

Kişisel Görüntü Kümelerinin İçsel Özellikler Kullanılarak Özetenmesi

Summarizing Personal Image Collections with Intrinsic Properties

Göksu Erdoğan, Bora Çelikkale, Aykut Erdem, Erkut Erdem

Bilgisayar Mühendisliği Bölümü, Hacettepe Üniversitesi, Ankara, Türkiye
{goksuerdogan, ibcelikkale, aykut, erkut}@cs.hacettepe.edu.tr

I. GİRİŞ

Özetçe —Görsel özetleme problemleri doğası gereği karmaşıktır. Zira özeten tanımı bir miktar belirsizlik taşımaktadır ve tek bir doğrudan söz edilemez. İyi bir özet, genellikle şu iki temel özelliğe bağlı olarak belirlenir: (i) kapsam ve (ii) çeşitlilik. Bir özeten iyi olması için kapsamının geniş olması gereklidir. Başka bir deyişle, özet verilen kümelerdeki kilit olayları ve kavramları içermelidir. Aynı zamanda iyi bir özetin çeşitliliğinin de iyi olması gereklidir yani aynı veya benzer olay ve kavramlar özette tekrarlanmamalıdır. Sosyal medya uygulamalarının yaygınlaşması ile kapsam ve çeşitliliğe ek olarak görüntülerin kalite, duyu uyandırma, popülerlik gibi içsel özellikler de önem kazanmıştır. Çalışmamızda kişisel görüntü kümelerini otomatik özetleme işlemi için, bir özeten sahip olması gereken iki temel özelliğin yanısıra görüntünün içsel özelliklerini de hesaba katan bir yöntem önerilmiştir. Geliştirilen bu yöntem çeşitli kişisel görüntü kümelerinin kitle kaynaklı toplanan özetleri ile kıyaslanması ve içsel özelliklerin dikkate alınmasının elde edilen özetleri iyileştirdiği gözlemlenmiştir.

Anahtar Kelimeler—*Görüntü Kümesi Özetleme, İçsel Özellikler*.

Abstract—Visual summarization problems are complex intrinsically. Because definition of a summary have some ambiguity and only one correct summary does not exist. A good summary consists of two main properties in general which are (i) coverage and (ii) diversity. A good summary should have high coverage. On the other words summary should consist of key events and concepts for given set. At the same time a good summary should also be diverse, i.e it should not consist of similar events and concepts. In addition to these two main properties, intrinsic properties such as quality, emotions, popularity of images increase in importance depending on prevalence of social media applications. We proposed an automatic summarization method which considers intrinsic properties of images in addition to coverage and diversity for personal image collection summarization. This developed method is compared with summaries of different personal image collections which are collected by crowdsourcing and it is observed that the taking intrinsic properties into account improves the summaries.

Keywords—*Image Set Summarization, Intrinsic Properties*.

Bu çalışma kısmen TÜBİTAK 113E497 nolu proje tarafından desteklenmiştir.

978-1-5090-1679-2/16/\$31.00 ©2016 IEEE

Günümüzde yaşanan teknolojik gelişmeler, akıllı telefonları hayatımızın önemli bir parçası haline getirmiştir. Bu gelişmeler telefonların öncelikli işlevi olan iletişimini sağlama özelliğinin yanında kamera olarak kullanılabilirliğine olanak sağlamıştır. Bunlara ek olarak sosyal medya uygulamalarının özendirici rolü görüntü verisinin her geçen gün büyük bir hızla artmasına sebep olmaktadır. Elde edilen son verilere göre her gün sadece Flickr, Facebook, Instagram ve Snapchat üzerinden 500 milyon resim paylaşmaktadır [9]. Doğum günümüzü kutlamak için büyük bir parti ayarladığımızı varsayıyalım. Her anımızı fotoğraflarla ölümsüzleştirmek isteriz ve buna bağlı olarak birçok fotoğraf çekерiz. Sonrasında etkinliğin basılı bir albümünü oluşturmak istediğimizde ya da sosyal medya hesabımıza yüklemek istediğimizde farklı kişilerle, farklı anlarda çekilmiş güzel fotoğraflardan oluşan bir kümeye özen gösteririz. Bunun için, bu bir yığın fotoğrafı tek tek elden geçirmenin yanısıra hangisinin daha güzel olduğuna karar vermeye çalışmak hem zaman alıcı hem de yorucudur.

Son zamanlarda görüntülerden bazı içsel özellikler çıkarıma dayanan yöntemler sıkça çalışmaya başlamıştır. Sosyal medya hesabımıza yüklediğimiz görüntülerin çok beğeni almasını ya da albümümüze koymak istediğimiz görüntülerin içimizde güzel duygular uyandırmasını isteyebiliriz. Bu durumda görüntülerden çıkaracağımız içsel özellikler, kişisel tercihlerimizi çıkarılacak özete yansıtabilememize olanak sağlar. Örneğin, Gygli vd. görüntülerin insanlar için ne kadar ilginç olduğunu bulmaya çalışmıştır [10]. Bir görüntünün ilginçliğini görüntünün estetikliğiyle ve nadirliğiyle ilişkilendirmiştir. Khosla vd. ise çalışmasında [3] bir görüntünün popüler olma ihtimalini tahmin etmeye çalışmıştır. Borth vd. tek bir özelliği bulmak yerine görüntünün insanda uyandığı birçok his ve duyguya ölçmeye çalışmıştır [4]. Bir başka çalışmada [11] ise resimlerin akılda kalıcık (ing. memorability) değerleri hesaplanmaya çalışılmıştır. Çalışmamızda görüntülerden çıkarılan içsel özellikler görüntü kümelerini özetleme yöntemini iyileştirmek için önerilmiştir. İçsel özelliklerin etkisi determinantsal nokta süreçleri (ing. determinantal point processes) (DPP) [2] ve çeşitlilik sıralama (ing. diversity ranking) [7]

yöntemleri ile çıkarılan özetler üzerinde gösterilmiştir.

Bu çalışmada katkılardımız şu şekilde özetlenebilir:

- 1) Kişiisel görüntü kümesi özetleme problemi için kullanıma açık etiketli bir veri kümesi bulunmadığından, veri kümesi toplama yoluna gidilmiştir. Veri kümesindeki her kişiisel görüntü kümesi için, dayanak olarak kullanmak üzere kitle kaynaklı (ing. crowdsourcing) olarak kullanıcılardan özetler toplanmıştır.
- 2) Kişiisel görüntü kümesi özetlerken görüntünün içeriğinin yanısıra görüntünün içsel özelliklerinin kullanılmasının çıkarılan özetlerin başarısına etkisi gösterilmiştir.

Bildirinin geri kalımı şu şekilde düzenlenmiştir: Bölüm II'de öncelikle kullanılan yöntemin detaylarından bahsedilmiştir. Daha sonra Bölüm III'te nice olümlerin sonuçlarına, karşılık geldikleri görsel özetlere yer verilmiş ve sonuçlar detaylı olarak yorumlanmıştır.

II. YÖNTEM

Çalışmada önerilen yöntem dört bölümde anlatılmıştır. Öznitelik Çıkarma bölümünde görüntülerden öznitelik çıkarma işlemi ve çıkarılan özniteliklerin hangi bilgileri taşıdıkları detaylandırılmıştır. Devamındaki İçsel Özellikler bölümünde görüntülerden çıkarılan mantıksal özelliklerden ve nasıl çıkarıldıklarından bahsedilmiştir. Son iki bölümde ise çıkarılan tüm özelliklerin iki farklı yöntemle özetleme işlemeye nasıl dahil edildiği anlatılmıştır.

A. Öznitelik Çıkarma

Görüntülerden her biri görüntüye dair farklı bilgileri taşıyan üç öznitelik çıkarılmıştır. Bu üç öznitelik, her görüntü için tek bir öznitelik vektörü elde etmek amacıyla arka arkaya eklenmiştir. Öznitelik çıkarma işlemi sözlük öğrenme gerektirdiğinden her bir görüntü kümesi için ayrı ayrı tekrarlanmıştır.

1) Renk bilgisi: Bu öznitelik, görüntünün genelindeki renk bilgisini ifade etmek için kullanılmıştır. Ancak genel olarak kullanılan histogram yaklaşımı yerine görüntülerin gerçek hayatı sıkça kullanılan 11 renkten (siyah, mavi, kahverengi, gri, yeşil, turuncu, pembe, mor, kırmızı, beyaz, sarı) hangisine denk geldiği bulunmuştur [5]. Görüntüden öncelikle 10x10 piksellik parçalar çıkarılmıştır. Daha sonra her bir parça, içinde en sık geçen renk ile ifade edilmiştir. Her görüntü için bu renklerin histogramları çıkarılmış ve L1 uzaklığı ile normalize edilmiştir.

2) Süper piksel bilgisi: Bu öznitelik, görüntünün içindeki benzer küçük parçaların ilişkisini ifade etmek için kullanılmıştır. Öncelikle görüntü kümesindeki tüm görüntülerden yoğun SIFT öznitelikleri çıkarılmış ve k-ortalama (ing. k-means) algoritması ile 200 kelimek görsel sözlük oluşturulmuştur. Daha sonra süper piksel'ler çıkarılmış ve her bir süper piksel alanının içinde kalan görsel kelimelerden bir histogram hesaplanmıştır. Böylelikle her görüntü süper piksel alanlarının içinde kalan görsel kelimeler cinsinden histogramlar olarak ifade edilmiştir. Bu histogramların sonuna yine aynı süper piksel alanından çıkarılmış kırmızı, yeşil, mavi (ing. RGB) renk histogramları eklenerken ara öznitelikler oluşturulmuştur. Bu ara özniteliklerin üzerinden ikinci kez k-ortalama ile 200

kelimelik görsel sözlük oluşturulmuş ve vektörler küme merkezlerine göre kodlanmıştır. Son olarak her görüntü için bu görsel kelimelerin histogramları çıkarılmış ve L1 uzaklığı ile normalize edilmiştir.

3) Konvolüsyonal derin ağlar (ing. Convolutional neural networks)(CNN) öznitelikleri: Bu yüksek seviye öznitelik sahne bilgisi ve nesnelere ait bilgileri taşırlar. Görüntü 0.5 örtüşme ile 10x10 parçaya bölünmüştür. Her bir parçadan önceden öğrenilmiş CNN sınıflandırıcılarının [6] 'fc7' katmanındaki değerleri alınmış ve bu değerler ara öznitelikler olarak kullanılmıştır. Daha sonra bu ara öznitelikler k-ortalama algoritması ile 300 kümeye ayrılmış ve çekirdek kodlaması (kernel codebook encoding) [12] yöntemi ile yumuşak atama yapılarak küme merkezlerine göre kodlanmıştır. Son olarak her görüntü için bu görsel kelimelerin histogramları çıkarılmış ve L1 uzaklığı ile normalize edilmiştir.

B. İçsel Özellikler

Görüntülerden içsel özellik olarak görüntünün popülerliğini ve insanda yarattığı duyguları ifade eden özellikler çıkarılmıştır.

1) Görüntünün Popülerliği: Görüntü popülerliğini ölçmek için Khosla vd.'nin [3] çalışmasından faydalanyanmıştır. Görüntünün içeriğinin görüntünün popülerliğine etkisini şu özellikleri kullanarak hesaplamıştır.

- Renk bilgisi ve basit özellikler. Örneğin, HSV renk uzayında ortalama renk değeri, renk histogramları vb.
- Düşük seviyeli görsel özellikler. Örneğin, Gist, doku, eğim, renk parçaları, CNN
- Yüksek seviye görsel özellikler. Örneğin, görüntüdeki nesneleri çıkarmak için bizim yaptığımız gibi CNN öznitelikleri kullanılmıştır.

2) Duyguların Bulunması ve Ölçülmesi: Görüntülerdeki duyguları ölçmek için Borth vd.'nin [4] çalışmasından faydalanyanmıştır. Borth vd. bu çalışmasında Flickr ve Youtube'dan getirilen görüntüler ve etiketlerden yola çıkarak 3000 kavramdan oluşan bir ontoloji inşa etmiştir. Seçilen kavramlar, güçlü his uyandıran, duygulara bağlantısı olan, sık kullanılan ve bulunma olasılığı yüksek olanlar arasından seçilmiştir. Bu kriterleri sağlamak için isim sıfat ikililerinden (İSİ) faydalanyanmıştır. Böylece tek başına isim veya sıfat kullandığında oluşan tarafsız durumları elemiştir. Örneğin, 'kelebek' tek başına duyu uyandırmazken 'renkli kelebek' dendigidende güçlü bir duyu uyandırmaktadır. Bu ontolojinin üzerine 1200 tane isimsıfat ikilisinin görüntülerdeki bulunma durumlarını tespit eden öğrenilmiş kavram algılayıcılar geliştirilmiştir. Her kelimeye -1 (negatif) ile +1 (pozitif) aralığında değişen his değerleri verilmiştir. İsim sıfat ikililerinin his değerleri ise 's' fonksiyonu ile hesaplanmıştır:

$$s(\text{İSİ}) = s(\text{sifat}) + s(\text{isim}) \quad (1)$$

Çalışmamızda bu kavram algılayıcı kütüphane aracılığıyla görüntülerin içeriği hisler ve bunların bulunma yanıtları hesaplanmıştır. Böylece her görüntü için 1200 uzunluğunda vektörler elde edilmiştir. Bu özelliğin özetleme işlemeye katılabilmesi ancak bir tamsayı değer olarak ifade edildiğinde mümkün olduğu için bu vektörden anlamlı bir değer çıkarma yoluna

gidilmiştir. Bunun için de Denklem 2'de verilen formül kullanılmıştır. Algılayıcı yanıtlarının vektörü σ elde edildikten ve s değeri Denklem 1 ile hesaplandıktan sonra, görüntünün duyuğu değeri q aşağıdaki formüle göre hesaplanmıştır:

$$q = \sum_{i=1}^{1200} \sigma_i \times s_i \quad (2)$$

Bu formüle göre insanda olumsuz duyuğu uyandıran görüntüler, olumlu duyuğu uyandıran görüntülere göre daha düşük duyuğu değerine sahip olacaktır.

C. Determinantsal Nokta Süreçleri(DPP)

N elemandan oluşan bir \mathcal{Y} kümesi düşünelim: $\mathcal{Y} = \{1, 2, 3, \dots, N\}$. \mathcal{Y} kümesinin alt kümeleri $2^{\mathcal{Y}}$ ile gösterilsin. \mathcal{Y} kümesinde DPP, herhangi bir $Y \subseteq \mathcal{Y}$ için olasılık ölçümüne denk gelir ve \mathcal{P} ile gösterilir [2]. L matrisi DPP kernelinin tamamını L_Y ise DPP kernelinin Y 'deki indisler için indirgenmiş reel simetrik NxN'lik benzerlikleri içeren alt matrisini ifade eder ve Y olasılığı

$$\mathcal{P}(Y \subseteq \mathcal{Y}) = \det(L_Y) \quad (3)$$

ilişkisiyle de ifade edilebilir. DPP'nin başka bir gösterim şekli ise marginal kernel K 'dır. K ve L kernelleri arasındaki geçiş kolaylıkla sağlanır: $K = (L + I)^{-1}L$. DPP'nin benzerlikleri nasıl modellediğini anlatmak için K kerneli daha uygundur çünkü bu kernel herhangi bir Y alt kümesinin olasılığını hesaplamak için gerekli tüm bilgileri içerir.

$$\begin{aligned} \mathcal{P}(i, j \in \mathcal{Y}) &= \begin{vmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{vmatrix} \\ &= K_{ii}K_{jj} - K_{ij}K_{ji} \\ &= \mathcal{P}(i \in \mathcal{Y})\mathcal{P}(j \in \mathcal{Y}) - K_{ij}^2 \end{aligned} \quad (4)$$

Örneğin, eğer $Y = \{i\}$ ise $\mathcal{P}(i \in \mathcal{Y}) = K_{ii}$ dir. Köşegendeki değerler elemanların bireysel olarak dahil olma olasılığını gösterir. Bu değerin 1'e yakın olması o elemanın nerdeyse her zaman DPP tarafından seçileceğini gösterir. Örneği biraz daha genişletmek adına $Y = \{i, j\}$ olduğunu yani Y kümesinin 2 elemanlı olduğunu düşünelim. Bu durumda, Denklem 4'te görüldüğü üzere, köşegen üzerinde bulunmayan elemanlar, ikişiler arasındaki negatif ilişkiye tanımlar. K_{ij} ne kadar büyük olursa i ve j elemanlarının beraber bulunma olasılığı azalır. Denklemdeki K_{ij} ifadesi DPP'nin ayrık kümeler seçmesini sağlar. K matrisinin \mathcal{Y} 'nin elemanları arası benzerliğin ölçümü olduğunu düşünürsek benzer elemanlar daha az birlikte bulunma eğilimindedir. Eğer $K_{ij} = \sqrt{K_{ii}K_{jj}}$ ise i ve j elementleri son derece benzerdir. Dolayısıyla nerdeyse emin bir şekilde beraber bulunamayacaklarını söyleyebiliriz.

Buraya kadar DPP'nin öncelikli işlevi olan ayrıklığı sağlamayı elemanlar arası benzerlik ölçümünerini içeren bir kernel kullanarak gerçekleştirdiğinden bahsettik. Bu kernel Gram matrisi olarak da gösterilebilmektedir: $L = B^T B$. Çalışmamızda B matrisinin sütunları görüntülerden çıkarılmış öz niteliklere karşılık gelmektedir. Ancak bu gösterim sadece ayrıklıkları ele almaktadır. Hem ayrıklıkları hemde içsel özelliklerini hesaba katmak için Denklem 5 kullanılmaktadır. Bu denklemdeki q_i içsel özelliklerden gelen kalite değeri, ϕ_i ise ayrıklık özellikleridir.

$$L_{i,j} = q_i \phi_i^T \phi_j q_j \quad (5)$$

Son olarak DPP'den sonuç çıkarabilmek için üstel zaman gerektiren normalizasyon ve örneklemme işlemlerinin gerçekleştirilmesi gereklidir ve bu amaçla Kulezsa ve Taskar tarafından önerilen polinomsal zamanlı örneklemme algoritması [8] kullanılmıştır.

D. Çeşitlilik Sıralama

Çeşitlilik sıralama yöntemi [7] temelinde altbirimsel bir fonksiyonu eniyilemeye çalışır. Altbirimsel fonksiyonlar 'azalan getiriler' özelliğini sağlayan fonksiyonlardır. Bu özelliğe göre $A \subseteq B \subseteq V - v$ ilişkisi bulunan bir kümede $f(A \cup v) - f(A) \geq f(B \cup v) - f(B)$ eşitsizliği sağlanmalıdır. Bu eşitsizliğin arkasında yatan mantık küçük kümeye yeni bir eleman eklediğimizde sağlanan kazanç, büyük kümeye aynı elemanı eklediğimizde sağlanan kazançtan büyük olmaktadır. Bu fonksiyonların en önemli özelliği ağızgözü yaklaşım kullanılarak en uygun sonuca yakın sonuçların bulunmasını garantilemesidir.

Çeşitlilik sıralama adından da anlaşılacağı gibi görüntülerde tekrar eden bilgileri azaltarak çeşitli elemanlardan oluşan bir alt kume seçmeye çalışır. En uygun dağılım birbirinden uzak ve eşit büyülüklükte elemana sahip 10 tane merkezin bulunduğu durumdur. Bu yöntem seçme sırasında merkeziliği korumaya çalışır. Merkeziliği sağlaması benzer elemanların bir araya toplanması açısından önemlidir. Ancak bu görüntülerin tüm kümeyi kapsaması için merkezlerin birbirlerinden uzak olmaları gerekmektedir. Öz nitelik vektörleri arasındaki ilişkiye Gauss benzerliği kullanarak ifade eder.

$$d_{xy} = \begin{cases} \exp(-\beta \|g(x) - g(y)\|^2), & \text{eğer } (x, y) \in \varepsilon \\ 0 & \text{diğer} \end{cases} \quad (6)$$

Seçme işlemini ise aşağıdaki altbirimsel fonksiyonu eniyileerek gerçekleştirir.

$$\begin{aligned} \max \sum_{x \in V} u(x) \\ a_x = \sum_{(x,y) \in \varepsilon} d_{yx} + z_x \text{ ise } u(x) = \frac{1}{a_x} \sum_{(x,y) \in \varepsilon} d_{yx} u(y) \\ s \in S \subset V, |S| \leq K \text{ için } u(g) = 0, u(s) = 1 \end{aligned} \quad (7)$$

$u(g)$ hiçbir eleman seçili değilken fonksiyonun değerinin sıfır olduğu gösterir. $u(s)$ ise bir eleman seçildiğinde onun değerinin 1 olduğunu gösterir. K tane eleman seçme hakkımız olduğunu düşünürsek, seçtiğimiz K tane eleman için u enerji fonksiyonunun en yüksek değerini alması sağlanmaya çalışılmaktadır. Enerji fonksiyonu karmaşık gözükmesine rağmen basitçe bir noktadaki enerjiyi komşu noktalardaki enerjilerin ağırlıklı bir toplamı cinsinden ifade etmektedir. Ayrıca z parametresi merkezilik ve çeşitlilik arasındaki dengeyi sağlamak için kullanılır. Bu parametre enerji kaybını gösterir ve aynı zamanda komşuluk çerçevesini belirler. z parametresinin büyük olması komşuluk çerçevesini daraltarak benzer noktalardan ayrılmış gibi ele alınmasına sebep olur. Dolayısıyla z parametresi bu durum göz önünde bulundurularak seçilmelidir. Bu çalışmada z parametresi ortalama bağlı olarak seçilmiştir.

Çeşitlilik sıralamada içsel özelliklerden gelen kalite değerleri ağırlık noktaları kullanılarak hesaba katılmıştır. Böylece kalite değerleri büyük olanlar daha büyük ağırlığa küçük olanlar ise daha küçük ağırlığa sahip olmuştur.



Şekil 1: Özetler:Birörnek (üst sol), K-ortalama (üst sağ), DPP (alt sol), Çeşitlilik Sıralama (alt sağ).



Şekil 2: Burning Man görüntükümesi.

III. DENEYLER

Kişisel görüntü kümelerini özetlemek için herkesin erişimine açık ve dayanak özetleri toplanmış bir veri kümesi bulunmadığı için kendi veri kümemizi toplamaya yöneldik. Veri kümemiz Flickr'dan İngilizce sorgular (Burning Man, İstanbul trip ...) ile çekilen görüntülerden oluşmaktadır. Sorgular belirlenirken belli bir etkinlik çerçevesinde olmalarına dikkat edilmiştir. Her biri 100 görüntünden oluşan 8 ayrı veri kümesi toplanmıştır (Şekil 2). İnsanlardan dayanak olarak kullanmak üzere veri toplamak için kendi web sitemizi gerçekleştirdik¹. Kullanıcılarından gördükleri görüntükümesini en iyi temsil eden 10 görüntüyü seçmeleri istenmiştir. Kullanıcılarla görüntüler zaman etiketlerine göre sıralı bir şekilde gösterilmektedir. Ayrıca iyi bir özeti iki temel özelliği olan kapsam ve çeşitlilik birer cümleyle tanımlanarak kullanıcı bilgilendirilmiştir.

Özetleme problemlerinin bir diğer zorluğu özeten başarısının nicel olarak ölçülmesidir. Bu çalışmada Tschiatschek vd. [1] tarafından önerilen, dokümanlarda kullanılan Rouge metriğinin görsel veriler için bir uyarlaması olan Görsel Rouge (Visual Rouge) (V-Rouge) ile başarımnicel olarak ölçülmüştür. Bu metrik insan algısıyla ilintili olduğu için tercih edilmiştir. Sadece insanlar tarafından çıkarılmış dayanak özetler bulunduğunda ölçülebilir. Her görüntükümesi için toplanan kullanıcı özetlerinden en yüksek Görsel Rouge değerine sahip olan 20 kullanıcı özeti seçilmiştir. Bu sayede özetler arası tutarlılık sağlanmaya çalışılmıştır.

Öncelikle Bölüm II'de bahsedilen DPP ve çeşitlilik sıralama yöntemleri ile görüntülerden çıkarılan içsel özelliklerin getirdiği kalite bilgisi de hesaba katılarak özetler çıkarılmıştır. Tüm özetlerin başarısı karşılık geldikleri V-Rouge değerile ölçülmüştür. Tablo I'de özetler ortalama V-Rouge değerine göre kıyaslanmıştır ve görüldüğü üzere bir tek durum hariç (DPP-popülerlik ikilisi) bütün durumlarda içsel özellikler, kullanılmadığı duruma göre sonuçları iyileştirmiştir. Daha sonra bu özetler, iki dayanak yöntemele çıkarılan özetler ile karşılaştırılmıştır. Birörnekleme ile seçilen özetler (ing. uniform) zaman etiketlerine göre sıralanmış görüntülerden eşit aralıklarla 10 görüntünün seçilmesiyle oluşturulmuştur. K-ortalama yönteminde görüntükümelere k-ortalama algoritması ile 10 kümeye bölünmüş ve merkezlere en yakın görüntüler seçilmiştir. Önerdiğimiz yöntemler Tablo II'de görüldüğü gibi dayanak yöntemleri geçmiştir. Tablo II'de verilen sonuçlar

Tablo I'de en iyi sonuçları üreten ayarlar kullanılarak hesaplanmıştır. Ayrıca Tablo II'de ortalama V-Rouge değerinin yanısıra yöntemlerin ortalama V-rouge değerlerine göre sıralanmasına bağlı olan Ortalama Sıra değeri gösterilmiştir. Sonuçlar gösteriyor ki insanların özetleme sırasında seçtikleri görüntüler duyu uyandırma kapasitesi ve popüler olma ihtimali yüksek olan görüntülerde içeriyor. Çıkarılan özetleri niteliksel olarak örneklemek adına Burning Man görüntükümesi için bütün yöntemlerle çıkarılan özetler Şekil 1'de gösterilmiştir.

Tablo I: İçsel özelliklerin etkisi.

İçsel Özellikler	DPP	Çeşitlilik Sıralama
-	0.747	0.761
Popülerlik	0.747	0.770
Duygu	0.751	0.768
Popülerlik & Duygu	0.753	0.764

Tablo II: Dayanak çalışmalarla karşılaştırma.

Görüntü Kümesi Adı	Birörnek	K-ortalama	DPP	Çeşitlilik Sıralama
A Trip to New York	0.759	0.750	0.732	0.779
A Visit to Italy	0.745	0.759	0.781	0.790
An Istanbul Tour	0.696	0.713	0.693	0.770
Birthday Party	0.809	0.812	0.821	0.772
Burning Man	0.662	0.731	0.727	0.733
London Olympics	0.667	0.735	0.720	0.748
St. Patrick's Day	0.775	0.750	0.792	0.752
World Cup	0.745	0.747	0.762	0.816
Ortalama Sıra	3.250	2.625	2.500	1.625
Ortalama V-rouge	0.732	0.750	0.753	0.770

KAYNAKLAR

- [1] S. Tschiatschek, R. Iyer, H. Wei, J. Bilmes, "Learning Mixtures of Submodular Functions for Image Collection Summarization", NIPS, 2014.
- [2] A. Kulesza, B. Taskar, "k-DPPs: Fixed-Size Determinantal Point Processes", ICML, 2011.
- [3] A. Khosla, A. Das Sarma, R. Hamid, "What Makes an Image Popular?", WWW, 2014.
- [4] D. Borth, R. Ji, T. Chen, T. Breuel, SF. Chang, "Large-scale Visual Sentiment Ontology and Detectors using Adjective Noun Pairs", ACM MMS, 2013.
- [5] J. van de Weijer, C. Schmid, J.J. Verbeek, "Learning Color Names from Real-World Images", CVPR, 2007.
- [6] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", ICLR, 2015.
- [7] G. Kim, E. P. Xing, L. Fei-Fei, T. Kanade, "Distributed Cosegmentation via Submodular Optimization on Anisotropic Diffusion", ICCV, 2011.
- [8] A. Kulesza, B. Taskar, "Determinantal Point Processes for Machine Learning", Foundations and Trends in Machine Learning, 2012.
- [9] M. Meeker, L. Wu, "Internet trends: Technical report", Perkins Caufield Byers, 2013.
- [10] M. Gygli, H. Grabner, H. Riemenschneider, F. Nater, L. V. Gool, "The Interestingness Of Images", ICCV, 2013.
- [11] A. Khosla, A. S. Raju, A. Torralba, A. Oliva, "Understanding and Predicting Image Memorability at a Large Scale", ICCV, 2015.
- [12] K. Chatfield, V. Lemtexpitsky, A. Vedaldi, A. Zisserman, "The devil is in the details: an evaluation of recent feature encoding methods", BMVC, 2011.

¹ vissum.cs.hacettepe.edu.tr