

AÇIKLAMALI GÖRÜNTÜ VERİTABANLARI KULLANARAK NESNE TANIMA VE ERİŞİMİ

Özge Can Özcanlı, Pınar Duygulu-Şahin, Fatoş T. Yarman-Vural

Bilgisayar Mühendisliği Bölümü
Orta Doğu Teknik Üniversitesi, Ankara
{ozge, duygulu, vural}@ceng.metu.edu.tr

Özetçe

Bu çalışmada görüntülerin otomatik olarak bölütlenmeleri sonucu elde edilen bölütlerin etiketlenmesi amaçlanmış ve ek açıklamalı (annotated) bir görüntü veri tabanına Kestirim En İyileme (KEİ) (Expectation Maximization) algoritması uygulanarak açıklamalar ile görüntü bölütleri arasındaki ilişki istatistiksel olarak modellenmiştir. Bu yöntemle verilen bir bölüt, en yüksek olasılığa sahip etiketle ilişkilendirilebilmektedir. Önerilen yöntemde performansın büyük ölçüde kullanılan özniteliklere bağlı olduğu gözlemlenmiş ve özniteliklerin birbirlerine göre önemlerinin analiz edilmesiyle, erişim ve tanıma performansının artırılması amaçlanmıştır.

1. Giriş

Gelişen teknoloji sayesinde, görüntü ve kelimeler bir arada bulunur hale gelmiştir. Web bunun en güzel örneklerinden biridir. Bir çok sayfa yazılarıyla çevrelenmiş resimler içermektedir. Optik Karakter Tanıma sayesinde resim içerisinde bulunan yazıları çıkarmak mümkün hale gelmiştir. Herşeyden öte hemen her resim anlamlı bir başlıkla tanımlanmıştır. Bütün bunların yanında özel olarak resimlerin kelimelerle elle ilintilendiği veri tabanları oldukça yaygındır. Örneğin; Corel görüntü veritabanında her resim birkaç kelimeyle tanımlanmıştır, bazı müze veri tabanlarında da resme ait özelliklerin ve bazen de gönüllüler tarafından hazırlanan açıklamaların metinleri bulunmaktadır.

Birlikte kullanıldıklarında ek açıklamalar, belirsizliği azaltmakta ve nesne tanıma yönelik klasik yöntemlerin dışında [1], yeni yaklaşımlar geliştirilmesini mümkün kılmaktadır.

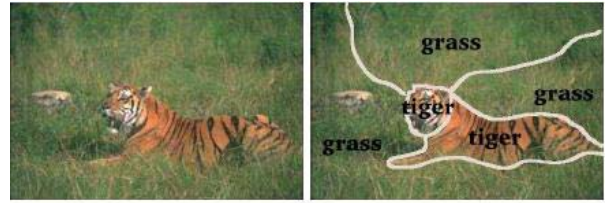
Bu çalışmada, ek açıklamalar kullanılarak görüntü bölütlerinin etiketlenmesiyle nesnelere tanınmasını öngören istatistiksel bir yöntem önerilmektedir. Ayrıca, etiketlenmiş bölütler kullanılarak, nesneye dayalı sorgular yapabilecek görüntü erişim sistemleri geliştirmek ve bu sistemlerin kullanıcıların ihtiyaçlarına daha fazla cevap verir hale gelmelerini sağlamak da mümkündür [3].

Bildirinin ikinci bölümünde önerilen nesne tanıma sistemi anlatılmakta, üçüncü bölümde kullanılan özniteliklerin birbirlerine göre önemleri incelenmekte ve son bölümde de çalışmanın sonuçları irdelenmektedir.

2. Bilgisayarla Çeviri Yöntemiyle Nesne Tanıma

Bahsedilen veri tabanlarında görüntüye ait kelimeler bilinmekte, ancak hangi kelimenin görüntüdeki hangi bölüte karşılık geldiği bilinmemektedir. Örneğin Şekil 1'de ek açıklaması -tiger-cat-grass- olan bir görüntü verilmiştir. Bu kelimelerin görüntüde olduğu bilinmemekte ama hangi bölütün -tiger- hangi bölütün -grass- olduğu bilinmemektedir. Önerilen yaklaşımda otomatik olarak elde edilen görüntü bölütlerine, açıklama kelimelerinin doğru bir şekilde yerleştirilmesi amaçlanmaktadır. Şekil 1'de görüntü bölütleri önerilen sistem kullanılarak kelimelerle eşlenmiştir.

Bölütlere eşlenik olan kelimelerin bulunması bölütlerden kelimelere çeviri şeklinde düşünülebilir. İşlemin bir dilin başka bir dile çevrilmesine benzerliği açıktır ve bir çeşit bilgisayarlı çeviri yöntemi olarak tanımlanabilir. Bu uygulamada çevrim yapılan dilin kelimeleri görüntü bölütleri, ikinci dilin sözlüğü ise veritabanındaki görüntülerin ek açıklamalarından çıkarılan belli sayıdaki kelime olarak düşünülebilir.

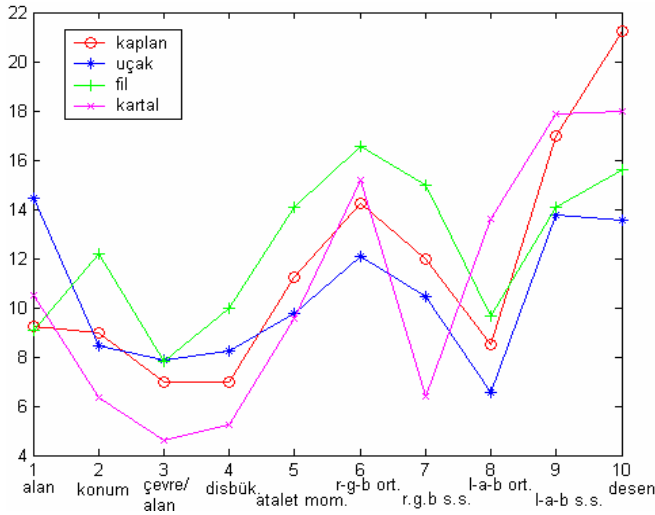


tiger cat grass

Şekil 1: Ek açıklaması -tiger-cat-grass- şeklindeki bir görüntü ve bölütlerinin etiketlenmiş hali. Ek açıklama kelimelerinin resimle eşleştiği bilinse de, hangi bölüte denk geldiği bilinmemektedir. Bu çalışmada amaç istatistiksel yöntemlerle bölütler ve kelimeler arasındaki ilişkileri öğrenerek, bölütlerin otomatik olarak etiketlenmesidir.

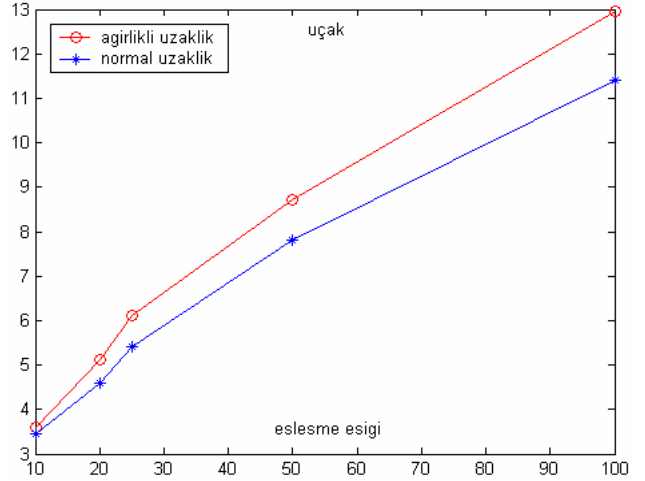
Bilgisayarla çeviri sistemleri, genelde bir dildeki kelimelerin başka dildeki kelimelere denk geldiği sözlükler (lexicon) tanımlamaya çalışmaktadır. Bizim bu çalışmada örnek aldığımız sistem Brown [6] tarafından önerilen örneklerden öğrenmeye dayalı istatistiksel bir yöntemdir. Sözlükler iki dilin birbirine paragraf düzeyinde eşleştirildiği metinlerden öğrenilir. Paragrafların denkliği bilinse de kelimeler arasındaki denklik bilinmemektedir. Yeterli sayıda veri olduğunda bu denklik problemini Kestirim En İyileme (KEİ) algoritmasını (Expectation Maximization) [7] kullanarak çözmek mümkündür. Bizim problemimizde de amaç, resim ve kelimelerin denkleştirildiği veri tabanlarından

dışbükeylik, atalet momenti, r-g-b ortalaması ve standart sapması, l-a-b ortalaması ve standart sapması ile ortalama yönelim enerjisi (doku) şeklinde olup toplam 30 elemanlı bir öznelik vektörü oluşturmaktadır. Elle etiketlenmiş görüntü veritabanında 301 bölüt ve kaplan, su, ağaçlar gibi toplam 14 sınıf (bölüt kategorisi) bulunmaktadır. Her bir öznelik grubunun sınıfların ayırt edilebilmesinde öneminin saptanması amacıyla görüntü erişimine dayalı bir değerlendirme yöntemi uygulanmıştır. Her sınıf ait tüm bölütlerle, her öznelik için ayrı ayrı sorgular yapılmış ve sınıf-öznelik ikilileri için ortalama kesinlik değerleri hesaplanmıştır. Bu kesinlik değerleri kullanılarak her sınıf-öznelik ikilisi için bir ağırlık değeri atanmıştır. Şekil 4 bazı sınıflar için özneliklerin yüzde olarak önemlerini (ağırlıklarını) bir grafikte göstermektedir.

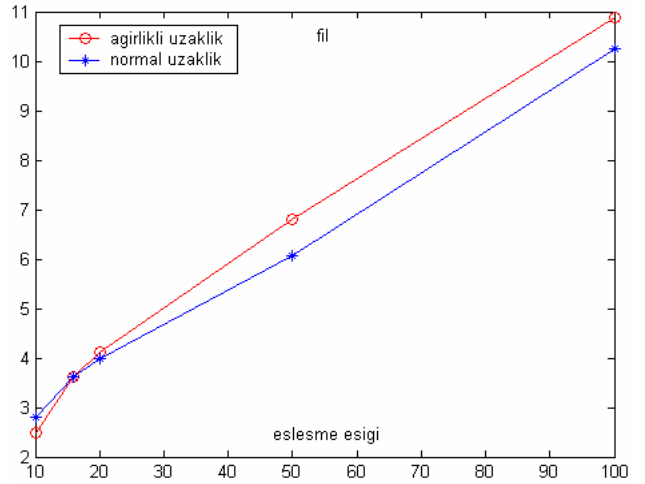


Şekil 4: Kaplan, uçak, fil ve kartal sınıfları için, öznelik uzayının 10 grubundan her birinin (metinde belirttikleri sıra ile) yüzde olarak önemi.

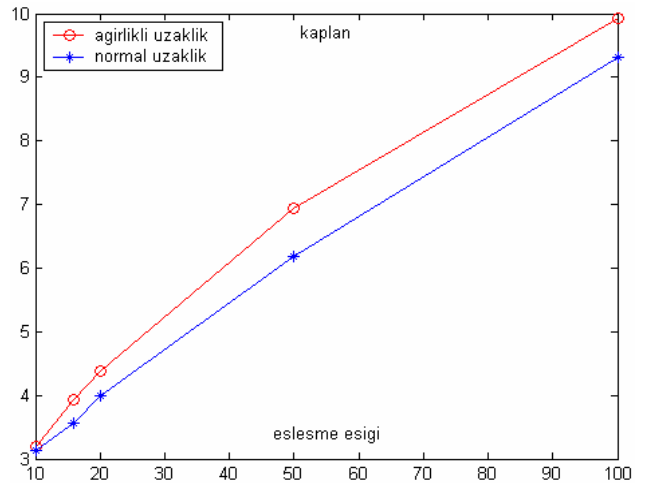
Aşağıdaki şekiller ise aynı sınıflar için öznelik vektörleri arasındaki uzaklıklar hesaplanırken, o sınıf için bulunan ağırlıkların kullanılmasıyla elde edilen erişim kesinliklerini gösteren grafiklerdir. Tüm grupların ağırlığı eşit ve bir olarak alındığında (normal uzaklıklarla) elde edilen kesinlikler de bu grafiklerde belirtilmiştir. Tüm grafiklerde yuvarlak noktalar ağırlıklı uzaklıkla, yıldızlı noktalar da normal uzaklıkla elde edilen değerleri göstermektedir.



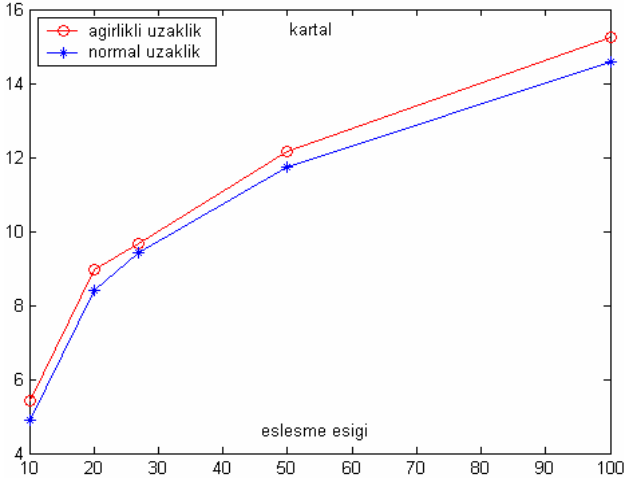
Şekil 5: Uçak sınıfının değişik eşiklerde normal ve ağırlıklı uzaklıklarla elde edilen erişim kesinlikleri.



Şekil 6: Fil sınıfının değişik eşiklerde normal ve ağırlıklı uzaklıklarla elde edilen erişim kesinlikleri.



Şekil 7: Kaplan sınıfının değişik eşiklerde normal ve ağırlıklı uzaklıklarla elde edilen erişim kesinlikleri.



Şekil 8: Kartal sınıfının değişik eşiklerde normal ve ağırlıklı uzaklıklarla elde edilen erişim kesinlikleri.

Grafiklerde görüldüğü gibi, öznelilikler sınıflar için ayırt edicilikleri oranında belirleyici olduklarında çok daha iyi erişim sonuçları elde edilmektedir. Topaklandırma için de benzer bir yaklaşım kullanılarak nesne tanıma yönteminin performansının artırılması mümkündür.

4. Sonuçlar

Bu çalışmada, bilgisayarla çeviri yöntemi kullanarak, ek açıklamalı bir veritabanında görüntü bölütleri ile belli bir kelime kümesi arasındaki ilişkinin istatistiksel olarak modellenmesi önerilmiş ve çıkarılan ilişkiler verilen bölütlerin etiketlenmesi için kullanılmıştır. Bu sayede yeni bir yaklaşımla nesne tanıma yapılması mümkün olmaktadır. Sistemin performansının artırılması amacıyla topaklandırma için kullanılan özneliliklerin göreceli önemleri analiz edilmiş ve özneliliklerin her sınıf için belirleyicikleri oranında değerlendirmelere katıldıkları takdirde topaklandırmanın iyileşeceği ve dolayısıyla etiketleme başarısının artacağı öngörülmüştür.

5. Kaynakça

- [1] Forsyth, D. A., and Ponce, J., *Computer Vision – A Modern Approach*, Prentice-Hall, 2002.
- [2] Shi J, Malik J., "Normalized cuts and image segmentation." *IEEE-PAMI*, 22(9):888-905, 2000.
- [3] Enser P. G. B., "Query analysis in a visual information retrieval context." *Journal of Documentation*, 51(2):126-170, 1995.
- [4] Barnard K., Duygulu P., Freitas N., Forsyth D. A., Blei D. and Jordan M., "Matching words and pictures." *Journal of Machine Learning Research*, in press.
- [5] Duygulu P., Barnard K., Freitas N., Forsyth D. A., "Object recognition as machine translation: learning a lexicon for a fixed image vocabulary." *Seventh European Conference on Computer Vision (ECCV)*, 4: 97-112, 2002.
- [6] Brown P. F., Della Pietra S. A., Della Pietra V. J., and Mercer R. L., "The mathematics of statistical machine translation: Parameter estimation." *Computational Linguistics*, 19(2):263-311, 1993.

- [7] Dempster A. P., Laird N. M., and Rubin D. B., "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society, Series B*, 1(39):1-38, 1977.
- [8] Duda R. O., Hart P. E., and Stork D. G., *Pattern Classification*, John Wiley and Sons Inc., 2001.
- [9] Corel Data Set, <http://www.corel.com/products/clipartandphotos>.