MPEG-7 UYUMLU VİDEO VERİ TABANLARI İÇİN ÖNEMLİ NESNELERİN OTOMATİK OLARAK BULUNMASI
Automatic Extraction of Important Objects for an MPEG-7 Compliant Video Database System

Muhammet Başatan, Uğur Güdükbay, Özgür Ulusoy

Bilgisayar Mühendisliği Bölümü
Bilkent Üniversitesi, Bilkent, Ankara
{bastan,gudukbay,oulusoy}@cs.bilkent.edu.tr

Özetçe

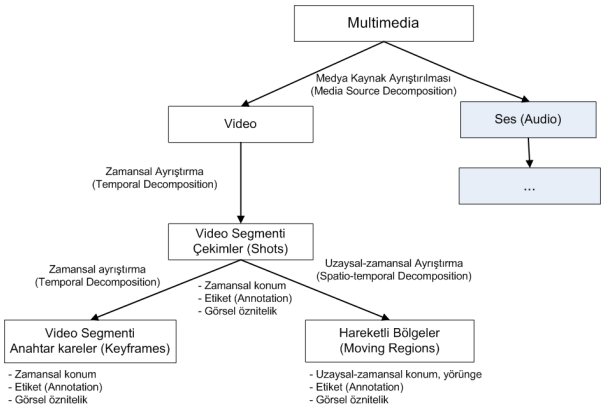
Bu çalışma, genel olarak nesneye dayalı endekslemeyi destekleyen, özel olarak MPEG-7 uyumlu veritabanları için, videolardan önemli nesnelerin otomatik olarak çıkarılmasını sağlayabilecek bir yöntem sunmaktadır. Şimdiye kadar yapılan benzer çalışmalar genellikle resimler üzerinde yoğunlaşmış ve sadece ilk bakışta dikkati çeken alanları bulmaya çalışmıştır. Önerilen yöntem ise videolar üzerinde çalışmak için tasarlanmış olup sadece ilk bakışta dikkat çeken bölgelerin değil, videonun endekslenmesi için önemli sayılabilecek bölgelerin de bulunabilmesini amaçlamaktadır. Bunun için önce video kareleri bölütlere ayrılmakta, sonra her bölüt için yerel ve genel renk, biçim, doku ve hareket bilgileri hesaplanmakta, son olarak bu özellikler kullanılarak eğitilmiş bir destek vektör makinesi (SVM) kullanılarak bölgelerin önemli olup olmadığına karar verilmektedir. İlk deney sonuçları önerilen yöntemin başarılı olduğunu ve elde edilen nesnelerin öncekilere göre anlamsal olarak daha iyi olduğunu göstermektedir.

Abstract

We describe a method to automatically extract video objects, which are important for object-based indexing of videos in an MPEG-7 compliant video database system. Most of the existing salient object detection approaches detect visually conspicuous image structures, while our method aims to find regions that may be important for indexing in a video database system. Our method works on a shot basis. We first segment each frame to obtain homogeneous regions in terms of color and texture. Then, we extract a set of local and global color, shape, texture and motion features for each region. Finally, the regions are classified as being salient or non-salient using SVMs trained on a few hundreds of example regions. Experimental results from news video segments show that the proposed method is more effective in extracting the important regions in terms of human visual perception.

1. GİRİŞ

MPEG-7 uyumlu bir video veri tabanı sistemi nesneye dayalı, oldukça gelişmiş mekansal ve zamansal (spatio-temporal) sorgulamaları destekleyebilir. Örnek olarak, belli renk, biçim, doku, yörünge ve etikete sahip bir nesnenin belli özelliklere sahip bir sahnede geçtiği video bölümlerine böyle bir sistem yardımıyla erişilebilir. Bunun için videonun uygun şekilde endekslenmesi gerekmektedir. Şekil 1'de gösterildiği gibi video önce zamansal olarak çekimlere ayrılır (shot boundary detection). Sonra her çekim içindeki önemli bölgeler bulunup (çekimlerin mekansal ve zamansal olarak ayrıştırılması) öznitelikleri hesaplanır ve MPEG-7'nin *Hareketli Bölge (Moving Region)* tanımlayıcısı ile tanımlanır. Çekimlerin geri planda kalan statik içeriği ise temsili karelerle veya onların daha küçük *Duran Bölgelere (Still Region)* ayrılıp özniteliklerinin çıkarılması ile temsil edilebilir. Bu işlemlerin büyük video veri tabanları için elle yapılması imkansızdır. Bu yüzden, özellikle çekimlerdeki önemli bölgelerin otomatik olarak bulunabilmesi sistemin pratikte kullanılabilir olması açısından çok önemlidir.



Şekil 1: Bir videonun MPEG-7 uyumlu bir veri tabanında saklanabilmesi için uzaysal ve zamansal ayrıştırılması, bölütlenmesi ve temsili.

Daha önce yapılan çalışmalarda genellikle resme ilk bakıldığında göze çarpan, dikkati çeken alanların bulunması konusu üzerinde yoğunlaşılmıştır. Bu konuda yapılan ilk önemli çalışma [1] ile sunulmuştur. Bu çalışmada ve takip eden benzer çalışmalarda genellikle resmin renk ve doku özelliklerinden *dikkat çekme haritaları* (saliency maps) elde

978-1-4244-1999-9/08/$25.00 ©2008 IEEE

edilmiş ve bu haritalar kullanılarak göze en fazla çarpan resim bölgeleri bulunmaya çalışılmıştır [2, 3, 4, 5, 6, 7, 8]. Video için önerilen ve hareket bilgisini kullanan yöntemler az sayıdadır [9, 10, 11]. Bu yöntemlerle elde edilen bölgeler çoğu zaman video endeksleme açısından önemsiz kalmaktadır. Bu çalışmada, var olan yöntemlerin yetersizliği göz önünde bulundurularak, video üzerinde çalışan; renk, biçim, doku gibi özelliklerin yanında hareket bilgisini de kullanan bir yöntem önerilmiştir.

2. ÖNEMLİ VİDEO NESNELERİNİN BULUNMASI

Önemli nesnelerin bulunması 5 aşamada gerçekleştirilir. (1) Videoların çekimlere bölütlenmesi (shot boundary detection) ve her çekimin ayrı ayrı işlenmesi. (2) Her çekim içinde, video karelerinin renk ve doku bilgilerine göre homojen bölütlere ayrılması (spatial segmentation). (3) Elde edilen her bölüt/bölge için renk, biçim, doku ve hareket özelliklerinin hesaplanması. (4) Elle etiketlenmiş, birkaç yüz örnekten oluşan eğitim seti kullanılarak eğitilmiş destek vektör makinesi (SVM) kullanılarak bölütlerin önemli olup olmadığına karar verilmesi. (5) Önemli bölgelerin çekim içinde takip edilmesi ve son önemli bölge kümesinin belirlenmesi.

2.1. Videoların Çekimlere Bölütlenmesi

Renk histogramına dayalı yöntemler basit olmakla birlikte iyi sonuç vermektedir. Bu çalışmada da çekimler art arda gelen video karelerinin HSV renk uzayında elde edilen histogramları karşılaştırılarak elde edilmektedir.

2.2. Video Karelerinin Bölütlere Ayrılması

Video karelerinin bölütlere ayrılması önerilen yöntemde kilit rol oynamaktadır. Çünkü elde edilen bölütlerin kalitesi daha sonraki aşamada bu bölütlerden hesaplanacak özellikleri ve sonuç olarak elde edilen önemli bölgelerin doğruluğunu doğrudan etkilemektedir.

Bu çalışmada kaynak kodları internette herkese açık olan JSeg resim bölütleme algoritması [13] önerilen sisteme adapte edilerek kullanılmıştır. Bu algoritma, önce resimdeki renkleri YUV renk uzayında birkaç renk sınıfına indirgemekte (color quantization), daha sonra resimde yerel pencereler üzerinde bir "iyi bölütleme" kriteri uygulayarak J-resmi (J-Image) adı verilen bir resim elde etmekte ve son olarak da çok ölçekli (multi-scale) J-resimlerini bölge büyütme (region growing) algoritması ile kullanarak renk ve doku olarak homojen bölütler elde etmektedir.

2.3. Önemli Bölgelerin Özellikleri

Video nesnelerinin hangisinin önemli olup veri tabanında saklanması gerektiği kişiden kişiye değişebilecek öznel bir konu olmakla birlikte bazı genel kurallar uygulanarak çoğu kişinin hemfikir olabileceği sonuçlar elde edilebilir. Bu çalışmada aşağıdaki genel kurallar yardımıyla önemli bölgeler diğerlerinden ayrıştırılmaya çalışılmıştır.

- Videolarda kameranın üzerine odaklandığı nesneler genellikle önemlidir. Örnek olarak, bir haber videosunda kamera, stüdyoda haber sunan kişiye odaklanmaktadır. Kameranın odaklandığı nesnelerin kontrastı yüksek olup kenarları daha keskindir. O yüzden bölgelerin değişinti (variance) ve entropi değerleri kullanılabilir.
- Görsel olarak dikkat çeken nesneler önemli olabilir. Çevrelerinden farklı özelliklere sahip bölgeler dikkat çekerler. O yüzden bölgelerin diğer bölgelerden ve bütün resimden ne kadar farklı olduğu hesaplanıp kullanılabilir.
- Hareketli nesneler önemli olabilir (örnek: yürüyen adam, hareketli araba/uçak).
- Resimde çok büyük yer kaplayan, çok küçük olan, çok ince ve uzun bölgeler genellikle önemli değildir. Çok yer kaplayan bölge arka plan; çok küçük, çok ince ve uzun bölgeler bölütleme hatalarından kaynaklanabilir. Dolayısıyla bölgelerin biçimsel özellikleri kullanılabilir.
- Önemli nesneler tutarlı olmalıdır (consistency); her çekim içinde en az belli sayıda karede bulunmalıdır (örneğin çekimdeki toplam kare sayısının % 10'u).

Bu kurallar yardımıyla her bölge için aşağıdaki özellikler hesaplanıp uzunluğu 18 olan bir öznitelik vektörü elde edilmektedir.

- Her bölgenin değişinti (variance) ve entropi değerleri.
- Her bölgenin X ve Y yönündeki ortalama hızları. Hızlar optik akıntı (optical flow) ile elde edilmektedir.
- Biçim özellikleri: bölge alanının tüm resim alanına oranı, en-boy oranı (aspect ratio), bölge alanının bölgenin MBR (Minimum Bounding Rectangle - En küçük sınırlayan dikdörtgen) alanına oranı.
- Kontrast özellikleri: bölgelerin ortalama renk değerlerinin komşu bölgelerin renklerinden farkları, diğer bütün bölgelerden farkları; bölgelerin komşu bölgelerden ortak sınır üzerindeki farkları (iki bölge arasındaki kenarın keskinliği); bölgelerin komşu bölgeler ve diğer bölgelerden değişinti (variance), entropi ve hız farkları.

2.4. Önemli Bölgelerin Seçilmesi

JSeg algoritması ile elde edilen bölütlerden pozitif ve negatif 300+ örnek seçilerek çıkarılan öznitelikler normalize edilmekte ve polinomsal çekirdeğe sahip bir destek vektör makinesi (SVM) eğitilmektedir [14]. Test aşamasında her bölgeden elde edilen öznitelikler SVM'e gönderilip çıktıya göre o bölgenin önemli olup olmadığına karar verilmektedir. Elde edilen bölgelerin sayısını azaltmak için, bölgeler SVM'deki ayırıcı düzleme uzaklıklarına göre puanlanarak sıralanmakta ve en yüksek puana sahip ilk N bölge seçilmektedir. N değişkeni, sistemin geri getirme (recall) ve kesinlik (precision) değerleri için bir kontrol parametresi olarak kullanılabilir; N büyük seçilirse geri getirme artarken kesinlik azalacaktır. Bu çalışmada, kullanılan veri kümesinin özellikleri dikkate alınarak N parametresinin değeri 5 olarak seçilmiştir. Böylece SVM 5'ten fazla bölge döndürdüğünde bunların ilk 5 tanesi dikkate alınmaktadır.

2.5. Önemli Bölgelerin Takibi

Hem bulunan bölgelerin MPEG-7 gösterimindeki yörünge değerleri, hem de önemli nesneler için belirlenen özelliklerden olan tutarlılık değerinin hesaplanabilmesi için elde edilen bölgelerin takip edilmesi gereklidir. Bu çalışmada, bulunan önemli bölgelerin art arda gelen video kareleri arasındaki takibi, bölgelerin renk histogramları kullanılarak yapılmıştır. Her önemli bölge için önce belli bir yerel pencere belirlenmekte; sonra sözkonusu bölge, bu pencere içine düşen bütün uygun bölgeler ile renk histogram uzaklığına göre karşılaştırılmakta ve uzaklığı belli bir eşik değerinin altında olan en yakın bölge ile eşlenmektedir. Eşleme işlemi başarısız olduğunda, önceki karede geçen önemli bölgenin bu karede olmadığı ya da bulunamadığı sonucuna varılmaktadır.

3. DENEY SONUÇLARI

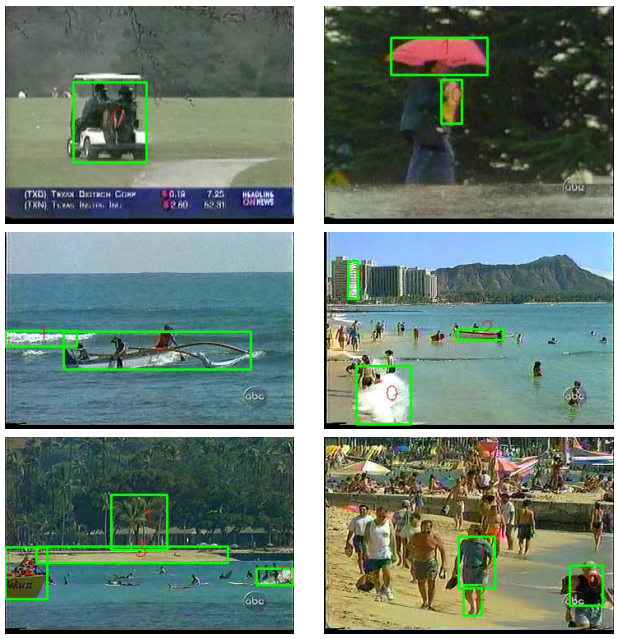
Bu bölümde kısa videolar üzerinde yapılmış deneylerde elde edilen sonuçlar verilmiş, sistemin performansı benzer bir sistemle karşılaştırılmıştır. Karşılaştırma için, resimlerde göze çarpan bölgeleri bulma konusunda ilk çalışmaları yapan araştırmacıların [1] geliştirdiği, MATLAB kaynak kodları internette herkese açık olarak bulunan yöntem (SM: saliency model) kullanılmıştır [15].

(a) (b)

Şekil 2: Görsel örneklerle iki yöntemin karşılaştırılması (bulunan ilk 5 bölge). (a) SM yöntemi, (b) Bu çalışmada önerilen yöntem (her bölge içindeki sayı o bölgenin bulunma sırasını göstermektedir).

Şekil 2'de SM ve bu çalışmada önerilen yöntem kullanılarak elde edilen önemli bölgelere örnekler verilmiştir. Örneklerde de görüldüğü gibi, önerilen yöntem görsel olarak daha anlamlı bölgeler bulabilmektedir. İlk resimde, önerilen yöntem yüz ve abajure ait bölgeleri başarı ile bulurken SM yöntemi ise pek başarılı olamamıştır.

İki yöntemi sayısal olarak karşılaştırmak için toplam 668 kareye sahip 2 video seçilip her iki yöntem bu videolar üzerinde çalıştırılıp, bulunan ilk 5 bölgenin doğrulukları görsel olarak doğru/yanlış/bulunamayan (correct/wrong/missed) şeklinde etiketlendikten sonra, geri getirme (recall) ve kesinlik (precision) değerleri hesaplandığında, SM yöntemiyle 0.70 ve 0.80 geri getirme seviyelerinde sırasıyla 0.50 ve 0.48; bu çalışmada önerilen yöntemle sırasıyla 0.70 ve 0.60 kesinlik değerleri elde edilmiştir.



(a) (b)

Şekil 3: Önerilen yöntemle bulunan bölgelere örnekler. $N = 5$.

Şekil 3'te, önerilen yöntemle elde edilen bölgelere değişik örnekler verilmiştir. Resim a-1'de golf arabası, a-2'de küçük tekne ve dalga başarı ile bulunabilmiş; resim b-1'deki yürüyen insanın şemsiyesi ve elindekiler farklı renkleri ve hareketlerinden dolayı bulunabilirken, kendisi arka planla çok yakın renkte olduğu için bulunamamış; yine resim b-3 çok karışık olduğu için iyi sonuç alınamamıştır. Sonuç olarak, bölütleme kalitesinin başarı oranını oldukça etkilediği çıkarımında bulunulabilir.

4. TARTIŞMA VE SONUÇLAR

Bu çalışmada MPEG-7 uyumlu veya nesne tabanlı endeksleme gerektiren herhangi bir video veri tabanı için önemli nesnelerin otomatik olarak bulunmasını sağlayacak bir yöntem önerilmiştir. Elde edilen nesneler MPEG-7 *Hareketli Bölge* (Moving Region) tanımlayıcısı ile video veri tabanında tutulabilecek ve veri tabanında nesne tabanlı arama yapmayı

sağlayacaktır. Deney sonuçları önerilen yöntemle elde edilen ilk sonuçların başarılı olduğunu göstermektedir.

Mevcut sistem çeşitli şekillerde geliştirilebilir. Kullanılan özniteliklerin geliştirilip öznitelik seçme teknikleriyle daha uygun bir öznitelik kümesi elde edilebilir. Önceki çalışmalarda kullanılan dikkat çekme haritaları (saliency maps) bölütleme algoritması ile entegre edilebilir. Hareket bilgisi optik akıntı ile değil daha doğru sonuçlar verebilecek önemli noktaların takibi ile hesaplanabilir. Son olarak farklı makine öğrenme teknikleri denenip en iyi sonuç veren teknik kullanılabilir.

Anlamsal bütünlüğe sahip nesnelerin elde edilmesi henüz çözülememiş bir problem olup hala araştırma konusudur. Bu çalışmada da bölütleme algoritması ile elde edilen benzer renk ve dokuya sahip bölgelerin önemli olup olmadığı bulunmaya çalışılmış, bu bölgelerin daha üst düzeyde anlamsal olarak birleştirilmesi problemine çözüm sunulmamıştır. Kullanılan öznitelikler yardımıyla elde edilen bölgelerin hiyerarşik olarak birleştirilmesiyle daha anlamlı bölgeler elde edilmeye çalışılabilir.

5. Teşekkür

Bu çalışma TÜBİTAK EEEAG-105E065 nolu projesi ile Avrupa Birliği 6. Çerçeve Programı FP6-507752 nolu projesi (MUSCLE) tarafından desteklenmektedir.

6. Kaynakça

[1] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, November 1998.

[2] Y.F. Ma and H.J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of the Eleventh ACM International Conference on Multimedia*, 2003, pp. 374–381.

[3] Ueli Rutishauser, Dirk Walther, Christof Koch, and Pietro Perona, "Is bottom-up attention useful for object recognition?," in *International Conference on Computer Vision Pattern Recognition*, July 2004, vol. 2, pp. 37–44.

[4] S. Kwak, B. Ko, and H. Byun, "Automatic salient-object extraction using the contrast map and salient points," in *Advances in Multimedia Information Processing, Lecture Notes in Computer Science (LNCS)*, 2004, vol. 3332, pp. 138–145.

[5] Feng Ge, Song Wang, and Tiecheng Liu, "Image-segmentation evaluation from the perspective of salient object extraction," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006, vol. I, pp. 1146–1153.

[6] Byoung Chul Ko and Jae-Yeal Nam, "Automatic object-of-interest segmentation from natural images," in *Proceedings of the 18th International Conference on Pattern Recognition*, 2006, pp. 45–48.

[7] Xiaodi Hou and Liqing Zhang, "Saliency detection: A spectral residual approach," in *IEEE Conference on Computer Vision Pattern Recognition*, June 2007, pp. 1–8.

[8] T. Liu, J. Sun, N. N. Zheng, X. Tang, and H.Y. Shum, "Learning to detect a salient object," in *IEEE Conference on Computer Vision Pattern Recognition*, June 2007, pp. 1–8.

[9] Guoping Qiu, Xiaodong Gu, Zhibo Chen, Quqing Chen, and Charles Wang, "An information theoretic model of spatiotemporal visual saliency," in *IEEE International Conference on Multimedia and Expo*, July 2007, pp. 1806–1809.

[10] Trent J. Williams and Bruce A. Draper, "An evaluation of motion in artificial selective attention," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2005, vol. 3, p. 85.

[11] O. Le Meur, D. Thoreau, P. Le Callet, and D. Barba, "A spatio-temporal model of the selective human visual attention," in *IEEE International Conference on Image Processing*, September 2005, vol. 3, pp. III–1188–91.

[12] Yang Liu, Christos-Savvas Bouganis, and Peter Y K. Cheung, "A spatiotemporal saliency framework," in *IEEE International Conference on Image Processing*, October 2006, pp. 437–440.

[13] Y. Deng and B.S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800–810, Aug 2001.

[14] Multi-Class Support Vector Machine, http://svmlight.joachims.org/svm_multiclass.html.

[15] Saliency Toolbox, http://www.saliencytoolbox.net.