

# ÇEKİM BÖLÜTLEME İLE İÇERİK TABANLI VİDEO KOPYA TESPİTİ

## CONTENT BASED VIDEO COPY DETECTION BY SHOT SEGMENTATION

Seda Tankız<sup>1,3</sup>, Tuğrul K. Ateş<sup>2,3</sup>, Ahmet Burak Can<sup>1</sup>

1. Bilgisayar Mühendisliği Bölümü  
Hacettepe Üniversitesi  
abc@hacettepe.edu.tr

2. Elektrik-Elektronik  
Mühendisliği Bölümü  
Orta Doğu Teknik Üniversitesi

3. TÜBİTAK Uzay Teknolojileri  
Araştırma Enstitüsü  
{seda.tankiz,tugrul.ates}@uzay.tubitak.gov.tr

### ÖZETÇE

Gün geçtikçe büyüyen dijital medya kaynaklarını kontrol etmek, gereksiz verilerin atılmasını sağlamak ve video telif haklarını korumak için kullanılan yöntemlerden birisi gizli damga yöntemidir. Bu bildiride, gizli damga yöntemine alternatif bir İçerik Tabanlı Video Kopya Tespit yöntemi sunulmuştur. Etkin video kopya tespit sistemleri, hem geometrik dönüşümlere hem de zamansal saldırılara karşı dayanıklı olmalıdır. Bu çalışmada, geometrik dönüşüm saldırılarına karşı ölçek bağımsız öznelik dönüşümü (SIFT) algoritmasının yerel tanımlayıcıları kullanılan ve zamansal saldırılara karşı videoların bölünmüş çekimleri üzerinden kopya araması yapılan bir video kopya tespit yöntemi önerilmektedir.

### ABSTRACT

Invisible watermarking is one of the possible approaches to control digital media resources which are growing day by day, eliminate redundant data, and maintain copyright of original videos. This paper presents a Content Based Video Copy Detection method alternative to invisible watermarking approach. Effective video copy detection systems should be robust against spatial and temporal transformations. In this study, a video copy detection method, which uses local descriptors of scale invariant feature transform (SIFT) algorithm against geometric transformation attacks and searches copies over segmented shots of videos against temporal attacks, is presented.

### 1. GİRİŞ

Son zamanlarda, iletişim teknolojilerinde kaydedilen gelişmeler nedeniyle sayısal video içeriklerine ulaşmak kolaylaşmıştır. Youtube gibi video servis sağlayıcılarının da etkisiyle dijital medya kaynakları büyümüştür. Bu durum sayısal kütüphanelere eklenen videoların kontrol altına alınması, analiz edilmesi, kopyalarının tespit edilmesi tekrarlarının elenerek gereksiz verilerin azaltılmasını zorunlu hâle getirmiştir. Bir video veritabanının bu özellikleri desteklemesi etkin bir arama işleminin gerçekleştirilmesi açısından da önemlidir.

Video kopyalarının tespiti, söz konusu videoların içerik ve telif haklarının korunması için de gerekmektedir. Video kopya tespiti için genellikle iki farklı yaklaşım kullanılmaktadır: sayısal video gizli damgası ve içerik tabanlı kopya tespiti. İlk yaklaşımda özgün videolara değiştirilemez parmak izi bilgisi eklenir ve kopyalanan videolarda bu bilgi aranır. Gizli damga yöntemi görsel dönüşümlere karşı hassas [1] olmakla birlikte kopyaları gezinen videolara gizli damga ekleyip kopyalarının

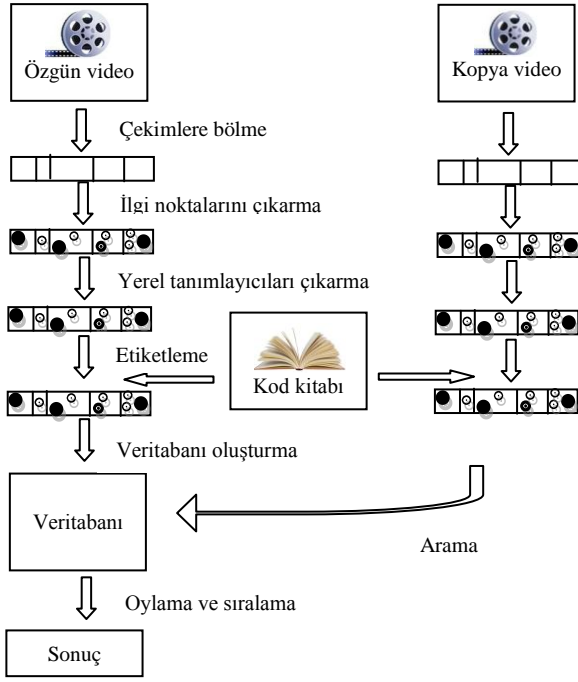
tespit edilmesi sağlanamaz. İçerik tabanlı kopya tespiti videolarda ek bilgi içerme gereksinimi duymadan, video içeriklerinden parmak izi çıkarma ve bu parmak izlerini karşılaştırma esasına dayanır. İçerik tabanlı kopya tespitindeki ana fikir, videonun kendisinin bir damga olduğudur ki kopya tespiti için yeterli benzersiz bilgiyi içerdiği düşünülür. İçerik tabanlı kopya tespitinin gizli damga yöntemi üzerindeki en önemli üstünlüğü özgün video piyasaya sürüldükten sonra da parmak izinin çıkartılabiliyor olmasıdır.

İçerik tabanlı kopya tespitinde genellikle parmak izi olarak adlandırılan bir takım özellikler çıkartılmakta ve veri tabanındaki parmak izleriyle karşılaştırma yapılmaktadır. Bu konuda yapılan çalışmalar global ve yerel parmak izlerini kullanmak üzere iki kısımda incelenebilir. Renk, parlaklık gibi global özellikler hızlı ve etkin çıkartılmakta [2,3,4] ve videoların aslı gibi kopyalarını bulmakta iyi çalışmaktadırlar. Fakat videoya logo eklenmesi, videonun yeniden ölçeklendirilmesi ya da kesilmesi durumlarında başarılı değillerdir [5]. Videonun özgünlüğüne yapılan söz konusu bu saldırılara karşı dayanıklı SIFT gibi yerel tanımlayıcılar kullanılmıştır [5,6,7,8,9]. Yerel tanımlayıcıların parmak izlerini çıkarmaları tüm kare üzerinde çalışan yöntemlere göre zaman almasına rağmen, bu işlem çevrim dışı gerçekleştirildiğinden sorun oluşturmamakta ve tüm karede çalışan yöntemlere göre daha başarılı olmaktadır. İşlem zamanını azaltmak amacıyla etkin dizinleme tekniklerinin kullanılabilmesi de yerel tanımlayıcı kullanımını olanaklı kılmaktadır [9,10].

Bu çalışmada, Ateş ve diğerlerinin [6] yaptığı çalışma temel alınarak, video içinde sadece çekim denen önemli kesimler için yerel tanımlayıcı bilgileri üzerinde kopya tespiti yapılmaya çalışılmıştır. Zamansal dönüşümlerine karşı dayanıklı içerik tabanlı kopya tespiti için, videoların çekimlere ayrılması ve her bir çekimde anahtar karelerin üzerinden ilgi noktalarının çıkartılması sağlanmaktadır. İlgi noktalarından çıkarılan tanımlayıcılar, bir kod kitabına göre etiketlenmektedir. Veritabanı, bu etiketler üzerine dizinlenen referans ilgi noktalarından oluşur. Her bir videonun bütününden değil de sadece belli kesimlerinden tanımlayıcılar çıkarıldığı için, veritabanında saklanması ve karşılaştırılması gereken etiket sayısı azalmaktadır. Bu çalışmada geliştirilen yöntemin testleri, TRECVID 2009 içerik tabanlı kopya tespiti test verileri [11] kullanılarak yapılmıştır.

### 2. ÖNERİLEN YÖNTEM

Bu bildiri [6]'da belirtilen çalışmayı temel almıştır. Fakat temel alınan bu çalışma basit ve hızlı işlem olanağı sağlayan sabit uzunluklu kayan pencereler kullandığı için hız değişimi ya da kesme ekleme/silme gibi zamansal dönüşümlere karşı



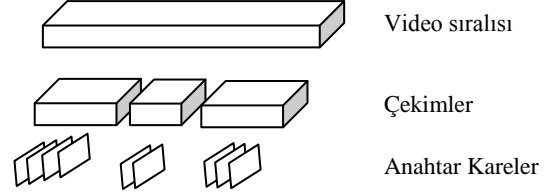
Şekil 1: Önerilen sistemin akış şeması.

dayanıklı değildir [9]. Bu problemi çözmek amacıyla bu bildiri çekim tabanlı video kopya tanıma sistemini sunmaktadır.

Önerilen sistem Şekil 1’de özetlenmiştir. Veri tabanında saklanacak referans videoları öncelikle çekim adı verilen daha küçük parçalara ayrıştırılır. Her bir çekimden, ilgili çekimi temsil edecek 1 ve 0,5 saniye aralıklarla anahtar kareler seçilir ve bu karelerden ilgi noktaları çıkartılır. Bu ilgi noktalarından, özgün videoya yapılan resim içinde resim, kırpma, kaydırma, logo ekleme, oran değiştirme gibi saldırılara karşı dayanıklı diğer yerel tanımlayıcılara göre en iyi performansa sahip [12,6] ölçek-bağımsız öznelik dönüşümü (SIFT) tanımlayıcısı kullanılarak yerel tanımlayıcılar elde edilir. SIFT yerel tanımlayıcılar, yüksek boyutlu öznelik kümesi içerdiğinden bir kod kitabı oluşturulmuştur. Kod kitabı  $k$ -ortalama ( $k$ -means) algoritması kullanılarak elde edilmiştir. Her bir yerel tanımlayıcıya kod kitabı üzerinden bir etiket atanır. Referans video veri tabanı, sorgulamayı hızlandırmak amaçlı bu etiketlerle dizinlenmiş ilgi noktalarını içinde barındıracak şekilde tasarlanmıştır. Sorgu videoları da aynı işlem süzgecinden geçtikten sonra sorgu videolarındaki etiketlere karşılık gelen ilgi noktaları veri tabanından çekilir ve oylama işlemi gerçekleştirilir [6]. Oylama sonuçları sıralanarak sonuç elde edilir.

## 2.1. Video Bölütleme

Bir video sıralısı birçok anlamlı sahneden oluşurken, her bir sahne ise birkaç tane çekime bölünebilir. Bir çekim, bir kamera tarafından çekilen, zaman içerisinde sürekli hareketi belirten ardışık kareler olarak tanımlanabilir [13]. Şekil 2’de gösterildiği gibi bir video sıralısının çekimlere bölünmesi ve bu çekimleri ifade edecek anahtar karelerin alınması, genelleme yaparak bir video sıralısını daha iyi anlatırken çoğu zaman kare bazında yapılan işlemlere göre daha doğru, tutarlı



Şekil 2: Video bölütleme.

öznelik çıkarılmasını ve veritabanında daha az verinin saklanmasını sağlamaktadır.

Çekim geçişleri ani ve kademeli olmak üzere iki şekildedir. Ani geçişleri bulmak kolayken kademeli geçişleri bulmak videoda devinim ile karıştığından zordur. Çekim geçişlerini bulup çekim sınırlarını çizmede Song [14] ve Petersohn’ın [15] çalışmalarındaki teknikler kullanılmıştır. İlk olarak *ayrık kosinüs dönüşümü* (DCT) uzayında ayrıt imgeler çıkartılmıştır. Ortalama imgelerden ve ayrıt imgelerden öznelikler çıkartılıp çekim değişimlerini bulabilmek için iki aşamalı kümeleme algoritması kullanılmıştır. İlk aşamada komşu iki ortalama imgelerdeki öznelik histogramları ve piksellerin farkları üzerinde iki-ortalama kümeleme ile ani geçişler bulunur. İkinci aşamada ise kademeli geçişler aşama aşama bulunmaktadır. Önce zincirleme geçişler bulunur ve bu arada devinimle karıştırılmaları önlenir. Sonra yavaş kararar/aydınlanma çekim sınırları tespit edilir. Son olarak da süpürerek geçişler bulunur. Çekim sınırları bulunduğundan sonra her bir çekimi ifade edecek 1 ve 0,5 saniye aralıklarla anahtar kareler çıkartılacak şekilde iki farklı deney yapılmıştır.

## 2.2. Benzerlik Ölçütü

Anahtar karelerin herhangi birinden çıkartılan ilgi noktalarından elde edilen tanımlayıcılara atanan etiketler  $c$  ile gösterilsin. Bir video parçası  $t$  zamanı,  $c$  etiketleri ve  $m$  ek bilgisiyyle ifade edilebilir. Referans ve sorgu video çekimleri arasındaki benzerlik karşılıklı anahtar karelerde aynı etikete sahip ilgi noktalarının ortalama benzerlikleridir. Benzerlik ölçütü Eşitlik (1)’de ifade edilmiştir.  $p(a,b)$  fonksiyonu ilgi noktaları arasındaki benzerliği ifade ederken,  $k(a)$  fonksiyonu ilgili karedeki ilgi nokta sayısını belirtmektedir.

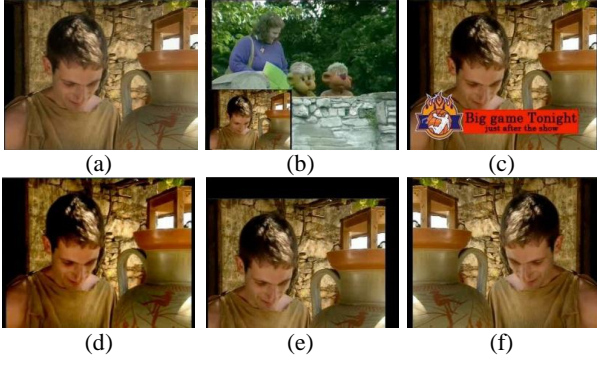
$$s = \frac{1}{\text{süre}} \sum_i \sum_j \delta(ti, tj) \delta(ci, cj) p(mi, mj) \quad (1)$$

$$\delta(a, b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases} \quad (2)$$

$$p(a, b) = \cos(\gamma_{\text{ön}_a} - \gamma_{\text{ön}_b}) \frac{\min(k(a), k(b))}{\max(k(a), k(b))} \quad (3)$$

Video parçası benzerliği karşılıklı karelerde aynı etikete sahip ilgi noktalarının benzerliklerinin ortalamasıdır. Eşitlik (1)’de  $p(a, b)$  işlevi, iki ilgi noktası arasındaki benzerliği ifade eder ve sistem gereksinimlerine göre tasarlanabilir. Bu çalışmada SIFT ilgi noktası çıkarıcısından elde edilen açı bilgisini ( $\gamma_{\text{ön}}$ ) ve kare bilgi karmaşıklığını ele alan (3) ölçütü esas alınmıştır. Kare karmaşıklığı, ilgi noktasının bulunduğu karedeki toplam ilgi noktası sayısı ( $k$ ) ile temsil edilebilir [6].

Sorgu videosundan elde edilen tüm etiketler, etiket numarası ile dizinlenmiş veritabanı içerisinde aranır ve eşleşen özgün video kısımları  $s$  puanı ile oylanır. Hem özgün hem de sorgu videolar aynı çekimleri içereceğinden, oylama aşamasında eşlenen anahtar kareler aynı olacaktır. Bu yöntem yakın kareler üzerinden benzerlik oylaması yapan [6]’ya göre daha dayanıklıdır.



Şekil 3: Özgün video karesi ve olası saldırıları: (a) Özgün kare, (b) resim içinde resim, (c) görüntü üzerine şekil, (d) karşıtlık değiştirme, (e) kaydırma, (f) ayna çevrimi.

### 3. DENEYLER

Bu bildiriye [6]'da olduğu gibi TRECVID 2009 içerik tabanlı kopya tespitinde kullanılmak üzere 400 saatlik özgün video ile süreleri 3 saniye ile 3 dakika arasında değişen sorgu videoları kullanılmıştır. Şekil 3'de özgün videolara gerçekleştirilen bazı dönüşümler [16] gösterilmektedir. Zamansal dönüşümlere daha dayanıklı olması amacıyla video, çekimlere bölünmüş ve her bir çekimi temsil edecek 1 ve 0,5 saniye aralıklarla anahtar kareler alınmak üzere iki farklı deney yapılmıştır. Özgün videolara eklenen boş kareler kontrol edilerek çekim sınırlarının yanlış bulunması engellenmiştir. Ayrıca ayna çevrimi dönüşümüne karşı her bir karenin ayna çevrim ilgi noktaları da çıkartılmıştır. İlgi noktaları 1024 kelime barındıran kod kitabıyla etiketlenmiştir. Başarı ölçütü olarak TRECVID içerik tabanlı video sezim değerlendirme metriği olan NDCR [11] kullanılmıştır. NDCR metriği sıfır ile bir aralığında bir sonuç dönmektedir ve sonucun sıfır olması en iyi performansı gösterirken, bir değeri sonuç dönülme durumu performansını verir. Dolayısıyla NDCR değerinin düşük olması performansın yüksek olduğu anlamına gelir.

Sınanan saldırı tipleri ve NDCR başarı değerleri Şekil 4'te gösterilmiştir. *Yapım sonrası çevrimler* saldırısı, altyazı ekleme, kırpma, ayna çevrimi, kaydırma, şekil ekleme gibi dönüşümleri içerir. *Tümünün birleşimi* saldırısında ise yapım sonrası çevrimler saldırısına ek olarak kare atma, karşıtlık değiştirme, sıkıştırma, bulanıklaştırma, görüntü ekleme, en-boy oranı değiştirme gibi saldırılar bulunmaktadır.

### 4. SONUÇLAR

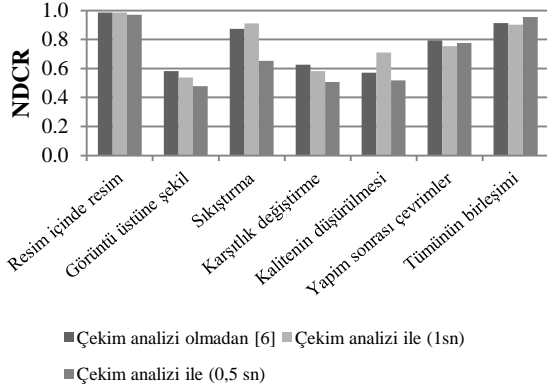
Bu bildiriye, video kopya tespiti sorununa, video çekimlerini kullanarak, anlamlı bölütleri yerel tanımlayıcılarla ifade eden içerik tabanlı bir çözüm önerilmiştir. Videoyu çekimlere bölütlemeye, birçok saldırıya karşı performansı artırmış ve sorgu sonuçlarının iyileşmesini sağlamıştır. Videonun çekimlere bölütlenmesi ve 1 saniyelik anahtar karelerin alınması, görüntü üstüne şekil, karşıtlık değiştirme, yapım sonrası çevrimler ve tümünün birleşimi saldırılarında daha iyi performansla sahip iken, sıkıştırma (çözünürlük ve bit oranında yapılan değişiklik) ve kalitenin düşürülmesi saldırılarında daha düşük performans göstermiştir. Bu durum, çekim sınırlarını

belirlemede kullanılan algoritmaların sıkıştırma, kalitenin düşürülmesi saldırılarına karşı yeterince dayanıklı olmadığını göstermektedir. Bu saldırılarda, çekim sınırları doğru şekilde belirlenemediği için kopya tespitindeki başarı düşmektedir. Buna karşın video bölütlerinden 0,5 saniyede bir anahtar kare çıkartılan deneyin performansı, hem [6]'daki sonuçlara göre hem de 1 saniyede bir anahtar kare çıkartılan deneye göre daha iyidir. Sadece tüm saldırıların birleşiminde kötü sonuç alınmaktadır. Bu durum da, çekim sınırlarını belirlemede kullanılan algoritmanın hatasından kaynaklanmaktadır.

Bu çalışmada, videonun tümünden etiket çıkarılması yerine, sadece çekimler taban alınarak kopya tespitinin mümkün olduğu sonucuna varılmıştır. Büyük video veritabanları üzerinde yerel tanımlayıcılarla gerçekleştirilen kopya tespit sistemleri, genelde tanımlayıcılarla ilgili çok miktarda verinin saklanması gerektirmektedir. Çekimler kullanılarak kopya tespit veritabanında daha az verinin saklanması mümkün olacaktır. Bu çalışma ilerideki çalışmalara alt yapı oluştururken, sonraki çalışmalarda sorgu karşılaştırma performansı arttırmak için üç boyutlu yerel tanımlayıcılar üzerinde yoğunlaşılacaktır.

### 5. KAYNAKÇA

- [1] Onur Kucuktunc, Muhammet Baştan, Uğur Gudukbay, Ozgur Ulusoy, "Video copy detection using multiple visual cues and MPEG-7 descriptors", *JVCIR*, vol.21, 2010, p 838.
- [2] C.Kim, B.Vasudey, "Spatiotemporal sequence matching for efficient video copy detection", *IEE Trans.Circuits Syst. Video Technology*, Vol. 15, p127-132, 2005.
- [3] Zhihua Xu, Hefei Ling, Fuhao Zou, Zhengding Lu, Ping Li, and Tianjiang Wang "Fast and robust video copy detection scheme using full DCT coefficients", *IEEE Int. Conf. on Multimedia & Expo*, 434-437, 2009.
- [4] H. Shen, X. Zhou, Z. Huang, J. Shao, and X. Zhou. "UQLIPS: Areal-time near-duplicate video clip detection system", *Proc. VLDB*, 1374-1377, 2007.
- [5] G. Willems, T. Tuytelaars, L. V. Gool, "Spatio-Temporal Features for Robust Content-Based Video Copy Detection", *Proc. ACM ICMR*, 283-290, 2008.
- [6] T. K. Ateş, E. Esen, A. Saracoğlu, M. Soysal, Y. Turgut, O. Oktay, A. A. Alatan, "Yerel Tanımlayıcılar ile İçerik Tabanlı Video Kopya Sezimi", *SIU*, 2010.
- [7] Apostol Natsev, Matthew Hill, John R. Smith, "Design and evaluation of an effective and efficient video copy detection system", *IEEE Int. Conf. on Multimedia & Expo*, 1353-1358, 2010.
- [8] Huamin Ren, Shouxun Lin, Dongming Zhang, Sheng Tang, Ke Gao, "Visual words based spatiotemporal sequence matching in video copy detection", *IEEE Int. Conf.on Multimedia & Expo*,1382-1385, 2009.
- [9] Chih-Yi Chiu, Cheng-Chih Yang, Chu-Song Chen, "Efficient and Effective Video Copy Detection Based on Spatiotemporal Analysis", *IEEE ISM*, 202-209, 2007.
- [10] Zhijie Zhang, Chongxiao Cao, Ruijie Zhang and Jianhua Zou, "Video Copy Detection Based on Speeded Up Robust Features and Locality Sensitive Hashing", *IEEE ICAL*, 13-18, 2010.



Şekil 4: NDCR metriği ile elde edilen deney sonuçları.

- [11] TRECVID 2009 Değerlendirme Kılavuzu  
<http://www-nlpir.nist.gov/projects/tv2009/tv2009.html>
- [12] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. PAMI*, Vol. 27, No. 10, pp. 1615-1630, 2005.
- [13] G. Boccignone, A. Chianese, V. Moscato, and A. Picariello, "Foveated shot detection for video segmentation", *IEEE Trans. CSVT*, 15(3):365-377, 2005.
- [14] Byung Cheol Song, Jong Beom Ra, "Automatic Shot Change Detection Algorithm Using Multi-stage Clustering for MPEG-Compressed Videos", *JVCIR*, 2001, 364-385.
- [15] Christian Petersohn, "Fraunhofer Hhi At Trecvid 2004: Shot Boundary Detection System", *TRECVID 2004, Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, Einsteinufer 37, 10587 Berlin, Germany*.
- [16] TRECVID CBCD dönüşüm resimleri  
<http://www-nlpir.nist.gov/projects/tv2009/tv2009.html>