

Metin Tabanlı İmge Getirme için Derlem Kullanarak Etiket Listesi Oluşturma

Creating Image Tags for Text Based Image Retrieval Using Additional Corpora

Nermin Samet^{1,2}, Samet Hiçsönmez^{1,3}, Fadime Şener⁴

¹Elektronik Sistemler ve Uçuş Disiplinleri Grubu, TÜBİTAK SAGE, Ankara, Türkiye

²Bilgisayar Mühendisliği Bölümü, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye

³Bilgisayar Mühendisliği Bölümü, Hacettepe Üniversitesi, Ankara, Türkiye

{nermin.samet,samet.hicsonmez}@tubitak.gov.tr

⁴Department of Computer Science, University of Bonn, Bonn, Germany

{sener}@informatik.uni-bonn.de

Özetçe —İmge getirme yöntemlerinin başarıları temelde doğrudan görüntü etiketlerine bağlıdır. Bununla beraber görüntüler kullanıcılar tarafından gürültü dedğimiz yanlış etiketlerle veya hiç etiket olmadan Web ortamına aktarılabilir. Diğer yandan Web'deki her görüntü için bir etiket kümesi oluşturulması zaman ve maliyet açısından pek de mümkün olmamaktadır. Bu çalışmada Web'de mevcut olan ciddi boyutlardaki metin derleminden yararlanılarak imge getirme problemi iyileştirmeyi amaçlayan yeni ve özgün bir yöntem sunulmuştur. Önerilen yöntem bu iki sorunu çözmeye yönelik olarak iki türlü çalışmaktadır; 1) gürültü içeren etiket listesi verilen bir görüntünün etiket listesinden ilgili olmayan kelimelerin çıkartılması, 2) etiket listesi olmayan bir görüntü için uygun bir etiket listesi oluşturulması. Bu çerçevede İstatistiksel Eşleme Modeli ve Varyans Modeli olmak üzere görsel kelime kümeleri (bag of visual words) üzerinden çalışan iki model tasarlanmıştır. Elde edilen etiket temizleme ve oluşturma sonuçlarına göre, önerilen yöntem son derece başarılı imge etiketleri oluşturmaktadır.

Anahtar Kelimeler—*imge etiket yaratma, imge etiket kaldırma, imge getirme, metin işleme*

Özet—Image retrieval methods mainly depend on image tags. Nonetheless, images are uploaded to WEB with noisy image tags or not a single tag at all. On the other hand, generating tags for every image on the WEB is a time consuming job. In this paper, we present a novel method to improve image retrieval problem using additional text corpora which resides in WEB. Proposed method works in two modes, each of which tries to solve following problems separately; 1) cleaning irrelevant image tags from a noisy image tag list, 2) generating a tag list for an untagged image. In this manner, we designed two models which operates on bag of visual words namely Statistical Matching Model and Variance Model. Test results show that proposed methods are very successful in terms of both image tag cleaning and tag generation.

Keywords—*image tag creation, image tag removal, image retrieval, text processing*

I. GİRİŞ

Günümüzde kameralar, cep telefonları vb. donanımların gelişmesi ve çoğalmaları ile son yıllarda Web ortamında saklanan görsel verideki muazzam artışa paralel olarak görüntü işleme ile ilgili sorunlar artmış ve çeşitlenmiştir. Bununla birlikte, imge getirme problemi, uzun yıllardır bilgisayarlı görünümün çokça araştırılan önemli bir problemi olmuştur. İmge getirme, temelde anahtar kelimeler kullanılarak gerçekleştirilmektedir. Bunun en güzel örneği arama motorları üzerinden yaptığımız görüntü erişimleridir. Google, Bing, Yahoo vb. arama



Image Tags: sail,sails,sailing,boat,boating,yacht,sail_boat,water,moray,moray firth,balintore,shandwick,seaboard,andysmith,andrew smith

(a)



Image Tags: sailing,water,boats,boating

(b)

Şekil 1: Çalışmamızda kullandığımız sportif etkinlikler veri kümesindeki [1] Yelkencilik (sailing) sınıfından seçtiğimiz iki görüntü. (a)Yelkencilik için doğru bir görüntü fakat etiket listesinde çokca yanlış etiket bulunmakta, (b)Yelkencilik için negatif bir görüntü fakat pozitif olarak işaretlenmiş.

motorları görsel verilere erişmek için yaygın olarak kullanılmaktadır. Şirket ve kişisel arşivlerden görüntü sorgularının yapılması da imge getirmenin kullanılabilceği yerlere örnek olarak verilebilir.

İmge getirme yaygın olarak, kullanıcılar tarafından sağlanan etiketler üzerinden gerçekleştirilmekte ve ardından içerik tabanlı iyileştirmeler yapılabilmektedir. İmge getirmede başarılı sonuçların alınması, kullanıcıların etiketlemede kullandığı kelimelerin doğruluğuna doğrudan bağlıdır. Bu nedenle, imge getirmedeki en önemli gereksinimlerden birisi görüntülerin doğru kelimelerle etiketlenmesidir. Fakat kullanıcılar görüntüleri Web ortamına yüklerken yanlış kelimeler kullanarak etiketleyebilir ve hatta görüntülere hiç etiket sağlamayabilirler. Bunun yanında, veri tabanlarındaki görüntülerin yeniden ve doğru bir şekilde etiketlenmesi zaman almakla beraber pahalı bir işlemdir.

Metin tabanlı arama motorları, görüntüyü tanımlayan anahtar kelime listesinde yanlış etiketlenmiş, eş anlamlı, sesteş ya da üst anlamlı kelimeler bulunduğu kullanıcıya hatalı sonuçlar dönebilir. Şekil 1'de çalışmamızda kullandığımız veri kümesinden bazı yanlış etiketleme sonuçlarını sunmaktayız. Bu bağlamda, çalışmamızda arama motorlarının metin tabanlı imge getirme sonuçlarını iyileştirmek adına yeni ve özgün bir yöntem sunmaktayız. Literatürde var olan imge

getirme çalışmalarından farklı olarak, yöntemimizde veri kümesinden ayrı olarak tamamlayıcı bir derlem kullanılmaktadır. Son dönemlerde bilgisayarlı görüde, görüntü ve metin verisinin birbirine tamamlayıcı olarak kullanılması sıkça görülmektedir. Bu artışın motivasyon kaynağı ise problemlerin sadece görsel bilgi kullanılarak çözülemediği noktada, tamamlayıcı metinlerden gelen bilgiyi kullanmanın performansı artırabilmesidir. Bunun tersi bir durum olan metin bilgisinin yetmediği noktalarda görüntülerden tamamlayıcı bilgi elde edilmesi de muhtemel bir kullanım şeklidir. Örneğin bir metin içerisinde geçen "fare" kelimesinin, bilgisayara ait bir donanım parçası ya da hayvan olduğunu anlamak için ilgili görüntüye ait görsel bilgi kullanılabilir.

Metinsel ve görsel bilginin beraber kullanıldığı çalışmalardan esinlenerek, biz de bu çalışmamızda, tamamlayıcı metin bilgisini, imge getirme problemine yardımcı olabilecek bir adım olan etiket temizleme ve etiket oluşturma problemlerini çözmeye kullandık. Önerilen yöntem, görüntü etiketi temizleme ve görüntü etiketi oluşturma olmak üzere iki problemi çözmeyi amaçlamaktadır.

1) Görüntü etiketi temizleme: Buradaki temel amaç verilen bir görüntüye ait görüntü içeren anahtar kelime listesinden o görüntünün içeriği ile alakalı olmayan kelimelerin, kelime listesinden çıkartılarak doğru etiketlerin kalmasını sağlamaktır. Görüntü etiketi temizleme işlemi için ilinti tabanlı metotlar önerilmiştir.

2) Görüntü etiketi oluşturma: Anahtar kelime listesi mevcut olmayan görüntüler için doğru, hatasız bir etiket listesi oluşturmak bir başka önemli ve zorlu problemdir. Bu amaçla bir etiketleme modeli oluşturulmuştur. Önerilen yöntem verilen bir görüntü için istenilen uzunlukta bir anahtar kelime listesi oluşturmaktadır.

Çalışmamızda, bu problemlerin çözümü için İstatistiksel Eşleme Modeli ve Varyans Modeli olmak üzere iki farklı ilinti modeli önerilmiştir. Yapılan testler önerilen yöntemin umut vaat edici ve gelecekte iyileştirmeye açık olduğunu göstermiştir.

Bildirimiz şu kısımlardan oluşmaktadır; Kısım II'de ilgili çalışmalar anlatılmış ve Kısım III'te önerilen yöntem açıklanmıştır. Kısım IV'te testlerde kullanılan veri kümesi ve Kısım V'te yapılan detaylı testler sonuçlarıyla beraber sunulmuştur. Bildirinin VI. kısmında ise elde edilen sonuçlar ve gelecekte yapılacak çalışmalar hakkında bilgiler verilmiştir.

II. İLGİLİ ÇALIŞMALAR

Görüntü anlamayı daha açık hale getirmede metin bilgilerinin kullanılması son yılların popüler araştırma konularından birisidir [1], [2], [3], [4]. Bu alandaki ilk çalışmalardan bir tanesi Barnard ve arkadaşları [5] tarafından yapılmıştır. Yazarlar bu çalışmalarında metinsel ve görsel veriyi birbirlerine tamamlayıcı olarak kullanarak görüntü bölütleme problemi üzerine çalışmışlardır. Benzer şekilde, Zitnick ve arkadaşları [6] objelerin ve onlara ait metinsel tanımlayıcıların uzamsal ilişkilerini anlamaya yardımcı olan bir yöntem sunmuşlardır.

Yakın zamanda, Socher ve Fei-Fei [1] görüntüleri Web'de bulunan geniş metin derlemi ile birlikte kullanarak görüntülerde nesnelere bölütleyen bir yöntem sunmuşlardır. Diğerlerinden farklı olarak, Socher ve Fei-Fei [1] bu çalışmalarında hizalanmamış metin bilgisini tamamlayıcı olarak kullanmışlardır.

Metin verisinin görsel veriye tamamlayıcı olarak kullanılması, video işlemede de son zamanlarda yaygınlaşmıştır. Şener ve arkadaşları [7] görüntüler ve metinler üzerinde semantik çözümleme yaparak bir video özetleme metodu sunmuşlardır. Bu çalışmada, aynı anda hem görsel hem de metinsel olarak videolar arasında ortak

olan aktivite adımları gözetimsiz bir şekilde tespit edilmiştir. Alayrac ve arkadaşları [8] ise seslendirilmiş videolardan konuşma tanıyıcılar kullanarak metin bilgisini çıkarıp, ardından metin ve görüntü verisini bütünleşik bir kümeleme yöntemi kullanarak kümelemiş ve videolar-daki ana adımları bulmak için kullanmışlardır.

Bu çalışmamızda, [1] ve [2] çalışmalarından esinlenerek görüntü veri kümesi ve bu görüntü kümesindeki sınıflar ile ilişkili tamamlayıcı bir derlem kullanarak metin tabanlı görüntü erişimi sonuçlarını iyileştirmeyi amaçlayan bir yöntem sunmaktayız.

III. YÖNTEM

Çalışmamızda, etiket temizleme ve oluşturma problemlerimiz için ilk adım olarak veri kümesindeki görüntülerden SIFT [9] öznelikleri çıkarılmıştır. Daha sonra özneliklere k-ortalama (k-means) kümeleme algoritması uygulanarak her görüntü, görsel kelime kümeleri (bag of visual words) [10] ile ifade edilmiştir. Metinsel kelimeler ve dağılımları Socher ve Fei-Fei [1] tarafından sağlanan Newyork derleminden [11] alınmıştır. Ardından İstatistiksel Eşleme ve Varyans modelleri oluşturularak görüntü etiketler temizlenmiş ve veri kümesindeki her sınıfın test görüntüleri için etiketler oluşturulmuştur. Önerilen İstatistiksel Eşleme ve Varyans modellerinin performansı hesapladığımız dayanak model ile karşılaştırılmıştır.

A. Nitelik Çıkarma

SIFT ilgi noktaları [9] bilgisayarlı görmede sıkça kullanılan öznelik çıkarma yöntemlerinden biridir. İlgi noktaları şekil olarak değişime en büyük tepkiyi veren noktalar. Bu amaçla, ilgi noktaları bulunup, bu ilgi noktalarının öznelik vektörleri hesaplanmıştır. Ardından her görüntü görsel kelime kümeleri [10] ile ifade edilmiştir.

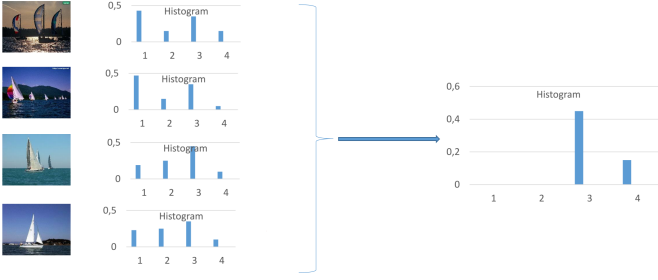
Buna göre, her görüntüye ait $[128 \times K]$ boyutunda SIFT ilgi noktaları matrisi çıkarılmıştır. Bütün görüntülerden alınan öznelikler k-ortalama kümeleme algoritması kullanılarak $k = 1000$ gruba kümelendirilmiştir. Bu kümelerin orta noktaları görsel kelimelere karşılık gelmektedir. Ardından, her görüntü için çıkarılan SIFT öznelik vektörleri ve hesaplanmış kelime noktaları arasında Öklid uzaklığı hesaplanıp, her öznelik vektörü kendisine en küçük uzaklığa sahip görsel kelime ile eşleştirilmiştir. Bu görsel kelime kümelerinin her görüntüye ait görülme sıklığı temel alınarak, o görüntüye ait $[1 \times 1000]$ boyutlu kelime kümesi dağılımı elde edilmiştir.

B. İstatistiksel Eşleme Modeli

$V = v_1, v_2, \dots, v_n$ veri kümemizdeki görsel kelimeleri, $T = t_1, t_2, \dots, t_m$ ise derlemdeki metinsel kelimeleri temsil etmektedir. Bu noktada modelimiz, derlemde bulunan t_j dışındaki kelimelere ait olan v_i görsel kelimesini reddeden koşullu olasılık değeri $P(v_i|t_j)$ 'yi kullanarak, eşleşmenin gerçekleştirilmesini sağlamaktadır. t_j 'ye özgü görsel kelimelerin seçilmesi v_i 'yi marjinalize ederek sağlanabilir.

$$P(v_i|t_j) = \sum_k \sum_l P(v_i|I_k, t_j)P(I_k|t_j)P(v_i|I_l, t_j)P(I_l|t_j), \quad (1)$$

Buna göre I_k ve I_l veri kümesindeki iki farklı görüntüye karşılık gelmektedir. Şekil 2 önerilen İstatistiksel Eşleme Modelini görsel olarak açıklamaktadır.



Şekil 2: Önerilen İstatistiksel Eşleme modelimiz. Derlemdeki her kelime için, örneğin *Yelkencilik*, görsel kelimelerin dağılımı tahmin edilerek bir dağılım hesaplanmıştır.

C. Varyans Modeli

Bu modelimizde varyans, görsel kelimelerin derlemdeki ilgili metinsel kelime baz alındığında ortalamadan ne kadar uzaklıktaki bir alana yayıldıkları bilgisini vermektedir. Diğer istatistiksel yaklaşımlarla karşılaştırıldığında varyans modeli matematiksel ve hesaplama karmaşıklığının sadeliği ile öne çıkmaktadır. Buna göre, bu modelimizde derlemdeki metinsel kelimelerin etiketi oldukları m adet görüntüye ait görsel kelimelerin $V = v_1, v_2, \dots, v_n$ her birinin varyans değerlerini hesaplanmıştır. Sonraki adımda ise ilgili metinsel kelime, varyans değeri en düşük olan ilk N görsel kelimenin dağılımı ile ifade edilmiştir.

$$Var_{v_i} = \sum_m \frac{(v_i - \mu_i)^2}{m} \quad (2)$$

D. Temsili Dağılım Hesaplanması

İstatistiksel Eşleme ve Varyans modellerimiz derlemdeki her kelime için görsel kelimeler arasından bir alt küme seçmektedir. Bu alt kümeleri elde ettikten sonra, derlemdeki her kelime için temsili bir dağılım hesaplamamız gerekmektedir. Bu amaçla görüntü veri kümemiz içinde, ilgili derlem kelimesi etiketine sahip bütün görüntüleri alıp, sadece seçilen görsel kelime alt kümesi üzerinden ortalama olarak, ilgili derlem kelimesi için temsili bir dağılım hesaplamaktayız. Daha sonra bu temsili dağılım ve her görüntüye ait görsel kelime dağılımları arasında Öklid uzaklığı hesaplayarak imge getirme yapmaktayız. Buna göre beklentimiz temsili dağılım ile arasında yüksek Öklid uzaklığı bulunan görüntülerin elenmesidir.

IV. VERİ KÜMESİ

Deneylerimizde Socher ve Fei-Fei tarafından sunulan Sportif Etkinlikler veri kümesini kullandık [1]. Bu veri kümesi içerisinde *Kürek*, *Badminton*, *Polo*, *Çim Bowlingi*, *Kar Kaykayı*, *Kroket*, *Yelkencilik* ve *Dağ Tırmanışı* olmak üzere sekiz farklı spor dalına ait görüntüler bulunmaktadır. Socher ve Fei-Fei [1], flicker.com sitesini araştırarak bu spor sınıfları ile ilgili resimleri toplamışlardır. Her sınıf ile ilgili metin derlemine elde etmek adına ise ilgili spor dalı hakkında bilgi veren New York Times makalelerini kullanmışlardır. Ancak *Çim Bowlingi*, *Polo*, *Kroket* ve *Dağ tırmanışı* hakkında yeterli sayıda makale olmadığı için, kendi çalışmalarında sadece *Badminton*, *Kürek*, *Yelkencilik* ve *Kar Kaykayı* sınıflarını kullanmışlardır. Benzer şekilde biz de çalışmamızda bu dört sınıfa ait veri kümesini kullandık. Kullandığımız veri kümesinde *Badminton* sınıfına ait 2016, *Kürek* sınıfına ait 991, *Yelkencilik* sınıfına ait 964 ve *Kar Kaykayı* sınıfına ait 1049 görüntü bulunmaktadır. Ayrıca veri kümesindeki her bir

görüntüye ait gürültü içeren etiket listesi bulunmaktadır. Şekil 1'de bu veri kümesindeki görüntü ve etiket çiftlerine örnekler bulunmaktadır.

V. PERFORMANS ANALİZİ

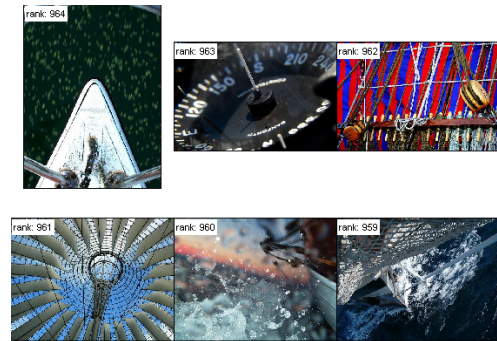
A. Gürültü Temizleme için Görsel Kelime Kümeleri Doğrulama

Bu deneyde amacımız görüntüler için hesapladığımız görsel kelime kümelerinin etkili olup olmadığını tespit etmektir. Veri kümesindeki dört sınıf için pozitif olarak işaretlenmiş birçok negatif görüntünün olduğunu gözlemledik. Şekil 1'de buna örnek göstermekteyiz. Örneğin, *Yelkencilik* sınıfına ait toplam 964 görüntüden 240 tanesi negatif görüntülerden oluşmaktadır. Bu aşamada veri kümemizdeki her sınıf için, görüntüleri modellerimizden elde ettiğimiz temsili dağılımları kullanarak sıraladık. Bu sıralamadan beklentimiz, her sınıftaki negatif görüntülerin sıralama listesinin sonlarında yer almasıdır.

Performans ölçümü yapabilmek için, veri kümesindeki tüm görüntüleri ilgili oldukları sınıfa ait olma ve olmama durumuna göre tekrardan etiketledik. Veri kümemizdeki bütün sınıflar için, İstatistiksel Eşleme ve Varyans Modellerimizi kullanarak temsili dağılım hesapladıktan sonra sıralama yaptık ve elde ettiğimiz bu sıralamayı ortalama kesinlik (Average Precision) kriterine göre değerlendirdik. Tablo 1'de dört sınıf için elde ettiğimiz ortalama kesinlik sonuçlarını sunmaktayız. İstatistiksel Eşleme Modelimizden elde ettiğimiz referans dağılımın Varyans modelinden daha başarılı olduğunu söyleyebiliriz. *Yelkencilik* ve *Badminton* sınıfları için model performanslarımızın diğer sınıflardan daha iyi olduğunu gözlemlemekteyiz. Bunun sebebi *Kürek* ve *Kar Kaykayı* sınıflarındaki pozitif etiketlenmiş negatif görüntü sayısının daha fazla olmasından kaynaklandığını düşünmekteyiz. Şekil 3'te ise görsel kelime kümelerinin başarısını görsel olarak görmekteyiz. Bu şekilde sıralama listesi tersten verilmiş ve negatif görüntülerin sıralanmış listede yoğun olarak son sıralarda getirildiği gözlemlenebilir.

Model	Görüntü Sınıfları			
	Badminton	Kürek	Yelkencilik	Kar Kaykayı
İstatistiksel Eşleme	0.739	0.3821	0.8109	0.3372
Varyans	0.6211	0.3700	0.7821	0.2659

Tablo 1: Veri kümemizdeki dört sınıf için, İstatistiksel Eşleme ve Varyans Modellerimizi kullanarak görüntü sıralama yaptık. Sonuçlarımızı ortalama kesinlik (Average Precision) kriterine göre sunmaktayız.



Şekil 3: Yelkencilik sınıfı için, İstatistiksel Eşleme Modelimizi kullanarak temsili dağılım hesapladıktan sonra sıralama yaptığımızda elde ettiğimiz görüntü sıralamasını tersten sunmaktayız.

B. Etiket Temizleme ve Oluşturma

1) *Dayanak Modelin Hesaplanması*: Önerdiğimiz yöntemler yeni ve özgün oldukları ve kullandığımız veri kümesi üzerinde benzeri bir çalışma mevcut olmadığı için yöntemimizin başarımını ölçmek için bir dayanak modeli hesapladık. Buna göre sadece derlemdeki kelimeleri kullanarak veri kümemizdeki görüntülerin etiketlerini temizledik. Eğer bir görüntüde derlemde bulunmayan herhangi bir kelime varsa bu kelime etiket listesinden kaldırıldı. Buradaki amacımız derlemin tek başına veri kümesini ne kadar iyi ifade ettiğini bulmak ve önerilen modellerimizi bu dayanak model ile karşılaştırmaktır.

2) *Etiket Temizleme*: Bu kısımda İstatiksel Eşleme ve Varyans modellerimizin etiket temizleme problemi için performanslarını ölçtük. Sonuçlarımızı Doğru-Pozitif Oranı ve Yanlış-Negatif Oranı kriterlerine göre değerlendirdik. Elde ettiğimiz sonuçlar Tablo II'de verilmiştir. Tabloyu incelediğimizde iki yöntemimizin de doğru görüntü etiketlerini tespit etmede başarılı olduğu görülmektedir (Doğru-Pozitif Oranı). Ancak sunulan modeller Yanlış-Negatif Oranı kriteri baz alındığında yüksek başarımlar vermemektedir. Bunun sebebinin, modellerimizin derlemdeki bazı kelimeleri temsil etmede yeterli olmamasına bağlıyoruz. Görsel kelime kümelerini değerlendirdiğimiz deneylerimize benzer şekilde, bu deneyde de yüksek performansların *Yelkencilik* ve *Badminton* sınıflarında ait olduklarını gözlemlemekteyiz.

		Görüntü Sınıfları			
		Badminton	Kürek	Yelkencilik	Kar Kaykayı
Doğru Poz. Or.	Dayanak	0.1898	0.1328	0.1109	0.1126
	İstatiksel Eşleme	0.2311	0.1590	0.1465	0.1155
Yanlış Neg. Or.	Varyans	0.2141	0.1378	0.1290	0.1021
	Dayanak	0.4236	0.2615	0.2966	0.3187
Yanlış Neg. Or.	İstatiksel Eşleme	0.4329	0.2943	0.2809	0.3211
	Varyans	0.4412	0.3019	0.3073	0.3457

Tablo II: İstatiksel Eşleme Modeli ve Varyans Modelinin etiket temizleme problemi için performans değerleri. Sonuçlarımızı Doğru-Pozitif Oranı ve Yanlış-Negatif Oranı kriterlerine göre değerlendirdik.

3) *Etiket Oluşturma*: Bu deneyimizde çalışmamızın bir diğer amacı olan etiket oluşturma sonuçlarımızı sunmaktayız. Buradaki amacımız New York derlemi içindeki her kelime için veri kümesindeki görüntüler kullanılarak tanımlayıcı görsel kelime kümeleri dağılımları hesaplamak ve yeni bir görüntü geldiğinde bu görüntüye otomatik etiket ataması yapabilmektir. Şekil 4'de Google arama motorundan indirilen test görüntüleri için İstatiksel Eşleme Modelimizin ürettiği en olası beş etiket verilmiştir. Etiketler incelendiğinde, modelimizin her sportif etkinlik için neredeyse hatasız ve resmi tanımlayıcı sonuçlar ürettiği görülmektedir.

VI. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada metin tabanlı imge getirme yönteminin sonuçlarını iyileştirmek adına özgün bir yöntem sunduk. Çalışmamızda Sportif Etkinlikler veri kümesini kullandık. Bu çalışmamızda literatürde var olan diğer çalışmalardan farklı olarak, veri kümemize tamamlayıcı olarak bir metin derlemi kullandık. Çalışmamızda, İstatiksel Eşleme Modeli ve Varyans Modeli olmak üzere iki yöntem sunduk. Deneylerimizin sonuçları incelendiğinde sunulan yöntemlerin görüntülerdeki gürültülü etiketleri elemeye ve yeni test görüntüleri için etiket üretmede başarılı olduğunu gözlemledik. Şu aşamada kullandığımız veri kümesinin yazarlarından aldığımız New York derlemi ile test ettiğimiz yöntemimiz için ileride daha geniş bir derlem ile kullanmayı



(a) Oluşturulan Etiketler: 'sport', 'badminton', 'action', 'shuttlecock', 'racket'



(b) Oluşturulan Etiketler: 'landscape', 'boat', 'sun', 'nature', 'reflection'



(c) Oluşturulan Etiketler: 'landscape', 'reflection', 'water', 'sailing', 'outdoor'



(d) Oluşturulan Etiketler: 'snowboarding', 'snowboard', 'cold', 'ski', 'nature'

Şekil 4: Örnek görüntüler için İstatiksel Eşleme Modelimizin ürettiği en olası beş etiketi sunmaktayız.

planlamaktayız. Ek olarak, görsel bilgiyi ifade ettiğimiz görsel kelime kümelerini farklı sözlük boyları ile denemeyi ve hatta literatürde yüksek performanslar gösteren Derin Öğrenme (Deep Learning) yaklaşımını uygulamayı düşünmekteyiz.

KAYNAKLAR

- [1] R. Socher and F. F. Li, "Connecting modalities: Semi-supervised segmentation and annotation of images using unaligned text corpora," *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*, pp. 966-973, 2010.
- [2] A. Karpathy and F. F. Li, "Deep visual-semantic alignments for generating image descriptions," *CoRR*, vol. abs/1412.2306, 2014.
- [3] C. L. Zitnick, D. Parikh, and L. Vanderwende, "Learning the visual interpretation of sentences," *IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013*, pp. 1681-1688, 2013.
- [4] S. K. Divvala, A. Farhadi, and C. Guestrin, "Learning everything about anything: Webly-supervised visual concept learning," *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, pp. 3270-3277, 2014.
- [5] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *J. Mach. Learn. Res.*, vol. 3, pp. 1107-1135, Mar. 2003.
- [6] C. L. Zitnick, D. Parikh, and L. Vanderwende, "Learning the visual interpretation of sentences," ser. ICCV '13. IEEE Computer Society, 2013, pp. 1681-1688.
- [7] O. Sener, A. R. Zamir, S. Savarese, and A. Saxena, "Unsupervised semantic parsing of video collections," *CoRR*, vol. abs/1506.08438, 2015.
- [8] J.-B. Alayrac, P. Bojanowski, N. Agrawal, J. Sivic, I. Laptev, and S. Lacoste-Julien, "Learning from narrated instruction videos," *arXiv preprint arXiv:1506.09215*, 2015.
- [9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [10] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering object categories in image collections," in *Proceedings of the International Conference on Computer Vision*, 2005.
- [11] L. D. C. E. Sandhaus., "The new york times annotated corpus.philadelphia."