

## Güçlü Temsil Yöntemleri ile Kişi Tanıma ve Takibi için Hibrit Bir Yaklaşım

\*Makale Bilgisi / Article Info

Alındı/Received: 09.11.2023

Kabul/Accepted: 31.07.2024

Yayımlandı/Published:02.12.2024

### A Hybrid Approach to Person Recognition and Tracking with Powerful Representation Methods

Gizem ORTAÇ KOŞUN<sup>1\*</sup>, Seçkin YILMAZ<sup>2</sup>, Rüya ŞAMLI<sup>1</sup>

<sup>1</sup> İstanbul Üniversitesi-Cerrahpaşa, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Yazılım Mühendisliği Ana Bilim Dalı, İstanbul, Türkiye

<sup>2</sup> Bursa Teknik Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği Bölümü, Bursa, Türkiye

© Afyon Kocatepe Üniversitesi

#### Öz

Halka açık ve kalabalık alanlarda yapılan gözetimlerde, otomatik bir takip sistemi olmadan manuel olarak gerçekleştirilen izlemelerde kişilerin takibi zor bir görevdir. Önerilen çalışma, bu sorunu çözmek ve literatüre katkı sağlamak amacıyla geliştirilmiştir. Hibrit sistem, gözetim sistemlerinde kişilerin tanımlanması ve takibi için tasarlanmıştır. Kamera sistemlerinin kaydettiği videolarda takip edilmek istenen kişi, Mask R-CNN yöntemi kullanılarak segmente edilir ve farklı tekniklerle çıkarılan özellikleri birleştirilerek daha güçlü bir temsil vektörü oluşturulur. Bu temsil vektörü, kişilerin videolarda otomatik bir sistemle aranmasını ve daha etkili bir şekilde takip edilmesini kolaylaştırır. Çalışmada, renk histogramları, Gabor filtreleri, yönlendirilmiş gradyan histogramı ve VGG16 mimarisinden oluşan özellik çıkarım teknikleri kullanılarak güçlü bir temsil oluşturulmaktadır. Bu temsil, kişilerin daha iyi tanımlanmasını sağlayarak videolarda daha hızlı ve etkin bir arama performansı sunar. Çalışma, güvenlik personelinin yüz tanıma zorluğu yaşadığı durumlarda olağan şüpheli kişilerin tanımlanması ve takibini kolaylaştırmak için önemli bir altyapı sağlar. Sorgulanan kişinin her bir video karesinde tespit edilip işaretlenmesini sağlayarak gerçek dünyada kullanılabilir bir uygulama örneği sunar. Elde edilen sonuçlar, kesinlik, duyarlılık ve F skoru metrikleri kullanılarak yorumlanmıştır.

**Anahtar Kelimeler:** Kişi tanıma, Kişi takibi, Gözetim videoları, Kişi yeniden tanımlama

#### Abstract

Tracking people is a difficult task in surveillance carried out in public and crowded areas and in manual monitoring without an automatic tracking system. The proposed study was developed to solve this problem and contribute to the literature. The hybrid system is designed for identification and tracking of individuals in surveillance systems. The person who is wanted to be followed in the videos recorded by the camera systems is segmented using the Mask R-CNN method and a stronger representation vector is created by combining the features extracted with different techniques. This representation vector makes it easier to search for people in videos and track them more effectively with an automated system. In the study, a powerful representation is created using feature extraction techniques consisting of color histograms, Gabor filters, directed gradient histogram and VGG16 architecture. This representation enables better identification of people, providing faster and more effective search performance in videos. The study provides an important infrastructure to facilitate the identification and tracking of usual suspects in situations where security personnel have difficulty in facial recognition. It provides an application example that can be used in the real world by ensuring that the interrogated person is detected and marked in each video frame. The results obtained were interpreted using precision, sensitivity and F score metrics.

**Keywords:** Person identification, Person tracking, Surveillance videos, Person reidentification

#### 1. Giriş

Günümüzde halkın güvenliği ve emniyeti konularının önemi her geçen gün artmaktadır. Özellikle havaalanları ve alışveriş merkezleri gibi kalabalık ve halka açık alanlarda, güvenlik endişeleri giderek artmaktadır. Bu nedenle, halkın güvenliğini sağlamak için önemli miktarda harcama yapılmakta ve çaba gösterilmektedir. Video gözetim sistemleri ise bu alanda kilit bir rol oynamaktadır. Adli soruşturmalar, suç tespiti, suç önleme ve kısıtlı

alanların korunması gibi çeşitli güvenlik konularını ele almak için günümüzde yaygın olarak kullanılan gözetim sistemleri, etkili bir güvenlik aracı olarak öne çıkmaktadır (Xiao vd. 2019).

Ülkemizde, kapalı alanların gözetimi için ağ kameralarıyla yapılan sürekli video kayıtları, olay veya anormallik olduğunda gözetim operatörleri tarafından manuel olarak izlenmektedir. Ancak, dağıtık kamera ağlarında insan operatörlerin becerilerine dayanan bir sistem kullanmak

maliyetli ve güvenilir değildir. Operatörlere genellikle aynı anda izleyebileceklerinden daha fazla kamera atanır ve hatta tek bir kamerada bile manuel izleme, dikkat eksikliklerine karşı savunmasız kalmaktadır. Ayrıca, operatörlerin bireysel deneyimi, odaklanma ve dikkat becerisi insan performansını belirleyen en önemli kriterlerdendir. Kamusal alan kamera ağları son yıllarda hızla büyüdüğü için, manuel izleme ve takip işlemlerinin ölçeklenebilir olmadığı giderek daha açık hale gelmektedir (Roshan vd. 2020).

Dağıtılmış çok kameralı bir gözetim sistemi için temel görevlerden biri insanları farklı konum ve zamanlardaki kamera görüntüleriyle ilişkilendirmektir. Kişi Yeniden Tanımlama (KYT, Person Re-identification - ReID) olarak bilinen bu durum; günlük hayatta yetkili/sınırlı erişim, sınır kontrolü, suç tespiti veya önlenmesi, adli arama gibi önemli derecede ihtiyaç duyulan ve rutin olarak gerçekleştirilen güvenlik görevlerindedir. Resmi olarak, farklı zamanlarda farklı kameralardan alınan bir kişinin görüntüleri arasındaki bağlantının kurulması olarak da tanımlanabilir (Chung vd. 2017). Geniş bir uygulama yelpazesine ve büyük ticari değere sahip olduğu için bilgisayarlı görü topluluğu içinde, otomatikleştirilmiş KYT çözümleri geliştirmeye artan bir ilgi vardır. Öte yandan, pek çok uygulamada bir kişiyi gerçekten benzersiz bir şekilde tanımlamak gerekli değildir. Aynı kişinin başka görüntülerde veya görüntü dizilerinde önceki veya ilerideki oluşumlarını belirlemek yeterlidir (Bäumel ve Stiefelhagen 2011).

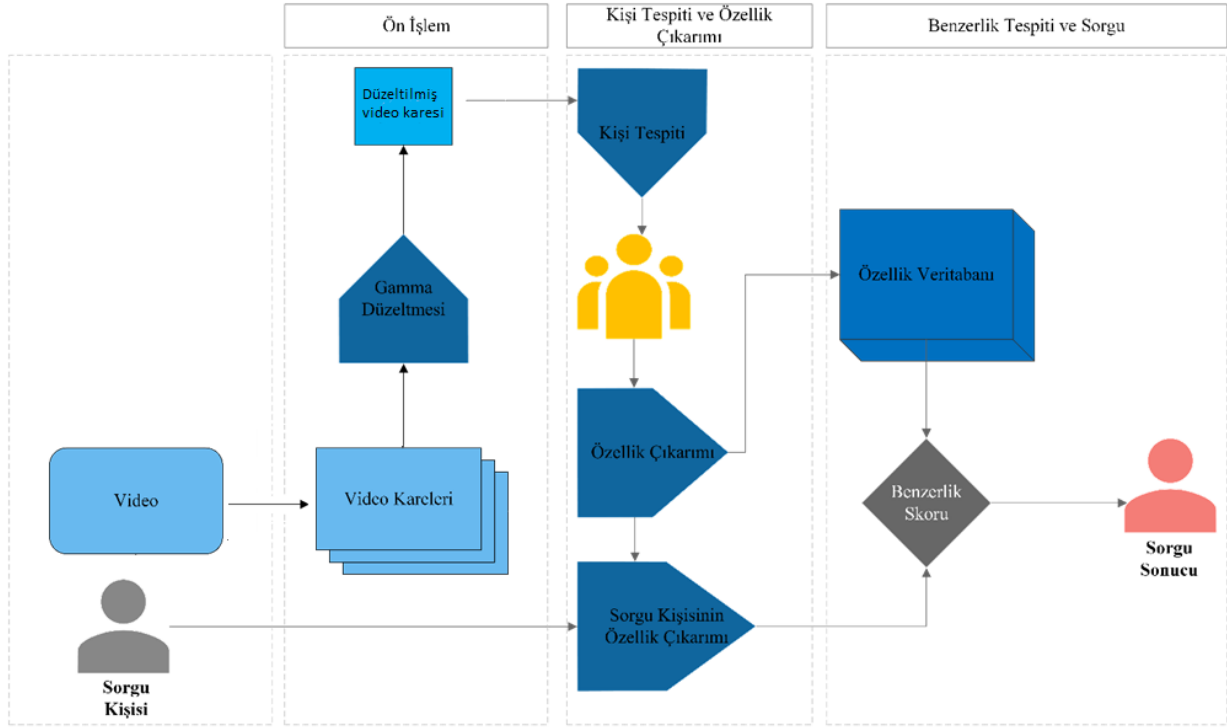
KYT için kişilerin benzersiz bir tanımlaması gerekmediğinden, genellikle güvenilir olmayan biyometrik özelliklerin kullanılması uygun değildir. Yüz ve yürüyüş gibi biyometrik özellikler, düşük çözünürlüklü kameralar nedeniyle video gözetim sistemlerinde kullanıma genellikle uygun değildir. Üniforma veya benzeri giysiler, yüz tanıma yapılamadığında veya kişiler için kullanılacak benzersiz bir özellik olmadığında da zorluğu artırabilir (Leng ve Tian 2019). Bu nedenle önerilen çalışmada bir kişinin tüm vücut görünümünün belirli bir zaman dilimi içinde değişmediği ve KYT için oldukça uygun olduğu gerçeği kullanılacaktır (Bäumel ve Stiefelhagen 2011).

KYT için literatürde kullanılan pek çok yöntem mevcuttur. Fakat bu yöntemler temel olarak "el yapımı" ve derin öğrenme sistemleri olarak ikiye ayrılmaktadır. El yapımı yöntemler histogramla renk özellikleri çıkarma ve Yerel İkili Model (Local Binary Pattern-LBP) ile doku çıkarma gibi düşük seviyeli özellikleri çıkarmak için kullanılan geleneksel yöntemlerdir. Renk histogramları, renk, desen, yürüyüş analizi ve basit mesafe ölçümleri faydalı olsa da düşük seviyeli öznelikler kullanmakta ve gerçek

uygulamalarda doğruluk ve güvenilirlik açısından yetersiz kalabilmektedir. Bu öznelikler ve metrikler, büyük sınıf içi ve sınıflar arası varyasyonlar durumunda etkili değildir. Öte yandan düşük seviyeli görsel özelliklerle karşılaştırıldığında uzun saç, mavi gömlek gibi tanımlayıcı insan özellikleri orta seviye tanımlamaları temsil etmektedir. Orta seviye tanımlamalar aynı kişi için daha tutarlıdır ve kişinin duruşu, açısı vb. çeşitliliklerine karşı daha dayanıklıdır. Fakat insan özelliklerini manuel olarak açıklamak maliyetli olduğundan, geniş bir özellik kümesi için yeterli eğitim verisi elde etmek zordur. Sonuç olarak, düşük seviyeli görsel özellikler hala kilit bir rol oynamaktadır ve karakteristik nitelikler çoğunlukla yardımcı özellikler olarak kullanılmaktadır (Su vd. 2016).

Son yıllarda ise Derin Öğrenme (Deep Learning-DL) uygulamalarının el yapımı özelliklerin bahsedilen eksikliklerinin üstesinden gelebildiği gözlemlenmiştir. Evrimsel Sinir Ağları (Convolution Neural Network - CNN) 2014 yılında KYT alanına sunulduğunda, özellik çıkarma yöntemleri tamamen değişmiştir. Yi vd. (2014) ve Wu vd. (2018) ise eğitim verilerinden otomatik olarak kişi görüntüsünün en uygun özelliklerini öğrenmek için bir Siamese sinir ağı kullanmıştır. Birçok yeni KYT çalışması, daha doğru sonuçlar elde etmek için derin öğrenmeye dayalı yöntemleri benimsemiş olsa bile derin özelliklerin avantajları kadar dezavantajları da vardır. Derin öğrenme modelleri, eğitim ve çıkarım aşamalarında genellikle yüksek hesaplama gücü gerektirir. Bunun yanında karmaşık ve yüksek boyutlu özellik vektörleri ürettiği için yorumlaması genellikle zordur. Ayrıca derin öğrenme modelleri eğitildikleri veri setlerine çok bağımlıdırlar. KYT için açıklama verilerinin de elde edilmesi yorucu ve maliyetli olduğundan, önceden eğitilmiş modeller genellikle herhangi bir uyarılma mekanizması olmadan yeni test ortamlarında kullanılmaktadır. Bu durum önemli performans düşüşlerine sebep olmaktadır (Almasawa vd. 2019).

El yapımı yöntemlerle derin öğrenme yöntemlerini birleştiren birkaç çalışmada ise renk, LBP ve Yerel Maksimal Oluşum (Latent Ordinal Model- LOMO) CNN ve Uzun Kısa Süreli Bellek (Long Short-Term Memory- LSTM) ile kullanılmıştır (Almasawa vd. 2019). Luo vd (2019) çalışmasında ise derin öğrenme tabanlı kişi yeniden tanıma sistemlerinde düşük seviyeli özneliklerin kullanımını ele alınmaktadır. Renk histogramları, HOG, LBP gibi geleneksel düşük seviyeli özneliklerin derin öğrenme modelleriyle birleştirilerek kullanılmasının faydalarını araştıran bu yaklaşım, derin öğrenme modellerinin performansını artırabileceğini ve KYT görevinde daha güçlü bir temel oluşturabileceğini göstermektedir.



Şekil 1. Önerilen algoritmanın akış şeması

Bu çalışmada da her iki yöntemin avantajlarından faydalanarak hem yüksek verimliliğe hem de genelleme kabiliyetine sahip ölçeklenebilir bir kişi KYT sistemi tasarlanması amaçlanmıştır. Ayrıca bu çalışma, düşük seviyeli öznelıkların derin öğrenme tabanlı sistemlerle entegrasyonunun, kişi yeniden tanıma performansını artırmak için önemli bir strateji olabileceğini vurgulamaktadır.

Bu kapsamda önerilen çalışmada kişi takibi ve çıkarımı konusunda farklı kişiler arasındaki ince farkları yeterince ayırt edici hale getirmek için özellik tanıma bağlamında “el yapımı” ve derin özellikler birleştirilerek, özellik tanımanın doğasında bulunan büyük sınıf içi varyasyona karşı daha dayanıklı olan birleştirilmiş bir temsil oluşturmak amaçlanmıştır. Önerilen yaklaşım derin modeller tarafından öğrenilen yüksek seviyeli veya bakış noktası değişikliklerinden kaynaklanan uzamsal değişikliklere karşı sağlamdır ve kişi özelliklerinin farklı görünüşlerini modelleyebilen zengin özellik temsillerini içermektedir.

## 2. Materyal ve Metot

Önerilen algoritmanın ana hatları Şekil 1’de gösterilmiştir. İlk aşamada gözetleme sistemlerinden alınan video ön işlem adımı olarak karelerine ayrılmaktadır. Elde edilen karelerde gama düzeltme (gamma correction) işlemi uygulanmıştır. Gama düzeltme, video veya dijital görüntülerde parlaklığı kodlamak ve çözmek için doğrusal olmayan bir işlemdir. Bilgisayar (ve diğer dijital ekran) ekranlarında parlaklığı düzgün şekilde görüntülemek için kullanılır. Bir kameranın içeriği yakalama şekli, bir ekranın

içeriği görüntüleme şekli ve görsel sistemin ışığı işleme şekli arasındaki farkları düzenlemeye yardımcı olmaktadır (Tharsanee vd. 2021).

### 2.1 Veri seti

Önerilen çalışmada farklı veri setleri kullanarak performans analizi yapılmıştır. Çalışmanın performansını değerlendirmek için diğer modern yeniden kimlik veri kümeleriyle karşılaştırıldığında nispeten küçük bir veri kümesi olan PKU-Reid test edilmiştir. Bu veri seti iki ayrı kamera görüntüsünden çekilen 1824 görüntü dahil 114 kişiyi içermektedir. Her kişi için, bir kamera görüntüsü altında sekiz farklı yönden sekiz görüntü yakalanmış ve 128 × 48 piksele normalleştirilmiştir (Ma vd. 2016). Ek olarak ücretsiz ve anonim bir stok video sitesinden (www.pexels.com) elde edilen ve kalabalık bir bölgede yürüyen kişileri gösteren bir havaalanı videosu üzerinde test edilmiştir. Kullanılan video toplamda 285 adet kareden oluşmaktadır.

### 2.2 Kişi tespiti ve segmentasyonu

Şekil 1’de gösterildiği gibi ön işlem aşamasından sonra kişi tespiti ve bölütleme işlemi uygulanmaktadır. Bunun için her bir video karesinde kişilerin tespiti, konumlarının belirlenmesi ve piksel bazında maskelerinin çıkarılması için Mask R-CNN algoritması kullanılmıştır. Mask R-CNN, nesne tespitinde öneri ve sınıflandırma olmak üzere iki aşamadan oluşmaktadır. Bu ana aşamalar Şekil 2’de gösterilmiştir. Mask R-CNN mimarisinde (Singh ve Shekhar 2018) giriş olarak bir görüntü verildiğinde ilk önce bir CNN kullanılarak bir özellik haritası çıkarılır.



görüntü bozulmasına ve ölçeklemeye karşı dayanıklıdır. Gözetleme videosunda kişinin yeniden tanımlanması için vücut özelliklerine küresel bir renk yaklaşımı seçilmiştir. Kişinin renk bilgisinin çıkarılması, yöntemin açık ve basit olmasını sağlar. Renkli histogram tabanlı görüntü almanın uygulanması kolaydır ve tanıma sistemlerinde iyi çalışılmış ve yaygın olarak kullanılmıştır (Zhang vd. 2017).

#### 2.4 Gabor filtresi ile doku özelliklerinin çıkarılması

Kullanılan ikinci özellik olan doku, bir görüntünün en önemli özelliklerinden biridir. Doku özellikleri de tanıma sistemlerinde yaygın olarak kullanılmaktadır. Bu çalışmada doku özelliklerinin çıkarımında Gabor filtreleme kullanılmıştır. Gabor filtrelerinin kullanımı, doku özelliklerinin çıkarılması için sinyal işleme en popüler yaklaşımlardan biridir. Bunlar, uzamsal ve frekans alanında filtrelemeye izin verir.

Doku analizi için gri düzey birlikte oluşum matrisleri, Tamura doku özelliği, Markov rastgele alan modeli, Gabor filtreleme ve yerel ikili modeller gibi çeşitli algoritmalar tasarlanmıştır. Uygulamada, doku özellikleri, ayırım gücünü artırmak için renk özellikleriyle birleştirilebilir, böylece iyileştirilmiş çıkarım performansı elde edilir.

Gabor filtreleri, memeli görsel kortikal hücrelerine çok benzer şekilde hareket eder ve farklı yönlerde ve farklı ölçeklerde özellikleri çıkarır. Farklı ölçeklerde ve yönlerde bir filtre bankası, bir görüntünün çok kanallı filtrelenmesine ve frekans ve yön bilgilerinin çıkarılmasına izin verir (Zekri vd. 2017). Gabor filtre ailesinin genel ifadesi Eşitlik 1,2 ve 3 ile ifade edilmiştir:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(\frac{-x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \quad (1)$$

$$x' = x \cos \theta + y \sin \theta \quad (2)$$

$$y' = -x \sin \theta + y \cos \theta \quad (3)$$

$\lambda$  kosinüsün dalga boyu faktörünü;  $\theta$ , Gabor fonksiyonunun yönünü kontrol eder. Sıfır derece theta değeri, Gabor fonksiyonunun dikey konumuna karşılık gelir.  $\phi$ , fazın ofset değerini ve  $\gamma$  ise uzamsal görüş açısını belirtir.  $\sigma$  değeri ise Gabor filtresinde kullanılan Gauss fonksiyonunun standart sapmasıdır (Li vd. 2010). Bu parametreleri değiştirerek frekans alanını neredeyse tamamen kapsayan bir filtre bankası elde edilir. Çok

kanallı filtre doku analizi alanında etkili bir yöntemdir. Filtre bankaları, bir görüntüyü, görüntünün her pikselini bir arazi kullanımı türünde sınıflandırmak için kullanılabilen doku özelliklerine ayırma özelliğine sahiptir. Çok kanallı filtreleme işlemi, retina üzerindeki görüntüyü, her birinin sınırlı bir frekans ve yönelim aralığında yoğunluk farklılıklarına sahip olan çeşitli filtrelenmiş görüntülere ayırıştıran insan görsel sisteminin özelliklerini taklit etmektedir. Bir Gabor filtre bankası, farklı radyal frekanslar ve yönelimlere sahip frekans alanını kapsayan bir grup Gauss filtresinden oluşmaktadır. Bu nedenle dokular filtrelere verdikleri yanıtlarla karakterize edilir ve bu filtrelerin her biri belirli bir frekans ve yönelim için tasarlanmıştır. Her filtreyi kullanarak  $I(x, y)$  görüntüsü üzerindeki Gabor dönüşümü Eşitlik 4'teki gibi tanımlanmıştır (Allison ve Wintz 2009):

$$\omega_{mn}(x, y) = \int (I(x_1, y_1) g_{mn} * (x - x_1, y - y_1) dx_1 dy_1) \quad (4)$$

Daha sonra  $m \times n \times 2$  boyutunda bir histogram oluşturulur ve her bir bölme ortalamayı ( $\mu_{mn}$ ) ve standart sapmayı ( $\sigma_{mn}$ ) temsil eder:

$$\mu_{mn} = \iint |\omega_{mn}(xy)| dx dy \quad (5)$$

$$\sigma_{mn} = \sqrt{\iint (|\omega_{mn}(x, y)| - \mu_{mn})^2 dx dy} \quad (6)$$

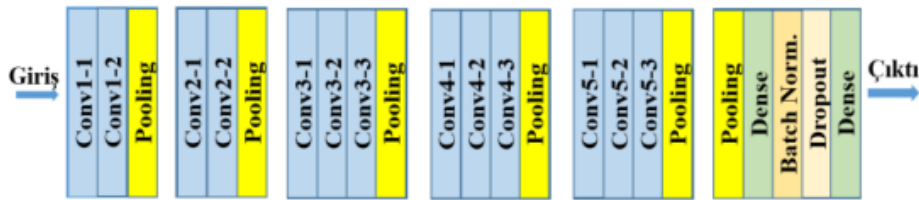
#### 2.5 Şekil özelliklerinin çıkarımı

Görüntülerde temsil edilen nesnelerin şekli, tanıma görevlerinde kullanılan en önemli özelliklerden biridir. Bu durum nesnelere tanımak için şeklin algısal olarak önemli olmasından kaynaklanmaktadır. Bazı durumlarda şekil, temsil edilen nesne hakkında renk, doku veya diğer özelliklerden daha fazla içsel bilgi içerir. Geometrik bir bakış açısıyla, biçim resmi olmayan bir şekilde, bir nesnenin bir görüntüdeki temsilinden ölçek, öteleme ve döndürme gibi benzer dönüşümler nedeniyle renk, doku ve efektlerin çıkarılmasının sonucu olarak tanımlanabilir. Yönlendirilmiş gradyan histogramı (Histogram of Gradient - HOG), nesne algılama amacıyla bilgisayarla görme ve görüntü işleme kullanılan bir özellik tanımlayıcısıdır. Teknik, bir görüntünün yerleştirilmiş bölümlerinde gradyan oryantasyon oluşumlarını sayar. Bu yöntem, kenar oryantasyonu histogramlarına, ölçekle değişmeyen özellik dönüşümü tanımlayıcılarına ve şekil bağlamlarına

benzer, ancak düzgün aralıklı hücrelerden oluşan yoğun bir ızgara üzerinde hesaplanması ve gelişmiş doğruluk için örtüşen yerel kontrast normalleştirilmesi kullanması bakımından farklılık gösterir. HOG tanımlayıcısının diğer tanımlayıcılara göre birkaç önemli avantajı vardır. Yerel hücreler üzerinde çalıştığı için, nesne yönelimi dışında geometrik ve fotometrik dönüşümlere değişmez. Bu tür değişiklikler yalnızca daha büyük uzamsal bölgelerde ortaya çıkacaktır. Ayrıca, kaba uzamsal örnekleme, hassas yönelim örnekleme ve güçlü yerel fotometrik normalizasyon, kabaca dik bir konumu korudukları sürece yayaların bireysel vücut hareketlerinin göz ardı edilmesine izin verir. HOG tanımlayıcısı, "hücre" olarak adlandırılan küçük bir uzamsal bölgenin pikselleri üzerindeki gradyan yönlerinin birikmesine ve sonraki amaçlar için dikkate alınacak özellik vektörünü sağlayan bir 1 boyutlu histogramın oluşturulmasına dayanır. Analiz edilecek görüntü  $L, N \times N$  piksel boyutunda hücrelere bölünür ve her pikseldeki gradyanın  $\theta_{x,y}$  yönelimi hesaplanır. Ardından  $\theta_{ij}$  yönelimleri  $i = 1 \dots N^2$  yani aynı  $j$  hücresine ait olanlar nicelenir ve bir  $M$ -kutu histogramında toplanır. Son olarak, elde edilen tüm histogramlar, bu algoritmik adımın nihai sonucu olan, yani sonraki işlemler için dikkate alınacak özellik vektörü olan benzersiz bir HOG histogramında sıralanır ve birleştirilir (Velmurugan ve Baboo 2011).

## 2.6 Derin özelliklerin çıkarımı

Geleneksel özelliklerin sınırlı temsil kabiliyeti ve derin sinir ağları yoluyla görüntü işlemedeki atılım nedeniyle, Evrimsel Sinir Ağları (Convolutional Neural Network – CNN) tabanlı görüntü alma son yıllarda hızla gelişmiştir. CNN, görüntü sınıflandırmasında büyük bir atılım yaptığından, birçok araştırmacı, eğitilmiş CNN özneliklerini kullanmaya çalışmaktadır. Derin öğrenme (Deep Learning- DL) tabanlı teknikler, eğitim hesaplamaları için önemli miktarda veri gerektirmesine rağmen, önceden eğitilmiş ağlar, çeşitli alanlarda özellik çıkarma ve görüntü alma için kolayca uyarlanabilir. Bugün, CNN tabanlı yöntemler, görüntü alımında geleneksel düşük seviyeli görüntü tanımlayıcılarının neredeyse tamamen yerini alarak, görüntü alımında yaygın bir uygulama olarak kabul edilmektedir. CNN, görüntü alımı için geleneksel düşük seviyeli görüntü tanımlayıcıların yerini almak için kullanılır ve 3 sınıfa ayrılır: (1) Önceden eğitilmiş CNN modelleri, (2) İnce ayarlı CNN modelleri ve (3) Hibrit CNN modelleri. Bu çalışmada VGG16 (Simonyan ve Zisserman 2014) mimarisi kullanılmıştır. VGG16, ImageNet ile önceden eğitilmiş ve 2014 yılında ILSVR yarışmasını kazanan evrimsel bir sinir ağı mimarisidir. VGG16 yüksek doğruluklu bir özellik çıkarıcı olarak hizmet verebilir (Liu vd. 2019). Bu modelin mimarisi Şekil 4'te gösterilmiştir.



Şekil 4. VGG16 mimarisi (Simonyan ve Zisserman 2015)

Şekil 4'te verilen özellik çıkarımında temel fikir,  $(224 \times 224)$  boyutlu görüntünün, çok küçük alıcı alanlara sahip filtrelerin kullanıldığı ReLU aktivasyonlarına sahip bir evrimsel katmanlar yığınından geçmesidir. Uzamsal havuzlama, bazı konvolüsyonel katmanlardan sonra gerçekleştirilen beş maksimum havuzlama katmanı tarafından gerçekleştirilir (Gkelios vd 2021). Tamamen bağlantılı son katman, 1000 nöron içeren çıkış katmanı olarak çalışır ve bunun 1000'e kadar sınıfla sonuçlanabileceğini belirtir. Yukarıda verilen teknikler ile özellik çıkarımları yapıldıktan sonra özellik vektörleri birleştirilir. Özellik vektörlerinin birleştirilmesinde aşağıdaki yaklaşım uygulanmaktadır.

## 2.7 Özellik vektörlerinin birleştirilmesi

Her bir video karesinde tespit edilip maskelenen ve segmente edilen kişi görüntülerinin yukarıda bahsedilen

(renk, doku, şekil, CNN tabanlı) öznelikleri çıkarılmış ve bu öznelik vektörleri birleştirilerek her bir kişi için tek bir öznelik temsili oluşturulmuştur. Sonuç olarak her bir video karesi için; segmente edilen kişilerin öznelik vektörlerini içeren bir galeri oluşturulmuştur.

## 2.8 Benzerlik metriği ve sorgulama

Her bir video karesinde tespit edilen kişi sayısı  $N$  olmak üzere;  $I_n$  segmente edilen kişi görüntüsünün öznelik vektörünü,  $y_n$  ise bu kişinin etiketini temsil etmektedir.  $I_q$  sorgu görüntüsünü,  $y^q$  sorgu görüntüsünün etiketini ve  $N$  adet galeri resmi  $I = \{(I_1, y_1), (I_2, y_2), \dots, (I_N, y_N)\}$  şeklinde temsil edilmektedir.  $N$  adet görüntünün benzerlik oranına göre sıralama listesi  $\{r_1, r_2, \dots, r_N\}$  olarak gösterilebilmektedir. Burada  $r_i, I_i$  görüntüsünün sıralanmış indeksidir. Bu amaç fonksiyonu doğru eşleşmelerin sıralama indekslerini minimize ederek,



doğru eşleşmeleri sıralamada üst sıralara taşımayı hedefler ve Eşitlik 7'deki gibi özetlenebilir:

$$\min \sum_{i=1}^N r_i f(y^q, y_i), f(y^q, y_i) = \begin{cases} 1 & y^q = y_i \\ 0 & y^q \neq y_i \end{cases} \quad (7)$$

Eşitlik 7'de iki amaç fonksiyonu, sorgu görüntülerinin galeri görüntüleri içindeki doğru eşleşmelerinin sıralamasını optimize etmeyi hedefler. İlk fonksiyon, genel sıralama performansını iyileştirmeye çalışmaktadır. Sorgulanacak kişi girdi olarak alındığında aynı öznitelik modelleri ile sorgu kişinin de öznitelik vektörü oluşturulmuştur. Sorgu kişinin öznitelik vektörü ile her karedeki kişilerin daha önceden oluşturulmuş ve veri tabanı haline getirilmiş öznitelik vektörleri arasındaki benzerlik oranı hesaplanmıştır. Benzerlik oranlarının hesaplamasında Cosine Benzerliği (Cosine Similarity – CS) kullanılmıştır. Bu metrik sıfırdan farklı iki vektör arasındaki bir ölçüdür (Rahutomo vd. 2012).

Bu iki vektör arasındaki açının kosinüsünü ölçmek için iç çarpım uzayını kullanır. Öklid nokta çarpım formülü kosinüs benzerliğini şu şekilde hesaplamak için kullanılabilir:

$$a \cdot b = |a||b|\cos\theta, \quad (8)$$

burada  $a$  ve  $b$  iki vektördür ve  $\theta$  aralarındaki açıdır.  $x = [x_1, x_2, x_3, \dots, x_n]$  Vektörünün Öklid normu veya Öklid uzunluğu ile aynı olan büyüklük veya uzunluk kullanılarak (Eşitlik 9), S benzerliği Eşitlik 10 kullanılarak hesaplanır:

$$|x| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}, \quad (9)$$

$$S = \cos\theta = \frac{A \cdot B}{|A||B|} = \frac{\sum_i^N A_i B_i}{\sqrt{\sum_i^N A_i^2} \sqrt{\sum_i^N B_i^2}} \quad (10)$$

Eşitlik 10'da A ve B iki vektördür. Cosine benzerliği -1 ve 1 arasında değerler almaktadır. Benzer vektörlerin oranı 1'e yakın, farklı vektörlerin benzerlik oranı ise -1'e yakındır.

Cosine benzerliği dışında çalışmanın performansını hesaplamak için çeşitli ölçümler kullanılmıştır. Bu metrikler karmaşıklık matrisi üzerinden hesaplanmıştır. Karışıklık matrisinde, Gerçek pozitif (GP), doğru şekilde sınıflandırılan pozitif etiketli verilerin sayısıdır, gerçek negatif, doğru şekilde sınıflandırılan negatif etiketli verilerin sayısıdır. Yanlış pozitif (YP), pozitif olarak

sınıflandırılan negatif etiketli verilerin sayısıdır. Yanlış negatif, yanlışlıkla pozitif olarak sınıflandırılan pozitif etiketli verilerin sayısıdır. En çok kullanılan performans ölçütleri doğruluk, duyarlılık, kesinlik ve f puanıdır. Doğruluk, doğru şekilde sınıflandırılmış verilerin tüm verilere oranı olan bir performans ölçüsüdür. Kesinlik, gerçek pozitifin tüm pozitif sınıflandırılmış verilere oranıdır. Duyarlılık, gerçek pozitifin tüm pozitif verilere oranıdır. F puanı, kesinlik ve duyarlılığın harmonik ortalaması olduğundan model performansını hesaplamak için güçlü bir ölçümdür (Miao vd. 2022). En çok kullanılan performans ölçütleri kesinlik (Eşitlik 11), duyarlılık (Eşitlik 12) F-Skorunun (Eşitlik 13) hesaplanmasında kesinlik ve duyarlılık sırasıyla K ve D olarak ifade edilmektedir. Formüller aşağıda verilmiştir:

$$\text{Kesinlik} = \text{GP}/(\text{GP}+\text{YP}) \quad (11)$$

$$\text{Duyarlılık} = \text{GP}/(\text{GP}+\text{YN}) \quad (12)$$

$$\text{F-Skoru} = 2 \cdot (\text{K} \cdot \text{D}) / (\text{K} + \text{D}) \quad (13)$$

### 3. Bulgular

Önerilen çalışmada her bir kişi görüntüsünün öznitelik vektörünü oluşturmak için kullanılan Gabor, renk histogramı, HOG ve VGG16 yöntemlerinin parametreleri şu şekildedir:

Gabor filtresi kullanılarak desen özelliklerinin çıkarımında kullanılan parametreler şu şekildedir: Filtre bankası bir yönlendirme  $\theta = (180^\circ)$  ve  $F = (0,1,0.5,0.8)$  olmak üzere üç radyal frekans değeriyle gerçekleştirilmiştir. Böylece frekans alanını kapsayan toplam 3 filtre elde edilmiştir. Ayrıca Gauss eğrisinin farklı standart sapma değerleri test edilmiştir. Bunlar çalışmada kullanılan iki değerden oluşan bir settir  $\sigma = (1ve3)$ .

Renk histogramlarının çıkarılmasında ise HSV renk uzayında renk kanalı (hue) için 8 bölme, doygunluk (saturation) için 12 bölme, değer (value) kanalı için 3 bölme kullanılmıştır. Böylece  $8 \times 12 \times 3 = 288$  boyutunda bir özellik vektörü elde edilmiştir. Görüntülerin tamamı için 3 boyutlu HSV renkli histogramı hesaplamak yerine görüntünün farklı bölgeleri için renk histogramları hesaplanmıştır. Küresel histogramlar yerine bölgelere dayalı histogramların kullanılması, yerelliği bir renk dağılımında simüle etmeyi sağlamıştır. Görüntü tanımlayıcı sol üst köşe, sağ üst köşe, sağ alt köşe, alt- sol köşe ve son olarak görüntünün merkezi olmak üzere görüntü tanımlayıcı görüntüyü 5 farklı bölgeye ayırmıştır. Bu sayede bu bölgelerin her birinin histogramı ayrı ayrı tanımlanmıştır. Her bir bölge  $8 \times 12 \times 3 = 288$  boyutlu bir

histogramla temsil edilmiştir. 5 bölge için genel özellik vektörü  $5 \times 288 = 1440$  boyutunda olmaktadır. Böylece her görüntü 1.440 sayı kullanılarak ölçülmüştür ve temsil edilmiştir.

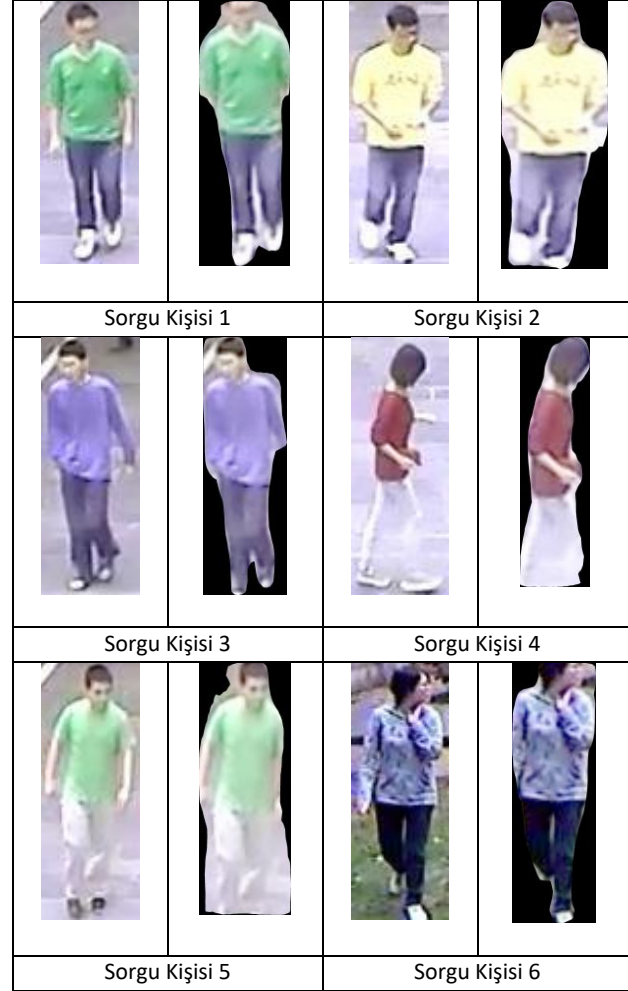
HOG tanımlayıcının uygulanması, görüntünün hücre adı verilen birbirine bağlı küçük bölgelere bölünmesini ve daha sonra her hücre için, her hücre içindeki piksellerin yönlendirilmiş gradyanlarının bir histogramının hesaplanmasını gerektirir. Daha sonra özellik vektörünü oluşturmak için bu histogramları birden fazla hücrede biriktirmek mümkündür. HOG tanımlayıcısındaki ilk gerçek adım, görüntü gradyanını hem x hem de y yönünde hesaplamaktır. Ardından görüntü hücelere ve bloklara bölünmektedir. Bir "hücre", her hücreye ait piksel sayısı ile tanımlanan dikdörtgen bir bölgedir. Bu çalışmada *pixels\_per\_cell* parametresi  $2 \times 2$  olarak seçilmiştir. Histogramları oluşturmadan önce ise yön sayısı belirlenmektedir. Yönlerin sayısı, ortaya çıkan histogramdaki kutu sayısını kontrol eder. Bu çalışmada yön sayısı 9 olarak seçilmiştir. Her hücredeki tüm piksellerin azalma yönlerini belirtilen sayıda yönlendirme bölmesinde gruplanıp, her bir bölmedeki gradyanların büyüklükleri toplandığında ortaya çıkan değer, bölmelerin yükseklikleri olacaktır. Kutu sayısı genellikle 9 olarak ayarlanmıştır, böylece her bir kutunun genişliği 20 derece olmuştur.

Son olarak VGG16 mimarisi ile çıkarılan özneliklerde ağın üst kısmına tamamen bağlı 3 katman dahil edilmemiştir. Böylece bu özellik için  $1 \times 512$  boyutunda öznelik vektörü oluşmuştur. Her bir karede tespit edilen kişi görüntülerinin farklı özellikleri birleştirilerek; kişilerin her birinin tek bir özellik vektörü ile temsil edilmesi sağlanmıştır. Çalışmada PKU-Reid veri setinden rastgele 26 kişinin farklı sayıda 5 farklı yönden çekilmiş rastgele görüntüleri galeri seti olarak kullanılmıştır. Galeri seti toplam 171 adet kişi görüntüsünden oluşmaktadır. 6 kişi ise sorgu kişisi olarak kullanılmıştır. Ön işlem aşamasında ilk olarak galeri seti ve sorgu kişilerinin görüntülerinin arka plandan temizlenmesi için Bölüm 2.2'de detaylı olarak açıklanan kişi segmentasyonu yöntemimiz uygulanmıştır. Bu sayede PKU-Reid veri setinden elde edilen galeri seti ve sorgu setinin görüntülerinde arka plan yok edilmiş, sadece kişilerin piksellerinin yer aldığı maskelenmiş görüntüler elde edilmiştir. Çizelge 1'de çalışmada PKU-Reid veri setinden seçilen sorgu kişilerinin arka plandan temizlenmeden önceki ve temizlendikten sonraki görüntüleri sunulmuştur.

İkinci olarak galeri seti ve sorgu görüntülerinin öznelik vektörleri oluşturulmuştur. 6 adet sorgu kişisi her bir galeri kişisinin özellik vektörleri ile tek tek karşılaştırılarak Cosine benzerliği hesaplanmıştır. Her bir sorgu kişisi ile

galeri görüntülerinin Cosine benzerliği hesaplandıktan sonra bu benzerlik oranları büyükten küçüğe sıralanmıştır. Çizelge 2'de her bir sorgu kişisi için benzerlik oranı en yüksek ilk 5 galeri resmi sonucu ve benzerlik oranları (CB kısaltması) gösterilmiştir. Doğru eşleşmeler yeşil ile yanlış eşleşmeler kırmızı ile işaretlenmiştir.





















**Çizelge 1.** PKU-reid veri setinden seçilen sorgu kişilerinin Mask-RCNN modeli kullanılarak arka plandan temizlenme sonuçları



Çalışmada PKU-Reid veri setinden oluşturulan galeri seti ve sorgu kişilerinin sonuçları için Ortalama Kesinlik (Average Precision) hesaplanmıştır. Ortalama Kesinlik değerleri tek bir sorgu için ayrı ayrı hesaplanan gerçek pozitif ve yanlış pozitif sorgu sonuçlarının ortalaması dikkate alınarak elde edilmiştir. Tespit algoritmasının performansını gösteren AP değeri sonuçları Çizelge 3'te gösterilmiştir. Sonuçlardan görüleceği üzere her bir sorgu kişisi için yüksek değerlere sahip Ortalama Kesinlik değerleri elde edilmiştir. Bu durum tespit algoritmasının başarımını göstermektedir. Ortalama benzerlik değerleri ise sorgu kişileri için en yüksek benzerlik oranına sahip ilk 5 sonuçtan doğru eşleştirme yapılan sonuçların ortalamasını yansıtmaktadır. Sonuçlardan anlaşılacağı üzere önerilen yöntem farklı açılara sahip kişi görüntüleri için yüksek eşleştirme performansı göstermiştir.



Çizelge 2. PKU-Reid veri setinden seçilen sorgu kişilerinin sorgu sonuçları

Sorgu Kişisi			Sorgu Sonuçları ve Cosine Benzerlik Sonuçları (CB)		
					
Sorgu Kişisi 1	CB : 0.852	CB : 0.869	CB : 0.861	CB : 0.912	CB : 0.913
					
Sorgu Kişisi 2	CB : 0.872	CB : 0.845	CB : 0.854	CB : 0.90	CB : 0.88
					
Sorgu Kişisi 3	CB : 0.803	CB : 0.772	CB : 0.837	CB : 0.806	CB : 0.809
					
Sorgu Kişisi 4	CB : 0.88	CB : 0.794	CB : 0.809	CB : 0.798	CB : 0.816
					
Sorgu Kişisi 5	CB : 0.805	CB : 0.792	CB : 0.782	CB : 0.804	CB : 0.857
					
Sorgu kişisi 6	CB : 0.821	CB : 0.857	CB : 0.833	CB : 0.871	CB : 0.898

**Çizelge 3.** Sorgu sonuçlarının ortalama benzerlik ve Ortalama Kesinlik değerleri

	Ortalama Benzerlik Değeri	Ortalama Kesinlik
Sorgu kişisi 1	0.881	1.0
Sorgu kişisi 2	0.872	1.0
Sorgu kişisi 3	0.805	0.804
Sorgu kişisi 4	0.820	0.804
Sorgu kişisi 5	0.804	0.809
Sorgu kişisi 6	0.865	1.0

Önerilen çalışmada performans karşılaştırması yapabilmek adına bir stok videosu kullanılmıştır. Çizelge 4'te havaalanı videosunda sorgulanan 4 kişinin gerçek pozitif tespitini içeren rastgele seçilmiş bir video karesinde işaretlenmiş sonuçları gösterilmiştir. Video bir havaalanında çekilmiştir. Video boyunca gösterilen açıda yerleştirilmiş bir kamera ile kişilerin yürüyüşü kaydedilmiştir. Video toplamda 285 kareden oluşmaktadır.

Çizelge 5'te ise örnek videoda tespit ve segmente edilen kişilerin görüntüleri gösterilmiştir. Verilen örnek video karesi haricinde video boyunca farklı kişiler de tespit edilmiştir. Videoda yer alan kişiler hareket halinde olduğu için her biri farklı sayıda video karesinde gözükmektedir. Kişi tespit ve segmentasyonu yöntemimizle (Bölüm 2.2) tespit edilen bu kişilerin her birinin video karesindeki

konum bilgisi takip yapabilmek için kaydedilmiştir. Çalışma kapsamında kullanılan videoda 4 örnek sorgu kişinin benzerlik sonuçları hesaplanmıştır. Çizelge 6'da videoda sorgulanan 4 kişinin video boyunca kaç karede yer aldığı, yer aldığı kareler boyunca sonuçların kesinlik (K), duyarlılık (D) ve F skorları sunulmuştur. Ek olarak sorgu kişinin en yüksek Cosine benzerliğine sahip olduğu ilk 5 sorgu sonucu gösterilmiştir.

**Çizelge 4.** Havaalanı videosunda sorgulanan kişilerin takibini gösteren örnek video kareleri



**Çizelge 5.** Örnek videoda tespit ve segmente edilen kişilerin görüntüsü



Elde edilen sonuçlar; önerilen algoritmanın sorgu sonuçları açısından yanlış pozitif tespitlerin göreceli olarak düşük bir oranda olduğunu göstermektedir. Özellikle video karelerinde kişilerin birbirlerini kapattığı durumlarda yanlış pozitif sonuçları artış göstermiştir. Bunun yanında ağırlıklı olarak aynı renk bilgisine sahip kişilerin de az miktar da olsa yanlış pozitif tespitlere neden olduğu gözlemlenmiştir.

Önerilen çalışmanın gerçek uygulamalar için kullanışlı olmasını sağlayan önemli bir özelliği ise her bir video karesinde sorgu kişisi ile eşleşen kişilerin takibinin yapılmasıdır. Her bir video karesinde sorgu kişisine en yüksek Cosine benzerliğine sahip olan kişi video karesinde işaretlenmektedir. Bu sayede çalışma kişi tanınmanın yanı sıra video üzerinden de takibe olanak sağlamaktadır.

















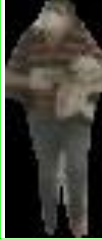







#### 4. Sonuçlar ve Tartışma

Bu çalışmada zorlayıcı bir problem olan kişi tanıma ve takibi problemi ele alınarak önerilen yöntem ile gerçek hayat uygulamalarına oldukça fayda sağlayacak bir altyapı geliştirilmiştir. Önerilen bu yöntemde kamera sistemleri videolarında belirli bir kişinin farklı özelliklerinin birleştirilerek daha güçlü bir temsil vektörüyle ifade edilerek gözetim videolarında aranması ve takip edilmesi sağlanmıştır. Gözetim videolarında tespit edilen kişilerin özellik çıkarımlarında renk, doku, şekil ve derin ağlarla elde edilen özelliklerden faydalanılmıştır. Bu sayede farklı seviyede özelliklerin birleştirilmesiyle daha etkin bir arama algoritması elde edilmiştir. Sonuç olarak bu çalışma ile güvenlik personelleri tarafından yüz tespiti yapılamayan olağan şüphelilerin takibini ve bir kişinin

kayıtlı bir videodan aranmasını kolaylaştırılması konusunda önemli bir alt yapı oluşturulmuştur. İlerleyen süreçlerde çalışmanın görüntü kalitesinin düşük olduğu gözetim videolarında da yüksek performans elde edilmesi

konusunda çalışılacaktır. Ek olarak tespit edilen kişilerin aksesuar/kıyafet bilgilerinin de çıkarılarak daha spesifik bir arama algoritması geliştirilmesi hedeflenmektedir.

**Çizelge 6.** Havaalanı videosuna ait 4 kişinin sorgu sonucu ve performans değerlendirmeleri

Sorgu Kişisi	Cosine Benzerlik Değeri (CB) En Yüksek İlk 5 Sorgu Sonucu					Sorgulanan Video Karesi Sayısı	Performans Sonuçları
						58	K : 0.8966 D : 1 F Skor : 0.9455
1	CB : 0.962	CB : 0.955	CB : 0.951	CB : 0.946	CB : 0.941		
						29	K : 0.8621 D : 1 F Skor : 0.9268
2	CB : 0.979	CB : 0.967	CB : 0.964	CB : 0.952	CB : 0.946		
						110	K : 0.65 D : 1 F Skor : 0.7879
3	CB : 0.908	CB : 0.901	CB : 0.906	CB : 0.893	CB : 0.891		
						186	K : 0.812 D : 1 F Skor : 0.895
4	CB : 0.957	CB : 0.933	CB : 0.903	CB : 0.905	CB : 0.9		

#### Etik Standartlar Bildirgesi

Yazarlar tüm etik standartlara uyduklarını beyan ederler.

#### Yazarlık Katkı Beyanı

Yazar 1: Metodoloji/Çalışma, deneysel tasarım, Analiz ve yorumlama, Araştırma, Yazma/orijinal taslak

Yazar 2: Denetleme/danışmanlık, Doğrulama

Yazar 3: Denetleme/danışmanlık, Doğrulama

#### Çıkar Çatışması Beyanı

Yazarların bu makalenin içeriğiyle ilgili olarak beyan edecekleri hiçbir çıkar çatışması yoktur.

#### Verilerin Kullanılabilirliği

Bu çalışma sırasında oluşturulan veya analiz edilen tüm veriler, yayınlanan bu makaleye dahil edilmiştir. Veri kümeleri istek üzerine mevcuttur. Bu makalenin sonuçlarını destekleyen ham veriler, yazarlar tarafından herhangi bir çekince olmaksızın sunulacaktır.

#### 5. Kaynaklar

Almasawa, M.O., Elrefaei, L.A., and Moria, K. 2019. A survey on deep learning-based person re-identification systems. *IEEE Access*, **7**, 175228-175247. <https://doi.org/10.1109/ACCESS.2019.2957336>

- Battal, A. ve Tuncer, A. 2022. Detection of Face Mask Wearing Condition for COVID-19 Using Mask R-CNN. *El-Cezeri*, **9(3)**, 1051-1060.  
<https://doi.org/10.31202/ecjse.1061270>
- Bäumli, M. and Stiefelhagen, R., 2011. Evaluation of Local Features for Person Re-Identification in Image Sequences. In *Proceedings of the International Conference on Advanced Video and Signal-based Surveillance (AVSS)*, 291-296.  
<https://doi.org/10.1109/AVSS.2011.6027339>
- Chung, D., Tahboub, K. and Delp, E.J. 2017. A two stream siamese convolutional neural network for person re-identification. In *Proceedings of the IEEE international conference on computer vision*, 1983-1991.
- Gkelios, S., Sophokleous, A., Plakias, S., Boutalis, Y. and Chatzichristofis, S.A. 2021. Deep convolutional features for image retrieval. *Expert Systems with Applications*, **177**, 114940.  
<https://doi.org/10.1016/j.eswa.2021.114940>
- He, K., Gkioxari, G., Dollár, P. and Girshick, R. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2961-2969.  
<https://doi.org/10.1109/ICCV.2017.322>
- He, K., Zhang, X., Ren, S. and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.  
<https://doi.org/10.1109/CVPR.2016.90>
- Leng, Q., Ye, M. and Tian, Q. 2019. A survey of open-world person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, **30(4)**, 1092-1108.
- Li, W., Mao, K., Zhang, H. and Chai, T. 2010. Designing compact Gabor filter banks for efficient texture feature extraction. In *Proceedings of the IEEE International Conference on Control Automation Robotics and Vision*, 1193-1197.
- Li, Y., Xie, S., Chen, X., Dollar, P., He, K. and Girshick, R. 2021. Benchmarking detection transfer learning with vision transformers.  
<https://doi.org/10.48550/arXiv.2111.11429>
- Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. and Zitnick, C.L. 2014. Microsoft coco: Common objects in context. In: *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, Proceedings, Part V 13*. Springer International Publishing, 740-755.  
[https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- Liu, F., Wang, Y., Wang, F.C., Zhang, Y.Z. and Lin, J. 2019. Intelligent and secure content-based image retrieval for mobile users. *IEEE Access*, **7**, 119209-119222.  
<https://doi.org/10.1109/ACCESS.2019.2935222>
- Luo, H., Gu, Y., Liao, X., Lai, S. and Jiang, W. 2019. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 0-0.  
<https://doi.org/10.48550/arXiv.1903.07071>
- Ma, L., Liu, H., Hu, L., Wang, C. and Sun, Q. 2016. Orientation driven bag of appearances for person re-identification. arXiv preprint arXiv:1605.02464.  
<https://doi.org/10.48550/arXiv.1605.02464>
- Miao, J., Zhu, W. 2022. Precision–recall curve (PRC) classification trees. *Evol. Intel.* **15**, 1545–1569  
<https://doi.org/10.1007/s12065-021-00565-2>
- Rahutomo, F., Kitasuka, T. and Aritsugi, M. 2012. Semantic cosine similarity. In *Proceedings of the International student conference on advanced science and technology-ICAST. South Korea: University of Seoul*,1.
- Roshan, S., Srivathsan, G., Deepak, K. and Chandrakala, S. 2020. Violence detection in automated video surveillance: Recent trends and comparative studies. *The Cognitive Approach in Cloud Computing and Internet of Things Technologies for Surveillance Tracking Systems*, Ch.11, 157-171.  
<https://doi.org/10.1016/B978-0-12-816385-6.00011-8>
- Simonyan, K. and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.  
<https://doi.org/10.48550/arXiv.1409.1556>
- Singh, J. and Shekhar, S. 2018. Road damage detection and classification in smartphone captured images using mask r-cnn. arXiv preprint arXiv:1811.04535.  
<https://doi.org/10.48550/arXiv.1811.04535>
- Su, C., Zhang, S., Xing, J., Gao, W. and Tian, Q. 2016. Deep attributes driven multi-camera person re-identification. In: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer International Publishing, 475-491.  
<https://doi.org/10.48550/arXiv.1605.03259>
- Tharsanee, R.M., Soundariya, R.S., Kumar, A.S., Karthiga, M. and Sountharajan, S. 2021. Deep convolutional neural network–based image classification for COVID-19 diagnosis. In *Data Science for COVID-19*.  
<https://doi.org/10.1016/B978-0-12-824536-1.00012-5>
- Velmurugan, K. and Baboo, S.S. 2011. Image retrieval using Harris corners and histogram of oriented gradients. *International Journal of Computer Applications*, **24(7)**, 6-10.  
<https://doi.org/10.5120/2968-3968>
- Wu, L., Wang, Y., Gao, J. and Li, X. 2018. Where-and-when to look: Deep siamese attention networks for video-based person re-identification. *IEEE Transactions on Multimedia*, **21(6)**, 1412-1424.  
<https://doi.org/10.48550/arXiv.1808.01911>

- Xiao, J., Li, S. and Xu, Q. 2019. Video-based evidence analysis and extraction in digital forensic investigation. *IEEE Access*, **7**, 55432-55442.  
<https://doi.org/10.1109/ACCESS.2019.2913648>
- Yao, H., Zhang, S., Zhang, D., Zhang, Y., Li, J., Wang, Y. and Tian, Q. 2017. Large-scale person re-identification as retrieval. *In Proceedings of the International Conference on Multimedia and Expo (ICME)*, 1440-1445.  
<https://doi.org/10.1109/ICME.2017.8019485>
- Yi, D., Lei, Z. and Li, S.Z. 2014. Deep metric learning for practical person re-identification, 34-39. ArXiv e-prints, 89.  
<https://doi.org/10.48550/arXiv.1407.4979>
- Zekri, K., Touzi, A.G. and Lachiri, Z. 2017. A comparative study of texture descriptor analysis for improving content based image retrieval. *In Proceedings of the International International Conference on Control, Automation and Diagnosis (ICCAD)*, 247-253.  
<https://doi.org/10.1109/CADIAG.2017.8075665>.

#### **İnternet Kaynakları**

- 1- Allison, D. and Wintz, M. 2009. Related Image Retrieval using Color and Gabor Filter Histogram Featurization. [http://www.cs.pomona.edu/~dkauchak/ir\\_project/witepapers/Image.pdf](http://www.cs.pomona.edu/~dkauchak/ir_project/witepapers/Image.pdf) (23.10.2023).