

YARI-OTOMATİK İMGE ETİKETLEME ARACI

SEMI AUTOMATIC ANNOTATION TOOL

Tuğçe Döngel¹, Hakan Çevikalp¹

1. Elektrik ve Elektronik Mühendisliği Bölümü
Makine ile Öğrenme ve Bilgisayarlı Görü Laboratuvarı,
Eskişehir Osmangazi Üniversitesi
tugce.dongel@hotmail.com, hakan.cevikalp@gmail.com

ÖZETÇE

İmge erişiminde kullanılan etiketli imgelerin etiketleme işlemi, tam-otomatik algoritmalar tarafından yapıldığında imge etiketleri genellikle hatalar ve eksik etiketler içermektedir. Bu etiketleri temizleme ve düzeltme işi genellikle manuel olarak kişiler tarafından yapılır. Farklı kişiler tarafından yapılan etiketleme işlemi de bu kişiler arasında koordinasyonu gerekli kılmaktadır. Bu bildiride, imge etiketleme için Matlab GUI ile tasarlanan yarı otomatik bir imge etiketleme aracı önerilmiştir. Önerilen bu çalışmada, ilk olarak verilen sorgu imgesinin ve etiketli imgelerin öznitelikleri karşılaştırılarak öznitelikler arasında “Chi-Squared” uzaklık ölçütü kullanılarak en kısa mesafe olanlar belirlenir. Daha sonra kullanılan uzaklık ölçütü göz önünde bulundurularak sorgu imgesine en çok benzeyen imgelerin etiket bilgileri kullanılarak, sorgu imgesinin etiketlenmesinde kullanılabilecek etiketlerden önem sırasına göre dizilmiş sıralı bir liste oluşturulur. Kullanıcı da oluşturulan bu etiket listesinden sorgu imgesiyle alakalı etiketleri seçerek etiketleme işlemi tamamlar. Listeyi döndürülen etiketlerin yeterli olmaması durumunda, geliştirdiğimiz araç aynı zamanda kullanıcıya liste dışında kendi etiketlerini de girmesine olanak vermektedir. Bu sayede önerilen yöntem ile hem zaman kaybı hem de insan algısının öznelliği en aza indirgenerek sistematik ve tutarlı etiketleme gerçekleştirilmiş olur.

ABSTRACT

When image labeling (annotation) process for image retrieval is performed by fully automatic algorithms, labels have noise, errors and missing labels. Correcting labels gathered automatically from web (using information around an image such as text, user-tags, etc.) is done manually using human labor. Therefore, coordination of different individuals is necessary for a consistent annotation. In this paper, a semi-automatic annotation tool designed with Matlab GUI has been proposed for an efficient and consistent image labeling. In the proposed framework, we first compare visual features of the query image and the labeled gallery images by using “Chi-Squared” distance. Then we create an ordered label list by using the labels of the closest images. The user finally selects the appropriate labels from the list and finishes the labeling process. The tool also allows one to enter new labels in case the returned labels are not enough to describe the image content. In this way the subjectivity of human perception and loss of time are reduced as well as consistency and coordination among different individuals’ annotations are accomplished.

1. GİRİŞ

Dijital fotoğraf çekme teknolojisinde son yıllardaki gelişmeler sonucunda hemen hemen her telefonda fotoğraf çekme standart bir hale gelmiştir. Buna paralel olarak hem dijital kameraların hem de telefonların fotoğraf çekme amacıyla kullanılmasını ve sosyal ağların fotoğraf ve video paylaşımına izin vermesi sonucu internet ortamındaki görsel imgelerin sayısında çok büyük bir artış yaşanmıştır. Örnek vermek gerekirse sadece Facebook ortamındaki resim sayısı 2010 yılında 10^{10} civarında olup bu sayı çok hızlı bir biçimde artmıştır. Buna Google, Flickr gibi siteler de eklendiğinde internet ortamındaki resim sayısı trilyonlarla ifade edilmektedir. Bu da beraberinde görsel olarak birbirine benzeyen imge ve videoların aranması gereğini ortaya çıkarmıştır.

İmge erişimi bu alanda yapılan birçok çalışmaya rağmen hala tam olarak çözülemeyen zorlu bir bilgisayarlı görü alanıdır. İmge erişimi bir sorgu imgesine benzeyen imgeleri birçok imge arasından bulup sıralı bir biçimde döndürmeyi içerir. İmge erişiminde en büyük problem imgenin görsel içeriğine bakarak imgeyi kendisini en iyi tasvir eden doğru anlambilimsel sözcüklerle (sınıflar) eşleştirebilmektir. İmgeleri etiketlemek için normalde imgede gözükken araba, ev, sahil gibi nesne kategorileri kullanılabildiği gibi nesne düzenleri, aktiviteler gibi bilgilerde kullanılarak daha zengin anlatımlar da kullanılabilir. Bazen de “doğa”, “sentetik”, ya da “doğal” gibi soyut kavramlar etiket olarak kullanılabilir.

Yakın bir zamana kadar imge erişimi genellikle sözcük (text) tabanlı olarak yapılmıştır. Fakat sözcük tabanlı imge erişimi bazı sorunlar doğurmuştur. İlk olarak internet ortamındaki tüm resimlerin doğru bir şekilde insanlar tarafından manuel olarak etiketlenmesi mümkün değildir. Bu sebeple imge etiketleri genellikle imgeyle beraber gelen metinden seçilen anahtar kelimelerden ya da imge dosya isimlerinden türetildiğinden bu etiketler gürültü ve hatalı etiketler de içermektedir. Bunun yanında bazı kelimelerin oldukça farklı birden fazla anlamları olabilir. Mesela arama motoruna “turkey” yazıldığında, arama motoru hem hindi resimleri hem de Türkiye’ye ait resimleri döndürmektedir. Özetlenecek olursa sadece sözcük bilgisi çoğu zaman bulunmak istenen imgelerin içeriğini tam olarak yansıtmamaktadır. Sözcük tabanlı yaklaşımlarda bir diğer sorun ise insan algısının öznellik sorunu olarak da bilinen görüntü anlambilimselliğinin öznelliğidir. Farklı koşullar altında farklı kişiler veya aynı kişi, aynı görsel içeriği farklı kavrayabilir. Buna insan algısının öznelliği denir [1]. Bu da imge erişimi için sorun oluşturabilir. Bu tür sorunların önüne

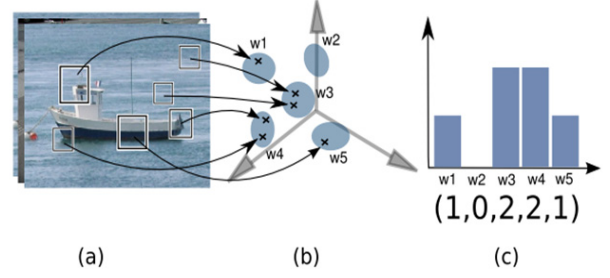
geçmek için sözcüklerle birlikte sorgu imgeleri ya da sadece sorgu imgeleri kullanılmaya başlanmıştır.

İmge erişimi için önerilen ilk yöntemlerde düşük seviye öznitelik çıkarma (low-level feature extraction) ve imge öznitelik vektörleri arasındaki benzerliği ölçmeye dayalı teknikler kullanılmıştır. Fakat düşük seviye öznitelik vektörleri arasındaki benzerlikleri kullanarak imgeleri doğru bir biçimde sınıflandırmak ya da imgenin içeriğini doğru bir şekilde anlamak mümkün değildir. Bu sorun anlambilimsel boşluk (semantic gap) problemi olarak da bilinir. Anlambilimsel boşluk probleminin çözümünü kolaylaştırmak için izlenecek yollardan biri tüm imgeleri etiketlemektir. Fakat imge sayısının çok fazla olması durumunda bu işlemi manuel olarak yapmak pratikte hem pahalı hem de zor bir iştir. Bu sebeple bu tür etiketleme işlemi yarı-otomatik ya da tam-otomatik algoritmalar tarafından yapılır. Buna bağlı olarak genellikle etiketlerin arasında hatalar ve gereksiz sözcükler vardır. Yarı-otomatik algoritmalar kullanılarak toplanan hatalı olabilecek etiketler kullanılmadan önce çözüm getirilmesi gereken önemli ve zorlu bir problem vardır: İçinde hatalar bulunan etiketli imgeler olabildiğince düzeltilmeli, yanlış ve gereksiz etiket bilgileri en aza indirgenmelidir. Etiketlerin temizlenmesinde insanlar kullanılmaktadır. Her bir imgenin etiketlerinin tek tek kontrol edilip düzeltilmesi pratikte hem zor hem de zaman alıcı bir iştir. Ayrıca etiketleme işinde çalışacak kişilerin elinde belirli bir kategori listesinin bulunmaması sebebiyle, farklı kişiler benzer resimleri farklı şekillerde etiketleyebilir. Örneğin “kedi” içeren bir resim etiketlenirken “kedi”, “kediler” gibi tekil ve çoğul kavramlar kullanılabilir ve bu da imge erişiminde sorunlara yol açabilir. Bu sorunun çözümü için bu bildiriye, uzman bir kişinin belli bir sayıda etiketleme yapmasıyla kategori listesi oluşturulmuş ve diğer kişilerin bu kategori listesini kullanarak daha pratik ve tutarlı bir şekilde etiketleme yapabilmesi için yarı otomatik imge etiketleme aracı tasarlanmıştır.

2. YÖNTEM

Yaptığımız çalışmada, İspanya'nın Valencia kentindeki “Universitat Politecnica de Valencia” kurumunda çalışan Prof. Roberto Paredes ve ekibinin “Overview of the ImageCLEF 2013 Scalable Concept Image Annotation Task” [2] projesi kapsamında “Relevant Image Search Engine (RISE) (<http://risenet.iti.upv.es/riase/>)” web sitesi için topladığı 30 milyon civarındaki imgeleri etiketlemek için kullandık. Bu imgelerin 250000 adedi tam otomatik algoritmalar tarafından imge etrafındaki metinler kullanılarak etiketlenmiştir ve bu etiketler arasında oldukça hatalı ve eksik etiketler mevcuttur. İlk olarak birbiriyle iletişim halinde çalışan 3 uzman kişi tarafından bu etiketler üzerinden geçilerek 2,5 ay içerisinde galeri setini oluşturacak etiketlerdeki hata oranı minimuma indirgenmiş 30.000 imgeden oluşan bir veri seti oluşturulmuştur.

Bir sonraki aşamada imgeler, “bag of words” modeli [3, 4, 5] ile betimlenmiştir. “Bag of words” modeli, her bir imgeyi sıralı temsili yamaların birleşimi olarak derlemektedir. Bu sebeple yöntem imgelerden belirli teknikler kullanılarak birçok yamanın seçimini gerektirir. Daha sonra bu yamalar SIFT gibi betimleyiciler kullanılarak sabit boyutlu yama vektörlerine dönüştürülür. Bir sonraki aşamada ise bu yama vektörlerinin nicemlenerek k adet temsili yama vektörlerinin betimlenmesidir. Son olarak Şekil 1’de görüldüğü gibi imgelerden gelen yamaların temsili yamalara olan



Şekil 1: Kelimeler çantası tabanlı imge betimleme. Her bir imgeden seçilen yamalar (a) betimleyiciler kullanılarak sabit boyutlu vektörlere dönüştürüldükten sonra her bir yamanın temsili yama vektörlerine olan uzaklıklarına bakılarak, yamalar temsili yamalara (şekilde $w_i, i = 1, \dots, 5$, olarak gösterilen 5 tane temsili yama vektörü vardır) atanır (b). Son olarak imge, atamaları yansıtan histogram vektörü ile temsil edilir (c).

uzaklıklarına bakılarak her bir yama bir temsili yamaya atanır ve imgeler bu atamalardan oluşturulan k boyutlu histogramlar olarak belirlenir. Çalışmamızda görsel sözlük boyutunu (visual vocabulary size) 5000 olarak belirlediğimizden her bir imge bu modelde 5000 boyutlu histogramlarla ifade edilmiştir.

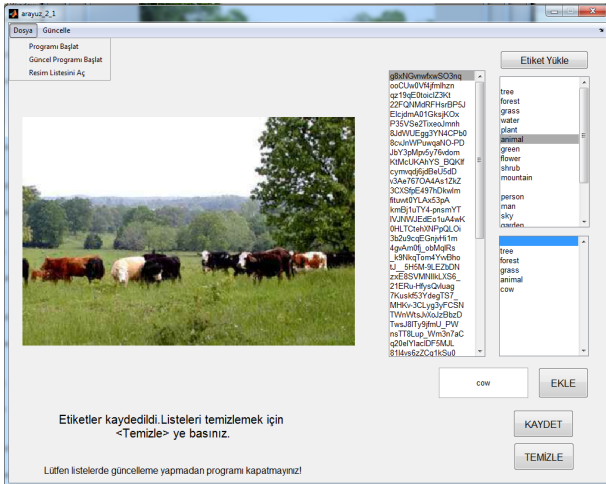
Tasarladığımız arayüzde etiketsiz bir sorgu imgesiyle manuel olarak etiketlenmiş sınırlı sayıda imgelemlerin 5000 boyutlu öznitelik vektörleri karşılaştırılmıştır. Karşılaştırma için iki imge histogramı arasındaki “Chi-Squared” uzaklıkları kullanılmıştır. Elimizde $I_1=(u_1, \dots, u_d)$ ve $I_2=(v_1, \dots, v_d)$ histogram vektörleri varsa, bu ikisi arasındaki Chi-Squared” uzaklığı aşağıda formül kullanılarak hesaplanır.

$$CSD(I_1, I_2) = \frac{1}{2} \sum_{m=1}^d \left[\frac{(u_m - v_m)^2}{u_m + v_m} \right] \quad (1)$$

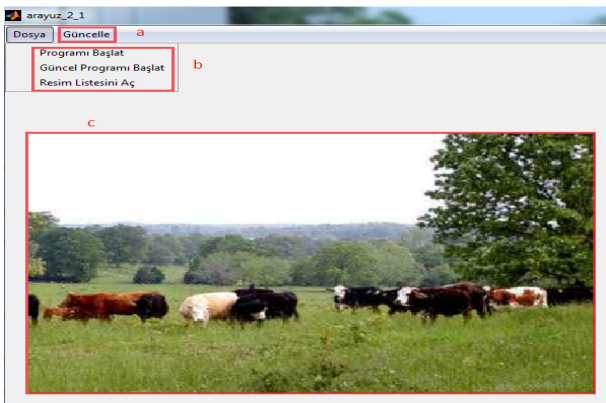
Öznitelikler, aralarındaki uzaklık minimum olandan maksimum olana doğru sıralanır ve aralarındaki uzaklık minimum olan 250 öznitelik vektörü ve bu vektörlere karşılık gelen imgeler belirlenir. Belirlenen bu özniteliklere karşılık gelen imgeler ve etiket bilgileri eğitim galerimizde mevcuttur. Bu etiket bilgileri okunarak her bir kategorinin kaç defa kullanıldığı sayılır ve kategoriler kullanım sayılarına göre çoktan aza doğru sıralanır. En çok kullanılan 20 kategori belirlenerek etiket listesi oluşturulur. Kullanıcı bu etiket listesinden sorgu imgesiyle alakalı olan etiketleri seçer ve sorgu imgesine uygun etiketlerin listelendiği bölüme ekler. Eğer bu etiketler dışında başka bir etiket eklenmek istenirse o da bu bölümdeki etiket listesine eklenebilir. Bu etiket listesi kaydedilerek imge etiketleme işlemi tamamlanmış olur.

Belirli bir sayıda etiketleme işlemi yapıldıktan sonra ya da program kapatılmadan önce etiketlenmiş imgelerin öznitelikleri ve etiketleri mevcut olan etiket bilgilerine eklenerek eğitim galerimizdeki etiket bilgilerinde güncelleme yapılır. Güncelleme yapılmasındaki amaç, sorgu imgesiyle daha fazla etiketli imge karşılaştırarak, sorgu imgesinin öznitelik vektörüne daha yakın öznitelik vektörleri bulmak ve dolayısıyla sonrası için daha doğru etiketlemeler yapmaktır. Tasarlanan arayüz Şekil 2’de gösterilmiştir. İlk olarak kullanıcı tarafından Şekil 3 (b)’de gösterilen menüden “Programı Başlat” sekmesine basılarak eğitim galerimizdeki etiket bilgileri yüklenir. Etiketli ve etiketsiz imgelerin özniteliklerinin bulunduğu dosyalar seçilerek öznitelikler yüklenir. Daha sonra etiketsiz imgelerin isimlerinin olduğu dosya seçilerek Şekil 4 (a)’da gösterilen imge isim listesi oluşturulur. İmge isim listesinden sırasıyla bir isim seçilir ve

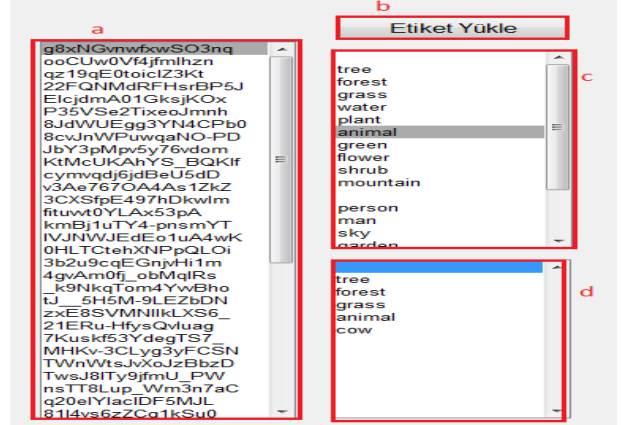
ardından Şekil 3 (c)'de gösterilen alanda imge görüntülenir. İmge görüntüledikten sonra Şekil 4 (b)'de gösterilen "Etiket Yükle" düğmesine basılır. Sorgu imgesi ve etiketli imge histogramları arasındaki uzaklığı minimum olan 250 öznitelik vektörü belirlenerek, bu vektörlerin etiket bilgileri okunur ve etiketler kullanım sayılarına göre çoktan aza doğru sıralanır. En çok kullanılan 20 kategori ismi Şekil 4 (c)'de gösterilen alanda listelenir. Bu listeden imgeyle alakalı etiketler seçilerek Şekil 4 (d)'de gösterilen alana yeni etiket listesi yazılır. Tüm bu işlemler yaklaşık 6.5 saniye sürmektedir. Eğer listelenen etiketler dışında bir etiket eklenmek istenirse eklenecek etiket, Şekil 5 (a)'da gösterilen metin kutusuna yazılır ve 'Ekle' düğmesine basılarak listeye eklenir. Etiket listesine etiket ekleme işlemi tamamlandıktan sonra Şekil 5 (b)'de gösterilen 'Kaydet' düğmesine basılarak hem imgenin ismi hem de etiket bilgileri excel dosyasına kaydedilir. Kaydedilme işleminden sonra Şekil 5 (c)'de gösterilen 'Temizle' düğmesi kullanılarak Şekil 4 (c) ve (d)'de gösterilen etiket listelerinin bulunduğu alanlar temizlenir ve böylece bir sorgu imgesi etiketlenmiş olur. Diğer sorgu imgeleri için de yukarıda anlatılan işlemler tekrarlanarak etiketleme işlemi yapılır. Belirli bir sayıda etiketleme işlemi yapıldıktan sonra ya da program kapatılmadan önce etiket bilgilerinin, imge isim listesinin ve özniteliklerin bulunduğu dosyaların güncellenmesi gerekmektedir. Şekil 3 (a)'da gösterilen 'Güncelle' düğmesine basılarak güncelleme işlemi yapılabilir.



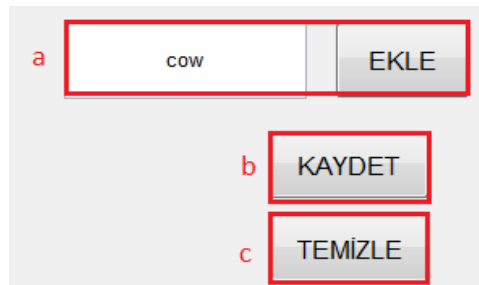
Şekil 2: Yarı otomatik imge etiketleme aracının görünümü.



Şekil 3: (a) "Güncelle" düğmesi. (b) "Programı Başlat", "Güncel Programı Başlat" ve "Resim Listesini Aç" sekmeleri. (c) Sorgu imgesinin gösterildiği alan.



Şekil 4: (a) İmge isim listesinin gösterildiği alan. (b) "Etiket Yükle" düğmesi. (c) Sorgu imgesini etiketlemede kullanılacak etiketlerin listelendiği alan. (d) Sorgu imgesine uygun etiketlerin listelendiği alan.



Şekil 5: (a) Eklenilecek diğer etiketlerin yazıldığı ve eklendiği alan. (b) Sorgu imgesine uygun etiketlerin kaydedilmesi için gerekli "Kaydet" düğmesi. (c) "Temizle" düğmesi.

3. DENEYSEL ÇALIŞMALAR

Yaptığımız arayüzün kullanılabilirliği hakkında fikir edinilmesi amacıyla, eğitim galerimizden rastgele seçilen her biri farklı kategorilere sahip 50 farklı imge üzerinde test edilmiştir. Bu amaçla kategorileri şüphe götürmeyen Şekil 6'da gösterilen imgelere benzer imgeler seçilmiştir.



Şekil 6: Test amaçlı seçilen bazı örnek imgeler.

Gerek galerimizde bulunan gerekse test olarak seçilen imgeler arasında büyük ölçüde ölçek, aydınlatma ve sınıf-ıçi değişimler bulunmaktadır. Ayrıca imgelerde arka plan son derece karmaşıktır ve bazı imgelerde nesne sınıfına ait örnek görüntüler başka nesnelere tarafından kapatılmıştır. Bu sebeple imgeleri betimlemek için daha önce belirtilen "bag of words" modeli [3, 4, 5] kullanılmıştır. Seçilen etiketsiz sorgu

imgesiyle etiketli imgelerin öznitelikleri “Chi-Squared” uzaklığı kullanılarak karşılaştırılmış, ardından sorgu imgesi ile etiketli imge histogramları arasındaki uzaklığı minimum olan 250 öznitelik vektörü belirlenir. Belirlenen öznitelik vektörlerinden ilk 20’sine ait imgeler sorgu imgesine en çok benzeyen imgeler olarak döndürülmüştür. Sistemin başarı performansı tamamıyla imge betimleme kullanılan uzaklık ölçütüne bağlıdır. Uzaklık ölçütü olarak “Chi-Squared” uzaklık ölçütünün seçilmesinin nedeni bu ölçütün imge histogramları için Öklid uzaklığına oranla çok daha iyi sonuçlar vermesidir [6,7]. Döndürülen 20 imge arasından sorgu imgesine benzer nesne içeren imgelerin sayısı bulunmuştur. Bunun yanında belirlenen öznitelik vektörlerine ait imgelerin etiket bilgileri kullanılarak en çok kullanılan kategoriler belirlenmiş ve sorgu imgesine ait etiketler bulunmuştur. Seçilen 50 farklı imge için bu işlemler tekrarlanarak toplamda 1000 imge döndürülmüştür. Döndürülen bu imgelerin 502 tanesi sorgu imgelerine benzer nesne içeren imgelerdir. Böylece bu arayüz kullanılarak %50,2 benzer imge döndürme başarısı elde edilmiştir. Döndürülen etiketler için kullanılan etiket sayısının kullanılmayan etiket sayısına oranı ise %15 civarındadır. İmge sayısı ve eğitim galerisindeki etiket bilgileri arttıkça, sorgu imgesine benzer imgelerin doğru olarak tanınma yüzdesi de artacaktır [6]. Şekil 7’de bir sorgu imgesine benzeyen imgelerin bulunmasına dair bir örnek verilmiştir.



Şekil 7: (a) Sorgu imgesi. (b) Sorgu imgesine benzeyen 20 imgeden ilk 3 tanesi.

4. TARTIŞMALAR

Yaptığımız bu çalışmada imgeler “bag of words” modeli ile betimlendikten sonra, imgelerin öznitelik vektörleri “Chi-Squared” uzaklık ölçütü kullanılarak karşılaştırılmıştır. Galerimizde etiketli imge sayısı 30000 civarındadır ve galeri setimizdeki etiketli imge sayısı arttıkça sistemin performansında artması beklenmektedir [8]. Fakat imge sayısı arttıkça sistemin hızı azalacağından, hızlandırma amacıyla imgelerin uygun Hash kodları kullanılarak kodlanması ve hızlı en yakın komşu algoritmalarının çalıştırılması gerekecektir. Ayrıca bugün bir çok başarılı imge erişimi algoritmasında “bag of words” modeliyle yada Hash kodları ile betimlenen imgeler döndürüldükten sonra, belirli sayıda imge seçip, bu imgeler ve sorgu imgesi arasındaki benzerlikler için, doğrudan her bir imgenin SIFT betimleyicileri yerel ve global kısıtlar kullanılarak karşılaştırma yapılmaktadır. Bu sayede çok daha başarılı

sonuçlar elde edilmiştir. Sistemimize yakın bir zamanda bu tür iyileştirmeleri eklemeyi de planlıyoruz.

5. SONUÇLAR

Bu çalışmamızda farklı kişilerin yaptığı imge etiketleme işlemine hız ve tutarlılık kazandırma amacıyla yarı otomatik bir imge etiketleme aracı Matlab GUI ile tasarlanmıştır.

Etiketsiz bir sorgu imgesiyle sınırlı sayıdaki etiketli imgelerin öznitelikleri karşılaştırılıp, “Chi-Squared” uzaklık ölçütü kullanılarak aralarındaki mesafe en az olanlar belirlenmiştir. Belirlenen özniteliklerin etiket bilgileri okunarak en çok kullanılan etiketler tespit edilmiş ve listelenmiştir. Kullanıcı tarafından bu etiketlerden sorgu imgesiyle alakalı olanlar seçilerek yeni etiket listesi oluşturulmuştur. Belli sayıda imge etiketlendikten sonra programın güncellenmesiyle etiket bilgilerinin daha verimli bir şekilde oluşturulması sağlanmıştır.

Tasarladığımız arayüz kullanılarak daha doğru ve daha hızlı bir imge etiketleme yapılmış olur. Kullanıcının etiketleri belli bir listeden seçmesi sebebiyle de insan algısı özneliği en aza indirgenmiş olur.

6. KAYNAKÇA

- [1] Y. Rui, T. S. Huang, and S. Mehrotra, “Relevance feedback techniques in interactive content-based image retrieval”, In Storage and Retrieval for Image and Video Databases (SPIE), pages 25–36, 1998.
- [2] M. Villegas, R. Paredes, “Overview of the ImageCLEF 2013 Scalable Concept Image Annotation Task”, In CLEF 2013 Evaluation Labs and Workshop, Online Working Notes. Valencia, (September 23-26 2013).
- [3] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, C. Bray, “Visual categorization with bag of keypoints,” ECCV Workshop on Statistical Learning for Computer vision, 2004.
- [4] H. Cevikalp, Z. Kurt, A. O. Onarcın, “The return of the king: The Fourier transform based descriptor for visual object classification,” IEEE Signal Processing and Communications Applications Conference, 2013.
- [5] S. Lazebnik, C. Schmid, J. Ponce, “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006.
- [6] H. Cevikalp, D. Larlus, F. Jurie, “A supervised clustering algorithm for the initialization of RBF neural networks classifiers,” IEEE Signal Processing and Communications Applications Conference, 2007.
- [7] H. Cevikalp, D. Larlus, M. Neamtu, B. Triggs, F. Jurie, “Manifold based local classifiers: Linear and nonlinear approaches,” Journal of Signal Processing Systems, vol. 61, pp. 61-73, 2010.
- [8] A. Torralba, R. Fergus, W. T. Freeman, “80 million tiny images: a large dataset for non-parametric object and scene recognition,” IEEE Trans. On PAMI, vol. 30 (11), pp. 1958-1970, 2008.