

Görsel Nitelik Öğrenmede Alt-Düzye Özneliklerin Karşılaştırılması

A Comparison of Low-level Features for Visual Attribute Recognition

Emine Gül DANACI
Bilgisayar Mühendisliği Bölümü
Hacettepe Üniversitesi
Ankara, Türkiye
n12129719@cs.hacettepe.edu.tr

Nazlı İKİZLER CİNBIŞ
Bilgisayar Mühendisliği Bölümü
Hacettepe Üniversitesi
Ankara, Türkiye
nazli@cs.hacettepe.edu.tr

Özetçe —Görsel nitelik öğrenme ve kullanımı, son yıllarda bilgisayarlı görü alanında sıklıkla araştırılmaya başlanmış bir konudur. Bu çalışmamızda, görsel nitelik öğrenmeye, hangi alt düzey özneliklerin daha anlamlı ve verimli sonuçlar verdiğini araştırmayı amaçlamaktayız. Bu kapsamda, renk ve şekil bilgisini farklı detaylarda ele alan alt düzey özneliklerin, nitelik sınıflandırmaya katkısı araştırılmış, ve deneysel olarak değerlendirilmiştir. Elde edilen sonuçlar, özellikle renk ve yerel şekil bilgisini bir arada kullanan colorSIFT özneliklerinin ve yine yerel şekil bilgisini kodlayan yerel ikili örüntü (LBP) özneliklerinin görsel nitelik sınıflandırmada etkili olduğunu ortaya koymaktadır.

Anahtar Kelimeler—Görsel nitelikler, alt-düzye öznelikler, colorSIFT, LBP.

Abstract—Recently, visual attribute learning and usage have become a popular research topic of computer vision. In this work, we aim to explore which low-level features contribute to the modeling of the visual attributes the most. In this context, several low-level features that encode the color and shape information in various levels are explored and their contribution to the recognition of the attributes are evaluated experimentally. Experimental results demonstrate that, the colorSIFT features that encode local shape information together with color information and the LBP features that encode the local structure are both effective for visual attribute recognition.

Keywords—Visual attributes, low-level features, colorSIFT, LBP.

I. GİRİŞ

Görsel nitelik öğrenimi, bilgisayarlı görü literatürüne son yıllarda dahil olmaya başlamış bir konudur. Görsel nitelikler, insanlar tarafından anlamlandırılabilen orta-düzye görsel kavramlar olarak nitelendirilebilir, ve en önemli özelliklerinden bir tanesi, kategoriler arasında paylaşılabilir olmalarıdır. Nitelikler kenar, ilgi noktası betimleyicileri gibi düşük seviye özneliklerden daha yüksek, yüksek seviye kategorilerden de daha düşük özelliğe sahiptir. Temel olarak görsel nitelikler, kategorilerin içindeki daha küçük kavramları ifade eder. Örneğin,

"ağaç, insan, gökyüzü" gibi genel nesne tanıma problemindeki sınıf etiketleri yüksek kategoride bilgileri içermektedir. Bunun yanında "doğal, yeşil, uzun saçlı" gibi kavramlar ise görsel niteliklere denk gelmektedir. Burada görüldüğü gibi bu nitelikler insanlar tarafından da anlaşılabilir. Aynı zamanda makineler tarafından da keşfedilebilir özelliğe sahiptir. Genel olarak görsel nitelikler, herhangi bir sıfat, materyal ya da fonksiyon nitelik olarak görülebilir.

Görsel nitelik öğrenme kapsamındaki temel amaç, nesneyi çeşitli açılardan tanımlayan orta düzey kavramlara ait sınıflandırıcıların öğrenilmesi ve bu sınıflandırıcıların sonrasında, daha önce hiç örneği görülmemiş sınıfları dahi öğrenmeye (sıfır örneklili öğrenme, zero-shot learning) yardımcı olmasıdır. Nitelikler aktif öğrenmede nitelik sınıflandırılması, nitelik-nitelik ilişkilendirilmesi ve nesne-nitelik ilişkilendirilmesinde kullanılmaktadır. Bununla birlikte her türlü aramalarda da kullanılmaktadır. Örneğin kaybolan bir insanı ya da herhangi bir suçluyu aramak için, veya herhangi bir ürünü aramak için kullanılmaktadır. Bu tip aramalar genellikle bir nitelik ile ilgili sorgu ile başlamaktadır ve sorguya göre getirilen sonuçlar üzerine devam etmektedir.

Nitelik modelleri üç tiptir. İlk modelde ikili sonuçlar üretilir (*Binary Classifier Confidence*). Bu model bir fotoğrafta gördüğümüz kişi için gülümsüyor ya da gülümsemiyor şeklinde iki sonuca indirgediğimiz durumlarda kullanılır. Bir diğer model niteliğin ne kadar güçlü olduğu (*Attribute Strength*) hakkında sonuçlar üretir. Aynı fotoğraf için bu kişi %60 olasılıkla gülümsüyor dediğimiz, niteliğe ait kuvvet bilgisi gösteren durumlardır. Diğer bir model ise göreceli niteliklerin (*Relative Attributes*) olduğu ve nitelikler arasında karşılaştırmalı sonuçların üretildiği modeldir.

Bu çalışmada yukarıda belirtilen modellerden ilki olan ikili sonuçlar üreten nitelik modeli ele alınmıştır. Çalışmamızda amaç, nitelik tanımlamak ve bir resimde bu niteliğin olup olmadığını öğrenme aşamasında, verilen görsel bir nitelik için hangi alt-düzye özneliklerin daha kullanılabilir ve etkin olduğu araştırılmasıdır. Bu bağlamda, verilen görüntüler üzerinde, farklı düzeyde renk ve şekil bilgileri içeren alt-düzye öznelikler çıkarılmış ve bu özneliklerin farklı kombinasyonları

yonlarının görsel nitelik öğrenmedeki etkileri araştırılmıştır.

Görsel nitelik değerlendirme açısından sıklıkla kullanılan aPascal [2] denektaşı veri kümesi üzerinde yapılan deneyler, colorSIFT ve yerel ikili örüntüler (LBP) özniteliklerinin, doku öznitelikleri ile birlikte kullanılması durumunda görsel nitelik sınıflandırması için faydalı olacağı göstermektedir.

Bildirinin geri kalanı şu şekilde düzenlenmiştir: Bölüm 2'de literatürdeki ilgili çalışmalardan bahsedilmiştir. Daha sonra Bölüm 3'te görsel nitelik sınıflandırma için kullanılan yöntemin ve çıkartılan özniteliklerin detayları verilmiştir. Bölüm 4'te, aPascal veri kümesi üzerinde yapılmış olan deneylerin sonuçları sunulmaktadır. Son olarak, Bölüm 5 çalışma sonucunda elde edilen sonuçları özetlemektedir.

II. İLGİLİ ÇALIŞMALAR

Önceki yıllarda yapılan çalışmalarda görsel niteliklerin farklı sınıflandırıcılarla öğrenilmesi, bu görüntülerin niteliklere sahip olup olmadığı ile ilgilenilmiş ve eğer varsa konumunun bulunmasına dair çalışmalar yapılmıştır [1]. Bu makaledeki araştırmanın ana kaynağı olan [2] çalışmada ise önceki çalışmadan daha kompleks anlamsal ya da materyal niteliklerin denetimli şekilde öğrenilmesine yer verilmiştir. Görüntüler için tahmin edilen nitelikler ile hem nesne tanıma hem de alışılmadık nesnelerin betimlenmesi gerçekleştirilmiştir. Nitelikler üzerine yapılan diğer bir çalışmada [10] ise belirli tipteki nesnelere için niteliklerin otomatik olarak bulunması sağlanmıştır.

Daha önce belirtilen çalışmalarda niteliklerin bulunmasına yönelik çalışmalar yapılırken, varolan nitelik sınıflandırıcılarına benzeyen yeni sınıflandırıcıların da otomatik olarak çıkartıldığı çalışmalar [7] yapılmıştır. Bir diğer çalışmada [3] ise hangi niteliğin bir görüntüde daha baskın olduğu ve hangi niteliğin o görüntü için daha fazla bilgi verebileceği bulunmuştur.

Nitelik üzerine yapılan çalışmalardan bir diğeri [4] ise göreceli nitelikleri içermektedir. Göreceli nitelikler için bütün eğitim kümesi içindeki veriler nitelikler bazında sıralanmış ve her nitelik için bir sıralama fonksiyonu öğrenilmiştir. Bu şekilde verilerin birbirleriyle kıyaslanabilmeleri sağlanmıştır.

Görsel niteliklerin kullanıldığı bir çok uygulama vardır. Bir çalışmada [5] görüntülerin ilginçlik ve estetiksel yönü nitelikler yardımıyla bulunmaya çalışılmıştır. Diğer bir çalışmada [6] ise kişilerin ne kadar kızgın olduklarına karar vermede yine görsel nitelikler kullanılmıştır. Görsel niteliklerin en çok kullanıldığı alan ise arama çalışmalarıdır. Bir çalışmada [8] belirtilen özelliklere göre insanlar aranmıştır. Diğer bir çalışmada [9] ise istediğimiz ürüne göreceli nitelikler ile nasıl etkin şekilde arama yapabileceğimiz üzerine çalışılmıştır.

III. YÖNTEM

Bu çalışmada nitelik seçiminde alt-düzyer özniteliklerin etkisi ölçülmüştür. Bu kısımda, ilk olarak kullanılan öznitelikler anlatılacak, daha sonrasında görsel nitelik öğrenmek için kullanılan sınıflandırma yöntemi detaylandırılacaktır.

A. Kullanılan Öznitelikler

Bu çalışmada verilen görüntünün farklı düzeydeki bilgilerini modellemek için farklı alt düzey özniteliklerin performansları değerlendirilmiştir. Bu kapsamda, materyal, şekil

ve renk bilgilerini kodlayacak alt öznitelikler üzerinde çalışılmıştır. Materyal öznitelikleri olarak doku tabanlı öznitelik olan texton kullanılmıştır [15]. Renk öznitelikleri kullanılmıştır. Renk öznitelikleri için görüntünün CIE Lab renk uzayındaki değerleri alınmıştır.

Şekil bilgisini elde edebilmek amacı ile, öncelikle HOG (Histogram of Oriented Gradients) [14] tanımlayıcılar çıkartılmıştır. HOG özniteliklerine ek olarak, yerel şekil ve renk bilgisini birleştiren ColorSIFT (Color Scale-invariant feature transform) [13] tanımlayıcılar ile yerel ikili örüntü tanımlayıcıları (LBP - Local binary patterns) da alt-düzyer öznitelikler olarak kullanılmıştır.

Parça bilgisini kullanabilmek amacı ile, bu özniteliklerin hepsi öznitelik torbası (*bag-of-features*) kodlaması kullanılarak ifade edilmiştir. Öznitelik torbası yaklaşımı bütün öznitelikler için tek tek uygulanmıştır. Renk öznitelikleri pikseller üzerinde yoğun şekilde örneklendirilmiş ve k-ortalamar (*k-means*) kümeleme yöntemi kullanılarak 128 orta merkez noktasına göre nicemlendirilmiştir. Doku texton öznitelikleri ise her piksel için ayrı ayrı hesaplanmıştır. Bu texton öznitelikleri, k-ortalamar sonucu 256 merkezine göre nicemlendirilmiştir. HOG tanımlayıcılar için 8x8 bloklardan oluşan, 4 piksellik adım büyüklüğüne sahip HOG uzaysal piramidi oluşturulmuştur. Daha sonra bu uzaysal piramit için k-ortalamar $k = 1000$ değeri ile çalıştırılmış, ve elde edilen 1000 küme merkezine göre nicemlendirilmiştir.

ColorSIFT [13] tanımlayıcılar SIFT tanımlayıcılarda göz ardı edilen renk bilgisini de SIFT tanımlayıcılara eklenmesiyle oluştururlar. ColorSIFT tanımlayıcılar için kullanılan yöntem de HOG tanımlayıcılar için kullanılan yöntem ile aynıdır ve 1000 k-ortalamar merkezine göre nicemlendirilmiştir. LBP tanımlayıcılar için de benzer şekilde 100 kmeans merkezine göre nicemlendirilerek öznitelikler oluşturulmuştur.

B. Sınıflandırma Yöntemi

Yukarıda anlatılan alt-düzyer öznitelikler üzerinde görsel nitelik sınıflandırıcıları oluşturma amacı ile Farhadi ve diğerlerinin uyguladığı yöntem [2] adapte edilmiştir. Bu yöntemde, eğitim aşamasında, eğitim veri kümesi üzerinde çıkarılan özniteliklerin her bir görsel öznitelik için sınırlayıcı kutular (bounding box) içinde bulunan öznitelikler birleştirilmiştir.

Öğrenme aşamasında nitelik sınıflandırıcıları için yararlı olan özniteliklerin seçimi yapılmıştır. Bu işlem için özniteliklerin küçük kümelerine sıfır dışı değerler verdiği için L1 normuyla düzenlenmiş lojistik regresyon kullanılmıştır [11]. Bu kısımda nesnelere ayırt edici özelliğe sahip olabilecek niteliklerin öğrenilmesi amaçlanmış ve bu özelliğe sahip olan öznitelikler seçilmiştir. Sonrasında seçilen bu öznitelikler üzerinden görsel nitelik sınıflandırıcıları lineer Destek Vektör Makineleri (DVM) kullanılarak öğrenilmiştir. Bu çalışmada toplam 64 tane nitelik sınıflandırıcısı kullanılmıştır. Bu sınıflandırıcı modellerinin oluşturulmasından sonra, her bir görüntü için bu sınıflandırıcılar çalıştırılmış ve hangi niteliklerin bu görüntüye ait olduğu araştırılmıştır.

IV. DENEYLER

A. Veri Kümesi

Deneylerde, görsel nitelik öğrenme için önemli denektaşı kümelerinden biri olan aPascal [2] veri kümesi kullanılmıştır.



Şekil 1: Veri kümesinden örnekler ve nitelikleri

Bu veri kümesi Pascal VOC 2008 veri kümesi kullanılarak nesne sınıflandırılması ve tanınması çalışmaları için de kullanılabilir şekilde oluşturulmuştur. İçeriğinde "insan, kuş, kedi, inek, köpek, at, koyun, uçak, bisiklet, bot, otobüs, araba, tren, şişe, yemek masası, kanepeler ve televizyon" gibi nesne kategorileri bulunmaktadır. Her nesne kategorisinde yaklaşık 150 ila 1000 adet nesne örneği bulunmaktadır ve toplamda 4340 adet görüntü vardır. aPascal veri kümesindeki görüntülerin eğitim ve deneme kümelerine ayrımı orijinal veri kümesi tanıtımında verildiği şekilde, 2113 görüntü eğitim için, 2227 görüntü ise test için kullanılmıştır.

Bu veri kümesi üzerinde, anlamsal görsel nitelikler için etiketlendirme Amazon Mechanical Turk kullanıcılarından sağlanmıştır. Bu kapsamda, toplamda 64 tane nitelik tanımlanmıştır. Bu 64 görsel niteliğin her biri için sınıflandırma modelleri oluşturulmuştur. Bahsi geçen nitelikler arasında şekil bilgisine dayalı 2D ve 3D gibi nitelikler, parça bilgisi olarak kola sahip, bacağa sahip, başa sahip gibi nitelikler ve materyal bilgisi olarak ağaçtan yapılma, camdan yapılma gibi görsel nitelikler bulunmaktadır. Veri kümesinden örnek resimler ve bu resimlerde bulunan görsel nitelikler Şekil 1'de gösterilmektedir.

B. Deney Sonuçları

aPascal veri kümesindeki deneyler kapsamında, önceki bölümlerde bahsi geçen öznelikler ve onların olası birleşim vektörleri üzerinden nitelik öğrenme performansı değerlendirilmiştir. Öznelik vektörlerinin değerlendirilmiş olan birleşimleri aşağıda verilmiştir.

- 1) Sadece renk özneliği
- 2) Sadece HOG tanımlayıcılar
- 3) Sadece texton özneliği
- 4) Sadece ColorSift tanımlayıcılar
- 5) ColorSift ve texton özneliği
- 6) ColorSift, HOG ve texton özneliği
- 7) HOG, texton ve renk özneliği
- 8) ColorSift, HOG ve renk özneliği
- 9) Sadece LBP özneliği
- 10) LBP, HOG ve renk özneliği
- 11) LBP, HOG ve ColorSift özneliği
- 12) LBP, HOG, ColorSift ve texton özneliği

Deneyler kapsamında, öznelik vektörlerinin birleştirilmesi erken birleştirme (early fusion) yöntemi ile art arda eklenme

Tablo I: Farklı alt düzey öznelik kullanımlarının görsel nitelik sınıflandırmadaki doğruluk değerleri

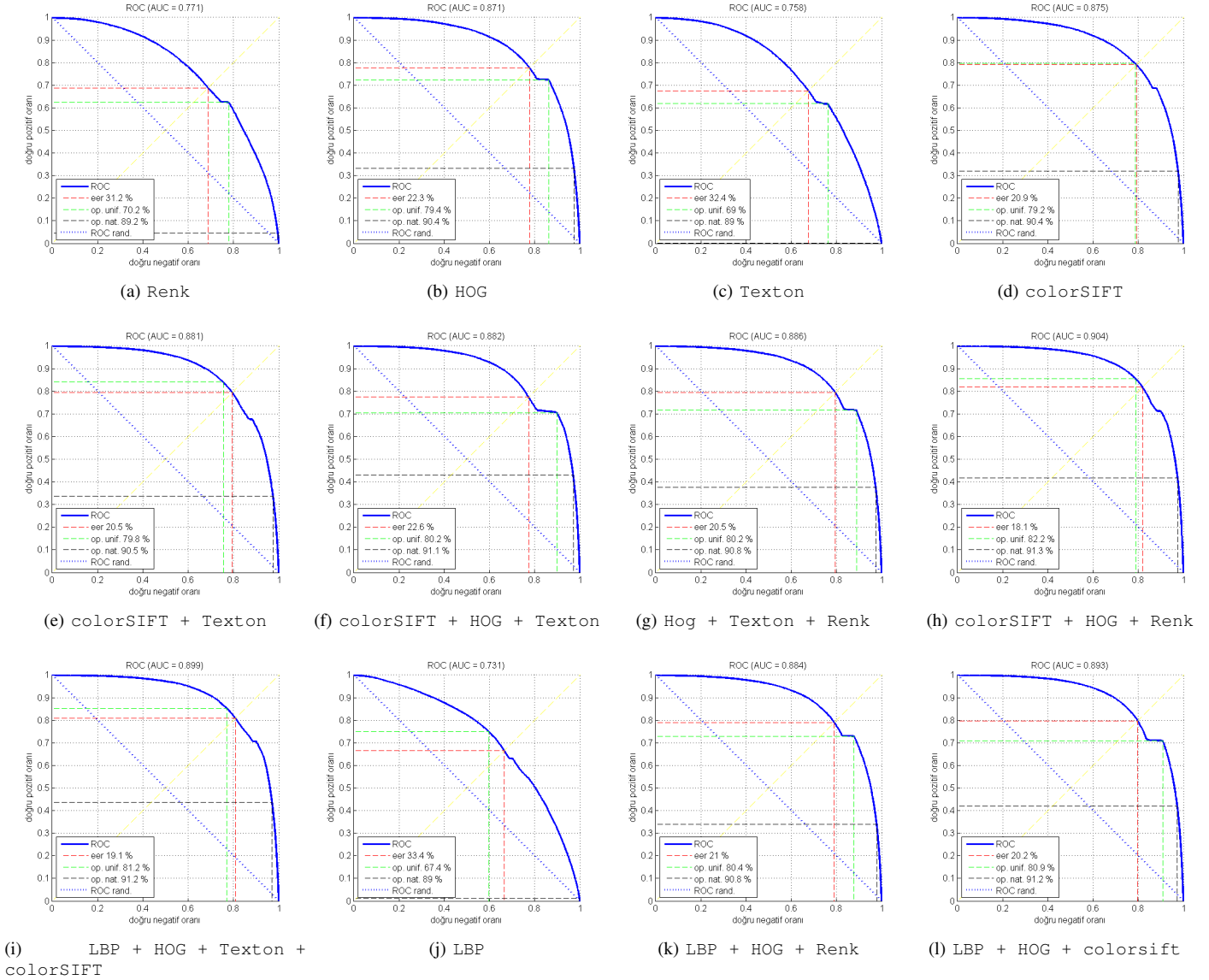
Öznelikler	ROC Alanı
LBP	0.731
Renk	0.771
Texton	0.758
HOG	0.871
ColorSift	0.875
ColorSift + Texton	0.881
ColorSift + HOG + Texton	0.882
HOG + Renk + Texton	0.886
LBP + HOG + Renk	0.884
ColorSift + LBP + HOG	0.893
ColorSift + LBP + HOG + Texton	0.899
ColorSift + HOG + Renk	0.904

suretiyle gerçekleştirilmiştir. Deneyler sonucunda elde edilen sonuçlar Şekil 2 ve Tablo I'de sunulmaktadır.

Elde edilen sonuçlara göre, tek öznelikler arasında en yüksek başarıyı ColorSIFT öznelikleri sağlamaktadır. LBP özneliklerinin diğer öznelikler ile birlikte kullanımı başarıyı artırmaktadır. En yüksek başarı ise, colorSIFT, HOG ve renk özneliklerinin birarada kullanılması ile elde edilmiştir. Yerel şekil ve renk bilgisinin birleşimindeki sonuçların daha iyi değerler verdiği görülmektedir. Sadece renk bilgisinin kullanımı ya da sadece yerel şekil bilgisinin kullanımı öğrenmede yetersiz kalmaktadır.

V. SONUÇLAR

Bu çalışma, alt-düzye özneliklerin, görsel nitelik öğrenme üzerindeki etkisini incelemektedir. Bu kapsamda, pekçok alt-düzye öznelik çıkartılmış, ve bu özneliklerin farklı birleşimlerinin nitelik sınıflandırma performansı üzerindeki etkileri araştırılmıştır. Yapılan deneyler sonucunda, özellikle yerel renk ve şekil bilgisinin bir arada kullanılmakta olduğu colorSIFT özneliklerinin nitelik sınıflandırmada performansa olumlu katkıda bulunduğu gözlemlenmiştir. İlerleyen çalışmalarda, farklı özneliklerin bu çerçevede değerlendirilmesi planlanmaktadır.



Şekil 2: Görsel nitelik öğrenme için farklı öznelik kullanımında elde edilen ROC eğrileri.

TEŞEKKÜR

Bu çalışma TÜBİTAK tarafından 112E149 no'lu Kariyer projesi kapsamında desteklenmiştir.

KAYNAKÇA

- [1] Ferrari, V. and Zisserman, A., "Learning Visual Attributes", "Advances in Neural Information Processing Systems", 2007
- [2] Farhadi, A., Endres, I., Hoiem, D., Forsyth, D. "Describing objects by their attributes" In: Proc. Comp. Vision Pattern Recogn. (CVPR). pp. 1778–1785, 2009
- [3] Turakhia, N. and Parikh, D., "Attribute Dominance: What Pops Out?", Computer Vision (ICCV), 2013
- [4] Parikh, D., Grauman, K., "Relative attributes", ICCV, pp. 503–510, 2011
- [5] Christie G., Parkash A., Krothapalli U., Parikh D., "Predicting User Annoyance Using Visual Attributes", CVPR, 2014
- [6] Dhar, S., Ordonez, V., Berg, T. L., "High level describable attributes for predicting aesthetics and interestingness", CVPR, pp. 1657-1664, 2011
- [7] Chen C. Y., Grauman K., "Inferring Analogous Attributes", CVPR, 2014
- [8] Vaquero, D. A., Feris, R. S., Tran, D., Brown, L. M. G., Hampapur, A., Turk, M., "Attribute-based people search in surveillance environments", WACV, IEEE Computer Society, pp. 1-8, 2009
- [9] A. Kovashka, D. Parikh, K. Grauman., "Whittlesearch: Image search with relative attribute feedback", CVPR, pp. 2973–2980, 2012
- [10] Berg, Tamara L., Berg, Alexander C., Shih, J., "Automatic Attribute Discovery and Characterization from Noisy Web Data", ECCV, 2010
- [11] Andrew Y. Ng. "Feature selection, l1 vs. l2 regularization, and rotational invariance", In ICML, 2004.
- [12] R.E. Fan, K.W. Chang, C.J. Hsieh, X.R. Wang, and C.J. Lin., "LIB-LINEAR: A library for large linear classification", Journal of Machine Learning Research 9, pp. 1871-1874, 2008
- [13] Abdel-Hakim, A. E., Farag, A. A., "CSIFT: A SIFT Descriptor with Color Invariant Characteristics", In CVPR, pp. 1978-1983, 2006
- [14] Dalal, N., Triggs, B., "Histograms of oriented gradients for human detection", In: CVPR, pp. 886–893, 2005
- [15] Varma M., Zisserman, A., "A statistical approach to texture classification from single images" Int. J. Comput. Vision, pp. 62:61–81, 2005