

---

## VİDEOLARDAKİ HAREKETLİ NESNELERİN TESPİT VE TAKİBİ İÇİN UYARLANABİLİR ARKAPLAN ÇIKARIMI YAKLAŞIMI TABANLI BİR SİSTEM

*Bahadır KARASULU\**

---

**Özet:** Gözetleme sistemleri temelinde bilgisayar bilimleri kapsamındaki video ve görüntü işleme araştırma alanları bulunmaktadır. Video işleme, belirli bir video görüntüsünde var olan sahne içerisindeki değişimleri incelemeye kullanılabilecek çeşitli yöntemleri içermektedir. Günümüzde video işleme bilgisayar bilimlerinin en önemli araştırma alanlarından birisidir. İki-boyutlu videolar; çoklu ortam içerik-tabanlı endekslemeye, bilgi elde etmeye, görsel gözetleme ve dağıtık çapraz-kamera ile gözetleme sistemlerinde, insan takibi, trafik izleme ve benzeri uygulamalardaki çeşitli bölütleme, nesne tespit ve takibinde kullanılmaktadır. Arkaplan çıkarımı (AÇ) yaklaşımı, hareketli nesne tespit ve takibi konusunda sıkça kullanılan yöntemlerden biridir. Literatürde bu konu ile ilgili benzer yöntemler de mevcuttur. Yapılan bu araştırma çalışmasında mevcut yöntemlere ek olarak daha etkin bir çözüme gidilmesi önerilmiştir. Uyarlanabilir arkaplan çıkarımı (UyAÇ) yaklaşımı kullanılarak oluşturulan modele göre bilgisayar ortamında nesne tespit ve takip sistemi yazılımı gerçekleştirilmiştir. İlgili video veri setleri ile deneysel çalışma yapılarak geliştirilen sistemin başarımı sınanmıştır. Deneysel sonuçlar ve tartışmaya çalışma içerisinde yer verilmektedir.

**Anahtar Kelimeler:** Arkaplan çıkarımı, nesne tespiti, nesne takibi, video işleme.

### **A System based on Adaptive Background Subtraction Approach for Moving Object Detection and Tracking in Videos**

**Abstract:** Video surveillance systems are based on video and image processing research areas in the scope of computer science. Video processing covers various methods which are used to browse the changes in existing scene for specific video. Nowadays, video processing is one of the important areas of computer science. Two-dimensional videos are used to apply various segmentation and object detection and tracking processes which exists in multimedia content-based indexing, information retrieval, visual and distributed cross-camera surveillance systems, people tracking, traffic tracking and similar applications. Background subtraction (BS) approach is a frequently used method for moving object detection and tracking. In the literature, there exist similar methods for this issue. In this research study, it is proposed to provide a more efficient method which is an addition to existing methods. According to model which is produced by using adaptive background subtraction (ABS), an object detection and tracking system's software is implemented in computer environment. The performance of developed system is tested via experimental works with related video datasets. The experimental results and discussion are given in the study.

**Keywords:** Background subtraction, object detection, object tracking, video processing.

---

\* Çanakkale Onsekiz Mart Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Terzioğlu Kampüsü, 17020, Çanakkale.

İletişim Yazarı: B. Karasulu (bahadirkarasulu@comu.edu.tr)

## 1. GİRİŞ

İnsanların güvenlik istekleri gün geçtikçe artmaktadır. İnsan-bilgisayar etkileşiminin artması ve bilgisayarlı otomasyon sistemlerinin insan hayatına etkisi düşünüldükçe, sağlıklı bir güvenlik sisteminin temel taşlarından birisi haline gelen gözetleme sistemleri daha da önem kazanmaktadır. Bir güvenlik sistemi için en temel bileşen genellikle bir gözcü veya gözetleme sistemi olarak ifade edilebilir. Böyle bir sistem, elektronik olarak (bilgisayar-temelli) bir yapı ile insanın bakmadığı zamanlarda da görüntü işleme teknikleri kullanarak gözetlenilmesi istenilen bölgeyi gözetlemekte ve hareketli nesnelerin tespit ve takibini (T-ve-T) yapabilmektedir. İnsan görüşü ile bilgisayarlı görü arasında temel bazı fonksiyonel benzerlikler bulunur. Bu iki sistemin de temel amacı uzamsal (konuma bağlı) veriyi yorumlayabilmektir. Böyle bir veri birden çok boyut ile endekslenilmekte ve büyük boyutlu depolama isteklerine sahip olabilmektedir (Karasulu, 2010).

Sayısal video işleme, içerik tabanlı video bilgi elde etme (Content-based video retrieval, İTVBE) sistemleri için temel teşkil etmektedir. İTVBE sistemleri, ele alınan videonun içeriğinin belirlenerek incelenebilir, ayrıştırılabilir ve sınıflandırılabilir hale gelmesine ihtiyaç duymaktadırlar. Bu yolla, nesnelerin ve nesnelerin içerisinde bulunduğu çevrenin özellikleri, durumları hakkında bilgi elde edilebilmesi mümkün olmaktadır. Hareketli nesne veya nesneleri tespit ve sonrasında da takip etme, bilgisayarlı görü kapsamında ele alınan büyük bir problemdir. Bilgisayarlı görü problemlerinin yorumlanmasında bir kesinlik bulunmamaktadır. Bu sebeple, videolarda hareketli nesne T-ve-T sürecinin belirli aşamalarının alt problemlere ayrıştırılması daha etkin bir yolla çözüm üretmeyi sağlamaktadır. Bu tarz alt problemler bir arada değerlendirilerek ana problemin çözümü gerçekleştirilebilmektedir. Videolarda nesnelerin sahne içerisinde bölütlenmesi için uzamsal bölütleme (spatial segmentation) ve zamansal bölütleme (temporal segmentation) kullanılmaktadır. Genellikle, uzamsal bölütleme ile nesne tespiti yapılırken, zamansal bölütleme ile de nesne takibi yapılmaktadır. Nesne T-ve-T sürecine, uzamsal ve zamansal bölütleme alt problemlerine sahip daha büyük bir problem olarak bakılabilir (Karasulu, 2010).

Video işleme, belirli bir video görüntüsünde var olan sahne içerisindeki değişimleri incelemede kullanılacak çeşitli yöntemleri içermektedir. Sahne, kendi içerisinde arkaplanda (background) veya önplanda (foreground) var olan çeşitli durağan veya hareketli nesneleri içerebilmektedir. Nesnelere ile sahne içerisindeki arkaplan ve diğer nesnelere arası etkileşimler ise video işleme'nin temel olarak ilgilendiği konulardır (Karasulu, 2010). Bir video dizisi bazı yapısal birimler içermektedir. Hiyerarşik yapıda en küçük birim video çerçevesidir (video frame). Video uygulamalarının, bir analiz işlemine başlamadan evvel, ilk işlem olarak bir videoyu video çekimlerine (video shot) bölümlenmeleri gerekir. Camara-Chavez ve diğ. (2008) çalışmasında, sürekli bir devinimi gösteren bir görüntü dizisi olarak video çekimini tasvir etmişlerdir. Bu devinim, tek bir kameranın tek bir işlemi ile yakalanmış bir devinim olmaktadır. Bu tarz çekimlere, tekli çekim (single shot) denilmektedir. Video çekimleri bir araya gelerek tam bir video dizisini oluşturmaktadır. Genelde, her bir video çekimi belirlenen anahtar çerçeveler (video key frames) tarafından gösterilmektedir. Bunlar, bir videodaki uzamsal ve zamansal özelliklere göre endekslenmektedirler (Karasulu, 2010).

Günümüzde video işleme bilgisayar bilimlerinin en önemli araştırma alanlarından birisi olarak kabul edilmektedir. Literatürdeki bir çok çalışmada nesne T-ve-T için, iki-boyutlu videolar, çoklu ortam içerik-tabanlı endekslemede, bilgi elde etmede, dağıtık çapraz-kamera ile gözetleme ve görsel gözetleme sistemlerinde, insan (yaya) takibi, trafik izleme ve üç-boyutlu televizyon (3DTV) uygulamalarında karşılaşılan çeşitli bölütleme yöntemleri kullanılmaktadır (Remagnino ve diğ., 2002; Foresti ve diğ., 2003; Erdem ve diğ., 2005; Karasulu, 2010). Genel olarak bakıldığında, video nesne takibi, ilgilenilen bir nesne veya nesnelerin bir video sahnesinden bölütleme yoluyla ayrıştırılması süreci olarak ifade edilebilir. Nesne takibi algoritmaları, karmaşık senaryolarla uğraşırken genellikle başarısız olabilmektedirler. Nesne

tanıma algoritmaları ise önceden-tanımlı olarak etiketlenmiş (pre-annotated) veri setleri ile oldukça iyi ve tutarlı sonuçlar verebilmektedir (Brdiczka ve diğ., 2006; Sàncheza ve diğ., 2009; Bennett ve diğ., 2008).

Arkaplan Çıkarımı'nın (Background Subtraction, AÇ) hareketli nesneyi sahne içerisinde konumlandırmadaki hızı, bu yöntemi araştırmacılar açısından oldukça sık kullanılan bir hale getirmektedir. Genel (global) bir eşik değeri (threshold) ile videodaki çerçeveler-arası basit fark AÇ yönteminin temel varsayımlarına karşı bir hassasiyeti de beraberinde getirmektedir. Bu varsayımlara göre, gürültüsüz bir arkaplana sahip sabit (durağan) bir kamera bulunması gerekmektedir. Buna karşın, gerçek hayattaki mevcut video gözetleme sistemlerinde kamera titreşimleri (jitter), ışıklandırma değişimleri vb. gibi çevresel etkiler bulunabilmekte ve AÇ varsayımlarına ters düşebilmektedir (Yılmaz ve diğ., 2006; Benezeth ve diğ., 2008). AÇ yöntemi ve diğer benzer yöntemler için genellikle bir sahne, nesne tespiti sırasında arkaplan modeli ismi verilen bir model ile gösterilebilmektedir. Bu yolla, ilgili yöntem ve buna ait algoritma her bir yeni gelen video çerçevesi ile arkaplan modeline göre arada oluşan farklılıkları tespit etmeye çalışmaktadır. Bu olguya, çerçeve farkları (frame differencing) denilmektedir. Video işlemede, sahnedeki bir değişimi sezme veya nesne hareketini kestirmek gibi işler için farklı anlardaki iki video çerçevesinin farkı ilgili pikseller arasında piksel-piksel (pixel-by-pixel) hesaplanmaktadır (Karasulu, 2010). Çevresel etkilere uyumlu bir sistem geliştirilerek böyle durumlarda nesne T-ve-T probleminde en uygun (optimal) çözümün getirilmesi gerekir. Ele alınan ana problem olan videolarda hareketli nesne/nesneleri T-ve-T için, alt problemler olarak video'nun her bir çerçevesi için görüntü bölütleme (image segmentation) işleminin nasıl yapılması gerektiği konusu bu çalışma kapsamında araştırılmıştır. Araştırma sonucunda, en uygun yöntem olarak eşikleme (thresholding) seçilmiş ve uygun eşik değerinin optimal bir biçimde seçilebilmesinin ortam şartlarında değişimlere uygunluk (uyarlanabilirlik) sağlamak konusunda önemli bir nokta olduğu ortaya çıkmıştır.

Geliştirilen nesne T-ve-T sisteminde de kullanılan AÇ yaklaşımı, literatürdeki diğer çalışmalarda da oldukça sık kullanılmaktadır. Başarılı bir yöntem olarak literatürde kabul görmüştür. Bu yaklaşımın çeşitli çevre şartlarındaki değişikliklere uygun hale getirilebilmesi için uyarlanabilir yapıda olması gerekmektedir. Bu çevre şartlarına etki eden değişiklikler arasında; video görüntülerindeki gürültü, parazitlenme, ışıklandırma koşullarının değişimi ve gölgelenme gibi olgular sayılabilir. Bu çalışma kapsamında, AÇ yönteminin yazılımsal gerçekleştirimi ve AÇ yönteminin Entropi-temelli olarak değişikliğe gidilmiş hali olarak "Uyarlanabilir AÇ" (Adaptive Background Subtraction, UyAÇ) yönteminin geliştirimi ve gerçekleştirimi yapılmıştır.

Bu çalışma altı bölümden oluşmaktadır. AÇ ve UyAÇ yöntemine ait detaylar ikinci bölümde yer almaktadır. Geliştirilen sistemde genel uyarlanabilir bir arkaplan modeline sahip bir nesne tespit altyapısı oluşturularak, bunu temel alan hareketli nesne takip sistemi gerçekleştirilmiştir. Çevre şartlarındaki değişiklikler başarıma doğrudan etki eden faktörlerdir. Bu takip sisteminin başarımına etki eden olgular sistem tarafından otomatik olarak belirlenerek, T-ve-T problemindeki ilgili parametreler olarak ele alınmaktadır. Bu yolla, ilgili parametreleri içeren bir yapı kullanılarak başarımın iyileştirilmesi sağlanmıştır. Başarımın iyileştirilmesi için uyarlanabilir bir altyapı oluşturulmuş ve herhangi bir değişiklik halinde sistemin kendini yeni duruma uyarlayabilmesi sağlanmıştır.

Çalışmanın ikinci bölümünde de bu altyapıya ait detaylara da yer verilmektedir. Bu çalışma kapsamında kamera ile elde edilmiş çeşitli video görüntüleri kullanılmış ve bu görüntüler içerisindeki hareketli nesnelerin uzam-zamansal tutarlılıkları göz önünde bulundurularak, çeşitli başarım ölçümü kriterlerine göre başarım değerlendirilmesi de yapılmıştır. Elde edilen başarım sonuçları çalışmaya ait deneysel sonuçlar olarak karşılaştırılmalı biçimde bu çalışmanın üçüncü bölümünde verilmektedir. Dördüncü bölümdeki tartışmada varılan sonuçlara ve konu hakkındaki yorumlara yer verilmiştir.

## 2. MATERYAL ve YÖNTEM

### 2.1. Geliştirilen Model ve Kullanılan Yöntem

Literatürdeki birçok çalışmada arkaplan'dan hareketli nesnenin ayrıştırılabilmesi için hem uzamsal (spatial) hemde zamansal (temporal) bölütleme (segmentation) kullanılmaktadır. Uzamsal bölütleme, sayısal görüntü bölütleme (digital image segmentation) temellidir. Buna göre bir görüntü, bu görüntüye ait bölgelere iki-seviyeli (siyah-beyaz) veya daha çok seviyeli olacak şekilde, yerel veya genel olarak bölütlenmektedir. Zamansal bölütleme, videoyu sahne, video çekimi (video shot) ve çerçeve olarak belirli hiyerarşik yapı elemanlarından oluşan aralıklara bölmektedir. Belirlenen zamansal bir aralıkta nesnenin rota ve durumu hakkında bilgi elde edilmesi bu yolla sağlanmaktadır (Karasulu, 2010).

Benzer birçok çalışma basit çerçeve farklarını (frame differencing) ve genel bir eşik değerini temel alarak arkaplan çıkarımı işlemini yapmaya çalışmakta ve uyarlanabilirlik yeterince dikkate alınmamaktadır. Bunun sonucunda çevre şartlarından oldukça fazla etkilenen ve sabit bir kameraya ihtiyaç duyan yaklaşımlar ortaya çıkmıştır. Bir video sahnesi genellikle bir arkaplan modeli (background model) ile ifade edilmektedir. Uygulanan yöntem veya algoritma, gelen her yeni video çerçevesinin (video frame) var olan arkaplan modelinden ne kadar farklılaştığını belirlemeye çalışır. Bu farklar üzerinden hareketli nesne, video çerçeveleri boyunca T-ve-T edilebilir. Zaman-yayılımlı (time-evolving) bir arkaplandaki değişimler, piksel tabanlı bir arkaplan modelinin oluşturulup bakımının/güncellenmesinin yapılması ile takip edilebilmektedir. AÇ yaklaşımı için önplan tespiti ve arkaplan modelinin bakımı oldukça önemli konulardır. Yukarıda söz edilen değerlendirmeler göz önüne alınarak, bu çalışmada ilk önce AÇ yöntemine dayanan yaklaşım incelenmiş ve bunun ardından bu yöntemin uyarlanabilir bir hale getirilmesi için nasıl değişikliğe uğratıldığı (modifiye edildiği) hakkında bilgilere yer verilmiştir.

#### 2.1.1. Arkaplan Çıkarımı Yaklaşımı

Genellikle, zaman-yayılımlı bir arkaplanın takip edilmesi için piksel-tabanlı bir arkaplan modeli oluşturulmakta ve sonrasında bunun bakımı (maintenance) ve güncellemesi (update) yapılmaktadır. Arkaplan modeli ile karşılaştırılan bir görüntü bölgesindeki herhangi bir önemli değişim, AÇ yöntemine göre hareketli bir nesne olarak ifade edilebilmektedir. Daha ileri bir işleme için, piksellerin ilgili bölge-içi devam eden değişimleri de işaretlenmektedir. Nesne ile ilgili birbirine bağlı bölgelerin elde edilmesinde bir bağlı bileşen analizi (connected component analysis) ve buna ait algoritma çoğunlukla uygulanmaktadır. Bu süreçte, arkaplan'ın bakımı en temel kısımdır. Zamanla değişen durumlarda, arkaplan'ın bakımı'nın yapılması, AÇ'nın başarımına etki edebilmektedir (Yılmaz ve diğ., 2006). Temel AÇ yöntemleri, genellikle boş bir sahneye ilişkin olan tek bir referans görüntüsünü alarak, bu görüntüyü arkaplan modeline temel olacak şekilde kullanmaktadır (Karasulu, 2010).

Genel bir AÇ algoritması, genellikle dört ana adımdan oluşmaktadır. Bu adımlar, Cheung ve Kamath (2004) çalışmasında; ön-işleme (pre-processing), arkaplan modelleme (background modelling), önplan tespiti (foreground detection) ve veri geçeri (data validation) olarak verilmiştir. Bazı basit görüntü işleme süreçleri, AÇ ön-işleme adımında, ham girdi olan video dizisini daha sonraki adımlarda kullanılmak üzere belirli bir formata dönüştürmektedir. Arkaplan modelleme adımında, yeni video çerçevesi bir arkaplan modelinin hesaplanması ve güncellenmesinde kullanılabilir. Bu model, sahne ister statik veya isterse dinamik olsun tüm arkaplan sahnesinin istatistiksel bir tanımını sağlamaktadır (Karasulu, 2010). Video işleme alanındaki birçok çalışmada araştırmacılar, arkaplandaki çevresel değişimlere karşı oldukça güçlü fakat tüm ilgilenilen hareketli nesnelere de tanımlayabilecek hassasiyette bir arkaplan modeli oluşturmaya çalışmışlardır (Benezeth ve diğ., 2008; Cheung ve Kamath, 2004; Karasulu, 2010). Bununla beraber, arkaplan modelleme teknikleri iki ana kategoriye ayrılabilirler:

Özyineli-olmayan (non-recursive) ve özyineli-olan (recursive). Bahsedilen tekniklere ait teknik detaylar literatürde ilgili çalışmalarda bulunmaktadır (Yılmaz ve diğ., 2006; Benezeth ve diğ., 2008; Cheung ve Kamath, 2004; McFarlane ve Schofield, 1995; Remagnino ve diğ., 1997; Heikkila ve Silven, 1999; Halevy ve Weinshall, 1999; Boulton ve diğ., 1999; Karmann ve Brandt, 1990; Friedman ve Russell, 1997; Power ve Schoonees, 2002; Gao ve diğ., 2000; Fuentes ve Velastin, 2003; Lee ve diğ., 2003; Carmona ve diğ., 2008; Karasulu, 2010).

Önplan tespiti adımı, video çerçevesinde bulunan ve arkaplan modelinde yeterince açıklanmamış durumdaki pikseller (Carmona ve diğ., 2008), ikili aday önplan maskesi (binary candidate foreground mask) olarak tanımlanırlar (Cheung ve Kamath, 2004). Veri geçirme adımı, AÇ yöntemi verilen aday maskeleri inceleyerek, geçerli olan hareketli nesnelere ilişkili olmayan pikselleri elemekte ve son önplan maskesini çıktı olarak vermektedir (Cheung ve Kamath, 2004; Karasulu, 2010).

AÇ yöntemlerinin temel prensipleri aşağıdaki basit formüle göre özetlenebilmektedir (Benezeth ve diğ., 2008):

$$\psi_e(m) = \begin{cases} 1 & \text{eğer } u(V_{m,e}, B_m) > \tau_\lambda \\ 0 & \text{diğer} \end{cases} \quad (1)$$

burada  $\psi_e$ ,  $e$  anındaki hareket etiketi olarak verilmektedir. Burada,  $m(x,y)$  uzamsal konumunun da bir fonksiyonudur. Ayrıca,  $e$  anındaki  $V_{m,e}$  video çerçevesindeki  $m$  pikseli ile arkaplana ait  $m$  pikseli arasındaki mesafe  $u$  ile gösterilmektedir. Bu durumda,  $\tau_\lambda$  terimi ise bir eşik değeridir. Birçok AÇ yöntemi arasındaki ana fark,  $B_m$  arkaplanı'nın nasıl modellendiği ve hangi  $u$  mesafe ölçütünün kullanıldığıdır.

AÇ yönteminin en önemli sınırlılığı durağan bir kameraya ihtiyaç duymasıdır. Kamera hareketleri, genellikle arkaplan modelini bozmakta, yanlış veya kısmi nesne tespitine neden olabilmektedir. Literatürde var olan bazı önplan tespiti yaklaşımlarından en çok kullanılanlardan birisi ilgili arkaplan kestiriminden önemli derecede farklı olan girdi pikselini kontrol etmektedir. Buna göre; ışıklandırma piksel parlaklığı ve buna ait arkaplan kestirimi incelenmektedir. Ayrıca, başka bir popüler önplan tespit şeması normalize edilmiş istatistikleri temel alan bir eşik değerinin uygulanmasına dayanmaktadır (Cheung ve Kamath, 2004).

### 2.1.2. Uyarlanabilir Arkaplan Çıkarımı

Önplan eşik değeri ve istatistiksel önplan eşik değerini kullanan birkaç çalışma literatürde mevcuttur. Bu tarz çalışmalarda, bu değerler birçok önplan tespit şeması tarafından deneysel olarak tanımlanmışlardır. İdeal olarak, eşik değerinin  $(x,y)$  uzamsal konumunun bir fonksiyonu olması gerekir. Örneğin, düşük kontrast (zıtlık) değerine sahip bölgeler için eşik değerinin küçük olması gerekmektedir. Bazen, düşük kontrast'a sahip sahnelerde nesne tespiti için bu durum avantajlı bir durum haline gelmektedir. Fuentes ve Velastin (2003) eşik değeri tanımlamak için, olası bir değişikliği (modifikasyon) de önermişlerdir (Cheung ve Kamath, 2004).

En bilinen yöntemlerden birisi, ortalama alma (averaging) yöntemi'dir. Bu yöntemde göre; arkaplan modelindeki her bir pikselin ortalaması ve standart sapması alınmaktadır. Bu işlem için belirli miktar video çerçevesinin elde bulundurulması, bir anlamda '*dinlenilmesi*' gerekmektedir. Böylece uygun bir istatistiksel model kurulabilecektir. Farklı istatistiksel modeller de, uygulanan T-ve-T yöntemi içerisinde kurulabilmektedir. İstatistiksel model'in iyi bir şekilde kurulması ile AÇ yöntemi nesne T-ve-T işlemini daha iyi yapacak hale getirilebilmektedir. UyAÇ sisteminde, AÇ yönteminin böyle bir istatistik model kurabilme yaklaşımı ele alınarak, kurulan model için belirli video çerçevesi boyunca '*dinleme*' işlemi gerçekleştirilmiştir. Ele

alınan herhangi bir video için, ilk 75 video çerçevesi boyunca, istatistiksel model oluşturacak şekilde sahne dinlemesi yapılmıştır. Video'nun geri kalan kısmındaki video çerçeveleri boyunca, Entropi-temelli bir biçimde otomatik olarak en uygun eşik değerini belirleme yoluyla ilgili parametrelerin ayarlanması ve belirli aralıklarla (her bir saniye başına video çerçevesi'nin belirli bir oranı kadar aralıklarla, örneğin 25 kare/sn veya 30 kare/sn gibi) istatistiksel arkaplan modelinin güncellenmesi yapılmıştır. Bu sayede arkaplan'da her bir saniye başına oluşabilecek herhangi bir çevresel değişim arkaplan modeline de parametrik olarak etki ettirilmiştir. Entropi-temelli nesnel bir değerlendirme ile en uygun (optimal) eşik değerinin belirlenmesi ve bu yolla görüntü bölütleme ile ilgili çeşitli çalışmalar literatürde de mevcuttur.

## 2.2. Entropi Yöntemi İle En Uygun Eşik Değeri Tespiti

Görüntü işleme ve bilgisayarlı görü alanlarında, en uygun (optimal) eşik değerinin belirlenmesi işlemi, sıklıkla otomatik görüntü bölütleme çalışmalarında kullanılmaktadır. Enformasyon teorisinde, Entropi kavramı bir sistemin makro durumunu tanımlamak için gerekli bilgi miktarını belirlemede kullanılmaktadır (Barbieri ve diğ., 2011). Böylece, Entropi kavramı, karmaşıklık kavramı ile ilişkili bir durum arz eder. Eğer bir sistem yüksek bir Entropi değerine sahip ise, bu sistemin durumlarını ifade etmek için daha çok bilgiye ihtiyaç bulunduğunu göstermektedir. Özel uygulamalara bağlı olarak, Entropi farklı biçimlerde de tanımlanabilir (Barbieri ve diğ., 2011). Kuantum mekanik alanında, örneğin Entropi'yi, von Neumann Entropisi (Neumann, 1955) ile ilişkili olarak, karmaşıklık teorisi için ise Kolmogorov Entropisi (Li ve Vitányi, 1997) ile ilişkili olarak değerlendirmek gerekebilmektedir (Barbieri ve diğ., 2011). Rasgele bir değişkenin minimum betimleyici karmaşıklığını belirlemede kullanılan biçimiyle Entropi, enformasyon teorisi kapsamında Shannon Entropisi olarak literatürde tanımlanmıştır (Cover ve diğ., 1991; Barbieri ve diğ., 2011). Kesikli bir rasgele  $p(x)$  dağılımı'nın Shannon Entropisi aşağıdaki basit formül ile tanımlanmaktadır (burada logaritma iki tabanı'nda alınmaktadır) :

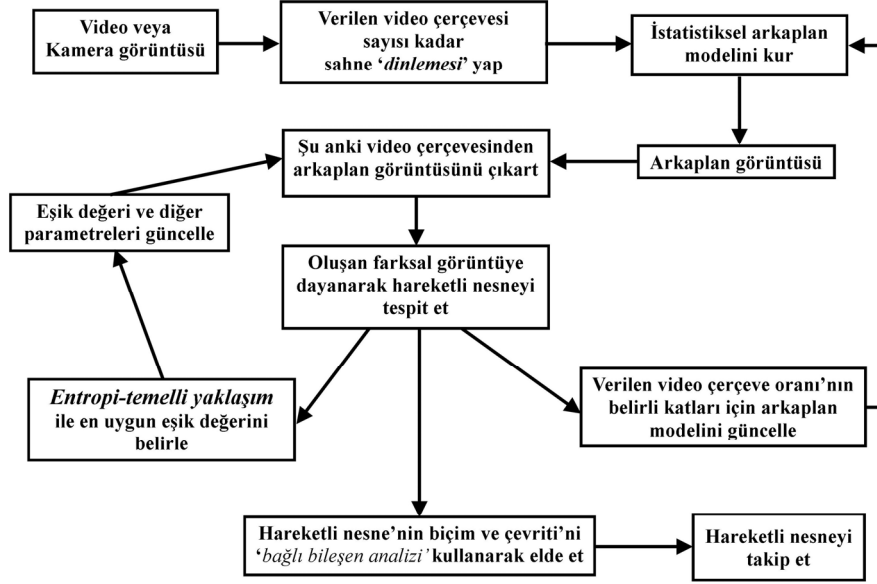
$$H(p) = -\sum_x p(x) \log p(x) \quad (2)$$

Görüntü analizi işleminde,  $p(x)$  bir görüntüdeki farklı renk bileşenleri'nin yoğunluklarını veya gri-seviyeleri'nin (gray-levels) dağılımlarını göstermektedir. Renkli bir görüntü'nün  $p(x)$  histogramı, verilen bir renk yoğunluğu'ndaki piksellerin sayılması ile elde edilir (Kırmızı (Red, R), Yeşil (Green, G) ve Mavi (Blue, B)). Bunlar 0 ilâ 255 arasında değer alabilir. Bu prosedür üç farklı renk histogramı'ndan oluşan bir kümeyi içermektedir. Bu açıdan bakıldığında, Entropi ele alınan sistem için oldukça iyi bir bilgi sağlayabilmektedir. Eğer bir görüntüdeki tüm pikseller aynı gri seviyesine veya renk bileşenlerinin aynı yoğunluğuna sahip ise, bu görüntü minimal bir Entropi değeri ile ifade edilebilir. Bunun haricinde, bir görüntünün her bir pikseli özel bir gri-seviyesi veya renk yoğunluğu ile ifade ediliyorsa, bu görüntü maksimum Entropi gösteriyor demektir (Barbieri ve diğ., 2011). Böylece görüntü içerisindeki nesnelerin dokuları'nın (object texture), Shannon Entropisi ile karakterize olduğu ortaya çıkmaktadır.

### 2.2.1. Uyarlanabilir Arkaplan Çıkarımı için Entropi-temelli Yaklaşım

UyaÇ yöntemi, Entropi-temelli yaklaşımı arkaplan modeli oluşturulması aşaması ve güncellenmesi aşamalarında etkin olarak kullanılmaktadır. Bu çalışma kapsamında geliştirilen UyaÇ yaklaşımı, AÇ yaklaşımı'nın değişikliğe uğratılmış hali (modifiye) olduğu için, AÇ yöntemini de kapsamaktadır. Buna göre; UyaÇ yöntemi sırasıyla bir girdi görüntüsünü (video'dan veya bir kamera'dan akış olarak) alarak bunun üzerinde çeşitli işlemler yapmakta ve hareketli nesne veya nesnelere tespit etmekte, arkaplan'dan ayırmakta ve bunların geliştirilen

sistem tarafından takip edilmesini sağlamaktadır. Şekil 1 'de Entropi-temelli UyAÇ yaklaşımına ait şematik verilmektedir.



**Şekil 1:**  
Entropi-temelli UyAÇ yaklaşımı için şematik.

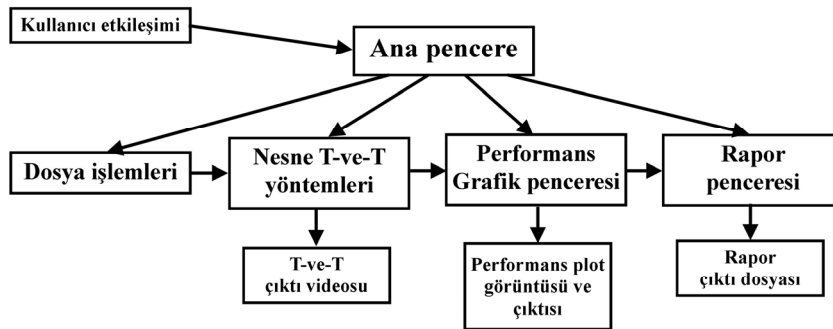
Şematikten de görüleceği gibi, UyAÇ yaklaşımı, verilen video çerçeve sayısı kadar sahne 'dinlemesi' yaptıktan sonra, arkaplan modelini kurmakta ve bu arkaplan modeline göre, şu anki video çerçevesinden arkaplan görüntüsünü çıkartmaktadır (frame subtraction). Daha sonra oluşan farksal görüntüyü temel alarak sahne içerisindeki hareketli nesneyi oluşturan pikselleri işaretlemektedir. Buna ek olarak, Entropi-temelli yaklaşım kullanılarak en uygun eşik değeri tespit edilmekte, bu piksel bölgesi (veya birden çok nesne varsa piksel bölgeleri) ve diğer parametreler güncellenerek, özyineli şekilde her yeni video çerçevesi için bu işlem hareketli nesne veya nesnelerin tam olarak yerinin belirlenmesi ile takip edilebilmesi için tekrarlanmaktadır. Bunun yanı sıra; verilen video çerçeve oranı'nın belirli katları için arkaplan modeli güncellenmekte ve bu katlara denk gelen video çerçevelerinden arkaplan görüntüsü elde edilerek sistem tarafından kullanılmaktadır. Hareketli nesne'nin biçim (shape) ve çevriti (contour), bağlı bileşen analizi yapılarak tam olarak elde edilmektedir. Bunun için geliştirilen sistem, kullanıcı isteğine de bağlı olarak poligon veya dışbükey zarf (convex hull) yaklaşımlarını uygulayabilmektedir. Geliştirilen yazılım sistemi, yukarıda bahsedilen AÇ yöntemini ve UyAÇ yöntemini (Entropi-temelli yaklaşım) destekleyerek, ikisi arasında T-ve-T için nasıl bir benzerlik veya farklılık oluştuğunun belirlenebilmesine de olanak sunmaktadır. Sistem, performans karşılaştırması yapılmasını sağlayan modülü sayesinde, başarımların (performance evaluation/benchmark) yapılabilmesine de olanak sunmaktadır. Geliştirilen yazılım sistemi'nin detayları üçüncü bölümde verilmektedir.

### 3. DENEYSEL SONUÇLAR

#### 3.1. Geliştirilen Yazılım Sistemi

Ele alınan videodaki herhangi bir nesnenin takip edilmesi için sayısal kameraya ve geliştirilecek takip yazılımına gerek duyulmaktadır. Nesne takip yazılımının geliştirilmesi için bir adet ve sınaması için bir adet olmak üzere toplamda iki adet masaüstü bilgisayar ve gerekli donanımlar (sayısal kameralar vs.) çalışma kapsamında kullanılmıştır. Tespit ve takip yazılımı Microsoft firması'nın Visual C++ (Visual Studio 2010) ortamında geliştirilmiş, başarımlarının yapılabilmesi için referans veri (ground-truth, GT) içeren video veri seti olarak CAVIAR video veri seti (nicel kıyaslama için) (CAVIAR, 2012) ve elle işaretlemesi yapılmış olan çalışma kapsamında oluşturulmuş referans video veri seti (nitel kıyaslama için) kullanılmıştır. Nitel değerlendirmeler, verilen video veri setindeki takip edilecek nesnelerin uzman tarafından belirlenmesi ve takip sisteminin çıktıları ile görsel kalite tabanlı olarak birebir karşılaştırılması yoluyla yapılmaktadır. Nicel değerlendirme, sayısal başarımların ölçüm sonuçlarına göre, önceden-ışaretlenmiş referans veri ile takip sisteminin çıktılarının sayısal sonuçlarının karşılaştırılması yoluyla yapılmaktadır. Geliştirilen nesne T-ve-T sisteminin başarımlarını, elde edilen gelen sonuçlar üzerinden bu çalışmada yorumlanmaktadır.

OpenCV yazılım kütüphanesi (Open Computer Vision Library), ilk olarak Intel firması tarafından geliştirilmiştir (Bradski ve Kaehler, 2008). Daha sonraki yıllarda ve günümüzde, Willow Garage isimli bir robotik temalı firma tarafından geliştirilmesine devam edilmektedir (Willow Garage, 2012). Bu kütüphane'nin kullanıcıya yönelik hedeflerinden birisi de bilgisayarda kullanımı kolay bir bilgisayarlı görü (computer vision) ortamı sağlamaktır. Böylece program geliştiricilerinin kolay ve hızlı bir şekilde uygulamaları geliştirilebilmesine yardımcı olmayı hedeflemektedir. OpenCV içerisinde gelen beşyüzden fazla farklı işlem yapan, ilgili görüntü işleme, bilgisayarlı görü, makine öğrenmesi fonksiyonları da bulunmaktadır. Birçok farklı alanda kullanılmaktadır. Bunlar arasında; İnternet'ten veya uydu'dan elde edilmiş haritalar, görüntü tarama, görüntü hizalama, tıbbi görüntüler, gürültü azaltma, nesne analizi, güvenlik ve saldırı tespit sistemleri, otomatik gözetleme (surveillance system), üretim kontrol sistemleri, kamera kalibrasyonu, askeri uygulamalar, insansız hava, toprak ve sualtı araçları, fabrikalarda ürün denetimi ve stereo-görü (stereo-vision) sayılabilir. Çalışma kapsamında geliştirilen sisteme ait şematik Şekil 2 'de görülmektedir.

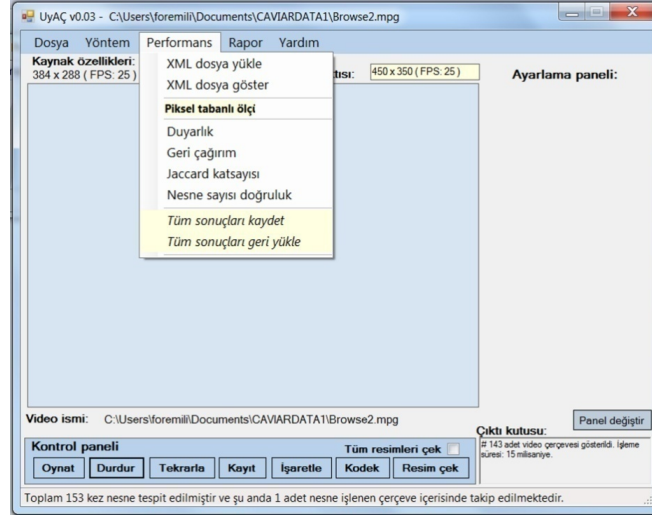


Şekil 2:  
UyaÇ nesne T-ve-T sistemi.

UyaÇ nesne T-ve-T sistemine ait şematikten de görüleceği gibi, geliştirilen yazılımsal sistem, 'Ana pencere' üzerinde yapılan ilgili son kullanıcı etkileşimlerini gerekli yöntemlerin çalıştırılması için kullanarak, video veya kamera akışlarından elde edilen görüntülerde hareketli nesne T-ve-T işlemini gerçekleştirmektedir. Bunun için gerekli video dosyası veya kamera



açma, kapatma, video çıktısı kaydetme gibi işlemleri, 'Nesne T-ve-T yöntemleri'nin çalıştırılmasını, ilgili yönteme ait 'Performans grafikleri'nin çizdirilmesini, sonuçta 'Rapor penceresi'nden ilgili yöntemin performans sonuçlarının son kullanıcıya gösterilmesi (ve çıktı dosyası olarak kaydedilmesini) sağlamaktadır. Aşağıdaki Şekil 3 'te, UyAÇ sistemi'nin ana penceresi üzerinden performans menüsü gösterilmektedir.



Şekil 3:

UyAÇ sistemi'nin ana penceresi.

### 3.2. Geliştirilen Sistemin Başarım Sınaması

Bir görsel sistemin, bilgi sistemleri kapsamında değerlendirilmesinde başlıca iki yaklaşım öne çıkmaktadır. Bunlar, nitel (Kalitatif, *Qualitative*) ve nicel (Kantitatif, *Quantitative*) olarak isimlendirilir. Bu yaklaşımlardan nitel olan, genellikle uzman bir kişinin çoğu zaman nesnel olmayan, daha çok öznel (subjektif) bir yaklaşımı temel alınarak ortaya koyduğu bir kıyaslama sürecidir. Bu tarz öznel yaklaşım, nitelik belirtmesine karşın bazı durumlarda nicelik belirten yaklaşımlardan daha iyi sonuç da verebilmektedir. Bu tarz yaklaşım temelde, nesne T-ve-T sürecinin bir uzman (insan) gözlemci tarafından geçmiş deneyimlerine dayanarak ve karşılaştırma yaparak vardığı kriterleri temel alan bir değerlendirme süreci olarak ifade edilebilir. Bu yaklaşımda uzman'ın belirlediği nesne sayısı, nesne sınır ve çevritleri, rota ve biçim gibi öğeler altın standart olarak kabul edilir ve buna uygun sonuçlar ele alınan (incelemeye tabi tutulan) algoritmik sonuç üzerinden göz ile yapılan kontrol ile kıyaslanır.

Nicel yaklaşım daha genele hitap etmek amacıyla, sayısal değerleri içermektedir. Nicel temelli bir değerlendirme düşünüldüğünde, sistemin toplam başarımı'nın (genel performansı'nın) değerlendirilebilmesi için ilk önce bazı kriterler ve ölçütler belirlenmesi gerekmektedir. Enformasyon sistemlerinin performanslarının ölçümünde ve örüntü tanıma (*pattern recognition*) alanında genellikle kullanılan Duyarlık (*Precision*) ve Geri çağırım (*Recall*) ölçütleri, elde edilen verinin gerçek veri ile ne kadar ilişkili olduğunu ölçmeye çalışır. Nesne sayısı doğruluk (*Object count accuracy*) ölçütü, gerçek sahne içerisinde gerçekten var olan nesne sayısı ile, algoritmik bir ölçüm ile elde edilmiş nesne sayısı arasında tutarlılığın bir ölçüsü olarak ifade edilmektedir. Bunların yanı sıra; Jaccard katsayısı (*Jaccard coefficient*) bir istatistik olarak ele alınır ve örnek alınan veri ile gerçek veri arasındaki çeşitliliği ve benzerliği ölçmeye çalışır. Çalışma kapsamında, nesnel bir değerlendirme yaklaşımı benimsenmiş olup, ölçümler oluşturulan yazılımsal sistem tarafından otomatik olarak alınmaktadır. Bu ölçümler yine aynı yazılımsal sistemin ilgili modülü sayesinde değerlendirilmekte ve sonuçlar son

kullanıcıya sayısal değerler olarak sunulmaktadır. Ayrıca nitel kıyaslama sonuçlarına da bu çalışmada yer verilmektedir. Bu çalışma için yapılan deneylerdeki sınamalar gerçek-zamanlı süreçler halinde yapılmıştır. İlgili nesne T-ve-T yöntemi başarımları gerçek-zamanlı olarak elde edilmekte fakat video-zamanında kaydedilmektedir. Böylece, başarımları ve bunlara ait grafikler video çerçeveleri tabanlı olarak oluşturulmaktadır. Başarımları, tüm seçilen sınamaya videoları üzerinden ilgili nesne T-ve-T yöntemi için ayrı ayrı yapılmıştır.

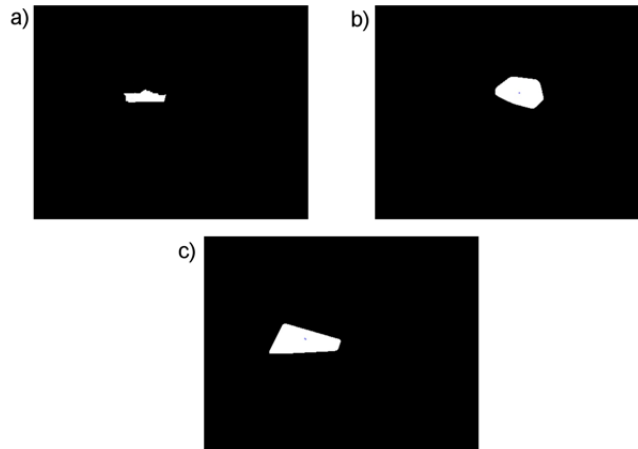
### 3.2.1. Nitel Kıyaslama

Nitel açıdan kıyaslama yapmak için örnek videolar içeren bir video seti'nin el ile oluşturulması gerekmektedir. Bu iş için çalışma kapsamında temin edilen iki adet USB bağlantılı web kamera ve bir adet taşınabilir kamera kullanılmıştır. Farklı ışık ortamları, durum ve sahnelerden elde edilmiş üç adet video deneylerde kullanılmıştır. Bu videoların detayları ve bunlar kullanılarak elde edilen nitel sonuçlar bu bölümde verilmektedir. Çalışma kapsamında kullanılmak üzere kaydedilmiş olan üç adet video, "*prj\_vid1*", "*prj\_vid2*" ve "*prj\_vid3*" isimleriyle ilgili kameralar ile kaydedilmiştir. Bu isimlendirme, çalışma içerisinde deneysel sonuçlarda da kullanılmaktadır. Bu videoların özellikleri aşağıdaki Tablo 1 'de verilmektedir. Tablodaki süreler "dakika:saniye" ve toplam video çerçeve sayısı olarak belirtilmiştir.

**Tablo 1. Nitel kıyaslama için elle oluşturulan video veri seti'ndeki videolar**

Video ismi	Süre	Toplam video çerçeve sayısı	Çözünürlük	Tip
<i>prj_vid1</i>	01:00	1798	480 x 270	MP4
<i>prj_vid2</i>	03:40	6596	480 x 270	MP4
<i>prj_vid3</i>	00:09	275	480 x 270	MP4

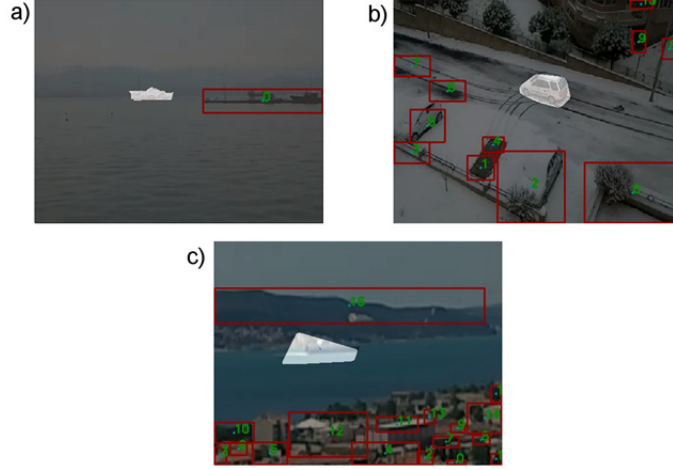
Geliştirilen yazılımsal sistem, OpenCV altyapısı sayesinde, hem MPEG hem de MP4 gibi sıkıştırma algoritması uygulanmış video tipleri ile çalışabilmektedir. Uzman kişinin sahne içerisinde işaret ettiği nesne veya nesnelere, GT nesnelere olarak (ground-truth object, GTO), T-ve-T programının algoritmik olarak tespit ettiği nesne veya nesnelere ise DT nesnelere (detected object, DTO) olarak çalışmamızda verilmektedir. Çeşitli GTO görüntüleri Şekil 4 'te ilgili videolar için siyah-beyaz görüntü (binary image) içerisinde beyaz renkli bölgeler (cisimler) olarak gösterilmektedir.



**Şekil 4:**

*Sınanan video için GTO nesnesi; a) "prj\_vid1" videosunun 766 no'lu çerçevesi, b) "prj\_vid2" videosunun 435 no'lu çerçevesi, c) "prj\_vid3" videosunun 53 no'lu çerçevesi*

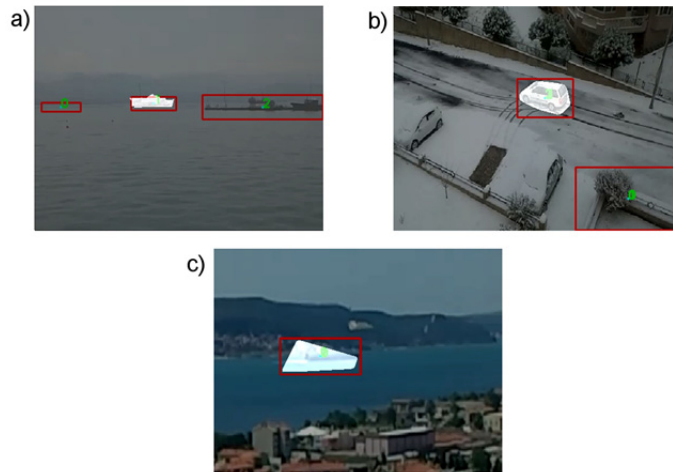
Elle oluşturulmuş videolarda uzman tarafından işaretlemesi yapılmış ilgilenilen nesne veya nesnelere aşağıdaki şekillerde beyaz cisimler olarak siyah zemin üzerinde görülmektedir. Uzmanın belirlediği referans cisim halindeki bu cisimlerin etrafına yeşil bir dikdörtgen konularak referans cisim ile algoritmik olarak elde edilen sonuç karşılaştırılmaktadır. Buna göre, AÇ yöntemi ile Şekil 5 (a) 'da "prj\_vid1" video çerçevesi #766 'dan alınan, Şekil 5 (b) 'de "prj\_vid2" video çerçevesi #435 'den alınan ve Şekil 5 (c) 'de "prj\_vid3" video çerçevesi #53 'den alınan görüntülerden elde edilmiş nitel sonuçlar görülmektedir.



**Şekil 5:**

*Sınanan video için AÇ yöntemi ile nitel sonuç; a) "prj\_vid1" videosunun 766 no'lu çerçevesi, b) "prj\_vid2" videosunun 435 no'lu çerçevesi c) "prj\_vid3" videosunun 53 no'lu çerçevesi,*

Diğer yandan, UyAÇ yöntemi ile Şekil 6 (a) 'da "prj\_vid1" video çerçevesi #766 'dan alınan, Şekil 6 (b) 'de "prj\_vid2" video çerçevesi #435 'den alınan ve Şekil 6 (c) 'de "prj\_vid3" video çerçevesi #53 'den alınan görüntülerden elde edilmiş nitel sonuçlar görülmektedir. UyAÇ yöntemi belirli videolardaki belirli ortam koşullarından (ışıklandırma, sis, kar yansımaları gibi diğer etmenlerden) daha az etkilenmekte ve bunlara göre kendini uyarlayabilmektedir.



**Şekil 6:**

*Sınanan video için UyAÇ yöntemi ile nitel sonuç; a) "prj\_vid1" videosunun 766 no'lu çerçevesi, b) "prj\_vid2" videosunun 435 no'lu çerçevesi, c) "prj\_vid3" videosunun 53 no'lu çerçevesi,*

### 3.2.2. Nicel Kıyaslama

Nicel kıyaslama, uzman kişi tarafından işaret edilen nesnelerin sayısal sonuç veren ölçütler ile kıyaslanması anlamına gelmektedir. Buna göre istenirse piksel bölgelerinin kesişim alanları üzerinden veya istenirse kapsayan kutuların (bounding box) alanlarının kesişimleri üzerinden ölçüm alınabilir. Yukarıda da belirtildiği gibi, uzman kişinin sahne içerisinde işaret ettiği nesne veya nesnelere, GT nesnelere olarak (GTO), T-ve-T programının algoritmik olarak tespit ettiği nesne veya nesnelere ise DT nesnelere (DTO) olarak belirtilmiştir. Tüm nicel ölçümler bu mantığa dayanılarak yapılmıştır. Buna göre bir GT nesnenin kapsayan kutusu ile bir DT nesnesinin kapsayan kutusu ne kadar doğrulukla örtüşüyor ise elde edilen ölçütlerde o kadar iyi sonuç verecektir. Literatürde, ölçütlerin bazıları tespit ve takipteki hataları ölçmede de kullanılabilir. Bu bakış açısıyla nesnel bir değerlendirme yapılabilmesi için ölçme işleminin de nesnel olması gerekir. Başarım ölçütleri olarak; Duyarlık (*Precision*), Geri çağırım (*Recall*), Nesne sayısı doğruluk (*Object count accuracy*) ve Jaccard katsayısı (*Jaccard coefficient*) ölçütleri çalışmamızda kullanılmıştır. Başarım değerlendirme sistemleri, GT verilerine ihtiyaç duymaktadırlar. Buna göre bu sistemler; T-ve-T algoritmasının ürettiği sonuçları verilen değerlendirme yöntemine göre değerlendirilerek sonuçları ölçütler aracılığıyla ortaya koyarlar. Tipik olarak, GT tabanlı ölçütler, doğru pozitifler (True Positive, TP), yanlış pozitifler (False Positive, FP), yanlış negatifler (False Negative, FN) ve doğru negatifler (True Negative, TN) oluştururlar. *Piksel-tabanlı* ölçütler'de TP ve TN tam doğru olarak sınıflandırılmış pikselleri gösterirken; FP ve FN, yanlış olarak önplan (FP) veya arkaplan (FN) olarak işaretlenmiş pikselleri gösterir. Genelde, her bir video çerçevesi için bu değerler hesaplanır ve bir genel değerlendirme ölçütü, tüm video dizisi için ortalama alınarak oluşturulur. *Nesne-tabanlı* ölçütler için TP, GT ile yeterince örtüşen tespit edilmiş nesne sayısını işaret ederken, FP, GT ile yeterince örtüşmeyen tespit edilmiş nesne sayısını işaret eder. FN ise herhangi bir otomatik olarak tespit edilmiş nesne tarafından GT nesnelere ile örtüşmeyen nesne sayısını işaret eder. Aşağıdaki matematiksel ifadelerde,  $N_{xx}$ , (burada  $xx$  ölçüt ismidir) hangi ölçüt'ün kaç adet olduğunu göstermektedir (Lazarevic-McManus ve diğ., 2008; Karasulu, 2010). *Duyarlık* (*Duy*) ve *Geri çağırım* (*GÇ*) ölçütleri için formüller:

$$Duy = P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad , \quad GÇ = R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (3)$$

olarak verilir. Burada,  $N_{TP}$ ,  $N_{FP}$  ve  $N_{FN}$ , sırasıyla doğru pozitifler, yanlış pozitifler ve yanlış negatifler olacak şekilde piksel veya nesne sayısını gösterir. Tek bir video çerçevesindeki referans alınan GT nesnesi kümesine  $GTO(t)$  ve programın algoritmik çıktısı şeklindeki DT nesnesi kümesine ise  $DTO(t)$  diyelim. Buna göre;  $t$  çerçevesindeki Nesne Sayısı Doğruluğu (*NSD*) ölçütü aşağıdaki şekilde ifade edilir:

$$NSD(t) = \begin{cases} \text{tanımsız} & , \text{ eğer } N_{GTO}(t) + N_{DTO}(t) = 0 \\ \frac{\text{Minimum}(N_{GTO}(t), N_{DTO}(t))}{N_{GTO}(t) + N_{DTO}(t)} & , \text{ diğer hallerde} \end{cases} \quad (4)$$

Burada  $t$  video çerçevesindeki  $N_{GTO}$  ve  $N_{DTO}$  sırasıyla GT nesnesi ve DT nesnesi sayısıdır. Bu ölçütte, sadece her bir çerçevedeki nesne tespit kutuları göz önüne alınmaktadır, fakat uzamsal bilgi göz önüne alınmamaktadır. NSD ölçütünün genel başarım sonucu tüm video çerçeveleri üzerinden ortalama  $NSD(t)$  alınarak bulunmaktadır (Kasturi ve diğ., 2009; Karasulu, 2010). Jaccard katsayısı (*Jaccard coefficient*) ölçütü bir benzerlik ölçüsü olarak ifade edilir. Literatürde

farklı versiyonları, *Tanimoto* katsayısı veya *Kappa* indeks, Benzerlik indeksi veya *Dice* katsayısı olarak da geçmektedir (Shan ve diğ., 2002; Dickinson ve diğ., 2009; Rosin ve Ioannidis, 2003; Yeh ve Fu, 2008). Bir referans görüntüden elde edilen ve algoritmik tespit sonucu elde edilen ilgili bölgeler sırasıyla,  $S_1$  ve  $S_2$  olsun. Buna göre;  $J_K$  aşağıdaki gibi ifade edilebilir:

$$J_K(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = \frac{|TP|}{|TP| + |FP| + |FN|} \quad (5)$$

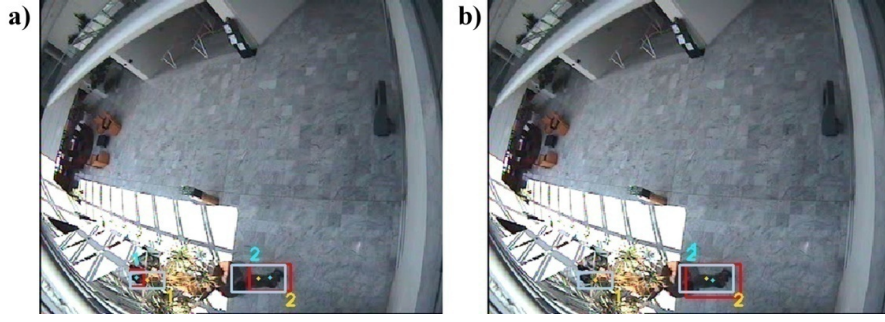
Yukarıda detayları verilen ölçütler çalışmamızda CAVIAR veri setine ait her bir video'nun her bir video çerçevesi için ölçülerek, ilgili videoların ilgili ölçüt için genel ortalamaları alınmıştır. Çalışma sonuçlarını elde ederken, Fransa'daki INRIA laboratuvar'larında üretilen CAVIAR projesine ait CAVIAR kıyaslama video seti (benchmark video set) kullanılmıştır (EC Funded CAVIAR project/IST 2001 37540). Video seti ve detaylarına ilgili web adresinden ulaşılabilir: " <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/> ". Çalışma kapsamında, CAVIAR veri tabanı (CAVIAR, 2010) ait ilk veri setinden (bu veri seti için INRIA giriş salonu) alınan küçük boyutlu videolar üzerinde sınaama işlemleri yapılmıştır. Bu videolar, yarı-çözünürlükteki PAL video standardı'na (384 x 288 piksel, saniye başına 25 video çerçevesi) uygun olarak MPEG-2 video sıkıştırma standardı'nda sıkıştırılmıştır. CAVIAR veri tabanı'nın ilgili websitesinde verilen video dosyaları 6 MB ila 12 MB, bazıları da 21 MB büyüklüğündedir. CAVIAR veri setine ait beş adet video bu çalışmanın deneylerinde kullanılmak üzere seçilmiştir. Bunlar sırasıyla aşağıdaki Tablo 2 'de verilmektedir.

**Tablo 2. Nicel kıyaslama için kullanılan CAVIAR video veri seti'nden seçilen videolar**

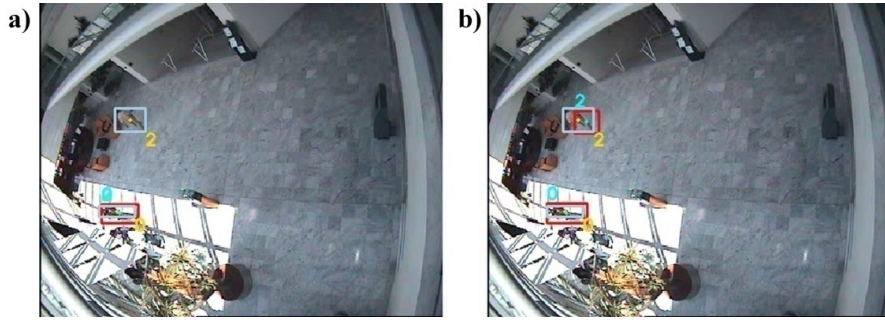
Video ismi	Toplam video çerçeve sayısı	XML GT dosya ismi	Çözünürlük	Tip
Browse_WhileWaiting1	792	bww1gt.xml	384 x 288	MPEG2
Meet_Split_3rdGuy	924	ms3ggt.xml	384 x 288	MPEG2
LeftBag_BehindChair	1068	lbbcgt.xml	384 x 288	MPEG2
Rest_FallOnFloor	1008	rffgt.xml	384 x 288	MPEG2
Walk2	1056	wk2gt.xml	384 x 288	MPEG2

Bu videolar kullanılarak on seferden fazla test yapılmıştır. Test sonuçları detaylı olarak bu bölümde verilmektedir. CAVIAR veri seti'nden seçilen videolar için DTO ve GTO nesnelere sırasıyla, kırmızı ve yeşil dikdörtgenler (kapsayan kutular) ile ilgili şekillerde aşağıda gösterilmektedir. Bu videolarda AÇ ve UyAÇ yöntemleri ayrı ayrı çalıştırılarak sonuçlar elde edilmiştir. Şekil 7 'de "Browse\_WhileWaiting1" videosu'nun #666 no'lu çerçevesi, Şekil 8 'de "Meet\_Split\_3rdGuy" videosu'nun #234 no'lu çerçevesi, Şekil 9 'da "LeftBag\_BehindChair" videosu'nun #530 no'lu çerçevesi, Şekil 10 'da "Rest\_FallOnFloor" videosu'nun #715 no'lu çerçevesi ve Şekil 11 'de "Walk2" videosu'nun #6 no'lu çerçevesine ait algoritmik olarak üretilen sonuçlar verilmektedir. Şekil 7 'den Şekil 11 'e kadar olan şekillerde, *a* şikkı AÇ yöntemi, *b* şikkı ise UyAÇ yöntemi ile elde edilen sonuçları göstermektedir. GTO nesnelere DTO nesnelere ile eşleştirildiğinin son kullanıcı tarafından anlaşılabilmesi için GTO nesnelere yeşil kapsayan kutularla, algoritmik sonuçlar ise kırmızı kapsayan kutular ile gösterilmiştir.

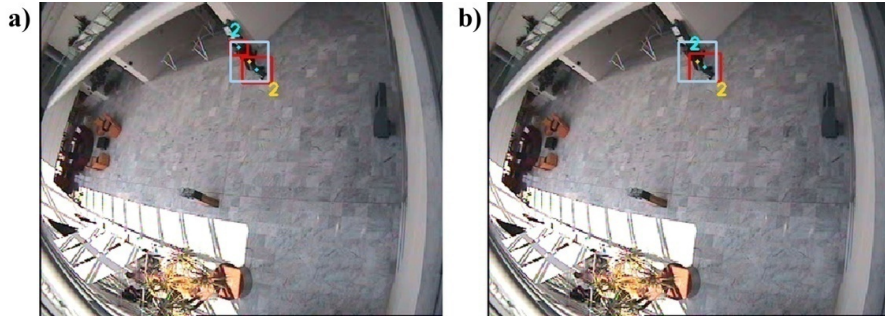




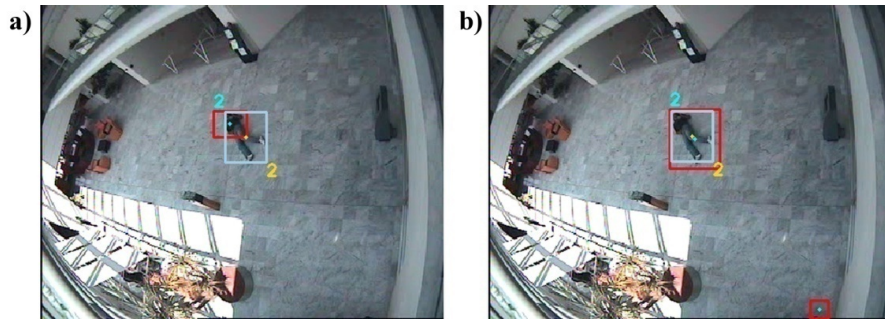
**Şekil 7:**  
*Sınanan video "Browse\_WhileWaiting1" için; a) AÇ yöntemi, b) UyaÇ yöntemi algoritmik sonucu.*



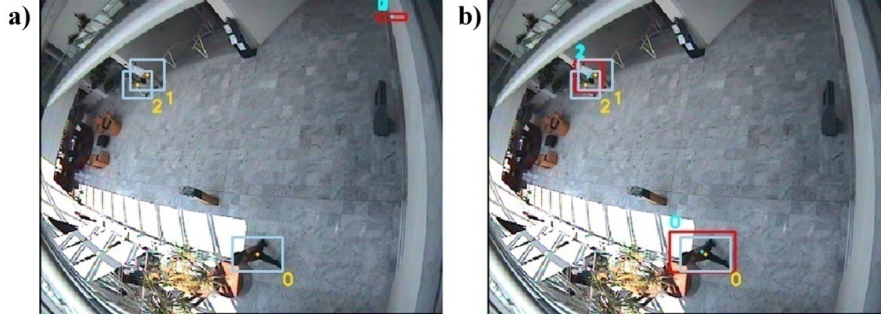
**Şekil 8:**  
*Sınanan video "Meet\_Split\_3rdGuy" için; a) AÇ yöntemi, b) UyaÇ yöntemi algoritmik sonucu.*



**Şekil 9:**  
*Sınanan video "LeftBag\_BehindChair" için; a) AÇ yöntemi, b) UyaÇ yöntemi algoritmik sonucu.*



**Şekil 10:**  
*Sınanan video "Rest\_FallOnFloor" için; a) AÇ yöntemi, b) UyaÇ yöntemi algoritmik sonucu.*



**Şekil 11:**

*Sınanan video "Walk2" için; a) AÇ yöntemi, b) UyAÇ yöntemi algoritmik sonucu.*

CAVIAR video veri seti'nden elde edilen bu algoritmik sonuçlar nicel değerlendirme için dört farklı ölçüt ile her bir video çerçevesi için ölçülüp, toplam video çerçeveleri üzerinden genel ortalama alınmıştır. İlgili değerler, ilgili video için test işlemi AÇ veya UyAÇ yöntemi ile yapıldığında, UyAÇ yazılımı'nın rapor gösterme penceresinden de görülebilmektedir. Buna göre, dört farklı ölçütün, beş farklı video için iki farklı yöntem (AÇ ve UyAÇ) ile elde edilmiş genel performans ortalamaları Tablo 3 ve Tablo 4 'te verilmektedir.

**Tablo 3. AÇ yöntemi ile elde edilen genel performans ortalamaları**

Video ismi	Duyarlık ölçütü	Geri çağırım ölçütü	Jaccard katsayısı ölçütü	Nesne sayısı doğruluk ölçütü
Browse_WhileWaiting1	<b>0,7981</b>	0,5020	0,3976	0,6515
Meet_Split_3rdGuy	<b>0,5996</b>	0,6306	<b>0,4300</b>	0,6546
LeftBag_BehindChair	<b>0,5027</b>	0,5704	0,3665	0,8109
Rest_FallOnFloor	0,3162	0,2528	0,1483	0,7719
Walk2	0,3407	0,1670	0,1702	<b>0,7823</b>

**Tablo 4. UyAÇ yöntemi ile elde edilen genel performans ortalamaları**

Video ismi	Duyarlık ölçütü	Geri çağırım ölçütü	Jaccard katsayısı ölçütü	Nesne sayısı doğruluk ölçütü
Browse_WhileWaiting1	0,7472	<b>0,5602</b>	<b>0,4009</b>	<b>0,7309</b>
Meet_Split_3rdGuy	0,5956	<b>0,6912</b>	0,4202	<b>0,6695</b>
LeftBag_BehindChair	0,4575	<b>0,5916</b>	<b>0,3724</b>	<b>0,8221</b>
Rest_FallOnFloor	<b>0,4710</b>	<b>0,5244</b>	<b>0,3402</b>	<b>0,8091</b>
Walk2	<b>0,5478</b>	<b>0,3221</b>	<b>0,2736</b>	0,3710

Bu ölçütler 1 değerine ne kadar yakın ise o kadar iyi sonuç elde edilmiş olarak yorumlanmaktadır. Tablolarda performans olarak "*daha iyi*" olan sonuçlar, italik ve kalın harflerle gösterilerek belirtilmiştir. Tablo 4 'den 15 adet ölçüm değerinin UyAÇ için AÇ yöntemine göre daha iyi çıktığı, Tablo 3 'den 5 adet ölçüm değerinin AÇ için UyAÇ yöntemine göre daha iyi çıktığı görülmektedir. Böylece, yukarıdaki tablolardan da anlaşıldığı gibi, UyAÇ yöntemi bazı videolarda daha iyi sonuç verirken, bazı videolarda AÇ ile yakın sonuçlar da vermiştir. Fakat UyAÇ yöntemi ile testlerden elde edilen genel sonuçlara bakıldığında, AÇ yöntemine göre üç kat daha fazla başarılı sonuçlar verdiği görülmektedir. Ayrıca çeşitli koşulların etkisi ile UyAÇ yöntemi parametrik olarak daha yeterli sonuçlara ulaşmaktadır.

#### 4. TARTIŞMA ve SONUÇ

Video gözetleme sistemleri (Video surveillance systems), İTVBE sistemlerinin temel aldığı alt sistemler olarak ele alınabilirler. Bu açıdan güvenlik, tanılama (diagnostic) ve benzeri konularda video görüntülerinin uzman kişi veya kişilerce incelenmesi oldukça uzun zaman almakta ve hataya karşı hassas bir yapı oluşturmaktadır. Bu işin yarı-otomatikleştirilmesi veya tam otomatikleştirilmesi en önemli konu olarak ortaya çıkmaktadır. Böyle bir nesne T-ve-T sistemi geliştirilmesi ile video gözetlemesi sonucu elde edilecek video verisi ile nitel ve nicel olarak daha tutarlı takip sonuçları elde edilebilmektedir. Bu sonuçlara göre ve bakılan konu başlığına göre, yapılan bu çalışmanın video ile tanılama yapılan diğer bilimsel alanlarda da (tıp, bilişim vs.) bilimsel gelişime katkı sağlaması düşünülmektedir. Geliştirilen sistemin kullanım alanlarına örnek olarak; bilgisayar bilimlerinde; çoklu ortam sistemlerinde (multimedia systems) ve bilgisayarlı görü (computer vision) alanlarında videodan nesne hareket ve durumlarının analizi ve çıkarılma (reasoning) konularında, tıp biliminde; beyin ve göz cerrahisi başta olmak üzere birçok alanda video ile tanılama konularında çalışan bilim adamları yararlanabilecek ve deneylerinde daha tutarlı sonuçlar elde ederek daha etkin çözümler oluşturabileceklerdir.

#### TEŞEKKÜR

Bu çalışma, Çanakkale Onsekiz Mart Üniversitesi (ÇOMÜ) Bilimsel Araştırma Projeleri Komisyon Başkanlığı tarafından desteklenen 2011/017 nolu "*Videolardaki Hareketli Nesnelerin Tespit ve Takibi İçin Uyarlanabilir Arkaplan Çıkarımı Yaklaşımı Tabanlı Bir Sistemin Geliştirilmesi*" projesi kapsamında elde edilen veriler kullanılarak gerçekleştirilmiştir.

#### KAYNAKLAR

1. Barbieri, A.L., Arruda, G.F. de, Rodrigues, F.A., Bruno, O.M. ve Costa, L. da F. (2011) An entropy-based approach to automatic image segmentation of satellite images, *Physica A: Statistical Mechanics and its Applications*, 390(3), 512-518, ISSN 0378-4371, doi: 10.1016/j.physa.2010.10.015.
2. Benezeth, Y., Jodoin, P.M., Emile, B., Laurent, H. ve Rosenberger, C. (2008) Review and evaluation of commonly-implemented background subtraction algorithms, *In: In 19th Int. Conf. of Pattern Recognition, (ICPR 2008)*, 1-4, doi: 10.1109/ICPR.2008.4760998.
3. Bennett, B., Magee, D.R., Cohn, A.G. ve Hogg, D.C. (2008) Enhanced tracking and recognition of moving objects by reasoning about spatio-temporal continuity, *Image and Vision Computing, Cognitive Vision-Special Issue*, 26(1), 67-81, doi:10.1016/j.imavis.2005.08.012.
4. Boulton, T.E., Michaels, R., Gao, X., Lewis, P., Power, C., Yin, W. ve Erkan, A. (1999) Frame-Rate Omnidirectional Surveillance & Tracking of Camouflaged and Occluded Targets, *In: In Proceedings of the Second IEEE Workshop on Visual Surveillance*, VS. IEEE Computer Society, Washington, DC, USA, 48-55.
5. Bradski, G.R. ve Kaehler, A. (2008) *Learning OpenCV: Computer Vision with the OpenCV Library*, O'Reilly Media, Inc. Publication, 1005 Gravenstein Highway North, Sebastopol, CA 95472, 555 p., ISBN: 978-0-596-51613-0.
6. Brdiczka, O., Yuen, P., Zaidenberg, S., Reignier, P. ve Crowley, J.L. (2006) Automatic acquisition of context models and its application to video surveillance, *In: In 18th Int. Conf. on Pattern Recognition (ICPR'06)*, Hong Kong, 1175-1178.



7. Camara-Chavez, G., Precioso, F., Cord, M., Phillip-Foliguet, S. ve Araujo, A.D.A. (2008) An interactive video content-based retrieval system. *In: 15th Int. Conf. on Systems, Signals and Image Processing (IWSSIP 2008)*, 133-136, doi:10.1109/IWSSIP.2008.4604385.
8. Carmona, E.J., Martínez-Cantos, J. ve Mira, J. (2008) A new video segmentation method of moving objects based on blob-level knowledge, *Pattern Recognition Letters*, 29(3), 272-285, doi:10.1016/j.patrec.2007.10.007.
9. CAVIAR, (2012) "Context aware vision using image-based active recognition", <http://homepages.inf.ed.ac.uk/rbf/CAVIAR>, (Erişim tarihi: 07.Eylül.2012).
10. Cheung, S.-C. ve Kamath, C. (2004) Robust techniques for background subtraction in urban traffic video, *Video Communications and Image Processing, SPIE Electronic Imaging*, San Jose, UCRL-JC-153846-ABS, UCRL-CONF-200706,.
11. Cover, T.M., Thomas, J.A., Wiley, J. ve W. InterScience (1991) Elements of Information Theory, Wiley, New York.
12. Dickinson, P., Hunter, A. ve Appiah, K. (2009) A spatially distributed model for foreground segmentation, *Image and Vision Computing*, 27(9),1326–1335.
13. Erdem, C.E., Ernst, F., Redert, A. ve Hendriks, E. (2005) Temporal stabilization of video object tracking for 3D-TV applications, *Signal Processing: Image Communication*, 20,151-167, doi: 10.1016/j.image.2004.10.005.
14. Foresti, G.L., Regazzoni, C.S. ve Varshney, P.K. (2003) Multisensor Surveillance Systems: the fusion perspective, Kluwer Academic Publishers, Dordrecht, 304 p., ISBN/ISSN 1-4020-7492-1.
15. Friedman, N. ve Russell, S. (1997) Image segmentation in video sequences: A probabilistic approach, *In: Proc. of the Thirteenth Annual Conf. on Uncertainty in Artificial Intelligence (UAI-97)*, Morgan Kaufmann Publishers Inc., San Francisco, California, USA, 175-181.
16. Fuentes, L. ve Velastin, S. (2003) From tracking to advanced surveillance, *In: Proc. of IEEE Int. Conf. on Image Processing (ICIP 2003)*, Barcelona, Catalonia, Spain, 3, 121-124.
17. Gao, X., Boulton, T., Coetzee, F. ve Ramesh, V. (2000) Error analysis of background adaption. *In: Proc. of IEEE Conf. on comp. vision and pattern recognition (CVPR'00)*, Hilton Head Island, South Carolina, USA, 1, 503-510.
18. Halevy, G. ve Weinshall, D. (1999) Motion of disturbances: detection and tracking of multi-body non-rigid motion, *Maching Vision and Applications*, 11(3), 122-137, doi: 10.1007/s001380050096.
19. Heikkila, J. ve Silven, O. (1999) A real-time system for monitoring of cyclists ve pedestrians, *In: Proc. of Second IEEE Workshop on Visual Surveillance*, Fort Collins, Colorado, USA, 74-81.
20. Karasulu, B. (2010) Videolarda Hareketli Nesne Tespiti Ve Takibi İçin Benzetimli Tavlama Tabanlı Bir Başarım Eniyileme Yaklaşımı, (Doktora Tezi), Ege Üniversitesi Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı.
21. Karmann, K.-P. ve Brandt, A. (1990) Moving object recognition using and adaptive background memory, 2, 289-307, *Time-Varying Image Processing and Moving Object Recognition, Cappellini V. (Ed)*, Elsevier Science Publishers B.V., Amsterdam, Netherlands, 346p.
22. Kasturi, R., Goldgof, D., Soundararajan, P., Manohar, V., Garofolo, J., Bowers, R., Boonstra, M., Korzhova, V. ve Zhang, J. (2009) Framework for Performance Evaluation of

- Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), 319-336, doi: 10.1109/TPAMI.2008.57.
23. Lazarevic-McManus, N., Renno, J.R., Makris, D. ve Jones, G.A. (2008) An object-based comparative methodology for motion detection based on the F-Measure, *Computer Vision and Image Understanding, Special Issue on Intelligent Visual Surveillance*, 111(1), 74-85, doi: 10.1016/j.cviu.2007.07.007.
  24. Lee, D-S., Hull, J. ve Erol, B. (2003) A Bayesian framework for gaussian mixture background modeling, *In: Proc. of IEEE International Conf. on Image Processing (ICIP 2003)*, Barcelona, Catalonia, Spain, 3, 973-976.
  25. Li, M. ve Vitányi, P. (1997) An Introduction to Kolmogorov Complexity and Its Applications, Springer.
  26. McFarlane, N. ve Schofield, C. (1995) Segmentation and tracking of piglets in images, *Machine Vision and Applications*, 8(3),187-193, doi: 10.1007/BF01215814.
  27. Neumann, J. von (1995), *Mathematische Grundlagen der Quantenmechanik (Mathematical Foundations of Quantum Mechanics)*. Springer, Berlin.
  28. Power, P.W. ve Schoonees, J.A. (2002) Understanding background mixture models for foreground segmentation, *In: Proc. of Image and Vision Computing*, Auckland, New Zealand, 267-271.
  29. Remagnino, P., Baumberg, A., Grove, T., Hogg, D., Tan, T.N. ve diğ. (1997) An integrated traffic and pedestrian model-based vision system, *In: In Proc. of 8th British Machine Vision Conference*, Essex, UK, 380-389.
  30. Remagnino, P., Jones, G.A., Paragios, N. ve Regazzoni, C.S. (2002) Video-Based Surveillance Systems: Computer Vision and Distributed Processing (Eds.), Kluwer Academic Publishers, Dordrecht, 296 p., ISBN/ISSN 0-7923-7632-3.
  31. Rosin, P.L. ve Ioannidis, E. (2003) Evaluation of global image thresholding for change detection, *Pattern Recognition Letters*, 24(14), 2345-2356.
  32. Sáncheza, A.M., Patricio, M.A., Garcia, J. ve Molina, J.M. (2009) A Context Model and Reasoning System to improve object tracking in complex scenarios, *Expert Systems with Applications*, 36(8), 10995-11005, doi: 10.1016/j.eswa.2009.02.096.
  33. Shan, Z.Y., Yue, G.H. ve Liu, J.Z. (2002) Automated Histogram-Based Brain Segmentation in T1-Weighted Three-Dimensional Magnetic Resonance Head Images, *NeuroImage*, 17 (3), 1587-1598, doi: 10.1006/nimg.2002.1287.
  34. Willow Garage (2012) "Willow Garage Robotic, OpenCV and Robot design", <http://opencv.willowgarage.com/wiki/>, (Erişim tarihi: 07.Eylül.2012).
  35. Yeh, J.-Y. ve Fu, J.C. (2008) A hierarchical genetic algorithm for segmentation of multi-spectral human-brain MRI, *Expert Systems with Applications*, 34(2), 1285-1295, doi: 10.1016/j.eswa.2006.12.012.
  36. Yilmaz, A., Javed, O. ve Shah, M. (2006) Object tracking: A survey, *ACM Comput. Surveys*, 38(4), Article 13, 45p, doi: 10.1145/1177352.1177355.

Makale 31.10.2012 tarihinde alınmış, 19.12.2012 tarihinde düzeltilmiş, 21.12.2012 tarihinde kabul edilmiştir.