T.C.
DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

EL HAREKETLERİNİN KINECT İLE KOMUT OLARAK ALGILANMASI

Julius BAMWENDA

YÜKSEK LİSANS TEZİ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI

DİYARBAKIR
Mart - 2018

T.C

DİCLE UNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ

DİYARBAKIR

Julius BAMWENDA tarafından yapılan "El Hareketlerinin Kinect ile Komut Olarak Algılanması" konulu çalışma, jürimiz tarafından Elektrik-Elektronik Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir

Jüri Üyesinin

Ünvanı Adı Soyadı

Başkan: Doç. Dr. Mehmet Siraç ÖZERDEM

Üye: Yrd. Doç. Dr. Muhammet Ali ARSERİM

Üye: Yrd. Doç. Dr. Serhat Berat EFE

Tez Savunma Sınavı Tarihi: 26/01/2018

Yukarıdaki bilgilerin doğruluğunu onaylarım.

.../...../20

Doç. Dr. Sevtap SÜMER EKER

ENSTİTÜ MÜDÜR V.

(MÜHÜR)

TEŞEKKÜR

Bu çalışma, Dicle Üniversitesi DÜBAP MÜHENDİSLİK.17.002 numaralı proje kapsamında desteklenmiştir. Desteklerinden ötürü DÜ Bilimsel Araştırma Projeleri Koordinatörlüğüne teşekkür ederim.

Bu tez çalışması boyunca birikimlerini ve tecrübelerini benden esirgemeyen değerli danışmanım Doç. Dr. Mehmet Siraç ÖZERDEM'e yol göstericiliği adına teşekkürler ederim.

Ayrıca maddi – manevi desteklerini benden esirgemeyen annem, babam ve eşime teşekkürü bir borç bilirim.

Ocak 2018
Julius BAMWENDA

İÇİNDEKİLER

Sayfa

TEŞEKKÜR ..... **I**

İÇİNDEKİLER ..... **II**

ÖZET ..... **III**

ABSTRACT ..... **IV**

ÇİZELGE LİSTESİ ..... **V**

ŞEKİL LİSTESİ ..... **VI**

KISALTMA VE SİMGELER ..... **VIII**

1. GİRİŞ **1**

1.1. Teknolojide Arayüz Gelişimi 1

1.2. El Hareketlerini Tanıma 2

1.3. Microsoft Kinect (MS Kinect) Sensörü 4

2. KAYNAK ÖZETLERİ **15**

3. MATERYAL VE METOT **27**

4. BULGULAR **43**

4.1. Uygulamadaki Donanım ve Yazılım Platformları 43

4.2 Uygulamadaki Veri Tabanı ve Kullanılan Arayüz 43

4.3. SVM Yöntemi ile Elde Edilen Sınıflandırma Sonuçları 44

4.4. YSA Yöntemi ile Elde Edilen Sınıflandırma Sonuçları 46

4.5. kNN Yöntemi ile Elde Edilen Sınıflandırma Sonuçları 51

5. TARTIŞMA VE SONUÇ **53**

6. KYNAKLAR **57**

ÖZGEÇMİŞ 59

ÖZET

EL HAREKETLERİNİN KINECT İLE KOMUT OLARAK ALGILANMASI

YÜKSEK LİSANS TEZİ

Julius BAMWENDA

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ BÖLÜMÜ
FEN BİLİMLERİ ENSTİTÜSÜ
DİCLE ÜNİVERSİTESİ

2018

Bu tez çalışmasında, el hareketlerinin MS Kinect ile komut olarak algılanması amaçlanmıştır. Bu amaç doğrultusunda; işaret dili tanıma, Kinect tabanlı ve Kinect kullanılmadan oluşturulan sistemler ile ilişkili çalışmalar incelenmiştir. Bu süreçte Kinect donanımının avantajları/dezavantajları ve kameralardaki farklılıklar incelenmiştir. Konu ile ilişkili önemli saptamalar ve önerilen yöntem ile elde edilen sonuçlar aşağıda açıklanmıştır.

Tez kapsamında toplam 24 el hareketi içeren imge seti kullanılmıştır. Bu el hareketlerinin 18 tanesi Amerikan İşaret Dili kapsamında yer alan el hareketleridir. Her el hareketine ilişkin 100 imge kayda alınmıştır. Sistemin güvenirliliğini artırmak amacıyla her hareket için oluşturulan 100 imge için üç farklı kişi görev almıştır. Böylelikle toplamda 24000 imge değerlendirilmeye alınmıştır. Her imgeden çevrimdışı (offline) RGB ve derinlik bilgileri kayda alınmış ve HOG yöntemi ile imgelerin öznitelikleri çıkarılmıştır. Oluşturulan öznitelikler ile sistemin eğitimi sağlanmıştır. El hareketlerinin komut olarak algılanmasında hareketlerin sınıflandırılması için üç farklı sınıflandırıcı yöntem uygulanmıştır. Bunlar; SVM, Çok katlı YSA mimarisi ve kNN sınıflandırıcılarıdır. Eğitme işleminden sonra, gerçek zamanlı olarak sistemin testi gerçeklenmiştir. Yapılan denemelerde her komut için 10 el hareketi ile test edilmiş ve doğru/yanlış şeklinde kayda alınmıştır. Elde edilen ortalama başarı performansları; kNN için %92.5, SVM için %95.83 ve YSA için %97.50 olarak elde edilmiştir. Bu performansın, diğer çalışmalara kıyasla kabul edilebilir düzeyde olduğu görülmüştür.

Anahtar Kelimeler: El Hareketi Tanıma, Kinect, HOG, Sınıflandırma

ABSTRACT

RECOGNITION OF HAND GESTURES WITH KINECT AS COMMANDS

MASTER THESIS

Julius BAMWENDA

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DICLE UNIVERSITY

2018

The aim of this thesis is hand gesture recognition (HGR) with MS Kinect as commands. For his aim, some topics such as sign language, systems with/without Kinect were examined. During the studies, advantage/disadvantage features of Kinect and the differences of different type of cameras were learned. The important clues related with subject and results are given below.

For hand gesture recognition, totally 24 hand gestures were used. 18 of them are inclusive of American Sign Language (ASL). For each hand gesture, 100 images were recorded using Kinect. 100 images were taken from three persons for increasing accuracy of the proposed system. So, totally 24000 images were used in this study. The offline images were recorded and then RGB and depth information of each image were extracted. The last feature extraction method of histogram of oriented gradients (HOG) was applied for each image. These features are used for hand gesture recognition. Three classification methods such as support vector machines (SVM), multilayer perceptron artificial neural networks (ANN) and k-nearest neighbors algorithm (kNN) were employed. After the training stage, the proposed system having three different classifiers tested in real time. 10 gestures per character were tested online and the each result such as true/false was recorded. The average of performance of kNN, SVM and ANN are 92.5%, 95.83% and 97.50% respectively. These obtained results are acceptable level, if we compare the results with other related studies.

Keywords: Hand gesture recognition, Kinect, HOG, classification

ÇİZELGE LİSTESİ

Çizelge No		Sayfa
Çizelge 1.1.	Windows SDK için OpenNI ve MS Kinect'in karşılaştırılması	8
Çizelge 1.2.	Yaygın kullanılan derinlik sensörleri arasındaki temel farklar	10
Çizelge 2.1.	SLR konulu makalelerden bazı temel çalışmalar. Yapılan uygulamalar tümüyle işaret dilini tanıma üzerine dayalıdır	16
Çizelge 2.2.	Kinect kullanılmadan elde edilen imgelerdeki el hareketlerinin tespitine ilişkin temel çalışmalar. Listelenen çalışmaların tamamı el hareketinin modellenmesi üzerinedir.	18
Çizelge 2.3.	MS Kinect kullanılarak elde edilen imgelerdeki el hareketlerinin tespitine ilişkin temel çalışmalar.	22
Çizelge 4.1.	SVM ile el hareketlerinin komut olarak algılanmasında elde edilen performans değerleri	46
Çizelge 4.2.	YSA ile el hareketlerinin komut olarak algılanmasında elde edilen performans değerleri	50
Çizelge 4.3.	kNN ile el hareketlerinin komut olarak algılanmasında elde edilen performans değerleri	51
Çizelge 5.1	El hareketlerinin sınıflandırılmasında kullanılan yöntemlerin, ortalama başarı performansları	54
Çizelge 5.2.	MS Kinect kullanılarak elde edilen imgelerdeki el hareketlerinin tespitine ilişkin temel çalışmalar ve tez kapsamında önerilen çalışma ile karşılaştırılması	54

ŞEKİL LİSTESİ

Şekil No		Sayfa
Şekil 1.1.	Arayüzlerin gelişimi	1
Şekil 1.2.	Bilgisayarla görünün diğer alanlar ile olan ilişkisi	3
Şekil 1.3.	Kinect'in genel mimarisinin gösterimi	4
Şekil 1.4.	MS Kinect V1 versiyonu ve bileşenleri	5
Şekil 1.5.	MS Kinect V2 versiyonu ve bileşenleri	5
Şekil 1.6.	Farklı görüş açılarının oluşturulması ve tanımlanması	6
Şekil 1.7.	Windows tabanlı Kinect'in görüş alanı	6
Şekil 1.8.	OpenNI platformundaki katmanlar	7
Şekil 1.9.	Derinlik bilgisinin belirlenmesi	8
Şekil 1.10.	Derinlik uzayı skalasının bir gösterimi	8
Şekil 1.11.	Kinect'den alınan bir derinlik algısı görüntüsü gösterimi	9
Şekil 1.12.	Renk akışının bir gösterimi	9
Şekil 1.13.	İskelete ait eklem konumlarının genel görünümü	10
Şekil 1.14.	Kinect iskelet bilgisi tarafından takip edilen kişilerin tanınması ve takibinin gösterimi	11
Şekil 1.15.	Ayakta durma moduna göre iskelet takibinin akış diyagramı	12
Şekil 3.1.	Tez kapsamında önerilen yöntemin işlem akışı	12
Şekil 3.2.	Çalışmada kullanılan el figürleri için örnek imgeler	28
Şekil 3.3.	Matlab platformunda Kinect üzerinden alınan bir imgenin ekran görüntüsü	28
Şekil 3.4.	Arka plandan ayrılmış, derinlik bilgisi içeren bir örnek figür	29
Şekil 3.5.	Arka plan çıkarma algoritmasının izlenen genel akış diyagramı	30

Şekil 3.6.	İmgeye ilişkin örnek bir histogram görüntüsü	32
Şekil 3.7.	Parlaklık yönlendirme ile ilişkili bir örnek uygulama	33
Şekil 3.8.	Bir örnek görüntüdeki gradyanlar ve histogramın oluşumu	34
Şekil 3.9.	Bloğun (3 × 3 hücre) kaydırılarak tüm imgede uygulanması	35
Şekil 3.10.	HOG kullanarak imgede özellik çıkarımına ilişkin bir örnek histogram	35
Şekil 3.11.	SVM model üretimi	36
Şekil 3.12.	Örnek bir optimal hiper düzlem	36
Şekil 3.13.	Bir yapay nöron modeli	38
Şekil 3.14.	Örnek bir n:h:m YSA mimarisi	39
Şekil 3.15.	Önerine YSA mimarisi	40
Şekil 3.16.	kNN ile sınıflamada k parametresinin önemi	41
Şekil 4.1.	Farklı el hareketlerine ilişkin örnekler ve kullanılan arayüz formu	45
Şekil 4.2.	280 iterasyon ile elde edilen başarı performansları	47
Şekil 4.3.	20 bölmeli skalada hata histogramı	47
Şekil 4.4.	Eğitim aşamasında iterasyonla birlikte görülen gradyan değişimi	48
Şekil 4.5.	Her bir set için elde edilen saçılım grafikleri	49
Şekil 4.6.	Elde edilen hata matrisi (Confusion matrix)	50

KISALTMALAR VE SİMGELER

ANN : Yapay Sinir Ağı
API : Uygulama Programlama Arayüzü
ASL : Amerikan İşaret Dili
CNN : Konvolüsyonel Sinir Ağı
DOF : Özgürlük Derecesi
FOV : Görüş Alanı
GUI : Grafiksel Kullanıcı Arayüzü
HGR : El Hareketi Tanıma
HMI : İnsan-Makine Arayüzü
HMM : Gizli Markov Modelleri
HOG : Yönlendirilmiş Degradelerin Histogramı
IR : Kızılötesi Işık Görüntüsü
KNN : K- En Yakın Komşu
MS : Microsoft
NUI : Doğal Kullanıcı Arayüzü
PNG : Taşınabilir Ağ Grafikleri
PUI : Kişisel Kullanıcı Arabirimi
RGB : Katkı renk modeli Kırmızı, Yeşil ve Mavi bileşenden oluşmaktadır
SLR : İşaret Dili Tanıma
SVM : Destek Vektör Makinesi
UI : Kullanıcı Arayüzü
3D : Üç Boyutlu

1. GİRİŞ

Bu tez çalışmasında, Microsoft Kinect ile el hareketlerinin komut olarak algılanması amaçlanmıştır. Kinect bir tür arayüz olması nedeniyle, bu bölümde arayüzlerin gelişimi, el hareketlerini tanıma ve Kinect donanımı tanıtılmıştır.

1.1. Teknolojide Arayüz Gelişimi

Bilgisayarlar, çevre donanımları ile birlikte hayatın her alanında yaygınlaşmıştır. Gelişimine göre farklı arayüzler tasarlanarak, bu donanımlar ile iletişime geçilmiştir. Bilgisayar içerikli donanımlar ile insan arasında kurulan ilk arayüz, 1960'lı yıllarda komut arayüzü (CLI, Command Line Interface) olarak bilinen platformlardan gerçeklenmiştir. Ardından grafiksel kullanıcı arayüzünün (GUI, Graphical User Interface) geliştirilmesi ile birlikte, klavyenin yanında fare (Mouse) gibi donanımlar da kullanılmaya başlanmıştır. Ancak teknolojide ve sensörlerde meydana gelen üst düzey değişim, klavye ve fare gibi donanımlara olan bağlılığımızı azaltmıştır. Bu bağımlılığın azalması, doğal kullanıcı arayüzünün (NUI, Natural User Interface) ortaya çıkmasına neden olmuştur. Doğal kullanıcı arayüzü için günümüzde verilebilecek en güzel örnek, dokunmatik ekranlardır. Dokunma olayı hiç olmadığı kadar insan – bilgisayar etkileşiminde önemli bir yere sahip olmuştur. Yeni gelişen teknolojiler sayesinde, bilgisayarlar insan hareketini yakalayabilme, takip edebilme ve etkileşime geçebilme yeteneğini kazanmışlardır. Bu teknolojik gelişim ile kişisel kullanıcı arayüz (PUI, Personel User Interface) dönemi başlamıştır. Söz konusu arayüzlerin 1960'lardan bu yana olan gelişimi Şekil 1.1.'de verilmiştir.



Şekil 1.1. Arayüzlerin gelişimi (Han, J. et al. (2016)

1.2. El Hareketlerini Tanıma

Doğal kullanıcı arayüzlerinden kişisel kullanıcı arayüzlerine (PUI, Personel User Interface) geçiş yapıldığında, hareket algılama çok popüler bir alan haline gelmiştir. Bu da kişi tanımanın temellerini oluşturmuştur. Bu gelişmenin akabinde, birçok araştırmacı yüz ve el hareketlerini tanıma konusunda geniş bir ölçekte çalışma yapmıştır, Ren, Z., Meng, J. & Yuan, J., (2011). El hareketlerini (HGR, Hand Gesture Recognition) tanıma konusu, doğal kullanıcı arayüzü kapsamında teknolojiye katkı sağlayacak nitelikte olması nedeniyle, bu tez çalışmasında el hareketlerinin tanınması üzerine çalışılmıştır. El hareketlerinin tanınma konusunun farklı alanlarda uygulandığını görebilmekteyiz. Örneğin, konuşma ve duyma sıkıntısı çeken insanlar ile iletişimde, işaret dili (SR, Sign Language) bir gereksinim haline gelmiş ve el hareketlerinin tanınması üzerine çalışılmıştır. Bunun ötesinde, medikal operasyonlar, su altı maceraları gibi sessiz iletişim gerektiren yerlerde de işaret dili ve dolayısıyla el hareketlerini tanıma konusu önemli bir konuma gelmiştir.

Gelişen teknoloji ile birlikte, işaret diliyle haberleşmedeki sıkıntıları aşmak için birçok çalışma yapılmıştır, Burges, C. J. C. (1998). Bu gelişmeler bizi de bu işin bir parçası olmaya itmiştir. Dolayısıyla bu tez kapsamında, farklı işaret dilini anlamaya yönelik bir sistemin oluşturulması üzerine çalışılmıştır. Tez kapsamında el hareketlerini tanıma amacıyla Microsoft firmasının ürettiği Kinect donanımı kullanılmıştır. Bu donanımın en önemli özelliği, görüntüdeki derinliğe dayalı çalışabiliyor olmasıdır. Bu derinlik ise Kinect donanımı kapsamında bulunan kızıl ötesi (IR, Infrared) kamerası ile elde edilmektedir. Çalışmada, segmente edilen sağ el ön görüntüsü arka plandan çıkarılmış ve ele ilişkin detayları içeren görüntü elde edilmiştir. Akabinde geliştirilen algoritma ile bu görüntünün derinliği belirlenmiştir. Tez kapsamındaki tüm yazılım çalışmaları, Kinect kullanıcı arayüzünde başarıyla gerçeklenmiş ve tüm detaylar diğer bölümlerde anlatılmıştır.

El hareketlerinin algılanmasında aşağıda listelenen özellikler, bu çalışmanın anahtar kelimeleridir.

a) Görüntüyü her durumda (karanlık ortamı dahil) yakalayabilme.

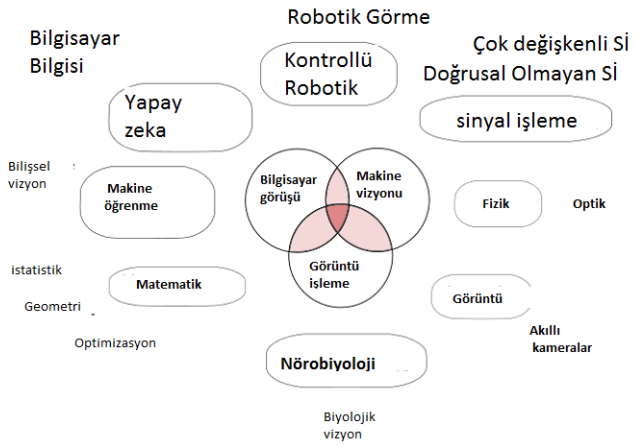
b) Sağ elin parmaklarını her harekette yakalayabilme.

c) Sağ eli segmente edebilmek için arka planı tümüyle izole edebilme.

d) Hareketleri gerçek zamanlı olarak yorumlayabilme ve gösterebilme.

Parmak hareketlerinin yukarıda listelendiği ölçüde tanınması, Kinect'in donanımsal özellikleri ile mümkün olabilmektedir. Örneğin, görüntülere ilişkin derinliğin belirlenmesinde Kinect'in IR kamerası devreye girer. Kameranın kızıl ötesi özelliğinden ötürü, sistem karanlık ortamdan etkilenmez. Sensör çıkışının 30Hz'lere kadar çıkabilmesi, sistemin gerçek zamanlı olarak çalışabilmesine olanak sağlar. Kinect ile cisimler 1.2 metreden 3.5 metreye kadar algılanabilir. Bundan dolayı, çalışmada el hareketlerine ilişkin tüm detayların yakalanabilmesi amacıyla; kamera, sağ ele yakın olacak şekilde konumlandırılmıştır. Yapılan çalışmalarda en uygun sonuçların 0.5 ile 0.8 metre aralığında elde edildiği görülmüştür.

El hareketinin tanınmasında en önemli sorun, elin pozisyonunun belirlenmesidir ki bu durum bilgisayar görüsü (computer vision) ile ilişkilidir. Görüntüyü anlama ve işleme için gerekli tüm teknik alt yapıyı sağlayan bilgisayarla görü alanı, yapay bir görme olayıdır. İnsanoğlu, sensörler gibi gözlerini kullanır ve çevresindeki farklı objeleri etiketler. Aynı şekilde bilgisayar görü alanında da insan kadar başarılı olamasa da, en azından makine görmesi olarak benzer bir misyonu yerine getirir. Birçok bilgisayarla görü uygulaması, özellikle yapılmak istenen işleve göre tasarlanır. Bilgisayarla görü kapsamında tezimizle ilişkili verilebilecek en uygun örnek; işaret dili ve el tanıma modelleridir. Bilgisayarla görünün, diğer alanlarla olan ilişkisi, Şekil 1.2.'de gösterilmiştir.



Şekil 1.2. Bilgisayarla görünün diğer alanlar ile olan ilişkisi, Kdnuggets.,(2016).

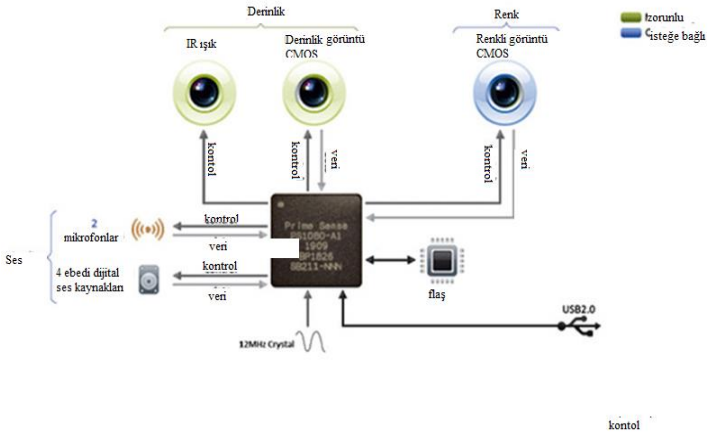
1.3. Microsoft Kinect (MS Kinect) Sensörü

Bu tez çalışmasında el hareketlerine ait görüntüleri yakalamada MS Kinect donanımı kullanılmıştır. Çalıştığımız sistem aynı zamanda birçok sistemle de gerçek zamanlı kullanılabilmektedir. MS Kinect hakkındaki genel bilgiler aşağıda belirtilmiştir.

Kinect donanımı, Microsoft tarafından Natal diye adlandırılan bir proje kapsamında 2006 yılında geliştirilmiştir. Bu projedeki temel amaç, Kinect ile Xbox 360 oyunlarının kontrolünü sağlamaktır. Proje başarıya ulaşmış ve wii adında bir kumanda sistemi geliştirilmiştir. Bu donanım cisimleri 3 boyutlu tarayabilme özelliğine sahiptir. Sonrasında Xbox daha iyi teknolojilerle gün yüzüne çıkmaya devam etmiştir, Zhang., (2012).

2010 yılında Microsoft Xbox 360 için ilk Kinect donanımı tasarlanmıştır. Donanım ilk etapta eğlence amaçlı üretilmiştir. Araştırmacılar Xbox 360'ı kullanarak yeni uygulamalar geliştirdikçe, Microsoft firmasına yeni ufuklar açmıştır. Sonrasında şirket programcıları ve uygulama geliştiricileri için yeni bir donanım tasarlanmıştır. Bu aşamada sonra pazarda V1 ve V2 diye adlandırılan iki Kinect versiyonu görülmeye başlanmıştır.

V1 ve V2 versiyonları arasında küçük bir farklılık vardır ancak destek amacıyla oluşturulan SDK yazılım platformu ile bu iki versiyonun limitleyici özellikleri ortadan kaldırılmıştır. Şekil 1.3.'de MS Kinect'in genel mimarisi verilmiştir.

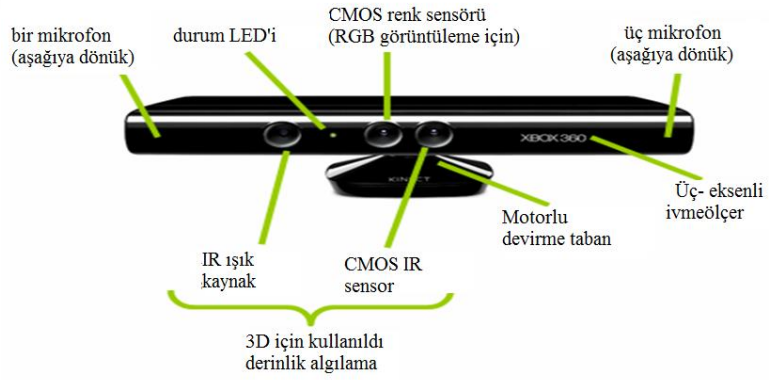


Şekil 1.3. Kinect'in genel mimarisinin gösterimi. Introduction-to-Kinect, (2012).

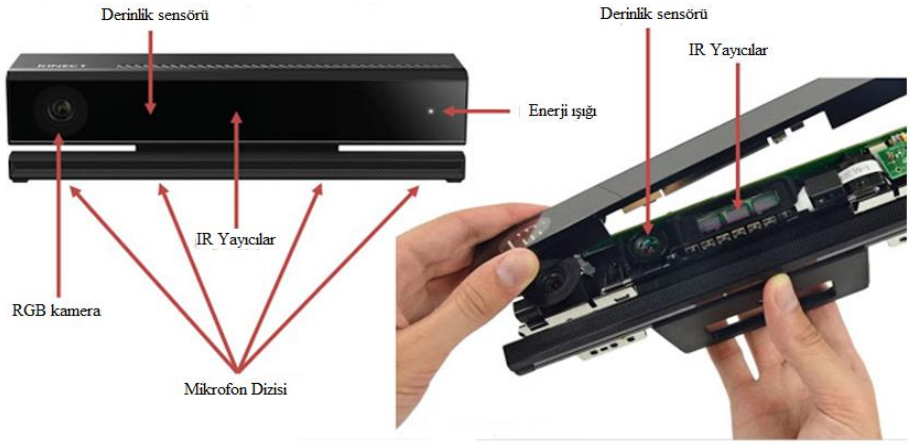
Kinect, gerçek zamanlı bir derinlik haritası oluşturma için derinlik sensörü ile birlikte bir ışık metodu kullanır. Derinlik haritası, donanım ile obje arasındaki ölçümlerin aralığını gösterir. Bu haritadaki derinlik ölçümlerinde, bazen dalgalanmalar olabilir ve bu gerçekliği önemli oranda etkiler.

Kinect' in bir kızıl ötesi (IR) projeksiyonu ve IR kamerası bulunur. Bu iki bileşen, derinlik seviyesini belirler. IR projeksiyon, objelerin üstüne düşecek şekilde kızıl ötesi ışık demeti gönderir. Oluşan örüntü, noktalar denizi şeklindedir. Bu noktalar kızıl ötesi aralığında olduğundan insan gözü tarafından algılanamaz. Noktalar sadece IR kamera ile fark edilir. IR kameraların diğer özellikleri normal RGB kamera ile aynı olmakla birlikte, tek farkı görüntüleri IR renk aralığında yakalamasıdır. IR kamera, nokta örüntülerini derinlik sensörü işlemcisine gönderir ve işlemci, derinliği algılamak için noktaların yerlerini tayin eder. Ayrıca Kinect kapsamındaki RGB kamera ile renkli görüntüler algılanır.

Kinect' de bulunan diğer bir özellik, aşağı doğru meyilli çoklu bir dizi mikrofon bulunur. Eğiklik açısı görüş alanı ile ilgilidir. Şekil 1.4 ve 1.5'de sırasıyla V1 ve V2 versiyonun donanımsal bileşenleri gösterilmiştir.



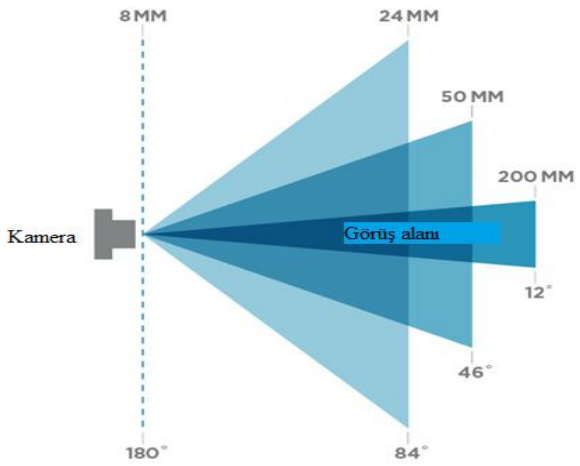
Şekil 1.4. MS Kinect V1 versiyonu ve bileşenleri, Han, J. vd. (2013).



Şekil 1.5. MS Kinect V2 versiyonu ve bileşenleri, Cook, R. Ve Cargill, T. (2014).

a) Görüş Alanı/Açısı

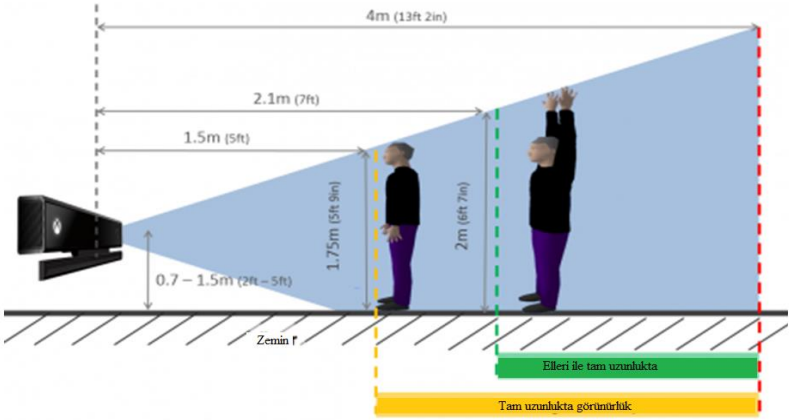
Kinect sensörleri normal kameralar gibi çalışır ve dolayısıyla kısıtlı bir görüş alanına sahiptir. Burada, Görüş Açısı (Field of View - FOV) olarak tanımlanan ve taranan cismin ne kadarının görüldüğü konusu karşımıza çıkar. Görüş açısının (FOV) farklı açılar ile gösterimi şekil 1.6'da gösterilmiştir. FOV parametresi tanımlanırken, hem yatay hem de dikey açılarının belirtildiği görülmektedir.



Şekil 1.6. Farklı görüş açılarının oluşturulması ve tanımlanması, Han, J. vd. (2013).

Kinect V2 versiyonu; 512x424 görüntü çözünürlüğü, 70.6x60 FOV (70.6° Dikey, 60° Yatay) açısı ve her derecede 7x7 piksel görüntü sağlar. Kinect'in renkli görüntü çözünürlüğü ise 84.1x53.8 FOV açısı ve her açı için 22x20 ortalama piksel görüntüsü sağlar. Windows tabanlı Kinect, objeleri yakın modda 40 cm mesafeden tanıyabilir. Bu durum, çalışmamızda dikkate alınmıştır. Tanınma ve takip için maksimum mesafe ise 80

cm'dir. Sistem için optimum mesafe 65 cm'dir. Bu aralıkta alınan görüntülerde daha az bozunum ve gürültü söz konusudur. Kullanıcı elini yatay olarak 60 ile 85 cm arasında, dikey olarak 40 ile 50 cm arasında hareket ettirebilir. Diyagram olarak MS Kinect'in görüş açısı, Şekil 1.7.'de gösterilmiştir.



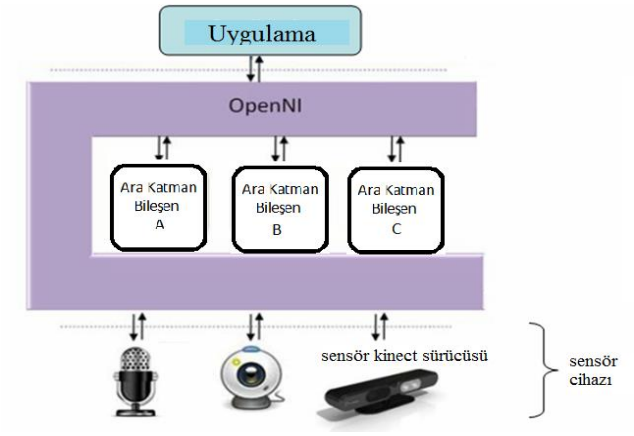
Şekil 1.7. Windows tabanlı Kinect'in görüş alanı, (docs.ipisoft.com)

b) Kinect için Yazılım Geliştirme Kiti (SDK)

Hareket tanıma, ses tanıma ve diğer amaçlar için kullanılan MS Kinect, yazılım Geliştirme Kiti (Software Development Kit - SDK) ile araştırmacıya uygulama platformu sunar. Kinect için resmi SDK yazılımı Temmuz 2011'de kullanıcılara sunulmuş ve SDK 1.0 sürümü olarak ilk kez Şubat 2012'de yayınlanmıştır. En son geliştirilen sürüm ise 2.0 dır.

Tez çalışmasında SDK 2.0 ücretsiz sürümü kullanılmıştır. Bu sürüm, Windows 8.0 ve 8.1 ile uyumludur. Günümüzde yaygın olarak kullanılan Windows 7 için SDK 1.6 sürümü kullanılabilir. SDK 2.0 sürümü, Kinect V2 için sürücüler ve arayüz içerir. Ayrıca kullanıcı için örnek kodlar da mevcuttur. Ayrıca SDK ile C++, C# ve VB.NET gibi yazılım dilleri kullanılabilmektedir.

Diğer bir açık kaynak SDK sürümü, 2010 yılında Xbox 360'ın tanıtımıyla çıkan OpenNI platformudur. Bu sürüm, Kinect'e ait her ihtiyacı karşılamakla birlikte iskelet takibi de yapabilmektedir. Yazılımın bu özelliği görülünce, birçok araştırmacı bu özelliği el ve ses tanıma uygulamalarında da kullanmıştır. Şekil 1.8'de OpenNI platformunun mimarisi gösterilmiştir.



Şekil 1.8. OpenNI platformundaki katmanlar

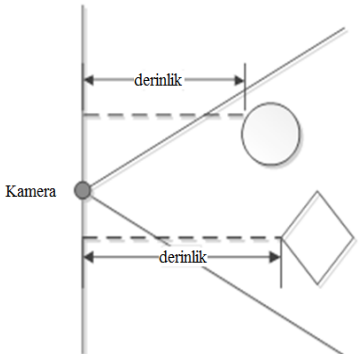
Windows tabanlı SDK platformu ile OpenNI platformunun karşılaştırılması Çizelge 1.1'de verilmiştir.

Çizelge 1.1. MS Kinect için geliştirilen Windows tabanlı SDK ve OpenNI platformlarının karşılaştırılması

	OpenNI	Microsoft SDK
Kamera kalibrasyonu	√	√
Otomatik vücut kalibrasyonu	×	√
Ayakta iskelet	√ (15 bağlantı)	√ (20 bağlantı)
Oturmuş iskelet	×	√
Beden hareketi tanıma	√	√
El hareketi analizi	√	√
Yüz izleme	√	√
Sahne analizörü	√	√
3 boyutlu tarama	√	√
Motor kontrolü	√	√

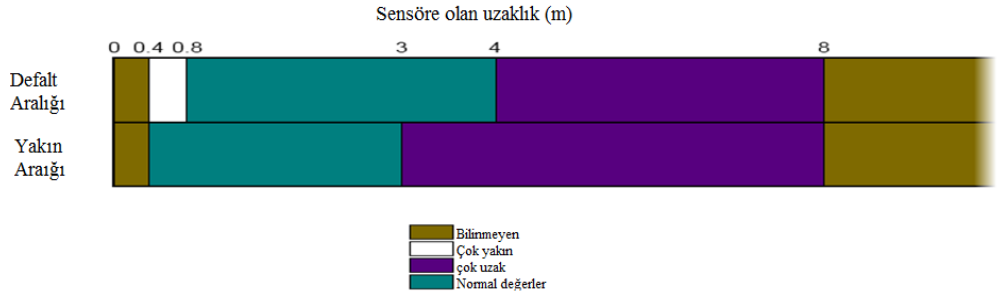
c) Derinlik Bilgisi

Donanımda derinlik algısı, yapısında bulunan derinlik kamerası ile elde edilir. Derinlik bilgisi piksel içeren birçok katmanın temsili sonucu oluşur. Kinect, kameranın yakınındaki objelerin mesafesini milimetre cinsinden ölçer (Şekil 1.9).



Şekil 1.9. Derinlik bilgisinin belirlenmesi

Bir objeye olan mesafenin belirlenmesi için objeye ait segmentasyon bilgisi ve mesafe pikselleri birleştirilir. Objeye ait segmentasyon bilgisi, takip edilen iskeletle ilgili kaydedilen bilgi ile ilintilidir. 16 bitlik derinlik bilgisinin 3 biti obje segmentasyonu ve 13 biti ise mesafe bilgisidir. Şekil 1.10'da gösterilen derinlik bilgisinden görüldüğü üzere, görüntü en fazla 8m mesafeden algılanabilmektedir.



Şekil 1.10. Derinlik uzayı skalasının bir gösterimi, msdn.microsoft.com (2017).

Derinlik ayarı farklı çözünürlükte (80x60, 320x240 ve 512x484 gibi) elde edilebilir. Bütün çözünürlükler için derinlik ayarı 30 fps' de elde edilir. Şekil 1.10.'da görüldüğü gibi kamerada yakın ve normal mod olmak üzere 2 farklı mod bulunur. Normal modda kamera sensörleri, 0.8 ile 4 metre arasında derinlik bilgisini algılar. Diğer taraftan yakın modda ise derinlik bilgisi, 0.4 ile 3 metre arasında algılanır. Aşınmaları azaltmak için yakın mod kullanımı tercih edilir. Şekil 1.11.'de sağ ele ilişkin yakın modda bir derinlik bilgisi gösterimi verilmiştir.



Şekil 1.11. Kinect'den alınan bir derinlik algısı görüntüsü gösterimi

Derinlik görüntüsünü elde etmede birçok sensör bulunmaktadır. Farklı sensörlere ilişkin özellikler ve temel farklılıklar Çizelge 1.2'de verilmiştir. Çizelgede görülen Kinect 2.0 versiyonu, belirli bir optik dalga formu aralığında ortama bağlı optik yol uzunluğu ölçümüne dayanır. Böylelikle ortamın yapısı ve özellikleri analiz edilebilmektedir.

Çizelge 1.2. Yaygın kullanılan sensörlerin derinlik açısından karşılaştırılması

kullanılan derinlik sensörleri

Sensör	Çözünürlük	Aralığı	Doğruluk	Açıklama
Kinect 1.0	320 x 240	0.8-4.0 m	4mm	20 Vücut eklemleri
Leap Motion	640 x 240	25-600mm	0.01mm	27 El eklemleri
Kinect 2.0 (TOF)	512 x 484	0.8-4.5m	1mm	25 Vücut eklemleri

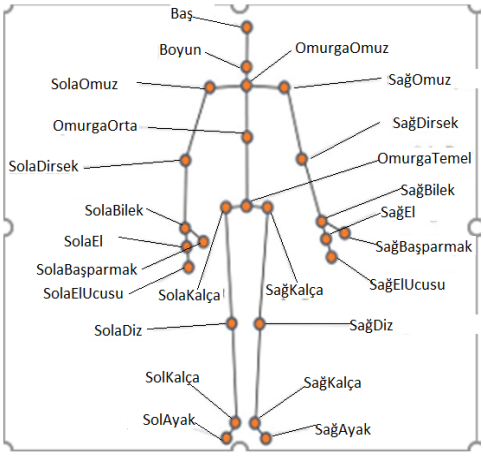
Renkli görüntü ayalarından Bayer veya YUV formatı seçilebilmektedir. Ancak kızıl ötesi seçilmesi durumunda, renk akışı kızıl ötesi akışına dönüşür. Dış kameranın kalibrasyonu, kızıl ötesi sistem sayesinde iyileştirilebilir. Karanlık ortamda görüntülerin yakalanmasında, kızıl ötesi ışık kaynağının benzerlerine göre daha gürbüz sonuçlar ürettiği görülmektedir. Görüntülerin analizinde, birçok araştırmacı tarafından IR kullanıldığı görülebilmektedir, Han.,Xa, (2013). Renk akışına ilişkin örnek bir örnek, şekil 1.12'de gösterilmiştir.



Şekil 1.12. Renk akışının bir gösterimi

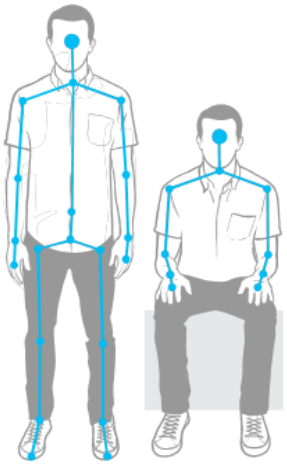
d) İskelet Takibi

Kinect donanımı ile gerçekleştirilen en iyi uygulamalardan biri iskelet takibidir. İskelet takibi sayesinde kişilerin tespiti, tanınması ve hareketlerinin takibi yapılabilmektedir, Song.,Lan, (2013). İskelet takibinin yapılabilmesi için kişinin sensör önünde bulunması gerekir. Özellikle kişinin başı ve vücudunun üst kısmı, sensör tarafından görülmesi gerekir. Kişilerin tanınması için özel bir duruş ve ayar bulunmamaktadır. Eklemlerin konumu, iskelet takibi için önemli bir parametredir. İlgili sensör, eklemlerin koordinatlarını X,Y ve Z ekseninde metrik türünden algılar (Şekil 1.13.).



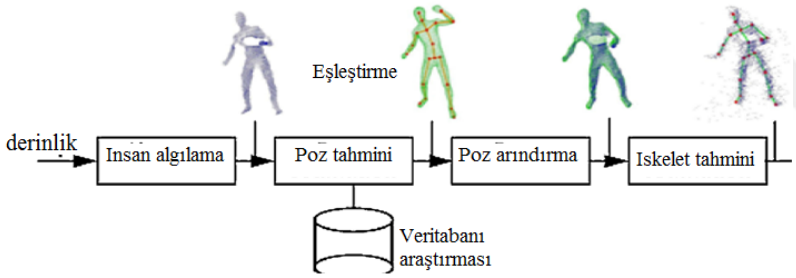
Şekil 1.13. İskelete ait eklem konumlarının genel görünümü

Takip etme algoritmalarında, genel olarak oturma ve ayakta (otomatik) durma şeklinde iki farklı mod mevcuttur (Fassel.,Mathias., (2013)). Kinect, ayakta durma (Otomatik) modunda toplamda 25 farklı eklem noktasını takip etmektedir. Oturma modu ise çoğunlukla kullanıcıların oturur pozisyonda takibi ve tanınması için kullanılır. Oturma modunda vücudun üst kısmına ait 10 farklı eklem noktası takip edilir. Modlara bakıldığında, yapılan takibe göre izlenen eklemler değişiklik gösterir. Ayakta durma (otomatik) modunda, kişilerin arka plana olan uzaklığı takip edilir. Oturma modunda ise kişiler arka plan görüntülerinden ayrılır. Bu aşamadan sonra, algoritma hareket takibine geçer. Oturma modu, ayakta durma moduna göre bellekte daha fazla yer bloke eder. Diğer taraftan yakın ve oturma modu, en iyi iskelet takibini sağlar. Takip sisteminde, pratik olarak belirli bir zaman içinde sadece bir mod kullanılabilir. Diğer bir değişle, sistem oturan bir kişiye takip ederken, eş zamanlı olarak ayakta bulunan bir kişiyi talip edemez. Kinect iskelet bilgisi ile kişilerin takibini sağlayabileceğine ilişkin bir örnek şekil 1.14'de gösterilmiştir.



Şekil 1.14. Kinect iskelet bilgisi tarafından takip edilen kişilerin tanınması ve takibinin gösterimi

Ayakta duran bir kişiye ilişkin iskelet takibinin gerçeklenmesinde, izlenen akış diyagramı şekil 1.15'de gösterilmiştir.



Şekil 1.15. Ayakta durma moduna göre iskelet takibinin akış diyagramı

İskelet takip bilgisi, çevresel gürültü ve sensör kaynaklı fiziksel kısıtlamalara bağlı olabilir. Daha düzgün parametre ayarı ile iskelet takip sisteminin daha iyi sonuç verebileceği unutulmamalıdır. İskelet takip algoritmaları, "Holt Double Exponential Smooothing" tabanlı yöntemler ile filtreleme işlemini içerir. Bu metot genelde ekonomi ile alakalı istatistiksel verilerin analizinde kullanılır. Filtre, diğer filtreleme algoritmalarına nazaran daha az gecikme ile daha düz sonuç elde edilmesini sağlar (Lan. Roanna (2015). Davranışların, parametrelere ve takip sistemine olan etkileri, Kinect for Windows, (2013) kaynağında anlatılmıştır.

Bu tez çalışmasında, Microsoft Kinect ile el hareketlerinin komut olarak algılanması amaçlanmıştır. Kinect tabanlı alınan sağ el görüntüleri, farklı makine

öğrenmesi algoritmaları ile sınıflandırılmıştır. Bu algoritmalar; Destek Vektör Makineleri (DVM), Yapay Sinir Ağları (YSA) ve K-means yöntemleridir. Bu yöntemler, Materyal ve Metot bölümünde açıklanmıştır.

2. KAYNAK ÖZETLERİ

Çalışmamızın amacı el hareketlerini gerçek zamanlı olarak tanıyabilen ve komut olarak algılayabilen bir sistem tasarlamaktır. Söz konusu alanla ilişkili literatür çalışması bu bölümde anlatılmıştır. Tez konusuna ilişkin literatürde görülen çalışmalar, üç bölümde değerlendirilebilir. Bu aşamalar; işaret dilini algılama, el hareketini tanıma ve MS Kinect ile el hareketini tanımadır. Bu aşamaların detayları aşağıda belirtilmiştir.

a) İşaret Dilini Tanıma

İşaret dili, işitme engelli olan insanlar tarafından kullanılan görsel bir dildir. İşaret dili tanımanın temel zorluğu, üç boyutlu (3D) uzayda işaret örneklerinin hem hareket hem de şekil bakımından farklılık göstermesidir. Bu sorunların çözümü için görüntü tabanlı ve algılayıcı tabanlı olmak üzere iki temel yaklaşım vardır. Görüntü tabanlı yaklaşımda, mesajın iletimi için basit donanımlar kullanırlar ve önişleme aşamasında önemli hesaplar gerektirir. Sensör tabanlı sistemlerde ise sensörlerle donatılmış eldivenler kullanılır. Sensör tabanlı sistemlerin yüksek maliyete sahip olduğu bilinmektedir. Bundan dolayı, bu çalışmada görüntü/imge tabanlı yaklaşım dikkate alınmıştır. Ren ve Meng tarafından imgelerdeki derinlik bilgileri kullanılarak, bir işaret dili tanıma sistemi önerilmiştir. Derinliğe dayalı bilgi tabanlı yaklaşımlar, genellikle renk veya 2D temelli yaklaşımlara oranla daha doğru sonuçlar üretebilmektedir, (Ren., Meng, (2013)). Bilindiği gibi, işaret dili için vücut bölümlerinden baş ve eller gereklidir. İşaret dilinin tanınması (SLR, Sign Language Recognition) konusunda literatürde geniş bir ölçekte çalışmaların yapıldığı görülebilmekte ve en temel çalışmalar Çizelge 2.1'de listelenmiştir. Ayrıca listelenen çalışmaların detaylarına aşağıda değinilmiştir.

Çizelge 2.1. SLR konulu makalelerden bazı temel çalışmalar. Yapılan uygulamalar tümüyle işaret dilini

tanıma üzerine dayalıdır

Kaynak	Yıl	Sınıf		Sınıflandırma Metodu	% Doğruluk	Donanım
		Statik	Dinamik			
Haro ve ark.	2008		X			
Kang.	2015		X	CNN	89.6	sen23D cam
Kishore ve ark.	2013	X		EFD ve ANN	92	phone cam
Ong ve Bowden	2002		X	K-mediod	97.4	Webcam
Starner ve ark.	1995		X	HMM		Webcam

Starner ve arkadaşları (1995), Gizli Markov Modellerini (HMM) kullanarak iki görsel tabanlı SLR sistemi önermişlerdir. El takibi için HMM ile sürekli hareket takibi yapılırken, çalışmada deri-renk eşleştirme yaklaşımı kullanılmıştır. Algoritmada, deri rengi pikseli bulunduktan sonra, aynı renk alanlarının tespiti için en yakın sekiz piksel dikkate alınmıştır. Yüz bölgesinin belirlenmesinde, eller hareket ederken kişiye ilişkin pozisyonunun neredeyse sabit olduğu varsayılmıştır. 2D video, örtüşme ayrımı ile sınırlandırılmıştır ve bu nedenle, iki elin örtüştüğü imgelerde ellerin ayrıştırılması sağlanamamıştır. Ancak tüm alanın her iki ele atanması ile bu sorun çözüme kavuşturulmuştur. Buna göre önerilen sistemlere ilişkin doğruluk oranlarının sırasıyla %92 ve %98 olduğu görülmüştür. Elde edilen sonuçlar, renk tabanlı hareketin sınıflandırılması ve izlenmesinde HMM yönteminin etkili olduğunu göstermiştir.

Ellerin takibini sağlamak için Ong ve Bowden (2002) tarafından önerilen diğer bir çalışmada, gri tonlamalı imgelerde el şekillerini saptayan hızlandırılmış kaskat sınıflandırıcı kullanılmıştır. Öncelikle farklı el şekilleri için çok sayıda imge kayda alınmıştır. Ardından K-medoid kümeleme algoritması kullanılarak, imgelerin gruplandırılması yapılmıştır. El imgeleri birinci katmandaki imge ile özetlenmiştir. Birinci katmanda gruplandıktan sonra, sınıflandırıcılar el şekillerini içeren görüntü bloklarını dikkate alarak, ilgili kümelerdeki diğer görüntülerle karşılaştırmak maksadıyla ikinci katmana aktarılmıştır. Çalışmanın sınıflandırma performansı, FloatBoost algoritması ile artırılmıştır. Denemede 5013 el imgesi kullanılmıştır. Ellerin algılanmasında birinci katmanın %98 doğruluğa, el şeklinin algılanmasında ikinci katmanın ise %97.4 doğruluğa sahip olduğu görülmüştür. Bu çalışmadaki yüksek performansın, doğrudan K-medoid kümeleme yöntemi ve FloatBoost algoritması ile ilişkili olduğu görülmüştür.

Kishore ve Prasad (2013), el hareketleri ile Hintçe işaret dilini sesli mesajlara dönüştüren bir sistem tasarladı. Araştırmalarında, otuz altı hareket işaretini düşünüyorlardı. Görüntüleri elde etmek için bir cep telefonu kamerası kullandılar. Gürültüyü azaltmak için görüntülerde uzaysal ortalama filtresi uygulandı. Sonuçta, görüntüler tek bir düzlemde düzleştirildi. El şekilleri için ikili görüntü özellikleri eliptik Fourier tanımlayıcıları kullanılarak çıkarıldı. Görüntülerin sınıflandırılması için sinir ağları, görüntülerden elde edilen özelliklerden hareketlerin tanımlanması için uygulanmış ve ortalama % 92 sınıflandırma doğruluğu elde edilmiştir. Tanınan işaretler, görüntünün giriş özellikleri ile ilgili sesli komutların yerini verir.

Kang (2015), işaret dili tanıma için önemli olan Amerikan İşaret Dili'nde (ASL) statik bir parmak imla oluşturma üzerine odaklanmaktadır. Diğer araştırmalar gibi, 24 sınıfın sınıflandırılmasına yönelik ortak yaklaşımın aksine otuz bir alfabe ve sayı da başarıyla sınıflandırılmıştır. 320 × 240 çözünürlüğe sahip Senz3D kameranın bir derinlik sensörünü kullanarak 31.000 derinlikli haritayı topluyorlardı, her el işaretinin 1000 görüntüden oluştuğunu ima ediyorlardı. Parmak yazım denetimi için kameraya en yakın nesnenin kullanıcının eli olduğunu varsaydılar. Bu onlara el bölütlemede yardımcı oldu. Beş konvolüsyon katmanı, üç max-pooling katmanı ve üç tamamen bağlı katmandan oluşan bir ağı kullanarak ön işleme tabi tutulmuş her derinlik görüntüsünden bir dizi boyutsal öznitelik vektörü çıkardılar. Veri setleri, beş farklı işletim modunda sinir ağları kullanılarak eğitilmiş ve test edilmiştir. Ortalama performans % 89.6 olarak elde edildi.

Haro ve diğerleri, (2008), İspanyolca için konuşma-işaret dili çeviri sistemi geliştirdi. Kimlik kartını yenileyen veya başvuran kişilere yardım ederken bir memur tarafından konuşulan cümlelerden oluşmaktadır ve çeviri sistemi üç modülden oluşmaktadır. Sözlü anlatımı bir sözcük sırasına indirgemek için konuşma tanıma cihazı kullanılmıştır. Ayrıca, kelime dizisini İspanyol İşaret Dili'ne (LSE) ait işaret dizisine dönüştüren doğal dil çeviri modülünü kullanır. Son olarak, işaret dizisini canlandıran bir 3D avatar. Doğal dil çeviri modülü için iki öneri sundu. Bunlardan ilki, % 31.60 SER (İşaret Hata Oranı) ve 0,5780 BLEU'ya (BiLingual Evaluation Understudy) ulaşan bir kural tabanlı çeviri modülünden oluşur. Bu öneride, her işaret için bir güven ölçümünü hesaplamak için konuşma tanıma aracından gelen güven ölçümleri kullanılmıştır. Bu güven ölçüsü, işaret animasyon işlemi sırasında çevrilen işaretin güvenilirliği hakkında kullanıcıyı bilgilendirmek için kullanılır. Doğal dil çevirisi için ikinci seçenek, paralel

yapıların eğitim amaçlı kullanıldığı istatistiksel bir çeviri yaklaşımına dayanmaktadır. En iyi konfigürasyon bir % 38.72 SER ve bir 0.4941 BLEU olarak rapor edilmiştir.

Kishore ve Prasad (2013), Hintçe işaret dilinin statik hareketlerini tanımaya yönelik bir yaklaşım önermişlerdir. Çalışmada, görüntü işleme teknikleri ve sinir ağları kullanılmıştır. Deneyler, İngiliz alfabe hareketleri ve sayıları üzerinde gerçekleştirilmiştir. Ayrıca Hintçe işaret dili, metin ve sesli mesajlara dönüştürülmüştür. Görüntü özelliklerinin çıkarılmasında güçlü Gabor filtreleri ve özellik vektör boyutunun düşürülmesi için ana bileşen analizi (PCA) kullanılmıştır. Son olarak, özellik matrisi paralel dağıtılmış işlem için sinir ağının girişine uygulanmıştır. Sınıflandırma işleminden sonra, ortalama % 98,2 oranında bir performans elde edilmiştir.

b) Diğer Cihazlarla El Hareketi Tanıma

El hareketlerini tanıma (HGR) işlemi, makine öğrenmesinde önemli sorunlardan biridir. Bir önceki bölümde de belirtildiği üzere, el hareketleri statik ve dinamik olmak üzere iki sınıfta değerlendirilmektedir. Bu bölümde, her iki sınıfa yer alan çalışmalar incelenmiştir. Kinect kullanılmadan elde edilen imgelerin sınıflandırılmasına ilişkin temel çalışmalar Çizelge 2.2'de listelenmiştir. Ayrıca listelenen çalışmaların detaylarına aşağıda değinilmiştir.

Çizelge 2.2. Kinect kullanılmadan elde edilen imgelerdeki el hareketlerinin tespitine ilişkin temel çalışmalar. Listelenen çalışmaların tamamı el hareketinin modellenmesi üzerinedir.

Kaynak	Yıl	Tanıma sınıfı		Sınıflandırma Metodu	% Doğruluk	Donanım
		Statik	dinamik			
Ari	2011	X		KBFNN	99.6	çevrimiçi veri setleri
Kainz ve Jakab	2014	X		DBN(DNN)		sıçrama hareketi
Molchanov, Gupta, Kim	2015		x	CNN		çevrimiçi veri setleri
Nagashree ve ark.	2015	X		SVM		web kamerası
Smedt ve ark.	2017		x	SVM/CNN	88/81	intel realsense
Vieriu ve ark.	2016	X		HMM	96.2	canon kamera
Suharjito F.B.	2016	X	x	SVM/HMM	91/89	web kamerası

Ren ve Meng (2011), aritmetik hesaplama ve taş- kağıt-makas oyunu olmak üzere iki gerçek yaşam problemine uygulanabilen bir el hareket tanıma sistemi önermişlerdir. Karşılaşılan başlıca güçlüklerin elle tespit ve hareket tanıma olduğu çalışmada belirtilmiş ve bu zorlukların Kinect ile aşıldığı vurgulanmıştır. El segmentasyonu ve şekil gösterimi ile izlenen görüntüler kaydında Kinect kullanılmıştır. Kinect'in düşük çözünürlüğünden

kaynaklı bozulmaların etkisini azaltmak için FEMD (Fingure-Earth Mover's Distance) olarak bilinen bir şekil-mesafe matrisi kullanılmıştır. Hem derinlik haritası hem de Kinect'in renkli görüntüsü kullanılarak, el şekilleri başarıyla tespit edilmiştir. Parmakları mükemmel bir şekilde tespit edebilmek için, eşik ayrıştırma ve yakın konveks şekil ayrıştırma olmak üzere iki parmak algılama algoritması uygulamıştır. Çalışmada %90.6 oranında bir performans elde edilmiştir.

Smedt ve arkadaşları (2017), 3D dinamik el hareket veri setini oluşturmuşlardır. Bunun oluşumunda, el iskeleti verileri ve derinlik görüntüleri kullanılmıştır. Bu çalışmada, iskelet tabanlı tanıma yöntemi ve anahtar çerçeve dizisi ile CNN algoritmaları uygulanarak örüntüler sınıflandırılmıştır. 3D el iskeleti verileri ile birlikte dinamik bir el hareketinin, bir zaman serisi olarak değerlendirilebileceği çalışmada belirtilmiştir. Ayrıca, 14 hareketin 28 hareketten daha yüksek bir performansa sahip olduğu gözlenmiştir. El iskeleti tabanlı yöntem ile 14 ve 28 farklı hareket için sırasıyla %88.24 ve %81.90 başarım oranları elde edilmiştir.

Kainz ve Jakab (2014), el tanıma için yeni bir izleme yaklaşımı önermiştir. Derinlik tanımada, görsel tabanlı izleme sistemi ile kombine SEMG tekniği yer almaktadır. İzleme sürecinde, el kaslarının EMG işaretleri ve imge işleme konularından yararlanılmıştır. Çalışmada hareket algılamada, leap motion donanımı kullanılmıştır. Leap Motion donanımının, el hareketlerini yakalamada etkin çalıştığı ancak sürekli parmak takibinde bazı el pozisyonlarında sorunlar yaşadığı görülmüştür. SEMG izleme modellerinin sınıflandırılmasında sinir ağları kullanılmış ve özellikle DBN (Deep Belief Network) ağının %96.34 ile en iyi performansı sergilediği belirtilmiştir.

Suharjito, Fitra B. Adinugraha (2016) tarafından yapılan çalışmada, Gizli Markov modeli ve Destek Vektör Makinesi (DVM) birleştirilmiş ve bir video akışında yer alan el hareketlerinin tanınması için uygulanmıştır. Bu uygulama ile DVM'nin el tanıma çalışmalarında kullanılabileceği görülmüştür. Gizli Markov Modelinin, mekansal-zamansal zaman serilerini modelleme açısından performansının yüksek olduğu literatürden bilinmektedir. Bu çalışma ile elde edilen başarımlar; statik el hareketi tanıma için ortalama % 91 ve dinamik el hareketi tanıma için ortalama %89 oranları elde edilmiştir.

Statik el hareketleri için yapılan diğer bir çalışma Ari (2011) tarafından önerilmiştir. Sınıflandırıcı olarak k-means tabanlı RBF sinir ağı kullanılmıştır. Radyal tabanlı fonksiyon ağı (RBFNN) en popüler sinir ağıdır. Popüler olmasının sebepleri; (i) mimarisi çok basit, sadece bir gizli katman içerir; (ii) gizli katmanda yerelleştirilmiş radyal taban fonksiyonu, özellik vektörünün girdi alanından gizli alana doğrusal olmayan dönüşümü için kullanılır; (iii) gizli nöronun duyarlılığı, taban fonksiyonuna ilişkin yayılma faktörünün ayarlanmasıyla sağlanır; (iv) hızlı ve yerel minima problemi yoktur, Haykin,(1999). Önişlem kısmı, üç aşamada gerçekleştirilmiştir. Hareket alanını resmin arka planından çıkarmak için Otsu bölütleme algoritması kullanılmıştır. Çalışmada morfolojik filtreleme uygulanmıştır. Doğru sınıflandırma için rotasyon, ölçek ve konumu değişmez olan Lokalize Kontur Dizisi (LCS), el hareketinin özellik kümesi olarak kullanılmıştır. Önerilen algoritma ile 25 statik el hareketine ilişkin görüntülerin % 99.60 oranı, doğru olarak saptanmıştır.

Statik el hareketleri üzerine diğer bir çalışma, Vieriu ve arkadaşları (2011) tarafından yapılmıştır. Çalışmada Canon Power Shot kamerası kullanılmış ve dokuz hareket sınıfı belirlenmiştir. Her sınıf için 50 imge kayda alındığından, toplamda 450 imge kullanılmıştır. Öznitelik çıkarımı aşamasında cildin renk segmentasyonu yapılarak, elin izolasyonu sağlanmıştır. Daha sonra imgenin gürültüden arındırılması için medyan filtreleme uygulanmıştır. Önerilen yöntem ile %96.2 oranında başarı elde edilmiştir.

Nagasharee ve arkadaşları (2015), el hareketlerini statik olarak algılayabilen yeni bir yaklaşım önerdiler. Veri kümesinin oluşturulmasında, bir web kamerası kullanılmıştır. Çalışmada, web kamerasının görüntüdeki derinliği algılayabilme özelliği kullanılmıştır. Önişlem aşamasında, eli ön ve arka zeminden ayırmak için bir sınır kutusu kullanılmıştır. Sınıflandırmada herhangi bir önemi olmayan piksellerin kaldırılması için imgeler yeniden boyutlandırılmıştır. Yeniden boyutlandırılmış imgelerdeki kenarların tespiti için Canny kenar algılama yöntemi kullanılmıştır. Yönlü Gradyan Histogram (Histogram of Oriented Gradients, HOG) ile elde edilen özellikler SVM ile sınıflandırılmıştır.

Parvathy ve Hema (2016) tarafından yapılan diğer bir çalışmada; ön işleme, arka plan çıkarma ve kenar algılama tekniklerine dayanan bir el hareketi bölütleme tekniği önerilmiştir. Başlangıçta diğer el hareketi tanıma yöntemlerinde olduğu gibi, el hareketleri web kamerası ile kayda alınmış ve önişlem olarak imgelere Yerel Olmayan Ortalama Filtre

uygulanmıştır. Arka plan için Gauss Karma Modeli (GMM) kullanılmıştır. Son olarak, kenar algılama için Canny kenar algılama yöntemi kullanılmıştır. Elde edilen sonuçlar diğer yöntemler ile karşılaştırılmıştır.

c) MS Kinect Tabanlı El Hareketi Tanıma

MS Kinect donanımının kullanılması ile el hareketlerinin belirlenmesinde büyük gelişmeler sağlandığı görülmüştür. Daha önceki bölümde de belirtildiği gibi, araştırmacıların çalışmalarında kullandıkları imgelere ilişkin (derinlik kamerası, kızıl ötesi ile temin edilen) çeşitli ayrıntılı veriler Kinect ile temin edilebilmektedir. Bu tez çalışmasında da Kinect donanımı kullanılmış ve el hareketlerinin komut olarak algılanması sağlanmıştır. Kinect kullanılarak elde edilen imgelerin sınıflandırılmasına ilişkin temel çalışmalar Çizelge 2.3'de listelenmiştir. Ayrıca listelenen çalışmaların detaylarına aşağıda değinilmiştir.

Çizelge 2.3. MS Kinect kullanılarak elde edilen imgelerdeki el hareketlerinin tespitine ilişkin temel çalışmalar.

Kaynak	Yıl	Yöntem	Sınıflandırma Yöntemi	% Doğruluk	Uygulama/Görev
Chen ve Wang n.d.	2017	dinamik	RNN	90	El hareketi modelleme
Chong ve ark	2016	statik	Karar ağaçları	94.7	Düzeltici hesaplama
Priyanka.	2016	statik			El hareketi modelleme

Dinh ve ark,	2013	statik	Karar ağaçları	97.8	Düzeltici hesaplama
Dominio ve ark.	2013	statik	SVM	97.5	El hareketi modelleme
Dong ve ark.	2011	statik	Rastgele Orman	90	İşaret dili
Li	2010	statik	K-ortalama	99.5	El hareketi modelleme
Molchanov Kim ve ark	2015	statik	ANN	90	Robotik
Nguyen ve Le	2015	statik	SVM ve RVM'ler		El hareketi modelleme
Ren ve arkadaşları	2011	statik		90.6	Düzeltici hesaplama
Song ve ark	2013	statik	SVM	97.7	El hareketi modelleme
Vo ve ark	2017	statik	SVM ve HMM	95	İşaret dili
Xie ve Shen	2015	statik	kNN	99.7	El hareketi modelleme
Keskin ve ark.	2011	statik	SVM	99.9	İşaret dili

Chen ve Wang (2017), çalışmalarında iskelet tabanlı bir dinamik el hareketi tanıma sistemi önermiştir. Şekil hareketlerini analiz etmek için el iskeletinden kemiklerin açıları belirlenmiş ve buradan parmak hareketlerine ilişkin özellikler elde edilmiştir. Ardından elin küresel hareket özellikleri belirlenerek, sınıflandırma işlemleri yapılmıştır. Sınıflandırma işleminde, parmak hareket özellikleri ve genel özellikler, iskelet dizisi ile birlikte RNN modeline uygulanmıştır. Performansı karşılaştırmak için iskelet tabanlı DHG-14/28 veri seti kullanılmıştır.

Chong ve arkadaşları (2016) tarafından, 3D nokta bulut tabanlı bir el tanıma yöntemi önerilmiştir. Derinlik kamerasından 3D ile ele ilişkin ham veriler çıkarılmıştır. Gerilmiş parmakların sayısı, parmaklar arasındaki açılar ve hareket bölgesi alan dağılımı olmak üzere üç özellik belirlenmiştir. Bu özelliklerin sınıflandırılmasında karar ağacı yöntemi kullanılmıştır. Ortalama doğruluk %94.7 olması nedeniyle, önerilen yöntemin hareket tanımada oldukça iyi olduğu gözlenmiştir.

Keskin ve arkadaşları (2011), çalışmalarında el için gerçek zamanlı iskelet uyum algoritmasını içeren bir derinlik görüntüsü önermişlerdir. Çalışmada, Kinect'ten alınan derinlik görüntüleri gerçek zamanlı olarak 30 fps'de değerlendirmeye alınmıştır. Parça yaklaşımıyla bir nesne tanıma yöntemi geliştirilmiştir. 21 farklı parça ile bir 3D el modeli oluşturulmuştur. Rastgele karar ormanları (Random Decision Forests, RDF) algoritması, derinlik görüntüleri ile eğitilerek el modeli oluşturulmuştur. Ardından, el iskeletinin ortak yerlerini tahmin etmek için sınıflandırma sonuçları, lokal mod bulma algoritmasına uygulanmıştır. Son aşamada ASL'nin on dijiti için SVM kullanılmıştır. Uygulama sonuçlarında, gerçek zamanlı canlı derinlikli görüntülerde % 99.9 oranında bir performans elde edilmiştir.

Dinh, Jeon ve lee, (2013) tarafından önerilen yaklaşımda, Kinect ile çekilen derinlikli imgelerde yer alan el parçalarına göre el hareketi belirlenmiştir. Önerilen yaklaşım, iki temel aşamadan oluşmuştur: Rasgele Ormanlar (Random Forests, RF) ile el parçalarının tanınması ve kural tabanlı el hareketi tanımasıdır. El derinliğinin yer aldığı bir veritabanı ve bunlara karşılık gelen el parçalarının etiketli bir haritası oluşturulmuştur. Ardından Rastgele Ormanlar yöntemi ile modelin eğitimi sağlanmıştır. Önerilen yaklaşım, sentetik ve gerçek veriler ile değerlendirilmiş ve beş denekten elde edilen 10 el hareketi üzerinden ortalama tanıma oranı %97.80 olarak elde edilmiştir.

Molchanav Kim ve arkadaşları (2015), araç sürücülerine ilişkin el hareketlerini tanıma amaçlı bir yöntem önermişlerdir. Çalışmada MS Kinect ile birlikte çoklu sensörler kullanılmıştır. Bu makale ile Kinect'in farklı bir alanda uygulanabileceği görülmüştür. Çalışmada sınıflandırıcı olarak ANN kullanılmıştır.

Vo ve arkadaşları (2017), Vietnam İşaret Diline (VSL) karşılık gelen sürekli (dinamik) hareketi tanımaya yönelik bir özellik çıkarma ve sınıflandırma yaklaşımı önermişlerdir. MS Kinect ile alınan derinlik bilgisi çalışmada kullanılmıştır. 5 gönüllü tarafından VSL'de 30 dinamik harekete karşılık gelen 3000 örnekten oluşan bir veri kümesi oluşturulmuştur. Hareketler, diğer araştırmacılarda olduğu gibi derinlik görüntüsünün dizisi ile temsil edilmiştir. Özellik çıkarmak için imgeler aynı büyüklükteki blokların üç boyutlu ızgarasına bölünmüş ve her birinden bir skaler değer elde edilmiştir. Sınıflandırma için SVM ve Gizli Markov Modeli (HMM) kullanılmıştır. Deneylerin ortalama doğruluk oranı %95 olarak elde edilmiştir.

tasarlamışlardır. Çalışmada, sekiz el hareketi ve her harekete ilişkin 80 örneğin yer aldığı bir veri seti oluşturulmuştur. Hareketlerin sınıflandırılmasında, Dynamic Time Warping ve Gizli Markov Modeli kullanılmıştır. Önerilen yöntem ile doğruluk tanıma oranının %99 olduğu görülmüştür.

Li (2010), işitme sorunları olan insanlar ile etkin bir şekilde iletişime geçmek maksadıyla bir el hareketini tanıma sistemi önermiştir. Sistem, el hareketini önceden tanımlanan el hareketleri ile karşılaştırarak tanır. Yöntem; el algılama, şekil tanımlama ve hareket tanıma olmak üzere üç aşamadan oluşmuştur. K-means kümeleme

algoritmaları uygulanarak, imgeleri sınıflandırılmıştır. Yaklaşık %99.5 oranında bir tanıma performansı elde edilmiştir.

El modellemesi yapmak karmaşık görevlerden biri olmakla birlikte, birçok araştırmacı bu zorluğun üstesinden gelmek için farklı yöntemler önermiştir. Bunlardan bir tanesi de Xie ve Shen (2015) tarafından yapılan çalışmadır. Çalışmalarında; Kinect görüntülerinden renk, derinlik ve iskelet parametrelerine dayalı bir el hareketini tanıma yöntemi önerilmiştir. Yöntemde, eller önce renk verisi ile arka plandan ayırt edilmiş; bu renk verisine göre arka plan bölgeleri imgeden ayırt edilmiştir. Ardından derinlik bilgisi ile cildin benzeri tüm gürültü bölgeleri çıkarılmış ve imgede iyileşme sağlanmıştır. Son olarak morfolojik yöntemler kullanılarak, eller ve kollar bölütlere ayrılmıştır.

Nguyen ve Le (2015) tarafından yapılan çalışmada, Kinect ile alınan el hareket imgeleri SVM ve RVM ile sınıflandırılmıştır. Çalışmada iskelet verileri kullanılmıştır. İskelet çerçevesindeki nokta hızı, eklem açısı ve eklem açı hızı olmak üzere üç farklı öznitelik verileri, SVM ve RVM sınıflandırma algoritmaları ile sınıflandırılmıştır. Elde edilen sonuçlar karşılaştırılmıştır.

Diğer bir ilgili çalışma ise Song ve arkadaşları (2013) tarafından yapılmıştır. Çalışmalarında, bir 3D el hareket tanıma yöntemi önermişlerdir. Kinect ile elde edilen imgelerde derinlik bilgileri dikkate alınmıştır. Ardından nokta bulutuna dönüştürülen el imgeleri bölütlenmiştir. Öznitelik olarak nokta bulutu üzerinde 3D moment invaryantları kullanılmış ve öznitelik vektörleri SVM ile sınıflandırılmıştır. Temel olarak kağıt, taş ve makas olmak üzere üç farklı el hareketine yoğunlaşılmıştır. Başarı performansı olarak %97.7 oranı elde edilmiştir.

Dominio ve arkadaşları (2013) ise yapmış oldukları çalışmada benzer yaklaşımlar izlemişlerdir. Ancak çalışmalarında ayrıca parmak ve avuç içi derinlik bilgisini de dikkate almışlardır. Araştırmada iki tür el pozu özelliklerini birleştiren bir el hareket tanıma yöntemi önerilmiştir. Sınıflandırmada SVM kullanılmış ve ortalama olarak, %97.5 oranında başarı performansı elde edilmiştir.

Dong ve arkadaşları (2011) tarafından Kinect ile elde edilen derinlik verileri kullanılmış ve alfabe işaretlerini tanıyabilen bir yöntem önerilmiştir. Her piksel için sınıflandırma algoritmasına dayanan bir derinlik-kontrast özelliği kullanılarak, el imgeleri bölütlenmiştir. Son olarak, çıkarılan özelliklerin sınıflandırılması için Random

Forest (RF) algoritması uygulanmıştır. Bu yöntemin performansını doğrulamak için, mevcut veri kümeleri kullanılmış ve 24 statik hareket %90 oranında doğru saptanmıştır.

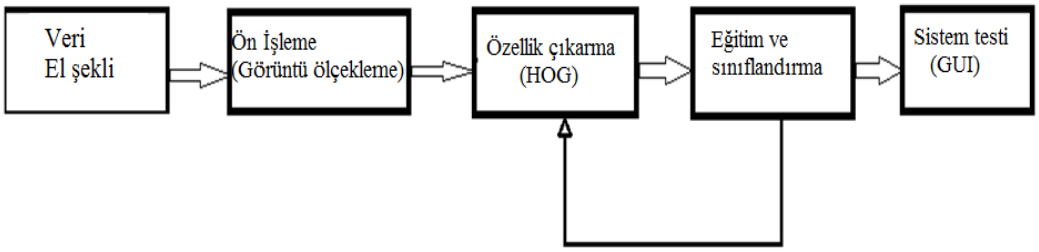
Bu tez çalışmasında da Kinect kullanılarak el hareket imgeleri kayda alınmıştır. Her el hareketi, bir karakteri simgeleyecek şekilde etiketlendirilmiştir. Her el hareketine ilişkin elde edilen öznitelikler farklı sınıflandırıcılar ile sınıflandırılmıştır.

2. KAYNAK ÖZETLERİ

3. MATERYAL VE METOT

Bu tez çalışmasında, MS Kinect ile el hareketlerinin komut olarak algılanması amaçlanmıştır. Bu bölümde ise söz konusu çalışmada kullanılan yöntemler ve önerilen yaklaşım anlatılmıştır. Tez kapsamındaki amacı gerçeklemek için önerilen yaklaşımında izlenen işlem akışı Şekil 3.1'de gösterilmiştir. Şekilden görüldüğü üzere, işlem akışı beş ana aşamadan oluşur; El Algılama, Önişleme, Özellik Çıkarma, Sınıflandırma ve sistemin testidir. Çalışmada, Kinect 2.0 ile ham veri akışları ve iskelet bilgiler kullanılmıştır.

Ayrıca önceki bölümlerde açıklandığı üzere; Kinect donanımında RGB kamera ve tam gövde 3D hareket yakalama kapasitesi sağlayan bir derinlik sensörü bulunur. Çalışmada bu donanımlardan elde edilen veriler ile birlikte, iskelet veri akışı değerlendirmeye alınmıştır. Söz konusu tüm işlem adımları, alt bölümlerde ayrıntılarıyla açıklanmıştır.



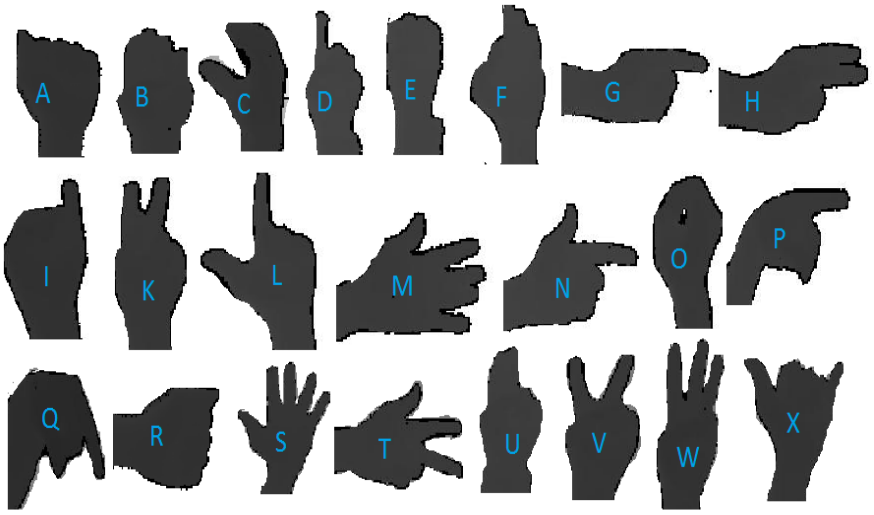
Şekil 3.1. Tez kapsamında önerilen yöntemin işlem akışı

a) El Hareketinin Algılanması

Yaklaşımda el derinliği görüntüsünü elde etmek için çeşitli adımlar izlenmiştir. Bu adımlar aşağıda listelenmiştir:

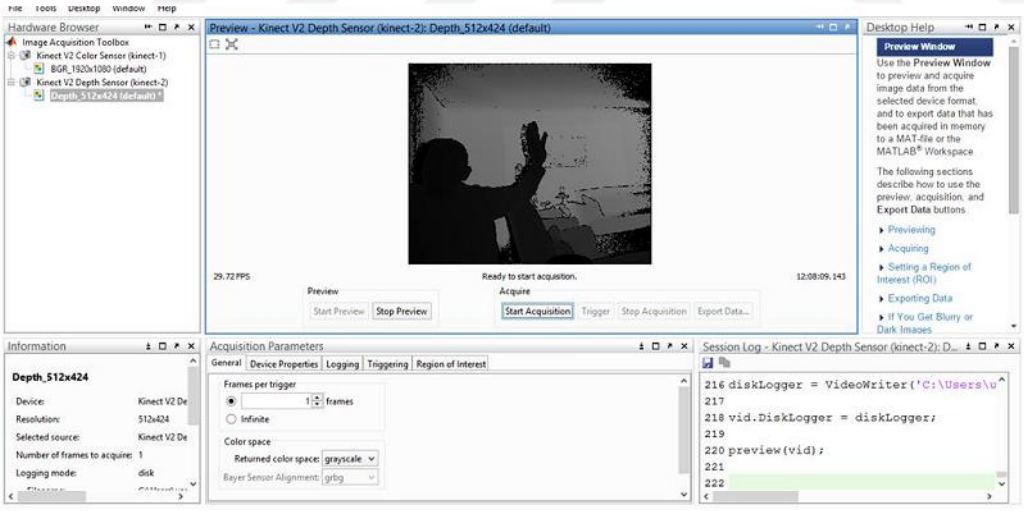
- Derinlik çerçevesinin üretilmesi ve yeniden ölçeklendirilmesi
- Sağ el pozisyonunun çıkarılması
- Sağ eldeki ilgi alanını (ROI) tanımlanması ve bölümlere ayrılması
- Arka planın bölümlenmesi için bir önişlemin yapılması
- El derinliği içeren bir veri setinin oluşturulmasıdır.

Eğitim verilerimizi temsilen kullanılan el hareketlerine ilişkin imgeler, Şekil 3.2'de gösterilmiştir. Veri seti, dinamik el hareketleri oldukları için J ve Z karakterlerini içermez. Karakter setinde, A'dan Y'ye kadar olan tüm karakterler bulunur. Karakter setinde yer alan el figürleri, Amerikan İşaret Dili temelinde yer alan figürlerdir. Ancak performansı artırmak için, birbirine yakın olan bazı figürlerde değişiklikler yapılmıştır. Figürü değiştirilen karakterler sırasıyla A, M, N, S, T ve Y dir.



Şekil 3.2. Çalışmada kullanılan el figürleri için örnek imgeler

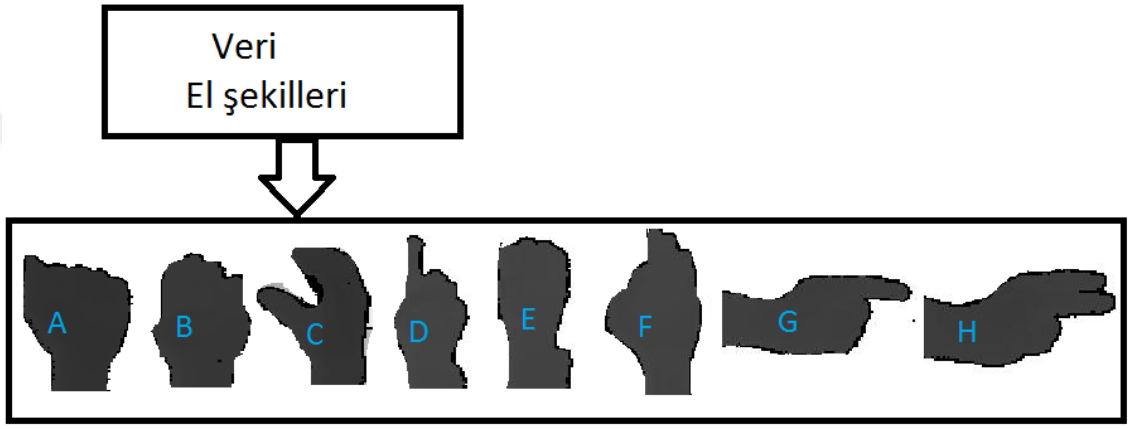
Tez çalışmasında Matlab yazılımı kullanılmıştır. Matlab'da bulunan ilgili araç ile Kinect'den veriler alınabilmektedir. Bu araç ile döngü içi işlem yapılabilmekle birlikte donanım tetiklenebilmekte, arka plan edinilebilmekte ve birden fazla cihaz ile senkronize çalışabilmektedir. Matlab platformunda Kinect üzerinden alınan bir imgenin ekran görüntüsü Şekil 3.3'de gösterilmiştir.



Şekil 3.3. Matlab platformunda Kinect üzerinden alınan bir imgenin ekran görüntüsü

Çalışma akışında, imgeler Kinect ile renk ve derinlik çerçeveleri ile birlikte gerçek zamanlı olarak kayda alınmıştır. Matlab'da yer alan İmge Algılama aracı (Image Acquisition Toolbox) ile Kinect üzerinden saniyede 30 kare imge toplanmıştır. Her kare, iskelet verilerini (depthMetaData) içerir. Ele ilişkin figürün saptanmasında iskelet

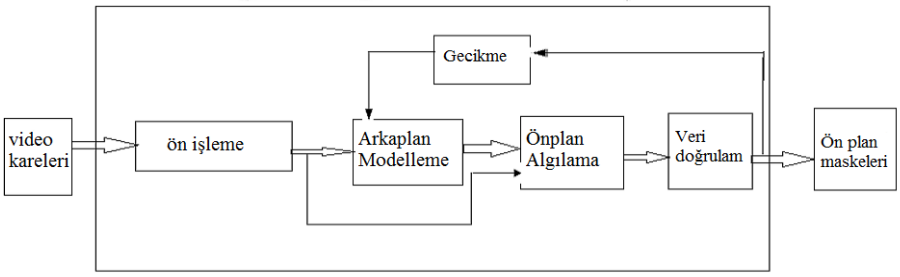
verisinin önemi oldukça büyük olduğu çalışmalarda görülmüştür. İskelet verileri, sağ ele ilişkin ortak pozisyonlarının belirlenmesini sağlamıştır. Kitle merkezi temel alınarak imgenin İlgi Alanı (ROI) belirlenmiştir. Ardından yansıma ve diğer faktörlerden kaynaklanabilecek etkileri azaltmak için arka plan çıkarma işlemi uygulanmıştır. Arka plan çıkarma işlemi bir sonraki alt bölümde anlatılmıştır. Bu işlemlerin sonunda, her imgenin derinlik verileri (handDeptImage) üretilmiş ve sonraki işlemler için depolanmıştır. Söz konusu akışta sonra elde edilebilecek örnek bir el figürü şekil 3.4'de gösterilmiştir.



Şekil 3.4. Arka plandan ayrılmış, derinlik bilgisi içeren bir örnek figür

b) Önişleme (Arka plan çıkarma)

Kinect ile derinlik algılamanın kalitesi doğal olarak gürültülü ve doğal koşullardan büyük ölçüde etkilenebilmektedir. Derinlik ölçümleri genellikle dalgalı formdadır ve derinlik haritalarında çok sayıda boşluk olabilmektedir. Kinect tarafından çekilen derin görüntülerin içinde, sensör tarafından ölçülmeyen tüm noktalar 0 olarak alınır. Ancak, gerçek derinlik değerini tahmin etmek için, eksik olan noktanın komşuları dikkate alınarak derinlik bilgisi tahmin edilebilmektedir. Bu varsayımla, eksik piksellerin değerleri belirlenmiştir. Bu alt çalışma, çerçeve farklılaştırma yöntemi olarak da adlandırılır. Ardından, derinlik algılamasındaki dalgalanmayı ortadan kaldırmak için derinlik dizisi üzerinde 4x4'lük bir pencere ile medyan filtre uygulanmıştır. Arkaplan çıkarmada izlenen genel akış diyagramı şekil 3.5'de gösterilmiştir.



Şekil 3.5. Arka plan çıkarma algoritmasının izlenen genel akış diyagramı

Yukarıdaki işlem akışını matematiksel olarak açıklamak için V(x,y,t) fonksiyonunu bir video dizisi olarak dikkate alalım. Burada t zamanı, x ve y ise piksel konumunu ifade eder. İmge algılama algoritması, bölütleme ile başlar. Bu işlem ile ön plan ve arka plan bölütlenir. Çalışmamızda, ilgi bölgesinin (sağ) tespiti için kitle merkezi kullanılmıştır. Bunu uygulamak için öncelikle bir resim, arka plan olarak alınmıştır. Ardından t anında alınan imge I(t) olarak, arka plan görüntüsü ise B olarak tanımlanmıştır. İmge çıkarma tekniği ile aritmetik hesaplamalar kullanılarak, bölütler imgelerden basitçe çıkarılmıştır. Şöyle ki I(t)'deki her piksel, P[I(t)] ile piksel değerini alır ve P[B] olarak belirtilen arka plan piksel değerinden çıkarılır. Çıkrama işlemi denklem (3.1)'de gösterilmiştir.

$$P[F(t)] = P[I(t)] - P[B] \tag{3.1}$$

Arka plan, t zamanındaki çerçeve olarak kabul edilir. Bu fark imgesi sadece iki karede değişen piksel konumları için bir miktar yoğunluk gösterir. Arka planın çıkarılmış olmasına rağmen; bu durumun, tüm ön piksellerin hareket ettiği ve tüm arka plan piksellerinin statik olduğu durumlar için geçerli olduğu görülmüştür. Çıkarmayı iyileştirmek için bu fark imgesi bir eşik değeri ile karşılaştırılmıştır. Buna göre yazılabilecek ifade denklem (3.2)'de gösterilmiştir.

$$[P[F(t)] - P[F(t+1)]] > Eşik \tag{3.2}$$

Bu, piksel yoğunluk farkının eşik değerine dayandırıldığı anlamına gelir. İmgenin hareket hızı, söz konusu işlem akışının doğruluğunu belirler. Bu da daha hızlı hareketler için daha yüksek eşiklerin gerekli olduğu anlamına gelir. Ardından, derinlik dizisine ortalama bir filtre uygulanarak işlem akışına devam edilir. Arka plan içeren görüntülerin alınması amacıyla, çeşitli resimlerin ortalamaları alınır. t anındaki arka plan görüntüsünü hesaplamak için

$$B(x,y,t) = \frac{1}{N}\sum_{i=0}^{N} V(x,y,t-i) \quad (3.3)$$

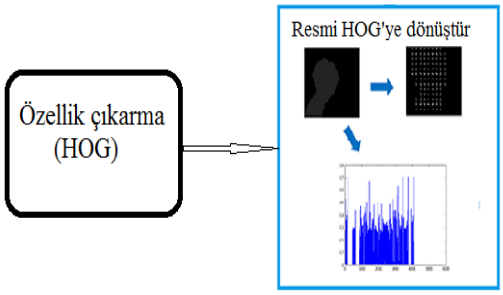
denklemi kullanılmıştır. Burada N, ortalama için alınan toplam görüntü sayısını ifade eder. Denklemdeki ortalamada, verilen imgelerin tüm ilgili piksellerinin toplamı yer alır. N terimi video hızını ifade etmekle birlikte, aynı zamanda saniyede alınan kare sayısını da ifade eder. Arka plan B(x,y,t) hesapladıktan sonra, t anı için V(x,y,t) imgesinde çıkarılır ve eşik değeri uygulanır. Bu yaklaşım denklem (3.4)'de gösterilmiştir.

$$[V(x,y,z) - B(x,y,t)] > Eşik \quad (3.4)$$

Bu işlem sonunda yumuşak geçişlere sahip imgeler elde edilmiş olunur. Bu aşamada sonra imgelerden özellik çıkarım işlemlerine başlanır.

c) El İmgelerine İlişkin Özelliklerin (Özniteliklerin) Çıkarılması

Bu çalışmada, Matlab tarafından desteklenen Yönlü Gradyan Histogram (Histogram of Oriented Gradients, HOG) algoritması kullanılmıştır. HOG ile imgedeki yönlendirilmiş eğrilerin oluşumlarının sayısı elde edilir. Tüm el hareketleri farklı yönlere sahip olduğundan, görüntü odaklı gradyanın HOG ile temsil edilebileceği görülmüştür. Dolayısıyla, bir el figüründe eşsiz odaklı eğimler söz konusu olacağından, her imgenin de benzersiz bir histograma sahip olacağı varsayılmıştır. Bu varsayımı dikkate alarak, histogram verisi ile el figürlerinin komut olarak belirlenmesi, bu tez hedefini oluşturmaktadır. Bir imgeye ilişkin örnek bir histogram, şekil 3.6'da gösterilmiştir. Bu çalışma ile el imgesindeki özellikler yönlü gradyan histogramı (HOG) elde edilmiş olunur. Bir histogram, grafiksel olarak sayısal verilerin dağılımını temsil eder. Bu yaklaşımla; nesnenin kaba şekli, geometrisi ve aydınlık düzeyi belirli bir seviyede



Şekil 3.6. İmgeye ilişkin örnek bir histogram görüntüsü

HOG ile nesne tanımlamanın, diğer yöntemlere göre bazı avantajlara vardır. HOG tanımlayıcı, yerelleştirilmiş hücreler üzerinde çalışır. Bu durum, HOG tanımlayıcı tarafından geometrik ve fotometrik dönüşümlerde sabit niceliklerin oluşturulmasına olanak sağlar. Ayrıca; kaba uzamsal örnekleme, hassas oryantasyona (yönlendirmeye) ilişkin örnekleme ve güçlü yerel fotometrik normalizasyon, nesnelerin dik bir konumu sürdürmesi halinde bireysel vücut hareketlerinin göz ardı edebilmesini sağlar. Yukarıdaki nedenler göz önüne alınır ise insan algılaması için HOG tanımlayıcılarının görüntülerde kullanılması uygun olduğu görülmektedir. Bu algoritmaların ayrıntılı açıklaması aşağıda listelenmiştir.

i) Renkli görüntüler gri ölçeğe dönüştürülür. İlk işlem, renk bilgisinin en aza indirgenmesini hedefler.

ii) Her bir pikseldeki parlaklık gradyanı hesaplanır. Ardından, her bir hücre için bir gradyan yönlendirme histogramı oluşturulur. Bu işlem, şekil değişikliklerine karşı güçlü olan özellik kazandırır.

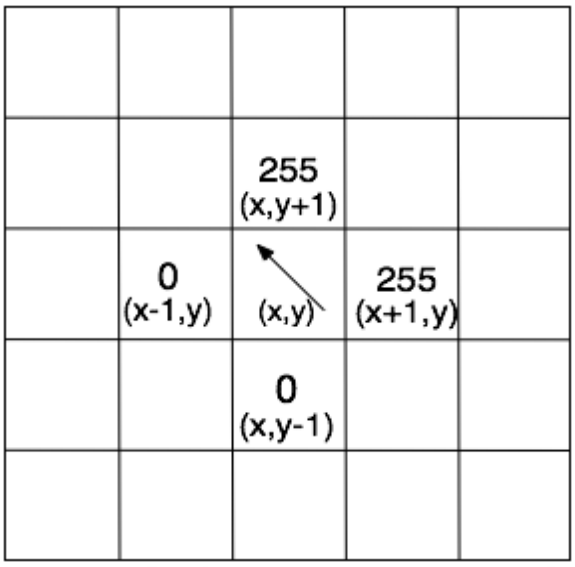
iii) Her bloğa ilişkin özellikler normalleştirilir. Bu işlem, özelliklerin aydınlatma değişikliklerine karşı dayanıklılığını artırır.

Her pikselin parlaklık gradyanı şöyle hesaplanır: Parlaklık gradyanı bir vektör olarak ifade edilmekle birlikte, parlaklık genliği ‘m’ ve oryantasyon (yönlendirme) ‘θ’ parametreleri cinsinden tanımlanır. Parlaklığın genliği olan m parametresinin görüntüdeki (x, y) koordinat sistemine göre tanımı, denklem (3.5)’de verilmiştir. Parlaklık yönlendirme ise (3.6) denkleminde olduğu şekilde tanımlanır. Her iki denklemde yer alan L, pikselin parlaklık değerini simgeler.

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \tag{3.5}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) - \frac{\Pi}{2} < \theta < \frac{\Pi}{2} \tag{3.6}$$

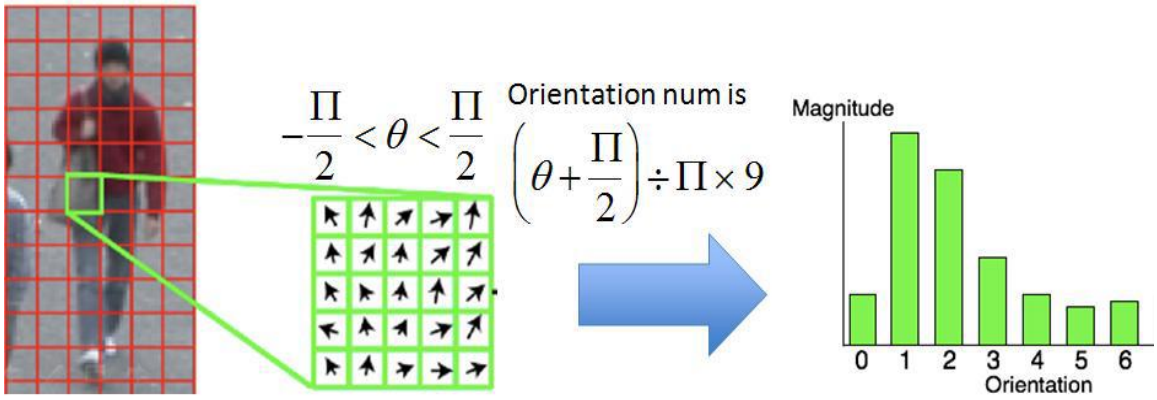
Tüm piksellere (3.5) ve (3.6) denklemleri uygulanması ile elde edilebilecek bir örnek, şekil 3.7’de gösterilmiştir.



Şekil 3.7. Parlaklık yönlendirme ile ilişkili bir örnek uygulama

Bu hesabın ardından, hücre alanında gradyan histogramı oluşturulur. Hücre alanı, 5x5 pikselden oluşturulur. Şekil 3.8'de, imgenin hücrelere nasıl bölündüğü konusunda fikir vermektedir. Gradyanın büyüklüğü ve yönünü kullanarak, parlaklık gradyan histogramı her bir hücre bölgesi için oluşturulur. Bu sırada yönlendirme bölmeleri (bin), 20° derecelik eşit aralıklarla 0° - 180° arasında olacak şekilde konumlandırılır. Bu durumda, bir hücrenin özellik vektörü dokuz boyut ile sınırlandırılır. İlgili hücreye gelen

bir örnek uygulama ve elde edilen histogram şekil 3.8'de gösterilmiştir.



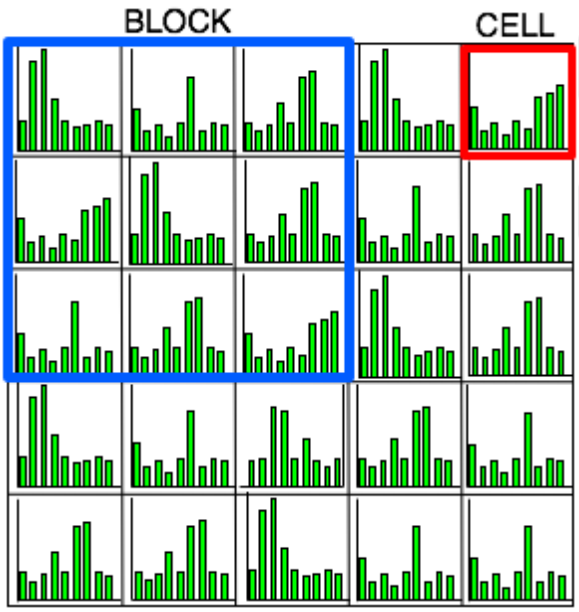
Şekil 3.8. Bir örnek görüntüdeki gradyanlar ve histogramın oluşumu

Bu işlem, nesnenin şeklindeki özellik miktarına ilişkin değişikliği güçlü bir şekilde sağlar. Bu durum, kenar çevirisi değişse bile hücre histogramının değişmemesinden kaynaklanır. Gradyan mukavemetleri geniş bir yelpazede değişir ki bu

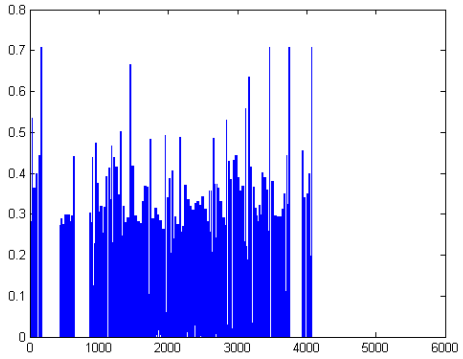
durum aydınlatmada ve ön-arka plan kontrastında yerel farklılıklara neden olur. Bundan dolayı, işlem normalizasyon gerektirir.

Çalışmamızda (3.7) denklemi ile normalizasyon işlemi yapılmıştır. v(n) bu çizginin uzunluğunu temsil eder. Bu çizginin uzunluğu, histogramın dikey ekseni ile ilgilidir. Çizginin yönü ise histogramın yatay ekseni ile ilgilidir. Bloktaki bu çizginin uzunluğu toplanır. Histogramın şekli bu işlemle düzenlenir ve aydınlatmadaki değişimlere karşı dayanıklıdır.

$$v(n) = \frac{v(n)}{\sqrt{\left(\sum_{k=1}^{3\times3\times9} v(k)^2\right) + 1}} \tag{3.7}$$



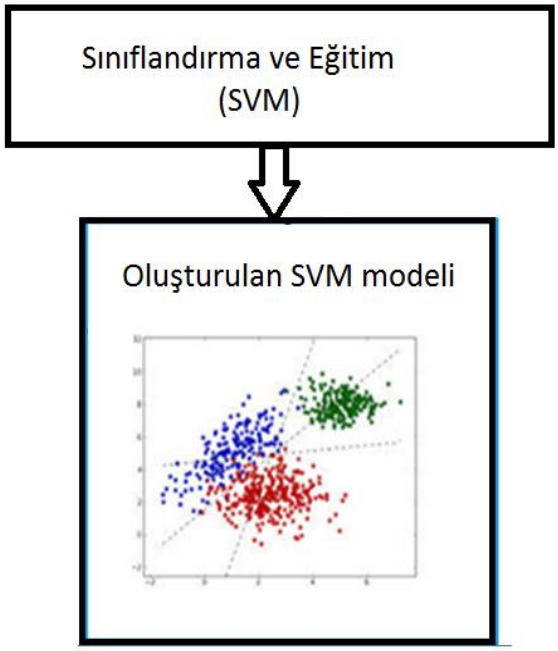
Şekil 3.9. Bloğun (3 × 3 hücre) kaydırılarak tüm imgede uygulanması



Şekil 3.10. HOG kullanarak imgede özellik çıkarımına ilişkin bir örnek histogram

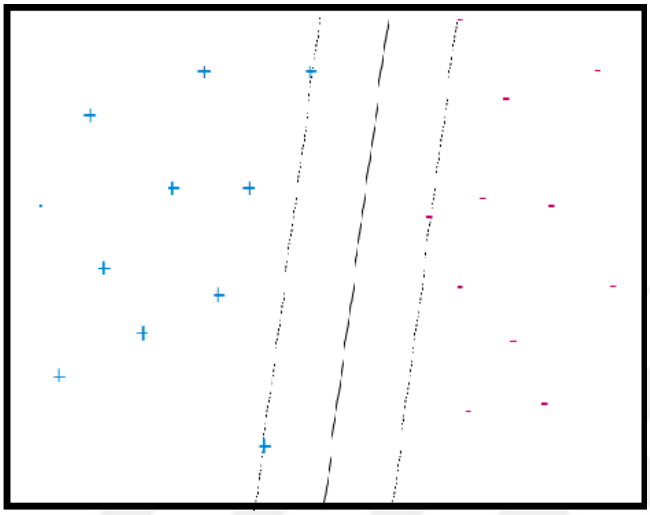
d) Sınıflandırıcı Olarak Destek Vektör Makinesi (SVM)

Bu uygulamanın hangi aşamada yapıldığı şekil 3.11'de gösterilmektedir. Destek Vektör Makinesi (SVM), Vapnik tarafından geliştirilmiş ve ilk defa denetimli öğrenmede Burges, C.J.C. (1998) tarafından kullanılmıştır. Temelde, bu yöntem ile iki sınıf ayrıştırılabilir ve öğrenme için destek vektörü ve çekirdekler kullanılır. Çekirdek tabanlı makine yaklaşımında, çalışılan alan uygun çekirdekler belirlenir.



Şekil 3.11. SVM model üretimi

SVM; diğer makinelerin aksine, sınıflandırma veya diğer görevler için bir hiper düzlem veya bir uzayda hiper yüzey seti oluşturabilir. Sınıflandırmanın içerdiği büyük marj parametresinin yüksek değerlikte seçilmesi durumunda, daha düşük hata ve yüksek genelleme elde edilebilmektedir. Örnek bir optimal hiper düzlem, Şekil 3.12'de gösterilmiştir.



Şekil 3.12. Örnek bir optimal hiper düzlem

SVM (ilkel) optimizasyon problemi için izlenilen yöntem:

$$max \frac{2}{w} subject\ to\ w^T X_i + b \begin{matrix} \geq 1 & if\ y_i = +1 \\ \leq -1 & if\ y_i = -1 \end{matrix} for\ \ i = 1 \ldots . . N$$

Veya eşdeğer

$$\min_w \| w\|^2\ subject\ to\ yi(W^T Xi + b) \geq 1\ fori = 1 \ldots \ldots . . N \qquad (3.7)$$

Bu problemin doğrudan çözülmesi oldukça karmaşıktır. Bu sorun, Lagrange çarpanları kullanılarak çözülebilir ve şeklinde tanımlanabilir.

$$L(w, b, \propto) = \frac{1}{2}\|w\|^2 - \sum_{t=1}^{N} \propto_t \ [y_t(w^t x_t + b) - 1]$$

Where $\propto_t \geq 0$ (3.9)

Yukarıdaki tanımlamada, Karush-Kuhn-Tucker koşulları uygulanır ve

$$maximize \sum_{t=1}^{N} \propto_t - \frac{1}{2} \sum_{s=1}^{N} \sum_{t=1}^{N} \propto_s \propto_t y_s y_t x_s^T x_t$$

$$subject\ to \sum_{t=0}^{N} \propto_t y_t = 0$$

$$where\ \propto_t \geq 0, \forall t\ \in [1, N] \qquad (3.10)$$

ifadesi elde edilir.

Bu ikili problem, ilgili kısıtlamaları daha basit bir şekilde ayarlayabilmektedir. İkili problem, sınırlı bir boyutsal alanı işaret etmesine rağmen, sadece ayrıştırılmış kümeler bu alanda doğrusal olarak sınıflandırılamadığı zaman görülür. Bu nedenle, sınıflamayı kolaylaştırmak için sonlu boyutlu alanın yüksek boyutlu alana dönüştürülmesi önerilmiştir. Bu işlemin sürdürülebilmesi için SVM tasarımının sağladığı olanak; kernel fonksiyonları yardımıyla veriler, orijinal alan değişkenleri türünden hesaplanabilmesidir.

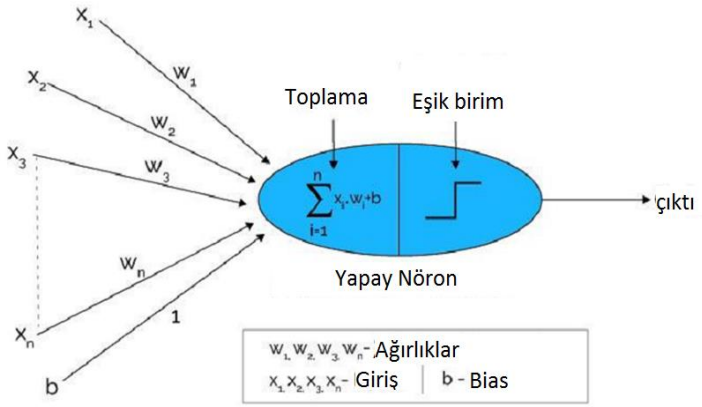
Bundan dolayı, çoklu sınıflandırma işlevini destekleyen LIBSVM yazılımı tez kapsamında kullanılmıştır. Matlab tabanlı LIBSVM yazılımı ile el hareketlerine ilişkin imgeler sınıflandırılmıştır.

d) Sınıflandırıcı Olarak Yapay Sinir Ağları (Artificial Neural Networks)

Sınıflandırma amaçlı kullanılan yapay sinir ağları (YSA) ile elde edilen sonuçlar, diğer sınıflandırıcı yöntemlerin sonuçları ile karşılaştırılmıştır. YSA mimarisi olarak çok katmanlı YSA tercih edilmiş ve en iyi sonuç üretmesi nedeniyle iki gizli katman ile mimari oluşturulmuştur. İki gizli katmanın daha iyi bir performans sergilemesinin nedeni, sınıf sayısının fazla olması ile ilişkili olduğu düşünülmektedir. Sonuçların, Naive bayes gibi sınıflandırıcılar ile karşılaştırıldığında ANN yönteminin daha iyi sonuçlar ürettiği gözlemlenmiştir. YSA yönteminin detayları aşağıda verilmiştir.

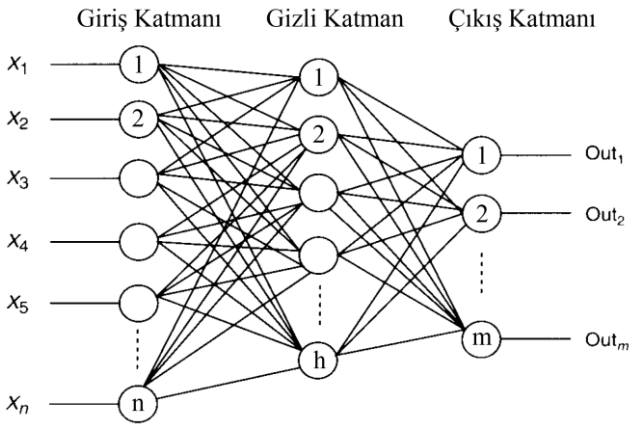
YSA; örüntü tanıma, kümeleme, sınıflandırma gibi bir takım makine öğrenme işlemlerini bilgisayar ortamında simülasyon olarak başarıyla gerçekleştirebilen bir yöntemdir. YSA yaklaşımı, biyolojik yapılardan esinlenilmiştir. YSA'lar, insan beyninin işlevini yerine getirecek nitelikte tasarlanmaya çalışılmıştır. Öğrenme aşamasında nöronlar arasında bulunan inter-nöron olarak adlandırılan bağlantı ağırlıkları uygun

değerler alır ve bilgiler bu ağırlıklar üzerinde depolanır. Bu yaklaşım, YSA yöntemini diğer yöntemlere göre farklı kılar. Şekil 3.13'de bir yapay nöron modeli gösterilmiştir.



Şekil 3.13. Bir yapay nöron modeli (Miklov, (2010)

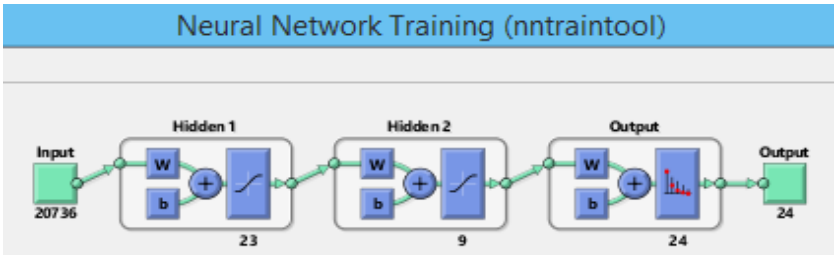
Sinir ağları, ileri ve geri yönlü işaret akışı içeren bir çalışma düzeneği vardır. Yapay ağda bulunan nöronların çevresinden gelen veriler, nöronlar için giriş niteliğindedir. Giriş verisi vektör formunda örüntü veya imge olabilir. ***n*** parametresi giriş sayısı olmak üzere, girişler matematiksel olarak ***x(n)*** şeklinde tanımlanabilir. Her girdi, karşılık gelen ağırlıklar ile çarpılır. Ağırlıklar, belirli bir sorunu çözmek için ağ tarafından kullanılan bilgi olarak tanımlanabilir. Ağırlıklar, bir şekilde sinir ağı kapsamındaki nöronların kendi aralarındaki ara bağlantı kuvvetini temsil ettiklerinden önemlidir. Ağırlıklı girdilerin tümü nöronda toplanır. Ağırlıklı toplam değeri sıfır ise çıktıyı sıfır değeri dışında bir değere taşımak ve sistemin yanıtını ölçeklendirmek için bias eklenir. Bias girişinin ağırlığı 1 olarak alınır. Sistemin yanıtını sınırlamak için aktivasyon fonksiyonu kullanılır. Bir aktivasyon fonksiyonu, istenilen çıktıyı elde etmek için kullanılan transfer fonksiyonları kümesine karşılık gelir. Transfer fonksiyonları soruna bağlı olarak hem doğrusal hem de doğrusal olmayan fonksiyonlar olabilir. Örnek bir YSA mimarisi Şekil 3.14' de verilmiştir.



Şekil 3.14. Örnek bir n:h:m YSA mimarisi

Şekil 3.14'ten görüldüğü üzere, girdiler giriş katmanındaki nöronlar üzerinden alınır. İkinci katman ise giriş ve çıkış katmanları arasında yer alan gizli katmandır. Çıkış katmanından YSA çıkışı elde edilir. Literatürde; Perceptron, Radyal tabanlı sinir ağları, Çok Katmanlı Algılayıcı, Recurrent Sinir Ağları, Uzun / Kısa Vadeli Bellek Ağları (LSTM), Hopfield Network, Boltzmann Makine Ağları, Derin Öğrenme Sinir Ağları gibi farklı mimarilerde sinir ağları görülebilmektedir. Tez kapsamında literatürde sıkça kullanılan ve performansı yüksek olan Çok katmanları algılayıcı mimarisi tercih edilmiştir. Örnek bir çok katmanlı algılayıcı mimarisi şekil 3.14'de gösterilmiştir. Giriş katmanında bulunan nöron sayısı doğrudan giriş örüntüsünün boyutu ile ilişkilidir. Çıkış

sonucunda en iyi sonucun 9 nöronla elde edildiği görülmüştür. Buna göre tez kapsamında kurulan mimari yapı 23:9:24 şeklindedir. Matlab platformu ile elde edilen mimari şekil 3.15'de gösterilmiştir.



Şekil 3.15. Önerine YSA mimarisi

Şekil 3.15'den görüldüğü üzere, her giriş örüntüsünün 23 pikselden oluşması nedeniyle giriş katmanındaki nöron sayısının 23 olduğu görülmektedir. 24 farklı karakter

sınıfı olması nedeniyle de çıkış katmanında 24 nöronun yer aldığı görülmektedir. Giriş kısmına uygulanan örüntü sayısının ise 20736 olduğu görülmektedir.

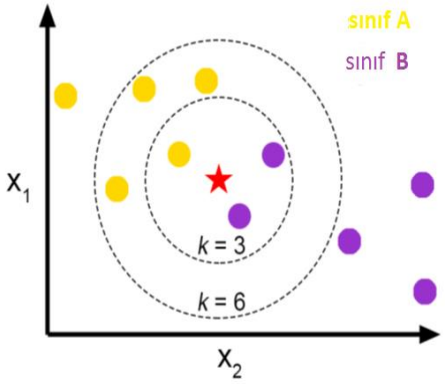
e) Sınıflandırıcı Olarak k-En Yakın Komşu Algoritması (kNN)

Veri setinin sınıflandırılmasında; SVM ve YSA yöntemleri ile birlikte kNN sınıflandırıcı da kullanılmıştır. Uygulanması kolay ve performansı yüksek olan bu yöntemin detayları aşağıda açıklanmıştır.

kNN yaklaşımı, komşu örüntülerin arasındaki mesafe ile sayısı arasındaki ilişkiye dayanan bir algoritmadır. Diğer bir değişle, örüntülerin bir biri arasında yakınlık ilişkilerine göre kümeleme işlemi yapılır. Doğrusal ayrıştırma yöntemi ile koordinat düzleminde çalışır. Çoğu araştırmacı, kullanımı basit ve etkili olması nedeniyle, kNN yöntemini diğer makine öğrenme algoritmalarına tercih eder.

kNN algoritmaları, verilerin yapısına duyarlıdır. Duyarlılığı artırmak, doğrudan mesafe metriği ve komşu sayısı gibi parametreleri iyi seçmek ile ilişkilidir. Örüntülerin boyutları düşük olduğu sürece algoritmanın daha iyi sonuçlar ürettiği bilinmektedir (Russel, S. J. ve Norvig, P., 2010).

kNN; sınıflandırılması istenen bir verinin, en yakın k komşuluktaki verileri temel alan bir yöntemdir. Bu algoritmada sınıfı bilinmeyen bir verinin, k en yakın komşulara olan benzerlikleri hesaplanır ve sınıflara ataması gerçekleştirilir. k en yakın komşu algoritmasında eğitim aşaması çok hızlıdır fakat test kısmı hem zaman hem de bellek açısından maliyetlidir. Örneğin k = 3 için yeni bir eleman sınıflandırılmak istenir ise bu durumda eski sınıflandırılmış elemanlardan en yakın 3 tanesi alınır. Bu elamanlar hangi sınıfa dahil ise, yeni eleman da o sınıfa dahil edilir. Sınıfın belirlenmesini kolaylaştırmak için ayrıca k değeri tek sayı olarak dikkate alınır. k değeri çoğunlukla denenerek saptanabilmektedir. Örneğin k=3 ve k=6 değerleri alınması durumunda, örüntünün ne kadar doğru sınıflandırılabildiğini şekil 3.16 da gösterilmiştir.



Şekil 3.16. kNN ile sınıflamada k parametresinin önemi

Mesafe hesabından ise genelde Öklid ve Mahalanobis uzaklık ölçütü kullanılır. Bu ölçütler aşağıdaki denklemler ile tanımlanabilmektedir.

Öklid uzaklık ölçütü (3.11) bağıntısı ile tanımlanabilir.

$$d(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1}^{p} \left| x_{ik} - x_{jk} \right|^{\frac{1}{2}} \right)^2 \tag{3.11}$$

Bağıntıda yer alan p değişken sayısı olup, xi ve xj ise gözlem çifti olarak tanımlanabilir. Öklid uzaklığı kullanılarak oluşturulan kümelerin yapısı değişken değerlerine dönüşüm uygulansa da özellik uzayı içerisinde değişmezdir (invariant).

Her biri *p* boyutlu *i*. ve *j*. gözlem vektörleri arasındaki karesel Mahalanobis uzaklığı,

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j) \tag{3.12}$$

uzaklığı hiper-elips yapıda kümeler oluşturur. Mahalanobis uzaklığının hesaplanmasında **S** örneklem kovaryans matrisinin tersinin bulunmasında problemlerle karşılaşılabilir. Verideki değişkenler ilişkili değilse **S** bir birim matris yapısına sahip olur ve Mahalanobis uzaklığı karesel Öklid uzaklığına eşdeğer olur.

4. BULGULAR

Bu tez çalışmasında, MS Kinect ile el hareketlerinin komut olarak algılanması amaçlanmıştır. Bu bölümde ise üçüncü bölümde anlatılan üç farklı sınıflandırma (YSA, SVM ve kNN) yöntemi ile elde edilen sonuçlar sunulmuş ve söz konusu yaklaşımların karşılaştırılması sağlanmıştır. Doğrudan sonuçların verilmesinden önce çalışmada

kullanılan donanım ve yazılımlar hakkında bilgiler verilmiş olmakla birlikte elde edilen veri kümesi hakkında detaylar sunulmuştur.

4.1. Uygulamadaki Donanım ve Yazılım Platformları

Tez kapsamındaki tüm uygulamalar, CPU (Intel Core i5-4200M) 2.50GHz, 8GB RAM ve 500GB sabit disk içeren bir bilgisayarda gerçekleştirilmiştir. Bilgisayarda kurulu olan işletim sistemi ise Windows 8'dir. Bu işletim sisteminin kullanılmasının sebebi ise MS Kinect donanımına ilişkin sürücülerini destekleyebilmesidir. Tez kapsamında imgelerin kaydı için MS Kinect donanımı kullanılmış olup, cihazın tüm teknik detayları tezin birinci bölümünde açıklanmıştır.

Tez kapsamındaki tüm çalışma kodları MATLAB R2016b platformunda yazılmıştır. Bu sürümün kullanılmasının nedeni ise MS Kinect V2'nin Matlab'ın R2014 ve üstü sürümleri ile çalışabilmesidir. Bu yazılımla birlikte Windows SDK V2 yazılım platformu kullanılmıştır. Bu yazılım, MS Kinect'i kullanmak için gerekli olan temel platformdur. Öte yandan, analizler için ise kullanılan diğer yazılım paketleri; SVMlight, Computer Vision System Toolbox, Image Acquisition Toolbox, Image Processing Toolbox, Neural Network Toolbox, Signal Processing Toolbox, Statistics and Machine Learning Toolbox, Bioinformatics Toolbox ve Matlab Wrapper.

4.2. Uygulamadaki Veri Tabanı ve Kullanılan Arayüz

Söz konusu yazılım platformları ile birlikte MS Kinect donanımı kullanılarak el hareketlerine ilişkin imgeler kayda alınmış ve gerekli olan veri tabanı oluşturulmuştur. 18 tanesi Amerikan işaret diline ait olan 24 farklı el hareketlerine ilişkin imgeler, veri tabanını oluşturmuştur. Amerikan işaret dilinin dışında yer alan 6 el hareketi, sınıflandırma algoritmalarının performansını artırmak amacıyla tarafımızdan değiştirilmiştir. J ve Z karakterleri dinamik harf olmaları nedeniyle veri tabanına eklenmemiştir. Veri setinin oluşturulmasında, el boyutlarındaki farklılıklardan dolayı ortaya çıkabilecek farklılıkları gidermek için üç farklı kişiden imgelerin kaydı alınmıştır. Her el hareketine ilişkin 100 imge oluşturulmuş ve dolayısıyla toplam 2400 derinlik görüntüsü oluşturulmuştur. Tez kapsamında kurulan bu veri tabanı ile birlikte önerilen sistemin eğitimi/testi için internet üzerinden elde edilen hem yapay hem de gerçek veri setleri kullanılmıştır. İnternet üzerinden elde edilen veri setleri ile önerilen sistemin performansı değerlendirilmiştir.

Çalışmalar gerçek zamanlı yürütülmüştür. Kinect saniyede 30 kare alır ve böylece doğru tanıma çerçevesi sayılır. Çalışmada kullanılan arayüz (GUI) dört penceren oluşmaktadır. Üst sol pencere, Kinect'in aldığı orijinal renkli görüntüsünü yansıtır. Üst sağ pencere, görüntünün derinlik bilgisini içerir. Derinlik bilgisi ile vücudun gövde ve el bölgeleri ayrıt edilebilmektedir. Kinect insanı tespit ettiğinde, derinlik verilerinden 25 iskelet noktası üretir. Sağ el iskelet noktaları (12) kullanılarak, sağ elin konumu belirlenir ve elin bulunduğu pencere segmente edilir. Segmente edilen bu imge, alt sol pencerede yansıtılır. Elin belirlendiği bu segmentler, bu tez kapsamında sınıflandırılarak, hangi komutu veya karakteri simgelediği saptanmıştır. Sınıflandırıcıların belirlediği sonuçlar ise arayüzün alt sağ penceresine yansıtılır. Farklı el hareketlerine ilişkin elde edilen sonuçlar şekil 4.1'de gösterilmiştir.

Gerçek zamanlım (real time) olarak alınan imgelerde yer alan el hareketlerinin komut olarak algılanması için önerilen yaklaşımlarda, performansı belirlemek için

$$Başarı\ performansı\ (\%) = \frac{Doğru\ sınıflandırılan\ \text{örüntü}\ sayısı}{Toplam\ \text{örüntü}\ sayısı} \cdot 100$$

bağıntısı kullanılmıştır, Kulkarni ve Lokhande, (2010).

4.3. SVM Yöntemi ile Elde Edilen Sınıflandırma Sonuçları

Çoklu sınıflandırmaya ihtiyaç duyulduğundan dolayı, LIBSVM yazılımı kullanılmıştır. Uygulamanın gerçeklenmesi ile elde edilen sonuçlar Tablo 4.3'de gösterilmiştir. Ortalama başarımın ise %95.83 olduğu görülmüştür.



Şekil 4.1. Farklı el hareketlerine ilişkin örnekler ve kullanılan arayüz formu

Çizelge 4.1. SVM ile el hareketlerinin komut olarak algılanmasında elde edilen performans değerleri

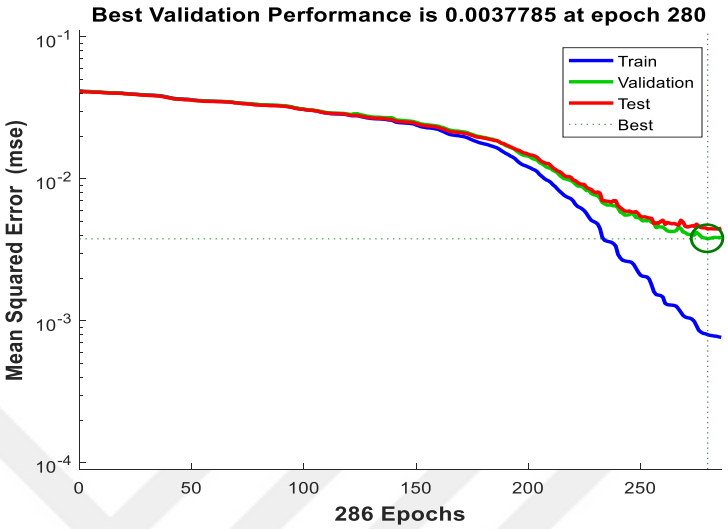
Karakter/Komut	Doğru Sınıflandırılan	Yanlış Sınıflandırılan	Başarı Performansı (%)
A	10	0	100
B	8	2	80
C	10	0	100
D	10	0	100
E	10	0	100
F	9	1	90
G	9	1	90
H	10	0	100
I	10	0	100
K	10	0	100
L	9	1	90
M	10	0	100
N	10	0	100
O	9	1	90
P	10	0	90
Q	10	0	100
R	9	1	90
S	10	0	100
T	10	0	100
U	10	0	100
V	10	0	100
W	10	0	100
X	9	1	90
Y	8	2	80
Ortalama Başarı Performansı			*95.83*

4.4. YSA Yöntemi ile Elde Edilen Sınıflandırma Sonuçları

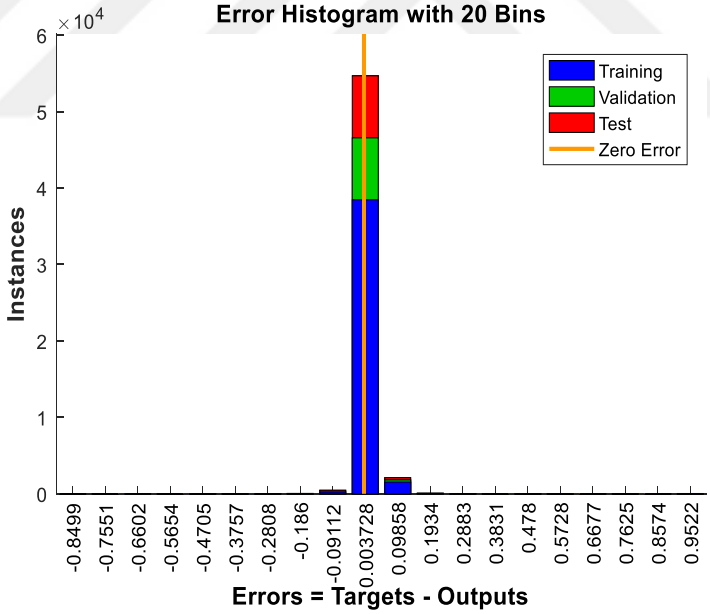
Çok katmanlı algılayıcı için eğitim fonksiyonu olarak scaled conjugate gradient backward propagation yöntemi kullanılmıştır. Ortalama karesel hata değeri, sınıflandırıcı için performans olarak değerlendirilmiştir. Veri seti; eğitim, test ve doğrulama setlerine bölünmüştür. Örüntüler rastgele olarak seçilmiştir. Farklı YSA mimarileri denenmiş ve en iyi sunucun detayları bu bölümde verilmiştir.

İki gizli katmalı bir mimari yapı ile en iyi sonuçların elde edildiği görülmüştür. Giriş katmanında 24 nöron, 1. gizli katmanda 23, 2. gizli katmanda 9 ve çıkış katmanında ise 24 nöron kullanılarak YSA mimarisi belirlenmiştir. Bu yapı ile elde edilen en iyi ortalama karesel hata değerleri; genel performans için 0.0018, eğitim performansı için

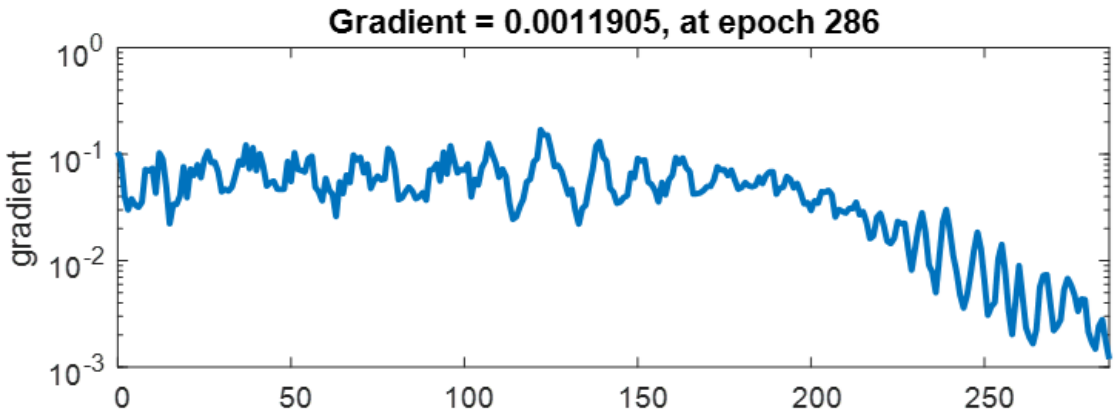
7.9794e-04, doğrulama performansı için 0.0038 ve test performansı için 0.0044 olmuştur (Şekil 4.2). Hata histogramı Şekil 4.3'de ve eğitim aşamasındaki gradyan değişimi şekil 4.4'de gösterilmiştir.



Şekil 4.2. 280 iterasyon ile elde edilen başarı performansları



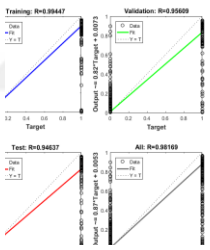
Şekil 4.3. 20 bölmeli skalada hata histogramı



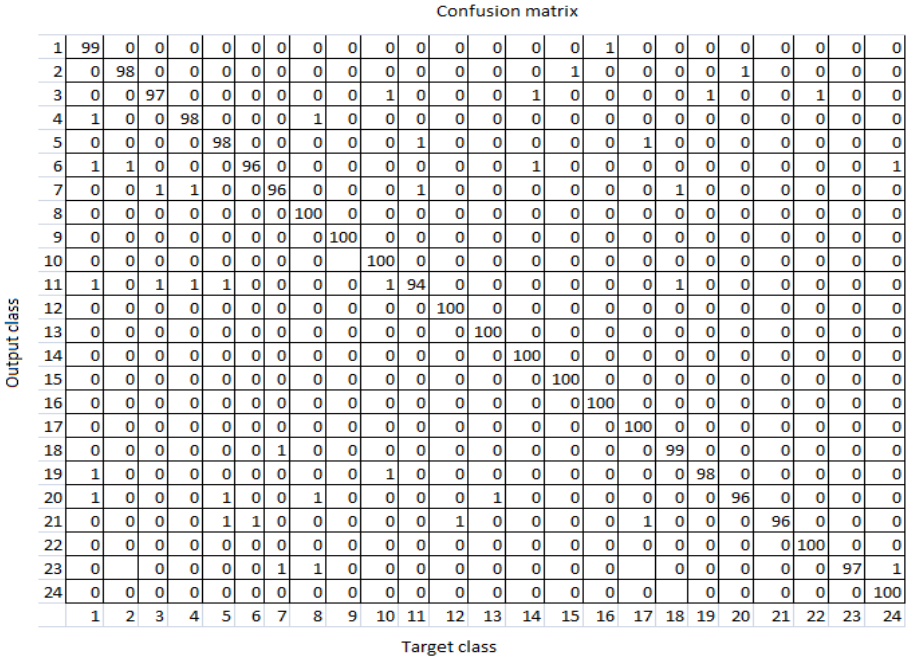
Şekil 4.4. Eğitim aşamasında iterasyonla birlikte görülen gradyan değişimi

YSA'nın eğitilmesi ile elde edilen ağırlıkların kullanılması ve elde edilen sonuçların saçılım grafikleri, önerilen yapının ne denli iyi olduğu konusunda fikir verebilmektedir. Bu konuda elde edilen saçılım grafikleri şekil 4.5'de gösterilmiştir. Tüm setler için korelasyon katsayısının (R), 1 değerine yakın olduğu görülmektedir. Elde edile hata matrisi (Confusion matrix) ise şekil 4.6'da gösterilmiştir.

Tüm karakterlerin/komutların sınıflandırılmasına ilişkin detaylı performans değerleri Çizelge 4.2'de gösterilmiştir. Ortalama başarımın ise %97.5 olduğu görülmüştür.



Şekil 4.5. Her bir set için elde edilen saçılım grafikleri



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	0	98	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
3	0	0	97	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0
4	1	0	0	98	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	98	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
6	1	1	0	0	0	96	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
7	0	0	1	1	0	0	96	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0		100	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	0	1	1	1	0	0	0	0	1	94	0	0	0	0	0	0	1	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0
18	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0
19	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	98	0	0	0	0	0
20	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	96	0	0	0	0
21	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	96	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
23	0		0	0	0	0	1	1	0	0	0	0	0	0	0	0		0	0	0	0	0	97	1
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	100

Şekil 4.6. Elde edilen hata matrisi (Confusion matrix)

Çizelge 4.2. YSA ile el hareketlerinin komut olarak algılanmasında elde edilen performans değerleri

Karakter/Komut	Doğru Sınıflandırılan	Yanlış Sınıflandırılan	Başarı Performansı (%)
A	10	0	100
B	9	1	90
C	9	1	90
D	10	0	100
E	10	0	100
F	10	0	100
G	10	0	100
H	10	0	100
I	9	0	100
K	10	0	100
L	10	0	100
M	10	0	100
N	10	0	100
O	9	1	90
P	9	1	90
Q	10	0	100
R	10	0	100
S	10	0	100
T	10	0	100
U	10	0	100
V	9	1	10
W	10	0	100
X	9	1	90
Y	10	0	100
Ortalama Başarı Performansı			*97.5*

4.5. kNN Yöntemi ile Elde Edilen Sınıflandırma Sonuçları

kNN yöntemi, parametrik olmayan (non-parametric) bir istatistiksel sınıflandırma tekniğidir. Yöntem, yeni örüntünün sınıfını belirlerken en yakın komşuluk değerlerini dikkate alır. En iyi performansı elde etmek için parametrelerin belirlenmesi, zamansal açıdan oldukça maliyetlidir. kNN yönteminde de ön plana çıkan noktaların; metrik uzaklık fonksiyonu ve dikkate alınacak komşu (k) sayısıdır. Bu tez kapsamında, K-katlı çapraz doğrulama ile en uygun k değeri belirlenmiştir.

kNN yöntemi ile tüm karakterlerin/komutların sınıflandırılmasına ilişkin detaylı performans değerleri Tablo 4.5'de gösterilmiştir. Ortalama başarımın ise %92.5 olduğu görülmüştür.

Çizelge 4.3. kNN ile el hareketlerinin komut olarak algılanmasında elde edilen performans değerleri

Karakter/Komut	Doğru Sınıflandırılan	Yanlış Sınıflandırılan	Başarı Performansı (%)
A	9	1	90
B	9	1	90
C	10	0	100
D	8	2	80
E	9	1	90
F	10	0	100
G	10	0	100
H	8	2	80
I	10	0	100
K	10	0	100
L	10	0	100
M	10	0	100
N	10	0	100
O	9	1	90
P	7	3	70
Q	9	1	90
R	10	0	100
S	10	0	100
T	9	1	90
U	10	0	100
V	10	0	100
W	8	2	80
X	8	2	80
Y	9	1	90
Ortalama Başarı Performansı			*92.5*

5. TARTIŞMA VE SONUÇ

Bu tez çalışmasında, MS Kinect ile el hareketlerinin komut olarak algılanması amaçlanmıştır. Bu amaç doğrultusunda; işaret dili tanıma, Kinect tabanlı ve Kinect kullanılmadan oluşturulan sistemler ile ilişkili çalışmalar incelenmiştir. Bu süreçte Kinect donanımının avantajları/dezavantajları ve kameralardaki farklılıklar incelenmiştir. Bu bölümde ise, konu ile ilişkili önemli saptamalar ve önerilen yöntem ile elde edilen sonuçlar, maddeler halinde listelenmiştir.

Önerilen Yöntem ile Elde Edilen Sonuçları

i) Bu tez çalışmasında, Kinect tabanlı sistemler ile karanlıkta çalışılabildiği, sağ el parmaklarının makul bir rotasyona göre tanınabildiği görülmüştür. Ayrıca bu işlemler yapılırken, imgede ön planı arka plandan tamamıyla izole edebildiği gözlemlenmiştir. Dolayısıyla Kinect tabanlı gürbüz sistemlerin oluşturulabileceği görülmüştür.

ii) Tez kapsamında toplam 24 el hareketi içeren imge seti kullanılmıştır. Bu el hareketlerinin 18 tanesi Amerikan İşaret Dili kapsamında yer alan el hareketleridir. Her el hareketine ilişkin 100 imge kayda alınmıştır. Sistemin güvenirliliğini artırmak amacıyla her hareket için oluşturulan 100 imge için üç farklı kişi görev almıştır. Böylelikle toplamda 24000 imge değerlendirilmeye alınmıştır.

iii) Her imgeden çevrimdışı (off-line) RGB ve derinlik bilgileri kayda alınmış ve HOG yöntemi ile imgelerin öznitelikleri çıkarılmıştır. Oluşturulan öznitelikler ile sistemin eğitimi sağlanmıştır. El hareketlerinin komut olarak algılanmasında hareketlerin sınıflandırılması için üç farklı sınıflandırıcı yöntem uygulanmıştır. Bunlar; SVM, Çok katlı YSA mimarisi ve kNN sınıflandırıcılarıdır.

iv) Eğitme işleminden sonra, gerçek zamanlı olarak sistemin testi gerçeklenmiştir. Yapılan denemelerde her komut için 10 el hareketi ile test edilmiş ve doğru/yanlış şeklinde kayda alınmıştır. Elde edilen ortalama başarı performansı Tablo 5.1'de gösterilmiştir.,

Çizelge 5.1 El hareketlerinin sınıflandırılmasında kullanılan yöntemlerin, ortalama başarı performansları

Sınıflandırıcı	Ortalama Başarı Performansı
kNN	% 92.50
SVM	% 95.83
ANN	% 97.50

v) Yapılan tez çalışmasının diğer çalışmalar ile karşılaştırılması Çizelge 5.2'de gösterilmiştir. Elde edilen başarı performansları karşılaştırıldığında, tez kapsamında elde edilen başarının kabul edilebilir düzeyde olduğu görülmüştür.

Çizelge 5.2. MS Kinect kullanılarak elde edilen imgelerdeki el hareketlerinin tespitine ilişkin temel çalışmalar ve tez kapsamında önerilen çalışma ile karşılaştırılması.

Kaynak	Yıl	Yöntem	Sınıflandırma Yöntemi	% Doğruluk	Uygulama/Görev
Tez kapsamında önerilen yaklaşım	**2018**	**Statik**	**HOG, SVM, YSA ve kNN**	**97.50**	**El hareketi modelleme**
Chen ve Wang n.d.	2017	dinamik	RNN	90	El hareketi modelleme
Chong ve ark	2016	statik	Karar ağaçları	94.7	Düzeltici hesaplama
Priyanka.	2016	statik			El hareketi modelleme
Dinh ve ark.	2013	statik	Karar ağaçları	97.8	Düzeltici hesaplama
Dominio ve ark	2013	statik	SVM	97.5	El hareketi modelleme
Dong ve ark.	2011	statik	Rastgele Orman	90	İşaret dili
Li	2010	statik	K-ortalama	99.5	El hareketi modelleme
Molchanov Kim ve ark	2015	statik	ANN	90	Robotik
Nguyen ve Le	2015	statik	SVM ve RVM'ler		El hareketi modelleme
Ren ve arkadaşları	2011	statik		90.6	Düzeltici hesaplama
Song ve ark	2013	statik	SVM	97.7	El hareketi modelleme
Vo ve ark	2017	statik	SVM ve HMM	95	İşaret dili
Xie ve Shen	2015	statik	kNN	99.7	El hareketi modelleme
Keskin ve ark.	2011	statik	SVM	99.9	İşaret dili

vi) Bu tez çalışmasında sadece statik el hareketlerini içeren imgeler kullanılmıştır. Dolayısıyla, dinamik yapısı nedeniyle J ve Z harfi dışındaki karakterleri simgeleyen el hareketleri, çalışma kapsamının dışında tutulmuştur. Bunun sebebi ise HOG yöntemi ile el hareketlerine ilişkin gradyan değişimin, bir dizi olarak elde edilememesidir.

vii) Tez kapsamında elde edilen başarı performansı ile el hareketlerinden oluşan komutların, çevre sistemlerini rahatlıkla kontrol edebileceği kanaati oluşmuştur.

Tez Çalışması ile İlişkili Önemli Saptamalar ve Öngörülen Gelecek Çalışmalar

i) MS Kinect donanımı, nispeten 640x480 gibi iyi bir çözünürlük ile derinlik verilerini sağlayabilmektedir. Çalışma esnasında donanımın derinlik bilgisini sağlayabildiği ancak Matlab'tan kaynaklı olarak nesnenin en az 80cm'lik mesafede olması gerektiği görülmüştür. Bunun sebebi ise Matlab'ın yakın modu desteklememesi olduğu düşünülmektedir. Öbür taraftan kullanılan Kinect donanımı ile el hareketlerinin 0.5m-0.8m aralığında tanınabildiği görülmüştür.

ii) MATLAB'da bulunan arayüzde, çeşitli el hareketlerindeki parmakların benzer şekilde durması, imgelerin ayrıştırılmasını zorlaştırmıştır. Bundan dolayı, imgelere ilişkin özniteliklerin derinlik bilgisine dayandırılarak oluşturulması, başarı performansını artırmıştır.

iii) Bu tez kapsamında olduğu gibi el hareketlerinin tanınmasında yeni yöntemler geliştirilebilir. Bu tez çalışmasında RGB renk alanı ve derinlik bilgisi kullanılmıştır. Ancak parlaklık etkisini ortadan kaldırmak için HSV (Hue, Saturation, Value) renk uzayı kullanılabilir ve böylece el hareketlerini tanıma performansı artırılabilir.

iv) Bu tez çalışmasında sadece statik el hareketlerini içeren imgeler kullanılmıştır. J ve Z karakterleri, dinamik el hareketlerini simgeleyen pozlar olması nedeniyle çalışmaya dahil edilmemiştir. Bu noktada, dinamik el hareketlerinin sınıflandırılması için kaynak özetleri bölümünde belirtildiği gibi HMM

yaklaşımı önerilir. Bundan dolayı, HMM ile sistem performansının artırılması düşünülmektedir.

v) El hareketlerini tanıma performansının artırılması için birden fazla Kinect donanımının kullanılabileceği düşünülmektedir. Hem Kinect hem de diğer türlü kameralara ilişkin özelliklerin birlikte kullanılması durumunda, başarı performansının artırılabileceği düşünülmektedir.

vi) Günümüzde sınırlı hareketlere sahip olmakla birlikte robot kollar ile çalışan sistemlerin varlığı bilinmektedir. Ayrıca Kinect'in kare-kare imge alıp, yüksek doğrulukta el imgelerinin pozisyonu saptayabilecek kapasiteye sahip olduğu bilinmektedir. Dolayısıyla, el hareketlerini robotlara adapte edebilecek ve aralarındaki kütüphanelerin oluşturulmasını sağlayacak Kinect tabanlı yöntemlerin üzerine çalışılması ile hareket sınırlamasının önüne kısmen geçilebileceği düşünülmektedir.

6. KAYNAKLAR

Anon, 2012. Using Kinect for Smart Costume Design. , (August).

Ari, S., 2011. A Static Hand Gesture Recognition Algorithm Using K-Mean Based Radial Basis Function Neural Network. , (i).

Burges, C. J. C. (1998) 'A Tutorial on Support Vector Machines for Pattern Recognition', Data mining and knowledge discovery, 2(2), pp. 121–167. doi: 10.1023/A:1009715923555.

Chen, X. & Wang, G., Motion feature augmented recurrent neural network for skeleton-based dynamic hand gesture recognition.

Chong, Y., Huang, J. & Pan, S., 2016. Hand Gesture Recognition Using Appearance Features Based on 3D Point Cloud. , (April), pp.103–111.

Cook, R. and Cargill, T. (2014) 'Say Hello to My Little Friend : JavaScript !' docs.ipisoft.com/User_Guide_for_Single_Depth_Sensor_Configuration msdn.microsoft.com. (2017) 'Coordinate Space Kinect for Windows 1.5, 1.6, 1.7, 1.8'

D, P.P., 2016. Hand Gesture Identification using Preprocessing , Background Subtraction and Segmentation Techniques. , 11(5), pp.3221–3228.

Dinh, D. et al., Hand Number Gesture Recognition Using the Recognized Hand Parts in Depth Images.

Dominio, F. et al., Hand Gesture Recognition with Depth Data Categories and Subject Descriptors.

Dong, C., Leu, M.C. & Yin, Z., American Sign Language Alphabet Recognition Using Microsoft Kinect.

Fässler, M. & Qi, G., 3D Reconstruction with Kinect – Project Proposal Timeline. , pp.3–4.

Han, J. et al. (2013) 'Enhanced Computer Vision with Microsoft Kinect Sensor : A Review', pp. 1–17.

Haro, L.F.D. et al., 2008. Speech to sign language translation system for Spanish.

Ibañez, R. et al., 2014. Advances in Engineering Software Easy gesture recognition for Kinect. , 76, pp.171–180.

Kainz, O. & Jakab, F., 2014. Approach to Hand Tracking and Gesture Recognition Based on Depth-Sensing Cameras and EMG Monitoring. , 3(1), pp.104–112.

Kang, B., Real-time Sign Language Fingerspelling Recognition using Convolutional Neural Networks from Depth map.

Keskin, C. et al., 2011. Real Time Hand Pose Estimation using Depth Sensors. , pp.1228–1234.

Kishore, P.V. V, Kishore, S.R.C. & Prasad, M.V.D., 2013. Conglomeration of Hand Shapes and Texture Information for Recognizing Gestures of Indian Sign Language Using Feed forward Neural Networks. , 5(5), pp.3742–3756.

Li, Y., 2010. Hand Gesture Recognition Using Kinect. , pp.196–199.

Mieee, P.V.V.K. et al., 2015. 4-Camera Model for Sign Language Recognition Using Elliptical Fourier. , pp.34–38.

Moise, I. et al. (no date) 'K-Nearest Neighbour Classifier Supervised data mining Classification → Decision Trees'.

Molchanov, P., Gupta, S., Kim, K., Kautz, J., et al., Hand Gesture Recognition with 3D Convolutional Neural Networks.

Molchanov, P., Gupta, S., Kim, K. & Pulli, K., Multi-sensor System for Driver ' s Hand-Gesture Recognition.

Motivation, H. (2016) 'Introduction To'. doi: 10.1108/S0749-742320160000019022. www.kdnuggets.com/2016/08/seven-steps-understanding-computer-vision.html

Nagashree, R.N. et al., 2015. Hand gesture recognition using support vector machine. , pp.42–46.

Nguyen, D. & Le, H., 2015. Kinect Gesture Recognition : SVM vs . RVM.

Ong, E. & Bowden, R., A Boosted Classifier Tree for Hand Shape Detection No Not Hand Yes Yes Yes Shape 1 (Yes / No) Detected Shape 1 (Yes / No) Detected Shape C.

Pisharady, P.K. & Saerbeck, M., 2015. Recent methods and databases in vision-based hand gesture recognition : A review. Computer Vision and Image Understanding, 141, pp.152–165. Available at: http://dx.doi.org/10.1016/j.cviu.2015.08.004.

Ren, Z., Meng, J. & Yuan, J., 2011. Robust Hand Gesture Recognition with Kinect Sensor. , pp.1–2.

Ren Z., Yuan J., Meng J., Zhang Z. Robust part-based hand gesture recognition using kinect sensor. IEEE Trans. Multimed. 2013;15:1110–1120.

Review, I., 2016. Hand Motion Gesture for Human-Computer Interaction Using Support Vector Machine and Hidden Markov Model Hand Motion Gesture for Human-Computer Interaction Using Support Vector Machine and Hidden Markov Model. , (May).

Russel, S. J. and Norvig, P. (2010) Artificial Intelligence a Modern Approach. doi: 10.1017/S0269888900007724.

Science, C. & Studies, M., 2015. A Survey of Hand Gesture Recognition. , 7782, pp.266–271.

Smedt, Q. De et al., 2017. SHREC ' 17 Track : 3D Hand Gesture Recognition Using a Depth and Skeletal Dataset.

Song, L. et al., 2013. Real-Time 3D Hand Gesture Recognition from Depth Image. , pp.1134–1137.

Starner, T., Weaver, J. & Pentland, A., 1995. Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video. , (466), pp.1–7.

Vieriu, R., Gora, B. & Gora, L., On HMM Static Hand Gesture Recognition. , pp.3–6.

Vo, D., Huynh, H. & Meunier, J., 2017. Dynamic Gesture Classification for Vietnamese Sign Language Recognition. , 8(3).

Xie, J. & Shen, X., 2015. Hand Posture Recognition Using Kinect. , pp.89–92.

ÖZGEÇMİŞ

1988 yılında Uganda'da doğdum. İlk, orta ve lise öğrenimimi Kampala Uganda'da tamamladıktan sonra Makerere Üniversitesi Teknologe Fakültesi Bilgisayar Bölümünden 2012 yılında mezun oldum. 2012 yılında Uganda'da özel bir bilgisayar okulunda öğretmen olarak çalıştım. 2013 yılında Türkiye'ye gelmeye karar verdim ve 2014 yılında Dicle Üniversitesi Mühendislik Fakültesi Elektrik - Elektronik Bölümünde Yüksek Lisans yapmaya hak kazandım. Evli ve bir çocuk babasıyım.

T.C.
DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ÇALIŞMASI İNTİHAL RAPORU FORMU

ÖĞRENCİ BİLGİLERİ	
ADI VE SOYADI	JULIUS BAMWENDA
ÖĞRENCİ NO	15805003
EĞİTİM – ÖĞRETİM YILI	2017 - 2018
YARIYIL	☒ Güz ☐ Bahar
ANABİLİM DALI	Elektrik-Elektronik Mühendisliği
PROGRAM	Yüksek Lisans
TEZ KONUSU	El Hareketlerinin Kinect ile Komut Olarak Algılanması
İNTİHAL RAPORU BİLGİLERİ	
RAPOR TÜRÜ	Tez Savunma Sınavı Sonrası
SAYFA SAYISI	72
BENZERLİK ORANI	%6
RAPORLAMA TARİHİ	24/03/2018

Yukarıda başlığı/konusu gösterilen tez çalışmamın kapak sayfası, giriş, ana bölümler, sonuç ve tartışma kısımlarından oluşan toplam 72 sayfalık kısmına ilişkin, 24/03/2018 tarihinde ~~şahsım~~/tez danışmanım tarafından *Turnitin* adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan intihal raporuna göre, tezimin benzerlik oranı %6 'tür.

Uygulanan filtrelemeler:

- ☐ Kabul/Onay sayfaları hariç,
- ☐ Kaynakça hariç
- ☐ Alıntılar hariç/dâhil
- ☐ Diğer

Dicle Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Programlarda Tez Çalışması İntihal Raporu Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edilmesi durumunda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

(İmza)
Julius BAMWENDA

(İmza) 26/03/2018 Doç.Dr. Mehmet Siraç ÖZERDEM **Tez Danışmanı**	(İmza) 26/03/2018 Prof.Dr. İbrahim KAYA **Anabilim Dalı Başkanı**