgisi kullanarak eylemin sınıfı tanınmıştır.

Tez çalışması kapsamında geliştirdiğimiz özgün hareket tabanlı öznitelikleri çıkartma yöntemi (Bölüm 4) kullanarak hareketli görüntülerden, hareket bilgisi elde edildikten sonra geliştirilen diğer sistemler gibi hareket bilgisini, duruş bilgisi birlikte kullanarak eylem tanıma sistemi oluşturduk. Geliştirdiğimiz bu özgün sistem aracılığıyla hareketli görüntülerdeki hareket örüntülerini çıkartan yöntemimizi hem tek başına hem de duruş bilgisi ile birlikte eylem tanıma probleminde kullanarak değerlendirme fırsatı yakalanmıştır. Weizmann ve KTH gibi oldukça tanınmış iki veri kümesi üzerinde geliştirdiğimiz özgün eylem tanıma sistemini test ettiğimizde her iki durumda da literatürdeki örneklerine göre başarılı tanıma oranlarına ulaştığımız gözlemlenmiştir [11].

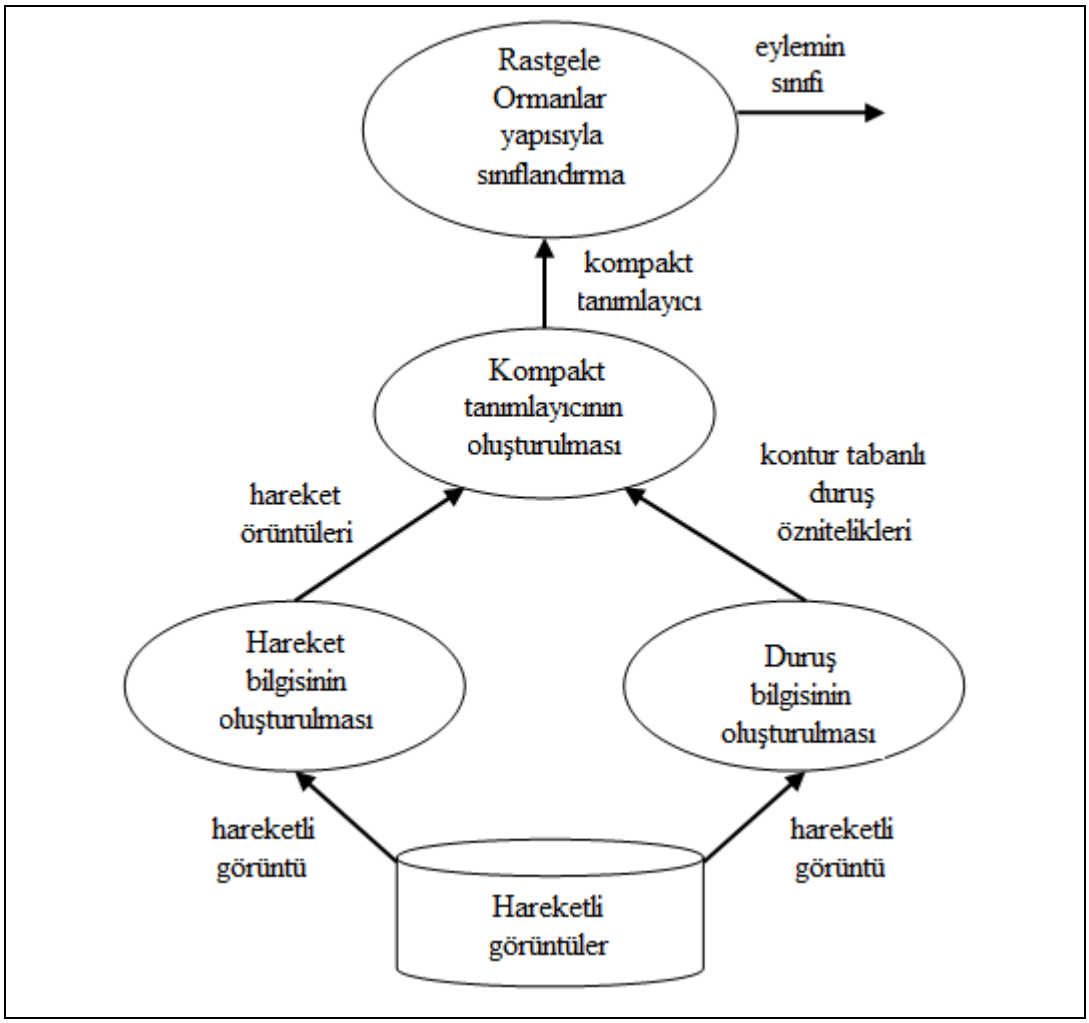
## 8.1. Tasarlanılan Sistemin Modeli ve Detayları

Geliştirilen sistem, test hareketli görüntülerinde bulunan eylemin sınıf bilgisini elde etmek üzere tasarlanmıştır. Sistem ana hatlarıyla üç farklı basamaktan oluşmaktadır. Bu basamaklar sırasıyla önişlemler, öznitelik çıkarımı, ve eylem tanımadır.

Ön işlemler basamağında kullanılan 2 farklı hareketli görüntü veri kümesinde ki (Weizmann ve KTH) hareketli görüntüleri oluşturan RGB resimlerde renk azaltması yapılarak 256 gri-seviyeli hale getirilmiştir. Bu gri seviyeli resimler G-KYM (Gri-Kırmızı Yeşil Mavi) olarak adlandırılmıştır. Daha sonra gri renkli görüntülerdeki gürültünün azaltılması için medyan filtreleme [103] kullanılmıştır. Medyan filtrenin boyutu kullanılarak 3x3 olarak alınmıştır. Hazır kullanılan yöntemden dolayı kontur tabanlı duruş özniteliklerinin elde edilebilmesi için aktör

silüet görüntülerinin hazır olması gereklidir [36]. Bunun elde edilebilmesi için arka plan çıkartımını takiben çeşitli morfolojik işlemler uygulanmıştır.

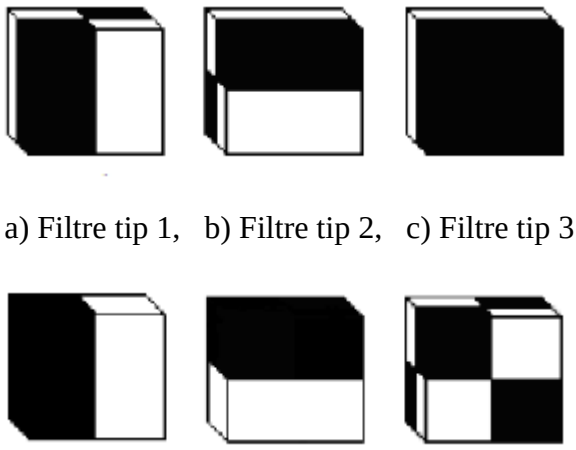
Öznitelik çıkartımı basamağında Şekil 8.1'de verilen veri akış diyagramından da görüldüğü üzere iki farklı öznitelik oluşturulmuştur. Bunlar sırasıyla hareket örüntüleri ve kontur tabanlı duruş özniteliklerdir. Hareketli görüntüdeki, hareket bilgisini elde edebilmek için özgün olan tabanlı öznitelikleri çıkartma yöntemi (Bölüm 4) kullanılmıştır. Aynı hareketli görüntülerdeki duruş bilgisini elde edebilmek için ise Cheema arkadaşları [36] oluşturulan kontur tabanlı duruş öznitelikleri çıkartılmıştır (Bölüm 2.1.2.1).




Şekil 8.1: Hareket ve duruş bilgisini kullanarak eylem tanıma veri akış diyagramı.

Hareket bilgisinin oluşturulmasında tez kapsamında Bölüm 4'de detaylıca anlatılmış olan yöntem kullanılmıştır. Fakat bu yöntemden farklı olarak kullanılan Haar benzeri filtre sayısı 16 yerine 24 olarak alınmıştır. Tüm filtre tiplerini gösteren

görsel Şekil 8.2'de verilmiştir [11]. Ek olarak tanımlanan 2 filtre tipi d) ve e) bentlerinde görselleri verilmiş olanlardır. Bu iki filtre tipi farklı zamanlar içerisinde x-y düzlemindeki değişikliklere yoğunlaştırılmış bir yapıdadır. 6 farklı filtre tipi için konumsal uzayda 4x4 ve 8x8 filtre büyüklükleri olmak üzere zamansal uzayda ise 4 ve 8 çerçeve zaman penceresi alınmak üzere 24 farklı filtre tanımlanmıştır. Belirtilen yöntemde diğer tüm işlemler aynı şekilde devam ettirildikten sonra verilen hareketli görüntünün her çerçevesi için kısa dönem hareket örüntüleri *KDHÖ* Eşitlik (4.2)'deki gibi hesaplanılmıştır. Bu örüntüler daha sonra kompakt tanımlayıcı oluşturulurken kullanılmıştır.




Şekil 8.2 Hareket ve duruş bilgisiyle eylem tanımada kullanılan filtre tipleri.

Duruş bilgisinin oluşturulmasında Cheema arkadaşları [36] tarafından oluşturulan kontur tabanlı duruş özniteliklerini oluşturan yöntem kullanılmıştır. Bu yöntem ile verilen hareketli görüntünün her bir çerçevesi için Eşitlik (2.4)'de tanımlanmış olan $\overline{D}$ vektörüyle duruş bilgisini tanımlanmıştır.

Kompakt tanımlayıcı ile duruş ve hareket bilgisi yerel olmaktan (çerçeve bazında) çıkartılmak istenmiş ve birleşik bir tanımlayıcı oluşturmak hedeflenmiştir. Hareketli görüntünün tüm çerçeveleri için hesaplanılan hareket örüntüleri *KDHÖ* Eşitlik (4.3) ve Eşitlik (4.4) kullanılarak küresel hareket gösterimine *KHG* dönüştürülmüş ve daha sonra ise Eşitlik (4.5) kullanılarak verilen hareketli görüntüdeki hareket bilgisini temsil eden *HAB* vektörü oluşturulmuştur.

Duruş bilgisi öznitelik vektörü olan $\overline{D}$ Eşitlik (2.4) ile hesaplanılmıştır. Bu vektör verilen hareketli görüntünün bir çerçevesindeki ($t$) duruş bilgisini göstermektedir. Tüm hareketli görüntü boyunca baskın olan duruşun tespit edilmesi ve diğer duruşların bu baskın duruştan farklılıklarının hesaplanılması için istatistiksel yöntemler olan ortalama ve varyansa başvurulmuştur. Kompakt tanımlayıcının $KT$(v) oluşturulması işleminde ise Eşitlik (5.5)'ten yararlanılmıştır:

$$KT(\mathrm{v}) = \mu\,(\overline{D}) \,\|\, \sigma\,(\overline{D}) \,\|\, HAB(\mathrm{v})\,. \tag{5.5}$$

Eşitlik 5.5'te $v$ incelenen hareketli görüntüyü, $HAB$(v) bu hareketli görüntünün hareket bilgisini, $\overline{D}$ ise hareketli görüntüdeki tüm duruş özniteliklerini temsil etmektedir.

Hareketli görüntüler için kompakt tanımlayıcılar oluşturulduktan sonra eylemin sınıflandırılması basamağına geçilmiştir. Bu basamakta kullanılan veri, eğitim kümesi ve test kümesi olarak iki farklı kısma ayrılmıştır. Eğitim kümesi kullanılarak oluşturulan Rastgele Ormanlar yapısıyla test örneğindeki eylemin sınıf bilgisi tahmin edilmiştir.

## 8.2. Deneysel Sonuçlar

Geliştirilen hareket ve duruş bilgisini kullanarak eylem tanıyan sistem [11] KTH veri kümesinde (Bölüm 3.1) ve Weizmann veri kümesinde (Bölüm 3.2.) test edilerek değerlendirilmiştir. Her iki veri kümesinde de test örneğinde bulunan aktör ile ilgili tüm hareketli görüntüler eğitim kümesinden çıkartılarak sistem eğitilmiştir. Test edilen örnekteki aktörü dışarıda bırakma (ing. leave-one-actor-out) yaklaşımı olarak da bilinen bu yöntem ile sistemin aktörler üzerine değil de eylemler üzerine yoğunlaşması sağlanmıştır.

Test aşamasında Weizmann veri kümesinde eylemlerin tanınması için hareket ve duruş bilgisi hem ayrı ayrı hem de birleşik olarak (kompakt tanımlayıcı) kullanılmıştır. Bu kullanımdaki amaç hareketli görüntüdeki eylemin tanınması aşamasında duruş ve hareket bilgisinin tanıma başarısına tekil olarak katkılarının incelenmesidir. Tekil olarak eylem tanıma başarısı incelenirken sınıflandırma aracı olarak Destek Vektör Makineleri, kompakt tanımlayıcı ile eylem tanıma başarısı incelenirken Rastgele Ormanlar yapısı kullanılmıştır.

Weizmann veri kümesinde bulunan 90 hareketli görüntüden 79'u sadece duruş bilgisini kullanarak başarılı bir biçimde tanınmıştır. Sadece duruş bilgisini kullanılarak yapılan bu tanımada %87,78 doğruluk oranına ulaşılmıştır. Aynı şekilde sadece hareket bilgisi kullanıldığında 90 hareketli görüntüden 82'si başarılı bir biçimde tanınarak %91,11 doğruluk oranı elde edilmiştir. Bu noktada hareket bilgisinin daha ayırt edici özellikler barındırdığını gözlemlenmiştir.

Kompakt tanımlayıcı kullanarak hem duruş hem de hareket bilgisi birleştirildiğinde ve bu tümleşik bilgi sayesinde hareketli görüntüler tanımlanarak sınıflandırma yapıldığında ise Weizmann veri kümesinde bulunan 90 hareketli görüntünün 85'inin eylemi başarılı bir biçimde tanınarak %94,44 doğruluk oranına ulaşılmıştır. Ayrıca bu deneyde %92444 ortalama duyarlılık oranı, ve %97,75 ortalama keskinlik elde edilmiştir. Bu noktada kompakt tanımlayıcının hem hareket hemde duruş bilgisinden yararlanarak daha başarılı bir sonuç elde ettiği gözlemlenmiştir. Weizmann veri kümesinde kompakt tanımlayıcı kullanılarak eylem bazında tanıma yapıldığında elde edilen sonuçlar Tablo 8.1'de detaylı olarak verilmiştir.

Tablo 8.1 incelendiğinde sistem tarafından en çok karıştırılan eylemin (dokuz hareketli görüntüden ikisinde hatalı tanıma yapılmıştır) sekerek ilerleme eylemi olduğu gözlemlenmiştir. Bu eylemde aktör tek ayağı üzerinde zıplayarak yol almaktadır. Dolayısıyla yaptığı bu eylem hem koşma hemde zıplama eylemine bir ölçüde benzemektedir.

KTH veri kümesinde kompakt tanımlayıcı kullanan sistemimiz ile tanıma yapıldığında %92,99 doğruluk oranı, %92,99 ortalama duyarlılık oranı, ve %92,97 ortalama keskinlik elde edilmiştir. Duruş ve hareket bilgisi kullanılarak bu veri kümesi üzerinde elde edilen sonuç her ne kadar literatürdeki çalışmalara yakın olsa da istenilen seviyede değildir. Bunun nedenleri ise kamera hareketi, görüntülerdeki gürültü, dokusal arka plan olarak sıralanmıştır. Eylemler bazında inceleme yapıldığında ise en çok karıştırılan iki eylemin rahat adımlarla yürüyüş-koşma ve koşma olduğu tespit edilmiştir. Bu iki eyleminde duruş ve hareket bilgilerinin birbirlerine benzerliği bu karıştırma da etkin rol oynamıştır. Kalan yanlış tanımalar incelendiğinde ise kendi içlerinde iki grupta toplanabileceği gözlemlenmiştir. Boks, alkış ve el sallama üçlüsü üst vücut hareketleri olarak kendi aralarında bazı hareketli görüntüler için yanlış tanıma sonuçları vermiştir. Diğer grupta ise koşma, yürüme ve rahat adımlarla yürüyüş-koşma eylem üçlüsü bulunmaktadır. Bunlar ise alt vücut

hareketleri olarak kendi arasında gruplanabilmektedir. Tablo 8.2'de KTH veri kümesinde kompakt tanımlayıcı kullanarak yapılan eylem tanıma sonuçları hareketli görüntü bazında karışıklık matrisi olarak verilmiştir.

Tablo 8.1: Weizmann veri kümesinde eylem tanıma karışıklık matrisi.

| | T.e.s. | İ.e.s. | Eğilme | Yürüme | Koşma | Z. İ. | Kollarını ve bacaklarını açarak ve kapatarak zıplama | Zıplama | Y.Y. | S.İ. |
|---|---|---|---|---|---|---|---|---|---|---|
| Tek elini sallama | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| İki elini sallama | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Eğilme | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Yürüme | 0 | 0 | 0 | 8 | 1 | 0 | 0 | 0 | 0 | 0 |
| Koşma | 0 | 0 | 0 | 1 | 8 | 0 | 0 | 0 | 0 | 0 |
| Zıplayarak ilerleme | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 1 |
| Kollarını ve bacaklarını açarak ve kapatarak zıplama | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 0 | 0 | 0 |
| Zıplama | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 0 | 0 |
| Yengeç yürüyüşü | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 0 |
| Sekerek ilerleme | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 7 |

Kompakt tanımlayıcı yapısı ile duruş ve hareket bilgisi birleştirilerek tanıma yapıldığında Weizmann veri kümesi üzerinde tanıma doğruluk oranında %3,3-%6,66 artış, KTH veri kümesi üzerinde tanıma doğruluk oranında %3,84-%7,12 artış sağlanmıştır. Bu artışlar sadece hareket bilgisi veya sadece duruş bilgisi kullanılmasıyla elde edilen sonuçlar ile karşılaştırma sayesinde elde edilmiştir. Bu noktada duruş ve hareket bilgisinin eylem tanımada ortaklaşa olarak kullanılmasıyla daha başarılı sistemler inşa edilebileceği gözlemlenmiştir. Karşılaştırmalı sonuçlar için Tablo 8.3 incelenebilir.

Tablo 8.2: KTH veri kümesinde eylem tanıma karışıklık matrisi.

| | Yürüme | Rahat adımlarla yürüme-koşma | Koşma | Boks | Alkış | El sallama |
|---|---|---|---|---|---|---|
| Yürüme | 95 | 3 | 2 | 0 | 0 | 0 |
| Rahat adımlarla yürüme-koşma | 5 | 85 | 10 | 0 | 0 | 0 |
| Koşma | 4 | 9 | 87 | 0 | 0 | 0 |
| Boks | 0 | 0 | 0 | 96 | 3 | 1 |
| Alkış | 0 | 0 | 0 | 2 | 96 | 1 |
| El sallama | 0 | 0 | 0 | 1 | 1 | 98 |

Tablo 8.3: Duruş ve/veya hareket bilgisiyle eylem tanıyan sistemlerin doğruluk oranları.

| **Sistem** | **Veri kümesi** | **Aktör adedi** | **Doğruluk oranı** |
|---|---|---|---|
| Bregonzio et al. [66] | Weizmann | 10 | % 96,66 |
| Zhang ve Tao [67] | Weizmann | 10 | % 93,87 |
| Baysal vd. [43] | Weizmann | 9 | % 92,6 |
| Ballan et al. [70] | Weizmann | 10 | % 92,41 |
| Cheema et al. [36] | Weizmann | 9 | % 91,6 |
| Dhillon et al. [72] | Weizmann | 10 | % 88,5 |
| Thurau [75] | Weizmann | 10 | % 86,6 |
| Bregonzio et al. [66] | KTH | 6 | % 94,33 |
| Zhang ve Tao [67] | KTH | 6 | % 93,5 |
| Ballan et al. [70] | KTH | 6 | % 92,1 |
| Baysal vd. [43] | KTH | 6 | % 91,5 |
| Donderac et al. [92] | KTH | 6 | % 85,1 |
| Dhillon et al. [72] | KTH | 6 | % 82,66 |
| Sadece hareket bilgisi | Weizmann | 10 | % 91,11 |
| Sadece duruş bilgisi | Weizmann | 10 | % 87,78 |
| Kompakt tanımlayıcı | Weizmann | 10 | % 94,44 |
| Kompakt tanımlayıcı | KTH | 6 | % 92,99 |

Geliştirilen kompakt tanımlayıcı kullanan sistem ile literatürde var olan bazı sistemlerin aynı veri kümeleri üzerindeki başarılı tanıma oranlarının karşılaştırması

Tablo 8.3'te verilmiştir. Bu karşılaştırmadan da anlaşılacağı üzere duruş ve hareket bilgisini tümleşik olarak kullanan sistemler [66], [67] sadece tek bir bilgi tipi (hareket veya duruş) kullanan sistemlere [36], [43], [72] oranla daha başarılı sonuçlar vermektedir. Bu tez kapsamında öne sürülen farklı bilgi kaynaklarından elde edilen bilgilerin efektif bir şekilde birleştirilmesiyle daha yüksek doğruluk oranına sahip tanıma yapılabileceği varsayımını destekler bir niteliktedir. Ayrıca geliştirilen sistemin literatürdeki diğer sistemler ile yakın sonuçlar vermesi ise geliştirilen sistemin literatürde yerini alabileceğinin bir kanıtıdır.

# 9. FİZİK TEDAVİ EGZERSİZLERİNİ ANALİZ EDEN TEMEL VE GELİŞTİRİLMİŞ SİSTEM

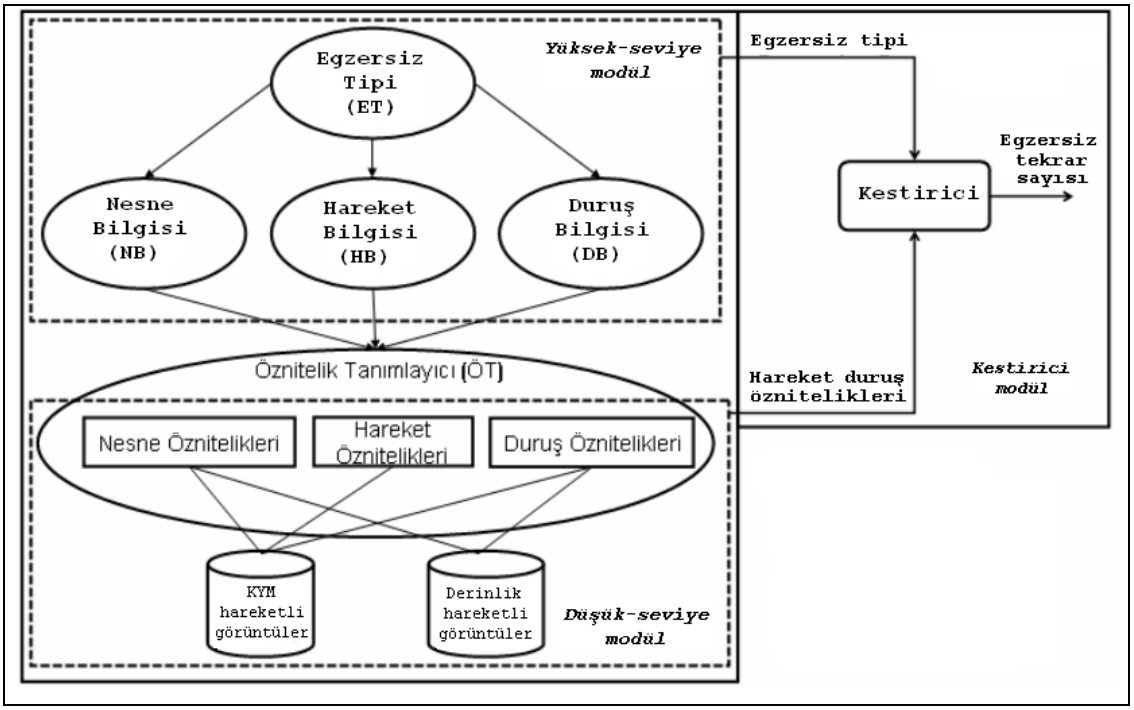
Fizik tedavi egzersizleri hastaların sağlıklarını yeniden kazanması için yapılan hareketler rutinidir. Bu egzersizler hastane ortamında fizyoterapi uzmanları eşliğinde yapılabileceği gibi çeşitli nedenlerden dolayı hastaların evlerinde tek başlarına bunları tekrar etmesi de istenilebilmektedir. Evde yapılan egzersizlerin kontrolü için hastadan yaptığı egzersiz ile ilgili ayrıntılarla doktorunu bilgilendirmesi beklenilmektedir. Bu noktada geliştirilebilecek bir sistem ile hastanın yaptığı egzersiz tipi ve adedi hakkında temel bilgilerin elde edilmesi iyileşme sürecine katkıda bulunacaktır. Yapılan tez kapsamında, fizik tedavi egzersizleri, hareketli görüntülerdeki eylemler gibi ele alınarak bu eylemin sınıflandırılması için iki farklı sistem sunulmuştur. Sunulan sistemler fizik tedavi egzersiz seansı içersinde tekrarlanan egzersizin tipi tanıdıktan sonra takip sürecine geçerek bu egzersizin aynı seans içerisinde kaç defa tekrar ettiğini tahmin etmek içinde özgün bir yöntem sunulmuştur.

Evde yapılan fizik tedavi egzersizleri kayıt altına alarak oluşturduğumuz EOFTE veri kümesinden (Bölüm 3.4) aynı seans içerisinde yapılan egzersizin tipini ve tekrar adedini belirleyecek iki sistem tasarlanmıştır. Bu tasarlanılan sistemler genel olarak egzersizdeki hareket bilgisini, duruş bilgisini ve etkileşimde bulunulan nesne inceleyerek bunları öznitelikler olarak tasvir eder. Elde edilen öznitelikler daha sonra üretken (ing. generative) bir Bayes ağı içerisinde kullanmakta ve böylece egzersiz tipini tanınmaktadır.

Tez kapsamında fizik tedavi egzersizlerini tanımak ve takip etmek için iki sistem tasarlanmış ve bunlar sırasıyla: temel sistem ve geliştirilmiş sistem olarak adlandırılmıştır. Temel sistemin [13] kullandığı yapı, öznitelikler bazından daha zenginleştirilerek geliştirilmiş sistem [12] elde edilmiştir. Bu farkları daha detaylı analiz etmek için her iki sistemde sırasıyla Bölüm 9.1 ve Bölüm 9.2 başlıkları altında incelenmiştir. Her iki sistemde de egzersiz tipi tespit edildikten sonra (takip aşamasında), seans içerisinde aynı egzersizden kaç tekrar yapıldığını tahmin etmek için özgün bir yaklaşım tasarlanmıştır. Egzersiz sayısını tahmin eden yaklaşım Bölüm 9.3 başlığı altında detaylandırılmıştır. Geliştirilen sistemlerden elde edilen deneysel sonuçlar ise Bölüm 9.4 başlığı altında değerlendirilmiştir.

## 9.1. Fizik Tedavi Egzersizlerini Analiz Eden Temel Sistem

Evde yapılan fizik tedavi seansları içerisindeki egzersizlerin tiplerini ve tekrar sayılarını bulmak için tasarlanılan temel sistem [13] üç ana modülden oluşmaktadır. İlk modül ile hareketli görüntü bazında düşük seviye (ing. low-level) öznitelikler elde edilmiştir. Bu öznitelikler ile hareketli görüntüdeki eylemin hareket bilgisine, duruş bilgisine ve nesne bilgisine erişilmiştir. Bu öznitelikler genel olarak öznitelik tanımlayıcısı olarak gruplandıktan sonra ikinci modül olarak yüksek seviye (ing. high-level) olarak tanımlanan üretken Bayes ağına verilerek egzersiz tipi tanınmıştır. Egzersiz tipi ve ilgili hareket-duruş öznitelikleri daha sonra son modül olan kestiriciye (ing. estimator) verilmiştir. Bu son modülde ise aynı seans içerisindeki egzersizin tekrar sayısını tahmin edilmeye çalışılmıştır. Tüm bu modüler yapının sistem tasarımı Şekil 9.1'de verilmiştir [13].




Şekil 9.1 Fizik tedavi egzersizlerini analiz eden temel sistemin modüler yapısı.

Hareket öznitelikleri, verilen bir hareketli görüntüdeki egzersizin hareket bazındaki özelliklerini tanımlamaktadır. Bu öznitelik tanımlayıcısı verilen her hareketli görüntü için Bölüm 4'de anlatıldığı şekilde oluşturulmuştur. Ayrıca iki farklı hareketli görüntü tipi için (KYM ve derinlik hareketli görüntüleri) ayrı ayrı hesaplanıp Eşitlik (4.5)'te verildiği gibi *HUB* hareket tabanlı öznitelik tanımlayıcısı

oluşturulmuştur. Hem derinlik hem de KYM hareketli görüntülerin ortak kullanılması durumunda her iki veri tipi için elde edilen *HUB*'lar yan yana birleştirilmek suretiyle tümleşik tanımlayıcı oluşturulmuştur.

Duruş öznitelikleri, verilen bir hareketli görüntüdeki egzersizi icra eden hastanın duruşu hakkında özellikleri tanımlamaktadır. Bu öznitelik tanımlayıcısı verilen her hareketli görüntü için Bölüm 6'da anlatıldığı şekilde oluşturulmuştur. Parametre olarak verilen hareketli görüntünün her 20 çerçevesinde bir defa çıkartılan duruş öznitelikleri daha sonra Eşitlik (6.2)'de olduğu gibi *DUB* duruş öznitelikleri tanımlayıcısı olarak elde edilmiştir. Hem derinlik hem de KYM hareketli görüntülerin ortak kullanılması durumunda her iki veri tipi için elde edilen *DUB*'lar yan yana birleştirilmek suretiyle tümleşik tanımlayıcı oluşturulmuştur.

Nesne öznitelikleri, verilen bir hareketli görüntüdeki egzersiz seansında hastanın etkileşimde olduğu nesne tipi hakkında bilgi vermektedir. Nesne özniteliği olarak tanımlanan bu özelliğin elde edilebilmesi için Bölüm 2.1.3.1'de anlatılan Fei-Fei'nin nesne tespit yöntemi [58] kullanılmıştır. Verilen hareketli görüntünün her 20 çerçevesindeki KYM görüntülerde, sisteme tanımlanan tüm nesne tipleri aranmış ve bulunan nesne tiplerinin sayısının tüm arama yapılan görüntü sayısına oranı hesaplanarak nesne tipi hakkında bilgi edinilmiştir.

Yüksek seviye modül içerisinde tanıma görevini yerine getirmek için tasarlanmış üretken Bayes ağı, verilen hareketli görüntüdeki egzersiz tipini *et* tüm egzersiz tipleri kümesi *ET* içerisinden bulmaya çalışmaktadır, $et \in ET$.

Üretken Bayes ağı Şekil 9.1'de görülen yüksek seviye modül içerisindeki grafiksel model olarak tasarlanmıştır. Bu grafiksel modelden anlaşılacağı üzere ağ içerisindeki rastgele değişkenler Egzersiz Tipi *ET*, Nesne Bilgi *NB*, Hareket Bilgisi *HB*, Duruş Bilgisi *DB*, Öznitelik Tanımlayıcı *ÖT* şeklinde sıralanmaktadır. Değişkenler arasındaki bağlantılarında bu grafiksel model esas alınarak yapılmıştır. Dolayısıyla üretken Bayes ağı ile bir hareketli görüntüdeki (*v*) egzersiz tipinin *et* tanıması işlemi ($L_v$) için seçilen öznitelik tanımlayıcısı $öt \in ÖT$ kullanılarak düzenlenen $L_v(öt)$ Eşitlik (9.1) ile ifade edilebilir:

$$L_v(öt) = \underset{et \in ET}{\arg\max} \sum_{HB,DB,NB} P(ET,HB,DB,ÖT). \tag{9.1}$$

Eşitlik (9.1)'de $P(ET,HB,DB,NB,ÖT)$ birleşik olasılık dağılım tablosunu temsil etmektedir. $P(ET,HB,DB,NB,ÖT)$ Şekil 9.1'de verilen koşullu olasılık bağlantıları kullanarak Eşitlik (9.2) ile düzenlenmiştir:

$$P(ET,HB,DB,NB,ÖT) \propto P(ET)\, P(HB|ET)\, P(DB|ET)\, P(NB|ET)\, P(ÖT|HB,DB,NB)\,. \tag{9.2}$$

Eşitlik (9.2)'de tanımlanan olasılıklar incelendiğinde $P(ET)$=0,125 olarak ortaya çıkmaktadır. Bunun nedeni kullanılan EOFTE veri kümesindeki egzersiz tip sayısının 8 olması ve eşit sayıda örnek içermesidir. Tablo 3.2.1 ve EOFTE veri kümesinin özellik bazındaki tanımları kullanılarak $P(HB|ET)$, $P(DB|ET)$, $P(NB|ET)$ kolaylıkla hesaplanılması rağmen $P(ÖT|HB,DB,NB)$ ifadesinin ilgili aksiyomlar kullanılarak açılması gerekmektedir. Bu açılım yapıldıktan sonra $P(ÖT|HB,DB,NB)$

$$P(ÖT|HB,DB,NB) \propto P(HB|ÖT)\, P(DB|ÖT)\, P(NB|ÖT), \tag{9.3}$$

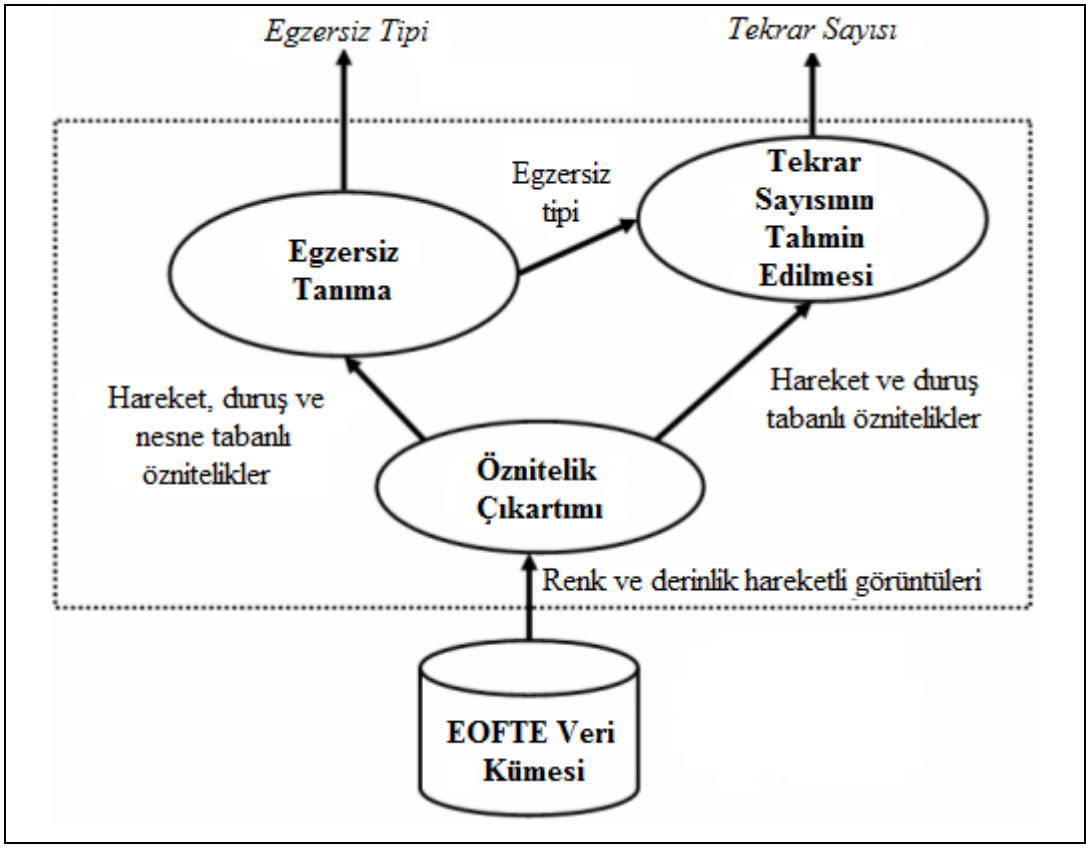
Eşitlik (9.3)'teki gibi ifade edilir. Eksik kalan ifadelerin: $P(HB|ÖT)$, $P(DB|ÖT)$, $P(NB|ÖT)$ hesaplanılması için düşük-seviye modülünden elde edilen öznitelikler kullanılmıştır. Düşük-seviye özniteliklerden ile eksik kalan ifadeler arasındaki bağlantı önceki paragraflarda anlatılmasına rağmen aralarındaki tam bağlantıyı sağlayabilmek sınıflandırma işleminde DVM yöntemi kullanılarak DVM skorları Gibbs dağılımı kullanılarak ilgili olasılıklara çevrilmiştir.

## 9.2. Fizik Tedavi Egzersizlerini Analiz Eden Geliştirilmiş Sistem

Fizik tedavi egzersizlerini analiz eden temel sistemin [13] kullandığı öznitelikler daha fazla zenginleştirilerek geliştirilmiş sistem [12] elde edilmiştir. Özetle, geliştirilmiş sistemde, kullanılan üretken Bayes Ağı temel sisteme oranla daha detaylandırılmış, öznitelik çıkarma yöntemleri daha çok çeşitlendirilmiş ve böylece daha başarılı sonuçlar elde eden bir sistem oluşturulmuştur.

Fizik tedavi egzersizlerini analiz ederek hareketli görüntülerdeki egzersizlerin tipleri ve tekrar sayılarını bulan geliştirilmiş sistemin veri akış diyagramı Şekil 9.2'de verilmiştir. Öznitelik çıkartımı modülünde verilen hareketli görüntüye ait düşük-seviye özellikler elde edilmektedir. Bu düşük-seviye özellikler hareket

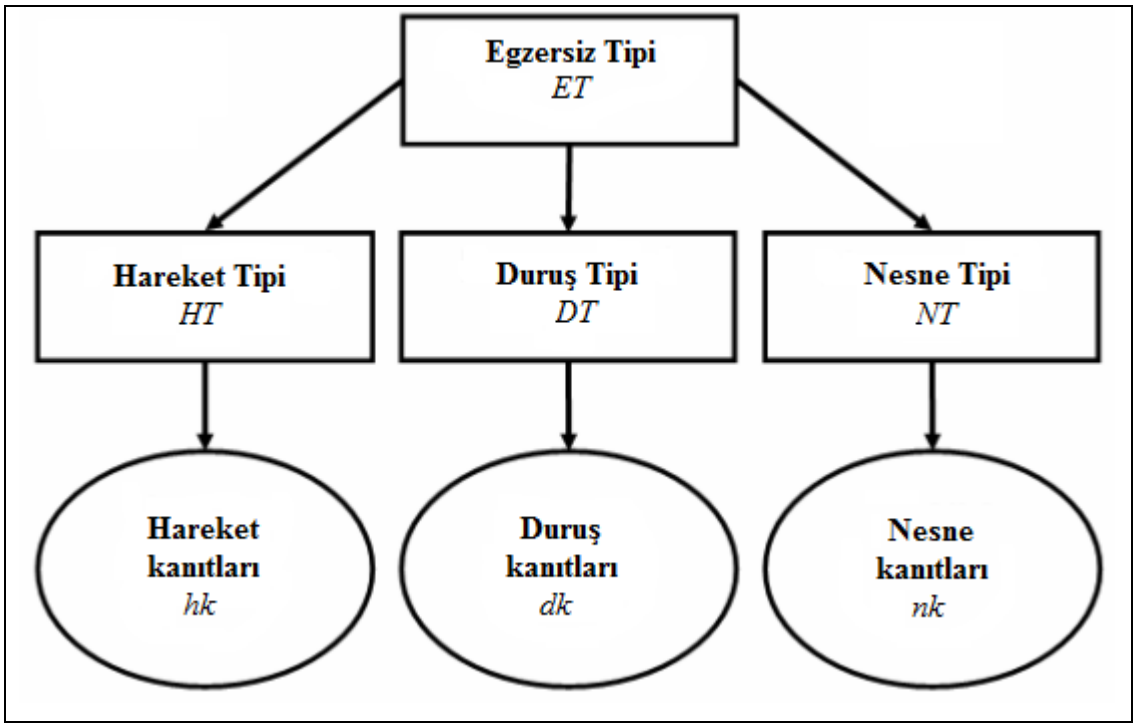
örüntüleri, duruş bilgisi, etkileşimde bulunan nesne tipi olarak sıralanmaktadır. Egzersiz tanıma modülünde üretken bir Bayes ağı kullanılarak hareketli görüntüdeki egzersizin sınıfı (tipi) tanınmaktadır. Bu işlem için gerekli olan kanıtlar ise makine öğrenmesi tabanlı sınıflandırıcıların düşük-seviye özellikleri yorumlamasından elde edilmektedir. Tekrar sayısını tahmin eden modül ise verilen hareketli görüntüdeki hareket ve duruş tabanlı özniteliklerini ayrıca egzersiz tanıma modülü sayesinde tespit edilen egzersiz tipini kullanarak aynı seans içerisindeki egzersizin kaç defa tekrar edildiğini tahmin etmektedir. Ayrıca Şekil 9.2'deki diyagram incelendiğinde temel ve geliştirilmiş sistem arasında modüller arasındaki veri akış yapısının aynı olduğu gözlemlenmektedir.




Şekil 9.2: Fizik tedavi egzersizlerini analiz eden geliştirilmiş sistemin veri akış diyagramı.

Üretken Bayes ağı Şekil 9.3'de verilen grafiksel model esas alınarak tasarlanmıştır. Bu grafiksel modelden anlaşılacağı üzere ağ içerisindeki gizli düğümler Egzersiz Tipi *ET*, Hareket Tipi *HT*, Duruş Tipi *DT*, Nesne Tipi *NT* şeklinde sıralanırken gözlemlenebilen düğümler ise Hareket kanıtları *hk*, Duruş kanıtları *dk*, Nesne kanıtları *nk* şeklinde sıralanmaktadır. Gözlemlenebilen

düğümlerdeki kanıtlar ilgili özniteliklerin makine öğrenmesi tabanlı sınıflayıcılar tarafından yorumlanması ile elde edilmektedir. Bu yaklaşım ile ağın karmaşıklığı azaltılmakta ve egzersiz tipinin tanınması işlemi daha kolay bir şekilde tanımlanmaktadır.




Şekil 9.3: Fizik tedavi egzersizlerini analiz eden geliştirilmiş sistemdeki üretken Bayes Ağının grafiksel modeli.

Şekil 9.3'te verilen Bayes ağı kullanılarak geliştirilmiş sisteme verilen bir hareketli görüntüdeki *v* egzersiz tipinin tanınması işlemi *L(v)* olarak tanımlanmıştır. *L(v)* ilgili öznitelik tabanlı kanıtları (hk, dk, nk) kullanarak hareketli görüntüdeki egzersiz tipini *et* $\in$ *ET* olmak üzere Eşitlik (9.4) ile tanımaktadır:

$$L(v) = \underset{et \in ET}{\arg\max} \sum_{HT,DT,NT} P(ET,HT,DT,NT,hk,dk,nk) \,. \tag{9.4}$$

Eşitlik (9.5)'te *P(ET,HT,DT,NT,hk,dk,nk)* birleşik olasılık dağılım tablosunu temsil etmektedir. *P(ET,HT,DT,NT,hk,dk,nk)* ifadesi Şekil 9.3'de verilen koşullu olasılık bağlantıları kullanarak aşağıdaki eşitlikteki gibi düzenlenmiştir:

$$P(et) \prod P(HT|et)\, P(DT|et)\, P(T|et)\, P(hk|HT)\, P(dk|DT)\, P(nk|NT). \tag{9.5}$$

Eşitlik (9.5)'te tanımlanan olasılıklar incelendiğinde $P(et)$=0,125 olarak ortaya çıkmaktadır. Bunun nedeni kullanılan EOFTE veri kümesindeki egzersiz sınıf sayısının 8 olması ve eşit sayıda örnek içermesidir. Tablo 3.2.1 ve EOFTE veri kümesinin özellik bazındaki tanımları kullanılarak $P(HT|et)$, $P(DT|et)$, $P(NT|et)$ kolaylıkla hesaplanılmaktadır. Örneğin *et* bir sandalye bacak egzersizi ise $P(DT=ayakta\ durma|et=sandalye\ bacak\ egzersizi)$ olasılık ifadesinin değeri 0'a eşittir çünkü veri kümesinde ayakta durulup yapılan bir sandalye egzersizi tanımlı değildir.

Eşitlik (9.5)'te gizli ve gözlemlenen düğümler arasındaki olasılık ifadeleri: $P(hk|HT)$, $P(dk|DT)$ ve $P(nk|NT)$ ise ilgili aksiyomlar kullanılarak Eşitlik (9.6)'daki gibi açılır. Eşitlik 9.6'da verilen olasılıklar $P(hk)$, $P(dk)$, $P(nk)$, $P(HT)$, $P(DT)$ ve $P(NT)$ incelenen hareketli görüntüler için aynı kabul edilmiş ve ilgili işlemlere dahil edilmemiştir. Fakat verilen bir hareketli görüntüdeki egzersizin tipinin bulunması $L(v)$ işlemi için: $P(HT|hk)$, $P(DT|dk)$ ve $P(NT|nk)$ ifadelerinin hesaplanılması gerekmektedir. Bu değerlerin elde edilmesi için makine öğrenmesi tabanlı sınıflandırıcılara başvurulmuştur.

$$
\begin{aligned}
P(hk|HT) &= \frac{P(HT|hk)\,P(hk)}{P(HT)}, \\
P(dk|DT) &= \frac{P(DT|dk)\,P(dk)}{P(DT)}, \\
P(nk|NT) &= \frac{P(NT|nk)\,P(nk)}{P(NT)}.
\end{aligned}
\tag{9.6}
$$

Veri kümesindeki tüm hareketli görüntülerden ilgili öznitelikler (hareket, duruş ve nesne tabanlı) çıkartıldıktan sonra bunlar birer kanıt gibi yorumlanmış ve yüksek sayıda özelliği başarılı bir şekilde sınıflandıran DVM sınıflandırıcıları kullanılarak *HT*, *DT* ve *NT* etiketleri atanmıştır. Bu sınıflandırma işlemlerinde verilen hareketli görüntü test örneği olarak ele alınmış ve geri kalan örneklerden aynı aktörün yaptığı egzersizler hariç tutularak eğitim kümesi oluşturulmuştur.

Klasik DVM sınıflandırıcıları etiket atama (sınıflandırma) işlemini yaptıklarında atanan etiket için bir olasılık değeri sunmazlar. Ama bu olasılık değeri $P(HT|hk)$ ve $P(DT|dk)$ ifadeleri için gereklidir. Bu noktada Gibbs dağılımı tabanlı bir

yaklaşım kullanılarak DVM skorlarından Eşitlik (9.7) aracılığıyla olasılık değerleri elde edilmiştir:

$$P(HT|hk) = 1/Z\, e^{-Q(HT,\, hk)}\,, \qquad P(DT|dk) = 1/Z\, e^{-Q(DT,\, dk)}\,, \tag{9.7}$$

Eşitlik (9.7)'de verilen Q(HT,hk) ve Q(DK,dk) potansiyel fonksiyonlarında ilgili DVM tahmin skorları kullanılmış ve *Z* normalizasyon sabiti 0,5 olarak alınmıştır.

## 9.3. Fizik Tedavi Egzersiz Seanslarında Tekrar Sayısını Bulma Yöntemi

Evde yapılan her fizik tedavi egzersizi seansında genellikle aynı egzersizin birden çok tekrarı bulunmaktadır [8], [81], [113]. Bu tekrar sayısı hasta ile ilgilenen fizik tedavi uzmanı tarafından belirlenmektedir [8]. Aynı egzersizin birden çok tekrar edilmesiyle o bölgenin eski ritmine kavuşturulması ve dolayısıyla iyileşme sürecinin kısalması hedeflenmektedir.

Tez kapsamında evde yapılan fizik tedavi egzersizlerinin tipleri hareketli görüntüler aracılığıyla tanınması hedeflenmiştir. Bu ana hedeften farklı olarak ikincil bir hedef konulmuştur. Bu ikincil hedef aynı seans içerisinde yapılan egzersizden kaç tane yapıldığı sorusunun yanıtını tahmin etmektir. Fizik tedavi egzersiz seanslarının takibi ve analizi aşamasında egzersiz tekrar sayısının bulunması hastanın sergilediği egzersiz performansını gözlemlemek açısından ilgili uzmana yararlı olmaktadır. Ayrıca hastanın kendi kendini kontrol edip evde yapması gereken egzersizleri düzenleyebilmesi açısından da egzersiz tekrar sayının bulunması önemlidir.

Tez kapsamında oluşturulan EOFTE veri kümesindeki egzersiz seanslarında tamamlanmış egzersizler bulunmaktadır. Tamamlanmış egzersizden kasıt egzersizin herhangi bir nedenden dolayı yarım bırakılmadan baştan sona kurallar dahilinde bitirilmiş olmasıdır. EOFTE veri kümesinde egzersiz seanslarında ayrı ayrı olarak hangi aktörün görev aldığı, hangi egzersiz tipinin kaç defa tekrar edildiği bilgisi tam olarak bulunmaktadır.

Aynı seans içerisinde yapılan egzersizin kaç defa tekrar edildiği bulmak için özgün bir yaklaşım geliştirilmiştir [12], [13]. Bu yaklaşım Şekil 9.1 ve Şekil 9.2'de görüldüğü üzere egzersiz tanıma modüllerinden gelen egzersiz tipi bilgisini ve

öznitelik çıkartımı modüllerinden gelen hareket ve duruş özniteliklerini kullanarak egzersiz tekrar sayısını tahmin etmektedir. Bu işlem için öncelikle zamansal bakımdan dilim tabanlı yeni bir küresel gösterim (Dilim Tabanlı Küresel Gösterim, DTKG) tanımlanmıştır. Sisteme verilen bir *v* hareketli görüntüsü için DTKG(τ), *v*'nin ilk çerçevesinden τ. çerçeveye kadarki çerçeveyi kapsayan yeni bir hareketli görüntü $v_\tau$ tanımlanması ile başlar. Daha sonra seçilen duruş ve hareket tabanlı özniteliklerin $v_\tau$ için çıkarılması ile devam eder. Elde edilen öznitelikler ile DTKG(τ) Eşitlik (9.8) ile tanımlanır:

$$\mathrm{DTKG}(\tau) = \mathrm{GKHG}(\mathrm{v}_\tau) \parallel \mathrm{DUB}(\mathrm{v}_\tau). \tag{9.8}$$

Eşitlik (9.8)'de kullanılan; hareket tabanlı öznitelik tanımlayıcısı *GKHG* Eşitlik (5.3) ile, duruş tabanlı öznitelik tanımlayıcısı *DUB* ise Eşitlik (6.2) ile tanımlanmıştır. Bu öznitelik tanımlayıcıları yerine literatürde varolan, tüm hareketli görüntüyü hareket ve duruş tabanlı tanımlayabilecek diğer yaklaşımlarda kullanılabilir.

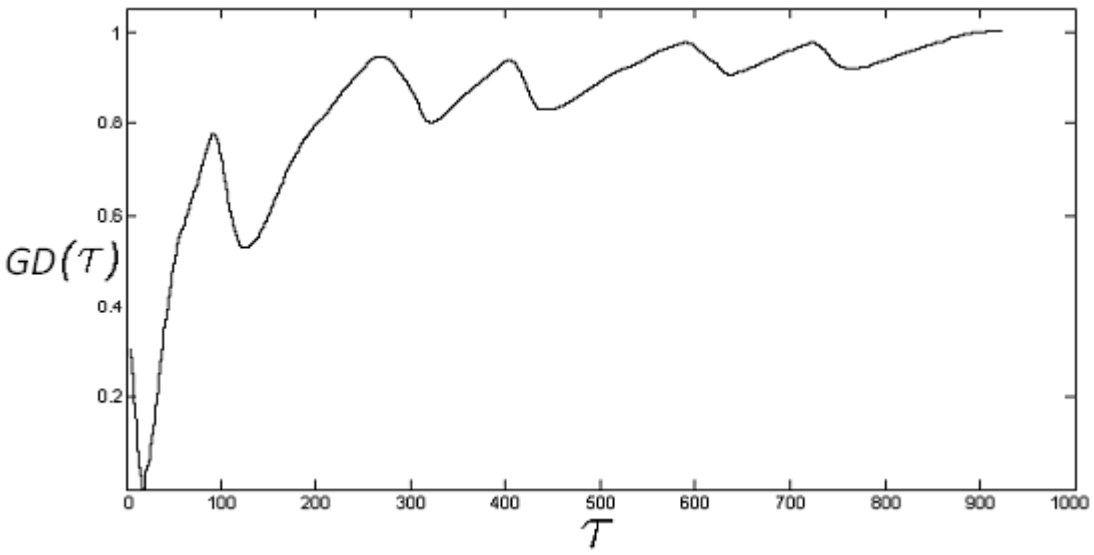
Tekrar sayısının tahmin edilmesindeki ana fikir; ardışık bir biçimde artan τ değerleri için *v* hareketli görüntüsünden yeni DTKG(τ) elde edilmesi ve bunların bulunan egzersiz tipine ne oranda benzediğinin yorumlanmasıdır. Egzersiz tipi hatırlanacağı üzere egzersiz tanıma modülü sonunda o seans için atanan egzersiz etiketidir. Seans içerisinde tekrarlanan egzersizin tekrar sayısı arttıkça atanan etiketteki egzersize benzeme oranı da artmaktadır. Bu benzerliği sayısal olarak tanımlamak için ikili DVM sınıflandırıcısı kullanılmıştır. DVM sınıflandırıcı verilen hareketli görüntünün tanınan egzersiz tipi mi yoksa diğer egzersizlerden mi olduğu sorusuna yanıt aramak üzere ikili çalıştırılmıştır. Bu benzerlik değerini tanımlamak için Eşitlik (9.9)'da tanımlanan DVM güvenirlik değeri *GD* (skor) kullanılmıştır:

$$\mathrm{GD}(\tau) = \sum_i a_i çf\,(dv_i, DTKG(\tau)) + sp. \tag{9.9}$$

Eşitlik (9.9)'da DVM güvenirlik değeri hesaplamasında: $a_i$ ile ağırlıklar, *çf* ile çekirdek fonksiyonları, $dv_i$ ile destek vektörleri, *sp* ile sapma (ing. bias) tanımlanmıştır. Güvenirlik değerlerinin ardışık bir şekilde artan τ değeri ile oluşturduğu örnek bir eğri Şekil 9.4'te verilmiştir. Dikkat edileceği üzere hareketli görüntü içerisinde egzersiz tekrarları arttıkça *GD* daha önce anlatıldığı üzere

artmaktadır. Bu artış sistemin hareketli görüntüdeki egzersiz tipini daha başarılı bir şekilde tanıdığının kanıtıdır.

Güvenirlik değerleri ve τ eğrisi kullanılarak tekrar sayısı tahmin edilmektedir. Bu tahmin için eğrinin yerel tepe noktalarında seans içerisindeki egzersizlerin bittiği varsayımından yola çıkılarak bu tepe noktalarının sayısı hesaplanılmıştır. Bu hesaplama için eğri ilk önce medyan filtre ile gürültülerinden arındırılmıştır. Daha sonra eğrinin türevinin sıfır olduğu noktalar sayılarak tekrar adedi tahmin edilmiştir. Hareketli görüntüdeki çerçeve sayısı binlere ulaştığında eğri düz bir çizgi haline gelebileceğinden bu durumlarda penceresel bir yaklaşım kullanılarak τ için başlangıç değeri farklılaştırılarak çözüm bulunabilir.




Şekil 9.4: Örnek bir egzersiz seansı için *GD(τ)-τ* eğrisi.

## 9.4. Deneysel Sonuçlar

Fizik tedavi egzersizlerini analiz eden temel ve geliştirilmiş sistemler çeşitli senaryolar ile test edilerek başarıları ölçülmüştür.Her iki sistemde, oluşturulan EOFTE veri kümesi kullanılarak test edilmiştir. İncelenen hareketli görüntü test örneği olarak alınmış, geriye kalan hareketli görüntülerden ise test örneğindeki aktörün hareketli görüntüleri çıkartılarak eğitim kümesi hazırlanmıştır. Bu işlem tüm hareketli görüntüler için tekrar edilerek sistemlerin EOFTE veri kümeleri üzerinde başarıları gözlemlenmiştir.

Fizik tedavi egzersizlerini analiz eden temel sistem hatırlanacağı üzere hareket tabanlı öznitelikleri elde etmek için Bölüm 4'te anlatılan Haar-benzeri özellikler ile hareket tabanlı öznitelik çıkarma yöntemini, duruş tabanlı öznitelikleri elde etmek için Bölüm 6'da anlatılan ızgarasal yapı ile duruş tabanlı öznitelik çıkarma yöntemini, etkileşimde bulunan nesne hakkında bilgi edinmek için ise Bölüm 2.1.3.1'de anlatılan Fei-Fei'nin nesne bulma yöntemini kullanmaktadır. Temel sistemin başarısını ölçmek için EOFTE veri kümesinden iki farklı senaryo hazırlanmıştır. İlk senaryoda sadece 256 bit gri renkli hareketli görüntülerden oluşan veri kümesi, ikinci senaryoda ise ilk senaryoya ek olarak derinlik hareketli görüntüleri de eklenen bir veri kümesi hazırlanmıştır. Burada amaçlanan, zaman ve renk bilgisi 2-boyutlu bilgi olarak kabul ettiğimizde, derinlik bilgisinin de eklenmesiyle 3-boyutlu bilgi kullanımının 2-boyutlu bilgi kullanımına oranla sistemin genel başarısına katkısının gözlemlenmesidir.

Fizik tedavi egzersizlerini analiz eden temel sistem EOFTE veri kümesinde bulunan 240 tane hareketli görüntüdeki egzersiz tiplerinden %90,8'ini %91,2 ortalama keskinlik oranı ile sadece renk hareketli görüntülerini kullanarak başarılı bir şekilde tanımıştır. Tablo 9.1'de karışıklık matrisi olarak egzersiz tipi başına başarılı tanınan hareketli görüntü sayıları verilmiştir. Tablo 9.1 incelendiğinde en çok birbirlerine karıştırılan egzersiz tiplerinin düz sarkaç ve dairesel sarkaç egzersizleri olduğu gözlemlenmiştir. Bunun nedeni ise derinlik bilgisi olmadan kamera karşısında yapılan dairesel sarkaç hareketi düz sarkaç hareketi olarak tanınmasıdır. Birbirlerine çok benzeyen bu iki egzersiz tipi arasındaki farkı rahatlıkla ortaya koyabilmek için temel sisteme derinlik hareketli görüntülerinin de verilmesi gerekmektedir.

Tablo 9.1: Sadece renk hareketli görüntüleriyle egzersiz tanıma sonuçları.

| ***Egzersiz Tipleri*** | **Sopa** | **Çapraz Sopa** | **Sırtüstü Bacak** | **Havlu** | **Düz Sarkaç** | **Dairesel Sarkaç** | **Sandalye** | **Topuk** |
|---|---|---|---|---|---|---|---|---|
| **Sopa** | 28 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Çapraz Sopa** | 3 | 27 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Sırtüstü Bacak** | 1 | 0 | 28 | 0 | 0 | 0 | 0 | 1 |
| **Havlu** | 1 | 0 | 0 | 29 | 0 | 0 | 0 | 0 |
| **Düz Sarkaç** | 0 | 0 | 0 | 0 | 25 | 5 | 0 | 0 |
| **Dairesel Sarkaç** | 0 | 0 | 0 | 0 | 3 | 27 | 0 | 0 |
| **Sandalye** | 0 | 0 | 0 | 0 | 2 | 1 | 27 | 0 |
| **Topuk** | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 27 |

Tablo 9.2 ile fizik tedavi egzersizlerini analiz eden temel sistem EOFTE veri kümesinde hem renk hem de derinlik hareketli görüntülerini kullanmasıyla elde edilen tanıma başarıları gösterilmiştir. Karışıklık matrisi olarak egzersiz tipi başına başarılı tanınan egzersiz seanslarının sayısının verildiği Tablo 9.2'de 240 farklı egzersiz seansından %96,25'ini (doğruluk oranı) başarılı bir biçimde tanındığı ortaya çıkmaktadır. Bu deneyde elde edilen ortalama keskinlik oranı ise %96,36'dır.

Derinlik hareketli görüntülerinin dahil olmadığı sonuçları içeren Tablo 9.1 ile karşılaştırıldığında, derinlik bilgisinin sistemin tanıma başarısını arttırdığı gözlemlenmiştir. Ayrıca derinlik bilgisinin kullanılmasıyla sistemin artık en çok karıştırdığı düz sarkaç ve dairesel sarkaç egzersizleri arasındaki karıştırmadan kaynaklanan hatalı tahmin sayısının azaldığı gözlemlenmiştir. Genel olarak temel sistemde derinlik bilgisinin kullanılması ile %5,45 oranında başarı artışı sağlanmıştır. Ayrıca karşılaştırma açısından Bobick ve Davis'in [14] tanımladığı sistem ile EOFTE veri kümesinde eylem tanıma yapıldığında %80,8 doğruluk oranıyla tanıma yapıldığı gözlemlenmiştir.

Tablo 9.2: Hem renk hem de derinlik hareketli görüntüleriyle egzersiz tanıma sonuçları.

| ***Egzersiz Tipleri*** | **Sopa** | **Çapraz Sopa** | **Sırtüstü Bacak** | **Havlu** | **Düz Sarkaç** | **Dairesel Sarkaç** | **Sandalye** | **Topuk** |
|---|---|---|---|---|---|---|---|---|
| **Sopa** | 29 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Çapraz Sopa** | 1 | 29 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Sırtüstü Bacak** | 1 | 0 | 29 | 0 | 0 | 0 | 0 | 0 |
| **Havlu** | 1 | 0 | 0 | 29 | 0 | 0 | 0 | 0 |
| **Düz Sarkaç** | 0 | 0 | 0 | 0 | 28 | 2 | 0 | 0 |
| **Dairesel Sarkaç** | 0 | 0 | 0 | 0 | 1 | 29 | 0 | 0 |
| **Sandalye** | 0 | 0 | 0 | 0 | 1 | 0 | 29 | 0 |
| **Topuk** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 29 |

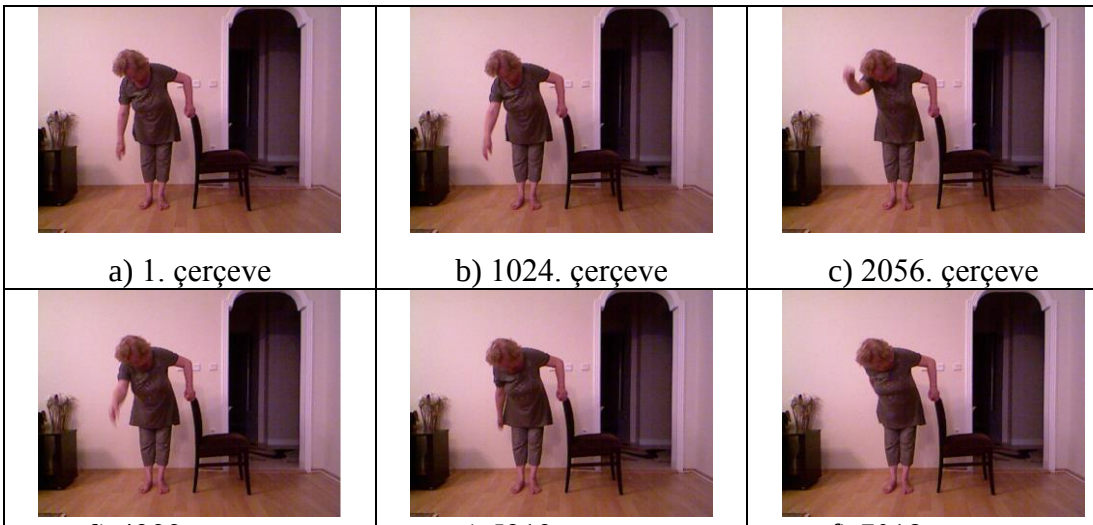
Tablo 9.1 ve Tablo 9.2 verilen sonuçlara ek olarak, her iki durumda da başarılı olarak tanınan dairesel sarkaç egzersizlerinden birinden alınan sıralı çerçeveler Şekil 9.5'da verilmiştir. Şekil 9.6'da ise sadece renk hareketli görüntülerinin kullanılması ile sistem tarafından yanlış egzersiz tipi olarak tanınan fakar hem renk hem de derinlik hareketli görüntülerinin kullanılması ile başarılı bir biçimde tanınan düz sarkaç egzersizinden sıralı çerçeveler verilmiştir.

a)

b)

c)

Şekil 9.5: Derinlik hareketli görüntüleri kullanılmadan başarılı tanınan dairesel sarkaç egzersizinden sıralı çerçeveler. a) Egzersiz seansı başlangıcından alınan örnek çerçeve, b) seans ortasından alınan örnek çerçeve, c) seans sonundan alınan örnek çerçeve.



| | | |
|---|---|---|
| a) 1. çerçeve | b) 1024. çerçeve | c) 2056. çerçeve |
| d) 4399. çerçeve | e) 5213. çerçeve | f) 7018. çerçeve |

Şekil 9.6: Hem renk hemde derinlik hareketli görüntüleri kullanılarak başarılı tanınabilen düz sarkaç egzersizinden sıralı çerçeveler. a)-f) egzersiz seansından sonuna kadar geçen sürede alınan örnek çerçeveleri zaman sırasına göre göstermektedir.

Fizik tedavi egzersizlerini analiz eden geliştirilmiş sistem öncelikle renk (256 seviye gri ton) ve derinlik hareketli görüntülerinden oluşan EOFTE veri kümesi üzerinde, tüm farklı öznitelik çıkartma yöntemlerini kapsayacak bir biçimde test edilmiştir. Hareket bilgisini oluşturmak için kullanılan öznitelikleri çıkarmak için Bölüm 5'te anlatılan tez kapsamında tasarlanan rastgele ormanlar yapısıyla geliştirilmiş hareket tabanlı öznitelik çıkarma yöntemi (*ROY_HÖ*); duruş bilgisini oluşturmak için Bölüm 6'da anlatılan tez kapsamında tasarlanan ızgarasal yapı ile duruş tabanlı öznitelik çıkarma yöntemi *(IY_DÖ)* ve Bölüm 2.1.2.1'de anlatılan Cheema arkadaşları [36] tarafından geliştirilen duruş tabanlı öznitelik çıkarma yöntemi; nesne bilgisini oluşturmak için ise de Bölüm 2.1.3.1'de anlatılan Fei-Fei'nin nesne tespit yöntemi [58], Bölüm 2.1.3.2'de anlatılan Viola ve Jones'un nesne tespit yöntemi [60] ve Bölüm 2.1.3.3'de anlatılan Gall ve Lempitsky'nin nesne tespit yöntemi [62] kullanılmıştır. Rastgele ormanlar yapısıyla geliştirilmiş hareket tabanlı öznitelik çıkarma yönteminde en fazla derinlik sayısı 5 belirlenmiş, ormandaki en fazla ağaç sayısı 24 seçilmiş ve her düğümde seçilen özellik sayısı ise toplam özellik sayısının karekökü olarak alınmıştır. Sistemde kullanılan diğer parametreler ise ilgili bölümlerde verilmiştir.

Fizik tedavi egzersizlerini analiz eden geliştirilmiş sistem için 240 farklı egzersiz seansındaki egzersiz tiplerinin tanınması sonucunda elde edilen başarılı tanıma oranları (doğruluk oranı) Tablo 9.3'te liste halinde verilmiştir.

Tablo 9.3: Geliştirilmiş sistemin çeşitli bilgi türleriyle elde ettiği tanıma oranları.

| **Hareket Bilgisi** | **Duruş Bilgisi** | **Nesne Bilgisi** | **Başarılı Tanıma Oranı** |
|---|---|---|---|
| *ROY_HÖ* | [36] | [58] | %98,33 |
| *ROY_HÖ* | *IY_DÖ* | [58] | %98,33 |
| *ROY_HÖ* | [36] | [60] | %97,9 |
| *ROY_HÖ* | [36] | [62] | %97,9 |
| *ROY_HÖ* | *IY_DÖ* | [62] | %97,9 |
| *ROY_HÖ* | *IY_DÖ* | - | %97,5 |
| *ROY_HÖ* | [36] | - | %95,4 |
| *ROY_HÖ* | *IY_DÖ* | - | %95,0 |
| *ROY_HÖ* | [58] | - | %93,33 |
| *ROY_HÖ* | [62] | - | %93,33 |
| *ROY_HÖ* | [60] | - | %92,9 |
| *ROY_HÖ* | - | - | %89,6 |
| - | [36] | [58] | %78,75 |
| - | [36] | [62] | %77,92 |
| - | *IY_DÖ* | [58] | %76,67 |
| - | *IY_DÖ* | [62] | %75,0 |
| - | [36] | [60] | %74,17 |
| - | [36] | - | %72,92 |
| - | *IY_DÖ* | [60] | %72,08 |
| - | *IY_DÖ* | - | %69,58 |

Fizik tedavi egzersizlerini analiz eden geliştirilmiş sistem sadece *ROY_HÖ* ile elde edilen hareket örüntüleri ile tanıma yaptığında %89,6 başarı oranına ulaşmış, sadece *IY_DÖ* ile elde edilen duruş tabanlı öznitelikleri kullanarak tanıma yaptığında ise %69,58 başarı oranına ulaşmış, *IY_DÖ* yerine Cheema arkadaşları [36] duruş tabanlı öznitelikleri kullandığında ise başarı oranı %72,92'ye ulaşmıştır. Bu sonuç ışığında hareket tabanlı özniteliklerin duruş tabanlı özniteliklere oranla egzersizleri

(dolayısıyla eylemleri) tanımada daha ayırt edici olduğu gözlemlenmiştir. Bir diğer gözlem ise Cheema arkadaşları [36] yönteminin *IY_DÖ* yöntemine oranla daha başarılı olduğudur. Ayrıca hareket bilgisi ve duruş bilgisinin ortak olarak kullanılması durumunda hareket ve nesne bilgisinin ortak olarak kullanımına oranla daha başarılı sonuçlar elde ettiği gözlemlenmiştir. Nesne bilgisi tek olarak egzersizleri ayırt edemediğinden, bu ikili karşılaştırma göz önüne alındığından egzersiz tanımada duruş bilgisinin nesne bilgisinden daha ayırt edici olduğu söylenebilir. Son olarak üç farklı bilgi türünün ortak kullanılmasıyla elde edilen sonuçlar incelendiğinde, en yüksek başarılı tanıma oranlarının bu üçlü birleşim ile elde edildiği gözlemlenmiştir. Bu gözlem ile tez genelinde ortaya konulan farklı bilgi kaynaklarından yararlanılarak daha başarılı egzersiz (eylem) tanıma yapılabileceği savı desteklenmiştir.

Tablo 9.3'te en başarılı tanıma oranları incelendiğinde %98,33 doğruluk oranı iki defa karşımıza çıkmıştır. Bunlarda kullanılan üç bilgi türünün sadece bir tanesinde (duruş tabanlı öznitelikler) farklılık olmasına rağmen aynı başarılı tanıma oranı elde edilmiştir. Bu sonuç bizlere üç farklı bilgi türü kullanıldığında tez kapsamında önerilen *IY_DÖ* yönteminin literatürdeki Cheema arkadaşları [36] duruş tabanlı öznitelik çıkarma yöntemi kadar başarılı olabildiğini göstermiştir.

Fizik tedavi egzersizlerini analiz eden geliştirilmiş sistemde EOFTE veri kümesinde sadece renk hareketli görüntüleri ve de hem renk hem de derinlik hareketli görüntüleri kullanılarak test edilmiştir. Buna ek olarak *ROY_HÖ* yöntemini test etmek için literatürden Bobick ve Davis'in [14] hareket özniteliklerini çıkartma yöntemi ile *ROY_HÖ* değiştirilerek kullanılmıştır. Elde edilen sonuçlar Tablo 9.4'te verilmiştir.

Tablo 9.4 incelendiğinde EOFTE veri kümesindeki 240 farklı egzersiz seansındaki egzersiz tiplerinde 220 tanesi sadece renk görüntüleri kullanılarak fizik tedavi egzersizlerini analiz eden geliştirilmiş sistem ile % 91,7 doğruluk oranıyla tanınmaktadır. Bu doğruluk oranı hem renk hem de derinlik hareketli görüntüleri birlikte kullanıldığında ise %98,3'ye yükselmektedir. Elde edilen iki sonuç incelendiğinde derinlik bilgisi ile başarılı eylem tanıma oranında %6,6'lık artış gözlemlenmektedir. Aynı artış temel sistemde de gözlendiği dikkate alındığında derinlik bilgisi ile kabaca renk ve zaman bilgileriyle birlikte 3 boyutlu bilgi kullanılmasının egzersiz (eylem) tanıma probleminde çok yarar sağladığı söylenebilir. Derinlik bilgisi kullanılmadığı durumda ortalama keskinlik oranı

%92,08 iken derinlik bilgisi kullanıldığı bu oran %98,36'ya ulaşmaktadır. Egzersiz tipine göre elde edilen sonuçlar ise Tablo 9.5 ve Tablo 9.6'da karışıklık matrisi olarak verilmiştir. Ayrıca tez kapsamında geliştirilen *ROY_HÖ* ile elde edilen hareket özniteliklerinin Bobick ve Davis'in [14] hareket öznitelikleriyle diğer iki bilgi türü aynı kalmak şartıyla karşılaştırıldığında, önerilen *ROY_HÖ* yönteminin daha başarılı sonuçlar verdiği gözlemlenmiştir.

Tablo 9.4: Geliştirilmiş sistemin farklı bilgi kaynakları ve veri türleriyle tanıma oranları.

| **Kullanılan Bilgi Türleri (Hareket, Duruş, Nesne)** | **Egzersiz Tanıma Başarı (Doğruluk) Oranı** |
|---|---|
| *ROY_HÖ* + [36] + [58], EOFTE sadece renk | %91,7 |
| *ROY_HÖ* + [36] + [58], EOFTE renk ve derinlik | %98,33 |
| [14] + [36] + [58], EOFTE sadece renk | %82,5 |
| [14]+ [36] + [58], EOFTE renk ve derinlik | %83,33 |

Fizik tedavi egzersizlerini analiz eden geliştirilmiş sistem ile % 91,7 başarı oranıyla sadece renk hareketli görüntülerini kullanarak elde etmiştir. Bu sonuç bize derinlik bilgisi kullanılmayan ortamlarda bile önerilen sistemin başarıyla çalışabileceği göstermektedir.

Fizik tedavi egzersiz seanslarında tekrar sayısını bulmak için tez kapsamında önerilen yöntemin başarısını ölçmek için iki farklı senaryo hazırlanmıştır. Bölüm 9.3'te anlatıldığı üzere tekrar sayısını tahmin eden yöntem o seanstaki egzersiz tipinin doğru olarak belirlenmiş olmasına ve kullanılan hareket ve duruş bilgilerinin ayırt edici olmasına bağlıdır. İlk senaryoda EOFTE veri kümesinde 240 egzersiz seansındaki egzersiz tiplerinin tamamen doğru olarak etiketlendiği varsayılarak tekrar sayısının ne kadar başarılı tahmin edildiği tespit edilmiştir. Bu durumda 220 seanstaki tekrar sayısı doğru tahmin edilerek %91,67 doğruluk oranı elde edilmiştir. Son senaryoda ise fizik tedavi egzersizlerini analiz eden geliştirilmiş sistemin tanıması sonucunda elde edilen egzersiz tip etiketleri kullanılmıştır. Son durumda sistemin tanıma sonucuyla birlikte tekrar sayısındaki doğruluk oranı da %87,5'e düşmüştür. Fakat bu oran bile tekrar sayılarının başarılı bir şekilde tahmin edildiği göstermiştir.

Tablo 9.5: Geliştirilmiş sistemin eylem tanıma karışıklık matrisi (sadece renk).

| ***Egzersiz Tipleri*** | **Sopa** | **Çapraz Sopa** | **Sırtüstü Bacak** | **Havlu** | **Düz Sarkaç** | **Dairesel Sarkaç** | **Sandalye** | **Topuk** |
|---|---|---|---|---|---|---|---|---|
| **Sopa** | 29 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Çapraz Sopa** | 2 | 28 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Sırtüstü Bacak** | 1 | 0 | 28 | 0 | 0 | 0 | 0 | 1 |
| **Havlu** | 1 | 0 | 0 | 29 | 0 | 0 | 0 | 0 |
| **Düz Sarkaç** | 0 | 0 | 0 | 0 | 25 | 5 | 0 | 0 |
| **Dairesel Sarkaç** | 0 | 0 | 0 | 0 | 3 | 27 | 0 | 0 |
| **Sandalye** | 0 | 0 | 0 | 0 | 2 | 1 | 27 | 0 |
| **Topuk** | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 27 |

Tablo 9.6: Geliştirilmiş sistemin eylem tanıma karışıklık matrisi (renk ve derinlik).

| ***Egzersiz Tipleri*** | **Sopa** | **Çapraz Sopa** | **Sırtüstü Bacak** | **Havlu** | **Düz Sarkaç** | **Dairesel Sarkaç** | **Sandalye** | **Topuk** |
|---|---|---|---|---|---|---|---|---|
| **Sopa** | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Çapraz Sopa** | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Sırtüstü Bacak** | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 0 |
| **Havlu** | 1 | 0 | 0 | 29 | 0 | 0 | 0 | 0 |
| **Düz Sarkaç** | 0 | 0 | 0 | 0 | 29 | 1 | 0 | 0 |
| **Dairesel Sarkaç** | 0 | 0 | 0 | 0 | 1 | 29 | 0 | 0 |
| **Sandalye** | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 0 |
| **Topuk** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 29 |

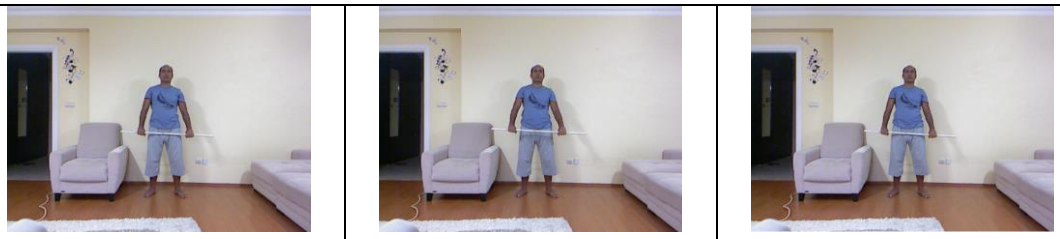
Egzersiz tekrar sayını tahmin eden yöntemin her iki senaryo içinde en çok hatalı tahminde bulunduğu egzersiz tipleri havlu ve dairesel sarkaç egzersizleridir. Havlu egzersizinde etkileşimde bulunan nesne olan havlunun görece ufak kalması, egzersizin başlangıç ve bitiş noktaları arasındaki çeşitli yanlış tespit hatayı arttırmıştır. Dairesel sarkaç egzersizinde ki tekrar sayılarının tahminindeki hatada güvenirlik değerleri-$\tau$ eğrisindeki sık dalgalanmalardan kaynaklanmaktadır. Son senaryo için elde edilen tahmin değerleri egzersiz tiplerine göre analiz edilerek Tablo

9.7'de verilmiştir. Bu tabloda her bir egzersiz tipi için EOFTE veri kümesindeki ortalama tekrar sayısıyla birlikte önerilen tekrar sayısını tahmin etme yönteminin ortalama hatası verilmiştir.

Egzersiz tekrar sayısını tahmin eden yöntem, üç tekrarı bulunan EOFTE veri kümesinden alınan bir sopa egzersizi hareketli görüntüsünde elde ettiği başarılı tahminin detaylı analizi yapılmıştır. Bu analiz kapsamında tekrar sayısını bulmakta kullanılan eğride tespit edilen üç tepe noktasına karşılık düşen çerçeveler Şekil 9.7'de verilmiştir. Tespit edilen bu tepe noktalarında egzersizin setindeki tekrarların tamamlanmış olması dikkat çekicidir. Bu açıdan sistemin başarılı bir biçimde istenilen hedefe ulaştığı söylenebilir.

Tablo 9.7: Egzersiz tekrar sayısının tahmin edilmesiyle ilgili sonuçlar.

| **Egzersiz Tipi** | **Ortalama Tekrar Sayısı** | **Ortalama Tahmin Hatası** |
|---|---|---|
| Sopa | 5,3 | 0,14 |
| Çapraz Sopa | 4,4 | 0,24 |
| Sırtüstü Bacak | 5,2 | 0,12 |
| Havlu | 4,6 | 1,28 |
| Düz Sarkaç | 5,4 | 0,43 |
| Dairesel Sarkaç | 4,6 | 1,46 |
| Sandalye | 4,8 | 0,72 |
| Topuk | 6,5 | 0,60 |
| *Ortalama* | *5,1* | *0,62* |



a) b) c)

Şekil 9.7: Tekrar sayısı üç olan bir sopa egzersizinde tahmin edilen tekrar başlangıç/bitiş çerçeveleri. a) birinci tekrar, b) ikinci tekrar, c) üçüncü tekrar başlangıç/bitiş çerçevelerini göstermektedir.

# 10. SONUÇLAR ve YORUMLAR

Hareketli görüntülerde insan eylemlerinin tanınması problemi bilgisayarla görme alanında güncel bir konudur. Konu kapsam olarak en basit web kameraları aracılığıyla sadece insan yüzlerindeki mimiklerin, el ifadelerinin tanınmasından başlayarak toplumsal olaylarda insan gruplarının eylemlerinin analizine kadar genişleyebilen bir yelpaze yayılmaktadır [1]-[4]. Seçilen problemin kapsamına göre kullanılan yöntemler farklılaşmakta ve diğer benzeri problemlere uygulanabilmesi zorlaşmaktadır. Literatürde varolan çeşitli yöntemler ise çeşitli kısıtlamalar altında çalışmakta ve gerçek hayatta (denetimsiz ortamlarda) kullanılması durumunda bir çok zorlukla karşılaşılmaktadır [1]-[3].

Tez çalışmasında hareketli görüntülerdeki insan eylemlerinin tanınması amacıyla temelde hareket ve duruş tabanlı öznitelik çıkarımı yapan özgün yöntemler tasarlanmasıyla birlikte genel olarak eylem tanıyan özgün sistemlerde oluşturulmuştur. Ayrıca, insan eylemlerinin tanınması kapsamında özel bir konu olan "evde yapılan fizik tedavi egzersizlerinin analizi" konusuna bu çerçeve içerisinde yaklaşılarak özgün bir sistem ortaya konulmuştur.

İlk olarak, hareketli görüntülerde eylemlerin tanınmasındaki en önemli bilgi kaynaklarından biri olan hareket bilgisi üzerine yoğunlaşılmıştır. Bu kapsamda hareketli görüntülerdeki ardışık çerçeveler arasındaki farkların aktörün hareketinden kaynakladığı varsayımından yola çıkılarak bu çerçeveler arasındaki kısa dönem hareket örüntüleri yakalanmıştır. Kısa dönem hareket örüntülerini tanımlayabilmek için 3-Boyutlu Haar benzeri filtreler kullanarak öznitelik çıkarımı yapan özgün bir yöntem ortaya konmuştur [10], [11]. Daha sonra kısa dönem hareket örüntülerinden hareketli görüntünün tümü hakkında anlamlı bir gösterim elde edebilmek için istatistiksel yöntemlere başvurularak küresel hareket tabanlı öznitelik tanımlayıcısı oluşturulmuştur [10], [11]. Tez çalışmasının merkezi olan bu yaklaşım ilerleyen dönemler içerisinde daha da zenginleştirilerek rastgele ormanlar yapısıyla geliştirilmiş hareket tabanlı öznitelik çıkarma (*ROY_HÖ*) yöntemi haline getirilmiştir [12]. Parametrik bir yapıda tanımlanan yöntem aracılığıyla çeşitli veri kümelerinde deneyler yapılarak *ROY_HÖ* yöntemi literatürdeki benzer amaç için tasarlanmış yöntemlerle karşılaştırıldığında, onlar kadar başarılı olduğu deneysel olarak kanıtlanmıştır.

İkinci olarak, eylem tanıyan sistemlerin ortak özellikleri incelendiğinde birden fazla bilgi türü (hareket, duruş, nesne, sahne v.b.) kullananların daha başarılı sonuçlar elde ettiği gözlemlenmiştir. Bu doğrultuda, fizik terapi egzersizlerinin analizi probleminde de kullanılması açısından yararlı olduğu düşünülen, ikinci bir bilgi türünü (duruş bilgisi) daha incelenerek geliştirilen sistemlere dahil edilmiştir [12]. Silüet sabit görüntülerde çalışan ızgarasal yapı ile duruş tabanlı özniteliklerin çıkaran (*IY_DÖ*) özgün yöntem geliştirilerek hareketli görüntülerdeki aktörün duruşu hakkında bilgi edinilmiştir.

Üçüncü olarak, 3-boyutlu Haar benzeri filtreler kullanarak hareketli görüntülerde elde edilen özniteliklerin yararını gözlemleyebilmek için farklı bir alan olan ve literatürde "nesne tanımada eylemin tanınmasının rolü" ve/veya "eylem tanımada kullanılan nesnenin tanınmasının rolü" olarak farklı şekillerde ifade edilebilen bu probleme yönelinmiştir [10]. Bunun için hareket ve nesne bilgisini karşılıklı olarak etkin bir şekilde yorumlayabilecek Bayes ağı merkezli bir sistem oluşturulmuştur. Elde edilen başarılı eylem/nesne tanıma oranlarına ek olarak tasarlanılan sistemin, insan vücut bölgelerine göre bölümleme yapmaması, aktörün elini takip etmemesi, zaman ekseninde herhangi bir normalizasyona gerek duymaması açısından literatürde varolan örneklere oranla daha az kısıtlama ile çalışabildiği gözlemlenmiştir [10].

Dördüncü olarak, 3-boyutlu Haar benzeri filtreler kullanarak hareketli görüntülerde elde edilen özniteliklerin, insan eylemlerini tanıma alanında başarısını gözlemleyebilmek için özgün bir sistem tanımlanmıştır. Bu sistemde hareket ve duruş tabanlı bilgi türlerini harmanlanarak eylem tanıma yapan bir yapı oluşturulmuştur. Bu yapı içerisinde duruş bilgisini elde edebilmek için Cheema arkadaşları [36]'un kontur tabanlı (ing. contour) yaklaşımı kullanılmıştır. Duruş ve hareket bilgisini tek bir çatı altında birleştirip hareketli görüntülerde yapılan eylemleri tanıyabilmek için sınıflandırıcı olarak rastgele ormanlar yapısı kullanılmıştır. Deneysel sonuçlar incelendiğinde literatürde varolan sistemlerin Weizmann veri kümesinde elde ettikleri doğru sınıflandırma oranının %96,66 ile %86,6 değiştiği, KTH veri kümesinde ise bu oranın %94,33-%82,66 aralığında olduğu tespit edilmiştir. Önerilen sistem ise Weizmann veri kümesinde %94,44, KTH veri kümesinde ise %92,99 oranında başarılı sınıflandırma yapmasıyla literatürdeki yerini almıştır. Bu çalışma ile ayrıca farklı bilgi türlerinin etkin olarak

birleştirilmesiyle yapılan eylem tanımadan tek bilgi türü kullanan yaklaşımlara oranlara daha iyi sonuçlar elde ettiği gözlemlenmiştir [11].

Beşinci olarak, tez çalışmasının özgün konusu olan "evde yapılan fizik tedavi egzersizlerinin analizi" hakkında yapılan araştırma sonucunda literatürde çok az sayıda örneğinin olduğu ve bunların genellikle özel rehabitilasyon merkezlerinde özel algılayıcılar v.b. ekipmanlar kullanılarak yapılan maliyetli sistemler olduğu görüldü. Her hastanın olanaklarının (zamansal, maddi, yaş durumu) bunlara yetmeyeceği göz önüne alındığında geliştirilen düşük maliyetli bir sistem ile egzersizlerin temel analiz bilgilerinin ilgili fizyoterapi uzmanına iletilmesinin hastanın iyileşme sürecini olumlu yönde etkileyeceği düşünülmüştür. Bunlara ek olarak fizik tedavi seanslarının hareketli görüntülerini barındıran hiçbir veri kümesine denk gelinmemiştir. Öncelikle ilk iş olarak veri kümesi için ilgili sağlık kuruluşlarına başvurularak sıkça karşılan ve evde yapılması için herhangi bir fazladan nesne gerektirmeyen (her evde bulunabilecek nesneler barındıran egzersizler) fizik tedavi egzersizleri seçildi. MS Kinect algılayıcı kullanarak 6 farklı aktörün 240 adet egzersiz seansının renk ve derinlik hareketli görüntülerinden oluşan EOFTE veri kümesi olarak kayıt edildi [12], [13]. EOFTE veri kümesi araştırmacılara açık hale getirildi.

Son olarak, fizik terapi egzersizleri birer eylem gibi yorumlanarak bunları tanıyabilecek, aynı seans içerisindeki egzersiz tekrar sayılarını tahmin edebilecek temel ve geliştirilmiş sistemler oluşturuldu [12], [13]. Bu sistemler fizik tedavi egzersizlerindeki üç ana bilgi türünü (hareket, duruş, etkileşimde bulunulan nesne) en etkin bir şekilde kullanmak üzere tasarlanmışlardır. Bayes ağı ile modellenen bir yapı içerisinde, üç farklı bilgi türünün öznitelikleri ve dolayısıyla bu öznitelikleri değerlendirerek bilgi türleri hakkında olasılık tabanlı sonuçlar üreten makine öğrenmesi tabanlı sınıflandırıcılar kullanarak, egzersiz seansında tekrar edilen egzersiz tipi başarılı bir şekilde tanınmıştır. Bu tanımayı takiben özgün bir yaklaşım ile aynı seans içerisindeki egzersizin tekrar sayısı elde edilmiştir. Geliştirilen sistemlerde kullanılan hareket ve duruş bilgilerini elde etmek için, tez kapsamında geliştirilen ilgili öznitelik çıkarma yöntemlerin kullanılmasıyla daha başarılı sonuçlar elde edildiği gözlemlenmiştir. Ayrıca tasarlanılan bu sistemlerin sadece renk hareketli görüntüleriyle çalışabilmesi ve %90'dan yüksek bir başarı oranı yakalamasının yanında renk zaman ve derinlik bilgisini (3-boyutlu bilgi) de birlikte kullanıldığında daha yüksek başarı oranlarına eriştiği gözlemlenmiştir [12].

# KAYNAKLAR


[1] Poppe R., (2010), “A Survey on Vision-based Human Action Recognition”, Image and Vision Computing, 28 (6), 976-990.

[2] Wang L., Hu W. H., Tan T. N., (2003), “Recent developments in human motion analysis”, Pattern Recognition, 36 (3), 585-601.

[3] Moeslund B. T., Hilton A., Kruger V., (2006), “A survey of advances in vision-based human motion capture and analysis”, Computer Vision and Image Understanding, 104 (2), 90-126.

[4] Gavrilla D., Davis L. S., (1996), “Tracking humans in action: A 3D model-based approach”, Proceedings of the Advanced Research Projects Agency Image Understanding Workshop, 737-746, Palm Springs, California, USA, 12-15 February.

[5] Zhu G., Xu C., Gao W., Huang Q., (2006), “Action recognition in broadcast tennis video using optical flow and support vector machine”, Computer Vision in Human-Computer Interaction, 89-98, Graz, Austria, 13-16 May.

[6] Blank M., Gorelick L., Shechtman E., Irani M., Basri R., (2005), “Actions as Space-Time Shapes”, The Tenth IEEE International Conference on Computer Vision, 1395-1402, Beijing, China, 17-20 October.

[7] Abbate S., Avvenuti M., Corsini P., Vecchio A., (2010), “Wireless Sensor Networks: Application-Centric Design”, 1st Edition, InTech Incorporation.

[8] Praveen K., Partvathi R., Venkata P., (2005), “Fundamentals of Physiotherapy”, 1st Edition, Jaypee Brothers Medical Publishers Limited Company.

[9] Zhou H., Hu H., (2008), “Human motion tracking for rehabilitation—A survey”, Biomedical Signal Processing and Control, 3 (1), 1-18.

[10] Ar I., Akgul Y.S., (2012), “A Framework for Combined Recognition of Actions and Objects”, International Conference on Computer Vision and Graphics, Lecture Notes in Computer Science, 7594, 264-271, Warsaw, Poland, 24-26 September.

[11] Ar I., Akgul Y. S., (2013), “Action recognition using random forest prediction with combined pose-based and motion-based features”, 8th International Conference on Electrical and Electronics Engineering, 315-319, Bursa, Turkey, 28-30 November.

[12] Ar I., Akgul Y. S., (2014), “A Computerized Recognition System for the Home-Based Physiotherapy Exercises Using an RGBD Camera”, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22 (6), 1160-1171.

[13] Ar I., Akgul Y. S., (2013), “A Monitoring System for Home-Based Physiotherapy Exercises”, 28th International Symposium on Computer and Information Sciences, Computer and Information Sciences III, 487-494, Paris, France, 28-29 October.

[14] Bobick F. A., Davis J. W., (2001), “The recognition of human movement using temporal templates”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (3), 257-267.

[15] Wang Y., Huang K., Tan T., (2007), “Human activity recognition based on R transform”, Proceedings of the Workshop on Visual Surveillance (VS) at the Conference on Computer Vision and Pattern Recognition, 1-8, Minneapolis, Minnesota, USA, 18-23 June.

[16] Souvenir R., Babbs J., (2008), “Learning the viewpoint manifold for action recognition”, Proceedings of the Conference on Computer Vision and Pattern Recognition, 1-7, Anchorage, Alaska, USA, 24-26 June.

[17] Efros A. A., Berd A. C., Mori G., Malik J., (2003), “Recognizing action at a distance”, Proceedings of the International Conference on Computer Vision, 726-733, Nice, France, 13-16 October.

[18] Danafar S., Gheissari N., (2007), “Action recognition for surveillance applications using optic flow and SVM”, Proceedings of the Asian Conference on Computer Vision, Lecture Notes in Computer Science, 457-466, Tokyo, Japan, 18-22 November.

[19] Jhuang H., Serre T., Wolf L., Poggio T., (2007), “A biologically inspired system for action recognition”, International Conference on Computer Vision, 1-8, Rio de Janeiro, Brazil, 14-20 October.

[20] Wiskott L., Sejnowski T., (2002), “Slow Feature Analysis: Unsupervised Learning of Invariances,” Neural Computation, 14 (4), 715-770.

[21] Nowozin S., Bakır G., Tsuda K., (2007), “Discriminative subsequence mining for action classification”, Proceedings of the International Conference on Computer Vision, 1-8, Rio de Janeiro, Brazil, 14-20 October.

[22] Klaser A., Marszalek M., Schmid C., Everingham M., Needham C., Fraile R., (2008), “A spatio-temporal descriptor based on 3d-gradients”, 19th British Machine Vision Conference, 1-10, Leeds, United Kingdom, 1-4 September.

[23] Laptev I., Marszalek M., Schmid C., Rozenfeld B., (2008), “Learning realistic human actions from movies”, Proceedings of the Conference on Computer Vision and Pattern Recognition, 8-16, Anchorage, Alaska, USA, 24-26 June.

[24] Hu M. K., (1962), “Visual pattern recognition by moment invariants”, Ireland Transactions on Information Theory, 8 (2), 179-187.

[25] Davis J. W., Bobick A. F., (1998), “Virtual PAT: a virtual personal aerobics trainer”, Workshop on perceptual user interfaces, 13-18, San Francisco, California, USA, 5-6 November.

[26] Ahad M. A. R., Tan J. K., Kim H., Ishikawa S., (2012), “Motion history image: its variants and applications”, Machine Vision and Applications, 23 (2), 255-281.

[27] Deans S. R., (2007), “The Radon Transform and Some of Its Applications”, 1st Edition, Courrier Corporation.

[28] Weinland D., Boyer E., Ronfard R., (2007), “Action recognition from arbitrary views using 3d exemplars”, 11th International Conference on Computer Vision, 1-7, Rio de Janeiro, Brazil, 14-20 October.

[29] Horn B.K., Schunck B.G., (1981), “Determining Optic Flow”, International Society for Optics and Photonics Technical Symposium East, 319–331 Washington, D.C., USA, 21-23 April.

[30] Chen M., Hauptmann A., (2009), “MoSIFT: Recognizing Human Actions in Surveillance Videos”, In Carnegie Mellon University-CS-09-161, Carnegie Mellon University.

[31] Collins M.J., (1996), “A new statistical parser based on bigram lexical dependencies”, 34th Annual Meeting of the Association for Computational Linguistics, 184-191, Santa Cruz, California, USA, 24-27 June.

[32] Yao A., Gall J., Fanelli G., Van Gool L. J., (2011), “Does Human Action Recognition Benefit from Pose Estimation ?”, 22th British Machine Vision Conference, 6-12, Dundee, UK, 29 August – 2 September.

[33] Thurau C., Hlavac V., (2008), “Pose primitive based human action recognition in videos or still images”, IEEE Conference on Computer Vision and Pattern Recognition, 1-8, Anchorage, Alaska, USA, 24-26 June.

[34] Goldenberg R., Kimmel R., Rivlin E., Rudzsky M., (2005), “Behavior classification by eigendecomposition of periodic motions”, Pattern Recognition, 38 (7), 1033-1043.

[35] Lu W. L., Little J. J., (2006), “Simultaneous tracking and action recognition using the pca-hog descriptor”, 3rd Canadian Conference on Computer and Robot Vision, 1-6, Quebec, Canada, 7-9 June.

[36] Cheema S., Eweiwi A., Thurau C., Bauckhage C., (2011), “Action recognition by learning discriminative key poses”, International Conference on Computer Vision Workshops, 1302–1309, Barcelona, Spain, 6-13 November.

[37] Wang C., Wang Y., Yuille A. L., (2013), “An approach to pose-based action recognition”, Computer Vision and Pattern Recognition, 915-922, Portland, Oregon, USA, 25-27 June.

[38] Gorelick L., Blank M., Shechtman E., Irani M., Basri R., (2007), “Actions as space-time shapes”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 (12), 2247-2253.

[39] Ikizler N., Duygulu P., (2009), “Histogram of oriented rectangles: A new pose descriptor for human action recognition”, Image and Vision Computing, 27 (10), 1515-1526.

[40] Xu R., Agarwal P., Kumar S., Krovi V. N., Corso J. J., (2012), “Combining skeletal pose with local motion for human activity recognition”, Articulated Motion and Deformable Objects, 7378(1), 114-123.

[41] Sung J., Ponce C., Selman B., Saxena A., (2012), “Human Activity Detection from RGBD Images”, IEEE International Conference on Robotics and Automation, 842-849, Minnesota, USA, 14-18 May.

[42] Rosales R., Sclaroff S., (2000), “Inferring body pose without tracking body parts”, Computer Vision and Pattern Recognition, 721-727, Hilton Head, South Carolina, USA, 13-15 June.

[43] Baysal S., Kurt M. C., Duygulu P, (2010), “Recognizing human actions using key poses”, 20th International Conference on Pattern Recognition, 1727-1730, Istanbul, Turkey, 23-26 August.

[44] Ullman S., Humphreys G. W., (1996), “High-level vision: Object recognition and visual cognition”, 2nd Edition, Massachusetts Institute of Technology Press.

[45] Rivlin E., Dickinson S. J., Rosenfeld A., (1995), “Recognition by functional parts”, Computer Vision and Image Understanding, 62 (2), 164-176.

[46] Stark L., Bowyer K., (1996), “Generic object recognition using form and function”, Machine Perception and Artificial Intelligence, 10 (1), 1996.

[47] Stark M., Lies P., Zillich M., Wyatt J., Schiele B., (2008), “Functional object class detection based on learned affordance cues”, 6th International Conference on Computer Vision Systems, 435-444, Santorini, Greece, 12-15 May.

[48] Moore D. J., Essa I. A., Hayes M. H., (1999), “Exploiting human actions and object context for recognition tasks”, The Proceedings of the 7th IEEE International Conference on Computer Vision, 80-86, Kerkyra, Greece, 20-27 September.

[49] Billard A., Calinon S., Dillmann R., Schaal S., (2008), “Handbook of robotics”, 1st Edition, Springer-Verlag Berlin Heidelberg.

[50] Kjellström H., Romero J., Kragić D., (2011), “Visual object-action recognition: Inferring object affordances from human demonstration”, Computer Vision and Image Understanding, 115 (1), 81-90.

[51] Peursum P., West G., Venkatesh S., (2005), “Combining image regions and human activity for indirect object recognition in indoor wide-angle views”, 10th IEEE International Conference on Computer Vision, 82-89, Beijing, China, 17-20 October.

[52] Wu J., Osuntogun A., Choudhury T., Philipose M., Rehg J. M., (2007), “A scalable approach to activity recognition based on object use”, 11th IEEE International Conference on Computer Vision, 1-8, Rio de Janeiro, Brazil, 14-20 October.

[53] Filipovych R., Ribeiro E., (2008), “Recognizing primitive interactions by exploring actor-object states”, IEEE Conference on Computer Vision and Pattern Recognition, 1-7, Anchorage, Alaska, USA, 24-26 June.

[54] Gupta A., Kembhavi A., Davis L. S., (2009), “Observing human-object interactions: Using spatial and functional compatibility for recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (10), 1775-1789.

[55] Parry R. H., (2004), “The interactional management of patients’ physical incompetence: a conversation analytic study of physiotherapy interactions”, Sociology of health & illness, 26 (7), 976-1007.

[56] Witkos J., Blonska-Fajfrowska B., Niebroj L., Gazdzik T., (2012), “Physiotherapy: Pressing Issues of Everyday practice”, 1st Edition, Lulu Enterprises Incorporated.

[57] Seymour J. C., Kerr K. M., (1996), “Community based physiotherapy in the Trent region: a survey”, Physiotherapy, 82 (9), 514-520.

[58] Fei-Fei L., (2007), “Bag of words models: Recognizing and learning object categories,” IEEE Conference on Computer Vision and Pattern Recognition, 1-8, Minneapolis, Minnesota, USA, 18-23 June.

[59] Lowe D. G., (1999), “Object recognition from local scale-invariant features”, 7th IEEE international conference on computer vision, 1150-1157, Ft. Collins, Colorado, USA, 23-25 June.

[60] Viola P., Jones M., (2001), “Rapid object detection using a boosted cascade of simple features”, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 511-518, Kauai, Hawaii, USA, 8-14 December.

[61] Gaoa C., Sang N., Tanga Q., (2010), “On Selection and Combination of Weak Learners in AdaBoost”, Pattern Recognition Letters, 31 (9), 991-1001.

[62] Gall J., Lempitsky V., (2009), “Class-specific hough forests for object detection”, Conference on Computer Vision and Pattern Recognition, 1022–1029, Fontainebleau Resort, Miami Beach, Florida, USA, 20-25 June.

[63] Shapiro L., George S., (2001), “Computer Vision”, 1st Edition, Prentice-Hall.

[64] Ballard D.H., (1981), “Generalizing the Hough Transform to Detect Arbitrary Shapes”, Pattern Recognition, 13 (2), 111-122.

[65] Kovashka A., Grauman K., (2010), “Learning a Hierarchy of Discriminative Space-Time Neighborhood Features for Human Action Recognition”, 23th IEEE Conference on Computer Vision and Pattern Recognition, 2046-2053, San Francisco, California, USA, 13-18 June.

[66] Bregonzio M., Xiang T., Gong S., (2012), “Fusing Appearance and Distribution Information of Interest Points for Action Recognition”, Pattern Recognition, 45 (3), 1220-1234.

[67] Zhang Z., Tao D., (2012), “Slow Feature Analysis for Human Action Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (3), 436-450.

[68] Berkes P., Wiskott L., (2005), “Slow Feature Analysis Yields a Rich Repertoire of Complex Cell Properties,” Journal of Vision, 5 (6), 579-602.

[69] Ikizler-Cinbis N., Sclaroff S., (2010), “Object, Scene and Actions: Combining Multiple Features for Human Action Recognition”, 11th European Conference on Computer Vision, 494-507, Hersonissos, Heraklion, Crete, Greece, 5-11 September.

[70] Ballan L., Bertini M., Bimbo A., Seidenari L., Serra G., (2009), “Recognizing Human Actions by Fusing Spatio-Temporal Appearance and Motion Descriptors”, 16th International Conference on Image Processing, 3569-3572, Cairo, Egypt, 7-10 November.

[71] Dollár P., Rabaud V., Cottrell G., Belongie S., (2005),"Behavior recognition via sparse spatio-temporal features”, 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 65-72, Beijing, China, 15-16 October.

[72] Dhillon P.S., Nowozin S., Lampert C.H., (2009), “Combining Appearance and Motion for Human Action Classification in Videos”, Computer Vision and Pattern Recognition Workshop, 22-29, Piscataway, New Jersey, USA, 20-25 June.

[73] Okuma K., Taleghani A., De Freitas N., Little J. J., Lowe D. G., (2004), “A boosted particle filter: Multitarget detection and tracking”, 8th European Conference on Computer Vision, 28-39, Slovansky Ostrov, Prague, Czech Republic, 11-14 May.

[74] Vermaak J., Doucet A., Pérez P., (2003), “Maintaining multimodality through mixture tracking”, 9th IEEE International Conference on Computer Vision, 1110-1116, Nice, France, 13-16 October.

[75] Thurau C., (2007), “Behavior Histograms for Action Recognition and Human Detection”, The 2nd Conference on Human Motion: Understanding, Modeling, Capture and Animation, 299-312, Rio de Janeiro, Brazil, 14-20 October.

[76] Broder A. Z., Glassman S. C., Manasse M. S., Zweig G., (1997), “Syntactic clustering of the web”, Computer Networks and ISDN Systems, 29 (8), 1157-1166.

[77] Shotton J., Fitzgibbon A., Cook M., Sharp T., Finocchio M., Moore R., Kipman A., Blake A., (2011), “Real-Time Human Pose Recognition in Parts from Single Depth Images”, Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, 1297-1304, Colorado Springs, USA, 20-25 June.

[78] Vliet P. M., Wulf G., (2006), “Extrinsic feedback for motor learning after stroke: what is the evidence?”, Disability and Rehabilitation, 28 (13-14), 831-840.

[79] Rumelhart D. E., Hinton G. E., Williams R. J., (1986), “Learning Representations by Back-propagating Errors”, Nature, 323 (1), 533-536.

[80] Domingo A., Ferris D. P., (2009), “Effects of physical guidance on short-term learning of walking on a narrow beam”, Gait Posture, 30 (4), 464-468.

[81] Koller-Hodac A., Leonardo D., Walpen S., Felder D., (2010), “A novel robotic device for knee rehabilitation improved physical therapy through automated process”, 3rd IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics, 820-824, Tokyo, Japan, 26-29 September.

[82] Nater F., Grabner H., Van Gool L., (2010), “Exploiting Simple Hierarchies for Unsupervised Human Behavior Analysis”, 23th IEEE Conference on Computer Vision and Pattern Recognition, 2014-2021, San Francisco, California, USA, 13-18 June.

[83] Soutschek S., Kornhuber J., Maier A., Bauer S., Kugler P., Hornegger J., Bebenek M., Steckmann S., Von Stengel S., Kemmler W., (2010), “Measurement of angles in time-of-flight data for the automatic supervision of training exercises”, 4th International Conference on Pervasive Computing Technologies for Healthcare, 1–4, Munchen, Germany, 22-25 March.

[84] Fitzgerald D., Foody J., Kelly D., Ward T., Markham C., McDonald J., Caulfield B., (2007), “Development of a wearable motion capture suit and virtual reality biofeedback system for the instruction and analysis of sports rehabilitation”, 29th international conference of the IEEE Engineering in Medice and Biology Society, 4870-4874, Lyon, France, 23-26 August.

[85] Jung Y., Kang D., Kim J., (2010), “Upper body motion tracking with inertial sensors”, Proceedings of the 2010 IEEE international conference on robotics and biomimetics, 1746-1751, Tianjin, China, 14-18 December.

[86] Gall J., Fossati A., Gool L. J. V., (2011), “Functional categorization of objects using realtime markerless motion capture”, Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, 1969-1976, Colorado Springs, USA, 20-25 June.

[87] Schuldt C., Laptev I., Caputo B., (2004), “Recognizing human actions: a local SVM approach”, Proceedings of the 17th International Conference on Pattern Recognition (ICPR), 3, 32-36, August.

[88] Web, (2015), https://www.microsoft.com/en-us/kinectforwindows/, (Erişim Tarihi: 05/01/2015)

[89] Xia L., Chen C. C., Aggarwal J. K., (2011), “Human detection using depth information by kinect”, Workshop on Human Activity Understanding from 3D Data in conjunction with CVPR, 15-22, Colorado Springs, USA, 20-25 June.

[90] Papageorgiou C. P., Oren M., Poggio T., (1998), “A general framework for object detection”, Proceedings of the 6th International Conference on Computer Vision, 555-562, Santa Barbara, California, USA, 23-25 June.

[91] Cui X., Liu Y., Shan S., Chen X., Gao W., (2007), “3d haar-like features for pedestrian detection”, IEEE International Conference on Multimedia and Expo, 1263-1266, Beijing, China, 2-5 July.

[92] Donderac R., Doermann D., Davis L., (2009), “Action Recognition Based on Human Movement Characteristics”, International Conference on Motion and Video Computing, 103-110, Utah, USA, 13-15 December.

[93] Viola P., Jones M., Snow D., (2005), “Detecting Pedestrians Using Patterns of Motion and Appearance”, International Journal of Computer Vision, 63 (2), 153-161.

[94] Breiman L., (2001), “Random forests,” Machine Learning, 45 (1), 5–32.

[95] Zhu X., Shaogang G., Chen C. L., (2012), “Comparing Visual Feature Coding for Learning Disjoint Camera Dependencies”, British Machine Vision Conference, 1-12, Guilford, UK, 3-7 September.

[96] Moosmann F., Triggs B., Jurie F., (2007), “Fast discriminative visual codebooks using randomized clustering forests”, Advances in Neural Information Processing Systems, 19 (1), 985-992.

[97] Vens C., Costa F., (2011), “Random forest based feature induction”, 11th IEEE International Conference on Data Mining, 744–753, Vancouver, Canada, 11-14 December.

[98] Lempitsky V., Zisserman A., (2010), “Learning to count objects in images,” Advances in Neural Information Processing Systems, 23 (1), 1324–1332.

[99] Adak B., Önen M., Tekeoğlu I., (1999), “Van ili merkez ilköğretim okullarında skolyoz taraması”, Türkiye Fiziksel Tıp ve Rehabilitasyon Dergisi, 4 (2), 12-26.

[100] Karakuş S., Kılınç F., (2006), “Postür ve sportif performans”, Kastamonu Eğitim Dergisi, 14 (1), 309-322.

[101] Samet H., Tamminen M., (1988), “Efficient component labeling of images of arbitrary dimension represented by linear bintrees”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 10 (4), 579-586.

[102] Dillencourt M. B., Samet H., Tamminen M., (1992), “A general approach to connected-component labeling for arbitrary image representations”, Journal of the ACM (JACM), 39 (2), 253-280.

[103] Marion A., (1991), “An Introduction to Image Processing”, 4th Edition, Chapman and Hall.

[104] Rubinstein Y. D., Hastie T., (1997), “Discriminative vs Informative Learning”, Knowledge Discovery and Data Mining, 5 (1), 49-53.

[105] Jensen F. V., (1996), “An introduction to Bayesian Networks”, 1st Edition, University College London Press.

[106] Pearl J., (1988), “Probabilistic Reasoning in Intelligent Systems”, 1st Edition, Morgan Kaufmann.

[107] Georgii H. O., (1988), “Gibbs measures and phase transitions”, 1st Edition, Walter de Gruyter.

[108] Yilmaz A., Shah M., (2005), “Actions Sketch: A Novel Action Representation”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 984-989, San Diego, California, USA, 20-26 June.

[109] Ikizler N., Cinbis R. G., Duygulu P., (2008), “Human action recognition with line and flow histograms”, 19th IEEE International Conference on Pattern Recognition, 1-4, Tampa, Florida, USA, 8-11 December.

[110] Liu J., Ali S., Shah M., (2008), “Recognizing human actions using multiple features”, IEEE Conference on Computer Vision and Pattern Recognition, 1-8, Anchorage, Alaska, USA, 24-26 June.

[111] Liu J., Kuipers B., Savarese S., (2011), “Recognizing human actions by attributes”, IEEE Conference on Computer Vision and Pattern Recognition, 3337-3344, Colorado Springs, USA, 20-25 June.

[112] Hu Y., Cao L., Lv F., Yan S., Gong Y., Huang T. S., (2009), "Action detection in complex scenes with spatial and temporal ambiguities", 12th IEEE International Conference on Computer Vision, 128-135, Kyoto, Japan, 27 September – 4 October.

[113] B. Ni, Y. Pei, S. Winkler, P. Moulin, (2012), "Kinect for rehabilitation", 6th International Convention on Rehabilitation Engineering and Assistive Technology, 1-2, Tampines, Singapore, 24-26 July.

# ÖZGEÇMİŞ

İlktan Ar 1982 yılında İstanbul'da doğdu. 2000 yılında başladığı Kadir Has Üniversitesi (KHAS), Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü'nü 2004 yılında başarıyla tamamlayarak, aynı yıl yüksek lisans eğitimine Yıldız Teknik Üniversitesi (YTÜ) Bilgisayar Mühendisliği bölümünde başladı. 2007 yılında yüksek lisans eğitimini başarıyla tamamlayarak, aynı yıl doktora eğitimine Gebze Yüksek Teknoloji Enstitüsü (GYTE), Mühendislik ve Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalına başladı. 2004 yılından bu yana Kadir Has Üniversitesi, Mühendiskik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği bölümünde araştırma görevlisi olarak çalışmaktadır.

# EKLER

## Ek A: Tez Çalışması Kapsamında Yapılan Yayınlar

Ar I., Akgul Y.S., (2012), “A Framework for Combined Recognition of Actions and Objects”, International Conference on Computer Vision and Graphics, Lecture Notes in Computer Science, 7594, 264-271, Warsaw, Poland, 24-26 September.

Ar I., Akgul Y. S., (2013), “A Monitoring System for Home-Based Physiotherapy Exercises”, 28th International Symposium on Computer and Information Sciences, Computer and Information Sciences III, 487-494, Paris, France, 28-29 October.

Ar I., Akgul Y. S., (2013), “Action recognition using random forest prediction with combined pose-based and motion-based features”, 8th International Conference on Electrical and Electronics Engineering, 315-319, Bursa, Turkey, 28-30 November.

Ar I., Akgul Y. S., (2014), “A Computerized Recognition System for the Home-Based Physiotherapy Exercises Using an RGBD Camera”, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22 (6), 1160-1171.

## Ek B: İngilizce-Türkçe Sözlük

| A | |
|---|---|
| Akış Histogramı | Histogram of Flow |
| Alçak-geçirgen filtre | Low-pass filter |
| Algoritma | Algorithm |
| Anahtar poz | Key pose |
| Arka plan | Background |
| Arka plan çıkarımı | Background subtraction |
| Art arda sıralı | Cascade |
| Aşağıdan-yukarı | Bottom-up |
| Ayrık | Discrete |
| Akış Histogramı | Histogram of Flow |
| Alçak-geçirgen filtre | Low-pass filter |
| Algoritma | Algorithm |
| Anahtar poz | Key pose |
| **B** | |
| Bayes Ağları | Bayesian Networks |
| Bilgisayarla Görme | Computer Vision |
| Birini dışarda bırakma | Leave-one-out |
| Birleşik olasılık | Joint probability |
| Bulanıklaştırma | Blurring |
| Büyüklük | Magnitude |
| Bayes Ağları | Bayesian Networks |
| Bilgisayarla Görme | Computer Vision |
| Birini dışarda bırakma | Leave-one-out |
| Birleşik olasılık | Joint probability |
| **Ç** | |
| Çekirdek | Kernel |
| Çerçeve | Frame |
| Çoklu Çekirdek Öğrenmesi | Multi-Kernel Learning |
| Çoklu Örnek Öğrenmesi | Multiple Instance Learning |

| | |
|---|---|
| **D** | |
| Dağılım | Distribution |
| Dedektör | Detector |
| Derinlik | Depth |
| Destek | Support |
| Destek Vektör Makinesi | Support Vector Machine |
| Doğrultma | Rectification |
| Doğruluk oranı | Accuracy |
| Dönme | Rotation |
| Dönüşüm | Transform |
| Duruş | Pose |
| Düğüm | Node |
| Düşük seviye | Low-level |
| Düzgün örnekleme | Uniform Sampling |
| **E** | |
| Egzersiz | Exercise |
| Eğitim kümesi | Training set |
| Eğri | Curve |
| Eklemeli Modeller | Additive Models |
| Eksen | Axis |
| En Yakın Komşuluk ara değerlemesi | Nearest Neighbor interpolation |
| En-boy oranı | Aspect ratio |
| Eniyileme | Optimization |
| Eşik seviyesi | Threshold |
| Eylem | Action |
| Egzersiz | Exercise |
| **F** | |
| Fisher Ayırma Analizi | Fisher Discriminant Analysis |
| Fizik Tedavi | Physiotherapy |
| **G** | |
| Geleneksel | Conventional |
| Genelleşmiş Hough Dönüşümü | Generalized Hough Transform |
| Global | Global |

| | |
|---|---|
| Görüntü | Image |
| Görüntü işleme | Image processing |
| Görünüş | Appearance |
| Gözetim | Surveillance |
| Gradyan | Gradient |
| Gradyanlar Histogramı | Histogram of Gradients |
| **H** | |
| Hareket Enerji Görüntüsü | Motion Energy Image |
| Hareket Geçmiş Görüntüsü | Motion History Image |
| Hareketli görüntü | Video |
| Hareket-tabanlı Ölçekle Değişmeyen Özellik Dönüşümü | Motion Scale-Invariant Feature Transform |
| Hız | Velocity |
| **I** | |
| Izgara | Grid |
| **İ** | |
| İlgi noktası | Interest point |
| **K** | |
| Kamera | Camera |
| Kanıt | Evidence |
| Karar | Decision |
| Kararlı | Stable |
| Karışıklık matrisi | Confusion matrix |
| Kelimeler Torbası | Bag of Words |
| Kenar | Edge |
| Kestirici | Estimator |
| Kısıtlama | Constraint |
| Kişisel | Private |
| Klip | Clip |
| Kontrast | Contrast |
| Kontur | Contour |
| K-Ortalamalar kümelemesi | K-Means clustering |
| Koşullu bağımsızlık | Conditional independence |

| | |
|---|---|
| Köşe | Corner |
| Küboid | Cuboid |
| **M** | |
| Makine Öğrenmesi | Machine Learning |
| Mantıksal Regresyon | Logical Regression |
| Medyan | Median |
| **N** | |
| Nesne | Object |
| Nesnelerin birbirlerini kapatması | Occlusion |
| **O** | |
| Optik Akış | Optical Flow |
| Ortalama | Mean |
| Oryantasyon | Orientation |
| **Ö** | |
| Öklid | Eucledian |
| Ölçek | Scale |
| Ölçekleme | Scaling |
| Ön plan | Foreground |
| Örnekleme | Sampling |
| Örüntü | Pattern |
| Özel | Private, Special |
| Öznitelik | Feature |
| Öznitelik çıkarma | Feature extraction |
| Özşekil | Eigenshape |
| **P** | |
| Parlaklık | Brightness |
| **R** | |
| Rastgele Orman | Random Forest |
| **S** | |
| Sepet | Bin |
| Sınıflandırma | Classification |
| Sınıflayıcı | Classifier |
| Silüet | Silhouette |

| | |
|---|---|
| **Ş** | |
| Şablon eşleştirme | Template matching |
| **T** | |
| Tahmin Etme | Prediction |
| Takip etme | Tracking |
| Tanıma | Recognition |
| Tanımlayıcı | Descriptor |
| Tekrar | Repetition |
| Temel Bileşen Analizi | Principal Component Analysis |
| Test kümesi | Testing set |
| **U** | |
| Uzamsal | Spatial |
| **Ü** | |
| Üretken | Generative |
| **V** | |
| Varyans | Variance |
| Veri kümesi | Dataset |
| **Y** | |
| Yaklaşım | Approach |
| Yapay Sinir Ağları | Neural Networks |
| Yukarıdan-aşağı | Top-down |

## Ek C: Türkçe-İngilizce Sözlük

| | |
|---|---|
| **A** | |
| Accuracy | Doğruluk oranı |
| Action | Eylem |
| Additive Models | Eklemeli Modeller |
| Algorithm | Algoritma |
| Appearance | Görünüş |
| Approach | Yaklaşım |
| Aspect ratio | En-boy oranı |
| Axis | Eksen |
| **B** | |
| Background | Arka plan |
| Background subtraction | Arka plan çıkarımı |
| Bag of Words | Kelimeler Torbası |
| Bayesian Networks | Bayes Ağları |
| Bin | Sepet |
| Blurring | Bulanıklaştırma |
| Bottom-up | Aşağıdan-yukarı |
| Brightness | Parlaklık |
| **C** | |
| Camera | Kamera |
| Cascade | Art arda sıralı, kademeli |
| Cell | Hücre |
| Classification | Sınıflandırma |
| Classifier | Sınıflayıcı |
| Clip | Klip |
| Clustering | Kümeleme |
| Computer Vision | Bilgisayarla Görme |
| Conditional independence | Koşullu bağımsızlık |
| Confusion matrix | Karışıklık matrisi |
| Constraint | Kısıtlama |
| Contour | Kontur |

| | |
|---|---|
| Contrast | Kontrast, karşıtlık |
| Conventional | Geleneksel |
| Corner | Köşe |
| Cuboid | Küboid |
| Curve | Eğri |
| **D** | |
| Dataset | Veri kümesi |
| Decision | Karar |
| Depth | Derinlik |
| Descriptor | Tanımlayıcı |
| Detector | Dedektör |
| Discrete | Ayrık |
| Distribution | Dağılım |
| **E** | |
| Edge | Kenar |
| Eigenshape | Özşekil |
| Estimator | Kestirici |
| Eucledian | Öklid |
| Evidence | Kanıt, delil |
| Exercise | Egzersiz |
| **F** | |
| Feature | Öznitelik |
| Feature descriptor | Öznitelik tanımlayıcısı |
| Feature extraction | Öznitelik çıkarma |
| Fisher Discriminant Analysis | Fisher Ayırma Analizi |
| Foreground | Ön plan |
| Frame | Çerçeve |
| **G** | |
| Generalized Hough Transform | Genelleşmiş Hough Dönüşümü |
| Generative | Üretken |
| Global | Küresel, global |
| Gradient | Gradyan |
| Gray-level | Gri seviyesi |

| | |
|---|---|
| Grid | Izgara |
| **H** | |
| Half Wave Rectification | Yarım Dalga Doğrultma |
| High-level | Yüksek seviye |
| Histogram of Flow | Akış Histogramı |
| Histogram of Gradients | Gradyanlar Histogramı |
| **I** | |
| Image | Görüntü |
| Image processing | Görüntü işleme |
| Interest point | İlgi noktası |
| **J** | |
| Joint probability | Birleşik olasılık |
| **K** | |
| Kernel | Çekirdek |
| Key pose | Anahtar poz |
| K-Means clustering | K-Ortalamalar kümelemesi |
| **L** | |
| Leave-one-out | Birini dışarda bırakma |
| Logical Regression | Mantıksal Regresyon |
| Low-level | Düşük seviye |
| Low-pass filter | Alçak-geçirgen filtre |
| **M** | |
| Machine Learning | Makine Öğrenmesi |
| Magnitude | Büyüklük |
| Mean | Ortalama |
| Median | Medyan |
| Motion | Hareket |
| Motion Energy Image | Hareket Enerji Görüntüsü |
| Motion History Image | Hareket Geçmiş Görüntüsü |
| Motion Scale-Invariant Feature Transform | Hareket-tabanlı Ölçekle Değişmeyen Özellik Dönüşümü |
| Multi-Kernel Learning | Çoklu Çekirdek Öğrenmesi |

| | |
|---|---|
| **N** | |
| Nearest Neighbor interpolation | En Yakın Komşuluk ara değerlemesi |
| Neural Networks | Yapay Sinir Ağları |
| **O** | |
| Object | Nesne |
| Occlusion | Nesnelerin birbirlerini kapatması |
| Optical Flow | Optik Akış |
| Optimization | Eniyileme |
| Orientation | Oryantasyon |
| **P** | |
| Pattern | Örüntü |
| Physiotherapy | Fizik tedavi, Fizyoterapi |
| Pose | Duruş |
| Prediction | Tahmin etme |
| Principal Component Analysis | Temel Bileşen Analizi |
| Private | Özel |
| **R** | |
| Random Forest | Rastgele Orman |
| Recognition | Tanıma |
| Rectification | Doğrultma |
| Repetition | Tekrar |
| Rotation | Dönme |
| **S** | |
| Sampling | Örnekleme |
| Scale | Ölçek |
| Scale Invariant Feature Transform | Ölçekle Değişmeyen Özellik Dönüşümü |
| Scaling | Ölçekleme |
| Silhouette | Silüet |
| Skip | Tek ayak üzerinde sekerek ilerleme |
| Spatial | Uzamsal |
| Spatio-Temporal | Uzam-Zamansal |
| Stable | Kararlı |
| Support | Destek |

| | |
|---|---|
| Support Vector Machine | Destek Vektör Makinesi |
| Surveillance | Gözetim |
| **T** | |
| Template matching | Şablon eşleştirme |
| Testing set | Test kümesi |
| Threshold | Eşik seviyesi |
| Time of Flight | Uçuş Zamanı |
| Top-down | Yukarıdan-aşağı |
| Tracking | Takip etme |
| Training set | Eğitim kümesi |
| Transform | Dönüşüm |
| **U** | |
| Uniform Sampling | Düzgün örnekleme |
| **V** | |
| Velocity | Hız |
| Video | Hareketli görüntü |