

# İçerik Tabanlı Görüntü Erişimi İçin Sahne Sınıflandırması

## Scene Classification for Content-Based Image Retrieval

Özge Çavuş, Selim Aksoy

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

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

### Özetçe

Son yıllarda çok geniş veri tabanlarının kullanımıyla birlikte içerik tabanlı görüntü indekslemesi ve erişimi önemli bir araştırma konusu haline almıştır. Bu çalışmada, görüntü indekslemesi için sahne sınıflandırmasını baz alan bir görüntü erişim sistemi tanımlanmıştır. Görüntülerden çıkarılan alt düzey öznitelikler görüntü indekslemesinde doğrudan kullanılmak yerine, bu öznitelikler sahne sınıflandırması için kullanılmış, ve görüntüler sınıflandırma sonucunda elde edilen anlamsal sınıf bilgileriyle indekslenmiştir. Sahne sınıflandırması için “kelime kümesi” (bag of words) doküman analizi yöntemi olarak bilinen tekniğin bir uyarlaması kullanılmıştır. Görüntü erişim sistemini insan algısıyla desteklemek ve anlambilimsel uçurumu en aza indirmek için erişim senaryosuna tek sınıf sınıflandırıcı bazlı ilgililik geri beslemesi eklenmiştir. Bunun için, ilgili görüntüleri çok iyi modelleyen, ilgili olmayan görüntülerden de bir o kadar uzak duran bir hiperküre oluşturan destek vektör veri tanımlaması kullanılmıştır. Önerilen yöntemler Corel veri kümesinde denenmiş ve başarılı sonuçlar elde edilmiştir.

### Abstract

Content-based image indexing and retrieval have become important research problems with the use of large databases in a wide range of areas. In this study, a content-based image retrieval system that is based on scene classification for image indexing is proposed. Instead of using low-level features directly, semantic class information that is obtained as a result of scene classification is used during indexing. The traditional “bag of words” approach is modified for classifying the scenes. In order to minimize the semantic gap, a relevance feedback approach that is based on one-class classification is also integrated. The support vector data description is used for learning during feedback iterations. The experiments using the Corel data set show good results for both classification and retrieval.

### 1. Giriş

Görüntü veri tabanlarının büyümesi ve yaygınlaşması ile birlikte içerik tabanlı görüntü indekslenmesi, çözümlenmesi ve erişimi önemli araştırma konuları haline almıştır. Günümüzde görüntü indekslemesinde kullanılan iki ana yöntem vardır. Bunlardan ilki ve geleneksel olanı görüntülerin içeriklerine göre elle verilen anahtar sözcüklerle ifade edilmesidir. Bu tür görüntülerin indekslenmesine yönelik olarak geliştirilen pek çok yöntem

olmasına rağmen, bu yaklaşımlar uğraştırıcı ve subjektif oldukları için başarılı sonuçlar vermemektedirler. Görüntü indekslemesinde kullanılan ve günümüzde çok popüler olan diğer bir yöntem ise alt düzey özniteliklerin kullanılmasıdır. Bu sayede görüntüler çeşitli öznitelik çıkarma algoritmaları kullanılarak elde edilen alt düzey özniteliklerle ifade edilir. Bu konuyla ilgili yapılan eski çalışmalar görüntünün bütününi baz alarak çıkarılan öznitelikler kullanmışlardır. Fakat görüntüyü sınırlı şekilde ifade edebilen bu öznitelikler ile görüntüdeki nesne çeşitliliği ve arka plan karışıklığı modellenememekte, ve bu nedenle de anlambilimsel uçurum olarak tanımlanan, görüntülerin alt düzey öznitelikleri ve üst düzey kavramları arasındaki mesafe artmaktadır. Günümüzde anlambilimsel uçurumu en aza indirmek için iki tür çözüm önerilmektedir. Birincisi görüntünün bütününi baz alarak çıkarılan öznitelikler yerine bölgesel tabanlı özniteliklerin kullanılması [1], bir diğeri ise görüntü erişimi senaryosu sırasında değişik öğrenme yöntemlerinin kullanılması ve bu sayede insan algısının da görüntü indekslemesine dolaylı olarak katılmasıdır [2, 3, 4].

Bu çalışmada, görüntülerden çıkarılan alt düzey öznitelikler görüntü indekslemesinde doğrudan kullanılmamıştır. Onun yerine bu öznitelikler sahne sınıflandırması için kullanılmış ve anlambilimsel uçurumu en aza indirmek amacıyla, görüntüler, sınıflandırma sonucunda elde edilen anlamsal sınıf bilgileriyle indekslenmiştir. Sahne sınıflandırması için görsel kelime kümesi doküman analizi yöntemi olarak bilinen yöntemin bir uyarlaması kullanılmıştır. Bu tür çalışmalarda dokümanlardaki kelimeler yerine görüntülerden değişik filtreleme yöntemleriyle elde edilen değişimsiz yerel betimleyiciler (visterm) kullanılmaktadır [5]. Fakat bu betimleyicilerin görsel çokanlamlılık denilen, bir betimleyicinin birden fazla nesneyi veya sahneyi tanımlaması probleminde yol açmalarından ötürü, bu çalışmada nesnelerin ve sahnelerin yerel özelliklerinin tanımlanması için bölütleme yöntemine başvurulmuştur. Bölütleme için görüntülerin renk ve doğrusal çizgi yapı özellikleri kullanılmıştır.

Sahne sınıflandırması ve içerik tabanlı görüntü erişimi çalışmalarında her ne kadar tatmin edici sonuçlara ulaşılsa da, her ikisi için de başarı ölçütü insan algısı bağımlıdır. Bu çalışmada, görüntü erişimi aşamasında ilgililik geri beslemesi yöntemlerini kullanan insan algısı destekli bir yaklaşım öne sürülmektedir. İlgililik geri beslemesi, kullanıcının kendisine dönen sonuçları ilgili veya ilgili değil olarak değerlendirmesi ve bu değerlendirme ışığında sistemin kullanıcıya yeni sonuçlar döndürmesini kapsamaktadır. İlgililik geri beslemesi alanında günümüzde pek çok çalışma mevcuttur. Bunlardan en tanınanları sorgu noktası hareketi [1, 4] ile özniteliklerin ağırlıklandırılması ve her sorgulamada bu ağırlıkların güncellenmesini [1] içermektedir. Bu

Bu çalışma TÜBİTAK 104E077 ve 105E065 numaralı projeler tarafından desteklenmiştir.

yöntemlerde karşılaşılan başlıca problem kullanıcının vermiş olduğu sınırlı sayıdaki geri besleme verisidir. Pek çok yöntem bu problemi sınıflandırma bazı tekniklerle çözmeye çalışmaktadır. En popüler yöntem destek vektör makinasının (Support Vector Machine) kullanılmasıdır [2]. Bu yöntemlerdeki sorun ise probleme iki sınıf sınıflandırması olarak yaklaşılmasıdır. Tahmin edileceği üzere, ilgili olarak etiketlenecek görüntülerin aynı sınıfa ait olacakları tartışmasızken, ilgili değil olarak etiketlenecek görüntülerin ait oldukları sınıflar farklılık gösterecektir. Bu nedenle, bu çalışmada, bu soruna tek sınıf sınıflandırması bazı çözümler aranmış ve tek sınıf veri tanımlama modeli olarak destek vektör veri tanımlaması (Support Vector Data Description) yöntemi kullanılmıştır.

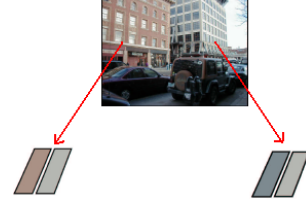
Bildirinin geri kalanı şu şekilde organize edilmiştir. 2. bölüm görüntülerin piksel bazlı, 3. bölüm ise doğrusal çizgi yapısı bazlı bölütlenmelerini anlatmaktadır. Kullanılan sahne sınıflandırması yöntemi ve görüntülerin indekslenmesi 4. bölümde yer almaktadır. 5. bölümde görüntü erişim senaryosunda kullanılan ilişilik geri beslemesi yöntemine yer verilmiştir. Son olarak 6. bölümde ise deneysel sonuçlar yer almaktadır.

## 2. Piksel Bazlı Bölütleme

Bu çalışmada görüntülerin bölütlenmesi için hem uzamsal hem de spektral bilgi kullanılmıştır. Uzamsal bilgi piksellerin  $x$  ve  $y$  koordinatlarını, spektral bilgi ise piksellerin HSV renk değerlerini içermekte olup, bu iki tür bilgi birleşik sınıflandırıcı yaklaşımı kullanılarak tümelştirilmiştir [6]. İlk aşamada bölütlenecek her görüntünün pikselleri için bir ilk etiketleme işlemi gerçekleştirilmiştir. Bu etiketleme işlemi için piksellerin spektral bilgileri üzerinde k-means öbekleme algoritması kullanılmış, etiketleme işlemi sonucunda her piksel bir  $t$  sınıfına dahil edilmiştir ( $t = 1, \dots, T$ ). Daha sonraki etiketleme aşaması için piksellerin hem spektral hem de uzamsal bilgileri kullanılmıştır. Bu aşamada etiketlenmiş piksellerin spektral bilgileri üzerinde en yakın ortalama sınıflandırıcısı, uzamsal bilgileri üzerinde Parzen sınıflandırıcısı eğitilmiştir. Aynı veri üzerinde, eğitilmiş sınıflandırıcılar çalıştırılarak her piksel için spektral,  $P_s(w_t|x_i)$ , ve uzamsal,  $P_u(w_t|x'_i)$ , olasılık değeri elde edilmiştir.  $x_i$ ,  $i$ . piksele ait HSV değerlerinin tutulduğu 3 boyutlu vektörü,  $x'_i$ ,  $i$ . piksele ait  $x$  ve  $y$  koordinatlarının tutulduğu 2 boyutlu vektörü,  $w_t$  ise  $t$ . sınıfı ifade etmektedir. Piksellere yeni sınıf etiketleri atamak için her iki olasılık değeri, çarpım katışım kuralı ile birleştirilmiş, ve birleştirilmiş olasılık değerini,  $P_b(w_t|x_i, x'_i)$ , enbüyüyen sınıfın etiketi  $i$ . piksele atanmıştır. Bu yeniden etiketleme işlemi en fazla 20 aşama olmak şartıyla pikseller kararlı bir etikete sahip oluncaya kadar devam ettirilmiştir.

## 3. Doğrusal Çizgi Yapısı Bazlı Bölütleme

Görüntülerdeki bazı nesneler sadece içerdikleri piksellerin spektral ve uzamsal özellikleri kullanılarak bölütlenemeyebilirler. Örneğin, binalar gibi insan üretimi nesneler, genellikle doğrusal parçaların birleşiminden oluşan bir yapı göstermektedir. Bu nedenle, renk özellikleriyle beraber, doğrusal çizgi yapılarından elde edilen özellikler de görüntülerin bölütlenmesinde kullanılabilir [7]. Aşağıda, doğrusal çizgi yapılarını kullanarak geliştirilmiş olduğumuz bölütleme yöntemi açıklanmaktadır.



Şekil 1: Farklı nesnelere ait iki doğrusal çizgi parçasının etrafındaki renk çiftleri ([7] numaralı kaynaktan alınmıştır).

### 3.1. Doğrusal Çizgi Parçalarının Özelliklerinin Çıkarılması

İlk olarak, Canny ayırt işleci kullanılarak görüntülerin ayırt haritaları oluşturulmuştur. Bu çıktılar üzerinde komşu ayırt piksellerinin oluşturduğu doğrusal yapılar belirlenerek görüntülerdeki doğrusal çizgi parçaları çıkarılmıştır [8]. Çıkarılan doğrusal çizgi parçaları, uç noktalarının koordinatlarıyla ifade edilmiştir.

Şekil 1'de görüldüğü üzere, insan yapımı bir nesneye ait doğrusal çizgi parçaları genellikle etraflarında biri sağından diğeri solundan gelen 2 ana renk içerirler ve bu renklerin aynı nesneden gelen doğrusal çizgi parçalarında ortak olması beklenir. Bu nedenle, insan yapımı olan nesneler için doğrusal çizgi parçalarının renk çifti bilgisi ayırt edici bir özelliktir, ve bu tip nesneler içeren görüntülerin bölütlenmesi için kullanılabilir. Bu özelliği kullanmak için, bu çalışmada, görüntülerden çıkarılan doğrusal çizgi parçalarına ait, sağ taraftaki ve sol taraftaki piksellerden ayrı ayrı olmak üzere ortalama RGB renk değerleri hesaplanmıştır. Her doğrusal çizgi parçası, uç noktalarına ait koordinatlarının yanı sıra, 3'ü sağ bölgeden 3'ü de sol bölgede olmak üzere 6 renk değeri ile ifade edilmiştir.

### 3.2. Doğrusal Çizgi Parçalarının Öbeklenmesi

Aynı nesneye ait doğrusal çizgi parçaları benzer renk ve konum özelliklerine sahip oldukları için, görüntüleri bölütleme amacıyla, her görüntüye ait doğrusal çizgi parçaları sırasıyla renk ve konum özelliklerine göre öbeklenmiştir. İlk öbekleme işlemi, doğrusal çizgi parçalarına ait 6 renk özelliği kullanılarak gerçekleştirilmiştir [7]. Her görüntüye ait doğrusal çizgi parçaları ortalama bağlı sıradüzensel öbekleme yöntemi (average link hierarchical clustering) kullanılarak en uygun sayıda öbeğe ayrılmıştır. Öbek sayısı sıradüzensel öbekleme için durdurma kuralı [9] kullanılarak elde edilmiştir.

Bazı görüntüler, ortak renk çiftlerine sahip farklı nesneler içerebilirler (aynı renklerde ve yan yana duran iki bina gibi). Bu tip problemleri elemek için ilk aşamada elde edilen çizgi öbekleri kendi içlerinde uç noktalarının konum bilgilerine göre tekrar öbeklenmiştir. Doğrusal çizgi parçalarının konum bilgilerine dayanan ikinci öbekleme işlemi iki aşamada gerçekleştirilmiştir. İlkinde, renk bazlı öbeklemede olduğu gibi ortalama bağlı sıradüzensel öbekleme yöntemi kullanılmıştır. Öbek sayısı, sıradüzensel öbekleme için durdurma kuralı [9] kullanılarak elde edilmiştir. Bu aşamanın sonunda bir nesneyi oluşturamayacak kadar az sayıda doğrusal çizgi parçasına sahip olan öbekler elenmiştir. Bazı öbekler, bir nesneyi oluşturabilecek kadar doğrusal çizgi parçasına sahip olmalarına rağmen içerdikleri çizgiler çok dağınık ve ayrıktır. Bu tip çizgi öbekleri biri nesnenin dışından diğeri içinden gelen ortak renk çifti değerlerine sahip olduklarından dolayı genellikle nesnelerin sınır bölgelerini oluşturmaktadırlar.

Bu tip öbekleri elemek için şöyle bir yaklaşım kullanılmıştır. Eğer bir çizgi öbeğini içine alan dışbükey zarf alanının o öbeğe ait çizgi sayısına oranı bir eşik değerinden büyükse, o öbek bir nesneyi ya da parçasını ifade etmek için uygun değildir. Diğer taraftan, bu oran bir eşik değerinden küçükse, o öbek bir nesneyi ya da parçasını ifade etmek için uygundur ve tekrar öbekleme işlemine tabi tutulmasına gerek yoktur. Bu iki eşik değeri arasında kalan orana sahip öbekler ise konum bazlı öbeklemede ikinci aşamaya geçeceklerdir. Bunun nedeni, bu tip öbeklerin hala tıkHz (compact) bir yapı sağlayamamış olmaları ve aykırı çizgiler içermeleridir. Bu nedenle, bu tip öbekler bu defa tek bağılı sıradüzensel öbekleme yöntemi (single linkage hierarchical clustering) ile kendi içlerinde tekrar öbeklenmişler ve bu sayede bu öbeklere ait ayrıncı çizgiler elenmiştir.

#### 4. Sahne Sınıflandırması

Bu çalışmada, “görsel kelime kümesi” tabanlı sahne sınıflandırması yöntemi [5] görüntü indekslemesi amacıyla kullanılmıştır. Biri piksel bazlı, diğeri doğrusal çizgi yapısı bazlı bölütleme sonucunda elde edilen iki farklı türde görüntü bölütleri oluşturulduktan sonra, her iki türdeki bölütler de farklı bölüt kod tablolarıyla ifade edilmektedir. Piksel bazlı bölütleme yöntemiyle (Bölüm 2) elde edilen bölütler, içerdikleri piksellerin HSV histogramlarıyla modellenmiş ve k-means öbekleme algoritması kullanılarak  $k_1$  adet bölüt türü içeren kod tablosu oluşturulmuştur. Doğrusal çizgi yapısı bazlı bölütleme yöntemiyle (Bölüm 3) elde edilen bölütler ise içerdikleri doğrusal çizgi parçalarına ait yönelim değerlerinin histogramıyla modellenmiş ve yine k-means öbekleme algoritması kullanılarak  $k_2$  adet bölüt türü içeren kod tablosu oluşturulmuştur. Bu iki tür bölütlerden elde edilen kod tabloları birleştirilerek  $k_1 + k_2$  bölüt türü içeren yeni bir kod tablosu tanımlanmıştır.

$k_1 + k_2$  bölüt türü içeren kod tablosunun oluşturulmasının ardından her görüntü için  $k_1 + k_2$  bölüt türünün histogramı hesaplanmıştır. Bu histogram ile modellenen her bir görüntü için, Bayesçi karar kuramı kullanılarak, bu görüntünün her bir sınıfa ait olma olasılığı hesaplanmış, ve bu olasılık değerlerinden oluşan yeni bir öznitelik vektörü ile görüntüler indekslenmiştir.

#### 5. İçerik Tabanlı Görüntü Erişimi ve İlgililik Geri Beslemesi

Bu çalışmada, olasılık değerleri ile indekslenen görüntüler ve bunlar arasındaki Öklid uzaklıkları kullanılarak bir görüntü erişimi sistemi gerçekleştirilmiştir. Bu sistemde tek sınıf sınıflandırması bazlı ilgililik geri beslemesi yöntemi kullanılarak insan algısı destekli bir yaklaşım öne sürülmüştür. İlgililik geri beslemesinde tek sınıf sınıflandırma yöntemi olarak destek vektör veri tanımlaması (SVDD) kullanılmıştır [10, 3]. SVDD, sınıflardan birinin iyi örneklenebildiği, diğlerinin çok daha dağınık olduğu durumlardaki sınıflandırmada iyi sonuçlar vermektedir [10]. SVDD, hedef sınıfı çok iyi modelleyen, ayırık sınıflardan da bir o kadar uzak duran bir hiperküreye tanımlar. Erişim senaryosunda kullanıcı ilk olarak arama yapmak istediği sınıf kategorisini seçer. Sınıflandırma işlemi sonucunda, aranan sınıf kategorisi ile otomatik olarak etiketlenmiş olan görüntülerden rastgele bir küme seçilir ve kullanıcıya sunulur. Kullanıcı kendisine sunulan görüntülerden birini sorgu görüntüsü olarak seçer

ve bu defa seçilen görüntüye en yakın görüntüler kullanıcıya döndürülür. Bundan sonraki sorgulamalarda kullanıcı kendisine dönen görüntüleri ilgili veya ilgili değil şeklinde etiketler. Kullanıcıdan alınan geri beslemenin ardından SVDD kullanılarak ilgili görüntüleri kapsayan ve onları ilgili olmayanlardan ayıran en uygun hiperküreye öğrenilir. İlgili görüntülerin bulunduğu bölgedeki görüntüler, hiperkürenin sınırlarına olan uzaklıkları bakımından büyükten küçüğe doğru sıralanır ve bu sıraya göre tekrar yeni bir geri besleme almak üzere kullanıcıya sunulur. Kullanıcı tatmin edici sonuçlar alınca kadar bu senaryo tekrarlanır.

#### 6. Deneysel Sonuçlar

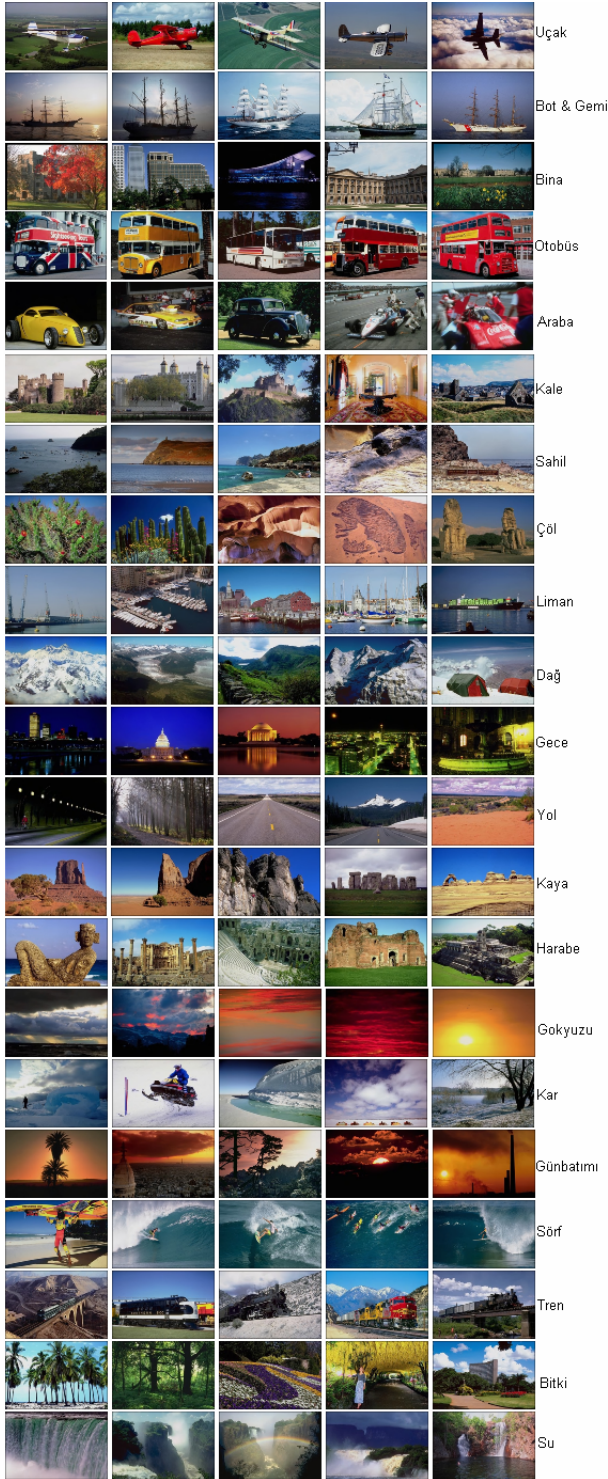
Bu çalışma Corel veritabanı üzerinde eğitilmiş ve sınanmıştır. 21 sınıf kategorisinden oluşan veri tabanı toplamda 4999 görüntü içermektedir. Her sınıftan rastgele seçilen toplam 3336 görüntü eğitim, geri kalan 1663 görüntü ise sınıma işlemleri için kullanılmıştır. Şekil 2 her sınıftan rastgele alınan örnek görüntüler içermektedir. Deneylerde, kod tablosu oluşturma aşamasında  $k_1$  ve  $k_2$  parametreleri için 1000 değeri kullanılmıştır, ve toplamda 2000 farklı bölüt türünü temsil eden bir kod tablosu oluşturulmuştur.

Sınıflandırma aşamasında sınıf kategorileri Bayesçi sınıflandırıcı ile modellenmiştir. Şekil 3’te sınıflandırma sonuçlarının yer aldığı hata matrisi gösterilmektedir. Daha önce bahsedildiği üzere, buradaki beklentimiz çok başarılı sınıflandırma sonuçları değildir. Şekil 2’de görüleceği ve sınıf kategori isimlerinden de anlaşılacağı üzere çok açıktır ki bir görüntüyü tek bir sınıfa dahil etmek mümkün olamayabilir. Örneğin bir dağ nesnesi, gökyüzü sahnesi ve su sahnesi aynı görüntü içinde yer alabilirler. Bu nedenle bu çalışmada, sınıfsal olasılık değerleri kullanılarak oluşturulan görüntü modelleri görüntü erişimi amacıyla kullanılmıştır.

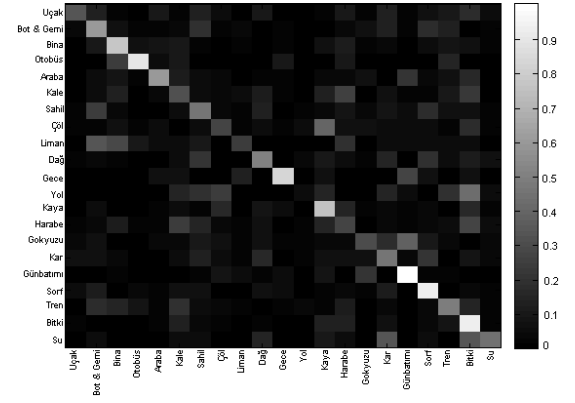
Sınıflandırma sonucunda elde edilen etiketler her sınıf kategorisinden otomatik olarak sorgular üretmek için kullanılmıştır. Veri kümesinden bilinen doğru sınıf etiketleri de 4999 görüntüyü her sorgu sonucuna göre sıralamak, ve her döngü sonucunda elde edilen ilk 30 görüntü içindeki her bir görüntüyü otomatik olarak, eğer sorgu görüntüsüyle aynı sınıf kategorisinde ise ilgili, geri kalanları ilgili değil olarak etiketleyerek geri besleme sağlamak için kullanılmıştır. Bu işlem, veri tabanında bulunan her görüntü için 4 defa geri besleme verilerek yapılmıştır. Şekil 4 her erişim döngüsü sonucunda elde edilen duyarlılık grafiklerini göstermektedir. En fazla artış ilk döngüde elde edilmiştir. Grafikten de görüleceği üzere geri bildirimle alınan sonuçlar geri bildirimsiz olanlara göre daima gelişme göstermiştir.

#### 7. Kaynakça

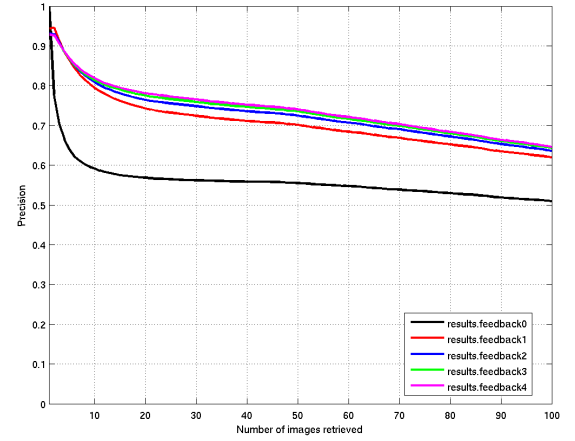
- [1] F. Jing, M. Li, L. Zhang, H. Zhang, and B. Zhang, “Learning in region-based image retrieval,” in *CIVR*, 2003.
- [2] G.-D. Guo, A. K. Jain, W.-Y. May, and H.-J. Zhang, “Learning similarity measure for natural image retrieval with relevance feedback,” in *CVPR*, 2001.
- [3] T. Onoda, H. Murata, and S. Yamada, “One class classification methods based non-relevance feedback document retrieval,” in *IEEE/WIC/ACM International Confe-*



Şekil 2: Corel veri tabanından örnek görüntüler.



Şekil 3: Sınıflandırma sonucunda elde edilen hata matrisi.



Şekil 4: İlk sorgu ve daha sonraki 4 geribesleme döngüsü için elde edilen sorgulama sonuçları. x eksenı erişilen görüntü sayısını, y eksenı ise duyarlılık (precision) değerlerini göstermektedir.

*rence on Web Intelligence and Intelligent Agent Technology*, 2006.

- [4] G. Giacinto and F. Roli, "Bayesian relevance feedback for content-based image retrieval," *Pattern Recognition*, vol. 37, no. 7, pp. 1499–1508, 2004.
- [5] M. Marszalek and C. Schmid, "Spatial weighting for bag-of-features," in *CVPR*, 2006.
- [6] P. Paclik, R. P. W. Duin, G. M. P. van Kempen, and R. Kohlus, "Segmentation of multi-spectral images using the combined classifier approach," *Image and Vision Computing*, vol. 21, no. 6, pp. 473–482, June 2003.
- [7] Y. Li and L. G. Shapiro, "Consistent line clusters for building recognition in CBIR," in *ICPR*, 2002.
- [8] A. Etemadi, "Robust segmentation of edge data," in *IEE Image Processing Conference*, 1992.
- [9] R. Mojena, "Hierarchical grouping methods and stopping rules: An evaluation," *The Computer Journal*, vol. 20, no. 4, pp. 359–363, 1977.
- [10] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.