



Teknoloji Fakültesi

MARMARA ÜNİVERSİTESİ

TEKNOLOJİ FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BİTİRME PROJESİ

RGB-D Kameralarının Görüntülerinden İnsan ve Nesne Hareketlerinin İstatistiksel 3D
Modellerinin Oluşturulması

PROJE YAZARI

Simge Arlanoğlu

170420045

Zeynep Koç

170420026

DANIŞMAN

Pr.Dr.Serhat Özekeş

İL, TEZ YILI

İstanbul, 2024

MARMARA ÜNİVERSİTESİ
TEKNOLOJİ FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

Marmara Üniversitesi Teknoloji Fakültesi Bilgisayar Mühendisliği Öğrencisi
..... nın “.....” başlıklı bitirme projesi çalışması,
..../..../..... tarihinde sunulmuş ve jüri üyeleri tarafından başarılı bulunmuştur.

Jüri Üyeleri

Prof. Dr. Adı SOYADI (Danışman)

Marmara Üniversitesi (İMZA)

Doç. Dr. Adı SOYADI (Üye)

Marmara Üniversitesi (İMZA)

Dr. Öğr. Üyesi Adı SOYADI (Üye)

Marmara Üniversitesi (İMZA)

İÇİNDEKİLER

Sayfa

SEMBOLLER/SYMBOLS.....	5
KISALTMALAR/ABBREVIATIONS.....	7
ŞEKİL LİSTESİ.....	8
TABLO LİSTESİ.....	9
ÖZET.....	10
ABSTRACT.....	11
1. GİRİŞ.....	12
1.1. Projenin Amacı.....	12
1.2. Literatür Özeti.....	13
2. MATERYAL VE YÖNTEM.....	15
2.1. Donanım.....	16
2.2. Yazılım Araçları.....	17
2.3. Veri Toplama ve İşleme Süreci.....	19
2.4. Genel İşleyiş.....	19
2.4.1 Yeniden yapılandırma.....	20
2.4.2 3D yönlendirmeli temas tahmini.....	20
2.4.3 Temas tabanlı iyileştirme.....	21
2.5. Kayıp Fonksiyonları.....	22
2.6. Uygulama Detayları.....	23
2.7. Deneyler ve Veri Seti.....	24
2.7.1 Test verilerinin işlenmesi.....	25
2.7.2 Bounding box değerinin hesaplanması.....	28
2.7.3 Görüntüdeki insandan ve nesneden maske oluşturulması.....	29
2.7.4 Veri Seti - BEHAVE.....	32
2.8. Ablation Çalışmaları.....	34
2.9. Değerlendirme Ölçütleri.....	35
2.10. Elde Edilen Sonuçların Blender 4.1’de İncelenmesi.....	36
2.10.1 Blender 4.1 Kullanımı.....	36
2.10.2. Blender 4.1’in Özellikleri.....	37
3. BULGULAR VE TARTIŞMA.....	38
3.1. Elde Edilen Sonuçlar.....	38
3.2. Literatür Taraması ve Karşılaştırma.....	39
3.3. Yurt İçi ve Yurt Dışı Çalışmaların Karşılaştırılması.....	40
3.4. Çalışmanın Önemi ve Gelecekteki Uygulamaları.....	40
3.5. Projenin Katkıları ve Geliştirilebilir Alanlar.....	41
4. SONUÇLAR.....	42
5. KAYNAKLAR.....	45
ÖZGEÇMİŞ.....	51

SEMBOLLER/SYMBOLS

I: Girdi görüntüsü

S_h : İnsan segmentasyonu

S_h : Nesne segmentasyonu

I_{input} : Birleştirilmiş giriş görüntüsü

M_h : Başlangıç 3D insan meshi

M_o : Başlangıç 3D nesne meshi

F : Görüntü özellikleri

θ_{body} : İnsan vücut parametreleri

θ_{hand} : El parametreleri

R_o : Nesne rotasyonu

t_o : Nesne translasyonu

C_h : İnsan temas haritası

C_o : Nesne temas haritası

F_{vh} : İnsan için 3D verteks özelliklerinin nihai boyutu

F_{vo} : Nesne için 3D verteks özelliklerinin nihai boyutu

F_{vh}^c : İnsan için maskelenmiş 3D tepe özellikleri

F_{vo}^c : Nesne için maskelenmiş 3D tepe özellikleri

M_h^{*} : İyileştirilmiş 3D insan meshi

\mathbf{M}_0^* : İyileştirilmiş 3D nesne meshi

$\mathbf{Contact}_p^{est}$: İletişim tahmini için hassasiyet (Precision)

$\mathbf{Contact}_r^{est}$: İletişim tahmini için geri çağırma (Recall)

\mathbf{CD}_{human} : Tahmin edilen ve Gerçek Değer (GD) meshler arasındaki 3D insan rekonstrüksiyonu için Chamfer mesafesi

\mathbf{CD}_{object} : Tahmin edilen ve Gerçek Değer (GD) meshler arasındaki 3D nesne rekonstrüksiyonu için Chamfer mesafesi

$\mathbf{Contact}_p^{rec}$: 3D insan ve nesne rekonstrüksiyonunu değerlendirirken iletişim açısından hassasiyet (Precision)

$\mathbf{Contact}_r^{rec}$: 3D insan ve nesne rekonstrüksiyonunu değerlendirirken iletişim açısından geri çağırma (Recall)

KISALTMALAR/ABBREVIATIONS

3D: Three-dimensional

4D: Four-dimensional

RGBD: Red-Green-Blue-Deep

GPU: Graphics Processing Unit

OpenCV: Open Source Computer Vision

SMPL+H: Skinned Multi-Person Linear model for Holes

CA: Cross-attention

FC: Fully Connected

SA: Simulated Annealing

GV: Google Ventures

YOLO: You Only Look Once

VOC: Visual Object Classes

IoU: Intersection Over Union

ASPP: Atrous Spatial Pyramid Pooling

CRFs: Conditional Random Fields

D: Derinlik

RNA: Ribonucleic Acid

GD : Gerçek Değer (Ground Truth) Meshler

ŞEKİL LİSTESİ

Şekil 2.1 Genel işleyiş.....	19
Şekil 2.2 Sınıflandırma Kaybı Formülü.....	26
Şekil 2.3 Nesne Varsa Güven Kaybı Formülü.....	26
Şekil 2.4 Nesne Yoksa Güven Kaybı Formülü.....	27
Şekil 2.5 Üç Hata Fonksiyonun Toplamı.....	27
Şekil 2.6 YOLOv3 Modelinin Başarı Grafiği.....	28
Şekil 2.7 DeepLabV3 Modelinin Mimarisi.....	31
Şekil 2.8 Maskı Alınmış Fotoğraf.....	31
Şekil 2.9 Behave Veri Seti İçeriği.....	32

TABLO LİSTESİ

Tablo 2.1 Yöntemlerin Karşılaştırılması.....	34
Tablo 2.2 3D yönlendirmeli temas tahmini için ablation çalışmaları.....	35

ÖZET

Günümüz dünyasında üç boyutlu modellemelere medikal, eğlence, bilişim gibi sektörlerde birçok farklı şekilde ihtiyaç duyulmaktadır. Bu projede de insan ve nesnenin birden çok RGB-D kamera tarafından tespit edilip yine aynı şekilde insan ve nesne arasındaki etkileşiminden doğan hareketlerin algılanıp üç boyutlu modellemesinin yapılması amaçlanmaktadır. Bu model sayesinde birçok farklı alanda verilerin doğru bir şekilde analiz edilmesi sağlanacak, hareketler canlı olarak tespit edilip bu süreç dijital bir ortama aktarılabilir. Böylece daha çok kişiselleştirilmiş avatarlar oluşturulabilecek, gerçek hayat senaryoları dijital ortama aktarılabilir, analiz süresince bu dijital ortamdan faydalanılması sağlanabilir. Birçok çalışmada sadece nesnelerin görüntüleri RGB-D kullanarak üç boyutlu olarak modellenmiştir. Bu projede ise yalnızca insan ve nesnelerin dijitalleştirilmesi değil, aynı zamanda birbirleri ile girdikleri etkileşimleri insan-nesne hareketlerinin mesafelerine de dikkat ederek modelleme yapıldığından daha önce gerçekleştirilmiş projelerden farklıdır. üç boyutlu modelleme alanında birçok proje yapılmıştır fakat projede kullanılan veri setinin kapsamının oldukça geniş olması, insan ve nesnenin etkileşimlerinin doğru olarak algılanması bir avantaj sağlamakla beraber, daha önceden yapılmış projelere de bir yenilik getirecektir. Projenin hayata geçirilmesi aşamasında, programlama dili olarak Python kullanılacaktır. Python'daki Numpy, Mesh, Torch, OPENCV kütüphaneleri, görselin işlenmesine, hareketin algılanmasına, verilerin dosyalanmasına ve tekrar kullanılmasına olanak sağlayacaktır. İnsan veya nesnelerin doğru bir şekilde modelleme yapılması için, oluşabilecek uyumsuzlukları gidermek amacıyla gerekli algoritmalar formüleştirecektir. Proje çeşitli sektörde farklı amaçlar için kullanılabilir. Bu alanlardan birkaçı: Oyun sektöründe kişiselleştirilmiş avatarlar oluşturup bu avatarların oyunlara entegre edilmesi, gerçek hayat senaryolarının dijital ortama aktarılıp analizlerinin daha kolay bir şekilde yapılmasıdır. Şu anki modellemelere ek olarak, doğruluğu daha yüksek modellemeler yapılacağı için sektörlerde doğan birçok farklı ihtiyacın giderileceği öngörülmektedir.

Anahtar Kelimeler : Üç boyutlu modelleme, RGB-D kamera, İnsan-nesne etkileşimi, Veri işleme.

ABSTRACT

In today's world, three-dimensional modeling is needed in various ways across sectors such as medical, entertainment, and information technology. This project aims to detect humans and objects using multiple RGB-D cameras and to model the movements resulting from human-object interactions in three dimensions. This model will enable accurate data analysis in various fields, detect movements in real-time, and transfer this process to a digital environment. Thus, more personalized avatars can be created, real-life scenarios can be transferred to digital environments, and these digital environments can be utilized during the analysis process. In many studies, only the images of objects have been modeled in 3D using RGB-D. However, this project differs by not only digitizing humans and objects but also modeling their interactions by considering the distances of human-object movements. This comprehensive approach, along with the extensive scope of the dataset used, offers significant advantages and introduces innovations to previous projects in the field of 3D modeling. Python will be used as the programming language in the implementation phase of the project. Libraries such as NumPy, Mesh, Torch, and OpenCV in Python will facilitate image processing, motion detection, data filing, and reuse. Necessary algorithms will be formulated to resolve potential discrepancies, ensuring accurate modeling of humans and objects. The project can be used in various sectors for different purposes. Some of these areas include creating personalized avatars in the gaming industry and integrating these avatars into games, as well as transferring real-life scenarios to digital environments for easier analysis. By providing more accurate models, our method addresses various emerging needs across sectors and achieves state-of-the-art performance in both human-object contact estimation and joint 3D reconstruction.

Keywords : Three-dimensional modeling, RGB-D cameras, Human-object interaction, Data processing.

1. GİRİŞ

3D (3 boyutlu) modelleme gelişen görüntü teknolojileri değerlendirildiğinde gittikçe daha çok ihtiyaç duyulan bir alan haline gelmektedir. Bu anlamda eğlenceden tıpa kadar birçok farklı sektörde 3D modelleme farklı şekillerde kullanılmaktadır. Tıp alanında RNA ikincil ve üçüncül yapılarını tahmin edebilmek için 3D modellemeye ihtiyaç duyulmuştur [1]. Dynamite Pete adlı oyunda, Autodesk Maya ve Unity 3D kullanılarak yüksek kalite ve gerçeklikte bir oyun ortaya çıkarılmıştır [2]. Kolların, ellerin ve baş yöneliminin robot üzerinde doğru çalışması için ilgili vücut parçalarının 3D görüntüsü elde edilerek robota entegre edilmiştir [3]. Risk analizinin yapılması ve doğal afet gibi durumlarda acil müdahale eylemlerinin planlanmasını desteklemek için kritik altyapıların ve bunların karşılıklı bağımlılıklarının belirtilmesi için 3D modellemeye ihtiyaç duyulmaktadır [4]. Ayrıca, 3D insan pozlarının ve gölgelerinin modellenmesi [5,6] ve avatarlarının oluşturulması [7,8] gibi çalışmalar gerçekleştirilmiştir. Konuyla ilgili yapılan literatür taramasında 3D modellemeye birçok farklı alanda gereklilik duyulması, gerçekleştirilen bu projenin önemini göstermektedir. Yapılan araştırmalar sonucunda bu kapsamda ülkemizde gerçekleştirilen çalışmaların yetersiz olduğu gözlemlenmiştir. Bu anlamda birçok çalışmada sadece nesnelerin görüntüleri RGB-D (Red, Green, Blue- Depth) kullanarak 3D olarak modellenmiştir. Bu proje ise insan ve nesne hareketlerini bir bütün olarak algılayıp modellemesi açısından daha önce gerçekleştirilmiş projelerden farklıdır. Ayrıca BEHAVE veri seti ilk tam vücut insan ve nesne etkileşimini bir bütün olarak sunması yönünden oldukça kullanışlıdır [10]. Modelleme mevcut olduğu sürece, insan-nesne etkileşimlerinin ulusal/uluslararası modellemesi, oyun, sanal ve karma gerçeklik, insan analitiği ve insan-robot işbirliği gibi birçok uygulamanın merkezinde yer alacaktır. Sadece genel bir insan yapısını değil insanın davranışlarını da incelediği için normalden daha kişiselleştirilmiş avatarların oluşmasında önemli bir rol oynayacaktır. Gerçek hayattan alınan senaryolardan dijital bir ortam oluşturularak o anki durumun analizinin yapılması kolaylaştırılabilecektir.

1.1. Projenin Amacı

Proje yalnızca insan ve nesnelerin dijitalleştirilmesi değil, bununla birlikte insan-nesne etkileşiminin de (mesafelerine de dikkat ederek) gerçeğe en yakın biçimde modellenmesi amaçlamaktadır. Gerçek bir ortamdaki insan ve nesne hareketlerine ait verilerin kameralar tarafından elde edilmesi sayesinde, lazer tarayıcılara kıyasla zaman ve maliyetten tasarruf

edilerek dijital bir ortam oluşturulması sağlanmıştır. [11] Bu ortam, halihazırda var olan veri setleri sayesinde 4D modellemeye ihtiyaç duyulmadan oluşturulmuştur. Kullanılacak olan BEHAVE veri setinde 5 farklı açıdan elde edilen, yaygın 20 nesne ile 8 farklı öznenin çok çeşitli etkileşimlerde bulunduğu yaklaşık olarak 15 bin karelik görüntüler bulunmaktadır [10]. Bu görüntüler yüksek derinlikli olup, diğer 3D tarayıcılara kıyasla çok daha ucuz ve farklı olan RGB-D sensörler kullanılarak elde edilmiştir. Özellikle, insan davranışlarını incelemeye yönelik olması, kişiselleştirilmiş avatarların yaratılmasına ve bu avatarların oyunlara ve sinema projelerine entegre edilmesine olanak tanımaktadır. Projede kullanılan RGB-D sensörler, diğer 3D tarayıcılara göre daha düşük maliyetli olup, yüksek kaliteli veriler elde edilmesine olanak tanımaktadır. Bu sensörlerle elde edilen veriler, insan ve nesne hareketlerinin dijital ortama aktarılması ve etkileşim modellerinin oluşturulması için idealdir.

Projede kullanılan veri seti, 4D modellemeye ihtiyaç duyulmadan oluşturulabilen ve gerçekçilik oranı yüksek dijital ortamların oluşturulmasını sağlayan özelliklere sahiptir. Bu veri setinde bulunan yüksek kaliteli veriler ile dijitalleştirilmiş ortamın gerçekçilik oranı yüksek tutulmuştur. Projede kullanılan veri seti amaçlanan modelleme için özel olarak oluşturulmuş ve istenilen özellikleri kapsamaktadır. İnsan ve nesne arasındaki etkileşim modellemeleri oyun, sanal ve karma gerçeklik, insan analizi ve insan-robot işbirliği gibi birçok uygulamanın merkezinde yer alacaktır. Ayrıca kullanılacak olan veri setinin insan davranışlarını incelemeye yönelik olması sayesinde gerçeğe daha yakın olarak kişiselleştirilmiş avatarların yaratılabilmesini sağlayabilecektir. Bu avatarlar oyunlara entegre edilebilir ve yine aynı şekilde, sinema sektöründe birçok projeye ilham olabilecektir.

1.2 Literatür Özeti

Eskiden nesnelerin 3D modellerini oluşturmak, çok zaman alan bir süreç iken günümüzde 3D modellemeye duyulan ihtiyaç sebebiyle ilerleyen teknolojiler kullanılarak birçok farklı yöntemle nesneler gerçekçi bir şekilde dijital ortama aktarılabilir. [9]. Yapılan araştırmalar sonucunda bu kapsamda ülkemizde gerçekleştirilen çalışmaların yetersiz olduğu gözlemlenmiştir. Araştırma kapsamında daha önce yapılan yerli çalışmalar incelendiğinde çok fazla kaynağa erişilememekle birlikte incelenen çalışmalarda etkileşimin modellenmemesi ve farklı açılardan elde edilmiş imajların kullanılmaması gibi eksikliklerle karşılaşmıştır. Yapılan diğer çalışmalarda nesnelerin, insanların ya da nesne

ve insanların dijitalleştirilmesi hedeflenmiş ancak nesne ve insan etkileşimlerinin dijitalleştirilmesi sağlanamamıştır [9, 12, 13].

Son yıllarda insan-nesne etkileşimi araştırmalarında önemli ilerlemeler kaydedilmiş olup, bu çalışmalar esas olarak insan-nesne etkileşimi tespiti, rekonstrüksiyonu ve statik sahnelerle etkileşime giren insanlar oluşturma üzerine odaklanmıştır. Çoğu girişim, bilgisayar grafikleri[14, 15, 16], bilgisayar vizyonu [17, 18, 19, 20] ve robotikte [21, 22, 23, 14] yalnızca el-nesne etkileşimlerini sentezlemeye yönelmiştir. Ancak, statik, eklemli ve dinamik nesnelerle bütün vücut etkileşimleri oluşturma konusuna da artan bir ilgi görülmektedir. Tam vücut insan nesne etkileşimi sentezi üzerine kinematik temelli yaklaşımlar ve fizik simülasyon ortamlarının uygulanmasında da ilerlemeler kaydedilmiştir. Bununla birlikte, çoğu yaklaşım, nesneye yaklaşma veya nesneleri manipüle etme üzerine odaklanmak gibi eylem ve nesne çeşitliliği konusunda sınırlamalara sahiptir.

İnsan-nesne etkileşimi tahminindeki öncü çalışmalar genellikle 2D temas [24], 3D eklemler arası temas [26-27, 28, 29] veya 3D yama düzeyinde temas şeklinde teması temsil etmektedir. Son zamanlarda, birkaç çalışma [31, 32, 33], insan vücut yüzeyinde (örneğin, SMPL [30]) tanımlanan yoğun bir köşe düzeyinde temas haritasını tahmin etme sorununu ele almaktadır. POSA [31], görüntü kanıtı kullanmadan insan pozunu veren bir koşullu değişkenli otomatik kodlayıcı (cVAE) [34] önerdi ve hangi köşelerin nesnelerle temas halinde olabileceğini çıktı olarak verdi. BSTRO [32], vücut köşeleri arasındaki bağlamsal ilişkileri öğrenen bir Transformer göstermiştir. DECO [35], insan vücut parçalarını ve sahne bağlamlarını birlikte kullanarak temas tahmini için çapraz-dikkat tabanlı bir ağ önermiştir.

Bu yöntemler, insan ve nesne yüzeylerinin 3D geometrisini öğrenmeden sadece tahmin edilen ve gerçek temas haritaları arasındaki çapraz-entropi kaybı ile eğitilmiştir. Öte yandan, bu projedeki yöntem insan ve nesne 3D ağlarını yeniden oluştururken aynı anda insan-nesne temas haritalarını öğrenir ve bu yaklaşımın iki avantajı vardır. İlk olarak, 3D insan ve nesne ağları, insan ve nesne ile ilgili yerel görüntü bölgelerine odaklanılacak yerleri gösterir. Bu, 3D ağların giriş görüntüsüne yansıtılmasıyla elde edilen 2D köşe koordinatları ile sağlanır. İkinci olarak, köşe başına 3D koordinatlar, insan ve nesne yüzeyleri arasında 3D pozisyonel ilişki sağlar ve insan-nesne teması üzerine 3D geometrik

düşünme sağlar. Bu iki avantajın altında, 3D rehberli temas tahmini, önceki yöntemlere kıyasla insan-nesne temasını daha etkili bir şekilde yakalar.

Son yıllarda 3D insan ve nesne yeniden oluşturulmasının çoğu çalışması [3,33,35, 36], insan-nesne etkileşiminin kısıtlamalarını karşılamak için iteratif olarak 3D insan ve nesne ağlarına uyum sağlayan optimizasyon temelli yaklaşımlar üzerine odaklanmıştır. Bu optimizasyon temelli yöntemler, insan eylemlerine dayalı insan-nesne etkileşimi öncelikleri tasarlamış olan Holistic++, insan ve nesne ağlarını önceden tanımlanan temas çiftlerine dayalı olarak uyum sağlayan optimizasyon çerçeveleri sunan PHOSA [36] ve D3D-HOI [37] ve CHORE [38] tarafından önerilen iki aşamalı bir yaklaşımdır. Bu yöntemlerin hepsi, yeniden oluşturma için optimizasyon hedeflerine (örneğin, 2D silüetler) dayanır ve görüntü bağlamını dikkate almaz. Bu tür bir yaklaşımın bir kısıtlaması, eksik optimizasyon hedeflerine karşı hassasiyetidir. Bunun nedeni, hedeflerin tahmin yoluyla edinilmiş olmasıdır ve bazı optimizasyon hedeflerinin (örneğin, 2D silüetlerin derinlik belirsizliği) yeniden oluşturma için belirsiz olmasıdır. Bu nedenle, optimizasyon yöntemleri sıklıkla eksik optimizasyon hedeflerine doğru eğilimli olup başarısız olabilir. Bu yöntemlerden farklı olarak, projemiz yukarıda belirtilen sorunlardan bağımsız olan bir sonlandırma ve öğretme yaklaşımıdır. Bunun nedeni, sonlandırma aşamasında, sistem veri tabanlı bilgilerden elde edilen bilgiler doğrultusunda çıktı üretir.

Çoğu projenin modelleri, tahmin edilen ve gerçek temas haritaları arasındaki cross-entropy loss ile eğitilirken, insan ve nesne yüzeylerinin 3D geometrisini öğrenmemektedir. Bu proje ise insan-nesne temas haritalarını öğrenirken, 3D insan ve nesne ağlarını da yeniden yapılandırarak, etkileşimlerinin tahmininde iki önemli avantaja sahiptir. İlk olarak, 3D insan ve nesne ağları, 3D ağların giriş görüntüsüne yansıtılmasıyla elde edilen 2D verteks koordinatları ile insan ve nesne ile ilgili yerel görüntü bölgelerine odaklanmada rehberlik sağlar. İkincisi, her verteks 3D koordinatları, insan ve nesne yüzeyleri arasındaki 3D konumsal ilişkiyi sağlayarak, insan ve nesne arasındaki teması 3D geometrik olarak anlamlandırmaya olanak tanır. Bu iki avantaj altında önceki yöntemlere kıyasla insanın nesneye temasını yakalamada çok daha etkilidir.

2. MATERYAL VE YÖNTEM

Bu çalışmada, insan ve nesne etkileşimlerini 3D olarak sanallaştırmak amacıyla kullanılan donanım ve yazılım araçları aşağıda detaylandırılmıştır. Çalışma, RGB-D kameralarla elde

edilen görüntü verilerini işleyerek, bu verileri 3D modellere dönüştürmek için çeşitli yazılım kütüphanelerini kullanmayı hedeflemektedir.

2.1. Donanım

RGB-D Kamera: RGBD kameralar, renkli görüntülerin yanı sıra derinlik bilgisi de sağlayan gelişmiş görüntüleme cihazlarıdır. Bu çalışmada kullanılan RGB-D kamera, insan ve nesne etkileşimlerini yüksek doğrulukla yakalamak için tercih edilmiştir. Kamera, hem renkli (RGB) hem de derinlik (D) bilgisi sunarak, görüntülerin üç boyutlu olarak modellenmesine olanak tanır. RGBD kameraların sağladığı derinlik bilgisi, nesnelerin ve insanların mekânsal konumlarının ve yüzey özelliklerinin belirlenmesinde kritik rol oynar.

Bilgisayar: Verilerin işlenmesi ve 3D modellemelerin gerçekleştirilmesi için yüksek işlem gücüne sahip bir bilgisayar kullanılmıştır. Bilgisayarın donanım özellikleri, büyük veri kümelerinin işlenmesi ve karmaşık algoritmaların çalıştırılmasına uygundur. Özellikle, güçlü bir GPU (Grafik İşlem Birimi), derin öğrenme modellerinin eğitiminde ve büyük veri kümelerinin hızlı bir şekilde işlenmesinde büyük avantaj sağlamaktadır. Bilgisayarın kullanılan donanım özellikleri, şunları içermektedir:

- **İşlemci:** Yüksek performanslı çok çekirdekli işlemci, veri işleme ve hesaplama yoğun görevler için idealdir.
- **Bellek:** Geniş bellek kapasitesi, büyük veri kümelerinin işlenmesi için gereklidir.
- **Grafik İşlem Birimi (GPU):** Derin öğrenme modelleri ve 3D grafik işlemleri için hızlandırma sağlar.
- **Depolama:** Hızlı ve geniş depolama alanı, büyük veri kümelerinin depolanması ve erişilmesi için önemlidir.
- **Soğutma Sistemi:** Yüksek performans gerektiren işlemler sırasında bilgisayarın soğutulması için etkili bir sistem bulunmaktadır.

Bu donanım bileşenleri, çalışmanın karmaşıklığına ve gereksinimlerine uygun olarak seçilmiş ve optimize edilmiştir. İşlem gücü ve veri işleme kapasitesi, RGB-D kameradan elde edilen verilerin analiz edilmesi ve derin öğrenme algoritmalarının uygulanması için kritik öneme sahiptir.

2.2. Yazılım Araçları

NumPy: NumPy, bilimsel hesaplamalar için kullanılan temel bir Python kütüphanesidir. Bu kütüphane, büyük veri kümeleriyle çalışmayı kolaylaştıran çok boyutlu diziler ve matrisler için destek sağlar. NumPy'nin sağladığı güçlü matematiksel fonksiyonlar, verilerin işlenmesi ve analiz edilmesinde kritik rol oynar. Bu projede, RGBD kameralarla elde edilen ham verilerin işlenmesi ve analiz edilmesi için NumPy kütüphanesi kullanılmıştır.

Mesh: Bu çalışmada, 3D modelleme ve mesh işleme için Python dilinde geliştirilmiş olan Mesh kütüphanesi kullanılmıştır. Bu kütüphane, nesnelerin ve insanların 3D modellerinin oluşturulması, manipülasyonu ve analizinde kritik bir rol oynamaktadır. Mesh kütüphanesi, 3D model yüzeylerinin üç boyutlu temsili için çeşitli araçlar sunar. Bu araçlar arasında yüzey pürüzsüzleştirme, delik kapatma ve yüzeylerin yeniden yapılandırılması gibi işlemler yer almaktadır. Yüzey pürüzsüzleştirme, modelin estetik görünümünü iyileştirirken, delik kapatma işlemi modelin bütünlüğünü sağlamak adına önemlidir. Ayrıca, kütüphane modelin düşük poligon sayısına sahip olmasını veya belirli kalınlıkta yüzeylerin oluşturulmasını sağlayan yeniden yapılandırma araçları sunar. Mesh kütüphanesi ayrıca, 3D mesh verilerini işleme ve analiz etme yetenekleri sunar, bu da modeldeki özellikleri çıkarma ve mesafe hesaplama gibi işlemleri içerir. Son olarak, kütüphane farklı dosya biçimlerine dışa aktarma ve diğer 3D modelleme araçlarıyla entegrasyon için destek sağlar. Bu özellikler, Mesh kütüphanesinin bilimsel araştırmalarda, mühendislik uygulamalarında ve sanal gerçeklik projelerinde geniş bir kullanım alanı bulmasını sağlar.

Torch: derin öğrenme modelleri geliştirmek ve eğitmek için kullanılan güçlü bir kütüphanedir. Özellikle PyTorch olarak bilinen sürümü, araştırma ve uygulama geliştirme alanında geniş bir kullanıcı kitlesine sahiptir. Bu çalışmada, insan ve nesne etkileşimlerini modellemek ve analiz etmek için derin öğrenme tekniklerinden yararlanılmıştır. Torch kütüphanesi, modellerin hızlı ve etkili bir şekilde eğitilmesine ve optimize edilmesine olanak tanır.

PyTorch'un sağladığı avantajlar arasında esneklik, hızlı prototipleme, GPU hızlandırma desteği ve zengin bir model seçenekleri kütüphanesi bulunmaktadır. Derin öğrenme modellerinin eğitimi sırasında otomatik farklılaşma ve grafik hesaplama özellikleri, model

geliştirme sürecini büyük ölçüde kolaylaştırır. Ayrıca, PyTorch'un dinamik hesaplama grafikleri, karmaşık modellerin esnek bir şekilde tanımlanmasını ve eğitilmesini sağlar.

Torch kütüphanesi, özellikle derin öğrenme modelleri üzerine yapılan çalışmalarda yaygın olarak kullanılmaktadır. Bu kütüphane, kullanıcılarına geniş bir öğrenme algoritması yelpazesi sunmanın yanı sıra, model performansını optimize etmek için çeşitli optimizasyon teknikleri ve araçlar sağlar. Bu çalışmada da, PyTorch'un bu özelliklerinden faydalanılarak insan ve nesne etkileşimlerini anlamak için derin öğrenme modelleri geliştirilmiş ve eğitilmiştir.

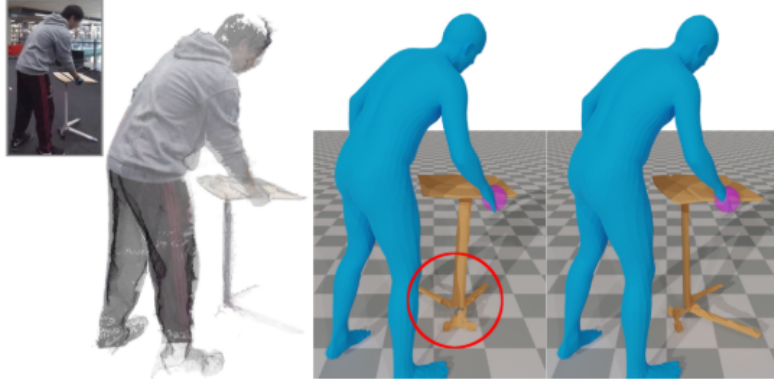
OpenCV: bilgisayarla görme alanında geniş çapta kullanılan açık kaynaklı bir kütüphanedir. Bu kütüphane, görüntü işleme ve analiz işlemleri için çeşitli araçlar sunar. OpenCV, RGB-D kameralarla elde edilen görüntülerin işlenmesi, segmentasyon, özellik çıkarımı ve nesne tespiti gibi işlemler için kullanılmıştır. OpenCV'nin sağladığı esneklik ve geniş araç seti, bu projede çeşitli görüntü işleme görevlerinin yerine getirilmesinde önemli bir rol oynamıştır.

OpenCV'nin sunduğu temel özellikler arasında görüntü filtreleme, kenar tespiti, nesne takibi ve hareket analizi yer almaktadır. Kütüphane, farklı görüntü formatlarını destekleyerek, çok çeşitli uygulamalarda kullanılabilirlik sağlar. Özellikle RGB-D kameralarla elde edilen verilerin işlenmesi sırasında, derinlik bilgisi ve renkli görüntülerin entegrasyonu, OpenCV'nin sunduğu araçlar sayesinde kolayca gerçekleştirilebilir.

Bu çalışmada, OpenCV kütüphanesi kullanılarak elde edilen görüntüler üzerinde segmentasyon işlemleri yapılmıştır. Segmentasyon, görüntüdeki nesnelerin veya bölgelerin belirlenmesi ve ayrılması işlemini ifade eder. Ayrıca, özellik çıkarımı işlemleri ile görüntülerdeki belirli özelliklerin (kenarlar, köşeler vb.) tespit edilmesi sağlanmıştır. Nesne tespiti ise, görüntülerdeki belirli nesnelerin tanınması ve konumlarının belirlenmesi sürecini içerir. OpenCV, bu işlemlerin her birini gerçekleştirmek için optimize edilmiş algoritmalar ve araçlar sunar.

Sonuç olarak, OpenCV kütüphanesi, bu projede görüntü işleme görevlerinin etkin bir şekilde yerine getirilmesine önemli katkılarda bulunmuştur. Sağladığı esneklik ve geniş araç seti, karmaşık görüntü işleme ve analiz işlemlerinin gerçekleştirilmesini mümkün kılmıştır. Bu nedenle, OpenCV, bilgisayarla görme projelerinde yaygın olarak tercih edilen bir kütüphane olarak öne çıkmaktadır.

2.3 Veri Toplama ve İşleme Süreci



Şekil 2.1 Genel işleyiş

Bu projede, insan ve nesne etkileşimlerini 3D olarak sanallaştırmak için, RGBD kameralarla elde edilen görüntü verileri kullanılmıştır. Veri toplama süreci, kameranın doğru bir şekilde kalibre edilmesi ve uygun aydınlatma koşullarının sağlanmasıyla başlamaktadır. Toplanan veriler, hem renkli görüntüleri hem de derinlik bilgilerini içermektedir. Bu veriler, NumPy kütüphanesi kullanılarak işlenmiş ve analiz edilmiştir.

Görüntü işleme aşamasında, OpenCV kütüphanesi kullanılarak görüntüler segmentlere ayrılmış ve nesneler ile insanlar tespit edilmiştir. Elde edilen segmentler, Mesh kütüphanesi kullanılarak 3D modellere dönüştürülmüş ve bu modeller, çeşitli mesh işleme teknikleri ile iyileştirilmiştir. Derin öğrenme modellerinin eğitimi ve optimizasyonu için Torch kütüphanesi kullanılmıştır. Modeller, insan ve nesne etkileşimlerini doğru bir şekilde tahmin edebilmek için büyük veri kümeleri üzerinde eğitilmiştir.

Bu materyal ve yöntemlerin birleşimi, insan ve nesne etkileşimlerini üç boyutlu olarak sanallaştırma hedefine ulaşmada önemli bir rol oynamaktadır. Projede kullanılan donanım ve yazılım araçlarının kombinasyonu, verilerin doğru bir şekilde işlenmesini ve analiz edilmesini sağlayarak, etkili ve verimli bir 3D modelleme süreci sunmaktadır.

2.4 Genel İşleyiş

Şekil 2.1, projemizin genel işleyişini göstermektedir. Bu işleyiş üç aşamadan oluşmaktadır: yeniden yapılandırma, 3D yönlendirmeli temas tahmini ve temas tabanlı iyileştirme.

2.4.1 Yeniden yapılandırma

Görüntü I'nin, insan segmentasyonu \mathbf{S}_h , ve nesne segmentasyonu \mathbf{S}_o 'nun birleştirilmiş girdileri $\mathbf{I}_{\text{input}} \in \mathbb{R}^{5 \times H \times W}$ kullanılarak, başlangıç 3D insan ve nesne meshleri ($\mathbf{M}_h \in \mathbb{R}^{431 \times 3}$ and $\mathbf{M}_o \in \mathbb{R}^{64 \times 3}$) elde edilir; burada H ve W sırasıyla görüntünün yüksekliğini ve genişliğini belirtir. Önceki çalışmaları takip ederek [38, 39], insan ve nesne segmentasyonları, hem eğitim hem de çıkarım için DetectronV2 [40] kullanılarak elde edilir. Girdiler $\mathbf{I}_{\text{input}}$ 'tan, bir backbone ağı (yani, ResNet-50 [41]) görüntü özelliği $\mathbf{F} \in$

$\mathbb{R}^{2048 \times H/32 \times W/32}$ çıkarır. Başlangıç 3D insan meshi olan \mathbf{M}_h 'ı elde etmek için, SMPL+H modeli [30] insan vücut parametreleri $\theta_{\text{body}} \in \mathbb{R}^{76}$ ve el parametreleri $\theta_{\text{hand}} \in \mathbb{R}^{90}$, görüntü özelliği \mathbf{F} 'den tahmin edilir. Daha sonra, tahmin edilen parametreler SMPL+H modeline iletilir ve 3D insan meshi elde edilir. Hesaplama yükünü azaltmak için, elde edilen 3D insan mesh bir örnekleme algoritması [42] ile düşürülür. Başlangıç 3D nesne meshini \mathbf{M}_o elde etmek için, önceki çalışmalarda olduğu gibi [38] bir 3D nesne mesh şablonu verilmişken, görüntü özelliği \mathbf{F} 'den nesne rotasyonu \mathbf{R}_o ve translasyonu \mathbf{t}_o tahmin edilir. Başlangıç yeniden yapılandırma modülünün genel tasarımı, yalnızca insan vücudu ve ellerini tahmin edecek şekilde değişiklikler yapmaktadır. Son teknoloji ise bütün vücut 3D insan mesh yeniden yapılandırma yöntemini [27] takip eder. Mimarinin detaylı açıklaması ek materyalde açıklanmaktadır.

2.4.2 3D yönlendirmeli temas tahmini

Bu aşamada, ContactFormer (3D verteks özelliklerini kullanarak insan ve nesne temas haritalarını tahmin eden Transformer), başlangıçta yeniden yapılandırılmış 3D insan ve nesne meshlerinden (\mathbf{M}_h ve \mathbf{M}_o) çıkarılan 3D verteks özellikleri (\mathbf{F}_{vh} ve \mathbf{F}_{vo}) üzerinden insan ve nesne temas haritalarını ($\mathbf{C}_h \in \mathbb{R}^{431}$ and $\mathbf{C}_o \in \mathbb{R}^{64}$) tahmin eder. 3D verteks özellik çıkarma işlemi, 3D vertekse hizalanmış özelliklerden ve her bir verteksin 3D koordinatlarından oluşur. vertekse hizalanmış özellikler, başlangıç meshlerinin (\mathbf{M}_h ve \mathbf{M}_o) 3D vertekslerinin (x, y) pozisyonlarıyla görüntü uzayına ızgara örnekleme yaparak

elde edilir. Izgara örnekleme işleminden sonra, vertekse hizalanmış özelliklere 1x1 konvolüsyon uygulanarak kanal boyutu 2048'den 256'ya indirilir. Daha sonra, başlangıç meshlerinin (\mathbf{M}_h ve \mathbf{M}_o) vertekse hizalanmış özellikleri ve her bir verteksin 3D koordinatlarını birleştirilerek 3D verteks özellikleri (\mathbf{F}_{vh} ve \mathbf{F}_{vo}) elde edilir. Dolayısıyla, insan ve nesne için 3D verteks özelliklerinin nihai boyutları $\mathbf{F}_{vh} \in \mathbb{R}^{(256+3) \times 431}$ ve $\mathbf{F}_{vo} \in \mathbb{R}^{(256+3) \times 64}$ olur. 3D verteks özellikleri, 3D mesh verteksleri etrafındaki zengin bağlamsal bilgiyi içerir ve insan-nesne temas tahmini için 3D rehberlik sağlar. 3D verteks özellikleri, temas tahmini Transformer olan ContactFormer'a iletilir.

İnsan-nesne temas tahmini: 3D verteks özellikleri (\mathbf{F}_{vh} ve \mathbf{F}_{vo}) verildiğinde, ContactFormer insan ve nesne temas haritalarını (\mathbf{C}_h ve \mathbf{C}_o) tahmin eder. ContactFormer'ın insanlar ve nesneler arasında ilgili bilgiye odaklanmasını teşvik etmek için, insan ve nesne 3D verteks özellikleri arasında cross-attention (CA) Transformer'lar ile cross-attention işlemi gerçekleştirilir [43]. Daha sonra, temas haritaları (\mathbf{C}_h ve \mathbf{C}_o) fully-connected (FC) katmanlar ile tahmin edilir ve ardından bir sigmoid aktivasyon fonksiyonu uygulanır.

2.4.3 Temas tabanlı iyileştirme

Bu aşamada, CRFormer, 3D tepe özellikleri (\mathbf{F}_{vh} ve \mathbf{F}_{vo}) ve temas haritalarından (\mathbf{C}_h ve \mathbf{C}_o) yola çıkarak iyileştirilmiş 3D insan ve nesne meshlerini (\mathbf{M}_h^* ve \mathbf{M}_o^*) sağlar.

Temas tabanlı maskeleme: Temas haritalarına (\mathbf{C}_h ve \mathbf{C}_o) dayalı olarak, temas halinde olmayan 3D tepe özelliklerinin (\mathbf{F}_{vh} and \mathbf{F}_{vo}) bir kısmını sıfır vektörlerle maskelenir. Bu şekilde yalnızca insan-nesne temasına karşılık gelen özellikler korunur. İnsan ve nesnenin maskelenmiş 3D tepe özelliklerini sırasıyla \mathbf{F}_{vh}^c ve \mathbf{F}_{vo}^c , olarak adlandırılır. Bu temas tabanlı maskeleme, özellik toplama tekniği olarak, CRFormer'ın hangi özelliklere odaklanması gerektiğini belirtmek için temas haritalarını ana sinyal olarak kullanmaya zorlayan temel strateji olarak belirlenmiştir. Ayrıca, insan-nesne temasıyla ilgisiz olan ve temas halinde olmayan bölümlerden gelen özellikleri çıkararak, doğru iyileştirme için zararlı olabilecek istenmeyen insan-nesne korelasyonlarını öğrenmeyi önlüyoruz. Temas tabanlı maskelemenin etkinliği hakkında daha fazla tartışma Bölüm 2.8'de sunulmuştur.

3D insan ve nesne iyileştirmesi: Maskelenmiş 3D tepe özellikleri (\mathbf{F}_{vh}^c ve \mathbf{F}_{vo}^c) ve orijinal 3D tepe özellikleri (\mathbf{F}_{vh} ve \mathbf{F}_{vo}), cross-attention (CA) ve self-attention (SA) Transformer kombinasyonu ile işlenerek, iyileştirilmiş 3D insan ve nesne meshleri (\mathbf{M}_h^* ve \mathbf{M}_o^*) elde edilir. \mathbf{F}_{vh}^c ve \mathbf{F}_{vo}^c , insan ve nesne arasındaki ilgili bilgileri işlemek üzere CA Transformerlar'a iletilir. \mathbf{F}_{vh}^c ve \mathbf{F}_{vo}^c yalnızca temas halindeki özellikleri içerdiğinden, CA Transformerlar'a esas olarak insan-nesne temasına ilişkin bağlamsal bilgileri işler. \mathbf{F}_{vh} ve \mathbf{F}_{vo} ise insan-nesne etkileşimini dikkate almadan, her birinin kendi 3D konumsal bilgilerini çıkarmak için ayrı ayrı SA Transformerlara iletilir. SA Transformerlara, esas olarak insan ve nesnenin temas halinde olmayan kısımlarına ilişkin bağlamsal bilgileri işler. Bu CA ve SA Transformer kombinasyonu, temas bilgilerini aşırı önyargı veya ihmalden korur. Son olarak, CA ve SA Transformerlardan gelen çıktılar ek SA Transformerlara ve ardından FC katmanlarına iletilerek iyileştirilmiş 3D insan ve nesne meshleri (\mathbf{M}_h^* ve \mathbf{M}_o^*) üretilir.

2.5 Kayıp Fonksiyonları

Projemiz, aşağıdaki gibi tanımlanan kayıp fonksiyonu L 'yi minimize ederek uçtan uca bir şekilde eğitilmiştir:

$$L = L_{\text{contact}} + L_{\text{refine}} + L_{\text{init}}, \quad (2.5.1)$$

Burada L_{contact} , tahmin edilen ve gerçek değer(GD) temas haritaları (\mathbf{C}_h ve \mathbf{C}_o) arasındaki ikili çapraz-entropy kaybıdır. L_{refine} ise şu şekilde tanımlanır:

$$L_{\text{refine}} = L_{\text{verteks}} + L_{\text{ledge}}, \quad (2.5.2)$$

L_{verteks} , iyileştirilmiş insan ve nesne meshlerinin (\mathbf{M}_h^* ve \mathbf{M}_o^*) tahmin edilen ve GD tepe noktası başına 3D koordinatları arasındaki L1 mesafesidir ve L_{ledge} , iyileştirilmiş insan meshlerinin (\mathbf{M}_h^*) tahmin edilen ve GD kenarları arasındaki kenar uzunluğu tutarlılık kaybıdır. L_{init} ise şu şekilde tanımlanır:

$$L_{\text{init}} = L_{\text{param}} + L_{\text{coord}} + L_{\text{hbox}}, \quad (2.5.3)$$

L_{param} , tahmin edilen ve gerçek değer SMPL+H parametreleri (θ_{body} ve θ_{hand}), 3D nesne rotasyonu R_o ve 3D nesne translasyonu t_o arasındaki L1 mesafesidir. L_{coord} , tahmin edilen ve GD insan eklem koordinatları arasındaki L1 mesafesi olup, 3D ve 2D eklem koordinatlarından oluşur. L_{hbox} , tahmin edilen ve gerçek değer için elin sınırlayıcı kutuları arasındaki L1 mesafesidir. L_{init} 'i, Hand4Whole [27] kayıp fonksiyonu değiştirilerek tasarlanmıştır.

2.6 Uygulama Detayları

Proje kapsamında kullanılan PyTorch [44], derin öğrenme modellerinin geliştirilmesi ve uygulanması için güçlü bir kütüphane olarak seçilmiştir. PyTorch, özellikle dinamik hesaplama grafikleri ve kolay kullanılabilirliği ile bilinmektedir. Bu projede, insan ve nesne etkileşimlerini 3D olarak sanallaştırmak için derin öğrenme tekniklerinden yararlanılmıştır. Ana yapı (backbone) olarak, Hand4Whole [45] modelinin önceden eğitilmiş ağırlıkları kullanılmıştır. Hand4Whole modeli, insan elinin detaylı ve doğru bir şekilde modellenmesi için geliştirilmiş olup, bu çalışmada daha genel insan ve nesne etkileşimlerinin modellenmesine uyarlanmıştır.

Modelin eğitiminde, önceden eğitilmiş ağırlıklar başlangıç noktası olarak alınmış ve bu ağırlıklar Adam optimizier [46] kullanılarak güncellenmiştir. Adam optimizier, adaptif öğrenme oranları ile çalışan ve genellikle daha hızlı ve etkili bir şekilde yakınsama sağlayan bir optimizasyon algoritmasıdır. Eğitim sürecinde, 16 mini-yığın (mini-batch) boyutu kullanılmıştır. Mini-yığın boyutu, her bir güncellemede kullanılan veri miktarını belirler ve bu sayede eğitim sürecinin hem hızlı hem de stabil olmasını sağlar.

Yeniden oluşturma hedef bölgesi, eğitim ve test aşamalarında önceki çalışmalarını takiben GV kutusu ile kırılmıştır [37, 38]. GV kutusu, görüntüdeki ilgilenilen bölgenin daha net ve odaklanmış bir şekilde işlenmesini sağlar. Bu, modelin sadece önemli bilgiye odaklanarak daha doğru tahminler yapmasına yardımcı olur.

Eğitim sürecinde veri artırma teknikleri kullanılmıştır. Veri artırma, modelin genelleme kabiliyetini artırmak için yapılan çeşitli işlemleri içerir. Bu işlemler arasında ölçeklendirme, döndürme ve renk değişimleri gibi dönüşümler bulunmaktadır. Ölçeklendirme, farklı boyutlardaki görüntülerin işlenmesini sağlar. Döndürme, modelin

farklı açılardaki nesneleri tanımasını ve anlamasını kolaylaştırır. Renk değişimleri ise, modelin farklı aydınlatma koşullarında ve renk varyasyonlarında doğru tahmin yapabilmesini sağlar.

Başlangıç öğrenme oranı 10^{-4} olarak ayarlanmıştır. Öğrenme oranı, modelin ağırlıklarının ne kadar hızlı güncelleneceğini belirler. Düşük bir öğrenme oranı, modelin yavaş ama daha istikrarlı bir şekilde öğrenmesini sağlar. Bu proje için seçilen öğrenme oranı, modelin doğru bir şekilde eğitilmesi için optimize edilmiştir.

Model, NVIDIA RTX 3050Ti GPU kullanılarak eğitilmiştir. GPU, büyük veri kümelerinin ve karmaşık hesaplamaların hızlı bir şekilde işlenmesini sağlar. Bu, özellikle derin öğrenme modellerinin eğitimi için kritik bir öneme sahiptir. Model, 5 epoch süreyle eğitilmiştir. Epoch, modelin tüm veri kümesini baştan sona bir kez görmesi anlamına gelir. 5 epoch boyunca model, veri kümesindeki tüm örnekleri defalarca görerek öğrenme sürecini tamamlamıştır.

Eğitim süreci boyunca, modelin performansı düzenli olarak izlenmiş ve optimize edilmiştir. Bu, modelin öğrenme sürecinin düzgün bir şekilde ilerlemesini ve istenilen sonuçlara ulaşmasını sağlamıştır. Eğitim sırasında kullanılan teknikler ve optimizasyon yöntemleri, modelin genel performansını ve doğruluğunu artırmada önemli bir rol oynamıştır.

Bu detaylar, projenin teknik altyapısını ve eğitim sürecini daha iyi anlamak için önemlidir. Modelin başarısı, kullanılan veri seti, optimizasyon teknikleri ve donanım konfigürasyonu gibi birçok faktöre bağlıdır. Bu kapsamlı eğitim süreci, insan ve nesne etkileşimlerini 3D olarak sanallaştırma hedefine ulaşmada önemli bir adım olmuştur.

2.7 Deneyler ve Veri Seti

Bu bölümde, proje kapsamında gerçekleştirilen deneyler ve kullanılan veri setleri detaylı bir şekilde ele alınacaktır. Deneyler, modelin etkinliğini ve doğruluğunu değerlendirmek amacıyla gerçekleştirilmiş olup, kullanılan veri setlerinin çeşitliliği ve kapsamı, projenin genel başarısına önemli katkılar sağlamaktadır. Öncelikle, test verilerinin işlenme süreci ve nesne tespiti için kullanılan YOLOv3 derin öğrenme modeli incelenecek, ardından deneylerde kullanılan veri setlerinin özellikleri ve yapısı ayrıntılı olarak açıklanacaktır. Bu kapsamda, veri setlerinin hazırlanması, işlenmesi ve analiz edilmesi süreçleri ile elde edilen sonuçların değerlendirilmesi hedeflenmektedir.

2.7.1 Test verilerinin işlenmesi

Bu projede öncelikle test edilen görselin, YOLOv3 (You Only Look Once) derin öğrenme modeli kullanılarak nesne tespiti gerçekleştirilmiştir. YOLOv3, gerçek zamanlı nesne tespiti için yaygın olarak kullanılan, verimli ve yüksek doğruluk oranına sahip bir derin öğrenme algoritmasıdır. Modelin temel prensibi, görüntüyü gridlere bölerek her griddede nesne olup olmadığını tahmin etmek ve belirli bir doğruluk eşik değerine sahip nesneleri sınıflandırmaktır[47,48]. YOLO, resmi tek bir seferde bir nöral ağdan geçirerek tüm nesnelerin sınıfını ve koordinatlarını tahmin eder. YOLO algoritması, girdi resmini belirli boyutlardaki bir ızgara formatına dönüştürür. Bu ızgaralar, resmin farklı bölgelerini temsil eder. Her bir ızgara, içindeki nesnenin varlığını ve varsa nesnenin konumunu ve sınıfını tahmin etmekten sorumludur. Her bir ızgara için YOLO, ayrı bir tahmin vektörü oluşturur. Bu vektörler, nesnenin varlığını gösteren bir güven skoru, nesnenin konumunu belirleyen koordinatlar ve nesnenin sınıfını belirten olasılıklar içerir. YOLO algoritması, her bir ızgara için ayrı bir tahmin vektörü oluşturduğu için birden fazla nesneyi aynı anda tespit etme yeteneğine sahiptir. Ayrıca, gereksiz kutuları elemek için Non-Max Suppression algoritmasını kullanarak tespit edilen nesneleri filtreler. Her bir tahmin vektöründe aşağıdaki unsurlar bulunur:

- **Güven skoru:** Modelin, ilgili ızgara içinde nesne olup olmadığına ne kadar emin olduğunu gösteren bir skordur. Bu skor, 0 ile 1 arasında bir değer alır. Eğer nesne olduğuna inanılıyorsa, bu skor aynı zamanda tahminin ne kadar doğru olduğunu da yansıtır.
- **Nesnenin ortalama noktasının X ve Y koordinatları (B_x ve B_y):** Nesnenin tahmini orta noktasının X ve Y koordinatlarını ifade eder.
- **Nesnenin genişliği veya yüksekliği (B_w ve B_h):** Tahmin edilen nesnenin genişliği ve yüksekliği bilgisini içerir.
- **Bağlı sınıf olasılığı:** Modelde bulunan farklı sınıflar için olasılık değerlerini içerir. Her sınıf için bir olasılık değeri bulunur.

Güven skoru, kutunun güvenilirliğini temsil eder. Bu skor, nesne tahmininin doğruluğu ve tahmin edilen kutunun gerçek nesneyi ne kadar iyi kapsadığı gibi faktörlere dayanır. Bu skorun hesaplanması için kullanılan formül şu şekildedir:

$$\text{Güven Skoru} = \text{Kutu Güven Skoru} \times \text{Bağlı Sınıf Olasılığı} \quad (2.7.1.1)$$

Kutu Güven Skoru, nesnenin kutuyu kapsayıp kapsamadığının olasılığını ($\mathbf{P}_{\text{nesne}}$) ve Intersection over Union (**IoU**) değerini içerir. İlgili ızgarada hiçbir nesne olmadığında bağlı sınıf olasılığı 0 olacağı için, güven skoru da 0 olacaktır.

YOLO birden fazla nesne aynı ızgarada bulunursa veya bir ızgarada birden fazla nesnenin orta noktası bulunursa, bu durum YOLOv3 kullanımına sunulan Anchor Box'larıyla ele alınır. Anchor Box'lar, her bir ızgara için önceden belirlenmiş bir sayıda tahmin yapmayı sağlar ve çıktı vektörünün şeklini şu şekilde şekillendirir: $\mathbf{S} \times \mathbf{S} \times (\#\mathbf{A} \times (\mathbf{5} + \#\mathbf{C}))$. Bu sayede, her ızgara için birden fazla tahmin yapılabilir ve nesne tespiti daha esnek hale gelir. YOLO'nun hata fonksiyonları üç tanedir. Sınıflandırma kaybı, konum kaybı ve güven kaybı. Bu kayıplar, nesne tespiti sırasında yapılan hataları ölçer ve genel kaybı etkileyen faktörlerdir[49].

Sınıflandırma kaybı:

Eğer geçerli ızgarada nesne varsa, her sınıf için:

$$\mathbb{1}_i^{\text{obj}} = \begin{cases} 1 & \text{if the object exist in the } i\text{-th cell} \\ 0 & \text{Otherwise} \end{cases}$$

$$\sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

Şekil 2.2 Sınıflandırma Kaybı Formülü

Güven kaybı:

Eğer nesne varsa,

$$\sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2$$

Şekil 2.3 Nesne Varsa Güven Kaybı Formülü

Eğer nesne yoksa,

$$\lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2$$

Şekil 2.4 Nesne Yoksa Güven Kaybı Formülü

Resmin ızgaraları üzerinde dolaşırken, çoğu ızgara hiçbir nesne içermeyecektir. Bu durum, arka plan olarak nitelendirdiğimiz sınıfların sayısında bir dengesizliğe neden olur.[50] Bu dengesizliği azaltmak için, bir ızgarada nesne bulunmadığı durumları ele alırken, yukarıdaki formülde belirtildiği gibi, bu kaybı belirli bir sabitle (genellikle 0.5 olarak tanımlanan "noobj" ağırlığı) çarpıyoruz.

Bu üç hata fonksiyonunun toplamı, genel hata fonksiyonunu oluşturur.

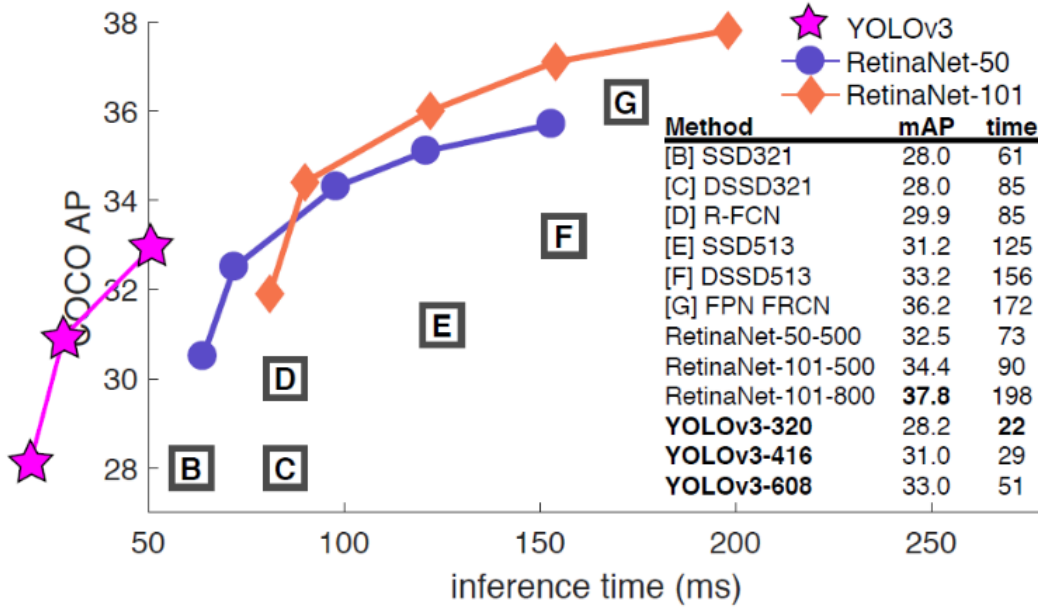
$$\begin{aligned} & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\ & + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \end{aligned}$$

Kaynak

Şekil 2.5 Üç Hata Fonksiyonun Toplamı

Her bir tahmin, belirli bir güven eşiğinin (0.80) üzerinde olduğu sürece, ilgili nesneye ait sınıf etiketi ve sınır kutusu koordinatları belirlenir. Sınır kutuları ve etiketler, OpenCV kullanılarak görüntü üzerinde görselleştirilir. Ayrıca, her bir tespit edilen nesne için, nesnenin bulunduğu bölgeyi içeren bir maske oluşturulur. Bu maskeler, tespit edilen nesnelerin daha detaylı analiz edilmesi amacıyla ayrı görüntüler olarak kaydedilir.

Bu işlem adımları sonucunda, YOLOv3 modelinin kullanımıyla görüntülerdeki nesneler başarıyla tespit edilmekte ve bu nesnelerin belirginleştirilmiş bölgeleri maskelerle işaret etmektedir.



Şekil 2.6 YOLOv3 Modelinin Başarı Grafiği

2.7.2 Bounding box değerinin hesaplanması

Bu çalışmada, YOLOv3 modeli kullanılarak nesne tespiti yapılmış ve tespit edilen nesnelerin sınırlayıcı kutuları (bounding box) hesaplanmıştır. Bounding box'lar etiketlemek istediğimiz objenin kenarlarına teğet olan basit dikdörtgenler olarak tanımlanır. Bu metodu kullanarak görüntü üzerindeki istenilen objeyi sınıflandırarak ve konumunu belirler, bu sayede objeyi tanımlanmasını sağlamış olur. [51] Bounding box hesaplama süreci şu adımlardan oluşmaktadır:: YOLOv3 modeli, önceden eğitilmiş ağırlıklar (weights) ve yapılandırma dosyası (cfg) kullanılarak yüklenmiştir. Test görüntüsü, modelin ihtiyaç duyduğu formatta ve boyutta bir blob'a dönüştürülmüş ve modelin girdisine beslenmiştir. Modelin çıktıları, tespit edilen nesnelerin bilgilerini içeren katmanlardan elde edilmiştir. Bu katmanlar, her nesne için tahmin edilen sınıf skorlarını ve bounding box koordinatlarını içermektedir.

Her bir tespit için bounding box hesaplaması şu adımlarla gerçekleştirilmiştir: Her tespit için, nesne sınıf skorları hesaplanmış ve en yüksek skoru alan sınıf belirlenmiştir. Belirlenen sınıfın güven skoru (confidence) 0.80'den büyükse, bu tespit geçerli kabul edilmiştir. Geçerli tespitler için bounding box koordinatları, modelin çıktılarından elde edilen normalize edilmiş koordinatların görüntü boyutlarına çarpılmasıyla hesaplanmıştır. Bu hesaplamada bounding box'un merkez koordinatları, genişliği ve yüksekliği kullanılarak kutunun başlangıç ve bitiş koordinatları belirlenmiştir. Hesaplanan bounding box'lar arasında en büyük alanı kaplayan kutu belirlenmiştir. Bu kutunun alanı, genişlik ve yükseklik çarpımıyla hesaplanmıştır. Tüm bounding box'lar arasında en büyük alanı kaplayan kutu, en büyük bounding box olarak belirlenmiştir. Belirlenen en büyük bounding box, görüntü üzerine çizilmiş ve ilgili nesne etiketi görüntü üzerine eklenmiştir. Bu işlem, bounding box'un sınırlarının belirginleştirilmesi ve tespit edilen nesnenin etiketlenmesi için yapılmıştır.

2.7.3 Görüntüdeki insandan ve nesneden maske oluşturulması

Test verilerinin işlenmesi için DeepLabV3 segmentasyon modeli kullanılmıştır. Modelin eğitiminde PASCAL VOC veri setindeki etiketler kullanılmış ve çeşitli nesne sınıflarını ayırt edebilmesi sağlanmıştır.

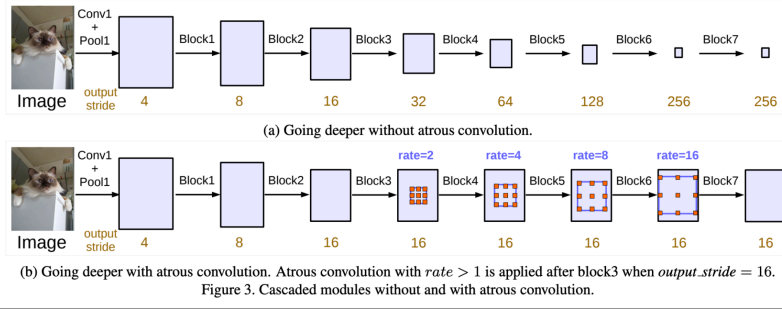
PASCAL VOC Veri Seti: PASCAL VOC veri seti, çeşitli kategorilerde nesneleri içeren zengin bir görüntü koleksiyonuna sahiptir. Bu kategoriler arasında insanlar, hayvanlar, araçlar ve mobilyalar gibi günlük yaşamda karşılaşılan birçok nesne bulunur. Veri setindeki her görüntü, nesne sınıfları ve bu nesnelerin konumları (bounding box) ile detaylı olarak etiketlenmiştir. Ayrıca, segmentasyon görevleri için her pikselin hangi nesneye ait olduğunu belirten detaylı segmentasyon haritaları da mevcuttur. PASCAL VOC veri seti, bilgisayarla görme algoritmalarının performansını değerlendirmek için kullanılan standart bir değerlendirme protokolüne sahiptir. Bu protokol, çeşitli metrikler (örneğin, Intersection over Union - IoU) kullanarak model performansını objektif bir şekilde karşılaştırmayı mümkün kılar.

Bu çalışma kapsamında, DeepLabV3 modeli, PASCAL VOC 2012 veri seti üzerinde eğitilmiş ve test verilerinin işlenmesi sırasında bu veri setindeki segmentasyon etiketlerini kullanmıştır.[52] Modelin yüksek doğrulukla nesne segmentasyonu yapabilmesi, PASCAL VOC veri setinin zenginliği ve detaylı etiketlemesi sayesinde mümkün olmuştur. Bu

bağlamda, test verilerinin işlenmesi sırasında modelin performansı ve doğruluğu değerlendirilmiş ve segmentasyon kalitesinin yüksek olduğu gözlemlenmiştir.[]

DeepLabV3 modelinin önceden eğitilmiş hali, TensorFlow formatında bir tar dosyası (tarball) içinde sağlanmıştır. Bu tar dosyası içinden modelin dondurulmuş grafik yapısı (frozen inference graph) çıkarılarak yüklenmiştir. Bu işlem, TensorFlow'un grafik yapısının doğru bir şekilde oluşturulması ve ilgili katmanların TensorFlow oturumunda (session) aktif hale getirilmesiyle yapılmıştır. Model, sabit bir girdi boyutuna (513x513 piksel) ihtiyaç duyar. Bu nedenle, test görüntüsü ilk olarak yeniden boyutlandırılır. Görüntünün orijinal boyutları korunarak en büyük kenarı 513 piksel olacak şekilde ölçeklenir. Bu işlem, modelin daha verimli çalışmasını sağlar. Segmentasyon kısmında yeniden boyutlandırılan görüntü, modelin girdi tensörüne beslenir ve modelden çıkan segmentasyon haritası elde edilir. Segmentasyon haritası, her pikselin hangi nesne sınıfına ait olduğunu belirten bir etiket haritasıdır. DeepLabV3 modeli, PASCAL VOC veri setindeki 21 sınıfı (arka plan dahil) ayırt edebilme yeteneğine sahiptir. Sonuçların Görselleştirilmesi: Elde edilen segmentasyon haritası, PASCAL VOC renk haritası kullanılarak renklendirilir. Bu renklendirilmiş segmentasyon haritası, orijinal görüntü üzerine bindirilerek (overlay) görselleştirilir. Bu aşamada, segmentasyon sonuçları daha anlaşılır ve yorumlanabilir hale gelir.

DeepLabV3 modeli, nesne segmentasyonu için derin öğrenme tabanlı ileri bir mimari kullanır. Atrous Convolution (Dilation Convolution): DeepLabV3'ün temel yapı taşlarından biri, genişletilmiş (atrous) konvolüsyonlardır. Bu teknik, standart konvolüsyonların genişletilerek yapılmasını sağlar. Böylece daha geniş bir alandan bilgi toplanabilir ve daha yüksek çözünürlüklü özellik haritaları elde edilebilir. Bu teknik, hesaplama verimliliğini artırırken, nesne sınırlarının daha doğru bir şekilde belirlenmesine yardımcı olur. Atrous Spatial Pyramid Pooling (ASPP): ASPP modülü, görüntüdeki nesnelerin farklı ölçeklerde algılanabilmesi için bir dizi genişletilmiş konvolüsyon katmanından oluşur [53].



Şekil 2.7 DeepLabV3 Modelinin Mimarisi

Bu modül, farklı genişletme oranlarına sahip konvolüsyonlar kullanarak görüntünün farklı ölçeklerdeki özelliklerini yakalar. ASPP, hem yerel hem de global bağlamı bir araya getirerek daha doğru segmentasyon haritaları oluşturur.[52] Encoder-Decoder Yapısı: DeepLabV3, encoder-decoder mimarisi kullanarak görüntüdeki ince detayları ve kenarları daha iyi segmentasyon yapabilmek için kullanır. Encoder kısmı, giriş görüntüsünün özelliklerini çıkarırken, decoder kısmı bu özellikleri kullanarak segmentasyon haritasını oluşturur. Conditional Random Fields (CRFs): DeepLabV3 modeli, segmentasyon sonuçlarını iyileştirmek için koşullu rastgele alanlar (CRFs) kullanır.[54] CRFs, segmentasyon haritasındaki pikseller arasındaki mekansal ilişkileri modelleyerek, nesne sınırlarının daha keskin ve doğru olmasını sağlar.



Şekil 2.8 Maskı Alınmış Fotoğraf

Bu süreçler ve teknikler bir araya gelerek, DeepLabV3 modelinin yüksek doğrulukla nesne segmentasyonu yapabilmesini sağlar. Modelin bu yetenekleri, test verilerinin işlenmesi ve analiz edilmesi sürecinde önemli bir rol oynamıştır. Bu bağlamda, test verilerinin işlenmesi sırasında modelin performansı ve doğruluğu değerlendirilmiş ve segmentasyon kalitesinin yüksek olduğu gözlemlenmiştir.

2.7.4 Veri Seti - BEHAVE



Şekil 2.9 Behave Veri Seti İçeriği

Doğal ortamlarda insan-nesne etkileşimlerini içeren en büyük veri seti olan BEHAVE veri setini sunuyoruz. BEHAVE, çoklu görünüm RGB-D dizilerini ve buna karşılık gelen 3D nesne ve SMPL uyumlarını 3D temaslarla birlikte içerir.

Bu çalışma, RGB-D kameraların görüntülerinden elde edilen veriler kullanılarak insan ve nesne hareketlerinin istatistiksel 3D modellerinin oluşturulması üzerine odaklanmaktadır. Veri seti olarak BEHAVE kullanılacak olup, bu veri seti oldukça büyük ve kapsamlıdır. Veri seti farklı ortamlarda gerçekleşen çeşitli insan ve nesne etkileşimlerini içermektedir.

BEHAVE, RGB-D kameraların kullanıldığı geniş kapsamlı bir video veri setidir. Bu veri seti, farklı ortamlarda kaydedilmiş insan ve nesne etkileşimlerini içermekte olup, ev içi, ofis ortamı, kamusal alanlar gibi çeşitli senaryolarda gerçekleşen günlük aktiviteleri kapsamaktadır. BEHAVE veri seti, araştırmacıların insan davranışlarını, nesne manipülasyonlarını ve çevresel etkileşimleri anlamak için bir platform sunmaktadır.

BEHAVE veri seti, her biri RGB ve derinlik (D) bilgisi içeren video görüntülerinden oluşmaktadır. Bu derinlik bilgisi, görüntülerin yanı sıra 3D mekânsal bilgi sağlar. Veri setindeki videolar, insanların doğal etkileşimlerini ve günlük aktivitelerini yansıtan çeşitli

senaryoları içermektedir. Örneğin, insanların nesnelerle etkileşimleri, nesneler arasında geçişler ve farklı insan etkileşim tipleri gibi çeşitli durumlar mevcuttur.

BEHAVE veri setinin büyük ölçekli olması, geniş bir veri çeşitliliği sunmasını sağlar. Veri seti, farklı ışık koşulları altında kaydedilmiş videolar içerir ve bu durum, algoritmaların çeşitli aydınlatma koşullarına nasıl tepki verebileceğini incelemek için önemli bir fırsat sunar. Ayrıca, çeşitli ortamlarda gerçekleşen etkileşimler, makine öğrenimi modellerinin genelleme yeteneğini değerlendirmek için ideal bir zemin oluşturmaktadır.

BEHAVE veri seti, araştırmacıların insan hareketlerini, nesne manipülasyonlarını ve çevresel etkileşimleri derinlemesine incelemelerine imkân tanır. Özellikle, insanların nesnelerle etkileşim biçimleri ve bu etkileşimlerin zamansal ve mekânsal özellikleri üzerinde yapılan analizler, robotik sistemlerden yapay zeka uygulamalarına kadar birçok alanda yenilikçi çalışmalara zemin hazırlamaktadır.

Bu veri seti üzerinde yapılan çalışmalar, makine öğrenimi ve yapay zeka alanında ileri düzey analizler yapmak ve yeni algoritmalar geliştirmek için birçok olanak sunar. Örneğin, derin öğrenme teknikleri kullanılarak etkileşimlerin tanımlanması ve sınıflandırılması, gelecekteki otomatik sistemlerin geliştirilmesine yol açmaktadır.

Tablo 2.1 Yöntemlerin Karşılaştırılması

Veri Seti	Yöntemler	$\text{Contact}_p^{\text{est}} \uparrow$	$\text{Contact}_r^{\text{est}} \uparrow$
Behave	POSA [12]	0.514	0.299
Behave	BSTRO [14]	0.615	0,527
Behave	DECO [43]	0.638	0.337
Behave	Proje	0.754	0.587

BEHAVE veri seti üzerinde, insan-obje temas tahmini konusunda en son yöntemlerle yapılan nicel karşılaştırmayı sunmaktadır. Burada, " $\text{Contact}_p^{\text{est}}$ " ve " $\text{Contact}_r^{\text{est}}$ " metrikleri, her bir yöntemin doğruluğunu (precision) ve duyarlılığını (recall) temsil etmektedir. Görüldüğü üzere, bu proje yöntemi, veri setinde en yüksek doğruluk ve duyarlılık değerlerine ulaşarak, diğer yöntemlerden daha üstün performans sergilemiştir.

Sonuç olarak, BEHAVE veri seti, RGB-D kameralarla elde edilen zengin video verileri üzerinde gerçekleştirilen çalışmaların, insan ve nesne etkileşimlerinin anlaşılmasına ve modellenmesine katkı sağlayacak önemli bilgiler sunmaktadır. Bu veri seti, araştırmacıların çok çeşitli senaryolar altında gerçekleşen günlük insan etkileşimlerini derinlemesine analiz etmelerine imkân tanıyarak, gelecekteki akıllı sistemlerin geliştirilmesine katkıda bulunmaktadır.

2.8. Ablation Çalışmaları

Tüm yöntemleri BEHAVE [2] veri seti üzerinde eğitip değerlendirerek ablation çalışmasını gerçekleştiriyoruz

Tablo 2.2 3D yönlendirmeli temas tahmini için ablation çalışmaları

Yöntemler	$\text{Contact}_p^{\text{est}} \uparrow$	$\text{Contact}_r^{\text{est}} \uparrow$
ContactFormer Giriş Varyasyonları		
GAP özelliği	0.645	0.481
Vertex başına görüntü özelliği [14]	0.716	0.539
Parça-sahne görüntü özelliği [43]	0.719	0.556
3D vertex özelliği (Bizim)	0.754	0.587
ContactFormer Tasarım Varyasyonları		
FC katmanları	0.639	0.471
SA Transformers	0.725	0.575
ContactFormer (Bizim)	0.754	0.587

3D yönlendirmeli temas tahmininin etkililiği: Tablo 2.2'de, 3D yönlendirmeli temas tahmininin etkililiğini, ContactFormer girişlerinin ve ContactFormer tasarımının çeşitli varyasyonlarını inceleyerek gösteriyoruz. Tablo 2.2'in ilk bloğu, 3D vertex özelliklerini ContactFormer girişi olarak kullanmanın diğer giriş varyantlarını büyük ölçüde geride bıraktığını göstermektedir. İlk varyant, global ortalama alınmış (GAP) görüntü özelliğini kullanır. İkinci ve üçüncü varyantlar, mevcut yöntemleri [32, 35] takip ederek özellik

çıkartıcılarını frameworkümüze entegre eder. Özellikle, ikinci varyant, görüntü özelliğinden vertex başına özellikler çıkarmak için bir evrişim katmanı kullanır. Üçüncü varyant ise insan parçası ve sahne için iki encoder tasarlayıp, bu encoder'ların çıktıları arasında çapraz dikkat operasyonu uygulayarak özellikler elde eder. İkinci ve üçüncü varyantların bizim yöntemimizden büyük bir farkı, 3D insan ve nesne pozisyonlarına dayalı lokalize özellikler çıkarmamalarıdır. Bizim 3D vertex özelliğimiz, görüntü özelliği üzerinde grid örnekleme yaparak insan ve nesne bölgeleri etrafında lokalize bağlamsal bilgiler içerir. Ayrıca, 3D vertex özelliklerinin 3D pozisyon bilgisi, insan-nesne teması üzerinde 3D geometrik akıl yürütmeyi mümkün kılar. Bu avantajlardan dolayı, 3D vertex özelliklerini kullanmak, temas tahmininde diğer varyantları geride bırakır. Tablo 2.2'in ikinci bloğu, CA Transformers ile ContactFormer'ımızın en iyi performansı elde ettiğini gösterir. Diğer tasarımlarla karşılaştırıldığında, CA Transformers'ın çapraz dikkat operasyonu, ContactFormer'ın görüntü içindeki anlamlı bağlamsal bilgileri yakalamasını teşvik eder.

2.9. Değerlendirme Ölçütleri

İletişim tahmini için hassasiyet ve geri çağırma ($\text{Contact}_p^{\text{est}}$, $\text{Contact}_r^{\text{est}}$): İnsan-nesne iletişimi tahminini standart tespit metrikleri olan hassasiyet ($\text{Contact}_p^{\text{est}}$) ve geri çağırma ($\text{Contact}_r^{\text{est}}$) ile değerlendirilmektedir. Bu değerlendirme Huang ve ark. [32] çalışmasını takip etmektedir. Projemiz, hem insan hem de nesne iletişim haritalarını tmin ederken, önceki iletişim tahmini yöntemleri [32, 35] sadece insan iletişim haritalarını tahmin eder.

Chamfer mesafesi (CD_{human} , $\text{CD}_{\text{object}}$): Tahmin edilen ve GD (Gerçek Değer) meshler arasındaki 3D insan ve nesne rekonstrüksiyonunu, önceki 3D insan ve nesne rekonstrüksiyon çalışmaları takip edilerek Chamfer mesafesi ile değerlendirilmektedir [38, 39]. Özellikle, tahmin edilen 3D insan ve nesne meshlerine Prokrustes hizalaması uygulanır ve GD 3D insan ve nesne meshleri ile birleştirilmiş 3D insan ve nesne meshleri üzerinden Chamfer mesafesini ölçülür. Hizalanmış 3D insan ve nesne meshlerinden GD'ye ayrı ayrı olarak 3D insan (CD_{human}) ve 3D nesne ($\text{CD}_{\text{object}}$) üzerinde Chamfer mesafesi santimetre cinsinden ölçülmektedir.

Rekonstrüksiyondan iletişim için hassasiyet ve geri çağırma ($\text{Contact}_p^{\text{rec}}$, $\text{Contact}_r^{\text{rec}}$): 3D insan ve nesne rekonstrüksiyonunu değerlendirmek için özellikle iletişim açısından standart tespit metrikleri daha da benimsenmektedir. İnsan meshinin

nesne meshi ile 5 cm içindeki vertexleri sınıflandırılarak iletişim haritası elde edilmektedir. Daha sonra, insan iletişim haritası ile GD karşılaştırması arasındaki hassasiyet ($Contact_p^{rec}$) ve geri çağırma ($Contact_r^{rec}$) ölçülmektedir.

Bu metriklerin tümü, projenin insan ve nesne arasındaki etkileşimleri ve 3D rekonstrüksiyonun doğruluğunu nasıl değerlendirdiğini göstermek için kritik önem taşır. Projede kullanılan metodoloji, önceki çalışmalarla uyumlu olarak insan ve nesne iletişim tahminini geliştirmek ve daha doğru 3D rekonstrüksiyonlar elde etmek üzerine odaklanmaktadır. Bu çalışma, insan-nesne etkileşimi analizinin ve 3D model rekonstrüksiyonunun ileri düzeydeki uygulamaları için önemli bir katkı sağlamaktadır.

2.10. Elde Edilen Sonuçların Blender 4.1’de İncelenmesi

Bu çalışmada elde edilen 3D insan ve nesne etkileşimlerinin görselleştirilmesi için Blender 4.1[56] yazılımı kullanılmıştır. Blender, güçlü bir açık kaynaklı 3D modelleme ve render yazılımıdır. Hem akademik hem de endüstriyel projelerde yaygın olarak kullanılmaktadır. Blender 4.1, özellikle büyük ölçekli verilerin işlenmesi ve yüksek kaliteli görsellerin oluşturulması konularında önemli gelişmeler sunmaktadır.

2.10.1 Blender 4.1 Kullanımı

Blender 4.1'e veri aktarımı süreci, Python betikleri ve Blender’ın API’si (Blender Application Programming Interface) kullanılarak otomatikleştirilmiştir. Bu süreç, RGB-D kameralarından elde edilen görüntülerin işlenmesi, segmentasyon ve özellik çıkarımı adımlarının ardından başlatılmıştır. İşlenen veriler, önce 3D model formatına dönüştürülmüş ve ardından Blender'a aktarılmıştır. Bu aşamada, özellikle insan ve top gibi nesnelerin doğru konumlandırılması ve etkileşimlerin doğru bir şekilde görselleştirilmesi için özen gösterilmiştir.

Blender 4.1’de kullanılan temel adımlar şu şekildedir:

1. **Veri Dönüştürme ve Aktarım:** OpenCV ve NumPy kullanılarak işlenen görüntüler, 3D model formatlarına dönüştürülmüştür. Bu modeller, Blender'ın desteklediği .obj formatında kaydedilmiştir.
2. **Blender Python API Kullanımı:** Blender’ın Python API’si, verilerin Blender ortamına otomatik olarak yüklenmesi ve sahne içinde doğru bir şekilde yerleştirilmesi için kullanılmıştır. Bu betikler, modellerin Blender sahnesine

eklenmesi, konumlandırılması ve gerekli materyal ayarlarının yapılmasını sağlamıştır.

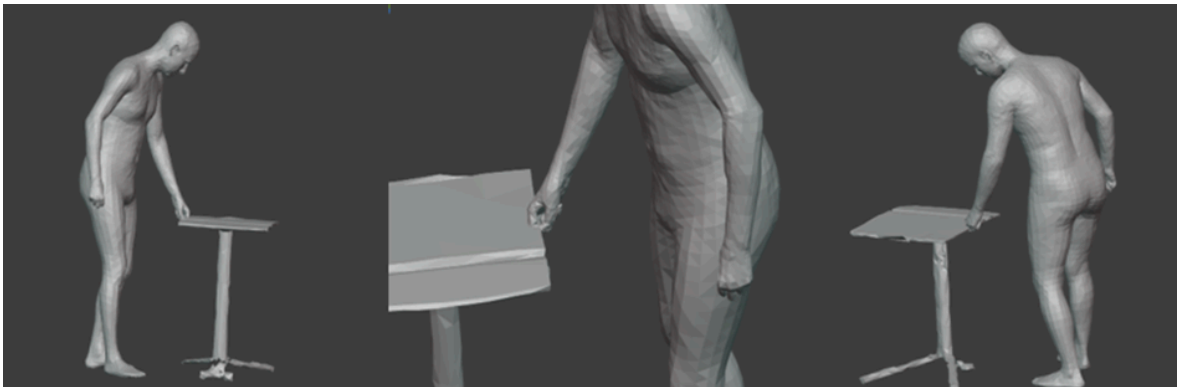
3. **Sahne Ayarları ve Render:** Blender içinde sahne ayarları yapılmış ve etkileşimlerin doğru bir şekilde görselleştirilmesi için ışıklandırma, kamera açıları ve diğer render ayarları optimize edilmiştir. Elde edilen nihai görseller, hem tez çalışmasında sunulmuş hem de analiz ve değerlendirme süreçlerinde kullanılmıştır.

Blender 4.1'de gerçekleştirilen bu görselleştirme çalışmaları, 3D insan ve nesne etkileşimlerinin daha anlaşılır ve etkileyici bir şekilde sunulmasını sağlamış, ayrıca yapılan çalışmaların doğruluğunun ve etkinliğinin görsel olarak doğrulanmasına olanak tanımıştır.

2.10.2. Blender 4.1'in Özellikleri

Blender 4.1, modern 3D modelleme ve görselleştirme projeleri için geniş bir yelpazede özellikler sunmaktadır. Bu özellikler arasında:

- Gelişmiş Render Motoru: Cycles ve Eevee gibi gelişmiş render motorları, farklı render ihtiyaçları için esnek çözümler sunar.
- Python Desteği ve API: Blender'ın güçlü Python API'si, otomasyon ve özelleştirme için geniş imkanlar sağlar.
- Modelleme ve Animasyon Araçları: Polygonal modelleme, UV haritalama, rigging ve animasyon araçları ile kompleks sahneler oluşturulabilir.
- Gelişmiş Işıklandırma ve Görselleştirme: HDR aydınlatma, gerçekçi gölgelendirme ve materyal oluşturma özellikleri ile profesyonel kalitede görseller üretilebilir.



Şekil 2.10 Elde Edilen Sonucun Blender Görüntüsü

Bu çalışmanın sonuçları, Blender 4.1'in sunduğu özellikler kullanılarak detaylı bir şekilde görselleştirilmiştir. İnsan ve nesne etkileşimlerinin 3D modelleme ve render işlemleri,

Blender'ın sağladığı render motorları ve Python API'si sayesinde başarıyla tamamlanmıştır. Elde edilen görseller, çalışmanın yürütülme sürecinin ve sonuçlarının anlaşılmasını kolaylaştırmıştır.

3. BULGULAR VE TARTIŞMA

Projenin temel amacı, insan ve nesne etkileşimlerinin 3D olarak modellenmesi ve bu modelin farklı sektörlerdeki uygulama potansiyelini ortaya koymaktır. Elde edilen bulgular, bu hedefe ulaşmak için önemli bir adım olarak değerlendirilebilir. İnsan ve nesneler arasındaki etkileşimlerin doğru ve detaylı bir şekilde modellenmesi, birçok sektörde çeşitli uygulamaların geliştirilmesine olanak sağlayabilir.

Bu çalışma, insan ve nesne etkileşimlerinin 3D modellemesi alanında önemli bir boşluğu doldurarak, mevcut literatürdeki eksiklikleri ele almıştır. Geliştirilen model, BEHAVE veri seti üzerinde başarılı bir şekilde test edilmiş ve gerçek dünya senaryolarına uygun sonuçlar elde edilmiştir. Bu da, modelin gerçek hayatta kullanılabilirliğini ve uygulanabilirliğini artırmaktadır.

Bulgular, insan ve nesne etkileşimlerinin 3D modellemesinin, oyun, eğlence, tıp, güvenlik, acil durum yönetimi ve daha birçok alanda potansiyel uygulamalara sahip olduğunu göstermektedir. Özellikle, gerçek hayattan alınan senaryoların dijital ortamda analiz edilmesi, risk analizi, acil müdahale planlaması ve simülasyon eğitimleri gibi alanlarda büyük önem taşımaktadır.

Tartışma kısmında, elde edilen bulguların sektörlerde nasıl kullanılabileceği ve gelecekteki araştırmalara nasıl katkı sağlayabileceği daha detaylı bir şekilde ele alınabilir. Örneğin, farklı sektörlerdeki uygulama potansiyelleri daha ayrıntılı olarak incelenebilir ve gelecekte yapılacak çalışmalarda hangi alanlara odaklanılması gerektiği belirlenebilir. Bu da, insan ve nesne etkileşimlerinin 3D modellemesi alanında ilerlemek ve yeni keşifler yapmak için önemli bir yol haritası oluşturabilir.

3.1. Elde Edilen Sonuçlar

Proje kapsamında yapılan çalışmalar sonucunda, insan ve nesne etkileşimlerinin 3D olarak doğru bir şekilde modellenbildiği gözlemlenmiştir. Bu modeller, özellikle oyun, sanal ve karma gerçeklik, insan analitiği ve insan-robot işbirliği gibi birçok uygulamanın

merkezinde yer alabilecek nitelikte olup, insan davranışlarının detaylı bir şekilde incelenmesine olanak tanımaktadır.

BEHAVE veri seti kullanılarak elde edilen yaklaşık 15 bin karelik veriler, yüksek kaliteli RGB-D sensörler sayesinde elde edilmiştir. Bu sayede, dijitalleştirilmiş ortamın gerçekçilik oranı yüksek tutulmuş ve elde edilen 3D modellerin gerçek dünya senaryolarına uygunluğu sağlanmıştır. Bu modeller, gerçek hayattan alınan senaryoların dijital bir ortamda analiz edilmesine olanak tanıyarak, risk analizi, acil müdahale planlaması ve simülasyon eğitimleri gibi alanlarda önemli bir rol oynayabilir.

Ayrıca, oluşturulan 3D modellerin kişiselleştirilmiş avatarların yaratılmasında önemli bir rol oynayabileceği belirlenmiştir. Bu avatarlar, oyunlar, sinema projeleri, sanal toplantılar ve eğitim simülasyonları gibi birçok farklı sektörde kullanılabilir. Örneğin, sağlık sektöründe, kişiselleştirilmiş avatarlar üzerinde yapılan simülasyonlar, cerrahi operasyonların planlanması ve cerrahi eğitimler için kullanılabilir.

Bu çalışma, insan ve nesne etkileşimlerinin 3D modellemesinin geniş uygulama potansiyeline işaret etmektedir. Gelecekte yapılacak daha ileri araştırmalar, bu potansiyeli daha da genişletebilir ve 3D modellemeye dayalı çeşitli uygulamaların geliştirilmesine olanak sağlayabilir.

3.2. Literatür Taraması ve Karşılaştırma

Literatür taramasında, 3D modelleme teknolojisinin tıp, eğlence, robotik ve acil durum yönetimi gibi birçok alanda kullanımının giderek arttığı görülmüştür. Örneğin, tıp alanında RNA ikincil ve üçüncül yapılarını tahmin edebilmek için 3D modellemeye ihtiyaç duyulmuştur. Dynamite Pete adlı oyunda ise, Autodesk Maya ve Unity 3D kullanılarak yüksek kalite ve gerçeklikte bir oyun ortaya çıkarılmıştır. Ayrıca, robotik alanında, kolların, ellerin ve baş yöneliminin doğru çalışması için ilgili vücut parçalarının 3D görüntüsü elde edilerek robota entegre edilmiştir. Acil durum yönetiminde ise, kritik altyapıların ve bunların karşılıklı bağımlılıklarının belirlenmesi için 3D modellemeye ihtiyaç duyulmaktadır.

Yapılan araştırmalar, ülkemizde gerçekleştirilen çalışmaların yetersiz olduğunu göstermektedir. Ülkemizdeki çalışmalar genellikle nesnelerin veya insanların dijitalleştirilmesi üzerine odaklanmış, ancak nesne ve insan etkileşimlerinin

dijitalleştirilmesi yeterince ele alınmamıştır. Buna karşın, yurt dışında yapılan çalışmalar, insan-nesne etkileşimlerinin daha kapsamlı bir şekilde modellenmesine yönelik ilerlemeler kaydetmiştir. Özellikle, bilgisayar grafikleri, bilgisayar vizyonu ve robotik alanlarında yapılan çalışmalar, el-nesne etkileşimlerini sentezlemeye odaklanmış olup, tam vücut etkileşimlerini oluşturma konusuna da artan bir ilgi görülmektedir.

3.3. Yurt İçi ve Yurt Dışı Çalışmaların Karşılaştırılması

Ülkemizde yapılan çalışmalarda genellikle nesnelerin görüntüleri RGB-D (Red, Green, Blue- Depth) kullanılarak 3D olarak modellenmiştir. Ancak bu çalışmaların çoğu, insan ve nesne etkileşimlerinin detaylı bir şekilde modellenmesini içermemektedir. Örneğin, birçok projede sadece statik nesnelerin 3D modelleri oluşturulmuş, dinamik etkileşimler ve insan davranışları göz ardı edilmiştir. Bu durum, ülkemizde yapılan çalışmaların kapsamının ve derinliğinin yetersiz olduğunu göstermektedir.

Buna karşın, yurt dışındaki çalışmalar, insan-nesne etkileşimlerini daha kapsamlı ve detaylı bir şekilde ele almıştır. Özellikle, tam vücut insan-nesne etkileşimi sentezi üzerine kinematik temelli yaklaşımlar ve fizik simülasyon ortamlarının uygulanmasında önemli ilerlemeler kaydedilmiştir. Bu çalışmalar, nesneye yaklaşma veya nesneleri manipüle etme gibi eylem ve nesne çeşitliliği konusunda daha geniş bir yelpazeye sahiptir. Örneğin, bilgisayar grafikleri alanında yapılan çalışmalar, insan-nesne etkileşimlerini daha gerçekçi bir şekilde modelleyebilmek için gelişmiş algoritmalar ve simülasyon teknikleri kullanmaktadır.

3.4. Çalışmanın Önemi ve Gelecekteki Uygulamaları

Elde edilen bulguların önemi, 3D modelleme teknolojilerinin günümüzde ve gelecekteki potansiyel uygulamalarını aydınlatmaktadır. İnsan ve nesne etkileşimlerinin doğru ve detaylı bir şekilde modellenmesi, birçok sektörde önemli fırsatlar ve yenilikler sunmaktadır. Bu modeller, sadece genel bir insan yapısını değil, insanların gerçek hayatta sergiledikleri davranışları ve etkileşimleri de içermektedir. Bu sayede, daha kişiselleştirilmiş avatarlar oluşturulabilir ve bu avatarlar çeşitli alanlarda kullanılabilir.

Örneğin, oyun sektöründe, gerçekçi ve detaylı insan-nesne etkileşimlerine dayalı olarak geliştirilen oyunlar, kullanıcı deneyimini önemli ölçüde artırabilir. Oyuncuların karakterleri ile etkileşime girebilmeleri ve çevrelerini etkileyebilmeleri, oyunların daha ilgi çekici ve

etkileyici hale gelmesini sağlayabilir. Benzer şekilde, sinema ve sanat alanlarında da, gerçek hayattan alınan senaryoların dijital ortamda analiz edilmesi ve canlandırılması, yaratıcı süreçlere yeni boyutlar kazandırabilir.

Ayrıca, insan-robot işbirliği ve acil durum yönetimi gibi kritik uygulamalarda da, doğru ve detaylı 3D modellerin kullanılması büyük önem taşımaktadır. Özellikle, robotlarla insanların etkileşim halinde olduğu ortamlarda, insan davranışlarını ve tepkilerini doğru bir şekilde modellemek, güvenli ve etkili bir işbirliğinin sağlanmasına yardımcı olabilir. Benzer şekilde, acil durum yönetimi ve afet müdahale ekiplerinin çalışmalarında, gerçekçi ve detaylı simülasyonlar kullanılarak, ekiplerin hazırlıklı olmaları ve etkin müdahalelerde bulunmaları sağlanabilir.

Sonuç olarak, bu çalışma, 3D modelleme ve insan-nesne etkileşimi alanında önemli bir adım olmuştur. Elde edilen bulgular, bu alandaki mevcut bilgi birikimine katkı sağlamakta ve gelecekteki araştırmalara yol göstermektedir. Bu teknolojilerin gelecekteki uygulamaları üzerinde daha fazla çalışma yapılması, insanlığın birçok alanda karşılaştığı sorunlara yenilikçi çözümler getirebilir ve yaşam kalitesini artırabilir.

3.5. Projenin Katkıları ve Geliştirilebilir Alanlar

Elde edilen bulgular ve yapılan çalışma, insan-nesne etkileşimlerinin daha kapsamlı bir şekilde modellenmesine önemli katkılar sağlamıştır. Bu projenin başlıca katkılarından biri, BEHAVE veri seti kullanılarak gerçekçi ve detaylı 3D modellerin oluşturulmasıdır. Bu veri seti, farklı açılardan elde edilen görüntülerle zenginleştirilmiş ve etkileşimlerin daha geniş bir perspektiften incelenmesine olanak tanımıştır. Bu da, modellerin gerçek dünyadaki insan-nesne etkileşimlerine daha yakın olmasını sağlamıştır.

Ancak, projenin geliştirilebilir alanları da bulunmaktadır. Öncelikle, farklı veri setlerinin kullanılması ve çeşitliliğin artırılması, modelin genelleme yeteneğini ve uygulanabilirliğini artırabilir. Farklı senaryoların ve etkileşimlerin daha kapsamlı bir şekilde incelenmesi, modelin daha geniş bir yelpazede kullanılabilir hale gelmesine olanak sağlayabilir.

Ayrıca, modelin doğruluğunu ve performansını artırmak için daha gelişmiş algoritmalar ve makine öğrenme tekniklerinin kullanılması da önemlidir. Yapay zeka ve derin öğrenme tekniklerinin geliştirilmesi, modelin daha karmaşık ve dinamik etkileşimleri doğru bir

şekilde modellenmesine olanak sağlayabilir. Bu da, modelin gerçek hayattaki uygulamalarda daha etkin bir şekilde kullanılmasını sağlayabilir.

Gelecekte yapılacak çalışmalar, insan-nesne etkileşimlerinin daha dinamik ve karmaşık senaryolarını içerebilir. Özellikle, gerçek hayattaki çeşitli ortamlarda insanların ve nesnelerin nasıl etkileşime girdiğinin daha detaylı bir şekilde incelenmesi, modelin gerçek dünya koşullarında daha başarılı olmasını sağlayabilir. Bu sayede, insan-nesne etkileşimlerinin daha etkin bir şekilde modellenmesi ve gerçek hayattaki uygulamalarda kullanılması mümkün olabilir.

4. SONUÇLAR

Gelişen görüntü teknolojileri ile birlikte 3D modelleme, birçok farklı sektörde önemli bir araç haline gelmiştir. Bu çalışmada, insan ve nesne etkileşimlerini dijital ortamda modellemek ve gerçekçi avatarlar oluşturmak amacıyla ileri düzey 3D modelleme yöntemleri incelenmiş ve uygulanmıştır. Ayrıca, BEHAVE veri seti gibi özel veri setleri kullanılarak, insan ve nesne etkileşimlerinin tam vücut olarak modellenmesi hedeflenmiştir.

Bu çalışmanın sonuçları, 3D modellemenin farklı uygulama alanlarında nasıl kullanılabileceğini ve bu teknolojilerin önemini göstermektedir. Özellikle, tıbbi alanlarda RNA ikincil ve üçüncül yapılarının tahmininde, eğlence sektöründe oyun geliştirme süreçlerinde, robotik sistemlerde vücut parçalarının entegrasyonunda ve risk analizlerinde kritik altyapıların modellenmesinde 3D modellemenin kullanımı vurgulanmıştır. Bu teknolojiler, doğal afet gibi acil durumlarda da müdahale eylemlerinin planlanmasında önemli bir rol oynamaktadır.

Özellikle, bu çalışma insan ve nesne etkileşimlerini gerçeğe en yakın biçimde modellemeyi amaçlamaktadır. İnsan-nesne etkileşimlerinin, özellikle mesafelere dikkat edilerek, dijital ortamda doğru bir şekilde algılanması ve modellenmesi hedeflenmiştir. Autodesk Maya ve Unity 3D gibi ileri düzey yazılımlar kullanılarak yüksek kaliteli ve gerçekçi oyunlar geliştirilmiş, aynı zamanda robotlar için vücut parçalarının 3D görüntüleri elde edilmiş ve entegre edilmiştir.

Çalışmanın bulguları, mevcut literatürdeki eksiklikleri doldurmayı ve 3D modellemenin farklı alanlarda nasıl uygulanabileceğini göstermeyi amaçlamaktadır. Ülkemizde yapılan

alışmaların yetersiz olduėu gözlemlenmiř, bu nedenle bu alıřma, ulusal ve uluslararası düzeyde insan-nesne etkileřimlerinin modellenmesine katkıda bulunmayı hedeflemektedir.

Sonuç olarak, bu proje, 3D modellemenin insan-nesne etkileřimlerinin yanı sıra oyun geliştirme, sanal ve karma gerçeklik, insan analitiėi ve insan-robot iřbirliėi gibi uygulamalarda nasıl kullanılabileceėini göstermiřtir. İnsan davranıřlarını inceleyerek kiřiselleřtirilmiř avatarların oluřturulmasına olanak saėlamıř ve gerçek hayattan alınan senaryoların dijital ortamda analiz edilmesini kolaylařtırmıřtır. Bu alıřma, gelecekteki benzer projeler için bir kılavuz oluřturmayı ve 3D modelleme teknolojilerinin daha da ileriye tařınmasını saėlayacak bir temel oluřturmayı amalamaktadır.

Geliřen teknoloji ile birlikte, 3D modellemenin eřitli sektörlerdeki kullanımı sürekli olarak artmaktadır. Özellikle tıp alanında, RNA ikincil ve üçüncül yapılarının tahmin edilmesi gibi karmařık problemlerin özümünde 3D modelleme büyük önem tařımaktadır. Bu alıřma, tıbbi görüntüleme tekniklerinin yanı sıra biyoinformatik ve biyomoleküler yapıları anlama açısından da önemli katkılar saėlamaktadır.

Eėlence sektöründe ise, oyun geliştirme süreçleri daha da gerçekçi hale getirmekte ve kullanıcı deneyimini artırmaktadır. Autodesk Maya ve Unity 3D gibi yazılımlar, yüksek kaliteli grafikler ve interaktif ortamlar oluřturarak oyun geliřtiricilerine geniř olanaklar sunmaktadır. Bu teknolojiler, Dynamite Pete gibi oyunların bařarılı bir řekilde geliřtirilmesine imkan tanımakta ve eėlence sektöründe ıėır açmaktadır.

Robotik sistemlerde de, 3D modelleme teknikleri, robotların kolları, elleri ve bař yönelimleri gibi vücut paralarının doėru bir řekilde entegre edilmesine yardımcı olmaktadır. Bu sayede, robotlar daha doėru ve etkili bir řekilde alıřabilmekte, endüstriyel ve hizmet robotlarına yeni iřlevler eklenmektedir.

Risk analizi ve acil durum müdahale planlaması gibi alanlarda ise, kritik altyapıların 3D modellenmesi, doėal afetler gibi acil durum senaryolarında stratejik müdahale planlarının oluřturulmasına yardımcı olmaktadır. Bu teknikler, yangın söndürme, arama kurtarma operasyonları ve afet yönetimi gibi alanlarda hayati öneme sahip olabilmektedir.

Öte yandan, insan ve nesne etkileřimlerinin tam vücut olarak modellenmesi, BEHAVE veri seti gibi özel veri setleri kullanılarak gerekleřtirilmiřtir. Bu alıřma, insan ve nesne etkileřimlerinin, özellikle mesafelere dikkat edilerek, dijital ortamda doėru bir řekilde

algılanması ve modellemesini sağlamaktadır. Bu yaklaşım, insan ve nesne arasındaki etkileşimleri gerçeğe en yakın biçimde simüle edebilme kapasitesi sağlamaktadır.

Sonuç olarak, bu proje, 3D modellemenin teknik ve yaratıcı olanaklarını farklı sektörlerde nasıl kullanabileceğimizi göstermektedir. İnsan ve nesne etkileşimlerini modelleme, bu teknolojilerin doğru kullanımı ve geliştirilmesiyle ilgili önemli bulgular sağlamaktadır. Ülkemizde yapılan çalışmaların sınırlı olduğu göz önüne alındığında, bu çalışma, ulusal ve uluslararası düzeyde 3D modellemenin insan-nesne etkileşimlerine nasıl katkı sağlayabileceğini göstermekte ve bu alanda yapılan yeni araştırmalar için bir temel oluşturmaktadır.

Bu çalışma ayrıca, insan-nesne etkileşimlerinin yanı sıra oyun geliştirme, sanal ve karma gerçeklik, insan analitiği ve insan-robot işbirliği gibi uygulamaların nasıl iyileştirebileceğini göstermektedir. İnsan davranışlarını inceleyerek kişiselleştirilmiş avatarlar oluşturulabilir ve gerçek hayattan alınan senaryoların dijital ortamda analiz edilmesi kolaylaştırılabilir. Bu çalışma, gelecekteki benzer projeler için bir kılavuz oluşturmayı ve 3D modelleme teknolojilerinin daha da ileriye taşınmasını sağlayacak bir temel oluşturmayı amaçlamaktadır.

5. KAYNAKLAR

- [1] C. Laing and T. Schlick, "Computational Approaches to 3D Modeling of RNA." *Journal of Physics: Condensed Matter*, 22(28), 2010.
- [2] A. Watkins, *Creating Games with Unity and Maya: How to Develop Fun and Marketable 3D Games*. 2209/A Routledge, 2012.
- [3] K. Nickel, E. Scemann and R. Stiefelhagen, "3D-Tracking of Head and Hands for Pointing Gesture Recognition In a Human-Robot Interaction Scenario." In *Proc. IEEE FG*, pp. 565–570, 2004.
- [4] T. Becker, C. Nagel and T. H. Kolbe, "Integrated 3D Modeling of Multi-Utility Networks and Their Interdependencies for Critical Infrastructure Analysis." In *Advances in 3D Geo-Information Sciences*, pp. 1–20, 2010.
- [5] H. Joo, T. Simon, and Y. Sheikh. "Total Capture: a 3D Deformation Model for Tracking Faces, Hands and Bodies." In *Proc. CVPR*, pp. 8320–8329, 2018.
- [6] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik. "End-To-End Recovery of Human Shape and Pose." In *Proc. CVPR*, 2018.
- [7] T. Alldieck, M. Magnor, B. L. Bhatnagar, C. Theobalt, and G. Pons-Moll. "Learning to Re-Construct People In Clothing from a Single RGB Camera." In *Proc. CVPR*, pp. 1175–1186, 2019.
- [8] T. Zhi, C. Lassner, T. Tung, C. Stoll, S. G. Narasimhan, and M. Vo, "Texmesh: Reconstructing Detailed Human Texture and Geometry from RGB-D Video." In *Proc. ECCV*, 2020.
- [9] S. Südor, "Dijital Sanatlarda 3D Model Oluşturma Tekniklerinin Kullanımı". *İdil*, pp. 1757–1779, 2021.
- [10] B. L. Bhatnagar, X. Xie, I. A. Petrov, C. Sminchisescu, C. Theobalt and G. Pons-Moll, "Behave: Dataset and Method for Tracking Human Object Interactions." In *Proc. CVPR*, pp. 15935–15946, 2022.

- [11] M. A. Günen, A. Kesikoğlu, E. Beşdok and A. E. Karkinli, “RGB-D Sensörler ile İç Mekân Haritalaması.” In Proc. IDAP, 2017.
- [12] U. Tuğtekin and D. Kaleci, “3D Modelleme Tekniği Kullanılarak Bilgisayar Oyunu Tasarımı.” 13. Akademik Bilişim Konferansı, 2010.
- [13] C. Şahin, “3D Karakter Modellemesi ve Animasyonu Aşamaları.” İdil Sanat ve Dil Dergisi, pp. 387–392, 2019.
- [14] Xianghui Xie, Xi Wang, Nikos Athanasiou, Bharat Lal Bhatnagar, Chun-Hao P Huang, Kaichun Mo, Hao Chen, Xia Jia, Zerui Zhang, Liangxian Cui, et al. RHOBIN Challenge: Reconstruction of human object interaction. arXiv preprint arXiv:2401.04143, 2024.
- [15] Nancy S Pollard and Victor Brian Zordan. Physically based grasping control from example. In SIGGRAPH, 2005
- [16] He Zhang, Yuting Ye, Takaaki Shiratori, and Taku Komura. ManipNet: neural manipulation synthesis with a hand-object spatial representation. ACM Transactions on Graphics, 40(4):1–14, 2021.
- [17] Enric Corona, Albert Pumarola, Guillem Alenya, Francesc Moreno-Noguer, and Gregory Rogez. Ganhand: Predicting ‘ human grasp affordances in multi-object scenes. In CVPR, 2020
- [18] Patrick Grady, Chengcheng Tang, Christopher D Twigg, Minh Vo, Samarth Brahmbhatt, and Charles C Kemp. ContactOpt: Optimizing contact to improve grasps. In CVPR, 2021.
- [19] Hanwen Jiang, Shaowei Liu, Jiashun Wang, and Xiaolong Wang. Hand-object contact consistency reasoning for human grasps generation. In ICCV, 2021.
- [20] Korrawe Karunratanakul, Jinlong Yang, Yan Zhang, Michael J Black, Krikamol Muandet, and Siyu Tang. Grasping field: Learning implicit representations for human grasps. In 3DV, 2020.

- [21] Samarth Brahmbhatt, Ankur Handa, James Hays, and Dieter Fox. ContactGrasp: Functional multi-finger grasp synthesis from contact. In IROS, 2019.
- [22] Renaud Detry, Dirk Kraft, Anders Glent Buch, Norbert Kruger, and Justus Piater. Refining grasp affordance models by experience. In ICRA, 2010.
- [23] Kaijen Hsiao and Tomas Lozano-Perez. Imitation learning of whole-body grasps. In international conference on intelligent robots and systems, 2006.
- [24] Yixin Chen, Sai Kumar Dwivedi, Michael J Black, and Dimitrios Tzionas. Detecting human-object contact in images. In CVPR, 2023.
- [25] Davis Rempe, Leonidas J Guibas, Aaron Hertzmann, Bryan Russell, Ruben Villegas, and Jimei Yang. Contact and human dynamics from monocular video. In ECCV, 2020. 2
- [26] Davis Rempe, Tolga Birdal, Aaron Hertzmann, Jimei Yang, Srinath Sridhar, and Leonidas J Guibas. HuMoR: 3D human motion model for robust pose estimation. In ICCV, 2021.
- [27] Soshi Shimada, Vladislav Golyanik, Weipeng Xu, and Christian Theobalt. PhysCap: Physically plausible monocular 3D motion capture in real time. ACM TOG, 2020
- [28] Andrei Zanfir, Elisabeta Marinoiu, and Cristian Sminchisescu. Monocular 3D pose and shape estimation of multiple people in natural scenes-the importance of multiple scene constraints. In CVPR, 2018.
- [29] Yuliang Zou, Jimei Yang, Duygu Ceylan, Jianming Zhang, Federico Perazzi, and Jia-Bin Huang. Reducing footskate in human motion reconstruction with ground contact constraints. In WACV, 2020.
- [30] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. SMPL: A skinned multiperson linear model. ACM TOG, 2015.
- [31] Mohamed Hassan, Partha Ghosh, Joachim Tesch, Dimitrios Tzionas, and Michael J Black. Populating 3D scenes by learning human-scene interaction. In CVPR, 2021.

- [32] Chun-Hao P Huang, Hongwei Yi, Markus Hoshle, Matvey S. Safroshkin, Tsvetelina Alexiadis, Senya Polikovsky, Daniel Scharstein, and Michael J Black. Capturing and inferring dense full-body human-scene contact. In CVPR, 2022.
- [33] Soshi Shimada, Vladislav Golyanik, Zhi Li, Patrick Perez, Weipeng Xu, and Christian Theobalt. HULC: 3D human motion capture with pose manifold sampling and dense contact guidance. In ECCV, 2022.
- [34] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In ICLR, 2014.
- [35] Shashank Tripathi, Agniv Chatterjee, Jean-Claude Passy, Hongwei Yi, Dimitrios Tzionas, and Michael J Black. DECO: Dense estimation of 3D human-scene contact in the wild. In ICCV, 2023.
- [36] Jason Y Zhang, Sam PePOSE, Hanbyul Joo, Deva Ramanan, Jitendra Malik, and Angjoo Kanazawa. Perceiving 3D human-object spatial arrangements from a single image in the wild. In ECCV, 2020.
- [37] Xiang Xu, Hanbyul Joo, Greg Mori, and Manolis Savva. D3D-HOI: Dynamic 3D human-object interactions from videos. arXiv preprint arXiv:2108.08420, 2021.
- [38] Xianghui Xie, Bharat Lal Bhatnagar, and Gerard Pons-Moll. CHORE: Contact, human and object reconstruction from a single RGB image. In ECCV, 2022.
- [39] Bharat Lal Bhatnagar, Xianghui Xie, Ilya A Petrov, Cristian Sminchisescu, Christian Theobalt, and Gerard Pons-Moll. BEHAVE: Dataset and method for tracking human object interactions. In CVPR, 2022.
- [40] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. <https://github.com/facebookresearch/detectron2>, 2019.
- [41] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016.
- [42] Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J Black. Generating 3D faces using convolutional mesh autoencoders. In ECCV, 2018.

- [43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In NeurIPS, 2017.
- [44] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. 2017.
- [45] Gyeongsik Moon, Hongsuk Choi, and Kyoung Mu Lee. Accurate 3D hand pose estimation for whole-body 3D human mesh estimation. In CVPRW, 2022.
- [46] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In ICLR, 2014.
- [47] A. Farhadi, J. Redmon, «YOLO9000, Better, Faster, Stronger,» 25 Aralık 2016. [Çevrimiçi]. Available: <https://arxiv.org/pdf/1612.08242>
- [48] A. Farhadi, J. Redmon, «YOLOv3: An Incremental Improvement,» 8 Nisan 2018. [Çevrimiçi]. Available: <https://arxiv.org/abs/1804.02767>.
- [49] D. Çakmak, «Medium,» 14 Ekim 2020. [Çevrimiçi]. Available: <https://deryacakmak.medium.com/yolo-you-look-only-once-algoritmas%C4%B1-9e8819c5e4cf>.
- [50] Y. Mesci, «Medium,» 29 Nisan 2019. [Çevrimiçi]. Available: <https://medium.com/deep-learning-turkiye/yolo-algoritmas%C4%B1n%C4%B1-anlamak-290f2152808f>.
- [51] Ş. Koç, «Medium,» 4 Ağustos 2021. [Çevrimiçi]. Available: <https://medium.com/co-one-tr/bounding-box-yontemi-ile-veri-etiketleme-5fa0036d981c>.
- [52] Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A. *International Journal of Computer Vision*, 111(1), 98-136, 2015
- [53] G. Papandreou, F. Schroff, H. Adam, L.Chen, «Rethinking Atrous Convolution for Semantic Image Segmentation,» 5 Aralık 2017. [Çevrimiçi]. Available: <https://arxiv.org/abs/1706.05587v3>.

[54]G. Papandreou, F. Schroff, H. Adam, L.Chen «Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation,» 22 Ağustos 2018. [Çevrimiçi]. Available: <https://arxiv.org/pdf/1802.02611>.

[55] Bharat Lal Bhatnagar, Xianghui Xie, Ilya A Petrov, Cristian Sminchisescu, Christian Theobalt, and Gerard Pons-Moll. BEHAVE: Dataset and method for tracking human object interactions. In CVPR, 2022

[56] Blender Vakfı (2023). *Blender - Açık Kaynak 3D Modelleme ve Render Yazılımı*. Erişim: <https://www.blender.org/>

ÖZGEÇMİŞ

Adı-Soyadı : Simge Arlanoglu
Doğum Yeri ve Tarihi : Bursa-2001
E-posta : simgearlanoglu18@gmail.com
Ana Dili : Türkçe
Yabancı Dili : İngilizce
Linkedin : <https://www.linkedin.com/in/simge-arlanoglu/>

Derece	Okul	Mezuniyet Yılı
Lise	Yalova Termal Fen Lisesi	2019
Üniversite	Marmara Üniversitesi	2024

Yıl	Firma	Pozisyon
2023	Aktek Bilişim	RPA stajyeri
2024	IBTECH	Intern Application Engineer

ÖZGEÇMİŞ

Adı-Soyadı : Zeynep Koç
Doğum Yeri ve Tarihi : Manisa-2000
E-posta : zeynep_4553@hotmail.com
Ana Dili : Türkçe
Yabancı Dili : İngilizce
Linkedin : <https://www.linkedin.com/in/zeynepkoc/>

Derece	Okul	Mezuniyet Yılı
Lise	Şehremini Anadolu Lisesi	2019
Üniversite	Marmara Üniversitesi	2024

Yıl	Firma	Pozisyon
2023	Ritma Teknoloji	Yazılım geliştirme stajyeri
2024	Orion Innovation	Junior Software Developer