

Rastlantısal Karar Ağaçlarıyla Nesne ve Renk Dağılımına Göre Sahne Sınıflandırılması

Scene Classification with Random Forests and Object and Color Distributions

Ahmet İşcen, Eren Gölge, Anıl Armağan, Pınar Duygulu
Bilgisayar Mühendisliği Bölümü
Bilkent Üniversitesi
Ankara, Türkiye
{ahmet.iscen, eren.golge, anil.armagan}@bilkent.edu.tr
duygulu@cs.bilkent.edu.tr

Özetçe—Bu çalışmada, görüntülerin içeriğindeki renkler venesnel olarak görüntüler için sahne sınıflandırması için bir yöntem sunulmuştur. Görüntülerdeki nesne ve renk içeriği göz önüne alınarak öznelik vektörleri oluşturulur. Bu öznelik vektörleri ile her bir sahne için, rastlantısal karar ağaçları kullanılarak, birer model elde edilir. Oluşturulan ağaç modelleri görüntünün sahnesinin sınıflandırılmasında kullanılmaktadır. Sunduğumuz yöntem, kategori bazlı sınıflandırmada Object Bank gibi en ileri yöntemlerle kıyaslanabilir sonuçlar vermektedir. Ayrıca, bilinen yöntemlerin hesaplama maliyeti görece yüksekken, sunduğumuz yöntem daha makul bir hesaplama maliyeti ile çalışır.

Anahtar Kelimeler — sahne sınıflandırması, rastlantısal karar ağaçları, bilgisayarlı görü, parça modelleri.

Abstract—We propose a method to recognize the scene of an image by finding the objects and the colors it contains. We approach this problem by creating a binary vector of detected objects and a histogram of the colors that the image contains. We then use these features to train a random forest classifier in order to determine the scene of each image. For class-based classifiers, our method gives comparable results with the state of art methods, such as Object Bank method, for the indoor scene dataset that we used. Additionally, while well-known methods are computationally expensive, our method has a low computational cost.

Keywords — scene recognition, random forests, computer vision, part based models

I. GİRİŞ

Dijital kameraların ve sosyal medyanın kullanımının

artması ile birlikte çok sayıda imge internet ortamında bulunmaktadır. Bu imgelerin etkin bir şekilde erişilebilmesi için, görüntülerdeki sahnelerin sınıflandırılması büyük önem kazanmıştır. Sahnelerin bulunabilmesi için verilen görüntünün anlamını ve içeriğini öğrenmek önemlidir. Ne yazık ki görüntüye ait etiket kelimeler veya metinsel açıklamalar gibi metinsel veriler gürültülü, yanlış ya da görüntü hakkında eksik bilgiler içeriyor olabilir. Bu nedenle metinsel veriler her zaman yeterli olmayabilir. Bu sorunu çözebilmek için görsel bilgileri çıkarmak gereklidir. Bu işlevi yapmak için birçok başarılı yöntem vardır. Bu yöntemler üç alt başlıkta sınıflandırılabilir: Renk, SIFT [2] ve HOG [4] gibi öznelikleri kullanan alt düzey yöntemler, bölütlenmiş bölgelerden çıkarılmış özellikleri kullanan orta düzey öznelik yöntemleri ve de son olarak nesnelere gibi yüksek seviye öznelikleri kullanan yöntemler, Nesne Bankası (Object Bank)[1] yöntemi vb. Bu çalışmada alt ve üst düzey yöntemleri birleştiren bir yöntem öneriyoruz.

Sunulan yöntemde ilk olarak belirli nesnelere verilen görüntüde var olup olmadığı kontrol edilir. Bu bilgiler ikili vektörlerde saklanılır. Nesnelere görüntüde bulunup bulunmamasına ek olarak, renk histogram seleleri bir görüntüyü temsil eden özneliklere eklenir. Çıkarılan nesne ve renk öznelikleri rastlantısal karar ağaçlarından çıkarılan modeller ile görüntü sahnelerinin sınıflandırılmasında kullanılmıştır.

II. İLGİLİ ÇALIŞMALAR

Bilgisayar ile görme alanında önemli bir yeri olan yöntemlerden biri SIFT[2] tanımlayıcıları yöntemidir. Bu yöntem görüntülerdeki yerel öznelikleri algılamak ve tanımlamak için sunulmuştur. Nesnelere yerel öznelikleri, o

nesnenin ilgi çekici noktaları olarak tanımlanmıştır. Daha sonra eğitim verilerinde verilen nesnelere ilgi çekici noktaları çıkartılarak her bir nesnenin ilgi çekici noktaları sınıma verilerinde o nesneyi algılamak için kullanılmaktadır. SIFT yöntemi kısmen örtülmüş veya ölçeği değiştirilmiş nesnelere algılanmasında başarılı bir yöntem olarak görülmüştür. Aynı zamanda bu yöntem bozulmuş ve aydınlatılması değiştirilmiş nesnelere algılanmasında da kısmen başarılı olarak görülmüştür.

Görüntülerde sahne sınıflandırmasında kullanılan bir diğer kabul edilmiş yöntem GIST[3] betimleyicileridir. GIST bütünsel betimleyicisi kullanılır. Bu yöntemde algısal boyutlar doğasallık, açıklık, pürüzlülük, yayılma vb. kullanılarak bir sahnenin baskın uzamsal biçimi gösterilir. Bu sayede bir sahnenin düşük boyutlu öznitelikleri çıkarılabilmektedir.

Sahne sınıflama problemiyle ilgili yakın zamanda önerilmiş bir çalışma olan Nesne Bankası (Object Bank) [1], nesnelere görüntüler üzerindeki varlıklarının ve pozisyonlarının sahnelere algılanmasında önemli olduğu varsayımından yola çıkar. Sahnelere üzerinde bulunan nesnelere uzamsal histogramlarını (Spatial Pyramid) [5] sahne görüntülerinin öznitelikleri olarak kullanan yöntem, daha sonrasında SVM ile sınıflandırma yapar. Fakat Object Bank sahne görüntülerinin öznitelik vektörlerinin büyüklüğü nedeniyle sonuç vermesi uzun zaman alan bir yöntemdir. Sunduğumuz yöntem ise kıyasla kısa bir öznitelik vektörü ile daha başarılı sonuç vermektedir. Üstelik çalışma süreleri kullanılan sınıflandırma algoritmasının SVM'e kıyasla daha süratli olmasıyla çok daha kısa sürmektedir.

III. ÖNERİLEN YÖNTEM

Sahne sınıflandırma yöntemimiz iki varsayımdan oluşur; bunlardan ilki, aynı sahnelerdeki imgelerin genel olarak aynı nesnelere içermesidir. Örnek olarak, oturma odası görüntüleri genellikle koltuk içerirken, mutfak görüntülerinde buzdolabı bulunması daha olasıdır. İkinci varsayım ise, aynı sahnelerdeki imgelerin genel olarak aynı renklerden oluşmasıdır. Buna da örnek olarak sera imgelerinin yeşil ağırlıklı olmasını, ya da havuz imgelerinin mavi ağırlıklı olmasını verebiliriz. Bu varsayımları göz önünde bulundurarak, imgelerdeki nesne ve renk dağılımları çıkartılıp öznitelik olarak kullanılmaktadır. Daha sonra bu öznitelikler, rastlantısal karar ağaçlarıyla beraber kullanılarak imgeler sınıflandırılmaktadır.

A. Nesne tespiti

Nesne tespiti, verilen bir model için, o modelin imgedeki varlığını belirtir. Nesne tespiti konusunda şu anda kullanılan en ileri tekniklerden bir tanesi Biçim Değiştirebilen Parça Modelleri [9] tekniğidir. Kısaca özetlemek gerekirse, BDPM tekniğinde her nesne modeli bir kök filtresinden, parça filtresi kümesinden, ve de herhangi bir parçanın çapa konumuna göre uzaklığını cezalandırmak için biçim değiştirme maliyetlerinden oluşur. Filtreleri yaratmak için HOG[4] öznitelikleri, öznitelikleri bulmak için ise çoklu ölçekli öznitelik piramidi kullanılır. Kayan pencere yöntemi kullanılarak, her penceredeki kök filtresi bir filtrenin öznitelik piramidindeki alt pencere ile nokta çarpımı, ve de aynı şekilde biçim filtreleri de, bir

filtrenin kök filtresine göre iki katı çözünürlükte HOG öznitelikleri ile nokta çarpımı hesaplanarak elde edilir. Her pencere için genel puan, kök ve parça filtre skorlarının toplamından, biçim değiştirme maliyetlerini çıkartılarak hesaplanır. Eğer bir pencere için olan skor belirlenen eşik geçerse, kullanılan modelin temsil ettiği nesne o konumda tespit edilir.

Biz BDPM modellerini kullanarak imgelerde nesnelere olup olmadıklarını bulduk. Bunun için, her bir nesne modelinin bir sütunu temsil ettiği ikili bir vektör kullanıp, görüntülerin nesne bilgilerini bu vektörlerin içinde kaydettik. Örnek olarak, eğer bir görüntüde "koltuk" nesnesi varsa, o nesneyi temsil eden vektör indisi 1 iken, eğer "koltuk" nesnesi yoksa, aynı indise sıfır değeri verilir.

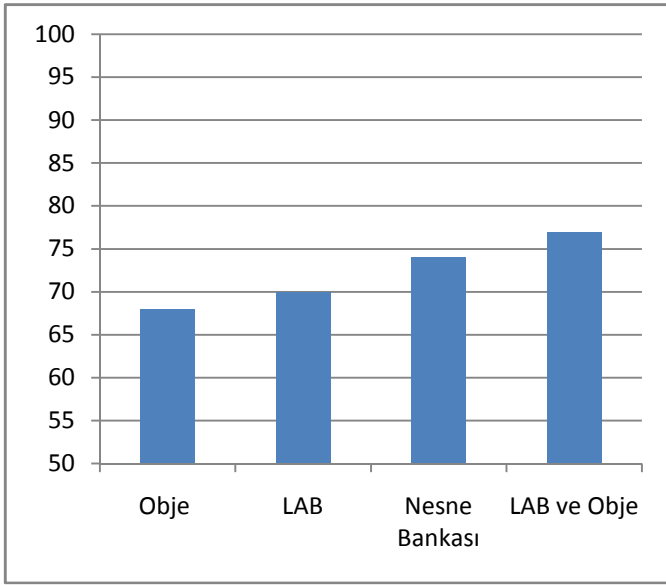
B. Renk tespiti

Diğer aşamada görüntülerdeki renk bilgisi derlenmektedir. Görüntülerin renk bilgilerini derlemek için iki farklı yöntem kullanılmaktadır; birincisi RGB renk uzayı, ikincisi ise LAB renk uzayıdır. RGB renk uzayı insanların renk algısına yakın olsa da, gerçek görüntülerdeki renk karşılaştırmasında pek başarılı olmamaktadır. LAB modeli ise RGB modeline göre insanın renk algısına daha yakın bir modeldir.

Biz literatürde daha sık kullanılan şekliyle A ve B boyutlarındaki değerleri değerlendirmenin renk bulma konusunda yeterli olacağına karar verdik. Renk değerlerini öznitelik olarak tutmak için, her bir imgedeki piksellerin RGB değerlerinin her bir boyutunu 4'er selelik histogramlara bölüp, onları küp olarak birleştirip, toplamda lineer olarak 64 sütunluk bir vektörde topladık. LAB'ın ise kullandığımız A ve B boyutlarını 8'er selelik histogramlara bölüp, aynı şekilde birleştirip, 64 sütunluk bir vektörde kaydettik. Bu seceler, belirli renklerin elimizdeki görüntüde oranını gösterir. Basit bir örnek olarak, havuz resmi düşünülebilir. Bu görüntünün renk histogramı çıkarıldığında havuzun rengi olan maviyi temsil eden seceler değeri diğer secelere göre daha yüksek olacaktır. Diğer renkleri temsil eden seceler yüzde dağılımına bakıldığında maviyi temsil eden secele göre düşük bir değere sahip olacaklardır. Son olarak, histogramlardaki değerlerin her birini görüntüdeki toplam piksel sayısına bölerek histogramları normalize ediyoruz.

C. Sınıflandırma

Sınıflandırma aşamasında her bir sahne için birer model elde edilir. Modellerin oluşturulmasında bir sınıfın tüm eğitim görüntüleri o sınıf için pozitif örnek olarak alınırken, aynı sayıda rastgele seçilmiş başka sınıflara ait görüntüler negatif örnekleri oluştururlar. Her bir görüntü öncesinde anlatıldığı gibi öznitelik vektörlerine çevrilerek temsil edilir. Derlenmiş olan öznitelik vektörleriyle sahnelerin sınıflandırılması, rastlantısal karar ağaçları (Random Forests)[6] kullanılarak yapılmaktadır. Karar ağaçlarıyla yapılacak sınıflandırma elimizdeki ikili öznitelik vektörlerinin yapısına da uyumlu görünmektedir ve bu varsayım yaptığımız deneyler sonucunda da diğer iyi bilinen sınıflandırma yöntemlerinden (SVM, Nöral Ağlar v.b.) iyi sonuç vermektedir. Rastlantısal karar ağaçları (RKA) için olarak sundukları genelleme yetisi ile karar ağaçlarının genel özellikleri olan, hızlı modelleme ve



Tablo 1: Sahne sınıflandırması için yöntemlerin karşılaştırılması.

sınıflandırma özellikleri gösterir. Bu özellikleri sayesinde bir çok sınıflandırma probleminde başarılı sonuçlar verir[7]. Özetle RKA tüm örnekler ve özneliklerden rastlantısal olarak seçilen örnek ve öznelik alt kümeleriyle farklı karar ağaçlarının modellenmesi ve bu modellerin verdiği oylama ile sınıflandırma yapar. RKA'nın alt uzaylardan yararlanması olması, elde edilen modellerin, bağımsız geçerlilik sınaması yapılmadan, ezber sıkıntısı yaşamasına engel olmaktadır. Uygulamamızda oylama yöntemi olarak, RKA ile yaratılan her bir ağaç sınıflandırma doğruluğu ile katsayılandırır. Bu şekilde görece daha doğru sonuç veren ağacın sonuca daha büyük katkı vermesi amaçlanmıştır.

IV. DENEYSEL SONUÇLAR VE TARTIŞMA

Deneylelerimizde veri kümesi olarak MIT Indoor Dataset¹ veri kümesini kullanıyoruz. Bu veri kümesi 67 tane kapalı alan kategorisi bulundurup, her kategoride en az 100 tane olmak üzere toplamda 15620 tane imge kapsar. Biz deneylelerimiz için her kategoriden en fazla 150, eğer kategorideki imge sayısı daha az ise kategorideki imge sayısı kadar rastgele imge seçip, daha sonra her bir kategori için bu imgelerin yüzde 70'ini eğitim kümesi, yüzde 30'unu da sınama kümesi olarak ayırdık.

Nesne tespitinin uygulamasında Object Bank projesi için önceden eğitilmiş olan 250 adet model kullanıldı. Bu modelleri kullanmamızın nedeni, kapalı mekan nesnelere bizim de kullandığımız MIT Indoor Dataset kullanılarak eğitilmesidir. Fakat, bu modellerin içinde açık hava veri kümesi kullanılarak eğitilen açık hava nesnelere de bulunması nedeniyle, ve de bizim çalışmamızda sadece kapalı mekan veri kümesi kullanıldığından dolayı, açık hava nesnelere kendimiz

¹<http://web.mit.edu/torralba/www/indoor.html>

eleterek sonuç olarak 169 adet kapalı alan nesne modelini kullandık. Daha sonra, her bir imge için bu nesnelere olup olmadıklarını 169 sütunluk ikili bir vektörde kaydettik. Son olarak, 169 sütunluk nesne vektörüyle 64 sütunluk renk vektörünü birleştirip her görüntü için toplam 233 boyutluk bir öznelik vektörü oluşturduk.

Rastlantısal karar ağaçlarıyla yapılan sınıflandırma işleminde RKA parametrelerinin doğru kalibrasyonu önem arz etmektedir. Parametreler değişik değerler ile yapılan deneylerin sonuçları ile belirlenmiştir. RKA parametreleri ile yapılan kalibrasyon denemeleri sonrasında en iyi sonuç tüm örnek kümesinin 0,8, tüm öznelik kümesinin de 0,6'sının alt küme olarak seçilmesiyle oluşturulan 500 farklı ağaç ile elde edilmiştir. Her bir sınıf ise öncesinde %70 eğitim görüntüleri, %30 test görüntüleri olarak ayrılmıştır.

Bu prosedürü yalnızca LAB renk öznelikleri, yalnızca obje öznelikleri, LAB ve obje özneliklerinin beraber kullanılarak deneysel sonuçlar elde ettik. (Tablo 1) Bu sonuçlar literatürde bilinen Nesne Bankası yöntemiyle üretilen özneliklerin sunduğumuz RKA sınıflandırmasıyla kıyaslanmıştır. Sonuçlarda görüldüğü üzere, daha kısa öznelik vektörleriyle, Nesne Bankasının uzun vektörlerine kıyaslanabilir sonuçlar elde edilmiştir. RKA ile kullanılan katsayısal oylama yöntemi de sonuçların doğruluğunda göreceli gelişim sağlar. Katsayısal oylama eşit ağırlıkla yapılan oylamadan, doğruluk oranını %73' den %77'e çekerek, daha iyi sonuç vermektedir. Ayrıca, sera ve havuz gibi belli renklerin etkin olduğu sahnelerdeki sınıflandırma yüzdesinin yüksek çıkması, renklerin nesnelere kullanıldığında sınıflandırmayı yükselteceğine dair olan varsayımımızı desteklemektedir. (Tablo 2)

V. ÖNGÜRÜLEN ÇALIŞMALAR

Sınıflandırma aşamasında renk ve nesne öznelikleri ayrı katmanlarda değerlendirilerek daha iyi sonuçlar elde edilebilir. İlk katmanda renk bilgisi kullanılarak elde edilen sınıflandırma bilgileri ikinci katmanda nesne öznelikleri dahil edilerek yapılan bir sınıflandırmayla nihai sonuç elde edilebilir. Bu sınıflandırma şeması daha iyi bir sonuç verebilir.

Görüntüler öznelik uzayında, ikili obje vektörü yerine bağlı öznelik yapısı (Relative Attributes) ile temsil edilebilir [8]. Alternatif olarak önerilen bu yöntem sınıflandırmada daha doğru sonuçlar verebilmesinin yanında, veriler için görülmemiş sahnelerin daha sonradan bağlı özneliklerine göre elimizdeki verilere eklenmesi durumunda onların da sınıflandırmasına da olanak sağlamaktadır.

TEŞEKKÜR

Bu çalışma TÜBİTAK 112E174 no'lu proje tarafından desteklenmektedir.

KAYNAKÇA

- [1] L. Li, H. Su, E. P. Xing, and L. Fei-fei, "Object Bank: A High-Level Image Representation for Scene Classification & Semantic Feature Sparsification," *Advances in Neural Information Processing Systems* 24 (2010).

[2] D. G. Lowe, "Object recognition from local scale-invariant features," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, no. [8, pp. 1150–1157 vol.2, 1999.

[3] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision* 42.3 (2001): 145–175.

[4] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, 2005.

[5] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, vol. 2, pp. 2169–2178.

[6] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.

[7] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," *Proceedings of the 23rd international conference on Machine learning - ICML '06*, pp. 161–168, 2006.

[8] D. Parikh and K. Grauman, "Relative attributes," *2011 International Conference on Computer Vision*, pp. 503–510, Nov. 2011.

[9] D. P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan "Object Detection with Discriminatively Trained Part Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1627 – 1645, Sept. 2010

bakery	0.71	gym	0.68	livingroom	0.76	staircase	0.81
auditorium	0.71	inside_subway	0.76	bathroom	0.86	studiomusic	0.72
bookstore	0.84	inside_bus	0.79	clothingstore	0.78	poolinside	0.83
bedroom	0.73	kitchen	0.71	grocerystore	0.77	mall	0.76
children_room	0.76	kindergarden	0.84	bowling	0.76	movietheater	0.75
casino	0.77	library	0.71	church_inside	0.86	nursery	0.67
buffet	0.83	laundromat	0.79	greenhouse	0.95	operating_room	0.75
closet	0.75	locker_room	0.68	deli	0.71	subway	0.73
cloister	0.81	lobby	0.72	office	0.79	toystore	0.83
corridor	0.81	classroom	0.62	meeting_room	0.69	restaurant	0.66
concert_hall	0.85	computerroom	0.82	prisoncell	0.83	trainstation	0.78
elevator	0.78	hospitalroom	0.85	airport_inside	0.79	tv_studio	0.73
dining_room	0.69	dentaloffice	0.82	restaurantkitchen	0.52	videostore	0.73
garage	0.56	laboratorywet	0.66	pantry	0.82	waitingroom	0.63
gameroom	0.76	fastfood_restaurant	0.76	museum	0.62	winecellar	0.80
florist	0.82	jewelleryshop	0.66	bar	0.79	warehouse	0.64
hairsalon	0.67	artstudio	0.61	shoeshop	0.64		

Tablo 2: Lab ve Obje öznelikleriyle kategorilere göre sahne sınıflandırma sonuçları