

Videolarda Optik Akış ve Gauss Karışım Modeli Kullanarak Nesne Tespiti

Hozafah Alkhattab¹, Fatih Bayram²

¹Afyon Kocatepe Üniversitesi, Fen Bilimleri Enstitüsü, Mekatronik Mühendisliği Anabilim Dalı, Afyonkarahisar, Türkiye.

²Afyon Kocatepe Üniversitesi, Teknoloji Fakültesi, Mekatronik Mühendisliği Bölümü, Afyonkarahisar, Türkiye.

e-mail: houzifa001@gmail.com ORCID ID: 0009-0000-5509-2393, fatihbayram@aku.edu.tr ORCID ID: 0000-0001-9578-9478.

Geliş/Received : 24.04.2025 ; Kabul/Accepted : 08.05.2025 ; Revize/Revised : 07.05.2025

Öz

Bu çalışma, videolarda nesne tespiti için Optik Akış (Optical Flow), Gauss Karışım Modeli (GMM) ve YOLOv2 kullanan bir yaklaşım önermektedir. Amaç, insan hareketlerini algılayarak tespit performansını artırmaktır. İki farklı yöntem ışığında deneyler gerçekleştirilmiştir: (1) Arka plan çıkarma ile nesne tespiti ve (2) arka plan çıkarılmadan doğrudan nesne tespiti. Deneylerde ölçülen parametreler, zaman performansı ve tanıma hassasiyetidir. Sonuçlar, arka plan çıkarma uygulanmadığında algılama doğruluğunun genel olarak daha yüksek olduğunu (örneğin, "Daria_walk" videosunda %84.47 doğruluk, işlem süresi 3.2553 saniye) ancak arka plan çıkarma yönteminin toplam işlem sürelerini azalttığını (örneğin, "Daria_walk" videosunda %80.23 doğruluk ve 2.999 saniye işlem süresi) göstermektedir. Benzer şekilde, "Daria_jump" videosunda arka plan çıkarma yöntemiyle %67.88 doğruluk ve 2.6983 saniye işlem süresi elde edilirken, arka plan çıkarılmadan %83.83 doğruluk ve 2.811 saniye işlem süresi elde edilmiştir. Her iki yöntem de tüm videolardaki insan nesnesini doğru şekilde tespit etmiştir. Bu araştırma, bilgisayarla görü alanındaki insan aktivitesi tanıma yöntemlerine katkı sağlamakta ve güvenlik, video gözetimi gibi uygulamalarda kullanılabilirliğini göstermektedir.

Anahtar Kelimeler

Optik akış; Gauss karışım modeli; Nesne tespiti; Arka plan çıkarma.

Object Detection in Videos Using Optical Flow and Gaussian Mixture Model

Abstract

This study proposes an approach that utilizes Optical Flow, Gaussian Mixture Model (GMM) and YOLOv2 for detecting objects in videos. The aim is to enhance detection performance by detecting human motion. Experiments were conducted using two different approaches: (1) object detection with background subtraction, and (2) direct object detection without background subtraction. The parameters measured in the experiments were processing time and recognition accuracy. The results indicate that, in general, detection accuracy is higher when background subtraction is not applied (for example, in the "Daria_walk" video, 84.47% accuracy with a processing time of 3.2553 seconds). However, the background subtraction method reduces overall processing time (e.g., in the "Daria_walk" video, 80.23% accuracy and 2.999 seconds processing time). Similarly, in the "Daria_jump" video, background subtraction resulted in 67.88% accuracy and 2.6983 seconds processing time, while without background subtraction, 83.83% accuracy and 2.811 seconds processing time were achieved. Both methods successfully detected the human subject in all videos. This research contributes to human activity recognition methods in the field of computer vision and demonstrates its applicability in areas such as security and video surveillance.

Keywords

Optical flow; Gaussian mixture model; Object detection; Background subtraction.

1. Giriş

Bilgisayarla görü ve görüntü işleme teknolojileri, son yıllarda büyük bir ilerleme kaydetmiş ve birçok sektörde kritik bir bileşen haline gelmiştir. İnsan hareket analizi, video gözetim sistemleri, otonom araçlar, robotik ve spor analizleri gibi alanlarda geniş bir kullanım yelpazesine sahiptir. Özellikle güvenlik ve gözetim uygulamalarında, insan hareketlerinin doğru bir şekilde algılanması ve nesnelerin tanınması büyük önem taşımaktadır.

Literatürde insan aktivitesi tanıma yöntemleri, geleneksel, derin öğrenme ve hibrit mimariler olmak üzere üç ana grupta incelenmektedir. Geleneksel yöntemler, genellikle özellik çıkarımı ve sınıflandırıcılar gibi el ile tasarlanmış tekniklere dayanırken, derin öğrenme yöntemleri, büyük veri kümelerinden otomatik olarak özellik öğrenme kapasitesine sahip modelleri kullanır. Hibrit mimariler ise, geleneksel yöntemlerin sağlam özellik çıkarım yeteneklerini derin öğrenme tekniklerinin güçlü sınıflandırma özellikleriyle birleştirerek daha verimli ve doğru sonuçlar elde etmeyi hedefler. Bu yöntemler, farklı veri türleri ve uygulama alanlarına göre farklı performanslar sergileyebilir ve her birinin avantajları ile sınırlamaları bulunmaktadır.

Bu kapsamda geleneksel yöntemlerin performansını ele alan çalışmalardan biri olan Attal ve arkadaşları (2015), k-En Yakın Komşuluk (k-NN), Destek Vektör Makineleri (SVM), Gauss Karışım Modelleri (GMM), Rastgele Ormanlar (RF), Naïve Bayes ve karar ağacı varyantları gibi denetimli yöntemlerin yanı sıra k-Ortalamalar (k-means) ve Gizli Markov Modelleri (HMM) gibi denetimsiz yöntemleri karşılaştırmıştır. Bu çalışmada, k-NN algoritmasının diğer yöntemlere kıyasla en iyi performansı gösterdiği belirtilmiştir. Benzer şekilde, Amjad ve arkadaşları (2021) tarafından sunulan bir çalışma, el yapımı özellikleri alt uzay havuzlamasıyla birleştiren iki düzeyli hiyerarşik bir yaklaşım önermiştir. Bu yöntem, atomik aktiviteleri tanımlamak için 17 el yapımı özellik kullanmakta ve ardından birleşik aktiviteleri tanımak için bu skorları kullanmaktadır. Ancak, bu yaklaşımın doğruluk oranı %79 (el yapımı) ve %62,8 (alt uzay birleştirme) olarak rapor edilmiştir.

Son yıllarda ise derin öğrenme yaklaşımları insan aktivitesi tanıma alanında ön plana çıkmıştır, özellikle Evrişimli Sinir Ağları (CNN) ve Yinelemeli Sinir Ağları (RNN) benzer görevlerde yüksek başarı göstermektedir. Pushpalatha ve Math (2022), CNN ve Kapılı Yinelemeli Birim (GRU) tabanlı hibrit bir model önererek %96,79 doğruluk oranı elde etmiştir. Khan ve arkadaşları (2024), CNN ve Uzun Kısa Süreli Bellek (LSTM) modellerini kullanarak çoklu modalite verilerini entegre eden bir sistem geliştirmiştir. Bu hibrit model, %94 doğruluk oranı ile yüksek performans sergilemiştir. Zebhi (2022), CNN'leri 2D Hızlı Fourier ve Dalgacık dönüşümleriyle uyarlayarak, ilk seviyede %97,25 ve ikinci seviyede %93,75 doğruluk oranları elde etmiştir.

Derin öğrenmenin başarısına rağmen, belirli uygulama senaryolarında geleneksel görüntü işleme teknikleri de önemli rol oynamaya devam etmektedir. Optik Akış tekniği, hareket eden nesnelerin hız ve yön bilgilerini belirlemek için yaygın olarak kullanılmaktadır. Optik Akış, videolardaki hareketli bölgeleri tespit ederek hem statik hem de dinamik ortamlarda başarılı sonuçlar sunar (Tao vd. 2017). Bunun yanı sıra, Gauss Karışım Modeli (GMM), arka plan modelleme ve nesne segmentasyonu alanında etkin bir çözüm sunmaktadır. GMM, özellikle arka plan çıkarma işlemlerinde sıkça kullanılan bir tekniktir ve farklı aydınlatma koşullarında başarılı sonuçlar vermektedir (Zhao ve Su 2017).

Bu amaçla, çalışmada iki farklı yaklaşım karşılaştırmalı olarak ele alınmıştır: (1) arka plan çıkarma işlemi ile nesne tespiti ve (2) doğrudan nesne tespiti. Arka plan çıkarma yöntemleri, hareketli nesnelerin daha belirgin hale getirilmesini sağlarken, doğrudan nesne tespiti yöntemleri daha kısa işlem süreleri sunabilir. Her iki yöntemin avantajları ve sınırlamaları, deneysel sonuçlarla birlikte değerlendirilmiştir.

Deneylerde kullanılan videolar, farklı insan hareketlerini içermekte ve literatürde sıkça karşılaşılan hareket tespit problemlerine uygun bir temel sunmaktadır. Örneğin, "Daria_walk" ve "Daria_run" videolarında, Optik Akış ve GMM

kullanılarak nesne tespiti başarısı incelenmiştir. Literatürdeki önceki çalışmalara kıyasla, bu çalışmanın yeniliği hem işlem sürelerini hem de tanıma doğruluğunu bir arada değerlendiren kapsamlı bir analiz sunmasıdır. Ayrıca, kullanılan yöntemlerin birbirleriyle karşılaştırılması, bu alandaki uygulama potansiyelini artırmayı amaçlamaktadır.

Bu bağlamda, çalışmanın amacı yalnızca insan nesnesini tanımak değil, aynı zamanda farklı yöntemlerin performansını detaylı bir şekilde incelemektir. Bu analizler, video gözetim, hareket analizi ve diğer bilgisayarla görü tabanlı uygulamalar için yol gösterici nitelikte olacaktır. Çalışmanın bulguları hem literatüre katkı sağlamakta hem de gelecekteki uygulamalar için yol gösterici niteliktedir.

2. Materyal ve Metot

2.1 Veri setleri

Deneylerde, farklı insan hareketlerini içeren dört video kullanılmıştır: “Daria_walk”, “Daria_jump”, “Daria_side” ve “Daria_run”. Bu videolar, her biri farklı hareketlerin analizine olanak tanıyacak şekilde seçilmiştir. Videoların kare sayıları sırasıyla 83, 67, 55 ve 41’dir. Her bir video, hem arka plan çıkarma işlemi uygulanarak hem de doğrudan nesne tespiti yöntemi kullanılarak analiz edilmiştir. Bu videolar Weizmann verisetinden alındı. Weizmann veriseti literatürde sıkça kullanılan bir veri kümesidir (Gorelick vd. 2005). Hareketli görüntülerde insan eylemlerinin tanınması alanında, önerilen farklı algoritmaların karşılaştırılmasında kullanılan Weizmann veri kümesinde 90 tane 180x144 çözünürlükte hareketli görüntü bulunmaktadır.

2.2 Kullanılan yöntemler

2.2.1 Optik akış

Optik akış, bir sahnedeki nesnelerin zaman içindeki hareketlerini analiz etmek amacıyla kullanılan bir görüntü işleme tekniğidir. Bu yöntem, ardışık video karelerinde yer alan aynı nesnenin konum değişikliklerini piksellerin hareketi üzerinden tahmin eder (Horn ve Schunck 1981). Genellikle, her bir

pikselin hareket vektörü (optik akış vektörü) hesaplanarak, sahnedeki genel hareket örüntüsü elde edilir.

Bu teknik, bilgisayarla görü uygulamalarında hareketli nesnelerin tanımlanması, izlenmesi ve hatta hareket yönü ve hızı gibi bilgilerin çıkarılması amacıyla yaygın olarak kullanılır. Tao vd. (2017) tarafından yapılan çalışmalarda, optik akışın insan hareket tanımada güvenilir bir özellik çıkarım yöntemi olduğu gösterilmiştir.

2.2.1.1 Lucas-Kanade ve alternatifleri

Optik akış hesaplamasında farklı algoritmalar kullanılsa da bu çalışmada Lucas-Kanade algoritması tercih edilmiştir. Bu yöntem, küçük hareketlerin varsayıldığı lokal pencerelerde yoğunluk değişimini analiz ederek hareketi tahmin eder (Lucas ve Kanade 1981). Basitliği ve düşük hesaplama maliyeti nedeniyle özellikle düşük çözünürlüklü videolarda tercih edilmektedir (Jain vd. 1995).

Alternatif bir yöntem olan Horn-Schunck algoritması, tüm görüntü alanını kapsayan bir global yaklaşım benimser. Bu yöntem, tüm sahnede sürekli bir akış alanı oluşturarak daha pürüzsüz sonuçlar verir, ancak yüksek hesaplama gücü gerektirir. Lucas-Kanade ise yerel alanlarda işlem yaptığı için daha hızlı ancak düşük ayrıntılı sahnelerde sınırlı performans gösterebilir.

2.2.1.2 Algoritmanın temel varsayımları

- a) **Aydınlık Sabitliği:** Bir pikselin yoğunluğu hareket sırasında değişmez.
- b) **Küçük Hareketler:** İki kare arasındaki hareketin küçük olduğu ve doğrusal bir modelle ifade edilebileceği varsayılır.
- c) **Komşuluk Sürekliliği:** Komşu piksellerin benzer hareket ettiği kabul edilir, bu da denklemleri çözmede istikrar sağlar.

Optik akış hesaplaması (Denklem 1), aydınlık sabitliği varsayımı (brightness constancy assumption) üzerine kuruludur. Bu varsayım, hareket eden bir pikselin yoğunluğunun zamanla değişmeyeceğini ifade eder. Yani:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (1)$$

Bu ifade, Taylor serisi açılımı ile birinci dereceden sadeleştirilerek şu denkleme indirgenir:

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \quad (2)$$

Burada:

- $u = dx/dt$, $v = dy/dt$: x ve y yönlerindeki hareket bileşenleri.
- $\partial I/\partial x$, $\partial I/\partial y$, $\partial I/\partial t$: görüntü gradyanlarıdır.

Bu denklem tek bilinmeyenli iki değişkenli olduğu için doğrudan çözülemez; bu nedenle Lucas-Kanade yöntemi gibi çözücüler, komşu pikselleri de hesaba katarak sistemi çözebilir.

2.2.1.3 Uygulama alanları

Optik akış, sadece insan hareketlerinin analizinde değil, aynı zamanda:

- Otomatik sürüş sistemlerinde engel tespiti,
 - Spor analizlerinde oyuncu takibi,
 - Drone'larda denge kontrolü,
 - Endüstriyel üretim hatlarında kalite kontrol
- gibi birçok alanda başarıyla kullanılmaktadır (Chauhan vd., 2013).

2.2.2 Gauss karışım modeli (GMM)

Gauss Karışım Modeli, video sahnelerindeki arka plan ve ön planı birbirinden ayırt etmek amacıyla kullanılan güçlü bir istatistiksel modelleme yöntemidir. Bu yöntem, her pikselin zaman içerisindeki yoğunluk dağılımını birden fazla Gaussian dağılımı ile temsil ederek, sahnedeki değişkenlikleri matematiksel olarak anlamlandırır. Böylece, statik ve dinamik arka planları kapsayan esnek bir yapı sunar (Stauffer ve Grimson 1999).

Özellikle video gözetimi, güvenlik sistemleri ve trafik analizi gibi uygulamalarda, arka plan modellemesi doğru nesne tespiti için kritik bir adımdır. GMM, bu tür ortamlarda ani aydınlatma değişimleri, gölge hareketleri veya dalgalanan su yüzeyleri gibi karmaşık arka planları başarıyla temsil

edebilmektedir (Zhao ve Su 2017; Zhou ve Zhang 2017).

2.2.2.1 Modelin yapısı ve bileşenleri

GMM, her pikselin değer dağılımını K adet Gaussian bileşen ile temsil eder. Her bileşen aşağıdaki parametrelerle tanımlanır:

- Ortalama (μ_k): Bileşenin merkez değerini ifade eder.
- Varyans (Σ_k): Dağılımın genişliğini ve şeklini gösterir.
- Ağırlık (π_k): Bileşenin modeldeki önem derecesini belirtir.

Matematiksel olarak, GMM'nin olasılık yoğunluk fonksiyonu Denklem 3 ile ifade edilir:

$$P(x) = \sum_{k=1}^K \pi_k \cdot \mathcal{N}(x | \mu_k, \Sigma_k) \quad (3)$$

Burada:

K : Modeldeki Gaussian bileşenlerinin toplam sayısını temsil eder.

Bu yapı sayesinde GMM, hem arka plan hem de zamanla değişen ön plan nesneleri için uygun bir modelleme ortamı sunar.

2.2.2.2 Model eğitimi

Model eğitimi GMM parametrelerinin (μ_k , Σ_k , π_k) tahmini için genellikle Expectation-Maximization (EM) algoritması kullanılır. Bu algoritma iki aşamadan oluşur:

1. Bekleme Aşaması (E-Step): Verilen parametreler ile her bir bileşenin veriler üzerindeki koşullu olasılığı hesaplanır. Bu, her bir verinin hangi bileşene ait olduğunun belirlenmesi için gereklidir.

2. Maksimizasyon Aşaması (M-Step): Model parametreleri, toplam olasılık hatasını minimize edecek şekilde güncellenir. Bu aşama, modelin parametrelerini verilerle uyumlu hale getirir.

Bu işlem iteratif olarak devam eder ve model parametreleri, gözlemleri en iyi açıklayacak şekilde

iyileştirilir. Zhao ve Su (2017), bu yöntemin özellikle düşük çözünürlüklü, gürültülü ve aydınlatma değişkenliği olan sahnelerde istikrarlı sonuçlar verdiğini raporlamıştır.

2.2.2.3 GMM'nin avantajları ve uygulama alanları

GMM, diğer arka plan modelleme yöntemlerine kıyasla aşağıdaki avantajları sunar:

- Aydınlatma değişimlerine karşı dayanıklıdır.
- Dinamik sahnelerde (örneğin: hareketli perde, ağaç yaprakları) ön planı doğru tespit edebilir.
- Uzun vadede öğrenme yeteneği ile çevresel değişimlere adapte olabilir (Chauhan vd., 2013).

Bu nedenlerle GMM, video bazlı insan tespiti, hareket analizi, akıllı gözetim sistemleri ve robotik görüş uygulamalarında yaygın olarak tercih edilmektedir.

2.2.2.4 Çalışmada GMM'nin kullanımı

Bu çalışmada, GMM yöntemi video karelerindeki arka planı modellemek ve hareketli nesneleri daha doğru şekilde ayırtmak amacıyla kullanılmıştır. MATLAB ortamında uygulanan GMM algoritması, optik akış ile birlikte entegre çalıştırılmış ve hareketli bölgeler başarılı şekilde tanımlanmıştır. Elde edilen birleşik maske, YOLOv2 nesne tespiti aşamasına giriş olarak sağlanmış ve sistemin genel doğruluk oranı üzerinde olumlu etkiler gözlemlenmiştir.

2.2.3 YOLOv2 nesne tespiti

YOLOv2 ("You Only Look Once" versiyon 2), gerçek zamanlı nesne tespiti için geliştirilmiş tek aşamalı (single-shot) bir derin öğrenme modelidir. Bu mimari görüntüyü yalnızca bir kez analiz ederek, içerisindeki tüm nesneleri sınıflandırır ve konumlarını tespit eder. Bu özellik, YOLO mimarisinin geleneksel böl-parça-sınıfla (region proposal + classification) mimarilerine göre çok daha hızlı olmasını sağlar (Redmon ve Farhadi 2017).

2.2.3.1 YOLOv2'nin özellikleri

YOLOv2, selefi YOLOv1'e göre birçok geliştirme içermektedir. Bunların başında:

- Anchor boxes kullanımıyla daha hassas kutu tahminleri,
- Batch normalization ile daha hızlı ve stabil öğrenme süreci,
- Darknet-19 isimli yeni bir özellik çıkarım ağı ile daha hafif ama etkili bir yapı yer alır.

YOLOv2 hem hız hem de doğruluk açısından dengeli sonuçlar sunar. Özellikle düşük gecikme gerektiren uygulamalarda (örneğin: video gözetimi, sürücü destek sistemleri, robotik navigasyon) tercih edilmektedir (Choi vd. 2022).

2.2.3.2 Darknet19 ve ön eğitim

Bu çalışmada, önceden "COCO" veri seti üzerinde eğitilmiş olan darknet19-coco modeli kullanılmıştır. Bu model, 19 katmanlı hafif bir konvolüsyonel sinir ağıdır ve düşük donanım kaynaklarında dahi çalıştırılabilecek şekilde optimize edilmiştir. COCO veri seti sayesinde geniş bir nesne yelpazesine dair ön bilgi içerdiğinden, daha küçük veri setleriyle bile başarılı tespit sonuçları alınabilmektedir.

2.2.3.3 YOLOv2'nin güçlü ve zayıf yönleri

YOLOv2'nin en önemli avantajı gerçek zamanlı çalışabilme yeteneğidir. GMM gibi geleneksel yöntemlerle karşılaştırıldığında:

- İşlem süresi çok daha düşüktür.
- Aynı anda birden fazla nesneyi tanıyabilir.
- Sabit çözünürlüklerde yüksek doğruluk sunar.

Ancak bazı zorlukları da mevcuttur:

- Özellikle küçük nesnelerin tespitinde hata oranı artabilir.
- Arka planla karışan renklerde yanlış pozitif tespitler görülebilir.
- Eğitim aşamasında geniş ve dengeli bir veri setine ihtiyaç duyar (Ali vd., 2017).

2.2.3.4 Çalışmada YOLOv2'nin kullanımı

Bu çalışmada YOLOv2 modeli, iki farklı şekilde uygulanmıştır:

1. Arka plan çıkarıldıktan sonra sadece hareketli bölgelerde nesne tespiti için.
2. Doğrudan tüm video karelerinde tam sahne tespiti için.

Her iki senaryoda da modelin başarımları metrikleri (doğruluk ve süre) karşılaştırmalı olarak analiz edilmiştir. GMM + Optical Flow ile elde edilen maske, YOLOv2 girişine uygulanarak hedeflenen nesnelere odaklanma artırılmış, bu da işlem süresini azaltırken doğruluğu optimize etmiştir.

2.3 DeneySEL prosedür

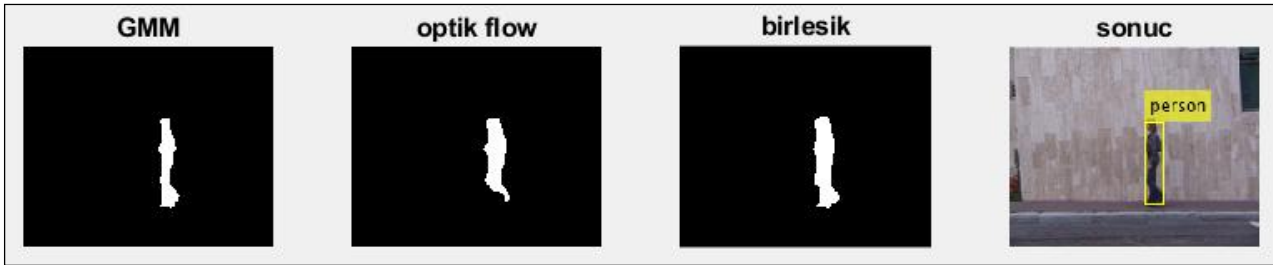
Her video için iki farklı yöntem uygulanmıştır:

a) Arka Plan Çıkarma ile Nesne Tespiti (APÇ): Bu yöntemde, arka plan çıkarma işlemi uygulanarak yalnızca hareketli bölgeler analiz edilmiştir. Optical Flow ve GMM, bu süreçte birlikte kullanılmıştır. Arka plan çıkarıldıktan sonra, YOLOv2 modeli ile hareketli bölgelerdeki nesneler tespit edilmiştir (Şekil 1-4). Bu

dört şekil, çalışmada kullanılan dört farklı video örneği için uygulanan adımları göstermektedir. Her bir şekilde sırasıyla şu adımlar yer almaktadır:

1. GMM (arka plan çıkarma sonucu)
2. Optical Flow (hareket yönü ve hızı)
3. Birleşik maske (GMM + Optical Flow birleşimi)
4. Nesne tespiti sonucu (YOLOv2 ile)

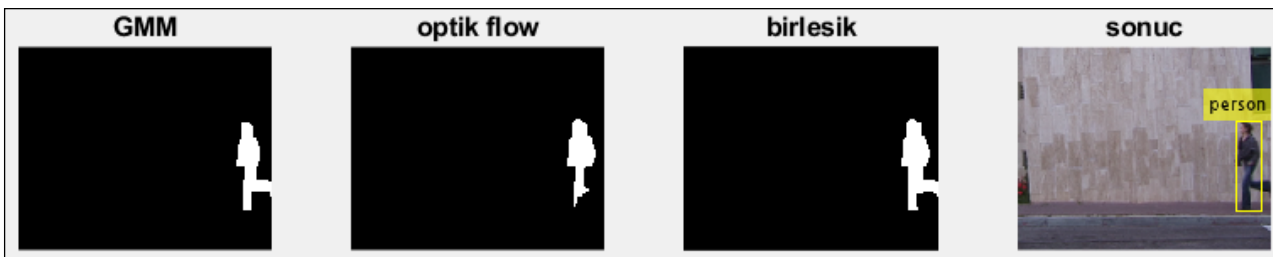
b) Doğrudan Nesne Tespiti (DNT): Bu yöntemde, arka plan çıkarma işlemi uygulanmadan tüm video kareleri YOLOv2 modeli ile analiz edilmiştir. Bu yaklaşım, işlem süresini azaltmakla birlikte, doğruluk oranını etkileyebilmektedir.



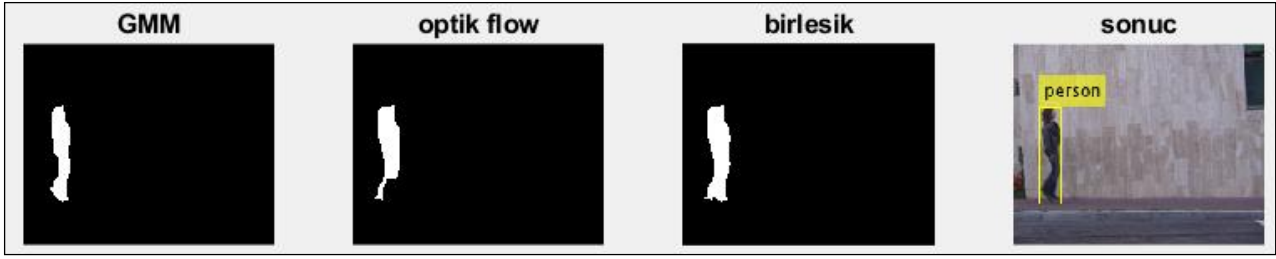
Şekil 1. Daria_walk videosu için, soldan sağa sırasıyla GMM, Optical Flow, birleşik maske ve nesne tespiti sonuçları gösterilmektedir.



Şekil 2. Daria_side videosu için, soldan sağa sırasıyla GMM, Optical Flow, birleşik maske ve nesne tespiti sonuçları gösterilmektedir.



Şekil 3. Daria_run videosu için, soldan sağa sırasıyla GMM, Optical Flow, birleşik maske ve nesne tespiti sonuçları gösterilmektedir.



Şekil 4. Daria_jump videosu için, soldan sağa sırasıyla GMM, Optical Flow, birleşik maske ve nesne tespiti sonuçları gösterilmektedir.

2.4 Performans ölçütleri

Bu çalışmada önerilen iki farklı nesne tespiti yönteminin başarımı, işlem süresi ve tanıma doğruluğu olmak üzere iki temel ölçüt kullanılarak değerlendirilmiştir. Bu ölçütler, gerçek zamanlı video işleme uygulamaları açısından hem hesaplama verimliliği hem de algılama güvenilirliği bakımından kritik öneme sahiptir.

2.4.1 İşlem süresi

İşlem süresi, sistemin bir video dosyasını işleyerek sonuç üretmesi için geçen toplam zamanı ifade eder. Bu ölçüm, her bir video için ayrı ayrı saniye cinsinden raporlanmıştır. İşlem süresinin düşük olması, özellikle:

- Gerçek zamanlı güvenlik kameraları,
- Otonom sistemler (örn. araçlar, dronlar),
- Endüstriyel üretim hatları gibi zaman hassasiyeti gerektiren uygulamalarda büyük avantaj sağlar.

Bu çalışmada, işlem süresi MATLAB ortamında zamanlayıcılar kullanılarak ölçülmüş ve sistemin tüm adımlarını kapsayacak şekilde hesaplanmıştır. Arka plan çıkarma gibi ön işlemler uygulandığında, video içeriğinin yalnızca sınırlı bölümleri işlendiğinden, toplam süre daha düşük çıkmaktadır.

2.4.2 Tanıma doğruluğu

Tanıma doğruluğu, YOLOv2 modeli tarafından yapılan nesne tespitlerinin ne kadarının doğru olduğunu ifade eder. Bu doğruluk, genellikle “Ortalama Doğruluk Oranı (mAP)” veya “Hassasiyet” gibi metriklerle ifade edilir. Bu çalışmada, kullanılan modelin her video için tahmin ettiği nesneler ile gerçek (etiketlenmiş) nesneler karşılaştırılarak doğruluk oranı hesaplanmıştır.

Doğruluğu etkileyen bazı temel faktörler:

- Arka planın karmaşıklığı,

- Hareketli nesnenin boyutu ve netliği,
- YOLOv2’nin sınırlı çözünürlükteki videolardaki performansı,
- Hatalı maskeleme (özellikle GMM’den kaynaklanan).

2.4.3 Ölçütlerin seçim nedeni

Bu iki ölçüt, sistemin hem verimliliğini hem de güvenilirliğini dengeleyen göstergeler sunar. İşlem süresi kısa ancak doğruluk düşükse sistem hızlı ama güvenilmezdir. Tersine, doğruluk çok yüksek fakat işlem süresi fazlaysa gerçek zamanlı uygulamalarda yetersiz kalabilir. Bu nedenle, her iki kriterin birlikte değerlendirilmesi, sistemin genel performansını adil biçimde ölçmek için önemlidir.

2.5 Kullanılan araçlar

Çalışmanın tüm deneysel süreçleri hem algoritmaların uygulanabilirliğini hem de performanslarının adil şekilde karşılaştırılmasını sağlamak amacıyla aynı yazılım ve donanım ortamında gerçekleştirilmiştir. Kullanılan araçlar, görüntü işleme, matematiksel modelleme ve nesne tespiti işlemleri için uygun olan, akademik çevrelerde sıkça tercih edilen bileşenlerden oluşmaktadır.

2.5.1 Yazılım ortamı

Deneylerde kullanılan temel yazılım ortamı MATLAB’dır. MATLAB, güçlü görüntü işleme kütüphaneleri, matris tabanlı hesaplama altyapısı ve görselleştirme araçlarıyla akademik çalışmalarda tercih edilen bir platformdur.

- **Optik Akış Uygulaması:** Lucas-Kanade algoritması, MATLAB’ın opticalFlowLK sınıfı kullanılarak gerçekleştirilmiştir.

- **GMM Uygulaması:** Arka plan modelleme için MATLAB'taki GMM fonksiyonları veya özelleştirilmiş kodlar kullanılmıştır. Bu sayede piksellerin zaman içerisindeki renk değişimleri çoklu Gaussian dağılımları ile analiz edilmiştir.
- **YOLOv2 Uygulaması:** Deep Learning Toolbox kullanılarak önceden eğitilmiş darknet19-coco modeli yüklenmiş ve video kareleri üzerinde nesne tespiti işlemleri yapılmıştır.

MATLAB ortamı sayesinde hem klasik hem de derin öğrenme tabanlı yöntemlerin aynı çatı altında entegre biçimde uygulanabilmesi sağlanmıştır.

2.5.2 Donanım yapılandırması

Tüm deneyler, sabit bir donanım konfigürasyonu altında gerçekleştirilmiştir. Böylece farklı yöntemler arasındaki işlem süresi farkları yalnızca algoritmaların verimliliğinden kaynaklanacak şekilde izole edilmiştir.

- **İşlemci:** Intel(R) Core(TM) i5-9300H CPU @ 2.40GHz
- **RAM:** 16 GB
- **GPU:** NVIDIA GeForce 1650
- **İşletim Sistemi:** Windows 11 x64

YOLOv2 modeli, MATLAB üzerinden CPU veya GPU üzerinde çalıştırılabilmektedir. Bu çalışmada, özellikle işlem sürelerinin kıyaslanabilirliğini sağlamak adına, tüm işlemler yalnızca CPU üzerinde yürütülmüştür.

3. Bulgular

Bu bölümde, çalışmada elde edilen sonuçlar detaylı bir şekilde sunulmuş ve tartışılmıştır. Her iki yöntemle yapılan analizlerin sonuçları, nesne tespit süresi ve tanıma doğruluğu gibi temel ölçütler açısından değerlendirilmiştir.

Arka plan çıkarma yöntemi kullanılarak yapılan analizlerde, nesne tespit süresi açısından avantaj sağlandığı görülmüştür. Örneğin, "Daria_run" videosunda bu yöntem ile nesne tespit süresi 2.171 saniye olarak ölçülmüştür. Bununla birlikte, aynı video için tanıma doğruluğu %76,46 olarak kaydedilmiştir. Bu sonuçlar, arka plan çıkarma

yönteminin işlem süresi açısından avantajlı olduğunu ancak doğruluk oranında bir miktar kayıp yaşanabileceğini göstermektedir.

Doğrudan nesne tespiti yöntemi, tespit doğruluğu açısından daha yüksek sonuçlar sağlamıştır. Örneğin, "Daria_walk" videosunda bu yöntem ile %84,47 doğruluk ve 3.255 saniye tespit süresi elde edilmiştir. Ancak bu yaklaşım, arka plan çıkarma yöntemine kıyasla daha uzun tespit süreleri gerektirmektedir (Tablo 1).

Tablo 1. Arka Plan Çıkarma (APÇ) ve Doğrudan Nesne Tespiti (DNT) Yöntemlerinin Süre ve Doğruluk Oranlarına Göre Karşılaştırılması

Video	Frame	APÇ	APÇ	DNT	DNT
		süre (s)	doğruluk (%)	süre (s)	doğruluk (%)
daria_walk	83	2.999	80.23	3.255	84.47
daria_side	55	2.286	85.00	2.419	85.77
daria_run	41	2.172	76.46	2.206	82.88
daria_jump	67	2.698	67.88	2.811	83.83

Her iki yöntemin performansı, farklı videolar ve hareket türleri için karşılaştırılmıştır. Genel olarak, doğrudan nesne tespiti yöntemi daha yüksek doğruluk oranlarına sahipken, arka plan çıkarma yöntemi nesne tespit süresi açısından daha avantajlıdır. Bununla birlikte, video boyutlarının daha büyük olduğu senaryolarda, arka plan çıkarma yönteminin sağladığı zaman avantajının daha belirgin hale gelmesi beklenmektedir. Literatürde yapılan çalışmalar da (Zhao ve Su 2017) video çözünürlüğünün artması durumunda, GMM gibi arka plan çıkarma yöntemlerinin işlem sürelerinde önemli iyileşmeler sağladığını göstermektedir. Bu bağlamda, daha büyük boyutlu videolar üzerinde yapılacak ileri çalışmalar, yöntemin bu avantajını daha net bir şekilde ortaya koyabilir. Ayrıca tüm videolarda, her iki yöntem de insan nesnesini doğru bir şekilde tespit etmeyi başarmıştır.

4. Tartışma ve Sonuç

Bu çalışmada, Optical Akış ve Gauss Karışım Modeli kullanılarak nesne tespiti performansı iki farklı yöntemle incelenmiştir. Elde edilen sonuçlar, arka plan çıkarma yönteminin işlem süresi açısından avantaj sağladığını, doğrudan nesne tespiti

yönteminin ise daha yüksek doğruluk oranları sunduğunu göstermiştir. Ancak video boyutlarının artması durumunda, arka plan çıkarma yönteminin işlem süresi avantajının daha belirgin hale gelebileceği öngörülmektedir (Zhao ve Su 2017).

Çalışmanın bir diğer önemli sonucu ise yöntemlerin sabit arka planlı videolar üzerindeki uygulama potansiyelidir. Özellikle güvenlik kameraları gibi sabit arka planlı sistemlerde, arka plan çıkarma yöntemi kullanılarak işlem süresinin ciddi oranda azaltılabileceği gösterilmiştir. Bu tür sistemlerde, kameranın ilk kurulumu sırasında arka planın belirlenmesi ve modele tanıtılması ile bu avantaj daha da optimize edilebilir. Böylece gerçek zamanlı gözetim sistemlerinde daha verimli bir nesne tespiti sağlanabilir.

Elde edilen sonuçlar ve literatürdeki çalışmalar doğrultusunda, nesne tespiti performansını daha da iyileştirmek için aşağıdaki öneriler sunulmaktadır:

- a. Model Güncellemeleri:
 - YOLOv2 yerine daha güncel modeller (YOLOv4, YOLOv5 veya YOLOv8) kullanılarak doğruluk artırılabilir.
 - Özellikle küçük nesnelerin tespitinde performans iyileştirmeleri sağlanabilir.
- b. Maskeler Üzerinde Geliştirme:
 - GMM ve Optical Flow sonuçlarının daha akıllı şekilde birleştirildiği mask fusion teknikleri uygulanabilir.
 - Segmentasyon bazlı bölgesel filtreleme kullanılabilir.
- c. Donanım Tabanlı İyileştirme:
 - FPGA gibi donanım hızlandırıcıları kullanılarak işlem süreleri önemli ölçüde azaltılabilir (Zhao ve Su, 2017).
- d. Veri Seti Çeşitliliği:
 - Daha karmaşık sahneler ve farklı kamera açılarında kaydedilmiş verilerle modelin genellenebilirliği artırılabilir.
 - Weizmann veri kümesine ek olarak UCF101 veya Kinetics gibi daha büyük veri setleriyle test yapılabilir.
- e. Gerçek Zamanlı Uygulama Prototipi:

- Önerilen sistem, bir gözetim kamerası üzerinde gerçek zamanlı olarak test edilerek saha koşullarındaki performansı ölçülebilir.

5. Referanslar

- Ali, S.T., Goyal, K. ve Singhai, J., 2017. Moving Object Detection Using Self-Adaptive Gaussian Mixture Model for Real-Time Applications. *Proceedings of the International Conference on Signal Processing and Embedded Systems (RISE)*, IEEE.
- Amjad, M., Khan, A., Siddiqi, M.H., Lee, S. ve Muhammad, N., 2021. Hierarchical human activity recognition using wearable sensors based on subspace pooling. *Sensors*, **21**(7), 2368.
- Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L. ve Amirat, Y., 2015. Physical human activity recognition using wearable sensors. *Sensors*, **15**(12), 31314-31338.
- Chauhan, A.K., Krishan, P. ve Kumar, D., 2013. Moving Object Tracking Using Gaussian Mixture Model and Optical Flow. *International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE)*, **3**(4), 243–246.
- Choi, H., Kang, B. ve Kim, D., 2022. Moving Object Tracking Based on Sparse Optical Flow with Moving Window and Target Estimator. *Sensors*, **22**(8).
- Horn, B.K.P. and Schunck, B.G., 1981. Determining optical flow. *Artificial Intelligence*, **17**, 1–3, 185–203.
- Jain, R., Kasturi, R. ve Schunck, B.G., 1995. *Machine Vision*. McGraw-Hill.
- Jiang, Y., Wang, J., Liang, Y. ve Xia, J., 2018. Combining Static and Dynamic Features for Real-Time Moving Pedestrian Detection. *Multimedia Tools and Applications*, Springer.
- Khan, N., Mehmood, W., Iqbal, T., Ahmad, S., Rehman, A. ve Rho, S., 2024. Multimodal human activity recognition using hybrid CNN-LSTM network. *Heliyon*, **10**(4), e22371.
- Lucas, B.D. and Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI)*, **2**, 674–679.
- Meghana, R.K., Chitkara, Y., Apoorva, S. ve Mohana, 2019. Background-Modelling Techniques for Foreground Detection and Tracking Using Gaussian Mixture Model. *Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE.
- Mohana ve Aradhya, H.V.R., 2020. Performance Evaluation of Background Modeling Methods for

- Object Detection and Tracking. Proceedings of the Fourth International Conference on Inventive Systems and Control (ICISC 2020), IEEE.
- Pushpalatha, K. ve Math, S., 2022. Hybrid deep learning model for human activity recognition using convolutional neural network and gated recurrent unit. *International Journal of Intelligent Technologies and Applied Statistics*, **15**(1), 1-13.
- Redmon, J. ve Farhadi, A., 2017. YOLO9000: Better, Faster, Stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, **1**, 6517–6525.
- Šilar, Z. ve Dobrovolný, M., 2013. The Obstacle Detection on the Railway Crossing Based on Optical Flow and Clustering. Proceedings of the Telecommunications and Signal Processing Conference (TSP), IEEE.
- Stauffer, C. and Grimson, W.E.L., 1999. Adaptive background mixture models for real-time tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, **2**, 246–252.
- Tao, D., Li, X., Wu, X. ve Maybank, S.J., 2017. Human action recognition based on multiple features using Gaussian mixture models. *IEEE Transactions on Image Processing*, **26**(3), 1377-1390.
- Thotapalli, P.K., Kumar, C.R.V. ve Reddy, B.C.M., 2021. Feature Extraction of Moving Objects Using Background Subtraction Technique for Robotic Applications. *International Journal of Intelligent Robotics and Applications (IJIRA)*, Springer.
- Yi, Q., 2018. Design of Moving Object Detection System Based on FPGA. Proceedings of the 10th International Conference on Communications, Circuits, and Systems (ICCCS), IEEE.
- Zebhi, B., 2022. 2D discrete wavelet and Fourier transform-based convolutional neural networks for human activity recognition. *Computers in Biology and Medicine*, **141**, 105072.
- Zhao, X. ve Su, Y., 2017. A dynamic approach for background modeling using Gaussian mixture models. *Computer Vision and Image Understanding*, **157**, 36-49.
- Zhou, D. ve Zhang, H., 2017. Modified GMM Background Modeling and Optical Flow for Detection of Moving Objects. Proceedings of the National Conference on Electronics and Computer Systems (NECS), IEEE.