

Nesne Takibi Yöntemlerinin Karşılaştırması ve Çekirdek Korelasyon Filtresinin Farklı Görünüm Modelleri ile Performans Analizi

Comparison of Object Tracking Methods and Performance Analysis of Kernelized Correlation Filter with Different Appearance Models

Elnura Musaoğlu^{1,2} ve Ceyda Nur Öztürk¹

¹Bilgisayar Mühendisliği Bölümü, Bursa Uludağ Üniversitesi, Bursa, Türkiye

²Özveri Ar-Ge Merkezi Özdilek Ev Tekstil San. ve Tic. A.Ş. Bursa, Türkiye
elnuraarslan@gmail.com, ceydanur@uludag.edu.tr

Özetçe—Nesne takibi hareketli görüntüler içerisinde her çerçevede nesne tespitine gerek kalmadan ilgi nesnelerinin belirlenmesini sağlar. Literatürde bu alanda yapılmış korelasyona, derin öğrenmeye ve farklı yöntemlere dayanan çeşitli çalışmalar mevcuttur. Bu çalışmalarda takip edilecek nesneyi modellemek için Ölçek Değişmez Özellik Dönüşümü (Scale Invariant Feature Transform - SIFT), Yönelimli Eğimlerin Histogramı (Histogram of Oriented Gradients - HOG), hareket modeli, evrişimli sinir ağı özellikleri gibi farklı görünüm modelleri kullanılmaktadır. Korelasyon yaklaşımı Fourier uzayında yapılan işlemlerle nesne takip sistemlerini gerçek zamanlı çalıştırmada etkindir. Korelasyon filtresini evrişimli sinir ağı ile kullanarak nesne takibi başarısı artırılabilir. Bu çalışmada amaçlanan farklı nesne takibi yöntemlerini karşılaştırmak ve farklı görünüm modellerinin nesne takibine etkisini analiz etmektir. Farklı ortam ve koşullarda hazırlanmış 100 görüntü serisi içeren nesne takibi karşılaştırma (OTB-100) veri seti kullanılmıştır. Performans ölçümü için kesinlik (precision) ve başarı (success) olmak üzere iki ölçüt hesaplanmıştır. OTB veri seti üzerinde yapılan deneylerden elde edilen sonuçlara bakıldığında korelasyon filtresi ile evrişimli sinir ağının birlikte kullanımının nesne takip başarısını artırdığı görülmüştür.

Anahtar Kelimeler — Nesne Takibi , Evrişimli Sinir Ağı, Çekirdek Korelasyon Filtresi, Derin Özellikler

Abstract— Object tracking enables determination of relevant objects in the moving images without the need for object detection in each frame. There are various studies that are based on correlation, deep learning and different methods in this field in the literature. In these studies, different appearance models such as Scale Invariant Feature Transform (SIFT), Histogram of Oriented Gradient (HOG), motion model, convolutional neural network features are used to model the object to be tracked. Having operations performed in Fourier space, the correlation approach is effective in making object tracking systems run in real time. Using the correlation filter with the convolutional

neural network, the success of object tracking can be increased. The aim of this study is to compare different object tracking methods and analyze the effect of different appearance models on object tracking. Object tracking comparison (OTB-100) data set which includes 100 images series prepared in different environments and conditions, was used. Two measures of precision and success were computed for performance measurement. When the obtained results of the experiments on the OTB data set were evaluated, it was observed that the combined use of the correlation filter and convolutional neural network increased the object tracking performance.

Keywords — Object Tracking, Convolutional Neural Network, Kernelized Correlation Filter, Deep Features

I. GİRİŞ

Nesne takibi bilgisayarla görme alanında önemli bir yere sahiptir. Hareketli ya da sabit kameralardan alınan bir dizi görüntü üzerinde hedef nesnelerin durumunun izlenmesi işlemi olarak tanımlanabilir. Nesne takibinin etkin olarak yer aldığı pek çok uygulama alanı bulunmaktadır. Bu uygulama alanlarına örnek olarak otonom araç ve robotlar, video izleme, insan-makine etkileşimi, tıbbi uygulamalar, artırılmış gerçeklik, video indeksleme, trafik kontrolü ve benzeri verilebilir. Nesne takibi çalışmalarında sıklıkla karşılaşılan problemler aydınlatma değişiklikleri, hızlı hareket, poz değişimleri, kısmi tıkanmalar ve arka plan karmaşıklığı gibi durumlardır. Bu gibi durumlarda nesne takibi zorlaşmaktadır. Mevcut çalışmaları incelendiğinde her zorluk durumu için başarılı bir algoritma geliştirmenin zor olduğu ve bu nedenle çalışmaların belirli bir alt problemin çözümüne odaklandığı görülmektedir.

Video görüntülerindeki hareketli nesnelerin tespiti ve takibi için literatürde farklı yöntemler mevcuttur. Optik akış, Kalman filtresi, korelasyon filtresi, parçacık filtresi bu yöntemlerde yaygınlaştıran yaklaşımlardır. Klasik yaklaşımlarda

alınan görüntü üzerinde basit bir hedef model oluşturularak ve daha sonraki görüntüler üzerinde bu model ile arama yapılarak nesne takibi işleminin gerçekleştirildiği görülmektedir. Ortalama kayma takipçisi [1] yönteminde hedef model renk histogramlarıyla modellenmekte ve sonraki çerçevelerdeki aday imge bölgelerinin renk histogramları ile karşılaştırma yapılarak hedef modelin konumu bulunmaktadır.

AdaBoost özellik çıkarma algoritmasının bir versiyonu olan Boosting yöntemi çevrim içi eğitim işlemini gerçekleştirmektedir. Bu özelliği ile takip edilen nesnenin görünüm değişikliklerine karşı hızlı adapte olabilmektedir. Nesne görünüm modellerinin hızlı hesaplanabilir olması bu yöntemin gerçek zamanlı çalışmasına imkân sağlamaktadır. Gerçek zamanlı Çoklu Örnekli Öğrenme (Multiple Instance Learning - MIL) yöntemi nesneyi arka plandan ayırt etmek için çevrim içi ayırt edici bir sınıflandırıcı eğitmektedir. Eğitim için tek bir örnek yerine bir dizi görüntü yaması kullanılmaktadır. MedianFlow çalışmasında izleme başarısızlık tespiti için ileri-geri hatasına dayanan yeni bir yöntem önerilmektedir. Geriye dönük izleme gerçekleştirilerek bir doğrulama yörüngesi oluşturulmakta ve söz konusu yörünge ile karşılaştırılmaktadır. Takip, öğrenme ve algılama (Tracking, Learning and Detection - TLD) yönteminin takip adımında nesnenin çerçeveden çerçeveye takibi, algılama adımında nesne görünüm modellerinin konumlandırılması ve takip adımının düzenlenmesi, öğrenme adımında ise algılama adımının hatalarını önlemek için güncelleme işlemleri gerçekleştirilmektedir. Ayrımcı korelasyon filtresi (Discriminative Correlation Filter - DCF) yöntemine dayanan Kanal ve Uzaysal Güvenilirlik İzleyicisi (The Channel and Spatial Reliability Tracker - CSRT) filtre boyutunu ayarlayabilmesi ile geleneksel DCF algoritmasını iyileştirmektedir [2].

Bolme ve ark. [3] çalışmasında hedeflerin görünümünü adaptif korelasyon filtreleri ile modellemekte ve her bir hedef için Gauss cevabı oluşturmaktadır. Çevrim içi model görünüm güncellemesi yaparak Fourier alanında korelasyona dayalı nesne takibi işlemini saniyede 669 görüntü işleyecek şekilde gerçekleştirmektedir. Henriques ve ark. [4] çalışmasında yönelimli eğimlerin histogramı özelliği ile hedef nesnenin dairesel kaymalarını hesaplayarak korelasyon tabanlı hedef nesne takibi işlemini gerçekleştirmektedir. Gauss kerneli kullanarak doğrusal korelasyonu çok kanallı korelasyon filtresine çevirmekte ve böylece Fourier alanında hızlı çalışacak korelasyon filtresi ile hedef nesne takibini gerçekleştirmektedir.

Bayes yöntemine dayanan Kalman filtresi modelin önceki bilgilerini kullanarak sistemin yeni durumunu tahmin etmeyi amaçlar. Yinelemeli Kalman filtresi tahmin ve doğrulama adımlarından oluşmaktadır [5]. Fan ve Wang [6] yaptıkları çalışmada Bayes yöntemine dayanan Kalman filtreleme yönteminin nesne sayısı küçük kaldığında çoklu hedef izlemede etkili iken, nesnelerin sayısı arttığında, kimlik bilgisi sayısı daha sık hale geldiği durumda ve yöntemin özyinelemeli yapısı nedeniyle düzeltilmesinin zorlaştığını ifade etmektedir.

Derin öğrenme kullanılarak yapılan hedef nesne takibi çalışmaları [7] [8] düşük seviyeli özellikler yerine evrişimli sınır ağının katmanlarından elde edilen yüksek seviyeli

özellikleri kullanarak hedef model oluşturma takip başarısını artırdığını göstermektedir. Ma ve ark. [8] çalışmasında hedef görünüm modeli için evrişimli sınır ağı katmanlarını hiyerarşik yapıda kullanılmaktadır ve çekirdek korelasyon filtresi uygulayarak nesne takibini gerçekleştirmektedir.

Bu çalışmada video görüntülerindeki nesnelerin takip edilmesi için farklı takip yöntemleri ve farklı görünüm modelleri kullanılmaktadır. Kalman filtresi, karesel hatanın minimum çıktısı (Minimum Output Sum of Squared Error-MOSSE) [3], çekirdek korelasyon filtresi (Kernelized Correlation Filter - KCF) [4], hiyerarşik evrişim özellikleri ve çekirdek korelasyon filtresi (Hierarchical Convolutional Features - HCF) [8] ile nesne takibi gerçekleştirilmektedir. Ayrıca OpenCV kütüphanesi kullanılarak Boosting, MIL, MedianFlow, TLD, CSRT yöntemleri de nesne takibi analiz ve karşılaştırmaları için kullanılmaktadır. KCF yöntemi için gri düzey yoğunluklar, RGB renk histogramı, HOG, VGG-19 katmanlarından elde edilen özellikler görünüm modeli olarak seçilerek farklı özelliklerin takip başarısına etkisi analiz edilmektedir.

II. MATERYAL VE METOT

A. Veri Seti

OTB-50 ve OTB-100 adlı iki bölümden oluşan OTB [9] veri seti tekli nesne takibi yöntemlerinde yaygın olarak kullanılan veri setlerinden birisidir. Farklı ortam koşullarında elde edilmiş 100 görüntü dizisinden oluşmaktadır. Bu çalışmada OTB-100 veri seti, taban gerçeklik için bu veri setinin sağladığı değerler kullanılmaktadır.

B. Nesne Takibi Yöntemleri

Nesne takibi amacıyla kullanılan yöntemlerden ilki MOSSE yöntemidir. Korelasyon filtresini kullanan MOSSE yönteminde işlem süresinin hızlandırılması amacıyla korelasyon işlemi Fourier uzayında gerçekleştirilir. Takip için kullanılan hedef nesnenin Fourier uzayındaki Gauss cevabı (G), görüntünün Fourier uzayına dönüştürülmüş formu (F) ve filtrenin Fourier uzayına dönüştürülmüş formu (H) ile hesaplanmaktadır (1). Denklem (1)'de kullanılan \odot sembolü eleman bazında çarpımı, $*$ ise kompleks eşleniği temsil etmektedir.

$$G = F \odot H^* \quad (1)$$

MOSSE gerçek evrişim çıktısı ile istenen evrişim çıktısı arasındaki hata kareleri toplamını minimize eden H filtresini hesaplamaktadır (2). Filtre takip işlemi sırasında çevrim içi güncellenerek hedefin görünüm değişikliklerine adapte olması sağlanır. Nesne takibi için kullanılacak filtre Fourier uzayında F ve G kullanılarak bulunur (3).

$$\min_{H^*} \sum |F_i \odot H^* - G_i|^2 \quad (2)$$

$$H^* = (\sum_i G_i \odot F_i^*) / (\sum_i F_i \odot F_i^*) \quad (3)$$

Bir diğer nesne takibi yöntemi ise hedef nesnenin dairesel kaymalarını hesaplayarak korelasyon tabanlı nesne takibi işlemini gerçekleştiren KCF'tir. Hedef nesnenin sonraki

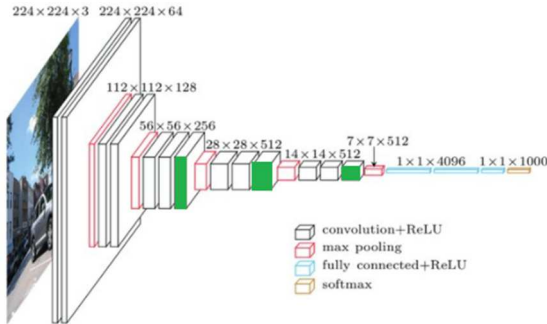
görüntüler üzerinde takip edilmesi için renk yoğunluk bilgisi ya da HOG özneliliği kullanılmaktadır. Görüntüdeki birinci dereceden türevlerin yönelim ve genlik değerlerinin histogramlarla ifade edilmesini sağlayan HOG nesne algılama alanında sıkça kullanılan bir özellik tanımlayıcıdır. KCF takip problemini tepe regresyon problemi olarak ele alır. Tepe regresyon basit bir kapalı form çözümünü kabul eder ve daha karmaşık yöntemlere yakın bir performans elde edebilir. Amaçlanan z sonraki görüntüde görüntü yamasının özellik vektörü iken x_i örnekleri ve bunların regresyon hedefleri y_i üzerindeki hata karesini en aza indiren bir $f(z) = w^T z$ fonksiyonu bulmaktır (4). Denklemden ifade edilen λ parametresi aşırı öğrenmeyi önlemek için kullanılan düzenleme parametresidir.

$$\min_w \sum_{i=1}^n (f(x_i) - y_i)^2 + \lambda ||w||^2 \quad (4)$$

Korelasyon filtresinin Fourier uzayında çözümlenmiş formu Denklem (5)'te ifade edilmektedir. Bu filtre hedef nesnenin sonraki görüntüdeki konumunu hesaplamada kullanılmaktadır.

$$w = \frac{x^* \odot y}{x^* \odot x + \lambda} \quad (5)$$

HCF hedef nesnenin takip edilmesi için derin öğrenme yöntemlerinden biri olan VGG-19 evrişimli sinir ağı ile çekirdek korelasyon filtresini kullanmaktadır. VGG-19 ağı Şekil 1'de gösterildiği gibi 16 evrişim, 3 tam bağlı (FC), 5 maxpool ve 1 softmax olmak üzere toplamda 19 katman içermektedir. Bu yöntemde Şekil 1'de yeşil renk ile ifade edilen evrişim katmanları çıktılar ile üç farklı korelasyon filtresi çalıştırılmaktadır. Hedefin konumunu hesaplamak için üç korelasyon çıktısının sonuçları hiyerarşik yapıda kullanılmaktadır. Son katmandan önceki katmanlara doğru hiyerarşik yapıyı oluşturan yöntemde amaç son katmanın korelasyon sonucuna fazla ağırlık vererek bunun takip sonucuna daha fazla etki etmesini sağlamaktır. Bunun nedeni evrişimli sinir ağı modelinde son katmanların nesne takibi başarısına etkisinin fazla olduğunun düşünülmesidir.



Şekil 1. VGG-19 derin öğrenme ağı [10]

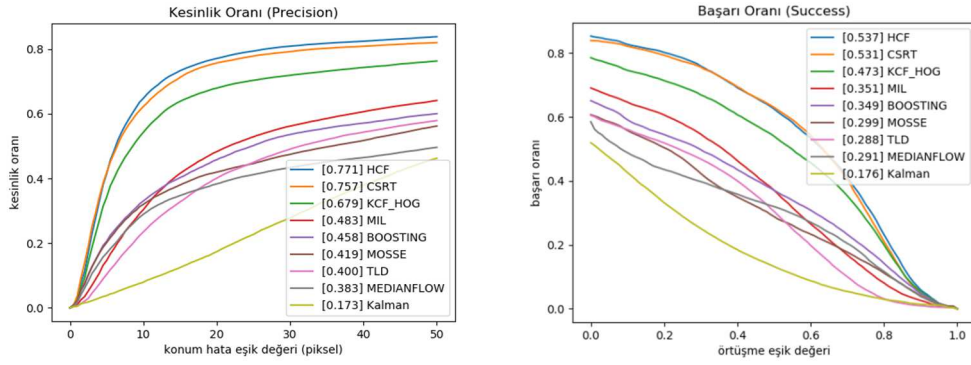
III. BULGULAR

Deney çalışmasının ilk aşamasında OTB-100 veri seti kullanılarak Kalman filtresi, MOSSE, HOG görünüm modeli ile KCF (KCF HOG), VGG-19 ağı 3 evrişim katmanından elde edilen özellikler ile KCF [HCF] ve OpenCV kütüphanesinin sağladığı takip yöntemleri karşılaştırılmaktadır.

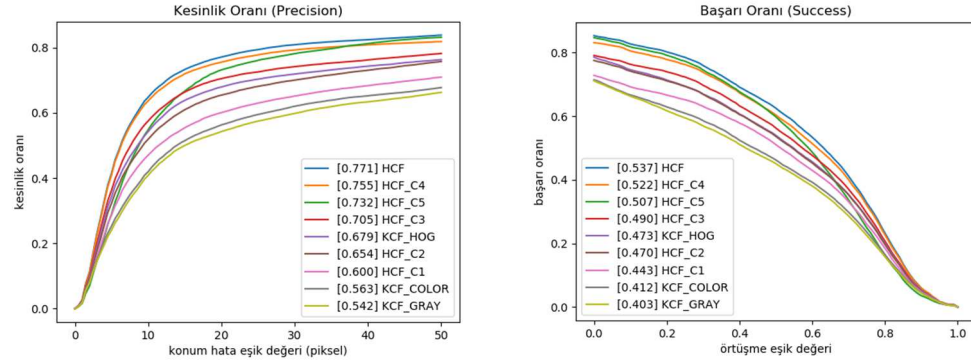
HCF yönteminde conv3_4, conv4_4 ve conv5_4 katmanları için kullanılan ağırlıklar sırasıyla 0.02, 1, 0.5'tir. Hedef konumunun belirlenmesi için kullanılan bu ağırlıklar ile sonuca en başarılı korelasyon çıktısının etki etmesi sağlanmaktadır. Deney çalışmasının sonraki aşamasında ise çekirdek korelasyon filtresinin farklı görünüm modelleri ile kullanımının takip başarısına etkisi analiz edilmektedir. Bu aşamada KCF yöntemi ile görüntü yoğunluklarının gri modeli, RGB renk histogram modeli ve ImageNet veri seti ile eğitilmiş VGG-19 ağında bulunan evrişim katmanlarından conv1_2, conv2_2, conv3_4, conv4_4 ve conv5_4 katmanlarından gelen özellikler kullanılmaktadır.

Deney çalışmasında 2 çeşit performans ölçütü yöntemi kullanılmaktadır. Bunlardan ilki kesinlik ikincisi ise başarı olarak ifade edilmektedir. Kesinlik ölçütü için hedef nesnenin tahmin konumu merkez noktası ile gerçek konum merkez noktası arasındaki fark Öklid yöntemi ile hesaplanmaktadır. Kesinlik performans ölçütü 20 piksel eşik değeri için hesaplanan mesafe ölçümleri içerisinde 20 sayısına eşit ya da küçük olanların sayısının toplam görüntü sayısına bölünmesi ile hesaplanır. İdeal nesne takibi yöntemlerinde istenen kesinlik değerinin büyük olmasıdır. Başarı izlenen hedeflerin çevreleyici kutuları ile gerçek çevreleyici kutuların kesişimlerinin ortak alana oranı (IoU) ile hesaplanmaktadır. Başarı ölçütü için kullanılan eşik değeri 0.5'tir. Her görüntü dizisi için tahmin edilen çerçeve ile gerçek çerçevenin kesişimlerinin ortak alana oranı 0.5 ve daha büyük olan değerlerin ilgili görüntü dizisindeki toplam görüntü sayısına oranı başarı değerini vermektedir. 100 veri seti için hesaplanan kesinlik ve başarı değerlerinin ortalaması performans ölçütündeki kesinlik ve başarı değerini vermektedir.

Şekil 2'de farklı takip yöntemlerinin performans değerlendirmesi gösterilmektedir. Kesinlik ve başarı sonuçları değerlendirildiğinde en başarılı yöntem HCF iken en başarısız yöntem Kalman filtresidir. Kalman filtresi 10031 fps (frame per second) ile en yüksek saniye başına çerçeve değerine sahip olmasına rağmen yüksek başarı oranı gerektiren çalışmalarda tercih edilmemektedir. HCF için evrişimli sinir ağının hiyerarşik katmanlarından gelen semantik ve uzaysal özelliklerin bir arada kullanılması ile nesne takibinde başarı oranının arttığı görülmektedir. Ancak VGG-19 ağının üç evrişim katmanı ile üç korelasyon filtresini çalıştırması fps değerini düşürdüğü için HCF yüksek işlem gücüne sahip GPU'lu bilgisayarın mevcut olması durumunda gerçek zamanlı uygulamalarda tercih edilebilmektedir. OpenCV kütüphanesindeki yöntemlerden en başarılı olan CSRT ortalama 0.757 kesinlik ve 0.531 başarı değerleri ile 2. en başarılı sonucu vermektedir ve 62 fps değeri ile gerçek zamanlı işlem gücü kapasitesine sahiptir. KCF_HOG yöntemi ise 175 fps ile CSRT yönteminden yaklaşık üç kat hızlı çalışmaktadır. MIL yöntemi ise düşük kesinlikleri ve 21 fps hızı ile yüksek başarı ve gerçek zamanlı çalışma gerektiren nesne takibi çalışmalarında tercih edilmemektedir. BOOSTING yöntemi, MOSSE'ye kıyasla daha başarılı sonuç vermektedir, ancak MOSSE yaklaşık 10 kat daha hızlı çalışmaktadır. TLD yönteminin başarı sonucu MOSSE'ye yakın iken yaklaşık 14 kat daha yavaş çalışmaktadır. MEDIANFLOW yöntemi ise yüksek 1395 fps işleyebilmesine rağmen Kalman filtresinden sonraki en başarısız yöntemdir.



Şekil. 2. Nesne takip yöntemlerinin kesinlik (solda) ve başarı (sağda) karşılaştırması



Şekil. 3. KCF tabanlı nesne takip yöntemlerinin kesinlik (solda) ve başarı (sağda) karşılaştırması

Çekirdek korelasyon filtresinin farklı görünüm modelleri ile kullanımının kesinlik ve başarıları sırasıyla Şekil 3'te gösterilmektedir. Görüntü gri düzey yoğunluk ve RGB renk histogramı modellerine kıyasla HOG özelliğinin çekirdek korelasyon filtresi ile kullanımı daha yüksek başarıya sahiptir. Görüntü gri düzey yoğunluk, RGB renk histogramı ve HOG için fps değerleri sırasıyla 216, 104 ve 175'tir. Çekirdek korelasyon filtresinin farklı görünüm modelleri ile gerçekleştirilen nesne takibi yöntemleri karşılaştırıldığında OTB-100 veri seti için en başarılı sonucu HCF'ten sonra VGG-19 ağının 4. evrişim katmanından gelen özellikler vermiştir. Son üç katmanın çekirdek korelasyon filtresi ile kullanımı KCF_HOG'dan başarılı iken ilk iki katmanın kullanımı daha başarısızdır.

IV. SONUÇ

Fourier uzayında hızlı işlem kapasitesine sahip çekirdek korelasyon filtresinin derin öğrenme ile birlikte kullanımı nesne başarı oranını arttırmaktadır. Aynı zamanda semantik ve uzaysal özelliklerin bir arada kullanılması durumunda nesne takibinde başarı oranı artmaktadır. Nesne takibi problemi için kullanılan MOSSE, Kalman filtresi, KCF, HCF ve OpenCV takip yöntemleri karşılaştırıldığında OTB-100 veri seti için en başarılı sonucu 0.771 kesinlik ve 0.537 başarı ölçümü ile HCF yönteminin verdiği görülmektedir. Çekirdek korelasyon filtresinin farklı görünüm modelleri ile gerçekleştirilen nesne takibi yöntemleri karşılaştırıldığında OTB-100 veri seti için en başarılı sonucu VGG-19 ağının 4. evrişim katmanından gelen özellik vermiştir. Aynı zamanda derin öğrenme modelinin son

evrişim katmanlarının ilk katmanlara kıyasla daha etkili takip sağladığı değerlendirilmiştir.

KAYNAKLAR

- [1] Comaniciu, D., Meer, P. 2002 Mean Shift: A Robust Approach toward Feature Space Analysis. IEEE Transactions on Pattern Analysis ve Machine Intelligence, 24: 603-619.
- [2] Brdjanin, A., Dardagan, N., Dzgal, D., Akagic, A. 2020. Single Object Trackers in OpenCV: A Benchmark. International Conference on Innovations in Intelligent Systems and Applications pp. 1-6.
- [3] Bolme, D. S., Beveridge, J. R., Draper, B. A., Lui, Y. M. 2010. Visual object tracking using adaptive correlation filters. Computer Vision and Pattern Recognition., pp. 2544-2550.
- [4] Henriques, J.F., Caseiro, R., Martins, P. and Batista, J., 2014. High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3), pp.583-596.
- [5] Welch, G., Bishop, G. 1995. An introduction to the Kalman filter, University of North Carolina, Department of Computer Science, TR 95-041.
- [6] Fan, L., Wang, Z. 2016. A Survey on Multiple Object Tracking Algorithm. IEEE International Conference on Information ve Automation, 1-3 Ağustos 2016, Ningbo, China.
- [7] Zhang, T., C. Xu, C., Yang, M. H. 2018. Learning multi-task correlation particle filters for visual tracking. IEEE Transactions on Pattern Analysis ve Machine Intelligence, 41(2).
- [8] Ma, C., Huang, C. J. B., Yang, X., Yang, M. H. 2015. Hierarchical convolutional features for visual tracking. IEEE International Conference on Computer Vision (ICCV), 7-13 Aralık 2015, Santiago, Chile.
- [9] Wu, Y., Lim, J., Yang, M.H. 2015. Object tracking benchmark. PAMI 37(9), 1834-1848.
- [10] Simonyan, K., Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.