Haber Videolarında İlgililik Geribeslemesiyle İçerik Tabanlı Erişim
Content-Based Retrieval of News Videos Using Relevance Feedback

Özge Çavuş, Selim Aksoy

Bilgisayar Mühendisliği Bölümü, Bilkent Üniversitesi, Bilkent, 06800, Ankara

{cavus,saksoy}@cs.bilkent.edu.tr

Özetçe

Gerek kamusal gerekse özel video arşivlerinin genişlemesi ve yaygınlaşmasıyla birlikte, özellikle haber videolarında içerik tabanlı erişim önem kazanmıştır. İlgililik geribeslemesi, içerik tabanlı erişim sistemlerinde kullanılan bir tekniktir. Bu makalede, farklı özniteliklerin, videolardan alınan görüntülerin farklı yerlerindeki önemini baz alarak kullandığımız ilgililik geribeslemesi tekniğini açıkladık. Bunu da ızgara kullanarak böldüğümüz resimlerdeki her bir parçaya her bir öznitelik için farklı bir ağırlık değeri vererek gerçekleştirdik. Bu ağırlık değerleri kullanıcın verdiği her geribeslemenin ardından güncellenmektedir. Bu geribeslemeler, kullanıcının dönen sonuçları pozitif ya da negatif olarak etiketlemesinden ibarettir. Her bir ağırlığın yeni değeri, pozitif ve negatif görüntülerin birbirlerine olan uzaklıklarının standart sapmasının, bütün pozitif görüntülerin birbirlerine olan uzaklıklarının standart sapmasına olan oranıyla bulunur. Tekniğin etkinliği, farklı özniteliklerin görüntünün farklı yerlerindeki önemini belirten ağırlık değerlerinin kullanımıyla, TRECVID video kümesindeki spor haber videoları üzerinde gösterilmiştir.

Abstract

Content-based retrieval in news video databases has become an important task with the availability of large quantities of data in both public and proprietary archives. We describe a relevance feedback technique that captures the significance of different features at different spatial locations in an image. Spatial content is modeled by partitioning images into non-overlapping grid cells. Contributions of different features at different locations are modeled using weights defined for each feature in each grid cell. These weights are iteratively updated based on user's feedback in terms of positive and negative labeling of retrieval results. Given this labeling, the weight updating scheme uses the ratios of standard deviations of the distances between relevant and irrelevant images to the standard deviations of the distances between relevant images. The proposed technique is quantitatively and qualitatively evaluated using shots related to several sports from the news video collection of the TRECVID video retrieval evaluation where the weights could capture relative contributions of different features and spatial locations.

1. Giriş

Son yıllarda teknolojideki hızlı gelişmeyle beraber gerek kamusal gerekse özel arşivlerde çok büyük miktarlarda çoğul ortam veri kullanımı gözlenmektedir. Özellikle görsel öğe, ses ve metin içeren haber videoları zengin içerikleri ve sağladıkları sosyal etki dolayısıyla çoğul ortam kaynakları arasında önemli bir yere sahiptir [1]. Bu kaynakların içerik tabanlı indekslenmesi, çözümlenmesi ve sorgulanması için gereken sistemlerin tasarımı gün geçtikçe daha da önemli bir araştırma konusu halini almaktadır.

İçerik tabanlı görüntü ve video sorgulama tekniklerinde ilk yapılan, görüntülerdeki alt düzey öznitelikleri çıkarmak ve oluşturulan öznitelik vektörlerini kullanarak görüntüler arası uzaklıkları hesaplamaktır [2, 3]. Bu konuyla ilgili çalışmalarda, görüntünün bütünü baz alınarak çıkarılan öznitelikler kullanılmıştır. Yakın zamanda gerçekleştirilen çalışmalarda ise öznitelik çıkarma daha çok bölgesel tabanlı çözümlemelere dayanmaktadır [4]. Görüntüyü parçalara ayırdıktan sonra bu parçalardan çıkarılan öznitelik vektörleri iki görüntü arasındaki uzaklığın ölçümünde kullanılmaktadır.

Her ne kadar görüntünün bütününden çıkarılan öznitelikler sonucunda sınırlı veriler üzerinde başarılı sonuçlar alındıysa da kapalı ve açık alanda kaydedilmiş olan haber videolarında aydınlatma, yer, duruş ve kapatılmayla ilgili pek çok sorun yaşanmaktadır. Bu nedenle, görüntünün bütününden çıkarılan ve görüntüyü sınırlı şekilde ifade edebilen öznitelikler ile görüntüdeki nesne çeşitliliği ve arkaplan karmaşıklığı modellenememektedir. Diğer taraftan az sayıda homojen bölümlere ayrılmış görüntüler (Corel veri tabanındaki gibi) üzerinde kullanılan pek çok tanınmış otomatik parçalama teknikleri de, düşük çözünürlüklü haber videolarında düşük veri kalitesi ve yüksek nesne çeşitliliği nedenleriyle iyi sonuçlar vermemektedir.

Bu çalışmada, ilk olarak görüntüleri birbirinden ayrık parçalara bölerek içeriklerini modelledik. Bireysel olarak görüntüdeki her bir bölge için renk, doku ve ayrıt gibi öznitelikleri çıkardıktan sonra, bu özniteliklere ve belli ağırlık değerlerine göre hesapladığımız görüntüler arası uzaklıkları aralarındaki benzerlikleri belirlemek için kullandık. Her bir özniteliğin her bir bölge için hesaplanmış olan ağırlık değeri, bu özniteliğin o bölge için önemini belirtmektedir. Kullanıcının sorgulama sonucu dönen sonuçlar arasından pozitif ve negatif olarak etiketlediği görüntülerle verdiği geribeslemesi, her sorgulama öncesinde bu ağırlık değerlerinin güncelleştirilmesi için kullanılır.

Önceki ilgililik geribeslemesi teknikleri, metinsel belge sorgulama literatüründen alınan sorgu noktası hareketi [4] ile özniteliklerin ağırlıklandırılmasını ve her sorgulamada bu ağırlıkların güncellenmesini [5, 6] içermektedir. Daha yeni çalışmalarda ise en uygun ağırlıkları veya öznitelik dönüşümlerini hesaplamaya çalışan eniyileme tabanlı teknikler [7, 8] ve de görüntü veri tabanını sınıflandırmayı öğrenmek için pozitif ve negatif geribesleme örneklerini kullanan destek vektör makinesi (support

Bu çalışma TÜBİTAK 104E077 ve DPT 2004K120720 numaralı projeler tarafından desteklenmiştir.

vector machine) gibi teknikler [4] yaygınlaşmaktadır. Fakat geribesleme örneklerinin yetersizliği nedeniyle eniyileme tabanlı teknikler pek uygulanabilir değildir [7]. Aynı zamanda destek vektör makinesi teknikleri de az sayıda örnek kullanılması nedeniyle yakınsaklık ve kararlılık sorunu yaşayabilirler [9].

Bizim bu makalede önerdiğimiz teknik, ağırlık güncellemeye dayanan ilgililik geribeslemesi kullanmaktadır [10]. Ağırlık değerleri, pozitif ve negatif görüntüler arasındaki uzaklıkların standart sapmasının, pozitif görüntülerin birbirlerine olan uzaklıkların standart sapmasına oranı seklinde bulunmaktadır. Bu tekniğin başarımı TRECVID video sorgulama yarışmasına [1] ait spor haber videoları kullanılarak değerlendirilmiştir.

Makalenin geri kalanı şu şekilde organize edilmiştir: 2. bölümde uzamsal ızgara planı üzerinden elde edilen alt düzey öznitelikle görüntü gösterimi anlatılmaktadır. Metinsel ve görsel sorgular kullanılan erişim senaryosu 3. bölümde yer almaktadır. 4. bölümde döngülü erişim için ilgililik geribeslemesine, son olarak 5. bölümde ise deneysel sonuçlara yer verilmiştir.

2. Görüntü Modellemesi

Bu çalışmada, görüntülerin uzamsal içeriklerini görüntüleri parçalara ayırarak modelledik. 5 sıra ve 7 sütun olmak üzere 35 ayrık parçalara böldüğümüz 352×240 çözünürlüklü video görüntülerinin her bir parçası için renk, doku ve ayrıt özelliklerine dayanan alt düzey öznitelikler çıkardık. Bu parçalardaki piksellerin RGB, HSV ve LUV değerlerinin ortalaması ve standart sapması o parçanın renk özniteliğini, bu piksellerin 3 farklı ölçekte ve 4 farklı yönelimde Gabor dalgacık tepkilerinin istatistikleri ise parçanın dokusal özniteliğini temsil eder. Canny ayrıt detektör çıktılarının gradyan yönelim değerlerinin histogramları ise ayrıt öznitelikleri olarak kullanılır. Ayrıt öznitelikleri 45 derecelik aralıklarla piksel yönelim değerlerini sayan 8 kutu ve ayrıt olmayan piksellerin sayısını tutan bir kutudan oluşan 9 elemanlı bir histogram kullanılarak özetlenmektedir.

Bu işlem sonucunda her bir görüntü parçası için 5 ayrı öznitelik vektörü elde edilir. Bu vektörler, RGB, HSV ve LUV istatistiklerinin her biri için 6, Gabor için 24, ayrıt yönelim histogramları için 9 değer içerir. Her bir öznitelik vektörünün bireysel bileşenleri, bütün öznitelik erimlerini yaklaşık olarak eşitlemek ve her biri için benzerlik hesaplamalarında hemen hemen aynı etki aralığını yakalamak için birim değişintisine normalize edilir.

3. Görüntü Erişimi

Ne yazık ki görüntü parçalaması kullanılsa bile kullanılan hiçbir öznitelik çıkarma algoritması tamamen başarılı sonuçlar vermeyebilir, birbirine görsel olarak benzer olmayan parçaların öznitelikleri birbirine yakın çıkabilir ve bu durumda sorgulama görüntüsünden ilgisiz olan görüntüler sorgulama sonucunda görülebilirler. Anlambilimsel uçurum (semantic gap) denilen bu problem, etkileşimli erişimi, benzerliğin ve insan algılamasındaki öznelliğin üst düzey içeriğini korumaya yönelik çok önemli bir araştırma problemi haline getirir.

Erişim senaryomuzda, kullanıcı ilk olarak sorgulamasına, metin sorgusu olarak bir ya da birden fazla anahtar kelime girerek başlar. Başlangıç olarak bu kelimelerin, otomatik konuşma tanıma teknikleriyle elde edilen video demeç özetlerinde aramaları yapılarak, kelimeleri içeren görüntüler sorgulama sonucu olarak döndürülür. Bu demeç özetleri, ham olan video metninin köklerine ayırma, etiketlendirme, frekans tabanlı süzgeçleme gibi işlemlerden geçirildikten sonraki halidir.

Metin tabanlı sorgulama sonrasında kullanıcı kendisine dönen görüntüler arasından bir ya da birden fazla görüntü seçer ve bu görsel sorgulama örnekleriyle beraber bütün veri tabanı üzerinde öznitelik vektörlerine dayandırılan görüntü tabanlı bir arama gerçekleştirilir. Bu ilk görsel sorgulamada, görüntüler arası benzerlikler, görüntünün her parçasına ait öznitelik vektörlerinin arasındaki Euclidean uzaklıklar bulunarak hesaplanır. İlk görsel sorgulama döngüsünde her bir görüntü parçasına ait her özniteliğin bu sorgulamaya aynı oranda katkısı olduğu varsayılır. Son olarak bundan sonraki sorgulamalarda kullanıcı kendisine dönen görüntü sonuçlarını pozitif ya da negatif olarak etiketler.

4. İlgililik Geribeslemesi

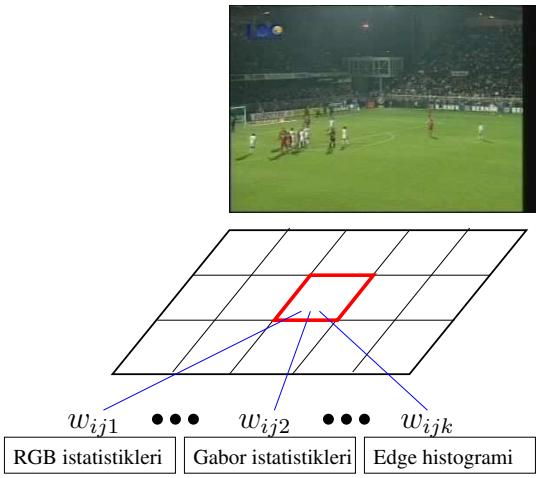
Veri tabanı aramasında, geribesleme bilgileri, görüntüler arası benzerlik hesaplama aşamasında her bir öznitelik vektörünün her bir görüntü parçasına olan katkısını değiştirmek için kullanılır. Bu katkı değişimi, her bir parça için her bir öznitelik vektörüne bir ağırlık değeri atanarak ve bu ağırlık değerlerini her döngü sonucunda alınan geribesleme bilgilerine göre güncelleyerek gerçekleştirilir. Şekil 1'de görüldüğü üzere i numaralı sırada ve j numaralı sütunda yer alan parçaya ait k numaralı özniteliğe atanan ağırlık değeri w_{ijk} ile gösterilmektedir ($i = 1, \ldots, 5$, $j = 1, \ldots, 7$ ve $k = 1, \ldots, 5$). Elimizde iki ayrı görüntü olduğunu varsayalım. İlk olarak i numaralı sıra j numaralı sütunda bulunan parçaların k numaralı öznitelik vektörleri arasındaki uzaklığı hesaplayalım ve değerine d_{ijk} diyelim. Bu durumda, iki görüntü arasındaki genel uzaklık ya da benzerlik hesapladığımız her bir parça arasındaki uzaklıklar kullanılarak

$$d = \sum_i \sum_j \sum_k w_{ijk} d_{ijk} \qquad (1)$$

olarak hesaplanır. Bu çalışmada her bir öznitelik bileşeni için ayrı bir ağırlık kullanılmamıştır. Bunun nedeni, böyle bir işlemin her bir iterasyonda çok sayıda parametrenin kestirimine gerek duyması ve dolayısıyla geribeslemede kullanılan örnek sayısının az olacağı göz önüne alındığında küçük örneklem problemi (small sample problem) ile karşılaşılmasını önlemektir.

Burada farklı öznitelik vektörleri ve görüntü parçaları arasındaki uzaklıkları hesaplarken şu varsayımı kullanmaktayız. Belirli bir parçanın belirli bir öznitelik vektörünün o anki sorgu görüntüsünde önemli bir öznitelik olması için, geribesleme olarak verilen pozitif görüntü örnekleri belirtilen parçanın bu öznitelik vektörü bakımından birbirlerine benzer olmalı, yani bu öznitelik vektörlerinin bu örnekler arasında değişintisi dar olmalıdır. Diğer taraftan da, pozitif ve negatif görüntü örnekleri bu öznitelik bakımından birbirlerinden farklı olmalı, yani bu öznitelik vektörlerinin pozitif ve negatif örnekler arasında değişintisi geniş olmalıdır. Bu nedenle ağırlık değerleri pozitif ve negatif görüntüler arasındaki uzaklıkların standart sapmasının pozitif görüntülerin birbirlerine olan uzaklıklarının standart sapmasına oranı şeklinde bulunur. İlk sorgulamada ise ağırlık değerlerinin her biri 1'e eşitlenir.

Sonuç olarak, ağırlık değerleri belirli bir sorgu görüntüsü için önemli olan öznitelikleri ve görüntü parçalarını yani görüntüdeki



Şekil 1: 3×5'lik bir ızgara planı örneği ve her bir parçaya denk gelen ağırlık değerleri. Deneylerde 5×7'lik bir plan ve 5 ayrı öznitelik vektörü kullanılmıştır.

konumları temsil eder. Örneğin Şekil 1 deki resimde seyirci kalabalığının bulunduğu üst kısımda dokusal öznitelikler önemliyken yeşil sahanın geniş yer kapladığı alt kısımlarda renk öznitelikleri daha ağır basmaktadır.

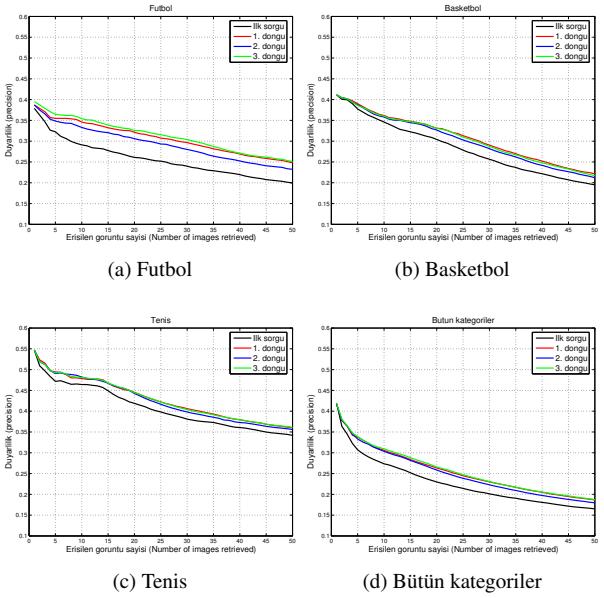
5. Deneysel Sonuçlar

TRECVID 2005 yarışmasına [1] ait 137 adet haber videosu üzerinde makalede bahsettiğimiz teknikleri uyguladık. TRECVID katılımcıları tarafından elle etiketlenen futbol, basketbol, golf, tenis ve Amerikan futbolu olarak 5 ayrı türde spor konuları içeren video görüntülerini, TRECVID veri tabanındaki 43907 ana çerçeve görüntüsü için ground truth olarak kullandık. Şekil 2'de her kategori için örnek video görüntüleri yer almaktadır.

Elde ettiğimiz bu ground truth'u otomatik sorgular üretmek, 43907 görüntüyü her sorgu sonucuna göre sıralamak, ve her döngü sonunda, elde edilen ilk 30 görüntü içindeki her bir görüntüyü otomatik olarak eğer sorgu görüntüsüyle aynı ground truth grubuna ait ise pozitif, geri kalanları ise negatif olarak etiketleyerek geribesleme sağlamak için kullandık. Bu işlem ground truth'da bulunan her bir görüntü için 3 defa geribesleme verilerek yapıldı.

Şekil 3'te ground truth'daki farklı kategoriler için değişik öznitelik kombinasyonları kullanıldığındaki duyarlılık grafikleri gösterilmiştir (yer sınırlaması nedeniyle sadece bu grafikler gösterilmiştir). Şekil 4 ve 5 ise renk özniteliklerini kullandığımız futbol ve basketbol kategori sorgulaması örneklerini göstermektedir. Bütün öznitelik kombinasyonları arasında, Gabor ve renk (RGB, HSV, LUV) öznitelikleri en yüksek ortalama duyarlılığı vermişlerdir. Kullanılan spor kategorilerinde ayrıt yapısının pek önemi olmaması nedeniyle, ayrıt yönelim özniteliklerinin kombinasyonlara katılması durumunda performans artışı gözlenmemiştir. Binalar, evler, ofislerin yer aldığı şehir görüntüleri içeren kategorilerde bu öznitelikler çok daha fazla önem arz edebilirler.

Duyarlılıkta en büyük yükselme ilk döngü sonucunda elde edilmiştir. Bunu takip eden döngülerde bu değer etrafında bir dalgalanma gözlemlenmiştir. Sonuçlara bakıldığında, ağırlık değerleri bir görüntü için farklı bölgelerde farklı öznitelikleri modelleyerek görüntüler arası benzerliklerin hesaplanmasında önemli bir başarı elde etmiş, bütün öznitelik kombinasyonları için geribildirimle alınan sonuçlar geribildirimsiz sonuçlara göre daima gelişme göstermiştir.



(a) Futbol (b) Basketbol

(c) Tenis (d) Bütün kategoriler

Şekil 3: İlk sorgu ve daha sonraki 3 geribesleme döngüsü için RGB, HSV, LUV renk öznitelikleri ve Gabor doku öznitelikleri kullanılarak elde edilen sorgulama sonuçları. x-ekseni erişilen görüntü sayısını, y-ekseni ise duyarlılık (precision) değerlerini göstermektedir.

6. Kaynakça

[1] U.S. National Institute of Standards and Technology, "TREC video retrieval evaluation (TRECVID)," 2005. [Online]. Available: http://www-nlpir.nist.gov/projects/trecvid/

[2] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 1, pp. 39–62, March 1999.

[3] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, December 2000.

[4] F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "An efficient and effective region-based image retrieval framework," *IEEE Transactions on Image Processing*, vol. 13, no. 5, pp. 699–709, May 2004.

[5] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Segmentation, Description, and Retrieval of Video Content*, vol. 8, no. 5, pp. 644–655, September 1998.

Şekil 2: Bütün ground truth grupları için örnek görüntüler. Yukarıdan aşağıya: futbol, basketbol, tenis, golf ve Amerikan futbolu.

(a) İlk sorgu sonuçları (19 doğru)

(b) Birinci döngü sonuçları (30 doğru)

Şekil 4: Örnek bir futbol sorgusu sonuçları (ilk 30 görüntü).

(a) İlk sorgu sonuçları (14 doğru)

(b) Birinci döngü sonuçları (20 doğru)

Şekil 5: Örnek bir basketbol sorgusu sonuçları (ilk 30 görüntü).

[6] S. Aksoy, R. M. Haralick, F. A. Cheikh, and M. Gabbouj, "A weighted distance approach to relevance feedback," in *Proceedings of 15th IAPR International Conference on Pattern Recognition*, vol. IV, Barcelona, Spain, September 2000, pp. 812–815.

[7] Y. Rui and T. Huang, "Optimizing learning in image retrieval," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, Hilton Head Island, South Carolina, June 2000, pp. 236–243.

[8] T. S. Huang and Z. S. Zhou, "Image retrieval with relevance feedback: from heuristic weight adjustment to optimal learning methods," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, Thessaloniki, Greece, October 2001, pp. 2–5.

[9] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Systems*, vol. 8, pp. 536–544, 2003.

[10] S. Aksoy and O. Cavus, "A relevance feedback technique for multimodal retrieval of news videos," in *Proceedings of EUROCON*, Belgrade, Serbia & Montenegro, November 2005.