



T.C.  
Ege Üniversitesi  
Mühendislik Fakültesi  
Elektrik Elektronik Mühendisliği Bölümü  
Lisans Bitirme Projesi Sonuç Raporu

## **GÖRÜNTÜ İŞLEME TABANLI GERÇEK ZAMANLI İNSAN KİTLESİ KONTROL SİSTEMİ**

### **REAL-TIME CROWD CONTROL SYSTEM BASED ON IMAGE PROCESSING**

Caner YILDIRIM  
05140000548, caneryldrim@gmail.com

Emir Kaan YERLİ  
05150000734, ekyerli@outlook.com

**Proje Danışmanı: Dr.Öğr.Üyesi Erkan Zeki ENGİN**

Bornova, İZMİR, Haziran 2020

<b>E.Ü. ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ BÖLÜMÜ</b> <b>LİSANS BİTİRME PROJESİ BİLGİ FORMU</b>	
1- Projenin Yapıldığı Dönem: 2019-2020 (Yıllık)	2- Rapor Tarihi: 21.06.2020
3- Projenin Başlangıç ve Bitiş Tarihleri: Ekim 2019 – Haziran 2020	
4- Projenin Adı: Görüntü İşleme Tabanlı Gerçek Zamanlı İnsan Kitlesi Kontrol Sistemi Project Name: Real Time Crowd Control System Based on Image Processing Projenin Toplam Maliyeti: Yazılım tabanlı bir proje olduğundan ve açık kaynak kodlarından yararlanıldığından herhangi bir maliyet oluşmamıştır.	
5- Proje Yürütücüsü Öğrenciler ve iletişim bilgileri (adres, e-posta, tel.) Caner YILDIRIM - caneryldrim@gmail.com - +905076658085 Emir Kaan YERLİ - ekyerli@outlook.com - +905342880330	
6- Projenin Yürütüldüğü Kuruluş: Ege Üniversitesi Elektrik – Elektronik Mühendisliği	
7- Destekleyen Kuruluş(ların ) Adı, Adresi ve Destek Miktarı: Proje için destek alınmamıştır.	
8- Öz: Günümüzde her türlü veri ve bu verilerin doğru analiz edilip, kullanılması çok daha önem arz eden bir seviyeye ulaşmış durumdadır. Belirli bir alana giriş ve çıkış yapan kişileri sayabilen kamera sistemleri ise bu bağlamda önemi gittikçe artan bir veri sağlayıcı olarak görülebilir. COVID-19, korona virüsü nedeniyle de belirli bir alandaki kişi yoğunluğunu takip etmenin önemi artarken bunu sağlayan sistemler maliyetli olmaları nedeniyle henüz yaygınlaşmamış durumdadır.  Bu projede belirli bir bölgedeki kişi yoğunluğunu takip edebilmek için bir alana giren ve çıkan kişileri tespit edip, toplam giriş ve toplam çıkış verilerini raporlayan ve standart bir CPU üzerinde gerçek zamanlı olarak çalışabilen bir sistem tasarlanmaktadır. Tasarımda özellikle düşük maliyet-düşük hata oranı gözetildi ve herhangi bir bölgeye kolayca uygulanabilir olması hedeflendi. Python üzerinde geliştirilen algoritmalarda, standart bilgisayar görme/görüntü işleme işlevleri için OpenCV ve sayımı yapılacak kişileri tespit etmek için ise derin öğrenme kütüphanesi MobileNet SSD (Single Shot Object Detection) Caffè (Convolutud Architecture for Fast Feature Embedding) modelinden yararlanıldı. Tespit edilen kişilerin gelişmiş izleme doğruluğu ile takip edilebilmesi için ise hem sentroid izleme hem de korelasyon filtreleri dahil olmak üzere iki ayrı nesne izleme algoritması kullanıldı. Derin öğrenme tekniklerinin sağladığı yüksek doğruluk ve gelişen teknoloji ile aşılın işleme gücü sınırları nedeniyle geleceğin kişi sayma teknolojileri için akla yatkın bir karar olarak değerlendirilebilir.  Anahtar Kelimeler: <i>Kişi Sayma, Görüntü İşleme, OpenCV, MobileNet SSD, Caffè, Derin Öğrenme, Yapay Sinir Ağları</i>	
9- Danışmanının Öğretim Üyesi Adı/Soyadı ve Görüşü: Dr.Öğr.Üyesi Erkan Zeki ENGİN	
10- Bölüm Kurulu Görüşü:	
11- Projenin Başarı Durumu:	Projenin Aldığı Not:

## ÖNSÖZ

Günümüzde her türlü veri ve bu verilerin doğru analiz edilip, kullanılması çok daha önem arz eden bir seviyeye ulaşmış durumdadır. Belirli bir alandaki giriş ve çıkışları sayabilen kamera sistemleri ise bu bağlamda önemi gittikçe artan bir veri sağlayıcı olarak görülebilir. COVID-19 virüs salgını nedeniyle belirli bir alandaki kişi yoğunluğunu takip etmenin önemi artarken bunu sağlayan sistemler henüz yaygınlaşmamış durumdadır. Projemizde bu açığı kapatabilmek amacıyla, düşük maliyet-düşük hata oranı gözetilerek, elde ettiği verileri hızlı, basit ve güvenilir şekilde sunan derin öğrenme tabanlı bir kişi sayma sistemi tasarlanmıştır.

Proje süresince bilgi ve tecrübesini bizden esirgemeyen Dr.Öğr.Üyesi Erkan Zeki ENGİN hocamıza sonsuz teşekkür ve minnetlerimizi sunmaktan onur ve mutluluk duyarız.

Caner YILDIRIM, Emir Kaan YERLİ  
Ege Üniversitesi, Mühendislik Fakültesi  
Elektrik-Elektronik Mühendisliği Bölümü  
İzmir, Haziran 2020

# ÖZET MAKALE

## GÖRÜNTÜ İŞLEME TABANLI GERÇEK ZAMANLI İNSAN KİTLESİ KONTROL SİSTEMİ

### REAL-TIME CROWD CONTROL SYSTEM BASED ON IMAGE PROCESSING

Caner YILDIRIM, Emir Kaan YERLİ  
Ege Üniversitesi  
Bornova, İzmir, 35100, Türkiye

#### Öz

Günümüzde her türlü veri ve bu verilerin doğru analiz edilip, kullanılması çok daha önem arz eden bir seviyeye ulaşmış durumdadır. Belirli bir alana giriş ve çıkış yapan kişileri sayabilen kamera sistemleri ise bu bağlamda önemi gittikçe artan bir veri sağlayıcı olarak görülebilir. 2020 ilk çeyreği ile ortaya çıkan ve yayılımı artan COVID-19 virüsü nedeniyle de belirli bir alandaki kişi yoğunluğunu takip etmenin önemi artarken bunu sağlayan sistemler maliyetli olmaları nedeniyle henüz yaygınlaşmamış durumdadır.

Bu projede belirli bir bölgedeki kişi yoğunluğunu takip edebilmek için bir alana giren ve çıkan kişileri tespit edip, toplam giriş ve toplam çıkış verilerini raporlayan ve standart bir CPU üzerinde gerçek zamanlı olarak çalışabilen bir sistem tasarlanmaktadır. Tasarımda özellikle düşük maliyet-düşük hata oranı gözetildi ve herhangi bir bölgeye kolayca uygulanabilir olması hedeflendi. Python üzerinde geliştirilen algoritmalarda, standart bilgisayar görme/görüntü işleme işlevleri için OpenCV ve sayımı yapılacak kişileri tespit etmek için ise derin öğrenme kütüphanesi MobileNet-SSD (Single Shot Object Detection) Caffé (Convolutated Architecture for Fast Feature Embedding) modelinden yararlanıldı. Tespit edilen kişilerin gelişmiş izleme doğruluğu ile takip edilebilmesi için ise hem sentroid izleme hem de korelasyon filtreleri dahil olmak üzere iki ayrı nesne izleme algoritması kullanılmaktadır. Derin öğrenme tekniklerinin sağladığı yüksek doğruluk ve gelişen teknoloji ile aşılın işleme gücü sınırları nedeniyle geleceğin kişi sayma teknolojileri için akla yatkın bir karar olarak değerlendirilebilir.

**Anahtar Kelimeler:** *Kişi Sayma, Görüntü İşleme, OpenCV, MobileNet SSD, Caffé, Derin Öğrenme, Yapay Sinir Ağları.*

#### Abstract

Today, all kinds of data and the correct analysis and use of these data have reached a much more important level. The camera systems that can count the people entering and exiting a certain area can be seen as an increasingly important data provider in this context. In addition, due to Covid-19, which emerged and spread with the first quarter of 2020, it becomes more important to follow the density of people in a particular area. However, the systems providing this have not become widespread due to their costs.

In this project, a system is designed to detect the people entering and leaving an area, to report total input and total output data, and to work in real time on a standard CPU in order to track the density of people in a particular region. The design was made especially considering the low cost-low error rate. It is also aimed to be easily applied to any building or area. Algorithms were developed on Python. OpenCV was used for computer vision / image processing functions. The Caffé (Convolutated Architecture for Fast Feature Embedding) model along with the deep learning library MobileNet-SSD (Single Shot Object Detection) was used to identify the people to be counted. Then, two separate object tracking algorithms, including sentroid tracking and correlation filters, were used to track the detected people with improved tracking accuracy. The choice of deep learning techniques can be considered a plausible decision for future person counting technologies due to the high accuracy it provides and the processing power limits exceeded by the rapid development of technologies in this field.

**Keywords:** *Person Counting, Image Processing, OpenCV, MobileNet SSD, Caffé, Deep Learning, Artificial Neural Networks.*

#### 1. Giriş

Günümüzde her türlü veri ve bu verilerin doğru analiz edilip, kullanılması çok daha önem arz eden bir seviyeye ulaşmış durumdadır. Belirli bir alana giriş ve çıkış yapan kişileri sayabilen kamera sistemleri ise bu bağlamda önemi gittikçe artan bir veri sağlayıcı olarak görülebilir. Bu sistemler,

deprem, yangın vb. acil durum senaryolarında güvenlik ve kurtarma operasyonlarının planlanmasında ve COVID-19 virüs salgını nedeniyle de belirli bir alandaki kişi yoğunluğunu takip etmede yardımcı bir araç olarak kullanılabileceğinden bu alanda ilerici yöntemler kullanarak ve uygulanabilirliği yüksek bir sistem tasarlanmasının önemi gitgide artmaktadır. Öyle ki her

erişim noktasının üstüne kurulan kişi sayıcıları ile giren veya çıkan tüm insanların iki yönlü bir sayımı gerçekleştirilebilir. Tüm girişleri ve çıkışları gerçek zamanlı olarak karşılaştırarak, bir odanın, belirli bir katın veya tüm bir binanın doluluk derecesi belirlenebilir, o bölge için yoğunluk dağılım haritası oluşturulabilir. Bir binanın tahliyesinde veya bölgesel bir afet durumunda ise o sırada binadaki kişilerin tam sayısı hakkındaki bu değerli bilgiler acil durum ekiplerine sunulabilir. En kapsamlı haliyle ise akıllı şehirler için anlık yoğunluk bilgisi sağlayacak bu sistem ile ulaşım, aydınlatma, havalandırma, güç tüketimi gibi birçok farklı alanda optimizasyon sağlanabilir.

Öte yandan mevcut sistemlerin maliyetleri ve güvenilirliklerine dair endişeler bu teknolojinin yaygınlaşmasına engel olmaktadır. Bu çalışmanın da kapsamında yer alan güncel uygulamalar ile bu sistemleri yüksek doğruluk oranları ve daha düşük maliyetler ile optimize ederek bu alandaki açığa kapatmak ve kesinlik gerektiren durumlara uygun hale getirmek hedeflenmektedir.

Bütün bunlar için geliştirilecek algoritma anlık olarak görüntüyü işleyebilmeli, birden fazla giriş ve çıkış anında hata yapmamalı, gölgelerden ve arka plandan bağımsız olarak güvenilir veri sağlayabilmelidir. Kameranın konumuna ve görüş açısına dikkat edilerek farklı saç rengine, kafa şekline sahip insanlar için algoritmalar geliştirilmeli ve bisiklet, bebek arabası, evcil hayvanlar gibi unsurların sayımı etkilememesi sağlanmalıdır. Bunun yanında bu bilginin acil durumlarda kullanılabilmesi için internet tabanlı bir iletişim sistemi tasarlanmalı, uygun görsel ve teknik raporlamalar oluşturulmalıdır.

### Hedefler

- Binalarda bulunan kişi sayısını, gerçek zamanlı olarak yüksek doğrulukla verebilen, uygulanabilir bir sistem tasarlanmalıdır.
- Sistemin mevcut güvenlik kameralarına entegre edilebilir ve mevcut internet altyapısını bağlanabilir olması gerekmektedir.
- Algoritma hızlı, basit ve güvenilir olmalıdır.
- İşlenen bilgilere acil durumlarda dahi internette kolayca ulaşabilmek için sistemin bir Ethernet ağının parçası olması gerekmektedir.
- Görüş alanı içerisinde veya sayım çizgisi yakınlarda uzun süre beklemeler, tekrarlı giriş-çıkışlar ihmal edilebilmelidir.

### Zorluklar

- Sayım bölgesinin yakınında çok sayıda insan olabilir. (Sayım bölgelerinde beklemeler olacaktır.)

- Ana girişteki aydınlatma koşulları güneşin gün içindeki konumunun değişmesi ya da binanın çalışma kuralları nedeniyle tutarsız olacaktır. Ana girişteki sayım kritik önem arz etmektedir. Bu alandaki ziyaretçileri saymak standart bir kapıdan veya düz bir koridordan geçen tek bir kişiyi tespit etmek için gereken basit bir durumdan çok daha zordur.

- Birbirine yakın bir grup içerisindeki her bireyin takip ettiği yolları hassas bir şekilde izleyebilen bir sistem tasarlamak zordur.

### 1.1. Literatür Özeti

Açık ve/veya kapalı alanlarda insanları bulmak ve izlemek geniş bir araştırmanın hedefi olmuştur. Bu çalışma kapsamındaki hedefler doğrultusunda literatürdeki çalışmalar da detaylı şekilde incelenmiştir. Birçok teknik bu pazarın gerekliliğini az çok başarılı bir şekilde ele almaktadır. İzlemenin gerçekleştiği koşullar, başarısı için kritik öneme sahiptir. Bu alandaki ilgili bazı çalışmalara aşağıda yer verilmektedir.

Basit ve etkili bir yaklaşım olarak Sexton bir segmentasyon algoritması kullanır [1]. Sistemlerini Parisli bir tren istasyonunda test ettiler ve %1 ila %20 arasında bir hata aldılar. Sistemleri, insanları arka plandan izole etmek için basit bir arka plan çıkarma yöntemi kullanır ve çerçeveler arasındaki eşleşmeyi yapmak için sadece damlaların ağırlık merkezlerini alır. Sistemin sağlamlığını artırmak için arka plan modeli sürekli olarak güncellenerek çevresel değişikliklerin sistem üzerindeki etkisini azaltır. Kamera, oklüzyonları azaltan ve damla tespit problemini basitleştiren üstten bir konuma asılmıştır.

Segmentasyon sonrası görüntü işleme üzerine yoğunlaşan bir başka yaklaşımda, damlaları basit bir arka plan çıkarma yöntemiyle ayıklar ve sonra çerçeveler arasındaki özelliklerini izlerler [2]. Her damlanın yolları saklanır ve sanal bir çizgiyi geçerken kesişim noktasını ve yönünü tespit etmek için kullanılır. Bu sistem oklüzyon problemleriyle uğraşmaz, bu nedenle kalabalık ortamlarda performansı büyük ölçüde azalmaktadır.

Stereo kamera kullanarak geçitten geçen insanları otomatik olarak sayma yöntemi önerilmiştir [3]. Önerilen yöntemde, stereo kamera kapının tavanına asılır ve kameranın optik eksenini, geçen insanların sadece yukarıdan gözlenebilmesi için ayarlanır. Çünkü bu sistem düzenlemesinde, eğer kapıda kalabalık bir insan grubu varsa, o zaman geçen insanların görüntü verileri, elde edilen görüntüler üzerinde birbirleriyle örtüşmez. Ek olarak, geçen insanların her bir yüksekliği, piramitleştirmeye dayalı algoritma uygulanarak ölçülmüş. Stereo kamera kullanarak yapılan bu sistemin en büyük avantajı, kamera tam tepede olduğu için birbirini kapatan görüntüler

olmayacaktır. Bu sayede doğruluk payı daha yüksek ölçümler yapılmaktadır.

Kişi sayma sistemlerinde yeni metot olarak birden fazla kamera kullanmayı önermektedir [4]. Bu yöntemde tek bir kamera için uyarlanabilir kayan pencere (adaptive sliding window) algoritması ile tasarlanır ve eşyazım (homography) teorisine göre birçok kamera arasındaki ortak görüntüler (aynı sahnenin farklı görünüşleri) hesaplanır. Tıkanma faktörünü göz önünde bulunduran nesne eşleştirme stratejisi, tek kameranın sonuçlarını iyileştirmek ve birden fazla kameranın sayımını elde etmek için tasarlanmıştır. Deney sonuçları, bu yöntemin karmaşık durumlarda iyi performans gösterdiğini göstermiştir. Ancak gerçek zamanlı performansı düşüktür.

## 1.2. Derin Öğrenme Mimarileri ile Kişi Sayma

Derin öğrenme alanındaki güncel çalışmalar ve uygulamalar kişi sayma sistemlerinde klasik yöntemlerle karşılaşılan zorlukların aşılmasına yardımcı olmaktadır. Klasik yöntemlerde özneliliklerin ayrıştırılması büyük miktarlarda tecrübe ve deney gerektirmekteydi. Ancak evrişimli sinir ağları ile insan görme sisteminin modellenerek bilgisayarlı görü teknolojilerinde yüksek doğrulukta sonuçlar gerçek zamanlı olarak elde edilebilir noktaya gelmiştir. Bunun için geliştirilen farklı model ve kütüphanelerin OpevCV ile tek bir noktada çalıştırılabilir olması da bu tekniklerin kullanılabilirliğini arttırmıştır. Bu modeller ile görüntülere ait özellikler daha derin katmanlarda öznelilikleri seçmek için işlem yapmayı gerek duymadan öğrenilebilmektedir.

Zhang ve çalışma arkadaşları 2015 yılında derin evrişimli sinir ağı modelinden yararlanarak kişileri çapraz sahne yöntemiyle saymayı önermiştir. Eğitim verisi, kişi sayımı ve kişi tespiti olmak üzere iki hedef içermektedir [5].

Boominathan ve çalışma arkadaşları CrowNet sistemini önererek yine evrişimli bir sinir ağı modeli kullanmıştır. Bir alanın yoğunluk haritasını tahmin edebilmek amacıyla derin ağ ve evrişimli sinir ağlarından beraber yararlanmışlardır. Bu model ile semantik bilgiler (beden, yüz) yakalanabilmiştir. Veri artırma (data augmentation) ile eğitim örneklerindeki eksikler kapatılmıştır. İlk veri artırma ile çoklu kişi örneklemedeki sorunlar çözülmüş, ikincisinde ise hataları azaltmak ve yoğun durumlar için modelin performansı iyileştirilmeye çalışılmıştır. Buradaki derin öğrenme mimarisi, VGG-16 ağına benzemektedir [6].

2016 yılında, Zhang ve çalışma arkadaşları, kişi sayımını değişen perspektiflerden oluşturulan görüntüler ile tahmin edebilmek amacıyla evrişimli

sinir ağı yapısı uygulamıştır. Çok katmanlı evrişimli sinir ağı modeli görüntüyü çoklu yoğunluk haritasına eşleştirmektedir. Bu yapının önemli özelliği, rastgele bir çözünürlük ve boyutta giriş yapılabilmesini sağlamasıdır. Çeşitli filtreler ile görüntüdeki değişimlere karşılık gelen özelliklerin öğrenilmesi sağlanır. Böylece bu yöntem ile herhangi bir boyuttaki görüntüdeki bozulmaları önleme imkânı ortaya çıkar [7].

Onoro, Hidra evrişimli sinir ağı (Hydra CNN) adında bir model ve ölçek temelli bir sayma modeli geliştirmiştir. Derin ve evrişimli tüm sinir ağlarını içinde barındıran bu sistem ile kişi yoğunluğu hesaplanabilmiştir. Altı katmanlı bir yapıdan meydana gelen CCNN (counting convolutional neural network-CCNN) modeli bu katmanların paralel olarak işlenmesiyle hedefe ulaşmaktadır. Nesne ve kişi sayma konusunda başarılı bir model olarak değerlendirilmektedir [8].

## 2. Sistemin Tanıtımı ve Çalışma Prensipleri

Bu çalışmada belirli bir bölgedeki kişi yoğunluğunu takip edebilmek için bir alana giren ve çıkan kişileri tespit edip, toplam giriş ve toplam çıkış verilerini raporlayan ve standart bir CPU üzerinde gerçek zamanlı olarak çalışabilen bir sistem tasarlanmaktadır. Python üzerinde geliştirilen algoritmalar, standart bilgisayar görme/görüntü işleme işlevleri için OpenCV ve sayımı yapılacak kişileri tespit etmek için ise derin öğrenme yapıları MobileNet ve SSD (Single Shot Object Detection) Caffé (Convolutated Architecture for Fast Feature Embedding) modelinden yararlanıldı. Bu iki tekniğin birlikte kullanılması kısıtlı kaynağa sahip cihazlardaki gerçek zamanlı performansı büyük ölçüde arttırmaktadır. Tespit edilen kişilerin gelişmiş izleme doğruluğu ile takip edilebilmesi için ise hem sentroid izleme hem de korelasyon filtreleri dahil olmak üzere iki ayrı nesne izleme algoritması kullanıldı. Derin öğrenme tekniklerinin sağladığı yüksek doğruluk ve gelişen teknoloji ile aşılın işleme gücü sınırları düşünüldüğünde geleceğin kişi sayma teknolojileri için akla yatkın bir karar olarak değerlendirilebilir.

Sayma işlemi yapmadan önce elbette nesnelerin(kişilerin) tespit edilmesi gerekmektedir. Nesnelerin bulunması için ise bir girdi görüntüsünü alıp, görüntünün içindeki nesneleri tespit etmek, tespit ettikten sonra da nesnenin bir insan olup olmadığının karar vermek gerekmektedir. Bu amaçla ise doğruluk ve hız kriterlerini karşılayan derin öğrenme algoritması, evrişimli sinir ağı CNN (ConvNet) tercih edilmiştir. CNN verilen görüntüleri ayırt etmek için bir insanı insan haline getiren benzersiz özellikleri kullanır. Aslında bu süreç beynimizde de bilinçsizce yapılmaktadır. CNN'de

aynı şeyi yapar. Ancak daha önce eğriler ve kenarlar gibi alt düzey özellikleri tespit ederler ve daha soyut kavramlara kadar bunları oluştururlar. Bu yaklaşımla ilgili sorun, ilgilenilen nesnelerin görüntü içinde farklı uzamsal konumlara ve farklı en boy oranlarına sahip olmasıdır. Bu nedenle, çok sayıda bölge seçmek zorunda kalınmaktadır ve bu hesaplamada sorun yaratabilir. Bu nedenle, R-CNN, YOLO, SSD gibi algoritmalar bu oluşumları bulmak ve hızlı bir şekilde kullanmak için geliştirilmiştir. Bu çalışma kapsamında teknikler değerlendirilerek gerçek zamanlı performans ve doğruluk kriterleri açısından SSD tekniği ile MobileNet yapısının bir arada kullanılmasına karar verildi. Bölüm 2.1 ve bölüm 2.2’de bu teknikler ile ilgili detaylı bilgilere yer verilmiştir.

Gerçek zamanlı olarak nesne algılaması için tasarlanmış olan SSD ile sınır çerçeveleri çizilerek çözünürlük düşürülmektedir. Ayrıca derinlemesine ayrılabilir evrişimden yararlanan MobileNet yapısıyla da ağıdaki parametre sayısı büyük ölçüde azaltılmaktadır. Böylece, CNN’e gönderilen ana resmin boyutunu azaltarak, sistem gereksiz işlem yükünden kurtarılmaktadır. SSD, özellik haritalarını çıkarmak için VGG16 ağını kullanmaktadır. Daha sonra Conv4\_3 katmanı ile nesneleri algılar. SSD, VGG16’dan sonra 6 yardımcı konvolüsyon katmanı daha ekler. Bunlardan beşi nesne tespiti için eklenmektedir. Bu katmanların üçünde 4 yerine 6 tahmin yapılır. Toplamda SSD, 6 katman kullanarak 8732 tahmin yapar. Sonuç olarak tespit edilen nesnelerin model ile eşleşmesine dayanarak tahmin değerleri sağlar. SSD bir nesne olup olmadığını belirtmek için 0 ile 1 arasında bir değer kullanır. SSD bu çalışmadaki renkli video görüntüleri gibi çok ölçekli özellik haritalarının doğruluğunu önemli ölçüde artırır. Nesne tespiti için kullanılan çok sayıda özellik eşleşme katmanının doğruluğunu artırmaktadır. Bu işlemlerin sonucunda çıkan nesne matrisi Caffé modeli ile CNN içerisinde konvolüsyon ve Relu işlemine uğrar, bunun sebebi katmanlardaki özniteliklere sahip nöronlardan geçirilirken aralarındaki benzerlikleri bulmaktır. Bulunan benzerlikler sonucunda her resmin içerisindeki nesnenin tanınması beklenmektedir.

Bir sonraki aşamada tanınan nesneler belli bir eşik değeri ile karşılaştırılır. Eğer insan olma ihtimali örneğin 0.65 üzerindeyse insan olarak etiketlenir. Etiketlenen bu nesnelerin merkez noktaları CentroidTracker sınıfının içine atanır. Bu nesneleri çerçeve içerisine almak yerine sadece nokta ile göstererek işlemciye daha az işlem yükü yüklenebilir. Etiketlenen nesnelerin merkez noktaları atandıktan sonra TrackableObject sınıfıyla merkezi noktaları her etiketlenmiş nesne için özelleştirilir. Bu sayede nesneleri takip etme işlemi yapılır. Nesnelerin her çerçeve arasındaki hareket yönüne bakılarak regresyon yöntemi uygulanır.

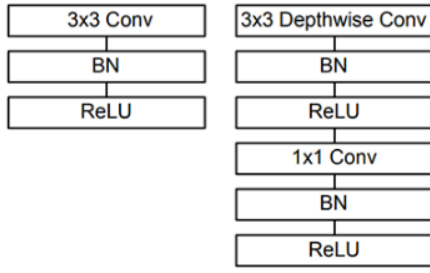
Regresyon yöntemi etiketlenmiş nesnenin belirlenen referans çizgisinin hangi yönüne hareket ettiğinin belirlenmesinde yardımcı olur. Tespit edilen kişinin ağırlık merkezinin hareket yönüne bakılarak giriş mi çıkış mı yapıldığı tespit edilir. Hareket yönüne göre giriş ve çıkış değerleri artırılır. Program sonlandırıldığında ise kullanıcıya bir kişi/zaman grafiği tablo olarak paylaşılır.

Algoritmanın çalışma adımları en basit şekliyle sıralanacak olursa,

- i. Gerçek zamanla bir video akışı veya kayıtlı herhangi bir video sisteme yüklenir.
- ii. Nesne Tespit Modülü: SSD nesne dedektörü kullanılarak sınırlayıcı kutular belirlenir. Sonrasında ise MobilNet Caffé evrişimli sinir ağından geçirilir ve eğitilmiş model ile girdileri karşılaştırarak nesnenin insan olup olmadığına dair bir tahmin elde edilir. Bu noktada sonuçları keskinleştirmek amacıyla düşük ağırlık yüzdesine sahip tahminler filtrelenir. Kişi olduklarına karar verilen nesnelere kimlikleri atanır.
- iii. Nesne Takip Modülü: Nesne takibi için dlib kütüphanesi ile sentroid izleme ve korelasyon tekniklerinden yararlanılır. Ağırlık merkezleri belirlenen ve kimlikleri atanan nesnelerin her karedeki bir önceki karedeki konumlarının korelasyon durumları incelenerek kişinin en olası mesafedeki konuma yöneldiği varsayılır. Bu şekilde kişilerin hareket yönleri izlenebilir.
- iv. Arayüz üzerinde, tanımlanan kişi nesnelerinin sentroidlerine kimlik bilgisi atanarak arka planda her kişi veri tabanına kaydedilir ve ardışık 40 karede tespit edilemeyen kişiler sistemden çıkarılır.
- v. Sistem çalışırken anlık olarak toplam giriş ve çıkış verileri ekrana yazdırılır. Sistemin sonlandırılmasıyla kişi/zaman raporu oluşturulur.

## 2.1. MobileNet

Nesne algılama ağları oluştururken genellikle VGG veya ResNet gibi mevcut bir ağ mimarisini kullanılmaktadır. Daha sonra ise bu nesne algılama hattı içinde kullanılır. Sorun şu ki, bu ağ mimarileri 200-500 MB boyutları ile çok büyük olabilir. Bu gibi ağ mimarileri, büyüklükleri ve sonuçtaki hesaplama sayıları nedeniyle kaynakları kısıtlı cihazlar için uygun değildir. Bu nedenle bu çalışmada, Google araştırmacılarının başka bir makalesi olan MobileNets’i (Howard ve diğerleri, 2017) kullanılmaktadır. Bu ağlara “Mobil Ağlar” diyoruz çünkü kaynak kısıtlı cihazlar için tasarlanmışlardır. MobileNets, geleneksel CNN’lerden derinlemesine ayrılabilir özelliği olan kıvrım kullanımı ile farklılık gösterir (Şekil 2).



Şekil 1 Toplu normalizasyon ve ReLU ile standart evrişimsel tabaka [8].

Derinlemesine ayrılabilir evrişimin ardındaki genel fikir evrişimi iki aşamaya ayırmaktır:

- $3 \times 3$  derinlemesine evrişim.
- Ardından  $1 \times 1$  noktasal evrişim.

Bu sayede ağdaki parametre sayısı büyük ölçüde azaltılmaktadır. Buradaki sorun ise doğruluk yüzdesi ile kaynak tasarrufu arasında bir takas yapılmasıdır. Bu dengeyi sağlayabilmek amacıyla MobileNet ve SSD tekniklerini bir arada kullanılabilir. Hem MobileNet mimarisini hem de Tek Atış Dedektörü (SSD) yapısını birleştirildiğinde, nesne algılamaya yönelik hızlı, verimli, derin öğrenme tabanlı bir yöntem elde edilmektedir.

## 2.2. SSD

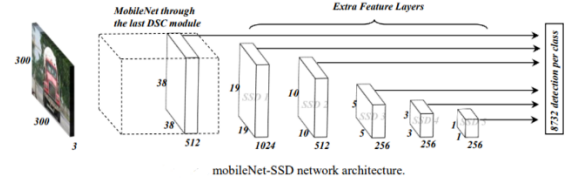
Derin öğrenme tabanlı nesne algılama söz konusu olduğunda, muhtemelen karşılaşacağınız üç temel nesne algılama yöntemi vardır. Bunlar;

- Faster R-CNNs (Girshick et al., 2015)
- You Only Look Once (YOLO) (Redmon and Farhadi, 2015)
- Single Shot Detectors (SSDs) (Liu et al., 2015)

R-CNN'ler derin öğrenme kullanarak nesne tespiti için gelişmekte olan bir yöntem olsa da “daha hızlı” uygulama ile bile (“R” “Bölge Önerisi” anlamına gelir) R-CNN'lerde algoritma 7 FPS sırasıyla oldukça yavaş olabilir. YOLO ise 40-90 FPS işleyebilecek şekilde çok daha hızlı olmasına rağmen arzulanan doğruluk seviyesini yakalayamamaktadır. Bu noktada çalışmada tercih edilen SSD ise bu ikisi arasında bir dengededir. YOLO 'dan daha yüksek doğruluk sağlaması ve gerçek zamanlı uygulamalara uygun olması sebebiyle bu teknik seçilmiştir.

SSD nesne tespiti iki bölümden oluşur:

- Özellik haritalarının ayıklanması,
- Nesneleri tespit etmek için evrişim filtrelerin uygulanması,



Şekil 2 SSD model mimari yapısı [9].

SSD, diğer tespit yöntemlerinden daha fazla yer, ölçek ve en boy oranını daha iyi kapsamak için birçok tahminde bulunur (8732). Bununla birlikte, birçok tahmin nesne içermez. Bu nedenle, sınıf güven puanı 0.1'den düşük olan tüm tahminler elenecektir.

SSD, aynı nesneye işaret eden yinelenen tahminleri kaldırmak için maksimum olmayan baskılama kullanır. SSD tahminleri güven skorlarına göre sıralar. En yüksek güven tahmininden başlayarak SSD, önceden tahmin edilen sınır kutularının aynı sınıf için mevcut tahminle 0.45'ten daha yüksek bir değere sahip olup olmadığını değerlendirir. Bulunursa, mevcut tahmin dikkate alınmaz.

## 2.3. Kamera Yerleşimi

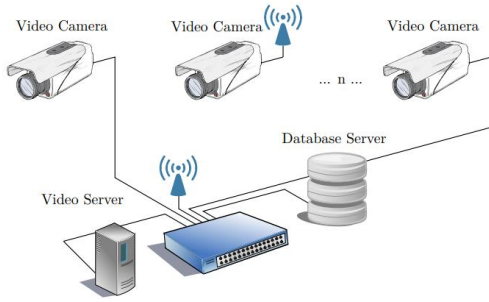
İnsan grupları kameranın görüş alanından geçtiğinde tıkanıklıklardan kaçınmak için tepeden görüş açısına sahip kamera kullanımı genellikle en iyi seçenektir. Bu yerleşim mümkün olduğunca tercih edilmelidir. Şekil 3'te görüldüğü gibi tepeden görüş açısı ile birbirlerine yakın hareket eden insan toplulukları çok daha kolay şekilde ayırt edilebilir. Bu da sistemin performansını büyük ölçüde etkilemektedir.



Şekil 3 Kişi takibinde farklı kamera açıları.



## 2.4. Sistem Mimarisi



Şekil 4 Kişi sayma sistemi için sistem yapısı [10].

Algoritmanın gerçek hayatta gerçekleştirilmesi için gerekli sistem mimarisi Şekil 4’te görülmektedir. Kablolulu veya kablosuz ağ kameralarından elde edilen video akışları standart bir CPU üzerinde çalışma kapsamında tasarlanan kişi sayma algoritmasından geçirilerek elde edilen verilerin anlık olarak takip edilebilir ya da ileride raporlamak için bir veri tabanında saklanabilir. Algoritmanın tek bir kamera ve işlemci içeren herhangi bir cihazla çalışabilir durumda olması uygulanabilirlik noktasında hedeflerin gerçekleştirildiğini göstermektedir.

## 2.5. Arayüz Tasarımı

Tasarlanan arayüz ile programın kolayca başlatılabilmesi, giriş olarak verilecek video akışının kolayca seçilebilmesi, elde edilen verilerin anlık olarak takip edilebilmesi ve bir rapor olarak sunulabilmesi sağlanmaktadır. Sistemin başlatılması için tasarlanan arayüz Şekil 5’te görülmektedir.



Şekil 5 Kişi sayma sisteminin başlatılması için tasarlanan arayüz.

Elde edilen sonuçları anlık olarak takip edebilmek için ise işlenen görüntü üzerine çalışan bir arayüz tasarlandı. Şekil 6’da bunun bir örneği görülebilmektedir. Bu kısımda aşağıdaki veriler çerçeveye yerleştirildi.

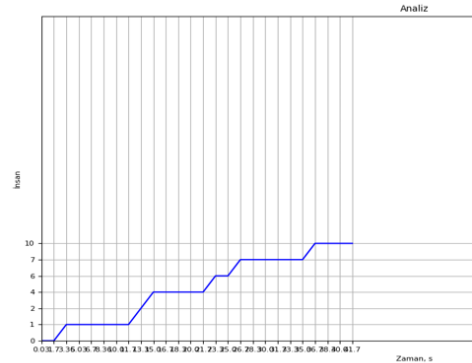
- Nesne kimliği: Her nesnenin sayısal tanımlayıcısı.

- Sentroid: Nesnenin merkezi, bir “nokta” ile temsil edilir.
- Sınır çizgisi: Giriş ve çıkışın temel alınacağı sınır.
- Giriş: Toplam giriş sayacı.
- Çıkış: Toplam çıkış sayacı.



Şekil 6 Gerçek zamanlı olarak kişi sayan sisteme ait arayüz görüntüsü.

Ayrıca bu verileri daha anlamlı hale getirebilmek amacıyla kişi/zaman grafikleri ile kullanıcıya bir rapor sağlandı. Şekil 7’de bunun bir örneği görülebilmektedir.



Şekil 7 Kişi sayma sisteminden elde edilen örnek bir rapor.

## 3. Yazılım Bazında Simülasyonlarla Sistemin İncelenmesi

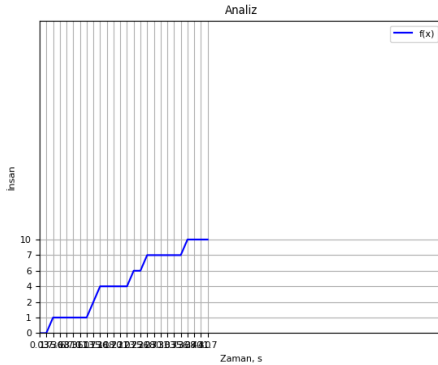
Ölçümler internet üzerinden erişilebilen çeşitli videolar ile yapılmıştır. Bunun sebebi Covid-19 salgını nedeniyle istenen test koşullarında video çekebilmenin imkânsız duruma gelmesidir. Bu nedenle bu kapsamlı bir arama yapılarak çeşitli test kriterlerini karşılayan video akışları bulunup bunlar üzerinde çalışılmıştır. Bu kısımda sırasıyla bu testlere yer verilmektedir.

#### ■ Test Videosu 1



Şekil 8 Birinci test videosuna ait görsel.

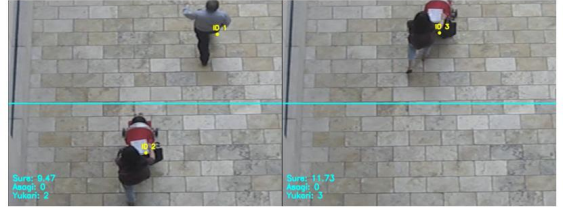
Şekil 8’de görülen birinci test videosu, arka planın yakın renk tonlarından oluşması, kameranın kayıt işlemi sırasında çok fazla titrememesi, içerisinde nesne ve gölge miktarının minimum düzeyde olması, son olarak da kamera açısının yer ile 90 dereceye yakın olması, geliştirilen model ağı için istenilen parametrelere en yakın örnektir. Bu nedenle verimliliğinin yüksek olması beklenilmektedir. Bu video 42 saniyeden oluşmaktadır, içerisinde toplam on nesne geçmektedir. Bunların yedisi aşağı yönlü, üçü yukarı yönlü hareket etmektedir. Bu örnekte sayma işlemi yapıldığında %100 doğruluk değerine ulaşılmıştır. Bu doğruluk oranının bu şekilde çıkmasının nedenleri arasında kamera açısının tepeden konumlanması ve keşfedilmesi gereken insan harici nesne sayısının minimum derecede olması nedeniyle işlemciyi yormaması olarak değerlendirilebilir. Şekil 9’da hangi saniyelerde kaç kişinin geçtiğine dair rapor görülmektedir.



Şekil 9 Birinci test video akışına ait çıktı.

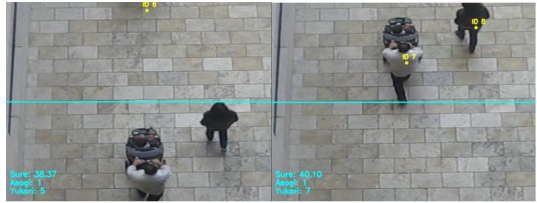
#### ■ Test Videosu 2

İkinci test video akışı birinci video akışı ile aynı ortamda gerçekleştirilmektedir. Ancak bu videoda üç farklı bebek arabasının bulunması nedeniyle, kişi sayma algoritmanızı testi için önemli bir başka kriteri sağlamaktadır. Şekil 10’da görüldüğü üzere bebek arabasıyla geçilmesinden sonra sadece ebeveyn için sayma işlemi gerçekleştirilmiştir. Bunun sebebi bebek arabasının içerisinde bir kişi olmayışıdır. Bu bir hata değildir.



Şekil 10 İkinci test videosuna ait görsel.

Şekil 11’da görüldüğü üzere bebek arabasıyla geçilmesinden sonra sadece ebeveynlerin sayılmasıyla hata gerçekleşmiştir. Buradaki hata, tasarlanan modele göre bebek arabasındaki kişiyi sayabilmek için gerekli eşik değerinin aşılamıyor olmasından kaynaklanmaktadır. Bu eşik değerinin düşürülmesi her ne kadar bu duruma çözüm sağlayabilir gibi görünsede farklı nesneleri de kişi olarak sayabilecek duruma gelmesi nedeniyle hata oranının arttıracaktır. Bu nedenle bu değer test koşulları için (ışık, kamera açısı, kare sayısı) optimize edilmesi gerektiği aşikardır.



Şekil 11 İkinci test videosu için bebek arabasındaki kişinin sayılamadığı durum.

İkinci test video akışına ait son örnekte ise, model ağı bebek arabasındaki çocuğu tespit edebilecek eşik değerini elde edebildiğinden sayma işlemine dahil etmiştir. Bu görseller kıyaslandığında Şekil 11’deki hatanın gölge segmentasyonundaki zorlayıcı koşullardan kaynaklandığı söylenebilir.



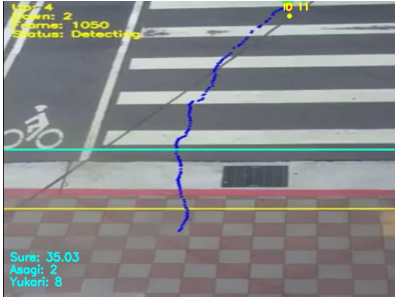
Şekil 12 İkinci test videosu için bebek arabasındaki kişinin doğru tespit edildiği durum.

Sonuç olarak ikinci örnek video da 69 saniye içerisinde toplam 16 kişi sayılmaktadır, ikisi aşağı yönlü, 14 tanesi yukarı yönlüdür. Başarı oranı 17 kişinin 16 kişisinin sayılması sebebiyle %94.11 şeklinde hesaplanmıştır.

#### ■ Test Videosu 3

Seçilen üçüncü örnek video akışında ise Kalman filtresi ve geleneksel yöntemler kullanan ticari bir

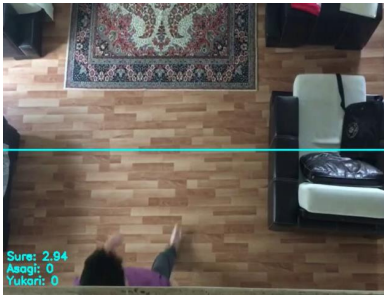
ürüne ait video tercih edilmiştir. Bu teste ait görsel Şekil 18’de görülebilmektedir. Burada sarı renkli yazılar bu yönteme ait sonuçları gösterirken turkuaz renkli yazılar projemize ait sonuçları göstermektedir. Bu test videosu ayrıca modelimizi geleneksel bir yöntem ile karşılaştırma imkânı sağlamaktadır. Kameranin açık bir alanda yerleştirilmesi ve açısı düşünüldüğünde sistemimiz için ideal koşulları karşılamayan bir konumdadır. Ayrıca arka planın farklı desenlerden oluşması nedeniyle görüntü üzerinde fazladan bir iş yükü meydana gelmektedir. Yine de görüntü iki modelden de geçirildiği zaman soldaki modelde toplam altı kişi sayılmıştır, bunların 4’ü yukarı, 2’si aşağı yönlüdür. Aynı görüntüyü tasarladığımız model ağından geçirildiğinde 10 kişi tespit edilmiştir. 35 saniyeden oluşan video akışında, sistemimiz tarafından yapılan tek hata tekerlekli sandalyedeki bir hastanın sayılamamasıdır. Bu hata ile birlikte modelimizin doğruluğunun %90.9 olmaktadır. Videonun alındığı model ise %54.5 doğruluk yüzdesine sahiptir.



Şekil 13 Üçüncü test video akışına ait görsel.

#### ■ Test Videosu 4

Seçilen dördüncü test video akışı, kişilerinin hızlarına, modelin vereceği tepkiyi ölçmek için proje çalışanları tarafından oluşturulmuştur. İlk olarak kameranın görüş açısı içerisinde yavaş bir şekilde geçilmiş daha sonraki aşamalarda kişinin görüş alanında daha hızla geçtiği durumlar ile model test edilmiştir. Şekil 14’te görülen örnekte kamera görüş açısından 2,16 saniyede girilip çıkmıştır. Kameranin görüş açısından bu hızla geçtiği durum için model başarılı bir şekilde kişiyi tespit edip sayma işlemini yapmıştır.



Şekil 14 Kişilerin görüş açısından geçiş hızlarının sistem üzerindeki testi. (Yavaş durum)

Aynı koşullarda gerçekleştirilen diğer bir örnekte ise kameranın görüş açısından 0,87 saniye de girilip çıkmıştır. Bu hızla geçildiği zaman model yine başarılı bir şekilde kişiyi tespit edip sayma işlemini yapmıştır. Sonraki aşamada ise kameranın görüş açısından 0,66 saniye de girilip çıkılarak test gerçekleştirilmiştir. Kişinin bu hızla hareket ettiği durum için model başarısız olmuştur. Nesneyi tespit edememiştir. Kameranin yere oldukça yakın olduğu ve nesneyi tanıyacak süresinin kısıtlı olduğu düşünüldüğünde bu sonuç anlaşılabilir olarak değerlendirilmektedir.

Bütün testler ve sonuçları değerlendirildiğinde kamera açısının modelde tasarlandığı şekliyle yere dik bir komunda olmasının performansı büyük ölçüde etkilediği görülmüştür. Elde edilen sonuçlara göre kameranın görüş açısında çok fazla nesne olması, işlemcinin iş yükünü arttırmakta ve kare hızı başına performansı düşürmektedir. Ayrıca beklendiği gibi, ışıklandırmanın ve görüş açısındaki gölge miktarının da doğruluk yüzdesini düşüren etkenlerden olduğu görülmüştür.

#### ■ Test Videosu 5

Son test videosu ile tasarlanan sistem bir mağazanın girişinden alınan görüntüler ile test edilmektedir. Bu video akışı görüntü içerisinde nesne miktarının fazlalığı, kamera açısının istenenden bir miktar farklı olması, güneş nedeniyle yüzeyden yansıyan ışık miktarı ve aynı anda birçok kişinin görüş içerisinde hızla geçmesi gibi durumlar değerlendirildiğinde sistemin zorlu koşullardaki performansını anlayabilmek için önemli bir örnektir.



Şekil 15 Sistemin bir mağaza girişinden alınan video akışı ile testine ait görüntü.

Bu video akışında 51 kişinin aşağı yönlü hareket ettiği, 25 kişinin ise yukarı yönlü hareket ettiği görülmüştür. Örnek videonun model ağıdan geçirilmesi sonucu Şekil 34’te görüldüğü üzere 48 aşağı yönlü sayma işlemi, 24 yukarı yönlü sayma işlemi yapılmıştır. Doğruluk yüzleri hesaplandığında, yukarı yönlü sayım için doğruluk yüzdesi %96 iken, aşağı yönlü sayım için doğruluk yüzdesi %94,11 olmaktadır. Görüntü içerisinde alınan video akışına ait gereksiz şekiller yer alması

başarı yüzdesini negatif yönde bir miktar etkilemektedir.

Tüm test videoları incelendiğinde, tasarlanan model ağının uygun parametreler verildiği takdirde %94'ün üzerinde doğruluk sağladığı sonucu ortaya çıkmaktadır. Tasarlanan sistem içerisinde kullanılan MobilNet-SSD teknolojisi gerçek zamanlı video akışları üzerinde nesne tespitinde 25-35 kara hızı ile beklenen performansı sağlamaktadır. Nesnenin sınıflandırılması aşamasında ise video işleme başarısı en yüksek olan derin öğrenme algoritması olan Caffe modeli kullanılarak kişi tespitindeki başarı en üst düzeye düzeye taşınmaktadır. Son olarak sentroid izleme ve korelasyon tekniklerinin bir arada kullanılmasıyla oluşturulan hareket takibi ile kişilerin yön tayinleri gerçekleştirilmektedir. Elde edilen sonuçlar bir arayüz ve tablo aracılığıyla kullanıcıya rapor olarak verilmektedir.

#### 4. Sonuçlar ve Tartışma

Kişi sayma teknolojilerinin derinlemesine araştırılmasıyla başlayan proje sürecinde sonuç olarak bir binada bulunan kişi sayısını takip etme probleminin giriş ve çıkışlara yerleştirilen kameralar yardımıyla veya güvenlik kameralarıyla çözüme ulaştırılabilir olduğu anlaşılmıştır. Proje kapsamındaki çalışmalar sonucunda tasarlanan algoritma, gelişmiş kişi algılama doğruluğu için derin öğrenme nesne dedektörleri (MobileNet ve SSD) ve gelişmiş izleme doğruluğu için hem sentroid izleme hem de korelasyon filtreleri dahil olmak üzere iki ayrı nesne izleme algoritması kullanarak, standart bir CPU üzerinde gerçek zamanlı olarak çalışabilen, hem "algılama" hem de "izleme" aşamalarını uygulayarak görüntüye giren veya görüntüden kaybolan kişileri izleyebilen ve hareket yönleri doğrultusunda bir alana giriş ve çıkış verilerini elde edebilen bir sistem tasarlanmıştır. Maliyeti ile uygulanabilirliği ve doğruluk yüzdesi yüksek bu sistem ile kişi sayma teknolojilerinin yaygınlaşacağı, bu teknolojilere olan farkındalığın artacağı ve ileride akıllı şehirler için önemli bir veri sağlayıcı olacağı düşünülmektedir.

Geliştirilen sistemi değerlendirmek için, internet ortamından elde edilen bir dizi görüntü üzerinde bazı analizler yapılmıştır. Elde edilen doğruluk sonuçları, istenen son hassasiyete oldukça yakın oldukları için sonuç tatmin edicidir. Ancak sistemin daha önceden tanımlanmış tüm durumlarda doğru bir şekilde değerlendirilmesi için ele alınması gereken çeşitli senaryolar olduğundan test süreci ileriye yönelik çalışmalar ile devam ettirilmelidir. Çözülmesi gereken ana sorun çeşitli ışık kaynakları ve gölgeler nedeniyle arka plan piksellerinin yanlış algılanabilmesidir. Bir dakikalık test videosunda sistem %5 ila 8 hata oranları ile çalışmaktadır. Sonuçlar doğrultusunda sistem mağaza, lokanta,

alışveriş merkezi, sınıflar vb. kapalı alanlarda kişi yoğunluğunu takip edebilir seviyededir. Ayrıca tüm vücudu tespit edebildiğinde ve çok fazla tıkanıklık olmaması durumunda, havalimanı, tren istasyonları gibi kalabalık mekanlarda da çalışabilir. Eğitilmiş derin öğrenme algoritması sayesinde hayvanları, bisikletleri ve benzer araçları saymamaktadır.

Bunların yanı sıra, elde edilen bilgiyi anlamlı hale getirme problemi ve bütün bir kampüsün ya da bütün bir şehrin yoğunluk haritasını oluşturabilme problemi ise projenin geliştirme sürecinde devam etmektedir. Örneğin sistem üzerinde geliştirmeler yapılarak Ege Üniversitesi Merkez kampüsündeki belirli yapıların ana giriş ve çıkışlarına bu sistem kurulumu büyük ölçekli durumlar için faydaları araştırılabilir. Sonrasında ise Covid-19 salgını nedeniyle kapalı alanlardaki kişi yoğunluğunu kontrol edebilmek amacıyla çalışmalar devam ettirilebilir.

#### Referanslar

- [1] G. Sexton, X. Zhang, G. Redpath, and D. Greaves, "Advances in automated pedestrian counting," in Proc. European Convention Security and Detection, 1995, pp. 106–110.
- [2] J. Segen, "A camera-based system for tracking people in real time," in Proc. 13th Int Pattern Recognition Conf, vol. 3, 1996, pp. 63–67.
- [3] Terada K., Yoshida D., Oe S., Yamaguchi J., "A Counting Method of the Number of Passing People Using a Stereo Camera".
- [4] Chang Q., Song Z., Shi R., Xu J., "A People counting method based on multiple cameras information fusion", 2013.
- [5] J. Shao, K. Kang, C. C. Loy, X. Wang, "Deeply Learned Attributes for Crowded Scene Understanding", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15), 4657-4666, 2015.
- [6] Y. Zhang, D. Zhou, S. Chen, S. Gao, Y. Ma, "Single Image Crowd Counting via Multi-Column Convolutional Neural Network", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16), 589-597, 2016.
- [7] D. Oñoro, R. R. Lopez-Sastre, "Towards Perspective-Free Object Counting with Deep Learning", Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2016.
- [8] C. Zhang, H. Li, X. Wang, X. Yang, "Cross-Scene Crowd Counting via Deep Convolutional Neural Networks", In Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition (CVPR'15), 833-841, 2016.

[9] Satish D. Pore, B. F. Momin, "Bidirectional People Counting System in Video Surveillance", 2017.

[10] G. Monteiro, J. Marcos, M. Ribeiro, and J. Batista, "Robust segmentation for outdoor traffic surveillance," in Proc. 15th IEEE Int. Conf. Image Processing ICIP 2008, 2008, pp. 2652–2655.

## İÇİNDEKİLER

ÖNSÖZ.....	I
ÖZET MAKALE.....	II
İÇİNDEKİLER.....	XII
TABLO VE ŞEKİL LİSTELERİ .....	1
ÖZ.....	2
1. GİRİŞ.....	3
1.1. Projenin Getirdiği Yenilik Unsurları .....	3
1.2. Literatür Özeti.....	5
1.3. Derin Öğrenme Mimarileri ile Kişi Sayma.....	8
2. GELİŞME .....	10
2.1. Derin Öğrenme.....	10
2.2. Derin Öğrenme ve OpenCV.....	11
2.3. CNN (Evrişimli Sinir Ağları).....	11
2.4. SSD (Single Shot MultiBox Detector).....	12
2.5. MobileNet .....	14
2.6. Caffè.....	15
2.7. Arka Plan Bakımı.....	15
2.8. Gölge Segmentasyonu .....	16
2.9. Bilgisayarlı Görme Araçları.....	17
2.10. Kullanılan OpenCV Kütüphaneleri.....	18
2.11. Algoritma Mimarisi .....	19
2.12. Arayüz Tasarımı.....	21
2.13. Sistem Mimarisi .....	23
2.14. Kamera Yerleşimi .....	23
2.15. Ölçümler .....	24
2.16. Performans .....	29
3. SONUÇ.....	31
REFERANSLAR.....	32
LİSANS BİTİRME PROJESİ ÖZDEĞERLENDİRME FORMU .....	34

## TABLO VE ŞEKİL LİSTELERİ

Şekil 1 Toplu normalizasyon ve ReLU ile standart evrimsel tabaka [8].	VI
Şekil 2 SSD model mimari yapısı [9].	VI
Şekil 3 Kişi takibinde farklı kamera açıları.	VI
Şekil 4 Kişi sayma sistemi için sistem yapısı [10].	VII
Şekil 5 Kişi sayma sisteminin başlatılması için tasarlanan arayüz.	VII
Şekil 6 Gerçek zamanlı olarak kişi sayan sisteme ait arayüz görüntüsü.	VII
Şekil 7 Kişi sayma sisteminden elde edilen örnek bir rapor.	VII
Şekil 8 Birinci test videosuna ait görsel.	VIII
Şekil 9 Birinci test video akışına ait çıktı.	VIII
Şekil 10 İkinci test videosuna ait görsel.	VIII
Şekil 11 İkinci test videosu için bebek arabasındaki kişinin sayılamadığı durum.	VIII
Şekil 12 İkinci test videosu için bebek arabasındaki kişinin doğru tespit edildiği durum.	VIII
Şekil 13 Üçüncü test video akışına ait görsel.	IX
Şekil 14 Kişilerin görüş açısından geçiş hızlarının sistem üzerindeki testi. (Yavaş durum)	IX
Şekil 15 Sistemin bir mağaza girişinden alınan video akışı ile testine ait görüntü.	IX
Şekil 16 Yapay sinir ağlarında örnek bir fonksiyon [14].	11
Şekil 17 SSD model mimari yapısı [17].	13
Şekil 18 Toplu normalizasyon ve ReLU ile standart evrimsel tabaka [18].	14
Şekil 19 Caffe modeli ile nesne tanımlama [18].	15
Şekil 20 RGB renk uzayında önerilen 3D koni modeli [7].	17
Şekil 21 Bilgisayarlı görme uygulaması olarak kişi sayan kamera [8].	18
Şekil 22 Kişi sayma sisteminin başlatılması için tasarlanan arayüz.	21
Şekil 23 Gerçek zamanlı olarak kişi sayan sisteme ait arayüz görüntüsü.	22
Şekil 24 Kişi sayma sisteminden elde edilen örnek bir rapor.	22
Şekil 25 Kişi sayma sistemi için sistem yapısı [15].	23
Şekil 26 Kişi takibinde farklı kamera açıları.	24
Şekil 27 Birinci test videosuna ait görsel.	24
Şekil 28 Birinci test video akışına ait çıktı.	25
Şekil 29 İkinci test videosuna ait görsel.	26
Şekil 30 İkinci test videosu için bebek arabasındaki kişinin sayılamadığı durum.	26
Şekil 31 İkinci test videosu için bebek arabasındaki kişinin doğru tespit edildiği durum.	27
Şekil 32 Üçüncü test video akışına ait görsel.	28
Şekil 33 Kişilerin görüş açısından geçiş hızlarının sistem üzerindeki testi. (Yavaş durum)	28
Şekil 34 Sistemin bir mağaza girişinden alınan video akışı ile testine ait görüntü.	29

## ÖZ

Günümüzde her türlü veri ve bu verilerin doğru analiz edilip, kullanılması çok daha önem arz eden bir seviyeye ulaşmış durumdadır. Belirli bir alana giriş ve çıkış yapan kişileri sayabilen kamera sistemleri ise bu bağlamda önemi gittikçe artan bir veri sağlayıcı olarak görülebilir. 2020 ilk çeyreği ile ortaya çıkan ve yayılımı artan COVID-19 virüs salgını nedeniyle de belirli bir alandaki kişi yoğunluğunu takip etmenin önemi artarken bunu sağlayan sistemler maliyetli olmaları nedeniyle henüz yaygınlaşmamış durumdadır.

Bu projede belirli bir bölgedeki kişi yoğunluğunu takip edebilmek için bir alana giren ve çıkan kişileri tespit edip, toplam giriş ve toplam çıkış verilerini raporlayan ve standart bir CPU üzerinde gerçek zamanlı olarak çalışabilen bir sistem tasarlanmaktadır. Tasarımda özellikle düşük maliyet-düşük hata oranı gözetildi ve herhangi bir bölgeye kolayca uygulanabilir olması hedeflendi. Python üzerinde geliştirilen algoritmalarda, standart bilgisayar görme/görüntü işleme işlevleri için OpenCV ve sayımı yapılacak kişileri tespit etmek için ise derin öğrenme kütüphanesi MobileNet-SSD (Single Shot Object Detection) Caffe (Convolutional Architecture for Fast Feature Embedding) modelinden yararlanıldı. Tespit edilen kişilerin gelişmiş izleme doğruluğu ile takip edilebilmesi için ise hem sentroid izleme hem de korelasyon filtreleri dahil olmak üzere iki ayrı nesne izleme algoritması kullanıldı. Derin öğrenme tekniklerinin sağladığı yüksek doğruluk ve gelişen teknoloji ile aşılacak işleme gücü sınırları nedeniyle geleceğin kişi sayma teknolojileri için akla yatkın bir karar olarak değerlendirilebilir.



# 1. GİRİŞ

Günümüzde her türlü veri ve bu verilerin doğru analiz edilip, kullanılması çok daha önem arz eden bir seviyeye ulaşmış durumdadır. Belirli bir alana giriş ve çıkış yapan kişileri sayabilen kamera sistemleri ise bu bağlamda önemi gittikçe artan bir veri sağlayıcı olarak görülebilir. Bu sistemler, deprem, yangın vb. acil durum senaryolarında güvenlik ve kurtarma operasyonlarının planlanmasında ve COVID-19 virüs salgını nedeniyle de belirli bir alandaki kişi yoğunluğunu takip etmede yardımcı bir araç olarak kullanılabileceğinden bu alanda ilerici yöntemler kullanarak ve uygulanabilirliği yüksek bir sistem tasarlanmasının önemi git gide artmaktadır. Öyle ki her erişim noktasının üstüne kurulan kişi sayıcıları ile giren veya çıkan tüm insanların iki yönlü bir sayımı gerçekleştirilebilir. Tüm girişleri ve çıkışları gerçek zamanlı olarak karşılaştırarak, bir odanın, belirli bir katın veya tüm bir binanın doluluk derecesi belirlenebilir, o bölge için yoğunluk dağılım haritası oluşturulabilir. Bir binanın tahliyesinde veya bölgesel bir afet durumunda ise o sırada binadaki kişilerin tam sayısı hakkındaki bu değerli bilgiler acil durum ekiplerine sunulabilir. En kapsamlı haliyle ise akıllı şehirler için anlık yoğunluk bilgisi sağlayacak bu sistem ile ulaşım, aydınlatma, havalandırma, güç tüketimi gibi birçok farklı alanda optimizasyon sağlanabilir.

## 1.1. Projenin Getirdiği Yenilik Unsurları

Mevcut sistemlerin maliyetleri ve güvenilirliklerine dair endişeler bu teknolojinin yaygınlaşmasına engel olmaktadır. Çalışmamızda bu sistemleri yüksek doğruluk oranları ve daha düşük maliyetler ile optimize ederek bu alandaki açığa kapatmak ve kesinlik gerektiren durumlara uygun hale getirmek hedeflenmektedir. Bu çalışmada belirli bir bölgedeki kişi yoğunluğunu takip edebilmek için bir alana giren ve çıkan kişileri tespit edip, toplam giriş ve toplam çıkış verilerini raporlayan ve standart bir CPU üzerinde gerçek zamanlı olarak çalışabilen bir sistem tasarlanmaktadır. Python üzerinde geliştirilen algoritmalarda, standart bilgisayar görme/görüntü işleme işlevleri için OpenCV ve sayımı yapılacak kişileri tespit etmek için ise derin öğrenme yapıları MobileNet ve SSD (Single Shot Object Detection) Caffee (Convolutud Architecture for Fast Feature Embedding) modelinden yararlanıldı. Bu iki tekniğin birlikte kullanılması kısıtlı kaynağa sahip cihazlardaki gerçek zamanlı performansı büyük ölçüde arttırmaktadır. Tespit edilen kişilerin gelişmiş izleme doğruluğu ile takip edilebilmesi için ise hem sentroid izleme hem de korelasyon filtreleri dahil olmak üzere iki ayrı nesne izleme algoritması kullanıldı. Derin öğrenme tekniklerinin sağladığı yüksek doğruluk ve gelişen

teknoloji ile aşılan işleme gücü sınırları nedeniyle geleceğin kişi sayma teknolojileri için akla yatkın bir karar olarak değerlendirilebilir.

Bütün bunlar için geliştirilecek algoritma anlık olarak görüntüyü işleyebilmeli, birden fazla giriş ve çıkış anında hata yapmamalı, gölgelerden ve arka plandan bağımsız olarak güvenilir veri sağlayabilmelidir. Kamera görüntüsü yukarıdan alınacağı için farklı saç rengine, kafa şekline sahip insanlar için algoritmalar geliştirilmeli ve bisiklet, bebek arabası, çanta gibi cisimlerin sayımı etkilememesi sağlanmalıdır. Bunun yanında bu bilginin acil durumlarda kullanılabilmesi için internet tabanlı bir iletişim sistemi tasarlanmalı, uygun görsel ve teknik raporlamalar oluşturulmalıdır. Mümkün olduğunca, mevcut güvenlik kameralarını kullanmak istenildiğinden, bölge izleme (region tracking) kullanarak geleneksel bir yöntem kullanmayı tercih edildi. Bu yöntem, daha önce bahsedilen diğer yöntemlerden en esnek olmasının yanı sıra başlangıçta minimum kalibrasyon ve eğitim gerektirir.

### **Hedefler**

- Çalışma kapsamında binalarda bulunan kişi sayısını, anlık olarak yüksek doğrulukla verebilen, uygulanabilir bir sistem gerekmektedir.
- Sistemin mevcut güvenlik kameralarına entegre edilebilir ve mevcut internet altyapısını bağlanabilir olması gerekmektedir.
- Algoritma hızlı, basit ve güvenilir olmalıdır.
- İşlenen bilgilere acil durumlarda dahi internetten kolayca ulaşabilmek için sistemin bir Ethernet ağının parçası olması gerekmektedir.
- Sistem yüksek kullanılabilirlik özelliklerine sahip olmalıdır. Sistemin 24 saat / 7 gün iş başında gözetimsiz çalışabilmesi gerekmektedir.
- Görüş alanı içerisinde veya sayım çizgisi yakınlarda uzun süre beklemeler, tekrarlı giriş-çıkışlar ihmal edilebilmelidir.

### **Zorluklar**

- Acil durum anlarında çok sayıda insan geçecektir.
- Sayım bölgesinin yanında çok sayıda insan olabilir. (Sayım bölgelerinde beklemeler olacaktır.)
- Ana girişteki aydınlatma koşulları ya güneşin gün içindeki konumunun değişmesi ya da binanın çalışma kuralları nedeniyle tutarsız olacaktır. Ana girişteki sayım kritik önem arz

etmektedir. Bu alandaki ziyaretçileri saymak standart bir kapıdan veya düz bir koridordan geçen tek bir kişiyi tespit etmemiz gereken basit bir durumdan çok daha zordur.

- Birbirine yakın bir grup içerisindeki her bireyin takip ettiği yolları hassas bir şekilde izleyebilen bir sistem tasarlamak zordur.
- İzleyeceğimiz alanda çok parlak ve çok karanlık alanlar olabileceğinden. Görüntünün her pikselinde iyi bir piksel bilgisine sahip olmak için, kamera çok iyi bir dinamik aralığa sahip olmalı ve bu nedenle, her piksel için iyi bir renk bilgisine sahip olmalıdır.

Bu hedefler ve zorluklar doğrultusunda bölüm 1.2.'deki şekilde literatürdeki çalışmalar derinlemesine incelendi ve daha önce gelişme raporunda bahsedildiği şekilde derin öğrenme tabanlı bir sistemde karar kılındı. OpenCV kütüphanesi görüntü işleme fonksiyonlarında büyük kolaylık sağlaması nedeniyle tercih edildi. Sistemdeki bir diğer önemli kriter ise güvenilir veri sağlayabilmek ve bunu acil durum anlarında raporlayabilmektir. Bu bağlamda ise yine oldukça kapsamlı bir literatür taraması yapılarak bölüm 2.5. ve 2.6'da bahsedilen derin öğrenme tabanlı MobileNet SSD Caffe modeli tercih edildi.

## **1.2. Literatür Özeti**

Proje kapsamındaki hedefler doğrultusunda literatürdeki çalışmalar detaylı şekilde incelenmiştir. Açık ve/veya kapalı alanlarda insanları bulmak ve izlemek geniş bir araştırmanın hedefi olmuştur. Birçok teknik bu pazarın gerekliliğini az çok başarılı bir şekilde ele almaktadır. İzlemenin gerçekleştiği koşullar, başarısı için kritik öneme sahiptir. Gerçekleştirilen projeye ilgili bazı çalışmalar aşağıdadır.

Tepegöz kamera görüntüsü kullanarak, insanların hareket yönlerini belirleyen ve sanal bir çizgiyi geçerken insanları sayan bir sistem tasarlanmıştır [1]. Tepeye yerleştirilen kamera ile, insan grupları kameranın görüş alanından geçtiğinde oluşabilecek tıkanma sorunu önlenir ve insanların yönünü belirlemek için bir uzay-zaman görüntüsü kullanılır.

Hashimoto sisteminde özel bir görüntüleme sistemi kullanarak (IR duyarlı seramikler, mekanik doğrama parçaları ve IR şeffaf lensler) insanları %95 oranında doğrulukla, 2 metrelik bir kapıdan geçerken sayabilen bir sistem geliştirdiler [2]. Ancak sistemin maliyeti nedeniyle uygulanabilirliği düşük seviye kalmıştır.

Tesei insanları izlemek ve oklüzyonları yönetmek için görüntü segmentasyonu, ilgilenilen bölgeleri çıkarmak için arka plan ayırt etme yöntemleri kullanır [3]. Damlları izlemek için

damla alanını, yüksekliği ve genişliği, sınırlayıcı kutu alanını, çevreyi ve ortalama gri seviyesini kullanırlar. Tüm bu özellikleri zaman içinde ezberleyerek, algoritma kutuları tıkanıklıktan kaynaklanan kabarcıkların birleştirilmesi ve ayrılması sorununu çözer. Kabarcıklar oklüzyon sırasında birleştiğinde, diğer özelliklerle yeni bir damla oluşturulur. Ancak, bu yeni damla, onu oluşturan kabarcıkların özelliklerinden gelen verileri saklar. Damlalar geri ayrıldığında, algoritma orijinal etiketleri orijinal damlalara atayabilir.

Shio'nun çalışması, insanın algısal gruplamasını simüle ederek arka plan segmentasyon algoritmasındaki insanların oklüzyonlarının saptanması üzerine odaklanmıştır [4]. İlk olarak, algoritma, çerçeve farkını kullanarak hareketin bir tahminini hesaplar ve bu verileri, arka plan çıkarma algoritmasının tıkanmış kişiler arasındaki sınırı belirlemesine yardımcı olmak için kullanır. Bu yöntem, genişlik, yükseklik, yön ve birleştirme/bölme adımı hakkında bilgiye sahip olan olasılıklı bir nesne modeli kullanır. Bir nesne modeli kullanmanın, segmentasyon için iyi bir gelişme ve oklüzyon problemini çözmenin olası bir yolu olduğu bulunmuştur. Ancak algısal gruplamanın kullanılması, örneğin, aynı yönde hareket eden bir grup insanın neredeyse eşit olduğu bazı durumlarda tamamen etkisizdir.

Basit ve etkili bir yaklaşım olarak Sexton basitleştirilmiş bir segmentasyon algoritması kullanır [5]. Sistemlerini Parisli bir tren istasyonunda test ettiler ve %1 ila %20 arasında bir hata aldılar. Sistemleri, insanları arka plandan izole etmek için basit bir arka plan çıkarma yöntemi kullanır ve çerçeveler arasındaki eşleşmeyi yapmak için sadece damlaların ağırlık merkezlerini alır. Sistemin sağlamlığını artırmak için arka plan modeli sürekli olarak güncellenerek çevresel değişikliklerin sistem üzerindeki etkisini azaltır. Kamera, oklüzyonları azaltan ve damla tespit problemini basitleştiren üstten bir konuma asılmıştır.

Segmentasyon sonrası görüntü işleme üzerine yoğunlaşılacak bir başka yaklaşımda, damlaları basit bir arka plan çıkarma yöntemiyle ayıklar ve sonra çerçeveler arasındaki özelliklerini izlerler [6]. Her damlanın yolları saklanır ve sanal bir çizgiyi geçerken kesişim noktasını ve yönünü tespit etmek için kullanılır. Bu sistem oklüzyon problemleriyle uğraşmaz, bu nedenle kalabalık ortamlarda performansı büyük ölçüde azaldı. Yol kesişme algoritması performansı kalabalık ortamlarda da olumsuz etkilendi.

Gary Conrad ve Richard Johnsonbaugh, oklüzyon problemini basitleştirmek için bir baş kamerası kullanmaktadır [7]. Aydınlatma değişikliği problemlerinin üstesinden gelmek için arka plan çıkarma kullanmak yerine ardışık kare farklılıkları kullanırlar. Sahip oldukları sınırlı

hesaplama gücü nedeniyle, algoritmaları görüntünün yalnızca küçük bir dikdörtgen alanında hareket edecek şekilde tasarlanmıştır. 7491 kişide %95,6 doğruluk oranı elde edebildiler.

Hoprasert toplanan görüntülerden elde edilen istatistikleri kullanarak parlak kısımları ve gölgeleri algılamayı önerdi [8]. Parlaklık ve renk sapması, pikselleri dört sınıfa ayırmak için dört eşik değerle kullanılır. Ortalama değeri [9] 'de referans görüntü olarak kullanan yöntem, dinamik arka plan için uygun değildir. Ayrıca, eşik değerleri, belirli bir algılama oranıyla parlaklık bozulması ve renk sapması histogramı temel alınarak tahmin edilir ve piksel değerlerinden bağımsız olarak tüm piksellere uygulanır. Bu nedenle, koyu piksel değerini gölge olarak sınıflandırmak mümkündür. Ayrıca, arka plan geçmişini kaydedemez.

İzlenen nesneleri temsil etmek ve eşleştirmek için nicelleştirilmiş yerel özellik tanımlayıcılarının histogramı kullanılmıştır [10]. Bu yöntemin, tanımlayıcıların önceden çıkarılabildiği görüntü alma uygulamalarında nesne eşleştirme ve sınıflandırma için etkili olduğu kanıtlanmıştır. Bu sistem gerçek zamanlı gereksinimlere yaklaşır.

Satish, insan sayma sisteminin nerelerde kullanılabileceğini ayrıntılı bir şekilde anlatılmıştır [11]. Güvenlik uygulamaları için tahliye durumunda olduğu gibi, herhangi bir zamanda gözetleme alanında kaç kişinin bulunduğunu bilmek, yaya trafiğini yönetmek, turistlerin akış tahminlerini bilmek önemlidir. Bu sistemler için İnsanları tespit etmek ve yönlerini analiz etmek iki önemli unsurdur. Sistem öncelikle insanları tespit eder, daha sonra kalman filtresi ile güzergahları takip ederek giriş mi çıkış mı yaptıklarını tahmin ederek sayacı günceller.

Yüz tanıma dayalı insan sayma sistemi, farklı ışık ve arka planlarda çalıştırılarak performansları gözlemlenmiştir [12]. NCC renk uzayına dayanarak, adayın cilt rengi ve boyutuna bakarak tespit eder. Yüz tanımadan sonra, kişinin yüzünü takip edip yüzünü sayarak izler. Kalabalık ortamlar için %80 doğruluk oranı korunmuştur. Görüntü işlemenin avantajları sıralanmıştır. Bu makalede insan tabanlı sayma sistemi kullanılmaktadır. Önerilen sistem dört bölümden oluşur. Hareketli kalabalığı bölümlendirme, ten rengi tespiti, yüz tanıma, izleme ve sayma şeklindedir. Bu bölümlendirmeler makale içinde ayrıntılı bir şekilde bölümler halinde anlatılmıştır. Örnekler verilmiştir. Sonuç olarak insan yüzüyle algılama yöntemi kapalı alanda %99 oranında doğruluk payı varken dışarıda %80 oranına kadar düşmektedir. İnsan yüzü, maske ve saç gibi etkenler hataya sebep olmaktadır.

Stereo kamera kullanarak geçitten geçen insanları otomatik olarak sayma yöntemini önerilmiştir [13]. Önerilen yöntemde, stereo kamera kapının tavanından asılır ve kameranın

optik eksenini, geçen insanların sadece yukarıdan gözlenebilmesi için ayarlanır. Çünkü bu sistem düzenlemesinde, eğer kapıda kalabalık bir insan grubu varsa, o zaman geçen insanların görüntü verileri, elde edilen görüntüler üzerinde birbiriyle örtüşmez. Ek olarak, geçen insanların her bir yüksekliği, piramitleştirmeye dayalı algoritma uygulanarak ölçülmüştür. Stereo kamera kullanarak yapılan bu sistemin en büyük avantajı kamera tam tepede olduğu için birbirini kapatan görüntüler olmayacaktır. Bu sayede doğruluk payı daha yüksek ölçümler yapılmaktadır.

Kişi sayma sistemlerinde yeni metot olarak birden fazla kamera kullanmayı önermektedir [14]. Bu yöntemde tek bir kamera için uyarlanabilir kayan pencere (adaptive sliding window) algoritması tasarlanır ve eşyazımı (homography) teorisine göre birçok kamera arasındaki ortak görüntüler (aynı sahnenin farklı görünüşleri) hesaplanır. Tıkanma faktörünü göz önünde bulunduran nesne eşleştirme stratejisi, tek kameranın sonuçlarını iyileştirmek ve birden fazla kameranın sayımını elde etmek için tasarlanmıştır. Deney sonuçları, bu yöntemin karmaşık durumlarda iyi performans gösterdiğini göstermiştir. Ancak gerçek zamanlı performansı düşüktür.

### **1.3. Derin Öğrenme Mimarileri ile Kişi Sayma**

Derin öğrenme alanındaki güncel çalışmalar ve uygulamalar kişi sayma sistemlerinde klasik yöntemlerle karşılaşılan zorlukların aşılmasına yardımcı olmaktadır. Klasik yöntemlerde özneliklerin ayrıştırılması büyük miktarlarda tecrübe ve deney gerektirmekteydi. Ancak evrişimli sinir ağları ile insan görme sisteminin modellenmesi ile bilgisayarlı görüş teknolojilerinde yüksek doğrulukta sonuçlar gerçek zamanlı olarak elde edilebilir noktaya gelmiştir. Bunun için geliştirilen farklı model ve kütüphanelerin OpevCV ile tek bir noktada çalıştırılabilir olması da bu tekniklerin kullanılabilirliğini artırmıştır. Bu modeller ile görüntülere ait özellikler daha derin katmanlarda öznelikleri seçmek için işlem yapmayı gerek duymadan öğrenilebilmektedir.

Zhang ve çalışma arkadaşları 2015 yılında derin evrişimli sinir ağı modelinden yararlanarak kişileri çapraz sahne yöntemiyle saymayı önermiştir. Eğitim verisi, kişi sayımı ve kişi tespiti olmak üzere iki hedef içermektedir [15].

Boominathan ve çalışma arkadaşları CrowNet sistemini önererek yine evrişimli bir sinir ağı modeli kullanmıştır. Bir alanın yoğunluk haritasını tahmin edebilmek amacıyla derin ağ ve evrişimli sinir ağlarından beraber yararlanmışlardır. Bu model ile semantik bilgiler (beden,

yüz) yakalanabilmiştir. Veri arttırma (data augmentation) ile eğitim örneklerindeki eksikler kapatılmıştır. İlk veri arttırma ile çoklu kişi örneklemedeki sorunlar çözülmüş, ikincisinde ise hataları azaltmak ve yoğun durumlar için modelin performansı iyileştirilmeye çalışılmıştır. Buradaki derin öğrenme mimarisi, VGG-16 ağına benzemektedir [16].

2016 yılında, Zhang ve çalışma arkadaşları, kişi sayımını değişen perspektiflerden oluşturulan görüntüler ile tahmin edebilmek amacıyla evrişimli sinir ağı yapısı uygulamıştır. Çok katmanlı evrişimli sinir ağı modeli görüntüyü çoklu yoğunluk haritasına eşleştirmektedir. Bu yapının önemli özelliği, rastgele bir çözünürlük ve boyutta giriş yapılabilmesini sağlamasıdır. Çeşitli filtreler ile görüntüdeki değişimlere karşılık gelen özelliklerin öğrenilmesi sağlanır. Böylece bu yöntem ile herhangi bir boyuttaki görüntüdeki bozulmaları önleme imkanı ortaya çıkar [17].

Onoro, Hidra evrişimli sinir ağı (Hydra CNN) adında bir model ve ölçek temelli bir sayma modeli geliştirmiştir. Derin ve evrişimli tüm sinir ağlarını içinde barındıran bu sistem ile kişi yoğunluğu hesaplanabilmiştir. Altı katmanlı bir yapıdan meydana gelen CCNN (counting convolutional neural network-CCNN) modeli bu katmanların paralel olarak işlenmesiyle hedefe ulaşmaktadır. Nesne ve kişi sayma konusunda başarılı bir model olarak değerlendirilmektedir [18].

## 2. GELİŞME

Bu bölüme projede kullanılan yöntem ve teknikleri teorik açıdan inceleyerek başlanmaktadır. Ardından bu yöntem ve teknikler kullanarak tasarlanan sistemden bahsedilmektedir. Sonraki kısımda ise sistemin ölçümleri ve performans çıktılarına yer verilmektedir.

### 2.1. Derin Öğrenme

Görüntü işlemede bir ortamdaki canlının kedi mi köpek mi olduğunu ayırt etmek için belirli özniteliklere dikkat etmemiz gerekir. Bunlar kulak tipi, ağız yapısı, yüz tipi, rengi ve tüy yapısı gibi; derin öğrenmede bunlardan farklı olarak bu öznitelikler katmanlar arasında kendiliğinden öğrenilmektedir.

Derin öğrenmede kullanılan yapıların temelinde yapay sinir ağları vardır. Yapay sinir ağları biyolojik sinir ağları model alınarak oluşturulmuştur. Burada sinapslardan gelen uyarılar farklı ağırlık katsayıları ile çarpılarak çıkışa yönlendirilir. Tekli veya çoklu katmanları olabilirler. Eğer birden daha fazla katmanı varsa çoklu katman olarak değerlendirilmelidir.

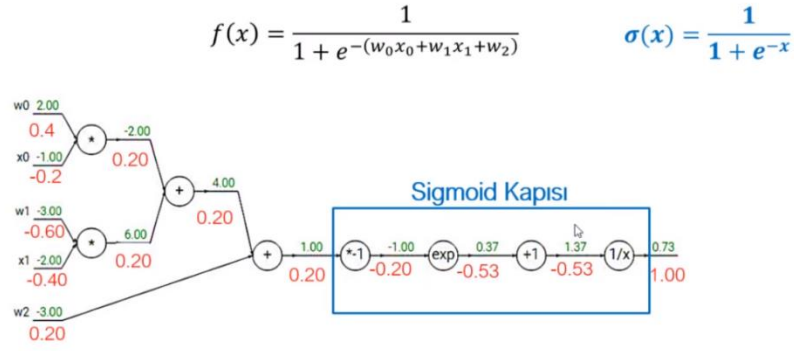
Sinir ağlarının formülü genelleştirildiğinde;

$$g(x|W_i, w_i, w_{i0}) = x^T W_i x + w_i x + w_{i0}$$

şeklinde elde edilir.  $x^T$  yapılmasının sebebi ağırlıklar ile aynı türden vektör olamamalarıdır. Bu durumda işlem yapılamayacağından girişin transpozu alınmaktadır.

Şekil 16'da örnek bir fonksiyon verilmiştir. Burada yeşil ile yazılanlar her adımın sonucudur. Kırmızı ile yazılanlar ise sonuçların türevleri alınarak geriye doğru gidilmesiyle elde edilen değerlerdir. Bu değerler bir sonraki işlemde ağırlık değerlerinin güncellenmesini sağlayacaktır. Bu olay hata oranı istenilen değere gelinceye kadar tekrarlanmaktadır. Makine öğrenmesinin arkasındaki akıl buna dayanmaktadır.





Şekil 16 Yapay sinir ağlarında örnek bir fonksiyon [14].

Özellikle düzenli ve etiketlenmiş veri çok kıymetlidir, eğer etiketlenmiş veri çok fazla değilse verimlilikleri pek fark etmemektedir. Fakat eğer etiketlenmiş veri miktarı çok fazlaysa derin sinir ağlarını kullanmak performansı büyük miktarda arttırmaktadır. Bu projede derin öğrenme kullanılmasıdaki motivasyonlardan bir diğeri de budur.

## 2.2. Derin Öğrenme ve OpenCV

Derin öğrenme tarafına OpenCV 2.4 sürümü ile yer verilmeye başlanmıştır. Görüntü sınıflandırma için birçok yeni kütüphane geliştirilmeye devam ederken, bu kütüphaneler ile modellerin oluşturulup içe aktarılabilmesi, farklı sınıflandırıcı ağların oluşturulabilmesi OpenCV 3.3 sürümü ile farklı kütüphanelerin birlikte çalışabildiği bir noktaya evrildi. Bu sürüm ile Caffé, TensorFlow, Torch DarkNet gibi yapılar için destek sağlandı. Ayrıca popüler; AlexNet, GoogLeNet, ResNet, SqueezeNet, VGG, ENet, VGG-based SSD, MobileNet-based SSD gibi ağlar içinde destek sağlanmaktadır.

## 2.3. CNN (Evrişimli Sinir Ağları)

Sinir ağlarının asıl konusu canlıların karar mekanizmalarının ve sinir sistemlerinin matematiksel olarak modellenmesidir. Sistemlere insansı bir karar mekanizması kazandırmak için bazı teknikler kullanılmaktadır. Evrişimli sinir ağları (convolutional neural networks-CNN), bilgisayarlı görü alanında en başarılı sonuçları vermesi nedeniyle günümüzde yaygın olarak kullanılmaktadır.

Geldiğimiz noktada, milyarlarca parametreyi hesaplamak için veriden farklı bilgiler çıkarmak gerektiği anlaşıldı. Görüntü Sınıflandırma (Image Classification), Nesne Bulma (Object Detection), Nesne Takip Etme (Object Tracking), Doğal Dil İşleme (Natural Language Processing), Stil Transferi (Style Transferring) gibi problemlerin çözümü için gereken

motivasyon yine yapay sinir ağı tabanlıdır. Verinin büyümesi ve veriden daha anlamlı bilgilere ulaşmak, öznetelik kestirimleriyle ilgili optimizasyon yapmayı zorunlu kılmaktadır. Klasik bir yapay sinir ağı modeliyle nöronlar ve katmanlar arasındaki bağlantılar ve öğrenilen parametreler çok büyük hesaplama zorlukları ortaya çıkarmaktadır. Bu yüzden sistemin başarısı koruyarak işlem yükünü azaltmak için CNN kullanılmaktadır.

### **2.3.1.İki Boyutlu Evrişim**

İki boyutlu bilgiye uygulanacak olan filtrenin x ve y eksenine göre simetrisi alınır. Tüm değerler matriste çarpılır ve tüm değerlerin toplamı çıkış matrisinin ilgili elemanı olarak kaydedilir. Buna çapraz korelasyon ilişkisi de denir. Giriş verisi (örneğin; görüntü) tek kanallı iken bu işlem basitçe yapılabilmektedir. Ancak giriş verisi farklı formatlarda ve kanal sayısında da olabilir.

### **2.3.2.Çok Kanallı Evrişim**

Renkli görüntüler, (RGB) 3 kanaldan meydana gelmektedir. Bu koşulda da evrişim işlemi 3 kanal için yapılır. Çıkış işaretinin kanal sayısı da uygulanan filtre kanalı/sayısı ile eşit olarak hesaplanır. Bu hesaplama işlemi sinir ağındaki bir katman olarak hayal edelim. Giriş görüntüsü ve filtre de aslında sürekli geri yayılımla güncellenen ağırlıklar matrisidir. Aktivasyon fonksiyonunu uygulanan çıkış matrisine en son skaler bir b (bias) değeri eklenir.

## **2.4. SSD (Single Shot MultiBox Detector)**

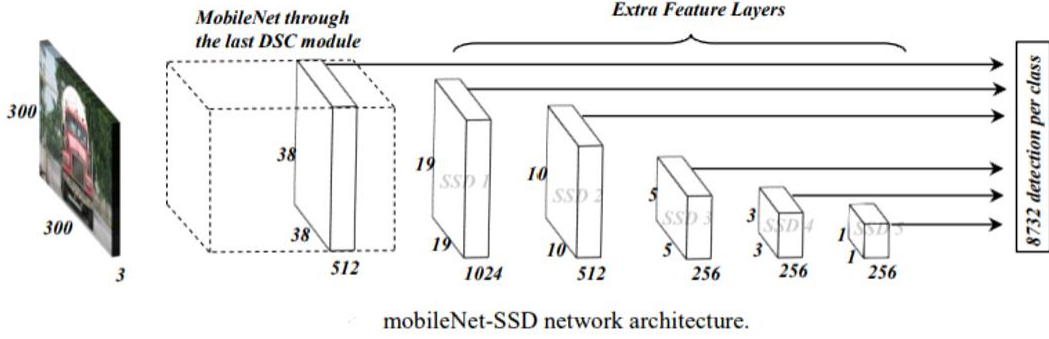
Derin öğrenme tabanlı nesne algılama söz konusu olduğunda, muhtemelen karşılaşacağınız üç temel nesne algılama yöntemi vardır.

- Faster R-CNNs (Girshick et al., 2015)
- You Only Look Once (YOLO) (Redmon and Farhadi, 2015)
- Single Shot Detectors (SSDs) (Liu et al., 2015)

R-CNN'ler derin öğrenme kullanarak nesne tespiti için gelişmekte olan bir yöntem olsa da, “daha hızlı” uygulama ile bile (“R” “Bölge Önerisi” anlamına gelir) R-CNN'lerde algoritma 7 FPS sırasıyla oldukça yavaş olabilir. YOLO ise 40-90 FPS işleyebilecek şekilde çok daha hızlı olmasına rağmen arzulanan doğruluk seviyesini yakalayamamaktadır. Bu noktada projede tercih edilen SSD ise bu ikisi arasında bir dengededir. YOLO'dan daha yüksek doğruluk sağlaması ve gerçek zamanlı uygulamalara uygun olması sebebiyle bu teknik seçilmiştir.

SSD nesne tespiti iki bölümden oluşur:

- iii. Özellik haritalarının ayıklanması,
- iv. Nesneleri tespit etmek için evrişim filtrelerin uygulanması,



Şekil 17 SSD model mimari yapısı [17].

SSD, VGG16'dan sonra 6 yardımcı konvolüsyon katmanı daha ekler. Bunlardan beşi nesne tespiti için eklenmektedir. Bu katmanların üçünde 4 yerine 6 tahmin yapılmaktadır. Toplamda SSD, 6 katman kullanarak 8732 tahmin yapar. Büyük ölçekli özellik haritaları doğruluğu önemli ölçüde artırır. Nesne tespiti için kullanılan çok sayıda özellik eşleme katmanının doğruluğunu arttırmaktadır.

SSD, diğer tespit yöntemlerinden daha fazla yer, ölçek ve en boy oranını daha iyi kapsamak için birçok tahminde bulunur (8732). Bununla birlikte, birçok tahmin nesne içermez. Bu nedenle, sınıf güven puanı 0.1'den düşük olan tüm tahminler elenecektir.

SSD, aynı nesneye işaret eden yinelenen tahminleri kaldırmak için maksimum olmayan baskılama kullanır. SSD tahminleri güven skorlarına göre sıralar. En yüksek güven tahmininden başlayarak SSD, önceden tahmin edilen sınır kutularının aynı sınıf için mevcut tahminle 0.45'ten daha yüksek bir değere sahip olup olmadığını değerlendirir. Bulunursa, mevcut tahmin dikkate alınmaz.

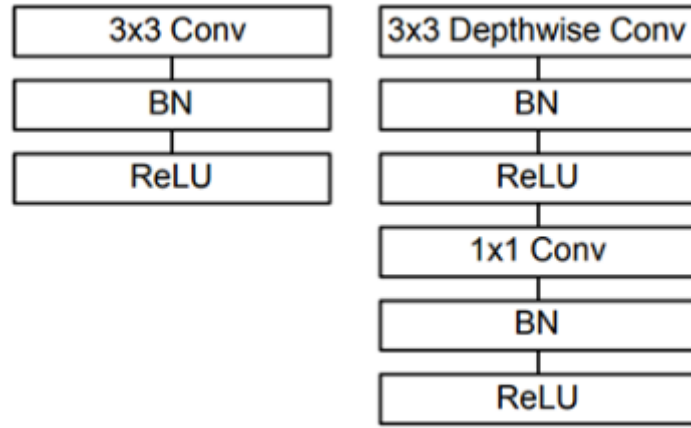
Doğruluğu artırmak için SSD şunları sunar:

- Nesne sınırlarını tahmin etmek için küçük evrişimli filtreler ve varsayılan sınır kutularına ofsetler.
- En boy oranlarındaki farkı işlemek için varsayılan kutular için ayrı filtreler.
- Nesne tespiti için çok ölçekli özellik haritaları.

## 2.5. MobileNet

Nesne algılama ağları oluştururken genellikle VGG veya ResNet gibi mevcut bir ağ mimarisini kullanılmaktadır. Daha sonra ise bu nesne algılama hattı içinde kullanılır. Sorun şu ki, bu ağ mimarileri 200-500 MB boyutları ile çok büyük olabilir. Bu gibi ağ mimarileri, büyüklükleri ve sonuçtaki hesaplama sayıları nedeniyle kaynakları kısıtlı cihazlar için uygun değildir.

Bunun yerine bu projede, Google araştırmacılarının başka bir makalesi olan MobileNets'i (Howard ve diğerleri, 2017) kullanılmaktadır. Bu ağlara “Mobil Ağlar” diyoruz çünkü kaynak kısıtlı cihazlar için tasarlanmışlardır. MobileNets, geleneksel CNN'lerden derinlemesine ayrılabilir özelliği olan kıvrım kullanımı ile farklılık gösterir (Şekil 2).



Şekil 18 Toplu normalizasyon ve ReLU ile standart evrişimsel tabaka [18].

Derinlemesine ayrılabilir evrişimin ardındaki genel fikir evrişimi iki aşamaya ayırmaktır:

- $3 \times 3$  derinlemesine evrişim
- Ardından  $1 \times 1$  noktasal evrişim.

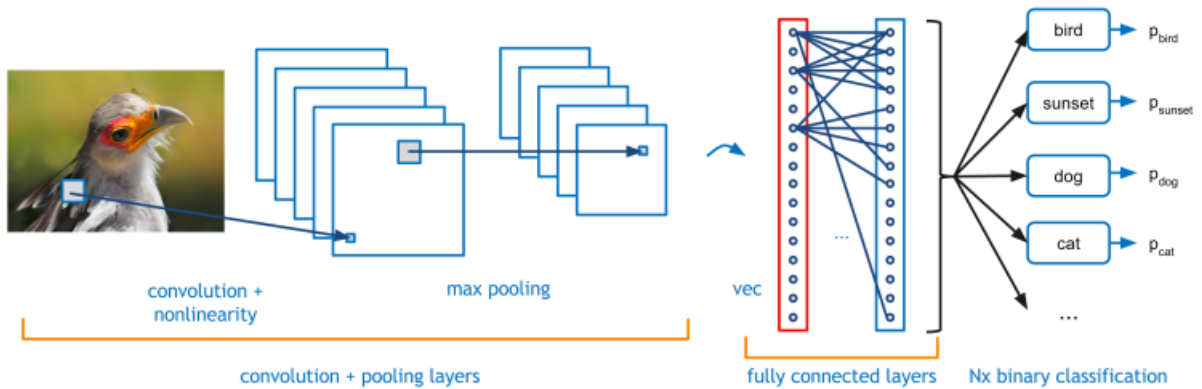
Bu sayede ağdaki parametre sayısı büyük ölçüde azaltılmaktadır. Buradaki sorun ise doğruluk yüzdesi ile kaynak tasarrufu arasında bir takas yapılmasıdır. Bu dengeyi sağlayabilmek amacıyla MobileNet ve SSD tekniklerini bir arada kullanıldı. Hem MobileNet mimarisini hem de Tek Atış Dedektörü (SSD) yapısını birleştirildiğinde, nesne algılamaya yönelik hızlı, verimli, derin öğrenme tabanlı bir yöntem elde edildi.

## 2.6. Caffe

Caffe,, hız ve modülerlik düşünülerek yapılmış bir derin öğrenme mimarisidir. Berkeley AI Research (BAIR) ve topluluk katılımcıları tarafından geliştirilmiştir. Yangqing Jia, projeyi UC Berkeley'deki doktora sırasında yaratmıştır. Caffe, BSD 2-Clause lisansı altında piyasaya sürülmüştür.

Proje kapsamında Caffe seçilmesindeki en önemli faktörler GPU-CPU'da paralel çalışabiliyor olması, ileri beslemeli ağlar ve görüntü işleme konusunda diğer algoritmaların bir adım önünde yer alması ve hassas ayarlamalar için önceden eğitilmiş modelleri barındırmasıdır. Şekil 19'da bu model ile nesne tanımlanması gösterilmektedir.

Ayrıca Caffe'nin sağladığı işleme hızı araştırma deneyleri ve endüstriye dağıtım için onu en uygun model haline getirmektedir. Caffe, tek bir NVIDIA K40 GPU ile günde 60 milyondan fazla görüntüyü işleyebilir. Araştırmalarda Caffe'nin mevcut en hızlı konveksiyon uygulamaları arasında olduğundan bahsedilmektedir [18]. Öte yandan Python için arayüzünün kullanışlı olması da Caffe'yi öne çıkarmaktadır.



Şekil 19 Caffe modeli ile nesne tanımlama [18].

## 2.7. Arka Plan Bakımı

Piksel seviyesi bileşeni, beklenen arka planın olasılıklı tahminlerini yapmak için Wiener filtrelemesi gerçekleştirir; Bölge düzeyinde bileşen ön plandaki nesnelerin homojen bölgelerini doldurur; Çerçeve düzeyindeki bileşen görüntüdeki ani, küresel değişiklikleri algılar ve arka planın daha iyi tahminleriyle takas eder.

Wiener filtresi, yakın geçmişe dayanan doğrusal bir tahmin edicidir. Öngörülen değerinden önemli ölçüde sapan herhangi bir piksel ön plan olarak bildirilir. Belirli bir piksel için doğrusal tahmin şu şekilde verilir:

$$s_t = - \sum_{k=1}^p a_k s_{t-k}$$

$s_t$ ,  $t$  karesindeki öngörülen piksel değeridir,  $s_{t-k}$  pikselin geçmiş bir değeridir,  $a_k$  tahmin katsayılarıdır. Böylece beklenen kare tahmin hatası;

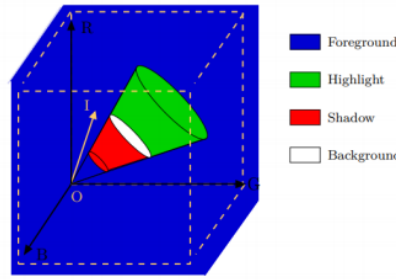
$$E[e_t^2] = E[s_t^2] + \sum_{k=1}^p a_k E[s_t s_{t-k}]$$

Burada tarif edilen çerçeve seviyesi bileşeni, daha önce gözlenen durumlara hızlı adaptasyon için bir yöntem gösterdiği için de dikkate değer bir öneme sahiptir. Sahne arka plan modellerinin temsili setleri tutulur ve daha sonra aralarında otomatik bir mekanizma değişir. Modellerin ne zaman değiştirileceğine karar vermek, görüntüdeki ön plan piksellerinin oranını izleyerek yapılır. Bu kesir 0,7'yi aşarsa, yeni bir modelin değerlendirilmesi tetiklenir. Seçilen model, geçerli görüntüdeki ön plan piksellerinin oranını en aza indiren modeldir. Modül, ışık anahtarı senaryosuna karşı başarıyla test edildi.

## 2.8. Gölge Segmentasyonu

Bu süreç sistemimizde kritik öneme sahiptir, çünkü bu süreç olmadan gölgeleri sahiplerinden ayırt edemeyiz.

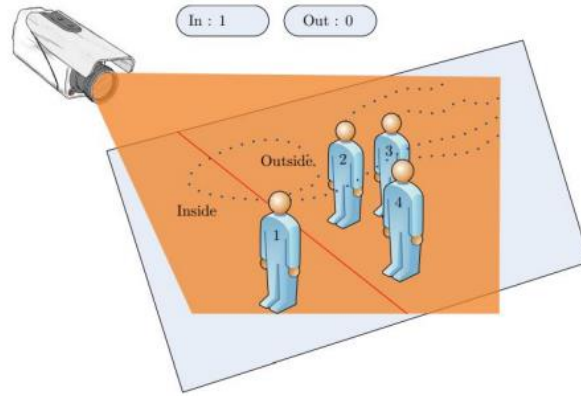
RGB renk uzayında parlaklığı renklilikten ayıran bir renk modeli önerilmektedir [13]. Bu modelde (Şekil 4.2) arka plan görüntüsü önce istatistiksel olarak piksel olarak modellenmiştir. Sonra her bir  $i$  piksel beklenen renk  $E_i$ , standart sapma  $s_i$  ve mevcut renk  $I_i$  ile temsil edilebilir.  $I_i$  ve  $E_i$  arasındaki fark parlaklık ( $\alpha_i$ ) ve kromatiklik  $CD_i$  bileşenlerine ayrılır.



Şekil 20 RGB renk uzayında önerilen 3D koni modeli [7].

## 2.9. Bilgisayarlı Görme Araçları

Kişi sayan bir sistem tasarlayabilmek için görüntü işleme yapabilen bilgisayar araçlarını öğrenerek, bu şekilde farklı görüntüleme algoritmalarının hızlı bir şekilde test edilmesi gerekliydi. Literatür taramasının ardından Açık Kaynak Bilgisayarlı Görme (OpenCV) tercih edildi. OpenCV, Linux, Windows ve Mac OS X'te çalışan ve Python, Ruby, Matlab ve diğer diller için ara yüzlerde aktif gelişim sağlayan açık kaynaklı bir bilgisayarlı görme kütüphanesidir. OpenCV, hesaplama verimliliği ve gerçek zamanlı uygulamalara büyük önem verilerek tasarlanmıştır. Optimize edilmiş C ve C ++ ile yazılmıştır ve çok çekirdekli işlemcilerin avantajlarından yararlanabilir. OpenCV'nin hedeflerinden biri, insanların hızlı bir şekilde görme uygulamaları oluşturmalarına yardımcı olan kullanımı kolay bir bilgisayar görme altyapısı sağlamaktır. OpenCV kütüphanesi ile, fabrikada ürün denetimi, tıbbi görüntüleme, güvenlik, kullanıcı arayüzü, kamera kalibrasyonu, stereo görüş ve robotik gibi birçok işlev başarılabılır. Bilgisayarla görme ve makine öğrenimi genellikle el ele gittiğinden, OpenCV genel amaçlı bir Makine Öğrenim Kütüphanesi de içerir. Şimdilik en büyük eksikliği sadece CPU üzerinden çalıştırılabiliyor olmasıdır.



Şekil 21 Bilgisayarlı görme uygulaması olarak kişi sayan kamera [8].

## 2.10. Kullanılan OpenCV Kütüphaneleri

- **Numpy:** NumPy, Python ile bilimsel hesaplama için temel pakettir. Diğer şeylerin yanı sıra: C / C ++ ve Fortran kodunu kullanışlı lineer cebir, Fourier transformu ve rastgele sayı yeteneklerini entegre etmek için gelişmiş bir N-boyutlu dizi nesnesi sofistike (yayın) fonksiyon araçları. NumPy, açık bilimsel kullanımlarının yanı sıra verimli çok boyutlu jenerik veri konteyneri. Rasgele veri türleri tanımlanabilir. Bu, NumPy'nin çok çeşitli veri tabanlarıyla sorunsuz ve hızlı bir şekilde entegre olmasını sağlar.

- **Dlib:** Dlib, gerçek dünya sorunlarını çözmek için C ++ 'da karmaşık yazılımlar oluşturmak için makine öğrenme algoritmaları ve araçlar içeren modern bir C ++ araç takımıdır. Hem endüstride hem de akademik çevrede robotik, gömülü cihazlar, cep telefonları ve büyük yüksek performanslı bilgi işlem ortamları dahil olmak üzere çok çeşitli alanlarda kullanılır. Dlib'in açık kaynak lisansı, herhangi bir uygulamada ücretsiz olarak kullanmanızı sağlar.

- **Imutils:** Çeviri, döndürme, yeniden boyutlandırma, iskeletleştirme ve Matplotlib görüntülerini OpenCV ile daha kolay görüntüleme gibi temel görüntü işleme işlevlerini kolaylaştıran bir dizi kullanışlılık kütüphanesi

- **Datetime:** Datetime modülü tarih ve saatleri işlemek için sınıflarsağlar. Tarih ve saat aritmetiği desteklenirken, uygulamanın odağı çıktı biçimlendirme ve manipülasyon için etkili öznitelik çıkarma üzerinedir.

- **Collections:** Bu modül, Python genel amaçlı yerleşik kaplara, dikte, liste, küme ve demetlere alternatifler sunan özel konteyner veri türlerini uygular.



- **Math:** Bu modül C standardı tarafından tanımlanan matematiksel fonksiyonlara erişim sağlar. Bu fonksiyonlar karmaşık sayılarla kullanılamaz.
- **Matplotlib:** Veri görselleştirmesinde kullandığımız temel python kütüphanesidir. 2 ve 3 boyutlu çizimler yapmamızı sağlar. Matplotlib genelde 2 boyutlu çizimlerde kullanılırken, 3 boyutlu çizimlerde başka kütüphanelerden yararlanılır.

### 2.11. Algoritma Mimarisi

Tasarlanan sistemde önemli olan kısımlar ortamdaki kişilerin tespiti, hareket yönlerinin takibi ve sayımın yüksek başarıyla gerçekleştirilmesidir. Programlama dili olarak veri analizi ve yapay zekâ çalışmalarının gelişimine öncelik etmesi, kütüphane bakımından zengin ve açık kaynaklı olması nedeniyle Python tercih edilmiştir.

Sayma işlemi yapmadan önce elbette nesnelerin(kişilerin) tespit edilmesi gerekmektedir. Nesnelerin bulunması için ise bir girdi görüntüsünü alıp, görüntünün içindeki nesneleri tespit eden, tespit ettikten sonra da nesnenin bir insan olup olmadığının tanımlanması gerekmektedir. Bu amaçla ise doğruluk ve hız kriterlerini karşılayan derin öğrenme algoritması, evrişimli sinir ağı CNN (ConvNet) tercih edilmiştir. CNN verilen görüntüleri ayırt etmek için bir insanı insan haline getiren benzersiz özellikleri kullanır. Aslında bu süreç beynimizde de bilinçsizce yapılmaktadır. Bölüm 2.3.'de daha detaylı bahsedildiği şekilde CNN'de aynı şeyi yapar. Ancak daha önce eğriler ve kenarlar gibi alt düzey özellikleri tespit ederler ve daha soyut kavramlara kadar bunları oluştururlar. Bu yaklaşımla ilgili sorun, ilgilenilen nesnelerin görüntü içinde farklı uzamsal konumlara ve farklı en boy oranlarına sahip olmasıdır. Bu nedenle, çok sayıda bölge seçmek zorunda kalınmaktadır ve bu hesaplamada sorun yaratabilir. Bu nedenle, R-CNN, YOLO, SSD gibi algoritmalar bu oluşumları bulmak ve hızlı bir şekilde kullanmak için geliştirilmiştir. Proje kapsamında bu teknikler incelenerek gerçek zamanlı performans ve doğruluk kriterleri açısından SSD tekniği ile MobileNet yapısının bir arada kullanılmasına karar verildi. Bölüm 2.4 ve bölüm 2.5'te bu teknikler ile ilgili detaylı bilgilere yer verilmiştir.

Gerçek zamanlı olarak nesne algılaması için tasarlanmış olan SSD ile sınır çerçeveleri çizilerek çözünürlük düşürülmektedir. Ayrıca derinlemesine ayrılabilir evrişimden yararlanan MobileNet yapısıyla da ağıdaki parametre sayısı büyük ölçüde azaltılmaktadır. Böylece, CNN'e gönderilen ana resmin boyutunu azaltarak, sistem gereksiz işlem yükünden kurtarıldı. SSD, özellik haritalarını çıkarmak için VGG16 ağını kullanmaktadır. Daha sonra Conv4\_3 katmanı ile nesneleri algılar. SSD, VGG16'dan sonra 6 yardımcı konvolüsyon katmanı daha ekler.

Bunlardan beşi nesne tespiti için eklenmektedir. Bu katmanların üçünde 4 yerine 6 tahmin yapılır. Toplamda SSD, 6 katman kullanarak 8732 tahmin yapar. Sonuç olarak tespit edilen nesnelerin model ile eşleşmesine dayanarak tahmin değerleri sağlar. SSD bir nesne olup olmadığını belirtmek için 0 ile 1 arasında bir değer kullanır. SSD bu projedeki renkli video görüntüleri gibi çok ölçekli özellik haritalarının doğruluğunu önemli ölçüde artırır. Nesne tespiti için kullanılan çok sayıda özellik eşleşme katmanının doğruluğunu arttırmaktadır. Bu işlemlerin sonucunda çıkan nesne matrisi Caffé algoritması ile CNN içerisinde konvülyasyon ve Relu işlemine uğrar, bunun sebebi katmanlardaki özneliteliklere sahip nöronlardan geçirilirken aralarındaki benzerlikleri bulmaktır. Bulunan benzerlikler sonucunda her resmin içerisindeki nesnenin tanınması beklenmektedir. Kullanılan kütüphaneler bu filtreleri optimum doğruluk düzeyinde seçebilmektedir.

Bir sonraki aşamada tanınan nesneler belli bir eşik değeri ile karşılaştırılır. Eğer insan olma ihtimali 0.65 (Bu değer kesin değildir) üzerindeyse insan olarak etiketlenir. Etiketlenen bu nesnelerin merkez noktaları CentroidTracker sınıfının içine girerek atanır. Bu nesneleri çerçeve içerisine almak yerine sadece nokta ile gösterilmesinin nedeni işlemciye daha az işlem yükü yüklemektir. Etiketlenen nesnelerin merkez noktaları atandıktan sonra TrackableObject sınıfıyla merkezi noktaları her etiketlenmiş nesne için özelleştirilir. Bu sayede nesneleri takip etme işlemi yapılır. Nesnelerin her çerçeve arasındaki hareket yönüne bakılarak regresyon yöntemi uygulanır. Regresyon yöntemi etiketlenmiş nesnenin belirlenen referans çizgisinin hangi yönüne hareket ettiğinin belirlenmesinde yardımcı olur. Tespit edilen kişinin ağırlık merkezinin hareket yönüne bakılarak giriş mi çıkış mı yapıldığı tespit edilir. Hareket yönüne göre giriş ve çıkış değerleri artırılır. Program sonlandırıldığında ise kullanıcıya kişi/zaman grafiği tablo olarak paylaşılır.

Projenin çalışma adımların en basit şekliyle sıralanacak olursa,

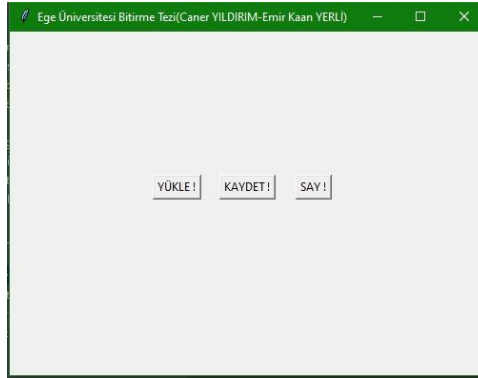
- i. Gerçek zamanla bir video akışı veya kayıtlı herhangi bir video sisteme yüklenir.
- ii. Nesne Tespit Modülü: SSD nesne dedektörü kullanılarak sınırlayıcı kutular belirlenir. Sonrasında ise MobilNet Caffé evrişimli sinir ağından geçirilir ve eğitilmiş model ile girdileri karşılaştırarak nesnenin insan olup olmadığına dair bir tahmin elde edilir. Bu noktada sonuçları keskinleştirmek amacıyla düşük ağırlık yüzdesine sahip tahminler filtrelenir. Kişi olduklarına karar verilen nesnelere kimlikleri atanır.
- iii. Nesne Takip Modülü: Nesne takibi için dlib kütüphanesi ile sentroid izleme ve korelasyon tekniklerinden yararlanır. Ağırlık merkezleri belirlenen ve kimlikleri

atanan nesnelerin her karedeki bir önceki karedeki konumlarının korelasyon durumları incelenerek kişinin en olası mesafedeki konuma yöneldiği varsayılır. Bu şekilde kişilerin hareket yönleri izlenebilir.

- iv. Arayüz üzerinde, tanımlanan kişi nesnelerinin sentroidlerine kimlik bilgisi atanarak arka planda her kişi veri tabanına kaydedilir ve ardışık 40 karede tespit edilemeyen kişiler sistemden çıkarılır.
- v. Sistem çalışırken anlık olarak toplam giriş ve çıkış verileri ekrana yazdırılır. Sistemin sonlandırılmasıyla kişi/zaman raporu çıkartılır.

## 2.12. Arayüz Tasarımı

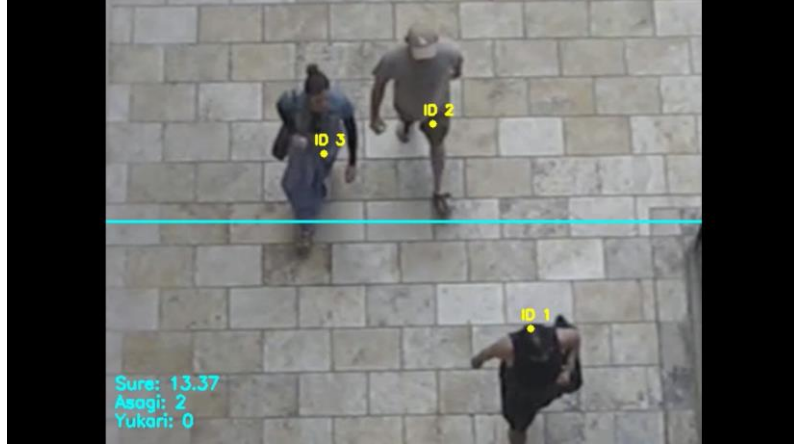
Tasarlanan arayüz ile programın kolayca başlatılabilmesi, giriş olarak verilecek video akışının kolayca seçilebilmesi, elde edilen verilerin anlık olarak takip edilebilmesi ve bir rapor olarak sunabilmesi sağlanmaktadır. Sistemin başlatılması için tasarlanan arayüz Şekil 22’de görülmektedir.



Şekil 22 Kişi sayma sisteminin başlatılması için tasarlanan arayüz.

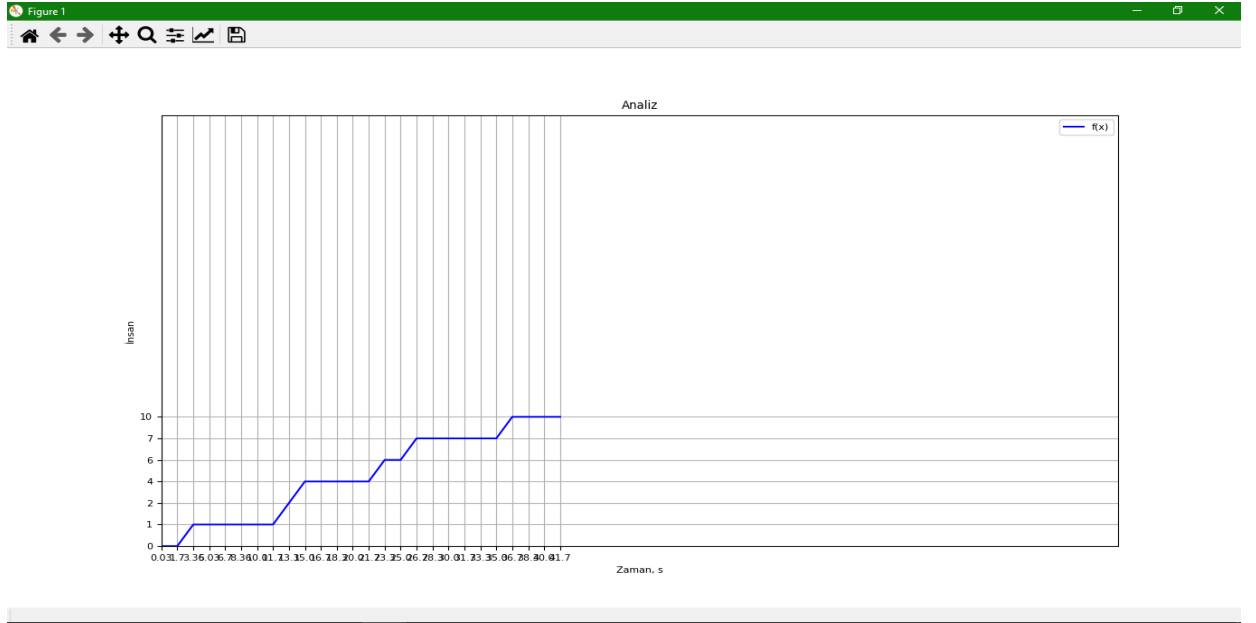
Elde edilen sonuçları anlık olarak takip edebilmek için görüntü üzerine çalışan bir arayüz tasarlandı. Burada aşağıdaki veriler çerçeveye yerleştirildi. Şekil 23’te bunun bir örneği görülebilmektedir.

- Nesne kimliği: Her nesnenin sayısal tanımlayıcısı.
- Sentroid: Nesnenin merkezi, bir “nokta” ile temsil edilir.
- Sınır çizgisi: Giriş ve çıkışın temel alınacağı sınır.
- Giriş: Toplam giriş sayacı.
- Çıkış: Toplam çıkış sayacı.



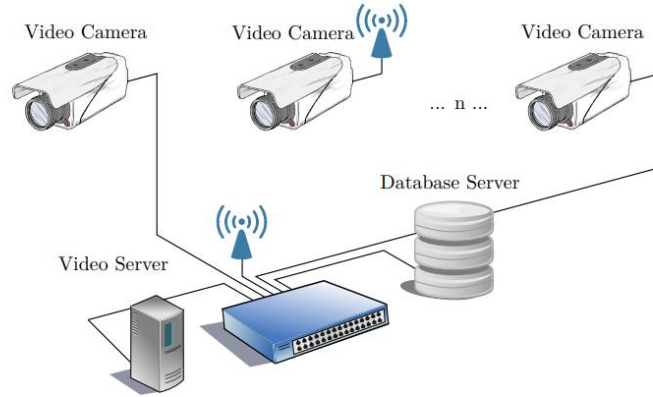
Şekil 23 Gerçek zamanlı olarak kişi sayan sisteme ait arayüz görüntüsü.

Ayrıca bu verileri daha anlamlı hale getirebilmek amacıyla kişi/zaman grafikleri ile kullanıcıya bir rapor sağlandı. Şekil 24’te bunun bir örneği görülebilmektedir.



Şekil 24 Kişi sayma sisteminden elde edilen örnek bir rapor.

### 2.13. Sistem Mimarisi



Şekil 25 Kişi sayma sistemi için sistem yapısı [15].

Algoritmanın gerçek hayatta gerçekleştirilmesi için gerekli sistem mimarisi Şekil 25'te görülmektedir. Kablolu veya kablosuz ağ kameralarından elde edilen video akışları standart bir CPU üzerinde proje kapsamında tasarlanan kişi sayma algoritmasından geçirilerek elde edilen verilen anlık olarak takip edilebilir ya da ileride raporlamak için bir veri tabanında saklanabilir. Algoritmanın tek bir kamera ve işlemci içeren herhangi bir cihazla çalışabilir durumda olması uygulanabilirlik noktasında hedeflerin gerçekleştirildiğini göstermektedir.

### 2.14. Kamera Yerleşimi

İnsan grupları kameranın görüş alanından geçtiğinde tıkanıklıklardan kaçınmak için tepeden görüş açısına sahip kamera kullanımı genellikle en iyi seçenektir. Bu yerleşim mümkün olduğunca tercih edilmelidir. Şekil 26'da görüldüğü gibi tepeden görüş açısı ile birbirlerine yakın hareket eden insan toplulukları çok daha kolay şekilde ayırt edilebilir. Bu da sistemin performansını büyük ölçüde etkilemektedir.



*Önden Görüş*



*Az Eğim ile Önden Görüş*



*Çok Eğim ile Önden Görüş*



*Tepeden Görüş*

*Şekil 26 Kişi takibinde farklı kamera açıları.*

## 2.15. Ölçümler

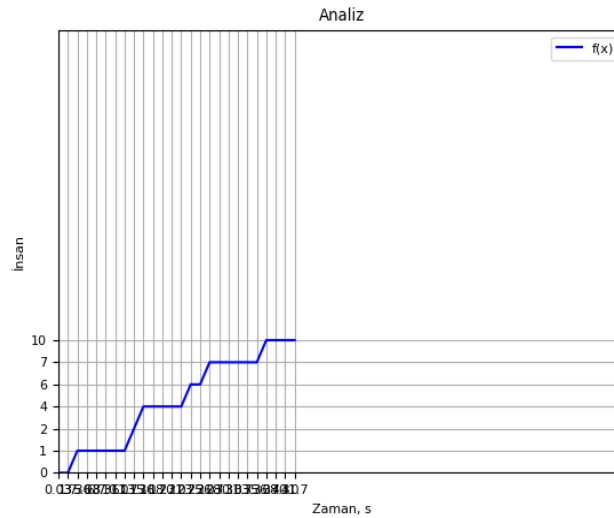
### ▪ Test Videosu 1

Ölçümler internet üzerinden erişilebilen çeşitli videolar ile yapılmıştır. Bunun sebebi Covid-19 salgını nedeniyle istenen test koşullarında video çekmenin imkânsız duruma gelmesidir. Bu nedenle bu kapsamlı bir arama yapılarak çeşitli test kriterlerini karşılayan video akışları bulunup bunlar üzerinde çalışılmıştır. Bu kısımda sırasıyla bu testlere yer verilmektedir.



*Şekil 27 Birinci test videosuna ait görsel.*

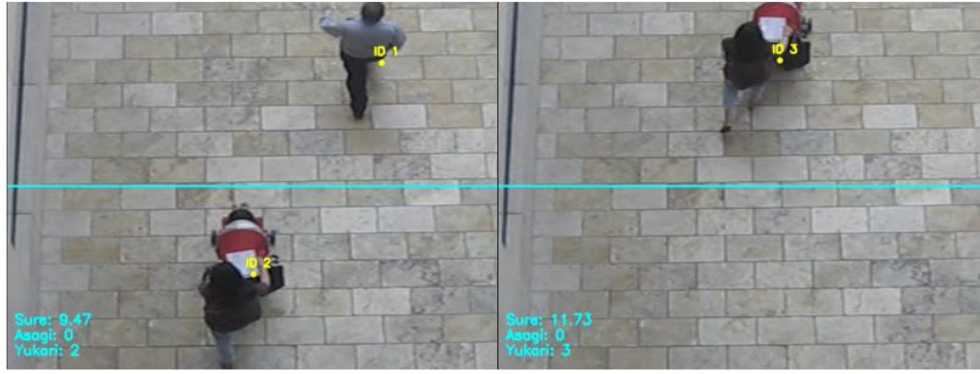
Şekil 27’de görülen birinci test videosu, arka planın yakın renk tonlarından oluşması, kameranın kayıt işlemi sırasında çok fazla titrememesi, içerisinde nesne ve gölge miktarının minimum düzeyde olması, son olarak da kamera açısının yer ile 90 dereceye yakın olması, geliştirilen model ağı için istenilen parametrelere en yakın örnektir. Bu nedenle verimliliğinin yüksek olması beklenilmektedir. Bu video 42 saniyeden oluşmaktadır, içerisinde toplam on nesne geçmektedir. Bunların yedisi aşağı yönle, üçü yukarı yönlü hareket etmektedir. Bu örnekte sayma işlemi yapıldığında %100 doğruluk değerine ulaşılmıştır. Bu doğruluk oranının bu şekilde çıkmasının nedenleri arasında kamera açısının tepeden konumlanması ve keşfedilmesi gereken insan harici nesne sayısının minimum derecede olması nedeniyle işlemciyi yormaması olarak değerlendirilebilir. Şekil 28’de hangi saniyelerde kaç kişinin geçtiğine dair rapor görülmektedir.



Şekil 28 Birinci test video akışına ait çıktı.

#### ▪ Test Videosu 2

İkinci test video akışı birinci video akışı ile aynı ortamda gerçekleştirilmektedir. Ancak bu videoda bebek arabasının bulunması nedeniyle, kişi sayma algoritmamızı testi için önemli bir başka kriteri sağlamaktadır. Şekil 29’da görüldüğü üzere bebek arabasıyla geçilmesinden sonra sadece ebeveyn sayma işlemi gerçekleştirilmiştir. Bunun sebebi bebek arabasının içerisinde bir insan görülmemesinden kaynaklanmaktadır. Bu bir hata değildir.



*Şekil 29 İkinci test videosuna ait görsel.*

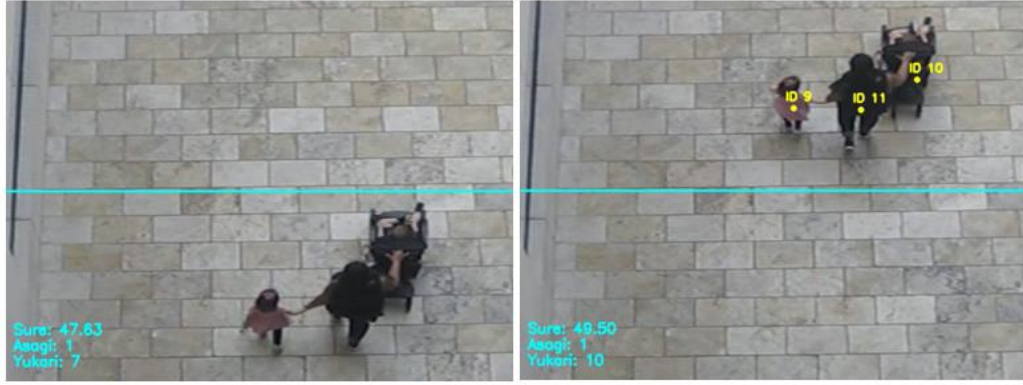
Şekil 30’da görüldüğü üzere bebek arabasıyla geçilmesinden sonra sadece ebeveynler sayılmıştır. Buradaki hata, tasarlanan modele göre bebek arabasındaki kişiyi sayabilmek için gerekli eşik değerinin aşılamıyor olmasından kaynaklanmaktadır. Bu eşik değerinin düşürülmesi her ne kadar bu duruma çözüm sağlayabilir gibi görünmesine rağmen farklı nesneleri de kişi olarak sayabilecek duruma gelmesi nedeniyle hata oranının arttıracaktır. Bu nedenle bu değerin test koşulları ışın (ışık, kamera açısı, kare sayısı) optimize edilmesi gerektiği aşıkardır.



*Şekil 30 İkinci test videosu için bebek arabasındaki kişinin sayılamadığı durum.*

İkinci test video akışına ait son örnekte ise, model ağı bebek arabasındaki çocuğu tespit edebilecek eşik değerini elde edebildiğinden sayma işlemine dahil etmiştir (Şekil 31). Bu görseller kıyaslandığında Şekil 30’daki hatanın gölge segmentasyonundaki zorlayıcı koşullardan kaynaklandığı söylenebilir.



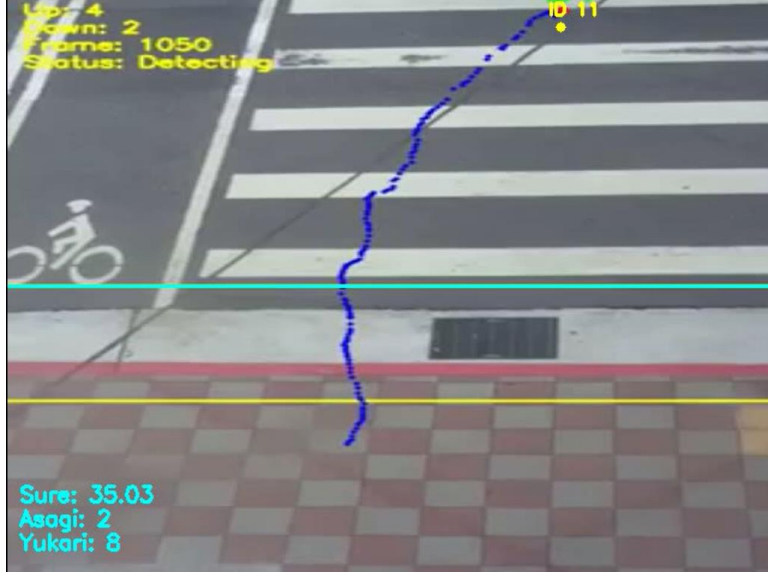


*Şekil 31 İkinci test videosu için bebek arabasındaki kişinin doğru tespit edildiği durum.*

Sonuç olarak ikinci örnek video da 69 saniye içerisinde toplam 16 kişi sayılmaktadır, ikisi aşağı yönlü, 14 tanesi yukarı yönlüdür. Başarı oranı 17 kişinin 16 kişinin sayılması sebebiyle %94.11 şeklinde hesaplanmıştır.

#### ▪ Test Videosu 3

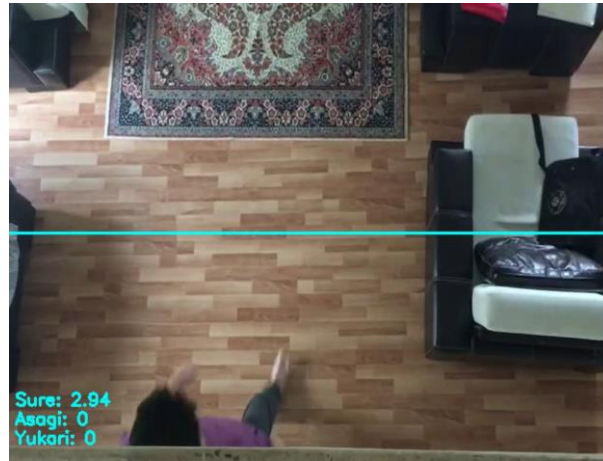
Seçilen üçüncü örnek video akışında ise Kalman filtresi ve geleneksel yöntemler kullanan ticari bir ürüne ait video tercih edilmiştir. Bu teste ait görsel Şekil 32’de görülebilmektedir. Burada sarı renkli yazılar bu yöntemle ait sonuçları gösterirken turkuaz renkli yazılar projemize ait sonuçları göstermektedir. Bu test videosu ayrıca modelimizi geleneksel bir yöntem ile karşılaştırma imkânı sağlamaktadır. Kameranın açık bir alanda yerleştirilmesi ve açısı düşünüldüğünde sistemimiz için ideal koşulları karşılamayan bir konumdur. Ayrıca arka planın farklı desenlerden oluşması nedeniyle görüntü üzerinde fazladan bir iş yükü meydana gelmektedir. Yine de görüntü iki modelden de geçirildiği zaman soldaki modelde toplam altı kişi sayılmıştır, bunların 4’ü yukarı, 2’si aşağı yönlüdür. Aynı görüntüyü tasarladığımız model ağından geçirildiğinde 10 kişi tespit edilmiştir. 35 saniyeden oluşan video akışında, sistemimiz tarafından yapılan tek hata tekerlekli sandalyedeki bir hastanın sayılamamasıdır. Bu hata ile birlikte modelimizin doğruluğunun %90.9 olmaktadır. Videonun alındığı model ise %54.5 doğruluk yüzdesine sahiptir.



Şekil 32 Üçüncü test video akışına ait görsel.

#### ■ Test Videosu 4

Seçilen dördüncü test video akışı ise, kişilerinin hızlarına, modelin vereceği tepkiyi ölçmek için proje çalışanları tarafından oluşturulmuştur. İlk olarak kameranın görüş açısı içerisinde yavaş bir şekilde geçilmiş daha sonraki aşamalarda kişinin görüş alanında daha hızla geçtiği durumlar ile model test edilmiştir. Şekil 33’de görülen örnekte kamera görüş açısından 2,16 saniyede girilip çıkmıştır. Kameranın görüş açısından bu hızla geçildiği durum için model başarılı bir şekilde kişiyi tespit edip sayma işlemini yapmıştır.



Şekil 33 Kişilerin görüş açısından geçiş hızlarının sistem üzerindeki testi. (Yavaş durum)

Aynı koşullarda gerçekleştirilen diğer bir örnekte ise kameranın görüş açısından 0,87 saniye de girilip çıkmıştır. Bu hızla geçildiği zaman model yine başarılı bir şekilde kişiyi tespit edip

sayma işlemini yapmıştır. Sonraki aşamada ise kameranın görüş açısından 0,66 saniye de girilip çıkılarak test gerçekleştirilmiştir. Kişinin bu hızla hareket ettiği durum için model başarısız olmuştur. Nesneyi tespit edememiştir. Kameranın yere oldukça yakın olduğu ve nesneyi tanıyacak süresinin kısıtlı olduğu düşünüldüğünde bu sonuç anlaşılabilir olarak değerlendirilmektedir.

- Test Videosu 5

Son test videosu ile tasarlanan sistem bir mağazanın girişinden alınan görüntüler ile test edilmektedir. Bu video akışı görüntü içerisinde nesne miktarının fazlalığı, kamera açısının istenenden bir miktar farklı olması, güneş nedeniyle yüzeyden yansıyan ışık miktarı ve aynı anda birçok kişinin görüş içerisinde hızla geçmesi gibi durumlar değerlendirildiğinde sistemin zorlu koşullardaki performansını anlayabilmek için önemli bir örnektir.

Bu video akışında 51 kişinin aşağı yönlü hareket ettiği, 25 kişinin ise yukarı yönlü hareket ettiği görülmüştür. Örnek videonun model ağıdan geçirilmesi sonucu Şekil 34'te görüldüğü üzere 48 aşağı yönlü sayma işlemi, 24 yukarı yönlü sayma işlemi yapılmıştır. Doğruluk yüzleri hesaplandığında, yukarı yönlü sayım için doğruluk yüzdesi %96 iken, aşağı yönlü sayım için doğruluk yüzdesi %94,11 olmaktadır. Görüntü içerisinde alınan video akışına ait gereksiz şekiller yer alması başarı yüzdesini negatif yönde bir miktar etkilemektedir.



Şekil 34 Sistemin bir mağaza girişinden alınan video akışı ile testine ait görüntü.

## 2.16. Performans

Bütün testler ve sonuçları değerlendirildiğinde kamera açısının modelde tasarlandığı şekliyle (Bölüm 2.14.) yere dik bir komunda olması performansı büyük ölçüde etkilediği görülmüştür.

Elde edilen sonuçlara göre kameranın görüş açısında çok fazla nesne olması, işlemcinin iş yükünü arttırmakta ve kare hızı başına performansı düşürmektedir. Ayrıca beklendiği gibi, ışıklandırmanın ve görüş açısındaki gölge miktarının da doğruluk yüzdesini düşüren etkenlerden olduğu görülmüştür.

Tüm test videoları incelendiğinde, tasarlanan model ağının uygun parametreler verildiği takdirde %94'ün üzerinde doğruluk sağladığı sonucu ortaya çıkmaktadır. Tasarlanan sistem içerisinde kullanılan MobilNet-SSD teknolojisi gerçek zamanlı video akışları üzerinde nesne tespitinde 25-35 kare hızı ile beklenen performansı sağlamaktadır. Nesnenin sınıflandırılması aşamasında ise video işleme başarısı en yüksek olan derin öğrenme algoritması olan Caffe modeli kullanılarak kişi tespitindeki başarı en üst düzeye düzeye taşınmaktadır. Son olarak sentroid izleme ve korelasyon tekniklerinin bir arada kullanılmasıyla oluşturulan hareket takibi ile kişilerin yön tayinleri gerçekleştirilmektedir. Elde edilen sonuçlar bir arayüz ve tablo aracılığıyla kullanıcıya rapor olarak verilmektedir.

### 3. SONUÇ

Kişi sayma teknolojilerinin derinlemesine araştırılmasıyla başlayan proje sürecinde sonuç olarak bir binada bulunan kişi sayısını takip etme probleminin giriş ve çıkışlara yerleştirilen kameralar yardımıyla veya güvenlik kameralarıyla çözüme ulaştırabilir olduğu anlaşılmıştır. Proje kapsamındaki çalışmalar sonucunda tasarlanan algoritma, gelişmiş kişi algılama doğruluğu için derin öğrenme nesne dedektörleri (MobileNet ve SSD) ve gelişmiş izleme doğruluğu için hem sentroid izleme hem de korelasyon filtreleri dahil olmak üzere iki ayrı nesne izleme algoritması kullanarak, standart bir CPU üzerinde gerçek zamanlı olarak çalışabilen, hem “algılama” hem de “izleme” aşamalarını uygulayarak görüntüye giren veya görüntüden kaybolan kişileri izleyebilen ve hareket yönleri doğrultusunda bir alana giriş ve çıkış verilerini elde edebilen bir sistem tasarlanmıştır. Maliyeti ile uygulanabilirliği ve doğruluk yüzdesi yüksek bu sistem ile kişi sayma teknolojilerin yaygınlaşacağı, bu teknolojilere olan farkındalığın artacağı ve ileride akıllı şehirler için önemli bir veri sağlayıcı olacağı düşünülmektedir.

Geliştirilen sistemi değerlendirmek için, internet ortamından elde edilen bir dizi görüntü üzerinde bazı analizler yapılmıştır. Elde edilen doğruluk sonuçları, istenen son hassasiyete oldukça yakın oldukları için sonuç tatmin edici olarak değerlendirilmiştir. Ancak sistemin daha önceden tanımlanmış tüm durumlarda doğru bir şekilde değerlendirilmesi için ele alınması gereken çeşitli senaryolar olduğundan test süreci ileriye yönelik çalışmalar ile devam ettirilmelidir. Çözülmesi gereken ana sorun çeşitli ışık kaynakları ve gölgeler nedeniyle arka plan piksellerinin yanlış algılanabilmesidir. Bir dakikalık test videosunda sistem %5-8 hata oranları ile çalışmaktadır. Sonuçlar doğrultusunda sistem mağaza, lokanta, alışveriş merkezi, sınıflar vb. kapalı alanlarda kişi yoğunluğunu takip edebilir seviyededir. Ayrıca tüm vücudu tespit edebildiğinde ve çok fazla tıkanıklık olmaması durumunda, havalimanı, tren istasyonları gibi kalabalık mekanlarda da çalışabilir. Eğitilmiş derin öğrenme algoritması sayesinde hayvanları, bisikletleri ve benzer araçları saymamaktadır.

Bunların yanı sıra, elde edilen bilgiyi anlamlı hale getirme problemi ve bütün bir kampüsün ya da bütün bir şehrin yoğunluk haritasını oluşturabilme problemi ise projenin geliştirme sürecinde devam etmektedir. Örneğin sistem üzerinde geliştirmeler yapılarak Ege Üniversitesi Merkez kampüsündeki belirli yapıların ana giriş ve çıkışlarına bu sistem kurulup büyük ölçekli durumlar için faydaları araştırılabilir. Sonrasında ise Covid-19 salgını nedeniyle kapalı alanlardaki kişi yoğunluğunu kontrol edebilmek amacıyla çalışmalar devam ettirilebilir.

## REFERANSLAR

- [1] K. Terada, D. Yoshida, S. Oe, and J. Yamaguchi, "A method of counting the passing people by using the stereo images," in Proc. Int. Conf. Image Processing ICIP 99, vol. 2, 1999, pp. 338–342.
- [2] K. Hashimoto, K. Morinaka, N. Yoshiike, and S. Kawaguchi, C.and Matsueda, "People count system using multi-sensing application," in Proc. Int Solid State Sensors and Actuators TRANSDUCERS '97 Chicago.Conf, vol. 2, 1997, pp. 1291– 1294.
- [3] A. G. Vernazza, "Long-memory matching of interacting complex objects from real image sequences," in no book, 1996.
- [4] A. Shio and J. Sklansky, "Segmentation of people in motion," in Proc. IEEE Workshop Visual Motion, 1991, pp. 325–332.
- [5] G. Sexton, X. Zhang, G. Redpath, and D. Greaves, "Advances in automated pedestrian counting," in Proc. European Convention Security and Detection, 1995, pp. 106–110.
- [6] J. Segen, "A camera-based system for tracking people in real time," in Proc. 13th Int Pattern Recognition Conf, vol. 3, 1996, pp. 63–67.
- [7] G. Conrad and R. Johnsonbaugh, "A real-time people counter," in Proceedings of the 1994 ACM symposium on Applied computing, ser. SAC '94. New York, NY, USA: ACM, 1994, pp. 20–24. [Online]. Available: <http://doi.acm.org/10.1145/326619.326649>
- [8] A statistical approach for real-time robust background subtraction and shadow detection, 1999.
- [9] A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in Proc. IEEE Conf. Computer Vision and Pattern Recognition
- [10] G. Monteiro, J. Marcos, M. Ribeiro, and J. Batista, "Robust segmentation for outdoor traffic surveillance," in Proc. 15th IEEE Int. Conf. Image Processing ICIP 2008, 2008, pp. 2652–2655.
- [11] Satish D. Pore, B. F. Momin,"Bidirectional People Counting System in Video Surveillance", 2010.
- [12] Chen T., Chen C., Wang D.,Kuo Y.," A People Counting System Based on Face-Detection", 2013.
- [13] Terada K., Yoshida D., Oe S.,Yamaguchi J.," A Counting Method of the Number of Passing People Using a Stereo Camera", 2014.
- [14] Chang Q., Song Z., Shi R.,Xu J.," A People counting method based on multiple cameras information fusion", 2015.

- [15] C. Zhang, H. Li, X. Wang, X. Yang, “Cross-Scene Crowd Counting via Deep Convolutional Neural Networks”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’15), 833-841, 2015.
- [16] J. Shao, K. Kang, C. C. Loy, X. Wang, “Deeply Learned Attributes for Crowded Scene Understanding”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’15), 4657-4666, 2015.
- [17] Y. Zhang, D. Zhou, S. Chen, S. Gao, Y. Ma, “Single Image Crowd Counting via Multi-Column Convolutional Neural Network”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’16), 589-597, 2016
- [18] D. Oñoro, R. R. Lopez-Sastre, “Towards Perspective-Free Object Counting with Deep Learning”, Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2017.

## LİSANS BİTİRME PROJESİ ÖZDEĞERLENDİRME FORMU

1. Projeniz tasarım boyutu nedir (prototip gerçekleştirme, benzetme veya analiz)?

Kişi sayma sistemleri üzerine kapsamlı literatür taraması yapılarak hedef ve kapsam doğrultusunda en uygun yöntemler kullanıldı. Projede hedeflenen kontrol mekanizmaları sağlanmış, nesne tespiti ve takibi için gerekli algoritmalar tasarlanmış ve toplam giriş ve çıkış bilgilerini kullanıcıya raporlayan bir arayüz tasarlanmıştır. Ancak proje kapsamında kullanılan teknikler ve oluşturulan algoritmalar Covid-19 virüs salgını nedeniyle internet ortamında erişilebilen videolar ile test edilmiştir. Proje kapsamında temel alınan kamera açısıyla çekilmiş videolarda ise istenen başarı yüzdesine ulaşılmıştır.

2. Kullandığınız tasarım yöntem (yöntemleri) açıklayınız:

Proje kapsamında en uygun araçlar ve kütüphaneler araştırılarak, tasarım aşamasın Python ve OpenCV tercih edildi. Proje ile ilgili hedefler ve zorluklar belirlendikten sonra ise kişi sayma sistemlerinde kullanılan teknolojiler literatür incelenerek gerçek zamanlı performanslarına, uygulanabilirliklerine, ölçeklenebilirliklerine ve doğruluk yüzdelere göre incelenerek en uygun yöntemde karar kılındı. Hedeflenen başarı yüzdesine erişmek amacıyla testler yapılarak tasarım aşamasında geliştirmeler yapıldı. Başarı oranının istenen seviyeye ulaşmasının ardından farklı ışık ve fps koşullarında testlere ve geliştirmelere devam edildi. Son olarak sistemin kullanılabilirliğini ve okunabilirliğini arttırmak amacıyla bir arayüz tasarlandı.

3. Kullandığınız veya dikkate aldığınız mühendislik standartları nelerdir?

Güvenli yazılım geliştirme standardı olan Common Criteria/ISO 15408 dikkate alınmıştır Ayrıca çevik yazılım metodolojine ait prensipler uygulanmıştır.

4. Kullandığınız veya dikkate aldığınız gerçekçi kısıtlar nelerdir? (Ekonomi, Çevre sorunları, Sürdürülebilirlik, Üretilebilirlik, Etik, Sağlık, Güvenlik, Sosyal ve politik sorunlar)

Projedeki asıl amacımız kişi sayma sistemlerini düşük maliyet ve yüksek doğruluk ile uygulayıp bu haliyle birçok yararlı uygulama alanına değinmekti. Öyle ki belirli bir alanın yoğunluk durumu akıllı şehirler için önemli ve anlamlı bir veri olmaktadır., Covid-19'un yayılma hızının çok daha yüksek olduğu lokanta, kafe, bar, alışveriş merkezi, mağaza, sınıflar gibi birçok kapalı alanda kitlelerin kontrollü giriş ve çıkışı sağlanabilir. Günlük, aylık, yıllık ortalama giriş çıkış verileri temel alınarak binaların enerji yönetiminde, perakendecilik sektöründe, pazarlama alanında, hastanelerde vb birçok sektör ve uygulamada anlamlı tedbir ve tasarruflar yapılabilir. Bu bağlamlarda projemiz kişilerin kimliklerini incelemekten sadece sayımlarını yaparak etik açıdan bir sorun teşkil etmezken, sağlık, ekonomi, güvenlik ve sosyal sorunlarda faydalı bir çalışma olma hedefini yerine getirmektedir.



5. Çalışmanızın daha önce yapılmış olan çalışmalardan ne farkları var (standart bakımından, kullanan malzeme bakımından v.s)

Günümüzde kişi sayma teknolojilerinde birçok alternatif bulunmaktadır ancak birçoğu uygulanabilirlik açısından eksik kalmaktadır. Derinlik algılayan kameralar yardımıyla ya da birden fazla kameradan elde edilen 3D görüntülerle kişi sayan sistemler bu alandaki en güncel örneklerdir. Bu sistemler daha çok perakende sektöründe ticari amaçlar ile kullanılmaktadır. Ancak teknolojinin ilerlemesi ile kameralar ucuzlamaya devam ederken ve görüntü işleme, nesne tespiti ile ilgili gelişmeler sürerken bu sistemler maliyetleri nedeniyle yaygınlaşmaktan uzaktadır. Derin öğrenme ile kişi tespiti alanındaki çalışmalarda bu açığı kapatmak için bir fırsat sunmaktaydı. Bu nedenlerle projemizde yüksek doğruluk, gerçek zamanlı takip sağlayan derin öğrenme tabanlı nesne tespiti modülleri ve iki katmanlı nesne takibi kullanılarak, standart bir kamera ve CPU üzerinde çalışabilen bir sistem geliştirilmiştir. Ayrıca bu teknolojinin kullanım alanları ile ilgili yeni fikirler üretilmiş ve mevcut fikirler geliştirilmiştir.

6. Proje çalışma takımınızı ve çalışma tecrübenizi açıklayın. (Takım çalışması faaliyetleri hakkında bilgi veriniz) (disiplin içi takım çalışması, çok disiplinli takım, bireysel çalışma)

Proje çalışma takımımız iki kişiden oluşmaktadır. Gerekli araçlar ve yöntemlerde karar kılındıktan sonra proje sürecine bu araçları (yazılım dillerini ve kütüphanelerini) öğrenerek başlandı. Eş zamanlı olarak literatürdeki yenilikler takip edilerek proje hedefleri ile örtüşen çalışmalar derinlemesine incelendi. Araçlarda uzmanlaştıkça ve testler devam ettikçe, proje çalışanlarının yetkinlikleri ve ilgi alanları da göz edilerek, sonuçlar doğrultusunda projeye yön verildi.

7. Proje yönetimini nasıl gerçekleştirdiniz. Açıklayınız.

a) İş yükü ve zamanı nasıl paylaştınız?

Proje çalışanları projenin ilk aşamalarında bütün araştırma sürecinde ortak rol almıştır. Ardından iş bölümü yapılmış ve her çalışan öğrendiklerini düzenli olarak diğer çalışan ile paylaşmıştır. Bu süreçte yapılması gerekenler, eksikler ve gerekli notları uzaktan paylaşabilmek için OneNote aracı kullanılmıştır. Ayrıca belirli aralıklarla proje danışmanı Dr.Öğr.Üyesi Erkan Zeki ENGİN ile Microsoft Teams üzerinden toplantılar yapılarak sürece yön verilmiştir.

b) Projede karşılaştığınız riskler ve önlemlerinizi

Projede karşılaşılan en büyük risk Covid-19 salgını nedeniyle test aşamalarında gerekli videoların elde edilememesi ile oluşmuştur. İstenen test koşulları tam olarak sağlanamasa da internet ortamından erişilen videolar ile çalışılmış ve sürece hızlıca adapte olunmuştur. Bu videolardaki farklı koşullarda çalışabilen ortak bir algoritma tasarlamakta projenin zorluğunu arttırmıştır. Ancak proje çalışanlarının

mevcut duruma hızlıca uyum sağlamasıyla bu zorluklar aşılarak projenin kapsamı da gelişmiştir.

c) Öneri raporunda öngörülen konu, yöntem, malzeme ve takvimden sapma/değişiklik oldu mu? Açıklayın.

Öneri raporunda kişi sayma tekniği olarak iki farklı kameradan alınan stereo görüntü ile çalışılması düşünülmekteydi. Ancak ilerleyen süreçte tek bir kamera ile çalışıp derin öğrenme tabanlı kişi tespiti kullanan bir sistemin projenin hedefleriyle daha fazla örtüştüğüne karar verildi. Böylece projenin uygulanabilirliği artmış oldu. Proje sürecinde Covid-19 nedeniyle planlanan test aşamalarını gerçekleştirmek zorlaşmış ve yeterli test koşullarının sağlanabilmesi için bu aşama teslim süresinin ilk olarak Ağustos 2020 olarak belirtilmesi nedeniyle ertelenmişti. Ancak teslim tarihinin tekrar Haziran 2020 olarak düzenlenmesi sebebiyle test aşamasında internet üzerinden erişilebilen videolar kullanılmıştır.