



POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)

URL: <http://dergipark.org.tr/politeknik>



Kalman filtresi ve küresel en yakın komşu yöntemi ile çok kişili gerçek zamanlı poz takibi

Multi-person real-time pose tracking using kalman filter and global nearest neighbor method

Yazar(lar) (Author(s)): Ahmet Samet HALICI¹, Ayşe DEMİRHAN²

ORCID¹: 0000-0002-8925-8205

ORCID²: 0000-0001-9227-9210

To cite to this article: Halıcı A. S., Demirhan A., “Kalman filtresi ve küresel en yakın komşu yöntemi ile çok kişili gerçek zamanlı poz takibi”, *Journal of Polytechnic*, 26(2): 889-899, (2023).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Halıcı A. S., Demirhan A., “Kalman filtresi ve küresel en yakın komşu yöntemi ile çok kişili gerçek zamanlı poz takibi”, *Politeknik Dergisi*, 26(2): 889-899, (2023).

Erişim linki (To link to this article): <http://dergipark.org.tr/politeknik/archive>

DOI: 10.2339/politeknik.1049933

Kalman Filtresi ve Küresel En Yakın Komşu Yöntemi ile Çok Kişili Gerçek Zamanlı Poz Takibi

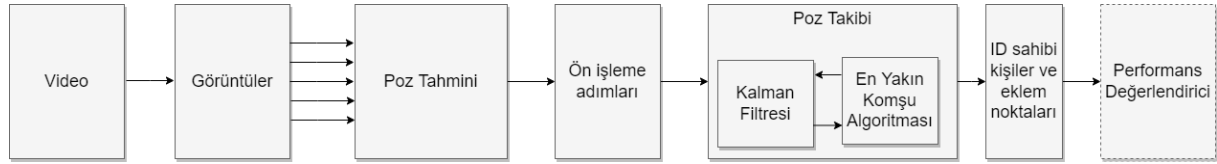
Multi-Person Real-Time Pose Tracking Using Kalman Filter and Global Nearest Neighbor Method

Önemli noktalar (Highlights)

- ❖ Gerçek zamanlı çok kişili poz takibi başarımını artıran ön işleme adımlarının geliştirilmesi / Development of preprocessing steps that increase performance of real-time multi-person pose tracking
- ❖ Poz takibi için kullanılacak en uygun poz tahmini yönteminin belirlenmesi / Determining the most appropriate pose estimation method that can be used for pose tracking
- ❖ Kalman filtresi (KF) ve küresel en yakın komşu (KEYK) ile bir poz takipçisi tasarımı / Design of a pose tracker using Kalman filter (KF) and global nearest neighbor (GNN)
- ❖ Geliştirilen poz takipçisinin performansının farklı poz tahmini yöntemleri ile değerlendirilmesi / Evaluation of the developed pose tracker's performance with different pose estimation methods

Grafik Özet (Graphical Abstract)

KF ve KEYK kullanılarak gerçek zamanlı çok kişili bir poz takibi algoritması geliştirilmiştir. Bu algoritma kapsamında geliştirilen ön işleme adımlarının farklı poz tahmini yöntemleri üzerindeki etkisi değerlendirilmiştir.



Şekil. Akış Diyagramı /Figure. Flow Chart

Amaç (Aim)

Geliştirilen çok kişili poz takip sistemini farklı poz tahmini algoritmalarıyla kullanarak tüketilen kaynak, başarımlar ve işlem süresi açısından değerlendirmektir. / To evaluate the developed multi-person pose tracking system in terms of consumed resource, success and processing time using different pose estimation algorithms.

Tasarım ve Yöntem (Design & Methodology)

Yukarıdan aşağı ve aşağıdan yukarı yaklaşımını kullanan poz tahmini yöntemleri için kişi takibi işleminde veri ilişkilendirme adımında KEYK, veri güncelleme adımında ise KF kullanılmıştır. Çeşitli ön işlemler kullanılarak poz tahmini çıktılarında düzeltme adımları uygulanmıştır. / GNN is used in the data association step and KF is used in the data update step for top-down and bottom-up pose estimation methods in the pose tracking process. Correction steps were applied to the pose estimation outputs obtained with various preprocessing steps.

Özgünlük (Originality)

PoseTrack veri kümesi üzerinde daha önce kullanılmamış olan DCPose poz tahminini kullanmak ve ön işleme adımları ile poz takibindeki başarımları artırmak. / Using DCPose pose estimation which has not been used before on the PoseTrack dataset and increasing the performance in pose tracking with the preprocessing steps.

Bulgular (Findings)

Ön işleme adımları başarımları sabit hızlı ve sabit ivmeli hareket modellerinde OpenPose için %22.6 ve %16.2, DCPose için %21.2 ve %21.8 artırmıştır. / Preprocessing steps increased performance by 22.6% and 16.2% for OpenPose, 21.2% and 21.8% for DCPose in constant velocity and constant acceleration motion models.

Sonuç (Conclusion)

DCPose, OpenPose'a göre daha başarılı olmuştur. Ön işleme adımları kullanıldığında her iki poz tahmini yöntemi ile de daha yüksek poz takibi başarımları elde edilmiştir. / DCPose, has been more successful than OpenPose. When the preprocessing steps are used, higher pose tracking performance is obtained with both pose estimation methods.

Etik Standartların Beyanı (Declaration of Ethical Standards)

Bu makalenin yazar(lar)ı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler. / The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Kalman Filtresi ve Küresel En Yakın Komşu Yöntemi ile Çok Kişili Gerçek Zamanlı Poz Takibi

Araştırma Makalesi / Research Article

Ahmet Samet HALICI^{1,2}, Ayşe DEMİRHAN^{3*}

¹Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik Elektronik Mühendisliği Programı, 06500, Ankara

²Komuta Kontrol ve Savunma Teknolojileri Bölümü, HAVELSAN A.Ş., Ankara, Türkiye

³Gazi Üniversitesi, Teknoloji Fakültesi, Elektrik Elektronik Mühendisliği Bölümü, 06500, Ankara

(Geliş/Received : 28.12.2021 ; Kabul/Accepted : 04.04.2022 ; Erken Görünüm/Early View : 07.06.2022)

ÖZ

Poz tahmini, kamera ile çekilen görüntülerde insan iskeletindeki anahtar noktaların piksel konumlarının tespit edilmesi amacıyla ortaya çıkmıştır. Poz tahmini yöntemlerinin çıktıları görüntüde tespit edilen tüm eklem noktalarının piksel değerlerini ilişkilendirdiği kişiye göre vermektedir. Videolarda kişilerin hareketlerini anlamlandırmak için ardışık görüntü kareleri boyunca kimliklendirilmeleri gerekir. Böylece kişilerin video boyunca ne zaman hangi hareketleri yaptığı tespit edilebilir. Bu çalışmada sabit hızlı ve sabit ivmeli hareket modeline göre Kalman filtresi kullanarak küresel en yakın komşu (KEYK) algoritması ile tasarlanan çok kişili poz takibi yönteminin verdiği sonuçlar incelenmiştir. Geliştirilen ön işleme adımlarının poz tahmini yöntemlerinin kalitesini artırarak poz takibine etkisi de tespit edilmiştir. Bu amaçla PoseTrack veri kümesi üzerinde DCPose ve OpenPose poz tahmini yöntemlerinin başarımları değerlendirilmiştir. Ön işleme adımları ile sistemin başarımlarının her iki yöntem için de yükseldiği görülmüştür. Gerçek zamanlı çalışabilen, başarılı bir poz tahmini yöntemi olan ve düşük kaynak tüketimine sahip OpenPose yöntemi ile literatürde en iyi sonuçları veren DCPose yönteminin sonuçları incelendiğinde çok kişili poz takibi konusunda DCPose yönteminin daha başarılı sonuçlar verdiği görülmüştür. 550 farklı video ile elde edilen sonuçlar ön işleme adımları uygulandığında başarımları sabit hızlı ve sabit ivmeli hareket modellerinde aşağıdan yukarı yöntemi OpenPose için %22.6 ve %16.02, yukarıdan aşağı yöntemi DCPose için %21.2 ve %21.8 artırmıştır.

Anahtar Kelimeler: Gerçek zamanlı poz takibi, poz tahmini, kalman filtresi, küresel en yakın komşu yöntemi, hareket tanıma için poz takibi.

Multi-Person Real-Time Pose Tracking Using Kalman Filter and Global Nearest Neighbor

ABSTRACT

Pose estimation has emerged in order to detect pixel positions of keypoints on the human skeleton in images taken with the camera. The outputs of the pose estimation methods give the pixel values of all the articulation points detected in the image according to the person they associate with. In order to make sense of people's movements in videos, people need to be identified across successive frames. Thus, it can be determined when people make which movements during the video. In this study, the results of a multi-person exposure tracking method that is designed with the global nearest neighbor (GNN) algorithm using the Kalman filter based on constant velocity and constant acceleration motion models were examined. The effect of the developed preprocessing steps that increase the quality of the pose estimation methods on the pose tracking has also been determined. For this purpose, the performance of DCPose and OpenPose pose estimation methods on the PoseTrack dataset was evaluated. It was observed that the performance of the system increased for both methods with the preprocessing steps. When the results of OpenPose method, which can work in real time, a successful pose estimation method and have low resource consumption, and DCPose method, which gives the best results in the literature, are examined, it is seen that DCPose method gives better results in multi-person pose tracking. The results obtained with 550 different video increased the performance in constant velocity and constant acceleration motion models by 22.6% and 16.02% for bottom-up method OpenPose and 21.2% and 21.8% for top-down method DCPose when preprocessing steps were applied.

Keywords: Real-time pose tracking, pose estimation, kalman filter, global nearest neighbor method, pose tracking for action recognition.

1. GİRİŞ (INTRODUCTION)

Çok kişili poz takibi insan hareketlerinin tanınması, video özetlemede önemli sahnelerin çıkartılması ve video içeriklerinin sınıflandırılmasında kullanılan önemli bir problemdir [1-4]. Poz takibi, insan eklem noktalarının

yörüngelerinin doğru tahmini, insan eylemi tanıma, insan etkileşimini anlama, hareket yakalama ve animasyonlar için faydalıdır. Özellikle kameraların rastgele hareketler yaparken kaydettiği videolarda çok kişili poz takibi oldukça zordur. Çünkü tekil kameralarda kişiler birbirinin önünden geçerken arkada kalan kişiler geçici olarak hiç görünmeyebilir. Kişilerin takibi görünmediği

*Sorumlu Yazar (Corresponding Author)
e-posta : ayseoguz@gazi.edu.tr

görüntü karelerinde de geçici olarak yapılabilmektedir. Poz takibi iki ana adımda gerçekleşir. Birinci adımda, kişilerin eklem noktalarının tespit edildiği poz tahmini yapılır. Bu adım bir sonraki adımda kullanılacak ölçümleri elde etmek için gereklidir. İkinci adımda, görüntü içinde tespit edilen çok sayıdaki eklem noktasının ait olduğu kişiye atanması ve zamana bağlı olarak o kişinin videoda bulunduğu sahne boyunca izlenmesi gerçekleştirilir [5].

Kamera ile çekilen bir görüntüde insan iskeletindeki eklem noktalarının çıkartılması işlemi poz tahmini olarak isimlendirilir. Poz tahmininde, görüntüdeki kişilerin önceden belirlenen eklem noktalarının piksel değerleriyle birlikte güvenilirlik değerinin bulunması amaçlanır. Güvenilirlik değeri 0-1 arasında değişir. Poz tahminin bazı uzuvların görüntüde olmadığı ya da kişinin bazı uzuvlarının önüne başka bir nesnenin geçtiği durumlarda da iyi çalışabilmesi amacıyla farklı yaklaşımlar geliştirilmiştir. Bu yaklaşımlar aşağıdan yukarı yaklaşım ve yukarıdan aşağı yaklaşım olarak isimlendirilir. Aşağıdan yukarı yaklaşım önce tüm vücut eklemlerini parça temelli olarak bulur ve bulunan eklem noktasının kime ait olduğunu belirleyerek o kişiye eklem noktası ataması yapar. Yukarıdan aşağı yaklaşım ise önce kişileri sınırlayan bölgeyi bir dikdörtgen alan biçiminde tespit eder ve daha sonra bu dikdörtgen bölgenin içinde yer alan eklem noktalarını bulmaya çalışır [6-8].

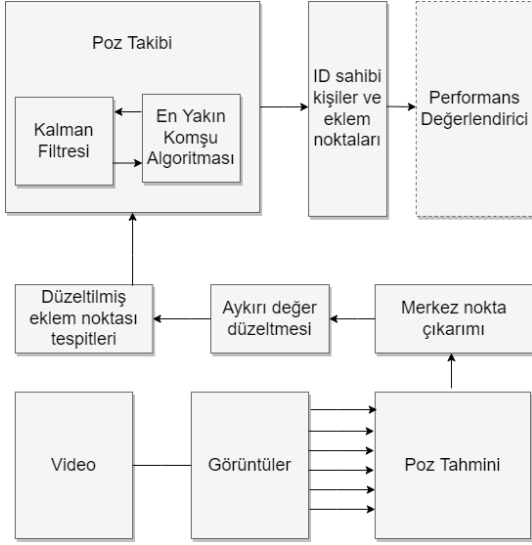
Poz tahmini adımında genellikle evrişimli sinir ağları kullanılmaktadır [9]. Aşağıdan yukarı yaklaşımını kullanan çalışmalara örnek olarak Cao ve diğerleri [10] tarafından yapılan çalışma gösterilebilir. Bu çalışmada OpenPose isimli yöntem geliştirilmiştir. Bu yöntemde eklem noktalarının ısı haritalarını ve eklemlerin aralarındaki ikili ilişkileri (parça yakınlık alanları) tahmin etmek için bir model oluşturulmuştur [10]. Eklem noktaları, aralarındaki parça yakınlık alanlarının ikili eşleşmelerine göre Hungarian algoritması [11] ile gruplandırılır. Hesaplama performansı, görüntüde tespit edilen kişi sayısından bağımsızdır. OpenPose yöntemi çok kişili poz tahminlerinde gerçek zamanlı olarak çalışabilir. Ladicky ve diğerleri [12] yönlü gradyanlar histogramı olarak adlandırılan öznelik çıkarımı ile insan uzuvlarını bölütlemiş ve uzuvların pozisyonlarını birlikte tahmin eden bir algoritma önermiştir. Yukarıdan aşağı yaklaşımını kullanan Sun ve diğerleri [8] HRNet olarak da bilinen farklı çözünürlükteki temsilleri paralel olarak işleyen ve bilgi paylaşımı yapan bir ağ yapısı kurarak çok ölçekli bir birleştirme uygulamıştır. HRNet eklem noktalarını potansiyel ısı haritası ile belirleyen başarılı ve önde gelen bir yöntem olarak kullanılmaktadır. Wang ve diğerleri [7] tarafından önerilen ve yukarıdan aşağı çalışan yöntem diğer yöntemlerden farklı olarak kişileri tespit eden ağdan bağımsız olarak tespit edilmeyen kişilerin de eklem noktalarını tahmin edebilir. Bu çalışmada kişileri çevreleyen kutunun içinde başka kişilerin de olabileceği ve bir kişinin diğerinin önüne geçtiği için tespit edilememiş olabileceği durumlar da göz önüne alınarak klip izleme ağı olarak isimlendirilen

bir ağ yapısı ile tahmin işlemi yapılmıştır. Ardışık görüntü karelerinde üst üste gelen tespitleri ilişkilendirerek tespit edilmeyen kişiler üzerinde de poz tahmini yapabilmeyi öneren bir çalışmadır [7]. Liu ve diğerleri [6] geleneksel görüntü tabanlı yaklaşımların video kareleri arasında zamansal bağımlılığı ve geometrik tutarlılığı göz ardı ettiğini öne sürerek DCPose olarak bilinen yeni bir yapı önermiştir. DCPose ardışık görevler icra eden üç modül içermektedir. İlk modül ardışık üç video karesi üzerinde bulunan eklem noktalarını birleştirir. İkinci modül mevcut görüntü ve ardışık görüntüler arasındaki poz artıklarını verimli bir şekilde elde etmek için geçici mesafeleri kullanarak çerçeveler arası eklem noktalarının farklarını hesaplar. Üçüncü modül ise eklem noktalarının ısı haritalarının yeniden örneklenmesi için farklı genişleme oranlarına sahip beş paralel evrişim katmanından oluşan bir poz düzeltme ağı ile son çıktıları üretir.

Poz takibi adımında çeşitli yöntemler mevcuttur. Poz takibi çalışmaları genellikle poz tahminlerini zaman içinde birbiri ile yeni bir ağ yapısı ile ilişkilendirerek yapmaya çalışmaktadır [7, 13-16]. Yukarıdan aşağıya yöntemler [15, 16] genel olarak kişilerin eklem noktalarını sınırlayan bölgeler arasında zamansal grafikler oluşturularak ve bölgelerdeki özellikleri kullanarak ilişkilendirme yapar. Aşağıdan yukarı yöntemler [13, 14] ise tespit edilen eklemler arasında uzam-zaman grafikleri oluşturularak yakınlık alanları ile ilişkilendirme yapar. Raaj ve diğerleri [14] çıkarım sırasında geçmiş bilgileri de dikkate alarak sadece o anki görüntüyü kullanır çünkü gerçek zamanlı çalışabilmesi amaçlanmıştır. Her görüntüde parça yakınlık alanları, eklem noktaları ve geçici yakınlık alanları arasında bağlantı oluşturan bir ağ yapısı ile kişi tespitlerini ilişkilendirir [14]. SimpleBaseline yöntemi önce her bir karede bağımsız olarak bir kişi tespiti çalıştırır ve ardından algılamalarını pahalı optik akış kullanarak zamansal benzerliğin tanımlandığı bir grafikte ilişkilendirir [16]. Detect-and-Track yöntemi ise kısa bir video klipte insanların eklem noktalarını tespit etmek için bir 3D Mask R-CNN yaklaşımını kullanır ve ardından tespit edilen sınırlayıcı bölgelerin konumlarını karşılaştırarak ardışık klipleri birbirine bağlar [15]. Wang ve diğerleri [7] ise bir görüntüde tespit edilen sınırlayıcı kutuları ardışık görüntü karelerine yayarak gözden kaçabilecek kişileri aramayı önermektedir.

Poz takibi yöntemleri arasında performansı yüksek ve maliyeti düşük olan yöntemi bulmak önemli bir problemdir. Bu çalışmada bu problemin çözümü için gerçek zamanlı çalışan düşük maliyetli bir izleyici Kalman filtresi ve KEYK kullanılarak geliştirilmiş ve OpenPose [10] ve DCPose [6] poz tahmini yöntemlerini kullanarak değerlendirilmiştir. Ön işleme adımlarının sistemin performansına etkisi araştırılmıştır. Aşağıdan yukarı bir yaklaşım olan OpenPose literatürdeki düşük işlem gücü ihtiyacına göre en başarılı yöntemdir. Yukarıdan aşağı bir yaklaşım olan DCPose ise daha çok işlem gücüne ihtiyaç duymakla birlikte PoseTrack veri kümesinde en başarılı sonuçları veren poz tahmini

yöntemlerinden biridir. Geliştirilen poz takibi yönteminin yapısı Şekil 1’de verilmiştir.



Şekil 1. Geliştirilen poz takibi yönteminin yapısı (The structure of the developed pose tracking method)

Çalışmanın devamı aşağıdaki gibi tasarlanmıştır: İkinci bölümde, çalışmada kullanılan veri kümesi, ön işleme adımları, poz takibi sistemini geliştirmede kullanılan teknikler olan Kalman filtresi ve KEYK ile performans değerlendirme ölçütleri açıklanmıştır. Üçüncü bölümde, çalışmadan elde edilen deneysel sonuçlar sunulmuştur. Dördüncü bölümde ise çalışmadan elde edilen sonuçlar tartışılmıştır.

2. MATERYAL VE METOD (MATERIAL and METHOD)

2.1. Veri Kümesi (Data Set)

Bu çalışmada, Pishchulin ve arkadaşları tarafından derlenen PoseTrack veri kümesi kullanılmıştır [17]. PoseTrack video tabanlı çok kişili insan pozunu tahmini ve poz takibi için kullanılan büyük ölçekli bir veri kümesidir. 66374 görüntü karesine sahip 550 adet videodan oluşur. Çizelge 1 PoseTrack veri kümesini oluşturmada kullanılan kaynakları, etiketli pozların sayısını, videoda birden çok kişinin bulunup bulunmadığını, video poz etiketlerinin kullanılabilirliğini ve videoların türünü göstermektedir [17]. Toplamda 153615 adet açıklamalı poz ve 22000’den fazla etiketli görüntü bulunmaktadır. Veri kümesinde etiketleme işlemi için VATIC [18] aracı kullanılmıştır. Bu araç sayesinde ardışık tüm görüntü karelerini etiketlemeden boş bırakılan görüntü kareleri nesne takibi yapılarak birleştirilebilmektedir. Çizelge 1’den görülebileceği gibi veri kümesi spor, film ve blog gibi çok farklı özelliklere sahip çok kişili videolar içermektedir. Şekil 2 veri kümesinde bulunan farklı özelliklerdeki videolardan elde edilmiş örnek görüntüler içermektedir. Veri kümesinde çok kalabalık, insanların kameraya göre birbirinin önünü kapattığı ve az kişili görüntü örnekleri Şekil 2’de bir

araya getirilmiştir. Bazı sahneler oldukça kalabalık ve karmaşık iken bazı sahneler ise az kişili ve sadedir.



Şekil 2. PoseTrack veri kümesi örnek görüntüleri ile elde edilen poz tahminleri. Üst iki satır OpenPose çıktıları, alt iki satır DCPose çıktılarıdır (Pose estimates obtained with the PoseTrack dataset sample images. Top two lines are OpenPose outputs, bottom two lines are DCPose outputs)

Çizelge 1. PoseTrack veri kümesini oluşturmada kullanılan kaynaklar ve videoların özellikleri [17] (Characteristics of the resources and videos used to create the PoseTrack dataset [17])

Dataset	# Poses	Multi-person	Video-labeled poses	Data type
LSP [25]	2,000			sports (8 act.)
LSP Extended [26]	10,000			sports (11 act.)
MPII Single Person [1]	26,429			diverse (491 act.)
FLIC [38]	5,003			feature movies
FashionPose [9]	7,305			fashion blogs
We are family [10]	3,131	✓		group photos
MPII Multi-Person [1]	14,993	✓		diverse (491 act.)
MS COCO Keypoints [28]	105,698	✓		diverse
Penn Action [51]	159,633		✓	sports (15 act.)
JHMDB [23]	31,838		✓	diverse (21 act.)
YouTube Pose [6]	5,000		✓	diverse
Video Pose 2.0 [39]	1,286		✓	TV series
Multi-Person PoseTrack [22]	16,219	✓	✓	diverse
Proposed	153,615	✓	✓	diverse

DCPose outputs)

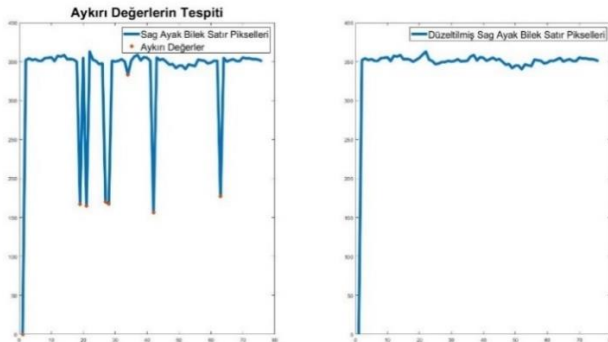
PoseTrack veri kümesi üç farklı görevde yöntemlerin karşılaştırılmasına imkân sağlar;

- Tek görüntü ve çok kişili poz tahmini
- Videolarda çok kişili poz tahmini
- Çok kişili poz takibi

2.2. Ön İşleme Adımları (Pre-Processing Steps)

Bu çalışmada ön işleme uygulamaları beş ana adımda gerçekleştirilmiştir. Birinci adım, eklem noktaları arasındaki aykırı değerlerin tespit edilerek düzeltilmesi işlemidir. Aykırı değerleri tespit etmek için bir videoda bulunan tüm eklem noktaları arasında inceleme yapılması gerekmektedir. Bir eklem noktasının

değerlerinde ardışık görüntüler arasında büyük değişimler olmamalıdır. Eğer aykırı değerler varsa ve poz tahmini algoritmasının çıktısı olarak elde edildiyse, bu değerler poz takibi adımına geçmeden önce düzeltilmelidir. Bu çalışmada, aykırı değerlerin tespit edilmesi için üç ölçekli mutlak medyan sapma kullanılmıştır. Bu yöntemde alt eşik, medyan ve üst eşik hesaplanmaktadır. Bu eşik değerlerinin dışında kalan değerler aykırı değer olarak tespit edilmektedir. Böylece piksel değerlerinde meydana gelebilecek aşırı değişimler tespit edilmektedir. Örneğin bir eklem noktasına ait ardışık 5 video karesinde tespit edilen satır piksel değerleri (50, 54, 59, 182, 53) şeklinde ise 4. görüntüdeki 182 değerinin poz tahmin yöntemi tarafından yanlış tahmin edildiği tespit edilerek bu aykırı değer için düzeltme işlemi yapılmaktadır. Aykırı değerlerin gerçek değerinin eğer önceki görüntü ile sonraki görüntü bir aykırı değere sahip değilse bu değerlerin arasında olması gerektiği ön görülmektedir. Bu nedenle aykırı değerlerin önceki ve sonraki değerleri aykırı değer değilse bu değerlerin aritmetik ortalaması alınarak düzeltme işlemi yapılmıştır. Şekil 3 tespit edilen aykırı değerler ve düzeltme sonucunda elde edilen değerleri göstermektedir. Grafiklerde x eksenini video kare numarasını y eksenini ise sağ ayak bileğinin bu karedeki satır pikselini göstermektedir. Sol tarafta aykırı değerlerin tespit edildiği grafik, sağ tarafta ise tespit edilen bu aykırı değerlerin düzeltildikten sonraki değerleri gösterilmektedir.

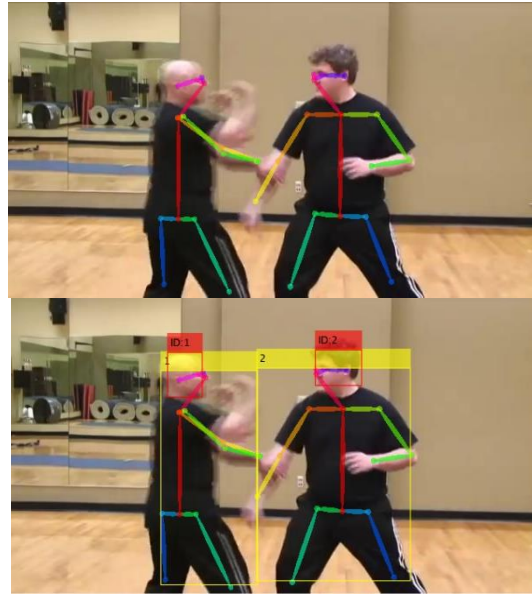


Şekil 3. Poz tahmini algoritmalarının çıktılarındaki aykırı değerlerin tespit edilmesi ve düzeltilmesi adımı (Step of detecting and correcting outliers in outputs of pose estimation algorithms)

İkinci ön işleme adımı merkez nokta çıkarımıdır. Bu çalışmadaki poz takibi yöntemi her kişi için bir merkez nokta belirlenerek bu noktanın video boyunca takip edilmesini amaçlamaktadır. Bu nedenle merkez noktanın, anormal değişimlerden uzak bir nokta olması gerekir. İnsanların el, kol, bacak ve ayak noktaları gövde ve boyun noktasına göre daha hareketlidir. Bu nedenle merkez nokta olarak gövdeye bağlı bir nokta ya da boyun noktası merkez olarak kabul edilmelidir. Bu çalışmada merkez nokta olarak boyun noktası kabul edilmiştir ve boyun noktasının görünmediği durumlarda tespit edilen diğer eklem noktaları kullanılarak vektörel boyun yaklaşımı olarak isimlendirilen yöntem ile boyun

noktasının konumu hesaplanmaktadır. Vektörel boyun yaklaşımı gövde ve kafa arasında ya da omuzlar arasındaki Öklid mesafelerini kullanarak ve gövdenin eğimini hesaplayarak boyun noktasının değerlerini tahmin eder [19]. Bu da poz takibi için eksik bilgi kullanılmamasını ve kararlı bir merkez noktası çıkarımının yapılmasını sağlar.

Üçüncü ön işleme adımı baş bölgesini sınırlayan alanın bulunmasıdır. Poz tahmini yöntemleri genellikle kulakları, gözleri, başın üst noktasını ve boynu tespit etmektedir fakat baş bölgesinin piksel konularının tespitini yapmamaktadır. Fakat PoseTrack veri kümesi, poz takibi başarımını ölçmek için bu bilgiyi istemektedir. Bu nedenle sonuçları değerlendirmeden önce bu ön işleme adımı ile analitik olarak baş bölgesini sınırlayan kutunun piksel değerleri hesaplanmıştır. DCPose baş bölgesinin sadece alt ve üst noktasını bulmaktadır. Kulakları tespit etmediğinden başın sağ ve sol tarafını bulamamaktadır. OpenPose ise yalnızca göz, kulak, burun ve boyun noktalarını bulmaktadır. Bu bilgileri kullanarak baş bölgesini sınırlayan alan DCPose ve OpenPose için ayrı ayrı tespit edilmiştir. Şekil 4 örnek bir video karesi için poz tahmini algoritmalarından alınan sonuçları ve bu sonuçlara ek olarak hesaplanan baş bölgesinin konumunu göstermektedir.



Şekil 4. Poz tahmininden elde edilen eklem noktaları ile baş bölgesinin bulunması (Finding the head region with the articulation points obtained from the pose estimation)

Baş bölgesini sınırlayan alanı tespit etmek için eklem noktaları içinde sabit iki tanesi arasında referans bir uzunluk belirlenir. Bu referans uzunluğa göre sol göz-sol kulak ve sağ göz-sağ kulak arasındaki Öklid mesafesi karşılaştırılır. Eğer sol göz ve sol kulak arasındaki mesafe sağ göz ve sağ kulak arasındaki mesafeden anormal şekilde büyükse referans uzunluk oranına göre kişinin kafasının bulunduğu görüntü koordinat sistemine göre başın dönük olduğu yönün çıkarımı yapılabilir. Kişinin

yönü Şekil 5'teki son baş görüntüsünde olduğu gibi sol tarafa bakıyorsa baş bölgesini sınırlayıcı alan başın arka tarafını da çevreleyebilmek için sağ tarafı da kapsamalıdır. Bu çalışmada doğru sınırlayıcı bölgeyi bulabilmek için genişleme yönü, geçici baş merkezi ve genişleme miktarı olmak üzere 3 adımlı hesaplama önerilmektedir. Genişleme yönü başın görüntü koordinat sistemine göre mevcut yönü ve referans uzunluğa göre



Şekil 5. Baş pozisyonlarına göre sadece poz tahmini ile hesaplanan baş bölgeleri ve tespit edilmesi gereken baş bölgeleri (Head regions calculated based on head positions using only pose estimation and head regions that should be detected)

Dördüncü ön işleme adımında videonun başrol oyuncularını arasında olmayan ve yanlış pozitif olma ihtimali yüksek kişilerin filtrelenmesi işlemi yapılır. Videolarda kişilerin tespit edilme yoğunluklarını hesaplayarak videoda belirli bir eşik değerinde görünen kişiler silinir. Örneğin bu eşik değeri %10 olarak belirlenirse ve video 70 görüntüden oluşuyorsa videoda 7 görüntü karesinden daha az bulunan kişiler hiç tespit edilmemiş varsayılır. Geçici olarak başlatılan kimliklendirilmiş kişiler filtrelenerek iptal edilir. Bu ön işleme adımı videolarda bulunan başrol oyuncularının tespit edilmesi için de kullanılabilir.

Beşinci ön işleme adımı ise sadece DCPose poz tahmini yöntemi için uygulanmaktadır. Bu adım el ve ayak bilekleri için bulunan sonuçların iyileştirilmesine yöneliktir. DCPose'un el ve ayak bilekleri sonuçları, PoseTrack veri kümesinin gerçek değerlerine göre Öklid mesafesi olarak bakıldığında kısadır. Örneğin ayak bileği için bu durumu düzeltmek amacıyla öncelikle kaval kemiğinin görüntü koordinat sistemine göre yönü hesaplanmalıdır. Bu yönde eklem noktasının morfolojik kemik yapılarına göre uzatılması gerekmektedir. İnsanların iskelet yapısına göre uyluk kemiği ve kaval kemiği uzunlukları ile çalışmayı Hasegawa ve arkadaşları [20] yapmıştır. Buna göre kemikler arası oran kullanılarak ayak bileğinin uzatılması gereken konum tahmin edilmiştir. Deneysel çalışmalar bölümünde bu işlem adımının sonuçları üzerindeki etkisi verilmiştir.

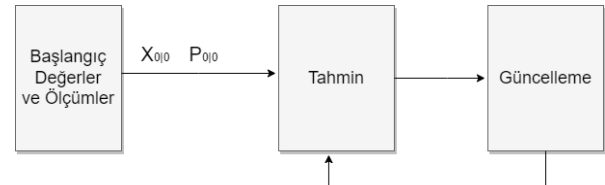
2.3. Poz Takibi (Pose Tracking)

Poz takibi problemi videolarda bulunan çok sayıda kişinin pozlarını (duruşlarını) tahmin ederek video kareleri boyunca bu kişileri takip edebilmek için poz ile ilişkili her bir anahtar noktaya benzersiz kimlik numaraları atanmasıdır. Bu amaçla çeşitli algılayıcılardan elde edilen veriler kullanılarak bilinmeyen parametreleri tahmin etmek gerekir. Bu çalışmada Kalman filtresi ve KEYK kullanılarak gerçek zamanlı çalışabilen bir takip modülü geliştirilmiştir.

durumuna bakılarak hesaplanır. Geçici baş merkezi tespit edilen mevcut noktaların en büyük ve en küçük değerlerinin ortalaması alınarak hesaplanır. Genişleme miktarı ise referans uzunluğa göre göz-kulak arasındaki Öklid mesafesi arasındaki ilişkiye bağlı olarak hesaplanır. Önerilen yöntem başın 3 boyutta gerçekleştirilebileceği dönüşlere dayanıklıdır.

2.3.1 Kalman filtresi (Kalman filter)

Kalman filtresi doğrusal ve doğrusallaştırılmış sistemlerde sistem değişkenlerini kullanarak durum tahmini yapabilmek amacıyla kullanılır. Filtreleme işlemi genel olarak tahmin ve güncelleme olmak üzere iki adımda yapılır. Tahmin adımında hareket denklemleri ve ölçümleri kullanılarak bir sonraki zaman adımındaki durum ve kovaryans değeri tahmin edilir. Güncelleme adımında ise ölçüm değerleri kullanılarak durum ve kovaryans değeri düzeltilerek güncellenir [21]. Kalman filtresinin çalışma prensibi Şekil 6'da akış diyagramı olarak verilmiştir.



Şekil 6. Kalman filtresi akış diyagramı (Kalman filter flowchart)

Şekil 6'daki döngü incelenirse $X_{0|0}$ ile gösterilen durum ve $P_{0|0}$ ile gösterilen durum kovaryansı için ilk değer ataması yapıldıktan sonra hareket modeline göre bir sonraki adımda kişilerin beklenen durumları ve beklenen kovaryans matrisi Eş. 1 ve Eş. 2 ile hesaplanır. Eş. 3 ile güncellenen zamandaki ölçüm değeri tahmin edilir [22]. Tahmin adımında Eş. 1-3 kullanılır. Güncelleme adımında ise Eş. 4-8 kullanılmaktadır [23, 24].

$$x_{k+1|k} = F_k x_{k|k} + G_u u_k \quad (1)$$

$$P_{k+1|k} = F_k P_{k|k} F_k^t + Q_k \quad (2)$$

$$z_{k+1|k} = H_{k+1} x_{k+1|k} \quad (3)$$

Eş. 1-3'teki $(k+1|k)$ gösterimi, k zaman anındaki değerlerle $k+1$ anındaki değerlerin tahmin edilmesidir. Bu durum a priori olarak da bilinir. $(k+1|k+1)$ gösterimi ise a posteriori olarak bilinir ve güncelleme sonrası tahmin anlamına gelir. Burada k zaman anını, F

durum geçiş matrisini, x durum vektörünü, G giriş geçiş matrisini, u giriş kontrol vektörünü, z ölçüm vektörünü ve H ölçüm matrisini ifade etmektedir. w değeri Q kovaryans matrisi ile sıfır ortalamaya sahip süreç gürültü vektörü, v ise R kovaryans matrisi ile sıfır ortalamaya sahip olan ölçüm gürültü vektörüdür. Güncellenen zaman anında ($k+1$) durumu düzeltmek için $k+1$ ölçümü ile k zaman anındaki ölçüm kullanılarak tahmin edilen değer arasındaki fark kullanılır. Güncelleme yapabilmek için önce ölçüm tahmin kovaryansı S_{k+1} (Eş. 4) hesaplanarak ardından Kalman kazancının K_{k+1} (Eş. 5) hesaplanması gerekir [23, 24].

$$S_{k+1} = H_{k+1}P_{k+1|k}H_{k+1}^T + R_{k+1} \quad (4)$$

$$K_{k+1} = P_{k+1|k}H_{k+1}^T S_{k+1}^{-1} \quad (5)$$

Hareket modeline göre tahmin edilen değer ölçümle güncellenir. Yani boyun noktasının tespiti ile hareket modelini kullanarak kişinin bir sonraki görüntüde nerede olacağı tahmin edilir ve bu tahmin ile bir sonraki görüntüde gelen boyun noktası değeri arasındaki farka göre düzeltme yapılır. Bu noktada ölçüm ile tahmin arasında doğrusal bir ilişki olduğu varsayılır. Eş. 6'daki $z_{k+1} - z_{k+1|k}$ bölümü ölçüm yeniliği (measurement innovation) olarak bilinmektedir (Eş. 13) [25]. Bu değerlerin ortalaması sıfır olmalıdır. Ölçüm yeniliği $k+1$ anındaki ölçüm değeri ile k anındaki ölçüm kullanılarak hesaplanan $k+1$ 'deki tahmin arasındaki fark olarak da ifade edilebilir [22, 23, 26].

$$X_{k+1|k+1} = X_{k+1|k} + K_{k+1}(Z_{k+1} - Z_{k+1|k}) \quad (6)$$

$$P_{k+1|k+1} = P_{k+1|k} - K_{k+1}S_{k+1}K_{k+1}^T \quad (7)$$

$$z_{k+1|k+1} = H_{k+1}Z_{k+1|k+1} \quad (8)$$

Deneylerde hareket modeli olarak sabit hızlı ve sabit ivmeli hareket modeli kullanılmıştır.

Sabit hızlı ve sabit ivmeli hareket modelleri doğrusal Kalman filtresi ile kullanılabilir. Sabit hızlı ve sabit ivmeli hareket modellerinde bir sonraki zaman adımında yapılan durum güncellemesi mevcut zamandaki durumun bir fonksiyonudur (Eş. 9 ve Eş. 10)

$$\begin{bmatrix} col_{k+1} \\ v_{col_{k+1}} \\ row_{k+1} \\ v_{row_{k+1}} \end{bmatrix} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} col_k \\ v_{col_k} \\ row_k \\ v_{row_k} \end{bmatrix} \quad (9)$$

$$\begin{bmatrix} col_{k+1} \\ v_{col_{k+1}} \\ a_{col_{k+1}} \\ row_{k+1} \\ v_{row_{k+1}} \\ a_{row_{k+1}} \end{bmatrix} = \begin{bmatrix} 1 & T & \frac{T^2}{2} & 0 & 0 & 0 \\ 0 & 1 & T & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & \frac{T^2}{2} \\ 0 & 0 & 0 & 0 & 1 & T \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} col_k \\ v_{col_k} \\ a_{col_k} \\ row_k \\ v_{row_k} \\ a_{row_k} \end{bmatrix} \quad (10)$$

Eş. 9 ve Eş. 10'da col_k değeri poz tahmini çıktısındaki merkez noktanın sütun değeridir. col_{k+1} ise hareket modeli kullanılarak $k+1$ anındaki durumun tahmin değeridir. Benzer şekilde row_k , row_{k+1} değerleri de merkez noktanın satır değerlerini ifade eder. v değerleri ise satır ve sütun pikselleri için hızları temsil eder. a değerleri ise satır ve sütun pikselleri için ivmeleri temsil eder.

Kalman filtresinde sistem durumları ile ölçümler arasındaki ilişkiyi tanımlamak için ölçüm modeli kullanılmaktadır. Ölçüm modeli durum vektörünün doğrusal bir fonksiyonudur. Poz tahmini çıktıları ölçümler olarak kullanıldığı için ölçüm modelinde merkez eklem noktasının satır ve sütun değerleri kullanılır. Ölçüm modeli olarak sabit hızlı hareket modeli için Eş. 11, sabit ivmeli hareket modeli için Eş. 12 kullanılır.

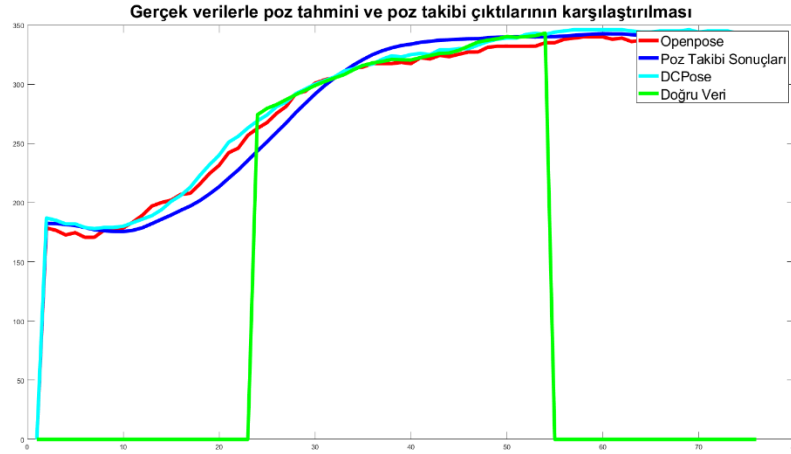
$$\begin{bmatrix} z_{col_k} \\ z_{row_k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} col_k \\ v_{col_k} \\ row_k \\ v_{row_k} \end{bmatrix} \quad (11)$$

$$\begin{bmatrix} z_{col_k} \\ z_{v_{col_k}} \\ z_{row_k} \\ z_{v_{row_k}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} col_k \\ v_{col_k} \\ a_{col_k} \\ row_k \\ v_{row_k} \\ a_{row_k} \end{bmatrix} \quad (12)$$

Poz tahmini çıktılarının gerçek değerlerden farkını belirleyebilmek için PoseTrack 2017 eğitim veri kümesindeki değerlerle DCPose ve OpenPose çıktıları kullanılmıştır. Bu farkların kovaryansı ölçüm gürültüsü olarak kullanılır. Veri kümesindeki bir videodan bir kişinin merkez noktası için poz tahmini değerleri Şekil 7'de gösterilmiştir. Bu tespitlerin gerçek değerden ne kadar saptığı incelenmiştir.

Kişilerin eklem noktalarının hareketleri ivmeli ya da hızlı hareketlerle tam olarak tanımlanamaz. Çünkü videoların içinde insanların hareketleri tamamen değişebilir ya da bir anda sabit olabilir. Kollar aniden dönüş hareketi yaparken gövde sabit gibi görünebilir. Bunun gibi çeşitli durumların da etkisini göz ardı etmemek ve sabit ivmeli hareketleri de yakalayabilmek için süreç gürültüsü matrisi kullanılabilir.

Bir ölçümün yeni bir hedef olup olmadığından emin olmak için geçici bir iz oluşturulmaktadır. Bu geçici iz belirli bir süre boyunca yeterli sayıda ölçüm alırsa onaylanmış bir iz olarak kabul edilir aksi takdirde silinir. M/N kriteri olarak tanımlanan bu çalışma mantığında kişilerin poz tahminlerinin ardışık N adet görüntünün M tanesinde ilişkilendirilmesi durumunda poz takibine başlamak için onaylanması gerekir. Bu şartı sağlayan kişilerin poz tahminleri izlerle ilişkilendirilir.



Şekil 7. Bir kişinin video boyunca merkez noktasının sütun değerlerinin gerçek verilerle poz tahmini ve poz takibi çıktılarının karşılaştırılması (Comparison of the column values of a person's center point throughout the video with actual data pose estimation and pose tracking output)

2.3.2 Veri ilişkilendirme (Data association)

Poz tahmini çıktıları olan eklem noktası ölçümlerini kimliklendirilmiş kişilere yani izlere atayabilmek için bir veri ilişkilendirme yöntemi kullanmak gerekir. Bu çalışmada bu amaçla düşük kaynak tüketimi nedeniyle KEYK algoritması kullanılmıştır. KEYK veri ilişkilendirmesi geçitleme (gating) ve veri atama (data assignment) olmak üzere iki adımdan oluşur. Geçitleme adımında ölçümlerle tahminler arasında mümkün olmayan eşleştirmelerin dikkate alınmaması sağlanır. Kovaryans değerine bağlı olarak bir tahminin konumu etrafında geçit (elipsoid) oluşturulur. Elipsoidin içinde olmayan ölçümler dikkate alınmaz. Ölçümlerle tahminler arasındaki fark Eş. 13'deki gibi hesaplanır [27].

$$\Delta\xi(t) = (z_{k+1} - z_{k+1|k}) \quad (13)$$

Burada $\Delta\xi(t)$ ölçüm ve ölçüm tahmini arasındaki vektörel farktır.

Kovaryans matrisi Eş. 4'deki gibi ve izlerle ölçümler arası d mesafesi Eş. 14'deki gibi hesaplanır.

$$d = \Delta\xi_k^T S^{-1} \Delta\xi_k \quad (14)$$

Oluşturulan elipsoidin sınırları D eşik değeri olarak kabul edilirse $d \leq D$ koşulunun karşılanması doğru bir ölçüm-iz ilişkilendirilmesi yapılabileceğini gösterir. Bu koşul ölçümün tahmin edilen izin hata elipsoidinin içinde olduğunu gösterir. Tek bir ölçüm, tek bir geçidin içine girdiyse doğrudan bir atama yapılır. Fakat bir geçidin içine birden çok ölçüm girebiliyorsa o zaman karmaşık durumlarda da en uygun ilişkilendirmeyi yapar. Veri atama adımında Munkres, Jonker-Volgenant, Auction gibi maliyet hesapları yapan yöntemler kullanılmaktadır. Daha net sonuçlar verdiği için bu çalışmada maliyet hesabı için Munkres yöntemi kullanılmıştır. Munkres yöntemi algılamalar ve izler arasında bir maliyet matrisi hesaplar. Bir iz için en düşük maliyete sahip algılama arasında atama işlemi yapılır [27-30].

Mevcut izlerle ilişkilendirilmeyen ölçümler için yeni izler oluşturulur. Ayrıca izlerin onaylanması ya da

silinmesi adımında ardışık N adet görüntünün M tanesinde izler ilişkilendirilirse onaylanma yapılacağından bu değerlerin belirlenmesi gerekmektedir. Belirtilen şartı sağlayan algılamalar izlerle ilişkilendirilir.

2.4. Performans Ölçütleri (Performance Metrics)

Performans değerlendirmesi yapabilmek için poz tahmini ve poz takibi yöntemlerinden veri kümesinde bulunan etiketler ile aynı formatta çıktılar oluşturmak gerekmektedir. Çok kişili poz takibi için bir görüntü karesindeki bir kişi ele alınırsa bu kişi için bulunması gereken veriler $x1, y1, x2, y2, score, trackID$ ve $annopoints$ 'tir. $x1$, baş bölgesini sınırlayan kutunun sol üst köşesinin piksel sütun değeri, $y1$, baş bölgesini sınırlayan kutunun sol üst köşesinin piksel satır değeri, $x2$, baş bölgesini sınırlayan kutunun sağ alt köşesinin piksel sütun değeri, $y2$, baş bölgesini sınırlayan kutunun sağ alt köşesinin piksel satır değeri, $score$, kişinin güvenilirlik puanıdır. Kişilerin güvenilirlik puanı 0-1 arasında değişmektedir. $annopoints$ değerinin ise poz tahmini ile elde edilen tespit edilmiş eklem noktalarının eklem numarası, sütun piksel değeri, satır piksel değeri ve görünürlüğü bilgilerini içermesi gerekmektedir. Örneğin bir görüntüde bir kişinin sağ ayak bileği görünmemiş ve diğer tüm eklem noktaları görünmüş olabilir. Bu durumda sağ ayak bileğinin görünürlük değerine 0, diğer eklem noktalarının görünürlük değerine 1 verilir.

Çok kişili nesne takibi performansını değerlendirmek için MOTP (Multiple Object Tracking Precision) ve MOTA (Multiple Object Tracking Accuracy) ölçütleri kullanılmaktadır. PoseTrack veri kümesi de farklı yöntemlerin performanslarını kıyaslamak için bu ölçütleri kullanır. MOTP ve MOTA değerleri Eş. 15 ve Eş. 16'deki gibi hesaplanır [31].

$$MOTP = \frac{\sum_{t,t} d_t^i}{\sum_t c_t} \quad (15)$$

$$MOTA = 1 - \frac{\sum_t (m_t + f_{pt} + mme_t)}{\sum_t g_t} \quad (16)$$

Çizelge 2. Ön işleme adımlarının farklı hareket modelleri ile poz takibi performansına etkisi (Effect of preprocessing steps on pose tracking performance with different motion models)

M/N Kriteri	Algoritma	MOTA Baş	MOTA Omuz	MOTA Dirsek	MOTA El Bilek	MOTA Bel	MOTA Diz	MOTA Ayak Bilek	MOTA Toplam	MOTP Toplam	Prec. Toplam	Rec. Toplam
Sabit Hızlı Hareket Modeli												
2/3	OpenPose	-14.6	33.3	13.9	11.3	21.1	16.8	13.0	11.6	56.3	58.9	69.6
	OpenPose ön işlemeli	33.5	46.2	32.6	27.6	38.3	32.6	28.9	34.2	62.5	75.5	52.8
	DCPose	46.7	49.0	21.3	7.7	36.8	19.3	-4.9	26.6	67.6	63.5	79.5
	DCPose ön işlemeli	55.5	56.3	42.6	39.0	52.7	49.0	35.9	47.8	67.2	77.9	72.1
2/5	OpenPose	-15.3	32.4	13.2	10.7	20.3	15.9	12.3	10.9	56.2	58.6	69.7
	OpenPose ön işlemeli	33.4	46.0	32.6	27.6	38.1	32.5	28.7	34.1	62.6	75.7	52.1
	DCPose	46.6	49.0	21.2	7.6	36.7	19.2	-5.0	26.5	67.7	63.4	79.6
	DCPose ön işlemeli	55.5	56.3	42.6	39.0	52.7	49.0	35.9	47.8	67.2	77.9	72.1
Sabit İvmeli Hareket Modeli												
2/3	OpenPose	-5.4	44.8	25.4	21.5	31.9	26.6	21.8	21.8	57.6	63.1	70.1
	OpenPose ön işlemeli	37.6	50.6	36.7	31.4	42.5	36.0	31.6	38.0	63.0	78.1	54.6
	DCPose	46.6	49.0	21.3	7.7	36.8	19.4	-4.7	26.6	67.7	63.4	79.5
	DCPose ön işlemeli	56.1	56.9	43.2	39.5	53.4	49.8	36.4	48.4	67.1	78.0	72.9
2/5	OpenPose	-6.0	44.1	24.8	21.0	31.1	25.9	21.2	21.2	57.6	62.8	70.2
	OpenPose ön işlemeli	37.7	50.7	36.7	31.4	42.7	36.0	31.6	38.1	63.0	78.2	54.6
	DCPose	46.3	48.7	20.9	7.2	36.3	18.9	-5.3	26.1	67.7	63.2	79.6
	DCPose ön işlemeli	56.0	56.9	43.2	39.4	53.4	49.7	36.4	48.4	67.1	78.0	73.1

Eş. 15 tüm çerçeveler üzerindeki eşleşen nesne-hipotez çiftleri için tahmini konumdaki toplam hatadır ve yapılan toplam eşleşme sayısı ile ortalaması alınır. Takipçinin nesne konfigürasyonlarını tanıma, tutarlı yörüngeler tutma vb. becerisinden bağımsız olarak, kesin nesne konumlarını tahmin etme yeteneğini gösterir. Burada c_t , t anı için bulunan eşleşme sayısıdır. Bu eşleşmelerin her biri için o_i nesnesi ve ona karşılık gelen hipotez arasındaki d_t^i mesafesi hesaplanır. Kalan tüm hipotezler yanlış pozitif olarak kabul edilir. Benzer şekilde, kalan tüm nesnelere kayıp olarak kabul edilir. Eş. 16'da m_t , fp_t , ve mme_t t görüntüsü için sırasıyla yanlış pozitiflerin, kayıpların ve uyumsuzluk hatalarının sayısıdır. g_t , t anındaki mevcut nesnelere sayısını ifade etmektedir [31].

3. DENEYSEL SONUÇLAR (EXPERIMENTAL RESULTS)

Poz takibi ile ilgili yapılan çalışmaların başarısını ölçebilmek için gerçek hayat problemlerini yansıtan videoların olduğu bir veri kümesine ihtiyaç vardır. PoseTrack [17] veri kümesi bu ihtiyacı karşılamaktadır ve en güncel çalışmaların sonuçları herkese açık şekilde yayınlanmaktadır. Poz tahmini aşamasında yaygın olarak kullanılan ve gerçek zamanlı çalışabilen OpenPose ile

son verilere göre poz tahmini için PoseTrack veri kümesi üzerinde en yüksek başarıyı elde eden DCPose yöntemleri kullanılarak PoseTrack veri kümesi üzerinde çok kişili poz takibi konusunda deneysel çalışmalar yapılmıştır.

Çizelge 2 PoseTrack 2017 veri kümesi kullanılarak OpenPose ve DCPose poz tahmin algoritmalarından elde edilen verilere poz takibi işlemi öncesinde uygulanan ön işleme adımlarının etkisini farklı M/N değerleri ve sabit hızlı ve sabit ivmeli hareket modelleri için göstermektedir. Çizelgeden de görüleceği gibi izlerin onaylanması için kullanılan M/N oranlarının sonuçlar üzerinde büyük bir etkisi olmamıştır. Fakat toplam MOTA değeri olarak 2/3 M/N kriteri ile daha iyi sonuçlar alınmıştır. Ön işleme adımları uygulanarak toplam çok kişili poz takibi MOTA değerinde sabit hızlı hareket modeli ile OpenPose için %22.6, DCPose için %21.2, sabit ivmeli hareket modeli ile de OpenPose için %16.2, DCPose için %21.8 performans artışı elde edilmiştir. DCPose yönteminin genel olarak OpenPose yöntemine göre daha başarılı olduğu görülmektedir. Fakat DCPose poz tahmini el/ayak bilekleri, dirsek ve diz noktalarında başarısız sonuçlar vermektedir. Görüntüde bu eklem noktaları olmasa bile DCPose bu eklem noktaları için bir değer tahmin etme eğilimindedir. Bu nedenle net olarak



Şekil 8. Örnek birer videoda OpenPose ile videonun (a) 22. ve (b) 40. karelerinde, DCPose ile videonun (c) 53. ve (d) 77. karelerinde, DCPose ile videonun (e) 84. ve (f) 149. karelerinde bulunan kişilerin takip sonuçları (In an example videos video tracking results of the OpenPose at (a) 22nd and (b) 40th frames of the video, DCPose at (c) 53rd and (d) 77th frames of the video, and DCPose at (e) 84th and (f) 149th frames of the video)

belirlenemeyen bilek noktaları için beşinci ön işleme adımı ile tahminde iyileştirme yapılmıştır. Çizelge 2’de bulunan MOTA Baş değerleri incelendiğinde ön işleme adımının bu bölgenin poz takibi performansını sabit hızlı hareket modelinde DCPose için %8.8, OpenPose için %48.1, sabit ivmeli hareket modelinde ise DCPose için %9.5, OpenPose için %43.0 artırdığı görülmektedir. Benzer şekilde diğer eklem noktalarındaki iyileşmeler de toplam MOTA değerindeki artış ile birlikte incelendiğinde birbirini kanıtlar niteliktedir.

Buizza ve diğerleri [21] de OpenPose kullanarak PoseTrack veri kümesi üzerinde MOTA hesaplamalarını yapmıştır fakat OpenPose kişilerin baş bölgesini doğrudan tespit etmemektedir. PoseTrack veri kümesi ile hesap yapılabilmesi için Bölüm 2.2’deki gibi baş bölgesinin hesaplanması hakkında bilgi verilmelidir. Çünkü bu durum toplam MOTA sonuçlarını etkilemektedir. Bu nedenle elde edilen sonuçlar doğrudan karşılaştırılabilir değildir. Bu çalışma, farklı

filtrelerin ve farklı hareket modellerinin poz takibine etkisini karşılaştırmaya değil ön işleme adımlarının performans üzerindeki etkilerini incelemeye odaklanmıştır. Buizza ve arkadaşları, sabit hızlı hareket modeli ve doğrusal Kalman filtresi ile toplam MOTA değerini 10.6 olarak hesaplamıştır. Fakat kullandıkları veri ilişkilendirme yöntemi bu çalışmadan farklıdır.

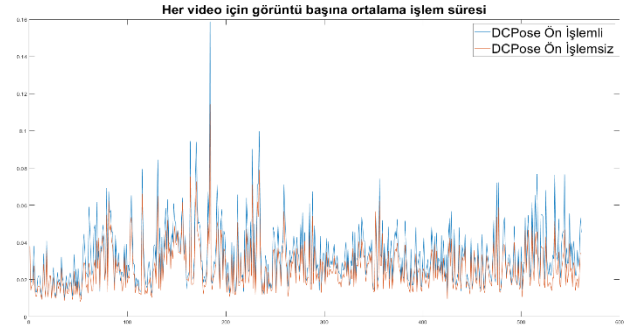
Şekil 8, peş peşe video karelerinde kişilerin takip edildiğini gösteren bir örnektir. Kişilere birer numara verilmiş ve takip eden karelerde bu numaralar korunarak kişilerin takibi yapılmıştır.

Şekil 9’da ön işleme adımlarının DCPose için işlem süresine etkisi görülmektedir. Mavi değerler ön işleme adımının ortalama sürelerini gösterir. Her video için poz takibi süresi hesaplanmıştır. Bu süre videodaki görüntü sayısına bölünerek bir görüntü karesi için ortalama takip işlem süresi bulunmuştur. Şekilde ön işleme adımının poz takibinde ne kadar işlem gecikmesi getirdiği işlem süreleriyle birlikte verilmiştir. Tüm görüntülerin poz

takibi için birim ortalama işlem süreleri hesaplandığında ön işleme adımı olmadan 26.1 milisaniye, ön işleme adımıyla birlikte 32.4 milisaniye gerekmektedir. Ön işlemler çalışma süresini 5.7 milisaniye yavaşlatmaktadır. Performansta %21.2 artış sağladığı için çalışma süresinden 5.7 milisaniyelik ödün vererek ön işleme adımlarının kullanılabilceği düşünülmektedir.

4. SONUÇLAR VE TARTIŞMA (RESULTS AND DISCUSSION)

Bu çalışmada gerçek zamanda çalışabilen çok kişili bir poz takibi yöntemi kullanılarak farklı poz tahmini yöntemlerinin başarımını artıracak çeşitli ön işleme adımları geliştirilmiştir. Çalışma sonuçları 550 farklı videodan oluşan PoseTrack 2017 veri kümesi ve bu veri kümesinin herkese açık olarak paylaştığı sonuç hesaplama yöntemi ile gerçek değeri bilinen 250 video için değerlendirilmiştir. OpenPose ve DCPose yöntemleri kullanılarak bu veri kümesi üzerinde kişilerin eklem noktaları tespit edilmiştir. Bu eklem noktaları Kalman filtresi ve KEYK yöntemi ile geliştirilen poz takibi modülünde kullanılmıştır. Poz takibi öncesinde yapılan ön işleme adımlarının sonuçlara etkisi incelenmiştir ve işlem süresi analizi yapılmıştır. Bu analizlerin sonuçlarına göre DCPose ile yapılan poz takibi sonuçları PoseTrack veri kümesi üzerinde daha başarılıdır. Ön işleme adımları çok kişili poz takibi toplam MOTA sonuçlarını sabit hızlı hareket modeli kullanılarak OpenPose poz tahmini yöntemi için %22.6, DCPose için %21.2 artırırken sabit ivmeli hareket modeli ile OpenPose için %16.2, DCPose için %21.8 artırmıştır. Poz takibinden önce yapılan işlemlerin her bir videoyu işlem süresi bakımından ne kadar etkilediği hesaplanmıştır. Elde edilen bu değerlerle ortalama bir video için ön işlemler ve ön işlemsiz poz takibi işlem süreleri karşılaştırıldığında tüm videolarda görüntü başına ortalama 5.7 milisaniye ilave işlem süresi gözlemlenmiştir. Buizza ve diğerleri sabit hızlı hareket modeli ile OpenPose yöntemini kullanarak toplam MOTA değerini 10.6 olarak elde etmiştir [21]. Bu çalışmada herhangi bir ön işlem adımı olmadan 11.6 toplam MOTA değeri elde edilirken geliştirilen ön işlem adımları MOTA değerini önemli derecede artırarak 34.2'ye yükseltmiştir. Çalışmada kullanılan veri tabanının resmi sayfası olan <https://posetrack.net/leaderboard.php> adresinden çok kişili poz takibi ile ilgili sonuçlar incelendiğinde ise DCPose ile sabit ivmeli hareket modeli kullanılarak elde edilen toplam MOTA değerinin PoseTrack veri kümesini yayımlayan Andriluka ve diğerlerinin sonuçlarından [17] daha başarılı olduğu görülmektedir. Bu çalışmada elde edilen sonuçlar listeye 23. sıradan girebilmektedir. Gelecek çalışmalarda ön işleme adımlarına ek olarak farklı poz takibi sonuçlarının birleştirilmesi ile daha başarılı sonuçlar elde edilebileceği ön görülmektedir. Ayrıca paralel poz takipçilerini birleştirerek yüksek doğruluğa sahip düşük kaynak tüketimi olan bir poz takibi yöntemi geliştirilmesi de hedeflenmektedir.



Şekil 9. Ön işleme adımlarının her bir videonun poz takip süresinin videodaki görüntü sayısına bölünmesi ile elde edilen ortalama işlem sürelerine etkisi (The effect of the preprocessing steps on the average processing times obtained by dividing the exposure tracking time of each video by the number of images in the video)

TEŞEKKÜR (ACKNOWLEDGEMENT)

Bu çalışma, Gazi Üniversitesi Bilimsel Araştırma Projeleri birimi tarafından FGA-2021-7113 nolu proje ile desteklenmiştir. Desteklerinden dolayı teşekkür ederiz.

ETİK STANDARTLARIN BEYANI (DECLARATION OF ETHICAL STANDARDS)

Bu makalenin yazar(lar)ı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

YAZARLARIN KATKILARI (AUTHORS' CONTRIBUTIONS)

Ahmet Samet HALICI: Deneyleri yapmış, deney sonuçlarını analiz etmiş ve makalenin yazım işlemini gerçekleştirmiştir.

Ayşe DEMİRHAN: Deney sonuçlarını analiz etmiş ve makalenin yazım işlemini gerçekleştirmiştir.

ÇIKAR ÇATIŞMASI (CONFLICT OF INTEREST)

Bu çalışmada herhangi bir çıkar çatışması yoktur.

KAYNAKLAR (REFERENCES)

- [1] Cote M., Jean F., Albu A.B., Capson D., "Video summarization for remote invigilation of online exams", *IEEE Winter Conference on Applications of Computer Vision*, Lake Placid, NY, 1-9, (2016).
- [2] Yan S., Xiong Y., Lin D., "Spatial temporal graph convolutional networks for skeleton-based action recognition", *AAAI Conference on Artificial Intelligence*, New Orleans, Louisiana, 7444-7452, (2018).
- [3] Üzen H., Hanbay K. "Yaya özellik tanıma için LM filtre temelli derin evrişimsel sinir ağı", *Politeknik Dergisi*, 23(3): 605-613, (2020).

- [4] Çalışan M., Talu M.F. "Comparison of methods for determining activity from physical movements", *Politeknik Dergisi*, 24(1): 17-23, (2021).
- [5] Snower M., Kadav A., Lai F., Graf H.P., "15 keypoints is all you need", *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Online, 6738-6748, (2020).
- [6] Liu Z., Chen H., Feng R., Wu S., Ji S., Yang B., Wang X., "Deep Dual Consecutive Network for Human Pose Estimation", *IEEE Conference on Computer Vision and Pattern Recognition*, Online, 525-534, (2021).
- [7] Wang M., Tighe J., Modolo D., "Combining detection and tracking for human pose estimation in videos", *IEEE Conference on Computer Vision and Pattern Recognition*, Online, 11088-11096, (2020).
- [8] Sun K., Xiao B., Liu D., Wang J., "Deep high-resolution representation learning for human pose estimation", *IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 5693-5703, (2019).
- [9] Bulat A., Tzimiropoulos, G., "Human pose estimation via convolutional part heatmap regression", *European Conference on Computer Vision*, Cham, Amsterdam, 717-732, (2016).
- [10] Cao Z., Simon T., Wei S.E., Sheikh Y., "Realtime multi-person 2D pose estimation using part affinity fields", *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, 7291-7299, (2017).
- [11] Kuhn H.W., "The Hungarian method for the assignment problem", *Naval Research Logistics Quarterly*, 2(1-2): 83-97, (1955).
- [12] Ladicky L., Torr P.H., Zisserman A., "Human pose estimation using a joint pixel-wise and part-wise formulation", *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, 3578-3585, (2013).
- [13] Jin S., Liu W., Ouyang W., Qian C., "Multi-person articulated tracking with spatial and temporal embeddings", *IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 5664-5673, (2019).
- [14] Raaj Y., Idrees H., Hidalgo G., Sheikh Y., "Efficient online multi-person 2D pose tracking with recurrent spatio-temporal affinity fields", *IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 4620-4628, (2019).
- [15] Girdhar R., Gkioxari G., Torresani L., Paluri M., Tran D., "Detect-and-track: Efficient pose estimation in videos", *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, 350-359, (2018).
- [16] Xiao B., Wu H., Wei Y., "Simple baselines for human pose estimation and tracking", *European Conference on Computer Vision*, Munich, Germany, 466-481, (2018).
- [17] Andriluka M., Iqbal U., Insafutdinov E., Pishchulin L., Milan A., Gall J., Schiele B., "PoseTrack: A benchmark for human pose estimation and tracking", *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, 5167-5176, (2018).
- [18] Vondrick C., Patterson D., Ramanan D., "Efficiently scaling up crowdsourced video annotation", *International Journal of Computer Vision*, 101(1): 184-204, (2013).
- [19] Çimen M.E., Boyraz Ö.F., Garip Z., Pehlivan İ., Yıldız M.Z., Boz A.F. "Görüntü işleme tabanlı kutu sayma yöntemi ile fraktal boyut hesabı için arayüz tasarımı", *Politeknik Dergisi*, 24(3): 867-878, (2021).
- [20] Hasegawa I., Uenishi K., Fukunaga T., Kimura R., Osawa M., "Stature estimation formulae from radiographically determined limb bone length in a modern Japanese population", *Legal Medicine*, 11(6): 260-266, (2009).
- [21] Buizza C., Fischer T., Demiris Y. "Real-time multi-person pose tracking using data assimilation", *IEEE/CVF Winter Conference on Applications of Computer Vision*, Online, 449-458, (2020).
- [22] Moon S., Park Y., Ko D.W., Suh I.H., "Multiple kinect sensor fusion for human skeleton tracking using Kalman filtering", *International Journal of Advanced Robotic Systems*, 13(2): 65, (2016).
- [23] Kalman R.E., "A new approach to linear filtering and prediction problems", *Journal of Basic Engineering*, 82(1): 35-45, (1960).
- [24] Brown R.G., "Introduction to random signal analysis and Kalman filtering", *John Wiley & Sons Inc.*, New York, (1983).
- [25] Bostanci E., Bostanci B., Kanwal N., Clark A.F., "Sensor fusion of camera, GPS and IMU using fuzzy adaptive multiple motion models", *Soft Computing*, 22(8): 2619-2632, (2018).
- [26] Unal M., Bostanci E., Guzel M.S., Unal F.Z., Kanwal N. "Evolutionary motion model transitions for tracking unmanned air vehicles", *New Trends in Computational Vision and Bio-inspired Computing*, 1193-1200, Springer, Cham, (2020).
- [27] Wang H., Zhang X., "Real-time vehicle detection and tracking using 3D LiDAR", *Asian Journal of Control*, 1-11, (2021).
- [28] Munkres J., "Algorithms for the assignment and transportation problems", *Journal of the Society for Industrial and Applied Mathematics*, 5(1): 32-38, (1957).
- [29] Güllü M., Polat H. "Text authorship identification based on ensemble learning and genetic algorithm combination in Turkish text", *Politeknik Dergisi*, 1-1, (2021).
- [30] Karasu S., Saraç Z. "Güç kalitesi bozulmalarının Hilbert-Huang dönüşümü, genetik algoritma ve yapay zeka/makine öğrenmesi yöntemleri ile sınıflandırılması", *Politeknik Dergisi*, 23(4): 1219-1229, (2020).
- [31] Bernardin K., Stiefelhagen R., "Evaluating multiple object tracking performance: The clear MOT metrics", *EURASIP Journal on Image and Video Processing*, 1-10, (2008).