Yerel Tanımlayıcılar ile İçerik Tabanlı Video Kopya Sezimi Content Based Video Copy Detection with Local Descriptors

Tuğrul K. Ateş^{1,2}, Ersin Esen^{1,2}, Ahmet Saracoğlu^{1,2}, Medeni Soysal^{1,2}, Yakup Turgut³, Ozan Oktay²
A. Aydın Alatan²

- 1. TÜBİTAK Uzay Teknolojileri Araştırma Enstitüsü
 - 2. Elektrik ve Elektronik Mühendisliği Bölümü Orta Doğu Teknik Üniversitesi
 - 3. Bilgisayar Mühendisliği Bölümü Orta Doğu Teknik Üniversitesi

Özetçe

Kopya video verilerinin tespit edilebilmesi için gelişmekte olan alanlardan biri olan İçerik Tabanlı Video Bulma, görünmez damgalamaya alternatif bir yaklaşımdır. İçerikten çıkarılan, saldırılara karşı gürbüz fakat ayrıştırıcı parmak izleri ile kopya tespitine dayanır. Bu çalışmada iki boyutlu yerel imge tanımlayıcıları ile büyük veritabanlarında kullanılacak bir video kopya sezim yöntemi önerilmektedir. Yöntem, video hacimleri içerisinde bulunmuş yerel tanımlayıcıların kümelenerek bağlı bulundukları kümeler üzerinden karşılaştırılmasına dayanmaktadır. Önerilen yöntem, ölçekbağımsız öznitelik dönüşümü (SIFT) tanımlayıcısı ile TRECVID 2009 İçerik Tabanlı Video Kopya Sezimi görevinde sınanmıştır. Bu bildiride, önerilen yöntem açıklanmakta ve farklı tanımlayıcılar ile küme boyutlarının, bulma başarımına etkileri deney sonuçları üzerinden incelenmektedir. Ayrıca, vöntemin üstünlükleri ve sakıncaları tartısılmakta, gelecekteki çalışmalar için öneriler sunulmaktadır.

Abstract

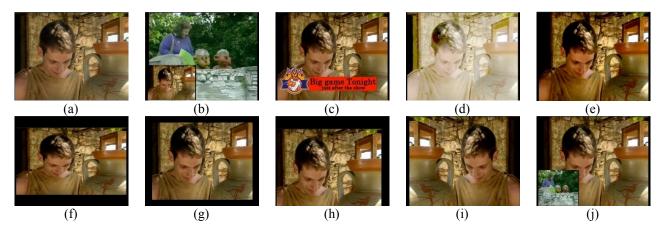
Content Based Copy Detection, one of the emerging areas on detection of video copies is an alternative approach to invisible watermarking. It depends on detection of copies with fingerprints extracted from the content, which are robust against attacks, yet discriminative. In this work, a video copy detection scheme, which will be used on large databases with two dimensional local image descriptors, is proposed. The method depends on clustering local descriptors and comparing them over their clusters. Proposed method has been tested with scale-invariant feature transform (SIFT) in TRECVID 2009 Content Based Copy Detection task. In this paper, the proposed method is explained and effects of alternative descriptors and cluster sizes on detection performance are investigated over experimental results. Furthermore, advantages and disadvantages of the method are discussed and recommendations for future works are made.

1. Giriş

Çoğulortam ve iletişim teknolojilerindeki hızlı gelişimin sonuçlarından biri, internet üzerinden ulaşılan sayısal video içeriğindeki büyümedir. Internet üzerinde YouTube, Google Video, Metacafe gibi kullanıcı eksenli video içerik veritabanlarının büyüklüklerinin üssel artışı ve H.264 gibi çoğulortam kodlama tekniklerinin sıradan verimli kullanıcıların gündemine girmesi, video içeriği üzerindeki fikir haklarının korunması sorununa farklı yaklaşımları gerektirmektedir. Bahsedilen büyüklükte video veritabanları üzerinde telif hakları ile korunan videoların kopyalarının tespiti, insan gücünü aşan büyüklükte bir sorundur. Bu soruna önerilen önemli çözümlerden biri gizli damgalama ile video üzerine gözle fark edilemeyecek sayısal parmak izleri eklenmesi ve bu parmak izlerinin kullanıcılar tarafından yüklenen videolarda aranmasıdır. Gizli damgalar, düşük görünürlükte, gürbüz ve ayrıştırıcı oldukları müddetçe ise yarar bir kopya sezim başarımı sunmaktadırlar.

Fikir haklarının korunması sorunu için gizli damgalama vönteminin kullanılmasının en büyük sıkıntısı, gizli kopyası aranacak videoların kaynağında damgaların eklenmesi gerekliliğidir. Dolayısıyla hali hazırda dolaşımda bulunan videolara ne gizli damga eklemek ne de bu şekilde onların kopyalarının bulunması mümkün değildir. Telif hakları ile korunan videoların dışında (örneğin sakıncalı ya da yasal olmayan içerik) bir kopya sezim sorunu ile karşılaşıldığında videoya önceden parmak izi eklenmiş olması çok düşük bir olasılıktır. Gizli damgalamanın tüm bu eksikliklerine cevap verecek, video üzerinde bir parmak izi bırakma gereksinimi olmayan edilgen bir kopya sezim bakıs açısı geliştirilmesi gerektiği açıktır. Son yıllarda önerilen bu edilgen yaklaşıma İçerik Tabanlı Kopya Sezimi adı verilir.

İçerik Tabanlı Kopya Sezimi, telif hakkı ihlallerinin ya da kopya videoların bulunmasında, veriler arasında karşılaştırma yapılacak parmak izinin içerikten oluşturulması esasına dayanır. Yaklaşımdaki temel düşünce verinin kendisinin damga özelliği taşıdığıdır. İçerik tabanlı kopya seziminde karşılaşılan temel sorun, görünmez damgalamada olduğu gibi,



Şekil 1: (a) Özgün görüntü ve içerik tabanlı video kopya seziminde gürbüzlük aranan saldırı tiplerinden bazıları: (b) resim içinde resim, (c) şekil ekleme, (d) parlaklık değiştirme, (e) karşıtlık değiştirme, (f) en-boy oranı değiştirme, (g) kırpma, (h) kaydırma, (i) ayna çevrimi ve (j) görüntü ekleme.

içerik olarak kopya içeren videoların, sayısal olarak özgün videodan farklılıklar içereceğidir. Kodlamadan kaynaklanan sayısal dizilim değişimi gibi basit ayrımlar olabileceği gibi, karşıtlık değiştirme, gürültü ekleme, geometrik dönüşümler ya da hız değiştirme gibi videonun özgünlüğüne yapılmış esaslı saldırılarla da karşılaşılabilir. Etken ya da edilgen her video kopya sezim yönteminin beklenen saldırılara karşı dayanıklılık göstermesi beklenir.

İçerik tabanlı video kopya sezim üzerine olan çalışmalar iki ana akıma ayrılırlar. Video içeriğinin tamamı ya da seçilen anahtar karelerin bütününden çıkarılan parmak izleri ile kopya sistemler, ilk ana akım olarak araması yapan değerlendirilebilir. Kim, çalışmasında [1] karelerin yalnızca 2x2'lik bölünmesinden elde edilen ortalama ton bilgisi eşlemesi ile bir kopya bulma yöntemi önermiştir. Yöntemin, renk ve parlaklık değiştirme gibi videonun özgünlüğüne saldırmayan değişikliklerde geçerli olabileceği kaydedilmiştir. Esen [2] ise video karelerini bir ağaç yapısıyla bölerek çıkarılan basit imge özniteliklerini nicemlemeyi önermiştir. Chiu [3], karelerden çıkarılacak özniteliklerin, zaman ekseninde Markov temelli olasılıksal bir modele oturtularak eşleştirilmesi ile, genel öznitelikli yöntemlerin, zamansal düzlemde gerçekleşebilecek saldırılardan korunabileceğini ifade etmiştir. Bu türdeki çalışmalar, öznitelik çıkarılacak video bölümlerinin seçiminde anahtar kare seçimi kullananlar [1, 3] ve hacim üzerinde çalışanlar [2] olarak iki alt sınıfa ayrılabilirler. Genel öznitelikli yöntemlerin kullanabileceği çok sayıda denenmiş ve olgunlaşmış parmak izi çıkarma yöntemi bulunmaktadır. MPEG-7 [4] gibi standartlar sayesinde sistemlerin geçerliliklerini artırmaları mümkündür. Görüntü ekleme, kırpma, ölçeklendirme, döndürme gibi videoların özgünlüklerine düzenlenebilecek saldırılar sonrası görüntünün genel özelliklerinde ortaya çıkan etkin değişiklikler, genel parmak izi kullanan yöntemlerin zayıf noktasıdır.

Video kopya sezim sorununda özgünlüğe yönelik saldırıları alt edebilmek için kullanılan yöntemlerde yerel parmak izi çıkarılmasına yönelim gerçekleşmiştir. Elde edilen yerellikle, videoların özgünlüklerinde gerçekleştirilen değişikliklerden bir ölçüde sakınmak amaçlanır. Law-To [5, 6] yerel noktaların zamandaki yörüngelerini hesaba alırken, Joly [7, 8] yerel parmak izleri ile olasılıksal karşılaştırma önerir. Yerel özniteliklerin başlıca sıkıntısı arama uzayını büyütmeleri ve arama süresini uzatmalarıdır.

Son zamanlarda ise hem yerel hem de genel özniteliklerin bir arada kullanılmasıyla [9, 10, 11, 12] özgünlük saldırılarını tam anlamıyla karşılamaya bir adım daha yaklaşılmıştır. Ayrıca, Saracoğlu'nun [13] da çalışmasında onayladığı gibi, ses üzerinden kopya bulmanın görüntüye kıyasla daha kolay olduğu ve bütün bir video kopya sisteminin ses verisini de değerlendirmesinin faydalı olacağı bilinmektedir.

Bu bildiride, video karelerinden çıkarılan yerel tanımlayıcı özniteliklerin küme numaraları ile çalışan, edilgen bir video kopya sezim yöntemi önerilmektedir. Öncelikle kopyası aranacak videonun seçilen karelerinde çeşitli ilgi noktaları ve bölgesel karşılıkları bulunur. Daha sonra bu bölgelerde çıkarılan tanımlayıcılar, daha önceden elde edilmiş kod kitabına bakılarak etiketlenir. Tarama, elde edilen yerel bölge etiketleri üzerinden yapılır. Bildirinin ikinci bölümünde önerilen yöntem anlatılmaktadır. Üçüncü bölümde TRECVID 2009 çalıştayının İçerik Tabanlı Kopya Sezimi [14] bölümünde kullanılan test verileri ile elde edilen sonuçlar verilmektedir. Son bölümde ise çalışmanın özetine ve geliştirilmesine yönelik fikirlere yer verilmektedir.

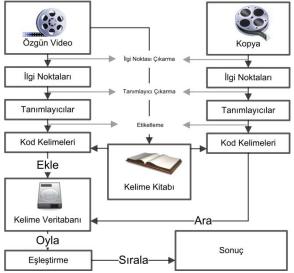
2. Önerilen Yöntem

Video kopya sezimi için yapılması gereken, özgün ve kopya videolardan çıkarılmış parmak izlerinin aynı içerik için eşleniyor olmasıdır. Yöntemimiz videolar içinden çıkarılmış iki boyutlu ilgi noktalarının eşlenmesini esas alır. Her çıkarılan ilgi noktasına atanacak bir etiket üzerinden karşılaştırma mümkün kılınır.

Herhangi bir video karesinden çıkarılmış ilgi noktalarından (p_i) , önce bir yerel öznitelik çıkarıcısı ile elde edilen tanımlayıcılar (f_i) , daha sonra da bu tanımlayıcılara verilmiş etiketler (c_i) elde edilir.

Elde edilen etiketlerin gürbüzlük sağlaması beklenen çeşitli saldırı tipleri Şekil 1'de yer almaktadır. Bu saldırılara karşı gürbüzlüğü belirleyen etmenler, kullanılan ilgi noktası ve yerel tanımlayıcı çıkarıcı ile etiketleyicinin yapısından kaynaklanır. Örneğin, sistemin görüntünün boyunun değiştirildiği saldırılara karşı koyabilmesi için kullanılan tanımlayıcıların ölçek bağımsız olması önerilmektedir. Döndürme saldırısının ise, sunum kalitesinin kopyalama işi sonrası çok darbe alacağı varsayımıyla küçük açılarda yapılacağı kabul edilebilir.

Etiketleme işi, tanımlayıcı vektörlerini kelime kodlarına indirgeyen bir kod kitabı üzerinden yapılır. Bu kod kitabı



Şekil 2: Yerel tanımlayıcılar ile içerik tabanlı video kopya sezimi yönteminin akış şeması.

herhangi bir ayrıştırma ya da kümeleme yöntemi ile elde edilebilir. Kod kitabının büyüklüğü, sistemin genelleyiciliği ile ayrıştırıcılığı arasında bir denge unsurudur.

Gösterimi (t_i, c_i, m_i) olarak verilen *göreli zaman-kelime kodu-ek bilgi* üçlülerinin bir listesi, sistemde bir video parçasını tanımlamak için yeterlidir. Ek bilgi olarak adlandırılan terim, ölçek ya da pozisyon gibi ilgi noktası ya da yerel tanımlayıcı çıkarıcılardan elde edilen bilgileri ifade eder. İki video parçasına $(\ddot{O} \text{ ve } K)$ ait kodlar üzerinden videoların benzerlik ölçütü (1)'de verilmektedir.

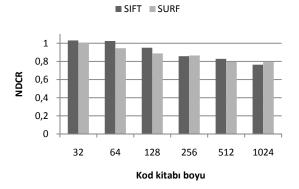
$$benzerlik_{\ddot{0},K} = \frac{1}{\ddot{sure}} \sum_{i} \sum_{j} \delta(t_i, t_j) \delta(c_i, c_j) p(m_i, m_j)$$

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

Video parçası benzerliği, bir diğer ifadeyle, karşılıklı karelerde aynı etikete sahip ilgi noktalarının benzerliklerinin ortalamasıdır. Eşitlikte p(a,b) işlevi, iki ilgi noktası arasındaki benzerliği ifade eder ve sistem gereksinimlerine göre tasarlanabilir. Bu çalışmada ilgi noktası çıkarıcısından elde edilen yönü $(y\ddot{o}n_l)$ ve kare karmaşıklığını ele alan (2) ölçütü esas alınmıştır. Kare karmaşıklığı, ilgi noktasının bulunduğu karedeki toplam ilgi noktası sayısı (k) olarak ifade edilebilir.

$$p(a,b) = \cos(y \ddot{o} n_a - y \ddot{o} n_b) \frac{\min(k(a), k(b))}{\max(k(a), k(b))}$$
(2)

Video benzerlik ölçütünün olası her karşılaştırma için kullanılması sistemin ölçeklendirilebilirliğini engeller. Kopyası aranacak videoların içerisindeki yerel tanımlayıcı etiketler, yalnızca eş etiketli ilgi noktaları ile eşleşeceği için, video veritabanını etiket numarası ile erişilebilen ilgi noktası veritabanına çevirmek, hızlı sorgulama için faydalı olacaktır. Kopyası aranacak video içerisinde karşılaşılacak her etiket için, o etikete sahip ilgi noktalarının veritabanından çekilerek oylanması yeterlidir. Toplanan oylar veritabanındaki karşılıklı video parçalarına verilerek veritabanındaki tüm zaman dilimlerinin, kopyası aranacak videoya benzerliği elde edilebilir. Daha sonra bu sonuçlar sıralanır ve en benzer video parçaları tarama sonucu olarak dönülür. Yöntemin akış özeti Şekil 2'de verilmektedir.



Şekil 3: Kod kitabı boyuna göre yöntemin NDCR başarımı.

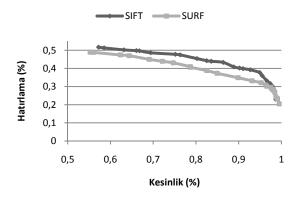
3. Deneyler

Yöntemin denenmesi, TRECVID 2009 İçerik Tabanlı Kopya Sezimi görevinde kullanılan 400 saatlik MPEG-1 kodlanmış CIF boyutunda özgün video ve bu videoların çeşitli saldırılar sonucu dönüştürülmesiyle elde edilen 3 saniye ve 3 dakika arasında olmak üzere 1407 adet sorgu videosu ile yapılmıştır. Sorgu videoları, özgün videolar arasında bulunanlar, özgün videolar arasında bulunmayan iki videonun birleşimi ile özgün videolar arasında bulunmayan görüntüler olmak üzere eşit sayıda üçe ayrılmıştır. Aralarında şekil-yazı bindirme, ölçekleme, kırpma, karşıtlık değiştirme, kare atma, kaydırma da olan saldırıların bir kısmı Şekil 1'te örneklenmektedir.

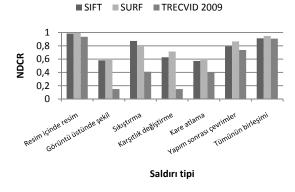
karelerinde. Videoların alınan saniyede bir karşılaştırılmak üzere ölçek-bağımsız öznitelik dönüşümü (SIFT) [15] ve hızlandırılmış gürbüz öznitelikler (SURF) [16] yöntemleriyle elde edilmiş iki farklı ilgi noktası topluluğu çıkarılmıştır. Ayrıca, ayna çevrimine karşı koyulması amacıyla, sorgu videolarının ayna terslerinden de ilgi noktaları elde edilmiştir. Bu ilgi noktalarından sırasıyla yine SIFT ve SURF için önerilmiş yerel tanımlayıcılar elde edilmiş ve bu tanımlayıcılardan özgün videolara ait 10⁵ adetiyle, çeşitli boylarda kelime kitapları oluşturulmuştur. Her bir kelime kitabı için tüm tanımlayıcı noktaların kod kelimesi bulunmuş ve her test düzeneği için 45'er milyon kod kelimesi ve onların ek geometrik bilgilerini içeren veritabanları oluşturulmuştur.

Yöntemin SIFT ve SURF ile kullanıldığında kod kitabı büyüklüğüne göre bütün sorgular için ortalama performansı Şekil 3'te gösterilmektedir. Başarım, her sorgu için yöntemin önerdiği ilk özgün video konumu için, hata ölçütü düzgelenmiş sezim maliyet oranı (normalized detection cost rate) [17] türünden ölçülmüştür. NDCR'ın sıfır olması, tam performansı temsil ederken, hiç bir sonuç dönülmediğinde alacağı değer bir olacaktır. Her iki durumda da 1024 kelimeli kod kitabı ile en iyi sonuçlar elde edildiği için diğer sonuçlar bu kod kitabı büyüklüğü üzerinden değerlendirilecektir.

Sınanan iki farklı ilgi noktası belirteci ve yerel tanımlayıcısı ile bütün sorgu videoları üzerinden elde edilen keskinlik-hatırlama eğrileri ise Şekil 4'te gösterilmektedir. Buradan çıkarılacak sonuç, SIFT tanımlayıcısının bahsedilen saldırı tiplerine, az da olsa, genel olarak daha dayanıklı olduğudur. Tanımlayıcıların her bir saldırı tipi için ayrı başarımları da Şekil 5'te görülmektedir. Saldırı tiplerinden olan *yapım sonrası çevrimler*, kırpma, kaydırma, altyazı ekleme, ayna çevrimi, şekil ekleme ve resim üstüne resim gibi



Şekil 4: Yöntemin farklı yerel tanımlayıcılar ile elde ettiği hatırlama-kesinlik eğrileri.



Şekil 5: Tüm saldırı tipleri için yöntemin ve en iyi yöntemlerin elde ettikleri NDCR başarımları.

videonun özgünlüğüne yönelik saldırıların karışımını içerir. *Tümünün birleşimi* saldırısında ise bunlara ek olarak bulanıklaştırma, kare atma, karşıtlık değiştirme, sıkıştırma, en-boy oranı değiştirme ve gürültü ekleme gibi sunuma yönelik saldırılar da bulunmaktadır. Aynı şekil üzerinde *TRECVID 2009 İçerik Tabanlı Kopya Sezimi* görevinde ayrı ayrı elde edilmiş en iyi NDCR'lar verilmektedir. Şekilde görülen *görüntü üstünde şekil, sıkıştırma, karşıtlık değiştirme* ve *kare atlama* saldırılarında önerilen ve yerel tanımlayıcı yöntemini geçen sistemlerin hepsi yerel ve genel öznitelikleri birleştiren karma yöntemler kullanmaktadırlar [10, 11, 12].

4. Sonuçlar

Bu çalışmada, giderek daha önemli hale gelen video kopya sezimi sorununun çözümü için yerel ilgi noktalarından çıkarılmış tanımlayıcıların kullanımını temel alan bir yaklaşım sunulmuştur. Yaklaşımda, ölçeklendirilebilirlik tanımlayıcıların kod kelimeleri ile etiketlenmesi ve bu etiketlerin karşılaştırması ile elde edilmiştir. Yapılan geniş kapsamlı deneyler, yaklaşımın salt genel öznitelik kullanan yaklaşımlara göre, beklenen tüm saldırılara karşılık verebildiği için, daha tercih edilir olduğunu fakat karma yöntemler kadar başarılı olamadığını göstermektedir. Buna rağmen, çalışmada sunulan şablon, genel özniteliklerin kopya bulma sistemine *ek bilgi* yordamı ile girilmesine olanak tanımakta olduğu için, ileride geliştirilecek yöntemlerin geçerliliği sınamak için bir alt yapı sunmaktadır.

5. Kaynakça

- Kim, C., Vasudev, B., "Spatiotemporal sequence matching for efficient video copy detection," *IEEE Transactions on CSVT*, vol.15, no.1, pp. 127-132, Jan. 2005
- [2] Esen, E., Saracoğlu, A., Ateş, T.K., Oskay Acar, B., Zubari, Ü., Alatan, A.A., "Kaba Öznitelikler ile İçerik Tabanlı Video Kopya Sezimi," *IEEE 17. Sinyal İşleme, İletişim ve Uygulamaları*, SIU 2009, 9-11 April 2009
- [3] Chiu, C.-Y., Chen, C.-S., Chien, L.-F., "A Framework for Handling Spatiotemporal Variations in Video Copy Detection," *IEEE Transactions on CSVT*, vol.18, no.3, pp.412-417, March 2008
- [4] Manjunath, B.S. Introduction to MPEG-7: Multimedia Content Description Interface, 2002, ISBN:0471486787.
- [5] Law-To, J., Gouet-Brunet, V., Buisson, O., Boujemaa, N., "Local Behaviours Labelling for Content Based Video Copy Detection," Proceedings of the 18th International Conference on Pattern Recognition, p.232-235, August 20-24, 2006
- [6] Law-To, J., Buisson, O., Gouet-Brunet, V., Boujemaa, N., "Robust Voting Algorithm Based on Labels of Behavior for Video Copy Detection," *Proceedings of the 14th annual ACM international conference on Multimedia*, October 23-27, 2006, Santa Barbara, CA, USA
- [7] Joly, A., Buisson, O., Frelicot, C., "Content-Based Copy Retrieval Using Distortion-Based Probabilistic Similarity Search," *IEEE Transactions on Multimedia*, vol.9, no.2, pp.293-306, Feb. 2007
- [8] Joly, A., Law-to, J., Boujemaa, N., "INRIA-IMEDIA TRECVID 2008: Video Copy Detection," *Online Proceedings of TRECVID* 2008 Workshop (Gaithersburg, MD, USA, November 17 – 18, 2008).
- [9] Ren, H., Lin, S., Zhang, D., Tang, S., Gao, K., "Visual Words Based Spatiotemporal Sequence Matching in Video Copy Detection," *IEEE Intl. Conf. on Multimedia and Expo 2009*, ICME 2009.
- [10] Cao, J., Zhang, Y.-D., Feng, B.L., Bao, L., Pang, L., Li, J.-T., Gao, K., Wu, X., Xie, H.-T., Zhang, W., Mao, Z.-D., "TRECVID 2009 of MCG-ICT-CAS," Online Proceedings of TRECVID 2009.
- [11] Liu, Z., Liu, T., Shahraray, B., "AT&T Research at TRECVID 2009 Content-based Copy Detection," *Online Proceedings of TRECVID* 2009.
- [12] Le, D.-D., Poullot, S., Crucianu, M., Wu, X., Nett, M., Houle, M.E., Satoh, S.I., "National Institute of Informatics, Japan at TRECVID 2009," Online Proceedings of TRECVID 2009.
- [13] Saracoğlu, A., Esen, E., Ateş, T.K., Oskay Acar, B., Zubari, Ü., Ozan, E.C., Özalp, E., Alatan, A.A., Çiloğlu, T., "Content based copy detection with coarse audio-visual fingerprints," Seventh *International Workshop on Content-Based Multimedia Indexing*, CBMI 2009, pp.213-218, 3-5 June 2009
- [14] TREC Video Retrieval Evaluation Home Page http://www-nlpir.nist.gov/projects/trecvid/
- [15] Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, 60, 2, pp. 91-110, 2004
- [16] Bay, H., Ess, A., Tuytelaars, T., Gool, L.V., "SURF: Speeded Up Robust Features," Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008
- [17] TRECVID CBCD Değerlendirme Dökümanı, http://www.itl.nist.gov/iad/mig/tests/trecvid/2009/doc/EventDet 09-EvalPlan-v03.htm#Evaluation