DERİN ÖĞRENME VE ANAHTAR POZ YÖNTEMİYLE İNSAN DURUŞU TAHMİNLEMESİ

HEDİYE NUPELDA KANPAK

HAZİRAN 2022

DİYARBAKIR

DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DERİN ÖĞRENME VE ANAHTAR POZ YÖNTEMİYLE İNSAN DURUŞU TAHMİNLEMESİ

HEDİYE NUPELDA KANPAK

DİCLE ÜNİVERSİTESİ LİSANSÜSTÜ EĞİTİM-ÖĞRETİM VE SINAV YÖNETMELİĞİNİN BİR PARÇASI OLARAK
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALINDA
YÜKSEK LİSANS TEZİ
OLARAK HAZIRLANMIŞTIR

HAZİRAN 2022

DİYARBAKIR

DERİN ÖĞRENME VE ANAHTAR NOKTA POZ YÖNTEMİYLE İNSAN DURUŞU TAHMİNLEMESİ

Hediye Nupelda KANPAK tarafından Dicle Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin bir parçası olarak hazırlanan bu çalışma, aşağıda bilgileri yazılı jüri üyeleri tarafından değerlendirilerek **Elektrik Elektronik Mühendisliği Ana Bilim Dalı**'nda **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Prof. Dr. Neslihan DALKILIÇ
Müdür, **Fen Bilimleri Enstitüsü**

Dr. Öğretim Üyesi M. Ali ARSERİM
Danışman, **Elektrik Elektronik Mühedisliği Bölümü, Dicle Üniversitesi**

Sınav Jürisi:

Dr. Öğr. Üyesi Hüseyin ACAR **(*)**
Elektrik Elektronik Mühedisliği Bölümü,
Dicle Üniversitesi

Dr. Öğr. Üyesi M. Ali ARSERİM **(**)**
Elektrik Elektronik Mühedisliği Bölümü,
Dicle Üniversitesi

Dr. Öğr. Üyesi Emrullah ACAR
Elektrik Elektronik Mühedisliği Bölümü,
Batman Üniversitesi

ONAY

Savunma Tarihi: 13 / 06 / 2022

(*) Sınav Jürisi kısmının birinci satırına Jüri Başkanının bilgilerini yazınız.
(**) Sınav Jürisi kısmının ikinci satırına Tez Danışmanının bilgilerini yazınız.

Dicle Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırlanan bu tez çalışmasında yer alan tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu beyan ederim. Ayrıca, bahse konu bu kural ve ilkelerin gerektirdiği üzere, bu çalışmada özgün olmayan tüm bilimsel içerikleri kurallara uygun biçimde alıntılayıp kaynak gösterdiğimi beyan ederim. Beyanımla çelişen herhangi bir delil bulunduğu takdirde tüm sorumluluğu üstleneceğimi kabul ederim.

Ad, Soyad: Hediye Nupelda KANPAK

İmza:

TEŞEKKÜR

Çalışmam boyunca benden bilgisini ve desteğini esirgemeyen tez danışmanım Dr. Öğr. Üyesi M. Ali ARSERİM'e, teşekkürlerimi sunarım. Akademik çalışmalarımda bana yardımcı olup destek olan tüm çalışma arkadaşlarıma teşekkür ederim.

Hayatımda en özel ve önemli yere sahip olan, başarılarımın alt yapısını oluşturan her konuda örnek aldığım babama, beni her koşulda destekleyip yanımda olan anneme ve evimizin miniği kardeşime teşekkürlerimi bir borç bilirim

İÇİNDEKİLER

TEŞEKKÜR v
İÇİNDEKİLER vi
ŞEKİLLER LİSTESİ viii
TABLOLAR LİSTESİ ix
SİMGELER VE KISALTMALAR LİSTESİ x
ABSTRACT xii
1. GİRİŞ 1
1.1 Problem Tanımı ve Literatür Özeti 1
1.2 Tezin Amacı 5
1.3 Tez Düzeni 6
2. ÖNCEKİ ÇALIŞMALAR 7
3. MATERYAL VE METOT 10
3.1 2B ve 3B Veri Kümeleri 10
3.2 İnsan Poz Tahminlemesi 11
3.2.1 Keras modellemesi ve jupyter not defterleri 12
3.3 Evrişimsel Sinir Ağları 12
3.3.1 Maksimum havuzlama ve dropout katmanları 15
3.3.2 Evrişimsel sinir ağının eğitilmesi 16
3.4 Tahminleme Yöntemleri 17
3.4.1 Doğru pozlama yöntemleri 20
3.4.2 Kişi katmanları yöntemleri 20
3.4.3 Anahtar noktalarının ortalama hassasiyeti yöntemleri 20
3.5 Oluşturulan Veri Seti 21
3.6 Tahminleme Aşamaları 22
3.6.1 Görsellerin alınması ve sistem girişi olarak kullanılması 24

3.6.2 Koşullara göre hareket tanıma 24
3.6.3 İlgi alanı çıkarımı 25
3.6.4 Rgb ve optik akış şekilleri 27
3.6.5 Sıralı yönlü gradyanlar ve destek vektör makinaları 31
4. BULGULAR VE TARTIŞMA 33
5. SONUÇLAR VE ÖNERİLER 40
KAYNAKLAR 42
ÖZGEÇMİŞ 46

ŞEKİLLER LİSTESİ

Şekil 3.1 Çalışmada yer alan veri setlerinden futbol maçlarına ait anlık görüntü tanıma sonucu toplanan veri seti 10
Şekil 3.2 Anahtar noktaları belirleme ve tahmin metodolojisi 11
Şekil 3.3 Evirişimli Sinir Ağı Yapısı 12
Şekil 3.4 Evrişim katmanı 13
Şekil 3.5 Veri kümelerinde veri tespiti için kullanılan Boru Şeması 14
Şekil 3.6 Denetimli Öğrenme Şematiği 17
Şekil 3.7 İnsan Duruşu Tahminlemesinde Yay 19
Şekil 3.8 Hesaplanan anahtar noktalar 22
Şekil 3.9 Tahminleme sisteminin yapısı 23
Şekil 3.10 Verilerin ilgi alanının analiz sonucunda çıkarılması 27
Şekil 3.11 RGB bileşenleri için ortak veri dönüştürme görüntüsü 28
Şekil 3.12 İki Boyutlu Anahtar Noktalar 29
Şekil 3.13. Küresel modelleme sistemi 32
Şekil 4.1 Anahtar noktalar için 16 referans görüntüsü 34
Şekil 4.3 Duruş Tahminlemesi Sonucu Elde Edilen Görüntüler 35

TABLOLAR LİSTESİ

Tablo 4.1 .Doğru Pozlama, Kişi Katmanları ve Anahtar Noktaları Yöntemleri ile Önden Görünüm Doğruluk Oranlarının Kıyaslaması 36
Tablo 4.2. Doğru Pozlama, Kişi Katmanları ve Anahtar Noktaları Yöntemleri ile Sağdan Görünüm Doğruluk Oranlarının Kıyaslaması 37
Tablo 4.3. Doğru Pozlama, Kişi Katmanları ve Anahtar Noktaları Yöntemleri ile Soldan Görünüm Doğruluk Oranlarının Kıyaslaması 38

SİMGELER VE KISALTMALAR LİSTESİ

Simge	**Açıklama**
k	Çekirdek Boyutu
s	Adım
p	Sıfır Doldurma
win	Girişim Uzamsal Boyutu
N	Çıktı Sayısı
yt	Temel Gerçeklik
yö	Öngörülen Çıktı
Q	Parametresel Vektörü
a	Öğrenme Hızı
J(Q)	Parametre Güncellemesi
dI	Parlaklık Şiddeti
α	Parlaklık
ε	Akıcılık

Kısaltma	**Açıklama**
ESA	Evrişimsel Sinir Ağları
ROİ	İlgi Alanı Çıkarımı
2D	İki Boyutlu Sistemler
3D	Üç Boyutlu Sistemler
RGB	Renk Katmanı
HOG	Kenar Özellikleri Çıkarımı
Surf	Boyuttan Bağımsız Çıkarım

ÖZET

DERİN ÖĞRENME VE ANAHTAR NOKTA POZ TAHMİNLEMESİ YÖNTEMİYLE İNSAN DURUŞU TAHMİNLEMESİ

Kanpak, Hediye Nupelda
Yüksek Lisans, Elektrik Elektronik Mühendisliği Bölümü
Danışman: Dr. Öğr.Üyesi M. Ali ARSERİM
Haziran 2022, 59 sayfa

İnsanlara ait video ve resimlerde yer alan insan duruşunu yorumlamak, hangi hareketlerde olduğunun belirlenerek tanınması ve anlamlandırılması insan duruşu tahminlemesinin en temel yapısını oluşturmaktadır. Bu tez çalışmasında resim ve videolarda yer alan görüntüler değerlendirilerek hareketin ne olduğuna, hangi amaçlar doğrultusunda yapıldığına karar veren bir sistem oluşturulmuştur. Böylelikle amaçlanan doğrultuda otomatik bir sistem halinde insan hareketlerinin tespiti yapılarak, istenilen zaman aralıkları belirlenerek, anlamlandırma işlemlerinin bütününü kapsayan bir çalışma yapılmıştır. Eldeki veri setlerinde yer alan görüntülerde ölçüt alınan kriterlerde anlamlandırılma aşamasından sonra hangi hareketlerin yapıldığı sınıflandırılmaya tabii tutulmaktadır. Tasarlanan sistemde hareketler anlamlandırma mekanizmasına bağlı olarak bir örüntü ifade edebilecek bir biçimde oluşturulmak zorundadır. Anlık veya durağan görüntülerin çıkarımı yapıldıktan sonra bilgi içermeyen zaman aralıkları sistemin dışında tutulduğunda bir sonuca varılmaktadır. Doğruluk paylarına bağlı kalınarak sistemin ne oranda verimli olabildiği elde edilmektedir. İnsanların yer aldığı resim ve videolarda konum koordinatlarından yola çıkılarak bir sonraki aşamada ne yapacağı veya iskelet sisteminin ne oranda değişeceği sistematik bir sistemde oluşturulmaktadır. Optik akış hesabı ile eklemler üzerinde yapılmakta olan analizler sayesinde hareket bilgisinin elde edilerek hareketin gücünü ifade eden analizler yapılmıştır. Görüntü de belirlenen bölge üzerindeki hareket bilgisinin eklemlerin tespiti ile belirlenerek hareket oluşturduğu alanlarda çalışmalar ilerletilmektedir. Bu alanlarla bağlı olarak çeşitli histogramlar geliştirilmektedir. Histogramlar iskelet çıkarımını ve hareket sınıflandırmasına kolaylık sağlamaktadır. Eklemlerin iskelet ağaç yöntemiyle tespiti ile ardışık şekillerde zaman kavramını da içeren betimleyiciler ile güvenilirliği esas alarak bu eklemler çerçevesinde kayan bir sınıflandırma mekanizması oluşturulmak istenmektedir. Bu çalışma itibariyle hareketsiz veya hareketli varlıkların otonom hareketini tanıyan ve anlamlandırabilen işlevsel bir yapı elde edilmesi hedeflenmektedir. Tahminlemede ki problemi çözümleyerek işe yarar ve kolaylaştırıcı mekanizma sağlaması açısından verimli bir yapı oluşturulmaktadır. İnsan vücudunun tespitinin daha net ve açıklayıcı biçimde hareketlerin anlamlandırılması çalışmalarında daha kolay ve anlaşılır bir yapı meydana oluşturulmuştur.

Anahtar Kelimeler: İnsan duruşu tahminlemesi, Hareketlerin tespiti, İnsan vücut tanıması ve takibi

ABSTRACT

HUMAN POSITION PREDICTION BY DEEP LEARNING AND KEYPOINT EXPOSURE ESTIMATION

Kanpak, Hediye Nupelda
Master of Science in Department of Electric Electronic Engineer
Supervisor: Dr. Öğr. Üyesi M. Ali ARSERİM
June 2021, 59 pages

Interpreting human postures that belong to humans will be the mainstay of predicting what movements they will take. In the pictures and videos that draw this thesis, a planning is made about the system that decides with which realistic images. A whole study of signification, the selections of its movements as an automated system with the intended planned, determined at targeted time intervals. Making sense of the criteria in the available data. In the design system, a display representation has to be created depending on the meaning of the movements. After deducing from the instantaneous or static representation, the information is obtained in a way that is taken over time intervals. The extent to which it can be effective is obtained by adhering to accuracy payments. A system is created to review the place to be reviewed, pictures and videos, based on the directions, or what will change in the frame system. Thanks to what can be done on the optics and on the joints, the movement information can be expressed on the analysis obtained. Concerning the demonstration of regional movement information in the image as well, it is being advanced with movement-related studies. Depending on these areas, various histograms are developed. Histograms have the possibilities of inferring the skeleton and accessing the means of transportation. These additions, which contain comprehensive information about the children's approaching timers with the equipment method of the parts, are requested to form a sliding representation of the joint. It can recognize a stationary or moving autonomous movement of this movement, and a vehicle that is being magnified in an enlarged manner. A useful structure is created to facilitate its utility and utility by solving the problem in estimation. Clear and explanatory drawings of children in childhood are easier and easier to draw a work in making sense of movements.

Keywords: Human posture estimation, Motion detection, Human body recognition and tracking

1. GİRİŞ

1.1 Problem Tanımı ve Literatür Özeti

Asıl hedeflenen çalışma veriler üzerindeki insanların iskelet yapısını oluşturmak için eklemlerin yerini koordinatlarına bağlı olarak belirlemektir. İnsan pozlama sisteminin iyileştirilmesi ve geliştirilerek sistematik hale getirilmesi ön planda tutulmaktadır. Temelde bu çalışma sistemin nasıl bütün olarak çalıştığının esasına dayanmaktadır. Tahminleme de hem iç hem de dış varyasyonlar çalışmayı doğru sonuçlandırılmayı doğrudan etkileyen faktörler arasında yer almaktadır.

Karmaşık görünüm modelleri kullanarak, öğrenme algoritmalarına dayanmaktadır. Bu algoritmalarla eğitim modelleriyle parametreler elde edilmektedir [1]. Veri kümelerinin ele alınarak görüntülerin tahminlemesinden bir sınıf oluşturulmaktadır. Tahminleme çalışmalarının sonucunda doğruluk payının yüksek seviyede doğruluk yüzdesi vermesi beklenmektedir.

Anlık görüntülerdeki insan hareketlerinin algılanması, yorumlanması, zaman aralıklarının anlamlandırılması açısından önem arz etmektedir. İnsan hareketlerinin içeriğinin çok çeşitli ve çok yönlü olması çalışmaları daha önemli kılmaktadır.

Hareketlerin algılanması, jest ve mimiklerin tanınması, videoların düzenlenmesi, videolarda yer alan içeriklerin kolaylıkla aratılmasının sağlanması oluşturulan ortak kıstaslar dâhilinde kolaylıkla gerçekleştirilmektedir ve tahminlendirme mekanizması büyük değişiklikler barındırmaktadır [2]. Bu değişikliklere bağlı olarak zorlu koşullarda dahil tüm şartlarda tahminlemeye uygun ortam oluşturulmaya çalışılmaktadır. Artış gösteren videolar için akıllı çözümleri de içinde barındıran mekanizma yapılmak istenmiştir [3]. İnsan hareketlerini analiz edebilen bir sistemin gerekli olan bilgilerin elde edilmesi konusunda yardımcı olabileceği görülmektedir [4].

Hareketlerin tahminlenmesi görünenin aksine daha zor ve komplekstir. İnsanların istemsizce yapmış olduğu doğal davranışlarda dâhil olmak üzere bu hareketler belirli periyotlar içinde olmasına rağmen kişiden kişiye bile farklılık göstermektedir. Işık, ortam, giysi gibi faktörlerin olumlu veya olumsuz etkileri de tahminlemeyi büyük ölçüde etkilemektedir. Aynı sınıf içerisinde aynı kişiye ait olan hareketlerin arasında bile büyük farklılıklar görülmektedir.

Resimlerin yanı sıra videoların da anlık görüntülerinin yer aldığı veri setleri de tahminleme yapılmak istenen yapılar arasında yer almaktadır. Verilerdeki hareket tanıma sınıflandırması kendi içinde çevrim içi ve çevrim dışı videolar olarak uygulamalarda kullanılmaktadır. Daha önce yapılmış olan çalışmalarda yer alan video ve kamera görüntülerinin bulunduğu uygulamalarda kullanılmaktadır [5]. Metro gibi toplu taşımalarda yer alan kameralardan araçların takibinin yapılması, suçlu veya şüphelilerin sokak ve caddelerde yer alan güvenlik kameralarından tespitinin yapılması, evdeki veya şirketlerdeki kameralardan hırsızların tespiti gibi örneklemeler çevrim içi uygulamalara örnek verilmektedir. Çevrim dışı uygulamalar ise çeşitli müsabakalarda kameralardan yakın çekim alanlarda sporcuların hareketlerinin tespitinin yapılması, haberlerde yer alan muhabirlerin ve insanların hareketlerinin tespiti, konserlerde ve çeşitli gösterilerde yer alan oyuncuların yapmış oldukları gösterinin ayrıntılarına bakılması gibidir [6].

Görüntülerdeki sabit veya hareketli olan insan vücudunun tahminlemesi sırasında kameranın farklı bakış açılarından hangi tür hareket çeşidinin olduğunun saptanması açısından eğitim ve test aşamaları gerekmektedir. Buradaki tahminleme çerçevesi, bir görüntünün nasıl göründüğünü öğrenmeyi desteklemektedir. Düşük düzeyde öğrenmeler ile gittikçe iyi sonuçlar veren bir sisteme geçiş sağlamaktadır. İnsan hareket tespitinde genel çalışma olarak videolardan görüntülerin çıkarılması, bunlara karşılık gelen hareketlerin sınıflandırılması esas alınmaktadır.

Etkili bir yaklaşım ise resimsel yapıdır. Resimde yer alan insan vücudunun görünümünün tanınmasında görüntüleri çerçeveden ayrıştıran nesnelerin, parça çiftlerinin üzerindeki geometrik kısıtlamalarla birlikte yerel parça şablonlarına dönüştürmektedir. Parçalar piksel konumları ve yönleri birer parametre sınıfı olarak elde edilen yapılarla resimlerden özellik çıkarımı yapılmaktadır. Daha sonrasında bu özelliklerden elde edilen bir sınıflandırma oluşturulmuştur.

Hareket tanımayı en zor ve kısıt koyan çalışmalar kamera hareketi ve istenilmeyen hareketlerin olduğu dinamiklerdir [7].

Tahminlemeyi zorlaştıran en temel etmenlerden bahsedilecek olursa insan vücudunun belirli kısımlarının veya tamamen insan vücudunun yer aldığı insan görüntüsünün içinde kaybolması, kamera açısı ve odaklandığı ilgi alanının değişmesidir Zamanlama, hareket başlangıç ve bitiş gibi durumlar her durumda farklılık oluşturabileceği gibi hareket tanımadaki genelleme durumunu kötü etkileyerek sınıflandırılmanın yapılmasını zorlaştıracaktır. Diğer her şey düzelse ve hiçbir sıkıntı yaşanmasa dahi zamansal boyut en büyük sıkıntıdır [8]. İnsan duruş tahminlemesi genel bir bakış açısı ile değerlendirilirse tanımlama kategorileri olarak düşük, orta, yüksek seviyelerde gerçekleştirilmesi ile hareket sınıfının etiket sistemiyle yaygın sınıflandırma yapması başarı yüzdesinin arttırılma beklentisini oluşturmaktadır [9].

Çarpık şablonlar kullanılarak görüntülerin aranması etkili çözüm mekanizmaları arasında yer almaktadır. Yönlendirilmemiş parçalara sahip resimsel yapıların eklemli modellemelerde hızlıdır ve bu çarpıtmalardan kaynaklı tahminleme yapıları dinamik oluşturulmakta olan programlamadan alınmaktadır. Görünüm değişiklikleri ortaya çıkabilmektedir. Kısmi tıkanıklıklar gibi diğer faktörlerle etkileşimler göz ardı edilmemesi gereken bağlılıklardır.

Büyük görünüm değişiklikleri olmaması adına küresel karışımları modellemek için yerel parça karışımları bir araya getirilerek eşit derece de olmasa bile yerel parça kombinasyonlarını bir araya getirmektedir. Karışımların birlikte oluşumu ile model öğrenmesinin tutarlı bir şekilde yönlendirilmiş kenar yapılarından ve uzuvlardan ilişkisel bir şekilde tahmin modeli oluşturulması sağlanmaktadır. Verimlilik açısından en etkin yöntem insan vücudundaki uzuvların kaba pikseli alınmasını sağlayan dedektörlerde yerel konumu belirlenmiş varsayılarak yeterli derece arama yapılmalıdır. Böylelikle hem tahminleme hem de tespit yapılmaktadır.

Poz tahminlerinin doğruluk yüzdeleri yerelleştirilmiş parçaların yüzdesi(PCP) ile gösterilmektedir. Siluet yöntemi görüntüdeki tüm pikselleri hesaba katmaktadır. Düşük seviyeli görüntü arasında özellik açıklamasıyla köprü oluşturmaktadır.

Fourier ve Hough dönüşümleri ile şeklin sınırlarına geçerek dış yapıları ile ilgili bilgilendirme sağlamaktadırlar. Bu çalışmada vücut tahmini için oluşturulan mekanizmada ki kilit noktalarının sistematik olarak tespitinin yapılması, geniş bir eklem aralığının bulunması, bu noktaların tespit edilmesi, görülmekte olan hataların nasıl giderileceğinin belirlenmesi ve anlık hareketlerin tahmin edilmesi amaçlanmaktadır. Parçalar temel alınarak oluşturulan mekanizmada, her eklem çerçevesinde bir odak noktası penceresi oluşturularak arka plan etkisi gibi minimal düzeyde olan faktörlerin tahminde ortaya çıkan diğer hata oranlarına göre daha düşük tutulması sağlanmaktadır.

Tahminlemenin ilk aşaması derin öğrenme yardımıyla 7 katman olarak başlatılmıştır. Görüntünün içeriğine odaklanılarak istenilen görüntünün elde edilmesi ve o görüntünün tahmin edilmesidir. Bu katmanların içeriğini oluşturduğu sinir ağı her bir eklemi ve eklem regresörününün tespiti için kullanılmaktadır. Formülasyon sistemleri için grafiklere yer verilmektedir. Eklemler arasındaki etkileşimleri grafikleştirerek yüksek çözünürlüklü analiz ile eklemler arasındaki etkileşimler görülerek yapılmış olan sistem sonuçları güçlendirilmiştir. Bu yaklaşımlar görünüş ve zamansal farklara bağlı tahminleme sonuçlarında iyi perfomans gösterdiği görülmektedir.

Sonuç olarak mevcut sistemdeki zorlukların basite indirgenmesi açısından önemli bir ilerleme sağlayacak, zaman kaybının önüne geçecek ve analiz yapılmasını daha anlamlı kılacak bir yapı oluşturulması büyük avantaj ve kolaylık sağlamaktadır.

1.2 Tezin Amacı

Bu çalışmada, Quan Hua modellemesiyle açık kaynak kodlu futbol maçlarına dair video ve anlık görüntüler kullanılarak insan duruş tespiti gerçekleştirilmiştir. Algoritmanın verdiği sonuçlar anahtar noktaların ortalama hassasiyeti metodu ile doğruluk oranları hesaplanarak yüzdesel başarı oranı elde edilmiştir.

Devasa boyutta artan resim ve videolar manuel takip ve algılamaktan çıkıp otomatik olarak algılayarak anlamlandırma işlemleri yapılmaya çalışılmıştır. Bu teknolojinin ilerlemesinde en önemli etken olan insan silueti, bu kayıtlardaki çeşitliliği oluşturan etkenlerin en başında yer almaktadır. Anlık takip düzeyine erişen kayıtlarda anlam çıkarımının hızlı bir mekanizma olarak otomatikleşmesine duyulmakta olan ihtiyaç kaçınılmaz olmaktadır. Resim ve videoların anlık olarak anlamlandırılmasına dair yer alan çalışmalar geniş bir yelpazeye yayılmaktadır.

Çalışmanın amacı evrişimsel sinir ağı(ESA), Opencv, Cv2, Keras ve Convnets metotları kullanılarak, İnsan duruşu tahminlemesi yapmaktır. Devasa boyutta artan resim ve videolar manuel takip ve algılamaktan çıkıp otomatik olarak algılayarak anlamlandırma işlemleri yapılmaya çalışılmıştır. Bu teknolojinin ilerlemesinde en önemli etken olan insan silueti, bu kayıtlardaki çeşitliliği oluşturan etkenlerin en başında yer almaktadır. Arka plandaki giysi ve vücut ölçülerindeki farklı koşullar başarıyı ne oranda etkiliyor, eğitim veri seti çalışmanın performansını nasıl arttırmaktadır gibi sorulara cevap bulunmaya çalışılmaktadır.

İnsan duruşu tahminlemesi geniş bir yelpaze içinde yer almaktadır. Buna ilişkin olarak görüntülerde yer alan insan siluetlerine ait görünümlerin eklem bölgeleri numaralandırılmaktadır. Bu eklem bölgeleri birleştirilerek iskelet ağaç sistemi oluşturulmaktadır. Buna ilaveten bu sistem çok çeşitli çalışmalara öncülük etmiştir.

Bu tez çalışmasında insan iskelet sisteminin hangi poz aralıklarında olursa olsun bir vücut tahminlemesi oluşturarak insan poz duruşu basit bir yapı oluşturulmaya çalışılmıştır. Süre açısından avantaj sağlayan bir sistem oluşturulmuştur.

1.3 Tez Düzeni

Bu tez çalışmasında, iskelet sistemi oluşturmak için, istenen anlık görüntülerin bulunarak birleştirilmesiyle ilgili olarak bir veri seti elde edilmiştir. Her verinin bu düzlemde kullanılıp tahminleneceği bir sistem tasarlanmış ve gerçekleştirilmiştir. Sistemin tasarımı anlatılırken öncelikle birinci bölümde çalışmanın ne alanda olduğu, neler içerdiği, yapılmasındaki amaç ve bu çalışmaya olan ihtiyacın sebeplerine yer verilmiştir. İkinci bölümde literatür özeti verilmiştir. Üçüncü bölümde materyal ve metot kısmı yer almaktadır. Dördüncü ve beşinci bölümde ise, insan duruşu tahminlemesi sonucu elde edilen bulgular ve sonuçlara yer verilmiştir.

2. ÖNCEKİ ÇALIŞMALAR

Tahminleme ve hareket tanıma konuları üzerine farklı yapı yöntemler kullanılarak çeşitli yaklaşımlar öne sürülmüştür. Bu yaklaşımların arasındaki en temel fark belirlenen kıstaslardır. Her çalışmaya özgü ve istenen sonuçlar neticesinde şekillenen kıstaslar kullanılarak ideal bir tahminleme metodu gerçekleştirilmektedir. Bu kıstaslar özellikle arka görünüm, arka plan, bakış açısı, eylemin gerçekleştirme şeklinden bağımsızlaşarak bir genellenebilen model olmalıdır. Veri seti hareket kavramını içeren özellikler ve zamansal kavramları da içermektedir. Hareket ve zaman kavramına ait veriler ardışık olarak işlenirler. Daha sonra zamansal olarak hangi pozisyon ve duruşun alınacağı belirlenir ve o anlık görüntü evrişimsel sinir ağı ile işlenir. Literatürde çeşitli yöntemler izlenerek iskelet ağaç metoduyla insan duruşu tahminlemesi yapılmaya çalışılmıştır.

Wang insan siluetlerine öklid uzantıları ile kademe oluşumu yaparak R transformunu meydana getirerek tahminleme yapmıştır [10]. Silüetlerin Öklid uzantıları ile kademeli mekanizma oluşturarak optik akış özelliklerini kullandığında kenar noktaları ve eklemlerin yerinin tespitinin yapılması genel anlamda bakıldığında en etkili yöntemler arasındadır.

Souvenir ve Babbs, öklid ile belirlenen R transformuna üçüncü bir boyut olarak zamanı katarak hareket tanıma çalışmaları yapmıştır [11].

Weinland, insan vücutlarındaki Öklid uzantılarını uzaklıklar ile eşleştirerek sınıflandırma yapmıştır. Bu çalışmaların sonrasında başka bir çalışmasında Chamfer adlı uzaklık ile sınıflandırma yaptığı çalışmaları mevcuttur [12].

Martinez ve diğerleri [13], taban çizgisi modeli ile raf dışı bir 2D detektörden giriş ile 2D poz çıkışı alarak tek bir görüntüdeki gerileme kaybı olan 3D pozunun tahminlemesini yapmıştır.

Zhou ve diğerleri [14], 2D ve 3D pozları aynı anda gerileterek [13] çalışmasına benzer performans sergileyen sistem oluşturmuştur. Bu çalışmada, ısı haritalarını en temel katmanda kullanılmaktadır ve bu harita ile eklemler belirlenmektedir.

Sun ve diğerleri, yaptıkları çalışmada eklemlerin 2D ve 3D ısı haritasındaki tüm konumlardaki olasılıklarının ağırlıklandırmışlardır [15].

Yapılan bir başka çalışmada, kinematik iskelet montajını benimsenerek ve 2D ve 3D eklem pozisyonları zamansal açıdan geriletilerek 2D ısı haritasını oluşturmuşlardı [16]. Bu şekilde 2D ve 3D sistemlerin güçlü olarak birbiriyle bağlantı kurması amaçlanmıştır.

Povllo ve diğerleri, ESA karelerinden dolambaçlı küçük ağa giriş olarak 2D pozlarla daha sonrasında 3D pozu tahminlemektedir [17]. Ayrıca özellik çıkarımında elde edilen veriler çeşitli makine öğrenmesi metotlarıyla modellenmiştir.

Bir başka çalışmada, t zaman aralığında silüetlerin takibini yapıp uzay zaman aralığında izlenimleri oluşturulmuştur ve Poisson denklemleri ile özellik çıkarımı yapılarak hareket tanımlanıştır [18].

Lu ve Little, yönlü gradyanlar histogramı üzerinde ana bileşen analizi uygulayarak hareket tanıma modeli tasarlamıştır [19].

Laptev ve Lindeberg, harris köşe tanımlayıcısını üç boyutlu şekle getirip, parça çıkarımı yaparak hareketleri tanımlamışlardır [20].

Willems, çalışmasında hızlandırılmış öznitelikler ile 3 boyutlu hareket tanımlayıcısı kullanmıştır [21].

Tekin ve arkadaşları, 3D pozunun kurtarılması için video dizisinin ard arda gelen karelerinden hareket bilgisini kullanmışlardır [22].

Pavlokos ve arkadaşları, renkli tek bir görüntüden uçtan uça öğrenme paradigması sayesinde en iki adımlı optimizasyon aşamalarından oluşan 2B ve 3B sistemler üzerinde çalışmışlardır [23].

Tung ve arkadaşları, kamera girişi için öğrenme tabanlı olan bir hareket yakalama sistemi tasarlamışlardır [24].

Sarafianos ve arkadaşları, görüntü ve videolardan 3 boyutlu insan tahminini teknolojiye ayak uydurması açısından çalışmışlardır [25].

Rhodin ve arkadaşları, 3D poz tahmini ve yöntemlerine dayalı çözümler önermişlerdir [26].

3. MATERYAL VE METOT

Bu bölümde; çalışmada kullanılan metot aşamalı olarak verilmiştir. İlk olarak insan duruşu tahminleme yöntemleri ve ardından önerilen metot ayrıntılı olarak açıklanmıştır. Yapılan işlemlerin sıralaması aşağıda sıralanmıştır.

1. Video ve Resimlerden Görüntü Alınması ve Sisteme Girdisi
2. Evirişimli Sinir Ağları
3. Koşullara göre Hareket Tanıma Aşamaları
4. Roi ve İlgi Alanı Çıkarımı
5. RGB ve Optik Akış Güç Şekilleri
6. Sıralanmış Yönlü Gradyanlar ve Destek Vektör Makineleri

3.1 2B ve 3B Veri Kümeleri

Bu tezde Quan Hua algoritması kullanılmıştır [27]. Bununla birlikte tez çalışmasında kullanılmak üzere futbol maçlarına ait çok sayıda farklı pozisyonlar içeren anlık görüntülerden veri seti oluşturulmuştur. Bu görüntüler farklı youtube videolarından elde edilmiştir. İlgili veri setine ait görüntülerden örnekler Şekil 3.1'de gösterilmiştir.

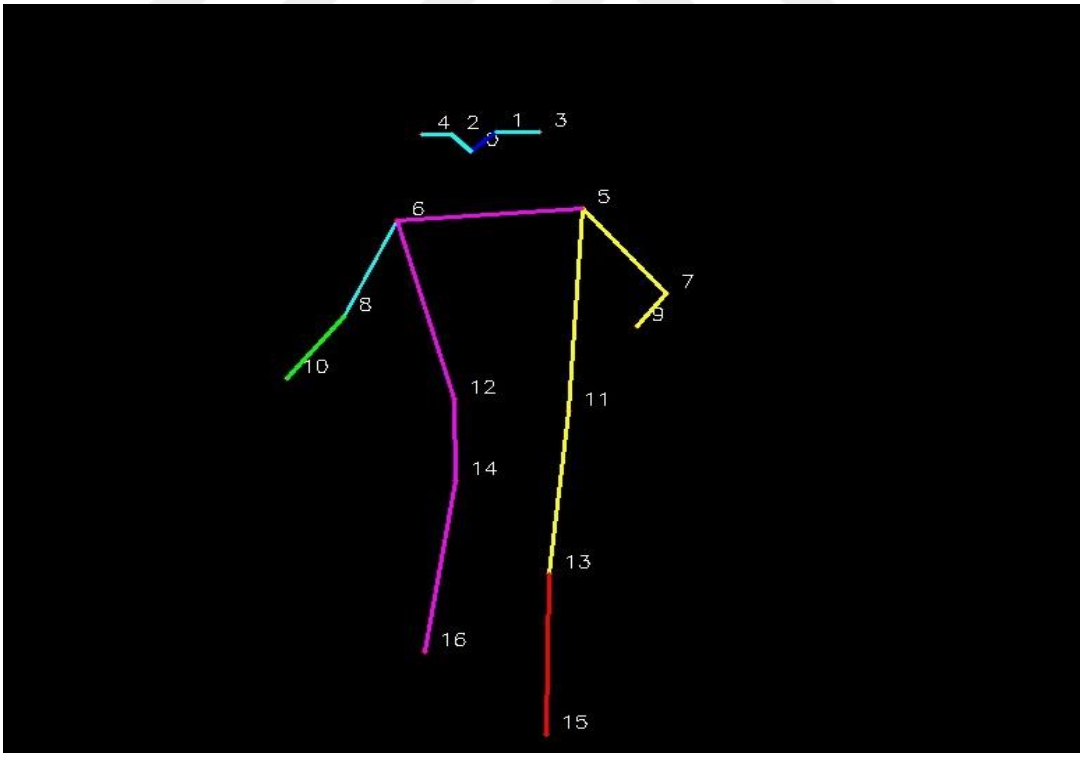


Şekil 3.1 Çalışmada yer alan veri setlerinden futbol maçlarına ait anlık görüntü tanıma sonucu toplanan veri seti

3.2 İnsan Poz Tahminlemesi

İnsan poz tahminlemesinin amacı, yapılan çalışmalara göre değişkenlik gösterebilir. 3B poz tahminlemeleri için 2D pozları alınıp derinlik görüntüleri kullanılarak çok sayıda çalışmalar yapılmıştır [28]. İnsan siluetindeki pozlar ana eklemler (bilek, omuz, kalça, diz, boyun, baş, göğüs kafesi ve leğen kemiği, dirsek, ayak bileği) ile temsil edilmektedir ve bu işlem Şekil 3.2' de gösterilmektedir. Bu tez çalışmasının odak noktası, bu eklemlerin görüntü koordinatlarını tahmin etmektir.

İnsan pozunun en zorlu yönlerine göre, insan görüntülerindeki büyük değişiklik (kıyafet, aksesuar, saç stilleri), ışık konumlanmasındaki değişiklik, insan vücudundaki fiziki farklılık, kısmi tıkanmalar, iskeletin karmaşıklığıdır. Pozların yüksek boyutluluğun en kolay çözümü anahtar eklemlerin numaralandırılarak poz tahminlemesi yapılmasıdır.



Şekil 3.2 Anahtar noktaları belirleme ve tahmin metodolojisi[29]

3.2.1 Keras modellemesi ve jupyter not defterleri

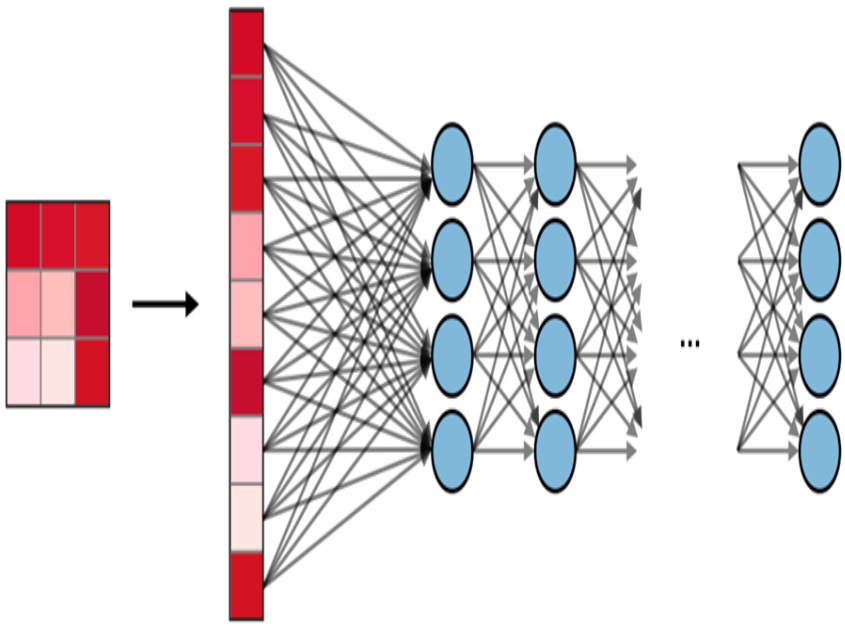
Python kaynak kodlu yazılım, derin öğrenme modelinin oluşturulması ve veri setlerinin eğitilmesi için olanak sağlar. Bununla birlikte bu yazılım ile zaman CPU ve GPU kullanarak evrişimsel sinir ağının eğitilmesine olanak sağlar.

Çoklu girdi ya da çoklu çıktıların, değişik ağ yapıları ile eğitilmesi mümkündür. Ayrıca Keras modellemesi ile çeşitli evrişimsel ağ yapıları oluşturulabilir.

Jupyter not defteri doğrudan internet tarayıcısıyla çalışmaktadır ve bu yapıyla derin öğrenme modellemeleri ile ilgili kodlar çalıştırılabilir. Jupiter not defterleri, makine öğrenmesi ve veri bilimi alanlarında kodlama için yaygın olarak kullanılmaktadır. Bu platform, evrişimsel sinir ağlarına ait python kodlarının bütünsel olarak çalıştırılabilmesiyle birlikte kodların satır satır işlenmesine olanak sağlar.

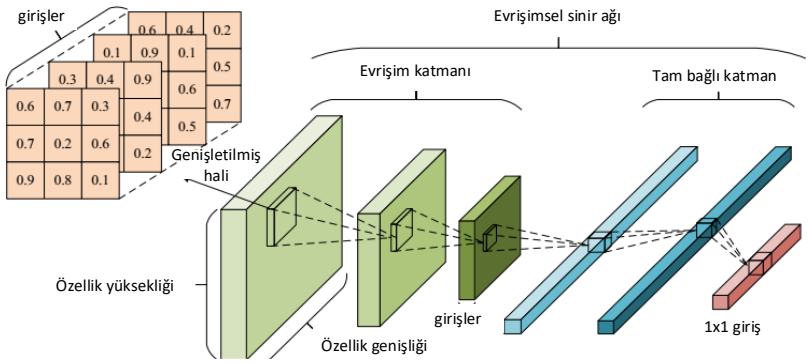
3.3 Evrişimsel Sinir Ağları

İnsan duruşu tahminlemesi çalışması yapıldığında, veriler farklı yollardan toplanabilmektedir. Evrişimsel sinir ağları sınıflandırma problemlerinde kullanılan etkili bir modeldir. Bu bölümde, ESA'ların çalışma yapısına değinilmiştir. Evrişimsel sinir ağlarına ait yapısal şema Şekil 3.3'te gösterilmektedir.



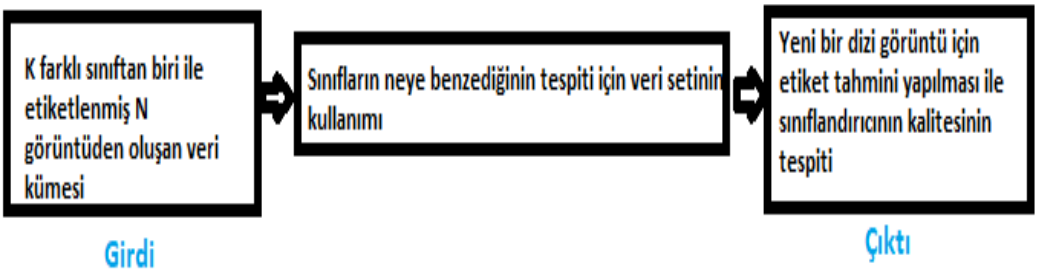
Şekil 3.3 Evirişimli Sinir Ağı Yapısı [30]

Evrişimsel sinir ağları çok büyük verilerin sınıflandırılması, tanınması gibi işlemler için kullanılmaktadır. Bunun için ağ parametreleri eğitilir ve başarı oranı maksimum olacak şekilde sinir ağı girişleri azaltılır. Bu yapıda çıkış katmanında yapay nöronlar bulunur. Bununla birlikte evrişim yapısal olarak 2 boyutlu olan görüntü işaretlerine de uygulanabilir. Evrişim süreci, kayan pencereler görüntü üzerinde iki boyutlu bir filtrenin evrişimi sayesinde hesaplanır ve bu işlem Şekil 3.4'te gösterilmiştir. Evrişim katmanları sayesinde daha az parametre sayısına ihtiyaç duyulmaktadır. Parametre sayısı, orijinal görüntü boyutundan bağımsızdır. Görüntü boyutu istenildiği gibi arttırılabilir ve buna bağlı parametre sayılarında değişiklik meydana gelmeyebilir.



Şekil 3.4 Evrişim katmanı [31]

Şekil 3.5'te yer alan şemada belirtilen aşamalar öncelikle n tane veriden oluşan görüntü kümesini girdi olarak kullanıp sınıfların oluşmasını sağlayacak alt yapı meydana getirmektedir. Sınıfları tam olarak kategorize etme amaçlı neye benzediğine dair tespitinin yapılmasına odaklı veri setinin kullanımıyla ön planda tutulmuştur. Yeni dizilerin verileri için tahminleme yapılmasıyla sınıflandırmanın kalite düzeyi ölçülmektedir. Boru sistemi olarak adlandırılan sistem, sınıflandırma kalitesinin üst seviyeye taşınmasına ve sınıflandırmanın kolaylıkla yapılmasını sağlamaktadır.



Şekil 3.5 Veri kümelerinde veri tespiti için kullanılan Boru Şeması

Boru şeması, görüntüleri anlamak, analiz etmek ve arka plandaki özelliklerin gizli temsillerini anlamaya çalışan sistemdir. Farklı kategorideki görüntüleri otomatik olarak sınıflandırma işleminde, hangi nesnelerin görüntüde var olduğunu algılayan algoritmalar kullanılmaktadır. Bunun temelinde derin öğrenme yapılarında yer alan Convnets yapıları vardır. Bu yapılar, bakış açısı değişimi, ölçek değişimi, varyasyon, deformasyon, tıkanmalar ve aydınlatma koşulları gibi zorlukları içeren görüntü kümelerini bir arada toplamaya yaramaktadır [32].

Verilere açıklama eklenmesi de önemli adımlardan sayılmaktadır. Veriler metin, resim, video gibi alanlarda anahtar kelimelere ihtiyaç duymaktadır. Gerekli metin, resim, video gibi nesnelere anahtar kelimeler atanır. Bu yüzden verileri etiketlemek gerekmektedir. Verileri etiketlemek doğrudan tahminlemenin başarım oranını etkiler. Böylelikle tahminleme yapılan verilerin doğru veri setlerine ayrılması sağlanarak etkili bir tahminleme gerçekleştirilebilir.

Görüntü sınıflandırması tahminleme yöntemlerinin ana kılavuzu gibidir. Ele alınan görüntülerin tek kategoride toplanarak etiketlenmiş veriyi test görüntüsü için kategorize edileni tahminlemek ve bu tahminlemenin doğruluk payını ölçmek gerekmektedir.

Bilgisayarla öğrenmede sınıflandırma için daha kolay yöntemler tespit edilmiştir. Belirlenen görüntülerin birçok örneği sağlanarak ve örnekler incelendiğinde her bir sınıfa ait görsel görünümünü eğiterek öğrenme algoritmaları geliştirilmektedir. Öncelikle görüntüler etiketlenir, daha sonra bu veri kümesi eğitim için toplanır ve bu veri setleri CPU veya GPU ile eğitilir.

Genellikle evrişimsel sinir ağlarıyla yapılan sınıflandırma için kullanılan veri kümeleri çok büyüktür. Özellikleri daha iyi belirlenmesi için veri büyütme yöntemlerine başvurulabilir.

Evrişimsel sinir ağının ana yapı taşı, öğrenilebilir filtrelere sahip katmanlardır. Girişlere filtreler uygulanır ve bir sonraki katmana geçiş yapılmaktadır [33].

Bu çalışmada uygulanan ağdaki evrişimsel sinir ağı katmanlarında 3 x 3 boyutunda filtreler kullanılmaktadır.

Tahminlemeden önce, katmanlarda kenarlar ve dokuların yerel özellikleri algılanır. Daha sonra yerel özellikler yüksek seviyeli özellikler ile birleştirilirler. Eğitim sırasında ağ güncellenerek en kullanışlı özellikleri içinde barındıran filtre parametreleri ağ tarafından öğrenilir.
Eğitimden sonra yapılan süreçte, esnek modelleme kullanılmaktadır. Esnek olan bir karışım modellemesi ile parça konumları arasında bir ilişki oluşturularak ortak noktalar belirlenir ve bu kısımlardan yakalama gerçekleştirilir. Eklemler arasındaki ilişkileri kodlayan resimsel şablonları oluşturarak tahminleme yapılmaktadır. Yapılan esnek şablon modellemesinde, eklemler arasında yapılan hesaplamalar modellemenin verimli bir şekilde yapılmasını sağlamaktadır. Ayrıca üssel olarak küresel karışım kümeleri ile modelleme yapılmaktadır.

3.3.1 Maksimum havuzlama ve dropout katmanları

Maksimum havuzlama işleminde en büyük değer döndürülerek diğer değerlerin güncellemesinde kullanılmaktadır. Maksimum havuzlama uzamsal boyutu kademeli düşürmek için evrişim katmanları arasına yerleştirilir. Parametre miktarını azaltma etkisine sahip bir yapıdır ve hesaplama miktarını azaltma görevi görmektedir. Parametreler azaldıkça düzenleyici etki artmaktadır ve ağın ezberleme yapmasını engellenmektedir.

Elde edilen çıktının uzamsal boyutunun hesabı (3.1)'de verilmiştir

$$\text{Çıktının uzamsal boyut hesabı} = \frac{[\text{win}-\text{k}+2\text{p}]}{\text{s}} + 1 \quad (3.1)$$

Burada;

"k" çekirdek boyutu, "s" adım, "p" sıfır doldurma, "win" girişin uzamsal boyutunu göstermektedir.

Dropout katmanı için 0 ile 1 arasında bir değer verilerek eğitimin performansını maksimize edilmek amaçlanır. Eğitimde bu değer miktarındaki bağlantıyı rastgele kapatmaktadır. Sinir ağlarındaki aşırı uyumu azaltmak için kullanılan düzenleme metodudur. Bu kısım, iki katman şeklinde oluşturulur. Zayıf katmanlarda gereksiz bilgi barındırılır, güçlü katmanlarda sürecin kısa sürede tanımlanması hedeflenir.

3.3.2 Evrişimsel sinir ağının eğitilmesi

Evrişimsel sinir ağında görüntü kümesinin eğitilmesi için girişler ağa uygulanır, katmanların içinde yer alan maksimum havuzlama kısmında çıktıların uzamsal boyut hesabı yapılır. Daha sonrasında bir öngörü elde edilene kadar, bu süreç ileri yayılma şeklinde devam etmektedir. Tahminde kullanılacak gerçek en yakın komşuluğa ne kadar yakınsa kayıp o kadar küçük olmaktadır.

Kayıp hesabı öklid aracılığıyla her bir nöronun uğradığı kayıp en yakın komşuluk mekanizmasına benzeyerek hesaplanması (3.2)'de gösterilmektedir.

$$\text{Öklid}(\text{Kayıp}) = \frac{1}{\text{N}} + (\text{i} = 1)\sum(\text{N})(\text{yt} - \text{yö})^2 \quad (3.2)$$

Burada;

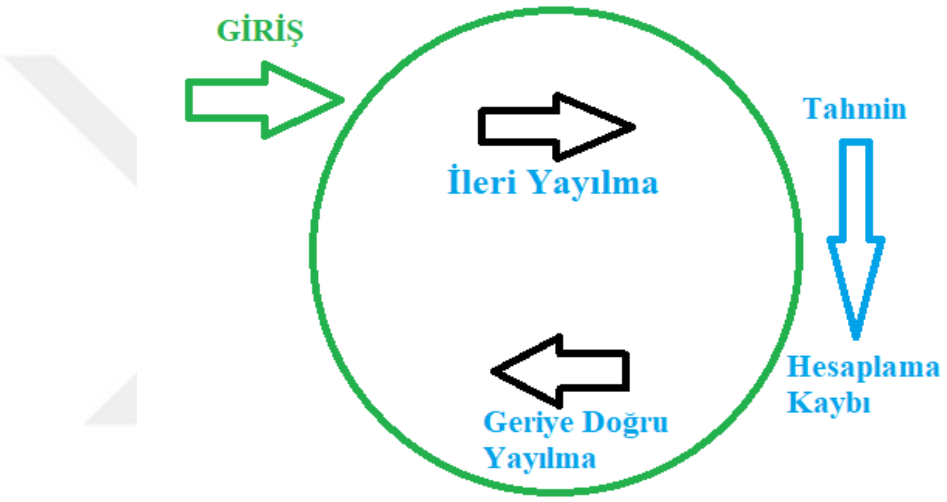
"N" çıktı sayısı, "yt" temel gerçeklik, "yö", öngörülen çıktı olarak nitelendirilir. Bu eşitlik algoritmanın ileri yayılma kısmında kullanılan hesaplamadır,

Geriye yayılmada kullanılan hesaplama ise (3.3)'de gösterilmektedir,

$$\text{Parametre Vektörü} = \text{Qi} = \text{Qi} - \frac{d}{dQi} J(Q) \quad (3.3)$$

Q, a ve J(Q) sırasıyla modelin parametresel vektörünü, öğrenme hızı, parametre güncellemesini göstermektedir.

Denetimli öğrenme sistemine ait şema grafiği Şekil 3.6'da gösterilmektedir. Giriş yapılan parametrelerden sonra ileri yayılma ve geriye doğru yayılma uygulanarak hesaplama kaybı tahmin sonuçlarından düşülerek bir çıktı alınmaktadır.



Şekil 3.6 Denetimli Öğrenme Şematiği

Test sisteminde kayıp azaldıkça eğitim süresi uzamaktadır. Eğitimde yer alan veriler sinir ağına uygulandıktan sonra gradyan iniş nedeniyle ağın eğitiminin olması ve hatanın azalması beklenmektedir.

3.4 Tahminleme Yöntemleri

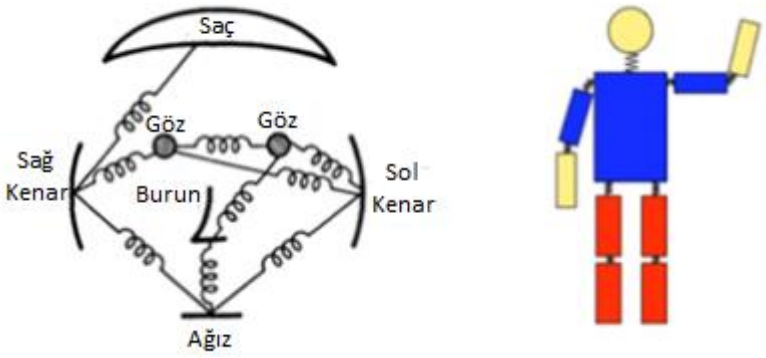
İnsan duruşu tahminlemesinde çok kapsamlı yöntemler kullanılmaktadır. Bu tez çalışmasında üç yöntem ele alınmıştır. Kullanılan yöntemler kıyaslanarak doğruluk yüzdesi en yüksek olan metodun çalışma alanı geniş bir yelpazede tutulmuştur. Kıstas alınan yöntemler oyunlarda, animasyonlarda, kameralarda görüntü tanımada yoğun bir şekilde kullanılmaktadır.

İnsan hareketlerinin incelemesi yapıldığında, örnek olarak yürüme işlemi ele alındığında, adım büyüklüğü, yürüme süresi, yürüme hızı, stili, tipi çeşitli farklılıklar göstermektedir. Bu tür hareketler incelenirken kamera ile kayıt işlemi esnasında, kameranın hareketi, istenilmeyen dinamik durumlarda ve ortamlarda çekimlerin devam etmesi, hareketleri tanımayı ve tahminlemeyi güçleştirmektedir. Kameranın kayıt açısı ve çektiği alanın ilgi alanı içinde olması görüntü işleme alanında çok önemlidir. Diğer yandan bahsedilen bu ayrıntılar, insan davranış tahminlemesinde genelleme yapılmasını zorlaştırmaktadır.

İki boyutlu poz tahmininde, bir RGB görüntüsündeki her bir insan vücudunda bulunan eklemler için (x,y) koordinatları temel alınarak tahminleme yapılmaktadır. 2 boyutlu sistemlerde, görüntüdeki tahminleme yapılacak nesneyi esnek bir yapıda olacak şekilde sürekli değişiklik gösterebilen bir mekanizma temsil etmektedir. Buradaki her bir parçaya karşılık gelecek şekilde bir eşleşme yapılmaktadır. Kullanılan parçalar yaylardan oluşmaktadır. Yayların buradaki temsil ettiği kısım, parçalar arasındaki bağlantıya odaklanarak gösterilmektir. Parçalar vücudun konumlarını gösterdiğinde ortaya çıkan şekil, yapılmak istenen poz tahmininin modellemesidir. Bu yöntem, modelleme yapılırken görüntü verilerine bağlı olmayan bir modellemeye sahip olma sınırlamasıyla ortaya çıkmaktadır.

Diğer yöntemde, Şekil 3.7'de yer alan deforme olabilen parça modellemesi yapılarak temel alınan görüntü ve videolardaki karmaşık eklemlerin arasındaki bağlantıyı ifade eden bir modelleme kullanılmaktadır.

Buradaki amaç değiştirilebilir veya bozulabilir parça modellerini, değiştirilebilir, konfigürasyonla düzenleyerek bir taslak veya şablon oluşturmaktadır. Bu modelde yay ile belirlenen parça şablonları yer almaktadır. Bu şablonlar nesneyi belirlemek veya algılamak için görüntüde yer alan kısımları eşleştirmektedir. Bu şekilde görüntü parçalardan oluşan şablonlar ile kolay ve iyi bir biçimde modellenebilmektedir.



Şekil 3.7 İnsan Duruşu Tahminlemesinde Yay[34]

Statik görüntüler incelendiği zaman insan tespiti ve insan pozu için yapılan yöntemde çarpık ve karışık modellemeler yerine kısaltılan veya dönüştürülen küçük şablonlar temel alınmaktadır. Esnek modellemede oluşturulan parçaların birbirlerine uzamsal bağlantılarının ve bu uzamsal ilişkilerde tespit edilen parçaların arasındaki ortaklığın oluşturduğu yapılarla birlikte kodlanarak yapı modellemesi oluşturulmaktadır. Buradaki esas amaç üssel küme temel alınarak yerel karışımların bileşimiyle farklı açılardan eklemlerin yerlerinin tespiti ile modellemedir.

Video görüntülerindeki poz tahmini için, bir RGB çerçevesinin x, y koordinatları ve zaman(t) ile ilgili bileşen ele alınarak, görüntü çerçeveleri değerlendirilmektedir. Veri olarak videolardaki görüntüler dikkate alındığında, kişinin sabit bir konumundan çok gerçek hareketini analiz edildiğinden dolayı daha fazla doğruluk payı elde edilmesini sağlamaktadır.

Videodaki kişilerin farklı görüşlere sahip açılardan ele alınması daha iyi sonuçlar vermektedir. Videoda rastgele belirlenen zamansal eşikler sayesinde vücut kontrol noktalarının konumu otomatik olarak tespit edilebilir. İnsan vücudundaki 2B noktalarının algılanmasını sağlamak bu yöntemi etkili kılan önemli hususlardan biridir. Anahtar noktalarının konumları çıkarıldıktan sonra referans video noktaları ile kıyaslanmaktadır [35]. Her video saniyesindeki anlık görüntülerin resmedilmesi sonucunda bir kıyas yapmak mümkündür. Ayrıştırma, kare kare tespit edilen anahtar noktaların analizi ve kriterleri temel alarak yapılmaktadır.

3.4.1 Doğru pozlama yöntemleri

Doğru pozlama yöntemi, vücut bölümlerinin doğru tahminleme yüzdesi olarak bilinmektedir. Kullanılan segment, uç noktaların içindeyse doğru bir tahminleme yapılmaktadır denilebilir. Tam ağaç modelinde üst kısımlar arasındaki bağlantıda yer alan üst beden sınırlamasından ötürü daha başarılı tahminleme yapılmaktadır.

Doğru pozlama yönteminde, geniş çaplı bir değerlendirme yapılmaktadır. Bu yöntemin işleyişi uzuvlara karşı yapılan kısaltma işlemlerine duyarlı ve doğru poz olasılığına dayalıdır. Bununla birlikte bu yöntem yerel vücut parçalarının yüzdesini ölçen bir yapıya sahiptir. Ayrıca tahminleme yapılacak vücut parçalarında kullanılan segmentlerin uç noktalarının % 50 diliminde yer alması gerekmektedir.
Doğru pozlama yönteminin zorlukları, tahminleme yapılacak uzvun ortalamasını puanlamaya yarayan uç noktaların yerel tanımının uzuv uç noktalarına denk gelmemesidir. Bu nedenle kısaltma miktarında azaltılabilinir.

3.4.2 Kişi katmanları yöntemleri

Kişi katmanları yönteminde, poz tahminlemesinde yer alan sorunları ele alan iki ölçüt vardır. Birincisi tahminlenen test görüntülerinin sınırlayıcı kutu ile açıklanmasıdır. Buradaki en önemli faktör, bu sınırlayıcı kutuların bir alt kümesinin bir dedektör aracılığı ile sert pozları da göz önünde bulundurarak değerlendirilmesinin yapılmasıdır.

İkinci Ölçüt sınırlayıcı kutu temelli kullanılan düzenlemelerin tahminleme için vücut eklemlerini belirlemede yer alan temel noktaları belirtmesidir. Bununla birlikte temel noktalar sınırlayıcı kutular ile örtüşmelidir. Burada öncelikle aday anahtar noktaları tanımlanır. Daha sonra kesin temel nokta koordinatları ile sınırlayıcı kutu genişliğinin doğru tahminlemesi yapılır.

3.4.3 Anahtar noktalarının ortalama hassasiyeti yöntemleri

Anahtar noktalarının ortalama hassasiyeti yönteminde, sınırlayıcı kutu kapsamında ve kendi yerel koordinatları içinde yer alırsa gerçek pozitiflik yani doğru adaylık sağlamaktadır. Değerlendirmenin sonucunda kaçırılan tespitler ve yanlışlıklar belirlenmektedir.

Anahtar kilit noktalarının her biri ayrı ayrı oluşturulmaktadır. Tahminlenen vücut parçalarının yani eklemlerin tespiti ve algılama doğruluğu hassas geri çağırma eğrisi ile gösterilmektedir.

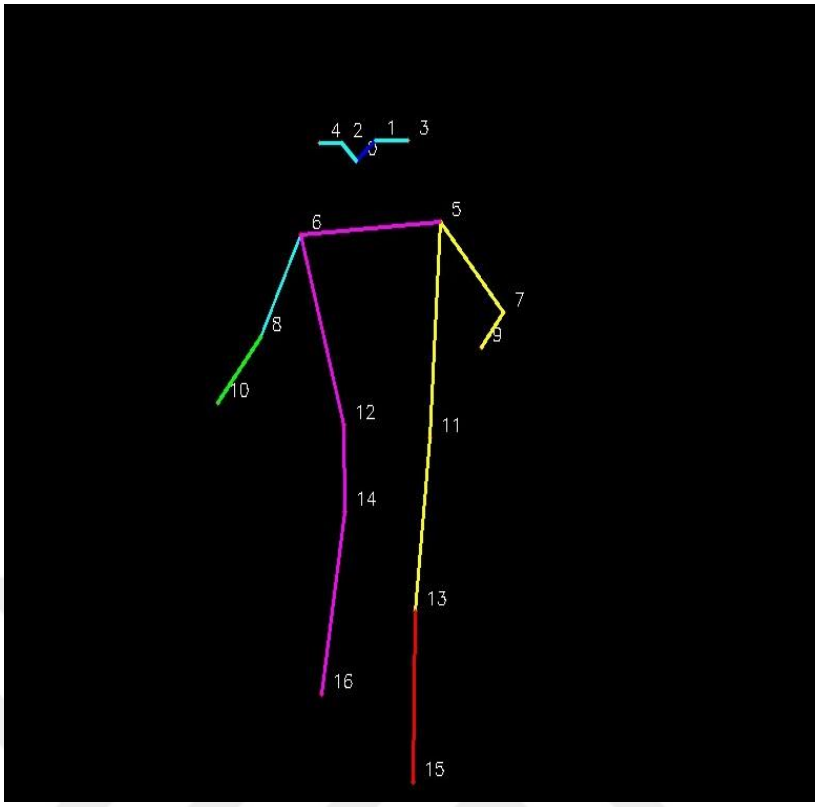
Görüntülerden alınan tahminleme verileri, evrişimsel sinir ağına uygulanır. Daha sonra koşullar belirlenerek hareket tanıma aşamaları oluşturulmaktadır görüntülerde tahminlemesinin yapılması istenen insan duruşunun belirlenmesi ROİ yardımıyla sağlanmaktadır. Bununla birlikte ilgi alanı çıkarımı yapılarak işe yaramayan ve çevresel faktörleri içeren görüntüler tahminleme mekanizmasının dışında tutulmaktadır. RGB bileşenleri eklemlerin yer aldığı koordinatları belirmede kullanılmaktadır. Ayrıca optik akış şekilleri sayesinde hareket halindeki eklem parçalarının tespiti sağlanmaktadır. Sıralanmış yönlü gradyanlar aracılığı ile kaba taslak oluşan iskeletin, elde edilengradyanlardan oluşturularak kenar özelliklerinin çıkarımı gerçekleştirilmektedir. Destek vektör makineleri yardımıyla da bu kenar özelliklerinin içinde yer alan eklemlerin net olarak tespiti yapılmaktadır.

Bu çalışma da, insan duruşu tahminlemesi yapılırken, evrişimsel sinir ağları, ROİ, RGB bileşenleri, destek vektör makineleri, ilgi alanı çıkarımı, optik akış güç şekilleri ve sıralanmış yönlü gradyanlar yöntemleri anahtar noktaların ortalama hassasiyeti yöntemi ile hesaplanmıştır.

3.5 Oluşturulan Veri Seti

Bu tezde oluşturulan veri seti kamuya açık olan 200 farklı görselden (video, anlık görüntü ve resimler) oluşmaktadır. Görüntüler, Youtube videolarından alınmıştır ve vücut ölçüsü, pozlama açısı, kameranın kişiye olan mesafe ve açısı ve arka plan bakımından büyük farklılık içermektedir. Kullanılan veri setinde Cristiano Ronaldo'ya ait futbol maçlarındaki görseller bulunmaktadır.

Sınırlayıcı karenin içerisindeki nesnelerin konumu ve ölçek hesaplanabilir. Hesaplanan anahtar noktalar Şekil 3.8'de yer almaktadır.



Şekil 3.8 Hesaplanan anahtar noktalar [36]

3.6 Tahminleme Aşamaları

Bu bölümde, insan poz tahmini sırasında uygulanmakta olan aşamalar anlatılmaktadır. Tahminleme aşamalarının akış diyagramı Şekil 3.9'da verilmiştir ve tahminleme işleminde bu diyagramdaki sıralamaya göre bir yol izlenmektedir.



Şekil 3.9 Tahminleme sisteminin yapısı

Sistemsel yapı, video çerçevelerinden hareketsiz arka planın çıkarılarak bu veriler ile evrişimsel sinir ağlarının eğitilmesi, daha sonrasında ilgi alanının ön plana koyulması sonucunda gerekli bilgilerin toplanması için yapılan çalışmalardır. Optik akış sonrasında filtreleme yapılarak parlaktan karanlığa koyu tona doğru sıralanmış örüntü gradyanlarıyla eklemlerdeki anahtar noktaları kesin ve net bir şekilde bulunur ve hareket sınıfı belirlenerek tahminleme yapılır.

3.6.1 Görsellerin alınması ve sistem girişi olarak kullanılması

Hareketlerin zamansal olarak değişen ortak ve ana nitelikler ile bir örüntü oluşturulmaktadır.

Görüntü ve videolardaki hareketlerin belirli zaman aralıklarında durağan hale getirilmesi tahminleme için en önemli noktalardan biridir. Bu hareket akışlarına bağlı olarak veri setleri şekillenmektedir. Bu çalışmada Cristiano Ronaldo'nun 200 adet gol atma anı veri olarak kullanılmıştır. Ayrıca Quan Hua'nın çalışmalarında yer verdiği temel sistematik kullanmaktadır [37]. Veri setinde yer alan görüntülerin kalitesi tahminlemenin daha kolay olmasını sağlamaktadır. Gol atma anlarının zamansal açıdan anlık olması veri setlerinin kalitesini düşürse de kullanılan verilerin zamansal olarak iyi ayarlanması tahminlemenin önceliklerden biridir. İstenilen ölçütlerin yer aldığı zaman aralıkları çalışmanın verimliliğine büyük katkı sağlamaktadır.

3.6.2 Koşullara göre hareket tanıma

Çalışmanın ilk aşamasında, video girişlerinden alınan görüntülerde odaklanılan görüntünün dışındaki arka planı kaldırmak ve ilgi alanını ön plana çıkarmak için gerekli bilgileri toplamak gerekmektedir. Örüntü esasından yola çıkılarak mekanizma oluşturulmaya çalışılmaktadır. Daha sonra optik akış mekanizması ile süzülerek hareket sınıfı belirlenmektedir. Açıktan koyuya doğru sıralanmakta olan gradyan desen birleşimlerinde anahtar noktalar bulunarak tahmin yapılmaktadır. İnsan duruşundaki farklılık, hareket şekilleri, kamera, ışıklandırma, perspektif etkisindeki farklılık sebebiyle çeşitlilik oluşmaktadır.

Değişken koşullarda sistem başarısı maksimum seviyede tutularak tahminlemenin sapması önlenmek amaçlanmaktadır. Bilginin çeşitliliğinin çok olması sınıflandırma açısından avantajdır.

Görüntüye girdi olarak kabul edildiğinde, kodlayıcı hareketin yönü, daralma olarak evrişim bloğu özellikleri çıkarılarak ana hatlarıyla belirtilmektedir. Koordinatlara bağlı regresörler sayesinde mimari sonuç oluşturmaktadır. Öncelikle kilit noktaları doğrudan kodlanılmak yerine o bölgede bir poz çerçevesi oluşturulmaktadır [38]. Özellik çıkarımıyla kategorize işlemlerine ihtiyaç duyulmaktadır. Bu kategorize işlemleri Convnets kullanılarak yapılmaktadır.

3.6.3 İlgi alanı çıkarımı

İlgi alanı çıkarımında, resim ve videolarda belirlenen bir şablon hedef alınarak bütünsel bir çerçevede tanımlama yapılmaktadır. Ayrıca bu çıkarım, şablonu oluşturan çerçeve içerisinde algılanan çizgi parçaları aracılığıyla tanımlanmaktadır ve tahminlemenin en önemli parçasıdır.

ROI ile üretilen çerçeve, vücudun temsili eden çubukları kapsayan bir dikdörtgen kutudur. Uzuvların veya eklemlerin aynı yerde görüntü bölgeleri veya ortak yerleri olabilmektedir. Yapılandırılan konumlar arasındaki ikili uzamsal ilişkileri dışarda tutulması olağandır.

Kinematik ağaç modeli, eklemler arasındaki bağlantıları belirten bir yapıdır. Sağ ve sol taraftaki uzuvlar birbirine benzer göründüğü için tahminlemede zorluk yaşanabilmektedir. Veri kümesinde genişletilmiş görüntü segmentleri yer alacak şekildeki ağaç modellemesi, baş, üst gövde, her kol için iki ve her bacak için iki olmak üzere on segmentli gövde modeli şeklindedir. Üst vücut dedektörü diğerlerine oranla daha güvenilir ve doğru sonuçlara ulaşılmasını sağlamaktadır.

Ağaç modelleri için üst gövdede, gövdenin her iki tarafında yer alan kollar için aday yerleri üretmek amacıyla sağ kol, gövde ve sol kol olarak üçlü kombinasyon temel alınmaktadır. İnsan vücudunda uzuvların belirlenmesi zor olabilir ve bunun sebebi olarak uzuvlarda ve uzuv kenarlarında kontrastın düşük olmasıdır.

Tahminleme yapılan vücutta üst gövde, baş ve kol bölümleri dedektörleri kullanılır ve kollarda üst ve alt kol ayrımı yapılmaz. Kontrast düşüklüğünden dolayı üst kol için sanal bir kol dedektörü kullanılabilir. Bu yüzden vücudun üst kısmının ve kolun alt kısmının arasındaki kısımlarda sanal bir dedektör oluşturulur ve böylece tahminlemenin başarısı arttırılır.

Sanal dedektör için tekli özellikler, ikili özellikler, bir çift segment için uzaklık unsurları gibi birtakım özellikler temel alınmaktadır. Burada, ikili özellikler birden fazla parça çiftleri ile oluşturulan detektörlerdir ve üst vücut, baş, sol üst kol, sağ üst kol, sol alt kol, sağ alt kol olarak 6 adet parça kullanılmaktadır. Bu parça çiftleri için ayrı ayrı uzaklık, görünüm, açı ve örtüşmezlik özellikleri hesaplanmaktadır. Bununla birlikte bir çift segment uzaklığı gerçek mesafe ile birleştirilen uç noktalarını temsil eden vektörden oluşmaktadır. Yapılan bu ayrıştırmaların asıl sebebi aktivitelerin tanınmasını arttırma amacıyladır.

Video anlık görüntülerinin insan vücudunu içeren kısmı, tahmin için ilgi odağıdır. Çıkarımlardaki ana faktör, arka plandan bağımsız olarak insan vücudunun odaklanmasıdır. Bir sonraki adım ise, eşitlik (3.4) ile RGB bileşenlerinden gri tonlama bileşeninin elde edilmesidir..

$$\text{gri tonlama bileşeni} = 0.299\text{R} + 0.587\text{G} + 0.114\text{B} \quad (3.4)$$

Arka plan çıkarma işlemi sonucunda, insan vücudunun çizgi hali kilit noktaların birleştirilmesiyle meydana gelir. Arka plan çıkarma ile ortam ışığı, parlaklık, dış etkenler etkisi en aza indirilmektedir.

Poz mekanizmasındaki sınırlayıcı kutuların esas alınmasındaki en büyük etken, tahmin edilecek insan vücuduna odaklanmanın zorluklarının geride bırakılıp tahminde bulunmayı kolaylaştırmasıdır. Bu sistemde, arka plandaki zorluklar minimum etkiye sahip görünmektedir. Geçici pencerelerle ilgi alanı çıkarımı Şekil 3.10'da gösterilmektedir.



Şekil 3.10 Verilerin ilgi alanının analiz sonucunda çıkarılması

Ayak bileklerinin dışarıda kalmaması için y koordinatı kutunun tabanında 15 piksel aşağıya çekilmektedir. Bununla birlikte sınırlayıcı kutu sistemi aşağıdaki gibi hesaplanmaktadır.

$$\text{Sınırlayıcı kutu} = \text{Ölçek x } 200 \text{ x } 1.2 \tag{3.5}$$

$$Ayarlanmış\ sınırlayıcı\ kutu = \text{nesnepozisyonu. x} - \frac{\text{kenar}}{2} \tag{3.6}$$

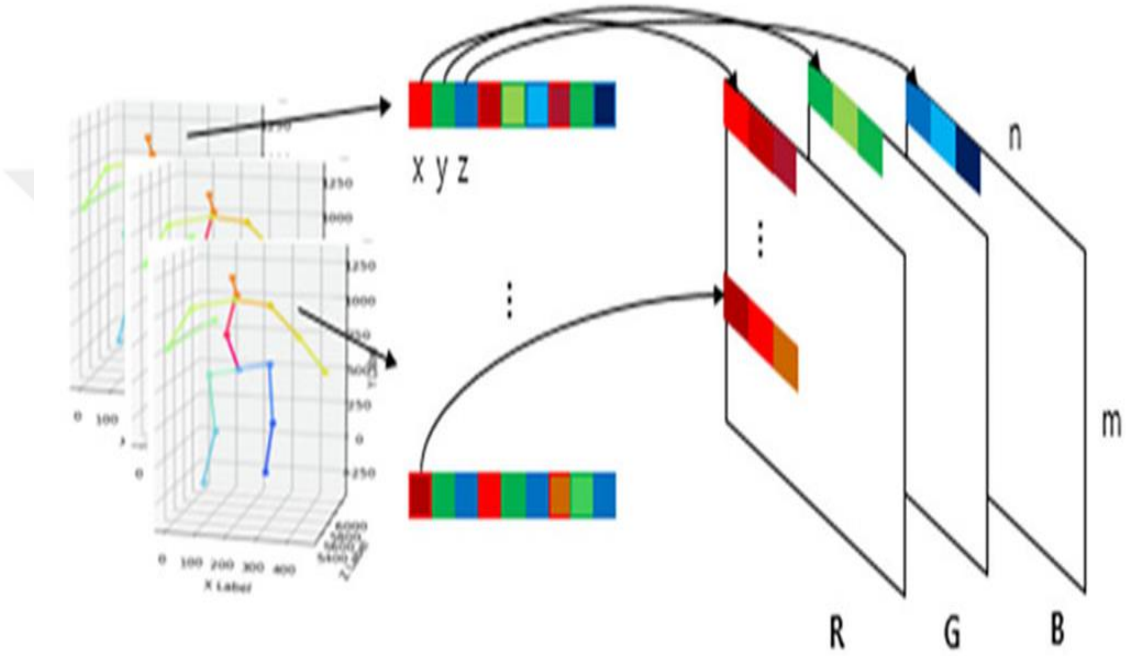
$$Ayarlanmış\ sınırlayıcı\ kutu = \text{nesnepozisyonu. y} - \frac{\text{kenar}}{2} + 15 \tag{3.7}$$

3.6.4 Rgb ve optik akış şekilleri

Moeslund, hareket kavramı çalışmasında hareketi 3 aşama olarak gruplandırmıştır. Bunlar, primitif eylem, eylem ve aktivite şeklindedir. Bununla birlikte insan vücudundaki uzuvların yaptığı hareketlere primitif eylem demiştir. Primitif eylemlerin oluşturduğu bütünsel eylem döngüsüne eylem, eylemlerin de oluşturduğu döngüye aktivite demiştir. Burada yer alan üç aşama tahminlemede de yer alan adımları kolaylıkla tanımlanmasına ve özelliklerinin çıkarılmasında büyük rol oynamaktadır.

Sonuç olarak bu resim ve videolardaki konum koordinatlarından hareketle, insanların bir sonraki aşamada nasıl tespit gösterileceği belirlenmeye çalışılmıştır[39].

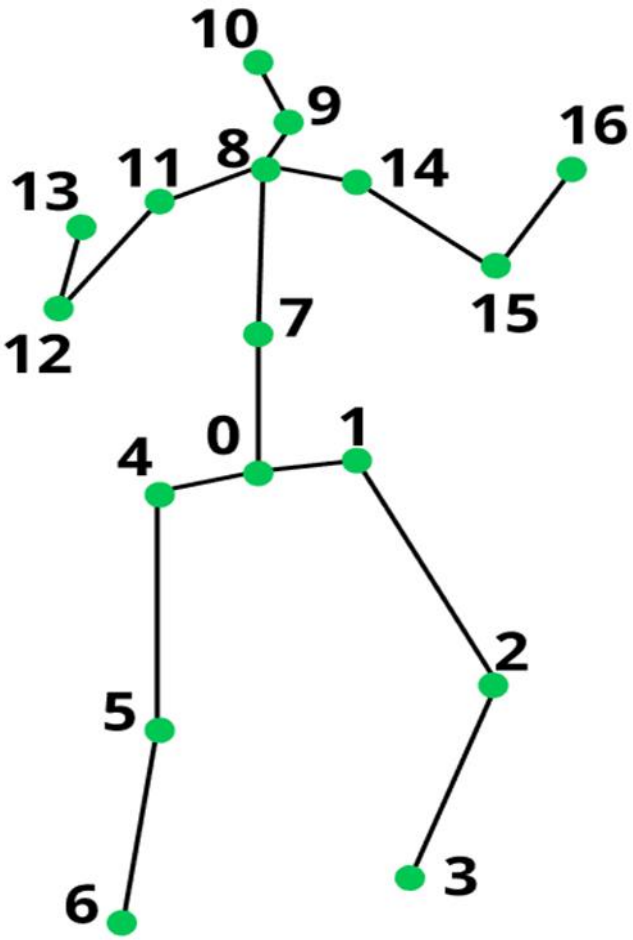
RGB bileşenleri ile görüntü analizi yapılarak konum koordinatları belirlenmektedir. Bu bileşenlerin görüntü analizi Şekil 3.11'de gösterilmektedir. Ortak konfigürasyonlar ile örnek veri setleri oluşturulabilmektedir.



Şekil 3.11 RGB bileşenleri için ortak veri dönüştürme görüntüsü [40]

Görüntülere dayalı insan eklem tespitleri esnek modelleme parçalarının sembolize edilmesi için bir örnek teşkil edebilir. Modeller karışım esnek bir yapı kullanılarak oluşturulmaktadır. Bununla birlikte bu işlem parçalar ile konumlar arasında ilişki kurularak gerçekleştirilmektedir. Ayrıca eklemler arasındaki ilişkiler belirlendikten sonra şablonlar oluşturulmaktadır.

Şekil 3.12'de optimize edilmiş eklemleri içeren bir ağaç yapısı gösterilmektedir. Yerel görünen eklemlerle ilgili her şey mekânsal ilişkilere ve bu eklemlerin oluşum şekillerine dayanmaktadır.



Şekil 3.12 İki Boyutlu Anahtar Noktalar [41]

Hareket, zaman ile 2 boyutlu düzlemin kesiştirilmesiyle oluşturulur. Öncelikle hareketin tespit edilmesi için görüntüde veya video da yer alan ilgi alanının kesitlerinin özelliklerinin net olarak belirlenmesi gerekmektedir. Burada doğru etiketin bulunması ile başarıya ulaşıla bilinir. Bununla birlikte yapılan işlemlerin, ilgi alanını kestirmesi gerekmektedir.

Optik akışın temel amacı, sadece insan vücudunun hareket yönünü tespit etmek yerine filtreleme işlemini gerçekleştirerek hareket halindeki eklem parçalarının mukavemetini belirlemektir. Böylece filtreleme işlemi, tahmin için gereksiz bilgilerin kaybolmasına yol açmış ve gerekli verilerin korunmasını sağlamıştır. Ayrıca hareketli görüntülerin kenar bilgilerinin birleştirilmesiyle desen oluşturulmaktadır.

Desen kullanılarak özellik çıkarımı yapıldığında, geçiş özelliklerini ortaya koyan histogramlar elde edilmektedir. Ancak görüntüdeki parlaklığın zamansal akışa bağlı olarak değişmediği ve her görüntü için zamanla değişmediği kabul edilmiştir [42].

Bu tez çalışmasında Hom-Schunk yöntemi kullanılmış ve yukarıda bahsedildiği gibi görüntüdeki parlaklığın zamansal akışa bağlı olarak değişmediği kabul edilmiştir ve eşitlik (3.8)'de gösterilmiştir.

$$(\frac{dI}{dt}) = 0 \tag{3.8}$$

Burada dI parlaklık şiddetinin zamansal türevidir. Ayrıca Zincir kuralı, (3.9)' daki gibi veri kümesindeki anlık görüntüye optik akış uygulandıktan sonra uygulanmaktadır.

$$\left(\frac{\mathrm{dI}}{\mathrm{dx}}\right) * \left(\frac{\mathrm{dx}}{\mathrm{dt}}\right) + \left(\frac{\mathrm{dI}}{\mathrm{dy}}\right) * \left(\frac{\mathrm{dy}}{\mathrm{dt}}\right) = 0 \tag{3.9}$$

Burada parlaklık şiddeti x, y koordinatlarına göre değişimi x, y koordinatlarının zamana göre değişimiyle hesaplanmıştır.

Eğer u, v eşitlik (3.10)'da gösterildiği gibi sırasıyla x ve y'nin zaman göre türevi olarak belirtilirse

$$\left(\frac{\mathrm{dx}}{\mathrm{dt}}\right) = \mathrm{u}, \left(\frac{\mathrm{dy}}{\mathrm{dt}}\right) = \mathrm{v} \tag{3.10}$$

(3.11) eşitliği

$$I'(x) * u + I'(y) * v = 0 \tag{3.11}$$

Elde edilir. Yukarıda elde edilen çıkarımlar, optik akışın özellikleri üzerinde tek başına kullanılmamaktadır. Temel piksellerin komşu piksellerinin benzer akış hareketlerine sahip olacağı düşünüldüğünde, sistemin çözümünü oluşturmak için (x, y) koordinat bileşenleri (3.12)'de gösterildiği gibi akıcılık kısıtına dahil edilmektedir,

$$\left(\frac{d^2u}{dx^2}\right) + \left(\frac{d^2u}{dy^2}\right) = \Delta^2 u \ , \left(\frac{d^2v}{dx^2}\right) + \left(\frac{d^2v}{dy^2}\right) = \Delta^2 v \ . \tag{3.12}$$

Optik akış değerlerini elde etmek için minimum parlaklık ve akıcılık kısıtlamaları gereklidir. Normal süre için sıfır hata olması gerekir. Ancak, bir parlaklık hatası oluşmakta ve bunu çözmek için arka planda denklem (3.13) kullanılmaktadır.

$$\varepsilon^2 = \iint \left(\alpha^2\varepsilon^2(\mathrm{cx}) + \varepsilon^2(by)\right)dxdy \tag{3.13}$$

Burada;

"ε" akıcılığı, "α" parlaklığı ve "b,c" katsayıları göstermektedir.

Optik akış yöntemiyle koordinat düzleminde değişim miktarı bulunur ve buna bağlı olarak x, y yönlerindeki değişim hesaplandıktan sonra optik akış güçleri hesaplanır. Bununla birlikte sınırlayıcı kutunun kenar özellikleri daha net belirlenir ve çevreye bağımlı olmadan insan vücut duruşunun net görünümü elde edilir ve akıcılık (3.14) eşitliğiyle bulunur.

$$\varepsilon^2 = (u^2 + v^2)^{\frac{1}{2}} \tag{3.14}$$

3.6.5 Sıralı yönlü gradyanlar ve destek vektör makinaları

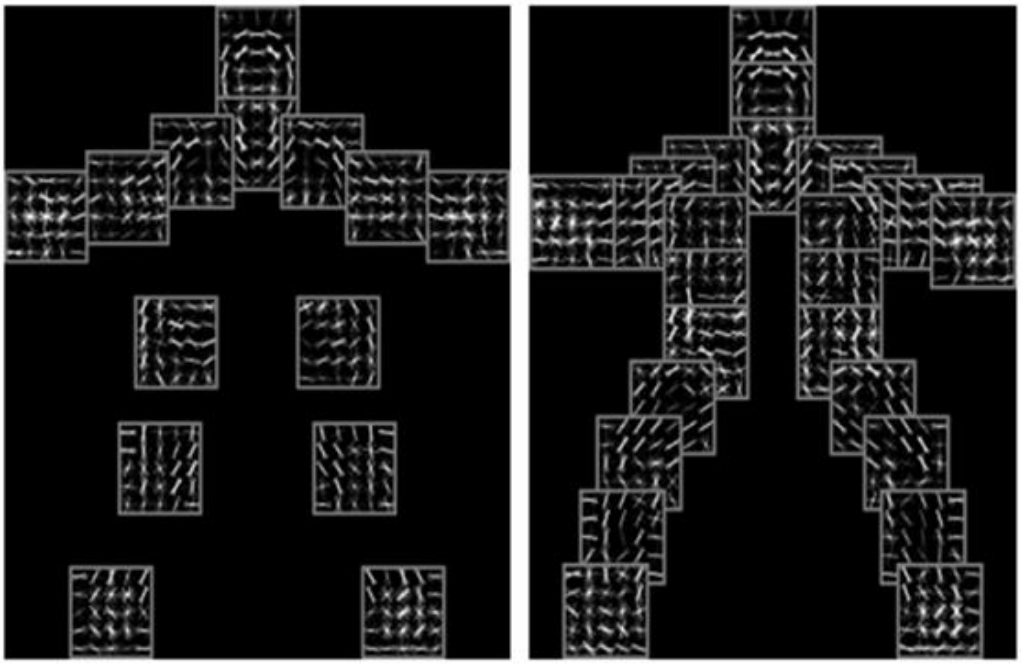
Optik akış gücü belirlendikten sonra hareket bilgilerini içeren kısım sıralı yönlü gradyanlardan oluşmaktadır. Görüntüdeki insan vücudunun kenar özelliklerinin çıkarılması HOG yöntemi ile.yapılmıştır.

Veriler arasındaki boşluklar değerlendirilerek veriler sınıflandırılmakta veya regresyon yapılabilmektedir. Destek vektör makineleri genellikle sınıflandırma için kullanılmaktadır. Kümeleme ya da sıra dışı verilerin tespiti için de uygulama alanları mevcuttur. Bu yöntem temel olarak verilerin birbirinden ayırt edilmesini sağlamaktadır. İnsan duruş tespitinde, destek vektör makineleri yerel lokalizasyon belirleme ve bu noktalarda yer alan eklemlerin tespitinde kullanılmaktadır. İki boyutlu insan duruş tahminlemesinde giriş ve çıkış değişkenlerini ayırmak için vektör değişken yöntemi kullanılmaktadır. Bu yöntem ile iki veya ikiden fazla veri etiketlerinin bulunduğu veri setlerindeki ayrım sağlanmaktadır.

Ayrım için gerekli özellikler veya kategoriler vektör değişkeni ile belirlenmektedir. Buna ilaveten bu yöntem kullanılarak en yakın noktalara denk gelen değerlerin ilgili sınıfa ait olmasını sağlamaktadır.

Vektör değişkenleri, noktaların konumuna bağlı olarak değişkenlik göstermektedir. Konuma bağlı en iyi maksimum durumu elde edildikten sonra sınıflandırma yapılır. Veriler arasındaki boşluğu maksimize etmek ve sınıflar arasına düzlem yerleştirilmesi, yeni gelecek verinin doğru kategori ve sınıfa ait olma olasılığını yükseltmektedir. Bu yöntemin en önemli avantajı daha iyi doğruluk değeri sunarak daha hızlı tahminin gerçekleştirilmesidir. Diğer yandan bu yöntemin dezavantajı ise eğitim süresinin fazla olmasıdır ve daha büyük veri setleri için daha fazla zaman harcanmaktadır.

Tahminleme yapılırken kısmi tıkanmalar ile karşılaşıldığında, küresel modellemeye başvurularak tıkanmaların önüne geçilmektedir. Öncelikle bir nokta belirlenir ve olası yakınlığa göre eklemler birleştirilmektedir. Şekil 3.13'te birleştirilen küreselleme modelleme sistemi şematize edilmiştir. Kullanılan parça şablonlarının sayısı doğru destek Vektör Makineleri ile tespit edilip arttırılırsa modellemedeki tahmin yüzdesi de aynı oranda başarı artışı göstermektedir.



a)14 sınırlayıcı kutu b)26 sınırlayıcı kutu

Şekil 3.13. Küresel modelleme sistemi [43]

4. BULGULAR VE TARTIŞMA

Bu bölümde, çalışma süresince elde edilen bulgulara yer verilmiştir. Tüm bu çalışmaların en temelinde belli başlı metotlar yer almaktadır. Bir hareket bilgisinin yanı sıra durgun görüntü bilgisinin aktarılması gerekmektedir

Öncelikle uzuvların birbirinden ayrılması ve karmaşıklığın olmaması için tüm eklem pozisyonları bulunmaya çalışılmıştır.

Vücut bölümlerinin belirlenmesi için eklem çiftleri belirlenmesi gerektiği görülmüştür Eklem çiftlerinin belirlenmemesi iskelet ağaç yönteminde karşıklık oluşturmuştur.

Sınırlayıcı kutuların x ve y koordinatları bulunurken öncelikle tepe bilgisi dâhil edilmelidir. Tepe bilgisi dâhil edilmediğinde algoritmanın, baş kısmındaki ağaç sistemini doğru tahminleyemediği görülmüştür. Sağ ve sol görüntüler ele alındığında uzuvların karıştırılmamasına dikkat edilmiştir. Bunun sebebiyet vereceği olumsuz sonuçlar önlenmeye çalışılmıştır. Bundan ötürü uzuvların koordinatlarının gerçekten doğru konumu verdiğinin testleri doğruluk oranlarına bağlı tespit edilmiştir.

İlgi alanı çıkarımı yapıldığında pozisyonlara bağlı ayak bileklerini görmediği durumlar tespit edilmiştir. Bu durumlarda ilgi alanındaki sınırlayıcı kutuyu piksel olarak büyütme işlemi yapılarak çözüm bulunmuştur.

Evrişimsel sinir ağlarında yer alan katmanların ölçü kriterleri ne kadar hassas ve özel olursa verileri o kadar iyi algılayıp daha doğru sonuçlar verdiği görülmüştür. Destek vektör makinaları ile uzuvlar küçük parça şablonları oluşturulup bir bütün elde edilmeye çalışılmıştır.

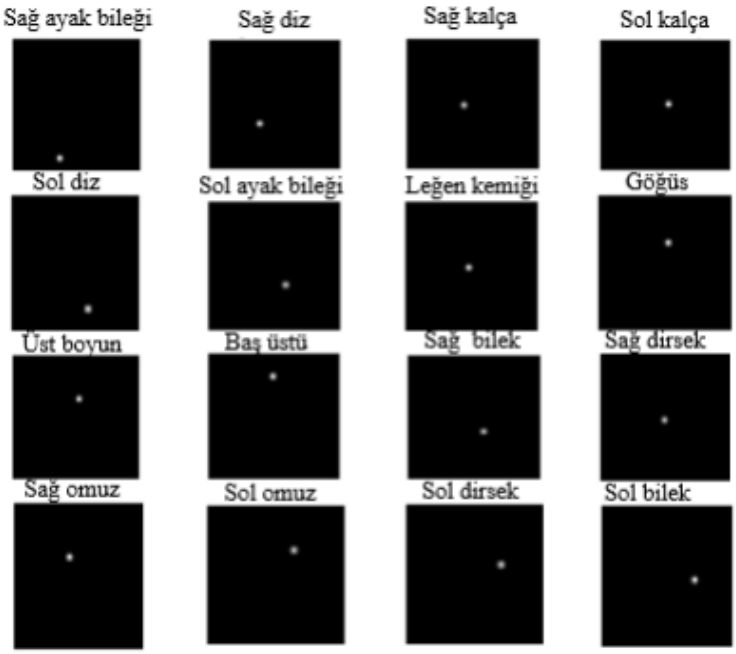
Evrişimsel sinir ağına eklenen Parse ve buff özellik kümelerinin eğitim aşamasında daha iyi performans sağlayacağı düşünülmüş ve bunlar dahil edildiğinde daha iyi eğitim sonuçları elde edilmiştir.

Sol vücut parçaları sağ görünümde daha az doğruluk gösterirken, sağ vücut parçaları sol görünümde daha az doğruluk göstermektedir. Arka görüş modelinde ise gözler ve burnu kapatan kilit noktalar daha az hassastır.

Optik akışın kullanılmasının amacı, insan vücudunun sergilediği hareketin yönünü tespit etmenin yanı sıra bir filtreleme işlemi uygulayarak hareket halindeki eklem parçalarını belirlemiştir. Böylelikle bu filtreleme işlemi tahminlemede gerekli bilgilerin korunmasına katkı sağlamıştır. Sonuç olarak hareketli görüntülerin kenar bilgilerinin birleşimi ile bir örüntü oluşturmuş ve özellik çıkarımı sağlayarak geçiş özelliklerini inceleyen histogramın oluşmasına ortam sağlamıştır.

Zaman kavramı da göz önünde tutularak yapılan tahminlemeler yüksek başarılı bir mekanizmanın ortaya çıkmasını sağlamıştır. Bu sayede istenilen anlık görüntü ve pozisyona bu tahminleme sistemi uygulanmıştır. Optik akış kullanımındaki hareket bilgisinin durgun görüntülerde HOG kullanımı ile kenar özelliklerin belirlenmesi sağlanmaktadır.

Şekil 4.1'te gösterilmekte olan referans görüntülerinin koordinatları belirlenmeye çalışılmıştır. Destek vektör makinelerinin bu koordinatlardaki sapmaları engellediği görülmüştür. Sağ ve sol omuz referans görüntüleri her türlü tahminleme sistemindeki ilk belirlenecek noktalardır.



Şekil 4.1 Anahtar noktalar için 16 referans görüntüsü [45]

Bu tez çalışmasında evrişimsel sinir ağı olarak Keras kullanılmıştır. Opencv programı içerisindeki Openpose kütüphanesinin kullanımıyla, görüntülerde İnsan vücudundaki 135 kilit noktasından temel olarak 17 kilit noktası alınmış ve tahminleme için gerçek zamanlı bir sistem oluşturulmuştur. Yazılan python programı ve Unity eklentileri ile eklemler arasındaki boru hatları etkinleştirilmektedir.

Veri seti, futbol maçlarının videolarından toplanmakta ve Cristiano Ronaldo'nun gol attığı anlardaki görüntülerinden oluşmaktadır. Tahminleme sisteminde kullanılmak üzere ön, sağ, sol ve arka görünümlerine göre anlık veriler alınmıştır. N-solid kinematik iskelet modeli, boru sistemiyle tahmin kalitesini en üst seviyeye çıkarmıştır. Bu sistemde ortam ışığı, anlık görüntüdeki parlaklık ve dış etkenler minimize edilmiştir. Tahminlemede kullanılan sistemin, veri setine uygulanmasına yönelik bazı örnekler Şekil 4.1'de gösterilmiştir. Şekilden görüldüğü gibi eklem çıkarımları ve iskelet ağaç modeli başarıyla oluşturulmuştur.



Şekil 4.2 Duruş Tahminlemesi Sonucu Elde Edilen Görüntüler

Veri setine uygulanan doğru pozlama, kişi katmanları ve anahtar tahminleme yöntemleri görüntüde bulunan kişilerin önden, sağ yandan ve sol yandan görünüşlerine uygulanmıştır. Bununla birlikte daha önceden bahsedilen 17 anahtar kilit noktası için bu yöntemlerin görünüm açıları için doğruluk oranları Tablo 1, 2, ve 3 te kıyaslanmıştır. Bu tablolardan, anahtar kilit yönteminin genel olarak daha başarılı olduğu görülmektedir. Görüntülerdeki kişilerin arkadan görünümleri için bahsedilen 3 yöntemin başarı oranları hemen hemen aynı olduğundan yer verilmemiştir.

Tablo 4.1 .Doğru Pozlama, Kişi Katmanları ve Anahtar Noktaları Yöntemleri ile Önden Görünüm Doğruluk Oranlarının Kıyaslaması

Anahtar Nokta Numaraları	Doğru Pozlama Yöntemi (%)	Kişi Katmanları Yöntemi (%)	Anahtar Noktaları Yöntemi (%)
0	91.5	90.5	92.7
1	88.9	87.2	90.6
2	89.6	89.1	91.3
3	88.7	88.4	89.5
4	83.7	84.3	87.6
5	89.4	89.9	90.2
6	82.6	85.4	89.3
7	86.1	87.8	90.5
8	87.6	88.1	89.6
9	88.4	89.0	90.9
10	89.6	91.5	89.9
11	90.1	89.7	90.3
12	85.3	87.8	89.6
13	90.1	90.2	90.7
14	84.3	85.4	89.8
15	90.9	90.5	91.1
16	90.7	89.5	90.1

Tablo 4.2. Doğru Pozlama, Kişi Katmanları ve Anahtar Noktaları Yöntemleri ile Sağdan Görünüm Doğruluk Oranlarının Kıyaslaması

Anahtar Nokta Numaraları	Doğru Pozlama Yöntemi (%)	Kişi Katmanları Yöntemi (%)	Anahtar Noktaları Yöntemi (%)
0	55.7	60.7	78.4
1	58.4	65.6	72.5
2	63.5	68.9	73.7
3	67.9	64.3	79.5
4	59.4	67.6	77.4
5	61.5	67.1	73.5
6	85.8	84.2	90.4
7	64.6	57.8	79.4
8	86.3	85.6	91.2
9	62.9	69.5	70.2
10	87.4	83.4	89.9
11	69.4	70.2	75.4
12	85.1	88.7	90.1
13	65.1	68.5	74.1
14	84.6	83.9	90.8
15	69.9	74.2	81.2
16	89.3	84.8	91.2

Tablo 4.3. Doğru Pozlama, Kişi Katmanları ve Anahtar Noktaları Yöntemleri ile Soldan Görünüm Doğruluk Oranlarının Kıyaslaması

Anahtar Nokta Numaraları	Doğru Pozlama Yöntemi (%)	Kişi Katmanlar Yöntemi (%)	Anahtar Noktaları Yöntemi (%)
0	53.7	62.7	77.3
1	57.4	63.6	73.6
2	62.5	69.9	74.8
3	64.9	61.3	78.5
4	57.4	68.6	79.5
5	71.5	61.1	91.7
6	84.8	80.2	70.3
7	63.6	59.8	93.4
8	87.3	81.6	90.6
9	67.9	70.5	92.8
10	84.4	80.4	85.4
11	64.4	70.2	92.4
12	85.1	83.7	71.1
13	71.1	75.5	90.6
14	86.6	81.9	72.0
15	68.9	71.2	93.2
16	83.3	80.8	70.7

Anahtar noktalarının ortalama hassasiyeti yöntemi en gerçekçi ve doğru bir değerlendirme oluşturmaktadır. Doğru pozlamada kullanılan dedektörler daha fazla adayın poz tahminlemesi için maksimum olmayan bastırma yöntemi ile başarımı azaltmaktadır. Kişi katmanları yöntemi maksimum olmayan bastırmadan etkilenmemektedir. Etkilenmemesinin sebebi kesin kanıt pencereleridir. Kişi katmanları yöntemi bu verileri kolay yorumlamak ve değerlendirmek açısından anahtar noktalarının ortalama hassasiyeti yöntemine göre daha hızlıdır. Görüntüde birden fazla aday kişi olduğunda kişi katmanları yöntemi daha fazla tercih edilmektedir.

5. SONUÇLAR VE ÖNERİLER

Bu tezde, insan pozu tahmin probleminin çözülmesinde literatürde mevcut olan çeşitli algoritmalar kullanılmıştır. Kullanılan veri seti bir futbolcunun gol atam anlarındaki fotoğraflardan oluşturulmuştur.

Bu çalışmada İlgi alanı çıkarımının oluşturulan sistemin başarımına pozitif yönde etkisi olmuştur. Böylece insan siluetlerinde dikdörtgen çıkarımı ile gereksiz olan arka plan ve gürültü ile karmaşaya sebep olacak bilgiler ortadan kaldırılmıştır.

Anahtar noktaların ortalama hassasiyeti yönteminin sağladığı en büyük avantaj destek vektör makinelerini de içinde barındırmasıdır. Bu vektörler eklemlerin ortalama koordinatlarını bularak istenilen düzeyde bir tahmin iskeleti oluşturulmasını sağlamasıdır. Kilit noktalarını belirlemek, eklemleri birbirine bağlamak, gövde kısımları pencere açıklıkları kullanmak, iskelet mekanizmasını fondan bağımsız hale getirmiştir.

Çalışmada Python yazılımında yer alan Opencv kütüphanesi kullanılarak evrişimsel sinir ağı oluşturulmuştur. Tezde yapılan çalışma başarım oranı literatürde olanlarla karşılaştırılmış ve başarım oranları Tablo 5.1' de gösterilmiştir. Tabloda yer alan 1. ve 3. çalışmalar bu tez çalışmasında elde edilen doğruluk oranı ile kıyaslandığında daha düşük, bunun yanı sıra 2. ve 4. çalışmalardaki doğruluk oranlarının daha yüksek olduğu görülmüştür. Bu farkların temel sebebi, tez çalışmalarında uygulanan yöntemlerdir. Bu tez çalışmasında elde edilen ortalama %92.2 doğruluk oranı evrişimsel sinir ağları, optik akış şekilleri ve destek vektör makinaları yöntemleri kullanılarak elde edilmiştir.

Tablo 5.1 Literatür ve yapılan çalışmaların kıyaslanması

	Literatürde yer alan çalışmalar	Ortalama Doğruluk Oranları (%)
1	Parekh, P., & Patel	89.6
2	Souvenir, R., & Babbs	92.7
3	Weinland, D., Ronfard	88.5
4	Rosebrock	95.3
5	Uygulanan Yöntem	92.2

Bu çalışmada özellikle dünyaca ünlü Cristiano Ronaldo'nun gol atma anlarındaki vücut yapısını çıkarılması hedeflenmiştir. İleriki çalışmalarda elde edilen sonuçların bir başka evrişimsel sinir ağıyla bu futbolcunun vücut şeklinin o anda topun kalenin sınırları içerisine gidip gitmeyeceğinin otomatik olarak bulunması çalışılacaktır. Bununla birlikte bu tip bir uygulama futbol ve basketbol maçları gibi bir topun hedefe ulaşması anında eylemi gerçekleştiren sporcunun vücut şeklinin bulunması için bilime bir katkı katacağı inanılmaktadır.

KAYNAKLAR

[1] P. Pareks and A. Patel, "Deep learning-based 2D and 3D human pose estimation: a survey," In Proceedings of Second International Conference on Computing, Communications, and Cyber-Security Springer, Singapore, s. 541-556, 2021.

[2] R. Souvenir and J. Babbs, "Learning the viewpoint manifold for action recognition," IEEE Conference on Computer Vision and Pattern Recognition., s. 1-7, 2008.

[3] Y. Yang and D: Ramanan, "Articulated human detection with flexible mixtures of parts," IEEE transactions on pattern analysis and machine intelligence, vol *35*, no 12, s. 2878-2890, 2012.

[4] Y. Wang, K. Huang and T. Tan ,"Human activity recognition based on r transform,", IEEE conference on computer vision and pattern recognition, s. 1-8, 2007.

[5] D. Ramanan,. "Learning to parse images of articulated bodies," Advances in neural information processing systems, vol. 19, 2006.

[6] J. Lee and B. Ahn,, "Real-time human action recognition with a low-cost RGB camera and mobile robot platform,", vol. 20, no. 10, s. 1-12, 2020 .

[7] D. Tran and D. Forsyth, " Improved human parsing with a full relational model," InEuropean Conference on Computer Vision, s. 227-240, 2010.

[8] D. Weinland, R. Ronfard and E. Boyer, "Free viewpoint action recognition using motion history volumes," Computer vision and image understanding, vol. 1 no. 104(2-3), s. 249-257, 2006.

[9] MC. Chang, H. Qi, X. Wang, H.Cheng and S. Lyu, "Fast Online Upper Body Pose Estimation from Video," vol. 1, s. 1-12, 2015.

[10] M. Eicher, V.Ferrari and Zurich, "Better appearance models for pictorial structures," vol. 2, no. 5, s. 1-11, 2009.

[11] Y. Yang and D.Ramanan, "Articulated human detection with flexible mixtures of parts," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 12, s. 2878-90, 2012.

[12] A. Toshev and C. Szegedy,. "Deeppose: Human pose estimation via deep neural networks," InProceedings of the IEEE conference on computer vision and pattern recognition, s. 1653-1660, 2014.

[13] S. Sethi, M. Kathuria and T. Kaushik, "Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread," Journal of biomedical informatics, vol. 120, s. 1-12, 2021.

[14] A. Rosebrock, “COVID-19: Face Mask Detector With Opencv, Keras/Tensorflow, And Deep Learning-Pyimage search,” .[online] PyImageSearch, 2020.

[15] A. Newell, K. Yang and J. Deng, “Stacked hourglass networks for human pose estimation,” InEuropean conference on computer vision, s. 483-499, 2016.

[16] N. Tanmay, A. Chen, “Using DeepLabCut for 3D markerless pose estimation across species and behaviors,” Nature protocols, vol. 14, no. 7, 12s.1-2, 2019.

[17] G. Taylor, J. Shotton and A. Hertzmann, “Metric regression forests for human pose estimation,”BMVC, s. 34-38, 2018.

[18] M. Andiluka, U. Iqbal, E. Insafutdinov, L. Pishchulin, A. Milan, J. Gall and B.Schiele, “Posetrack: A benchmark for human pose estimation and tracking,” In Proceedings of the IEEE conference on computer vision and pattern recognition, vol.20, no.5, s. 5167-5176, 2018.

[19] J. Wang and S. Payandeh, “Hand motion and posture recognition in a network of calibrated cameras,” Advances in Multimedia, s. 1-26, 2017.

[20] E. Remelli, S.Han, S. Honari, P. Fua and R.Wang, “Lightweight multi-view 3d pose estimation through camera-disentangled representation,” InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, s. 6040-6049, 2020.

[21] M. Zhao, T.Li, M. Abu Alsheikh, Y. Tian, H. Zhao, A. Torralba and D. Katabi, InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition, s. 7356-7365, 2018.

[22] N. Sarafianos, B. Boteanu, B. Ionescu and IA. Kakadiaris, “3d human pose estimation: A review of the literature and analysis of covariates”. Computer Vision and Image Understanding, vol.152, s. 1-20, 2016.

[23] G. Rogez and C.Schmid, “Mocap-guided data augmentation for 3d pose estimation in the wild,” Advances in neural information processing systems, s. 1-9, 2016

[24] W. Chen, H. Wang, Y. Li, H. Su, Z.Wang and C. Tu, “Synthesizing training images for boosting human 3d pose estimation,” IEEE Fourth International Conference on 3D Vision (3DV), s. 479-488, 2016.

[25] M. Fabbri, F. Lanzi, S.Calderara, S.Alletto and R. Cuchiara, “ Compressed volumetric heatmaps for multi-person 3d pose estimation, ” InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, s. 7204-7213, 2020.

[26] R. Liu, “Attention Based Temporal Convolutional Neural Network for Real-Time 3D Human Pose Reconstruction,” University of Dayton, 2019.

[27] Z. Cao, T Simon, SE Wei and Y Sheikh “Realtime multi-person 2d pose estimation using part affinity fields,” InProceedings of the IEEE conference on computer vision and pattern recognition, s. 7291-7299, 2017.

[28] X Kang, B Song, and F Sun, “A deep similarity metric method based on incomplete data for traffic anomaly detection,” In IoT. Applied Sciences,vol. 9, no. 1, s. 1-18, 2019.

[29] A Sengupta, I Budvytis and R Cipolla, “Synthetic training for accurate 3d human pose and shape estimation in the wild,” arXiv preprint arXiv:2009.10013, 2020.

[30] K Duan,S Bai, L Xie,H Qi,Q Huang and Q Tian, “Centernet: Keypoint triplets for object detection,” InProceedings of the IEEE/CVF international conference on computer vision, s. 6569-6578, 2019.

[31] VT Hoang, VD Hoang and KH Jo, “Realtime multi-person pose estimation with rcnn and depthwise separable convolution,” In2020 RIVF International Conference on Computing and Communication Technologies (RIVF), s. 1-5, 2020.

[32] N Sarafianos,B Boteanu, B Ionescu and IA Kakadiari, “3d human pose estimation: A review of the literature and analysis of covariates,” Computer Vision and Image Understanding, vol. 152, s. 1-20, 2016.

[33] C. Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke and A Rabinovich, “Going deeper with convolutions,” InProceedings of the IEEE conference on computer vision and pattern recognition, s. 1-9, 2015.

[34] K He, X Zhang, S Ren and J Sun, “Deep residual learning for image recognition,” InProceedings of the IEEE conference on computer vision and pattern recognition, s. 770-778, 2016.

[35] D Silver, A Huang, CJ Maddison, A Guez, L Sifre and G Van Den Driessche, “Mastering the game of Go with deep neural networks and tree search,” Nature, vol. 529, no 7587, s. 484-489, 2016.

[36] V Dumoulin and F Visin, “A guide to convolution arithmetic for deep learning,” arXiv preprint arXiv:1603.07285, 2016.

[37] PF Felzenszwalb and DP Huttenlocher, “Pictorial structures for object recognition,” International journal of computer vision, vol. 61, no. 1, s. 55-7, 2005.

[38] W Zeng, W Ouyang, P Luo, W Liu and X Wang, “3d human mesh regression with dense correspondence,” InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, s. 7054-7063, 2020.

[39] X Zhou,Q Huang ,X Sun,X Xue and Y Wei, “Towards 3d human pose estimation in the wild: a weakly-supervised approach,” InProceedings of the IEEE International Conference on Computer Vision, s. 398-407, 2017.

[40] M Andriluka, L Pishchuli, P Gehler and B Schiele, “2d human pose estimation: New benchmark and state of the art analysis,” InProceedings of the IEEE Conference on computer Vision and Pattern Recognition, s. 3686-3693, 2014.

[41] B Tekin, A Rozantsev, V Lepetit and P Fua (2016). “Direct prediction of 3d body poses from motion compensated sequences,” InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition. s. 991-1000, 2016

[42] M Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, “2D human pose estimation: New benchmark and state of the art analysis,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit, s. 3686–3693, 2014.

[43] B. Tekin, A. Rozantsev, V. Lepetit, and P. Fua, “Direct prediction of 3D body poses from motion compensated sequences,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit, s. 991–1000, doi: 10.1109/CVPR.2016.113, 2016.

[44] G Pavlakos , L Zhu, X Zhou and K Daniilidis, “Learning to estimate 3D human pose and shape from a single color image,” InProceedings of the IEEE conference on computer vision and pattern recognition, s. 459-468, 2018.

[45] H Rhodin, J Spörri, I Katircioglu, V Constantin, F Meyer, E Müller, M Salzmann and P Fua, “Learning monocular 3d human pose estimation from multi-view images,” InProceedings of the IEEE conference on computer vision and pattern recognition, s. 8437-8446, 2018.

ÖZGEÇMİŞ

Kişisel Bilgiler		
Soyad, Ad		Hediye Nupelda KANPAK
Web sayfası (Research Gate, Academia, vs.)		
Eğitim Bilgileri		
Derece	Kurum	Mezuniyet Yılı
Yüksek Lisans		
Lisans	D.Ü. Müh. Fak. Elk. Elktr. Müh.	2019
Lise	Toplukonut Anadolu Lisesi	2015
İş Deneyimi		
Dönem (Yıl)	Şirket, Kurum	Görev
2020	Dicle Elektrik Dağıtım A.Ş	Mühendis
Yabancı Dil		
İngilizce		
Yayınlar		
1. KANPAK, H. N., & ARSERİM, M. A. (2021). Human posture prediction by Deep Learning. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, *12*(5), 775-782.		
2.		
3.		
4.		
Özel İlgiler		

DİCLE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
TEZ BENZERLİK BİLDİRİMİ FORMU

Öğrencinin Adı, Soyadı	Hediye Nupelda KANPAK		
Öğrenci No	19805008		
Ana Bilim Dalı	Elektrik Elektronik Mühendisliği(Tezli YL)		
Program Türü	Proje ☐	Yüksek Lisans ☒	Doktora ☐
Tez Danışmanı (Ünvanı, Adı, Soyadı)	Dr. Öğretim Üyesi M. Ali ARSERİM		
(Varsa) II. Tez Danışmanı (Ünvanı, Adı, Soyadı)			
Tez Başlığı	Derin Öğrenme ve Anahtar Nokta Poz Yöntemiyle İnsan Duruşu Tahminlemesi		
RAPOR BİLGİLERİ			
Raporlama Aşaması	Tez Savunma Sınavı Sonrası		
Sayfa Sayısı	59		
Raporlama Tarihi	20.06.2022		
Benzerlik Oranı (%)	%4		

Yukarıda bilgileri verilen tez çalışmamın toplam 61 sayfalık kısmına ilişkin,24/05/2022.. tarihinde şahsım/tez danışmanım tarafından TURNİTİN isimli intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan intihal raporuna göre, tezimin benzerlik oranı % 4 olarak tespit edilmiştir.

Uygulanan filtrelemeler:

☐Başlangıç Bölümleri (Kabul ve Onay sayfası, Teşekkür sayfası, Özet/Abstract) hariç

☐Kaynaklar hariç

☐Alıntılar hariç/dâhil

☐Diğer (Açıklayınız)

Tezimin benzerlik oranı, Dicle Üniversitesi Fen Bilimleri Enstitüsü İntihal Raporu Uygulama Esaslarında belirtilen üst sınır benzerlik oranını aşmamaktadır. Benzerlik oranım üst sınır benzerlik oranının altında olsa dahi aksinin tespit edilmesi durumunda her türlü yasal sorumluğu kabul ettiğimi ve hukuki sonuçlarına razı olduğumu bildirir,
gereğini arz ederim.

Öğrencinin Adı, Soyadı: Hediye Nupelda KANPAK
Tarih: 20.06.2022
İmza:

Danışman Dr. Öğr. Üyesi M. Ali ARSERİM Tarih: 20.06.2022	İmza:
Ana Bilim Dalı Başkanı V. Dr. Öğr. Üyesi M. Ali ARSERİM Tarih: 20.06.2022	İmza: