

İnsan Kalabalıklarının Baskın Kümeler Tabanlı Analizi

Dominant Sets Based Analysis of Human Crowds

Burçak Asal, Aykut Erdem, Erkut Erdem
Bilgisayar Mühendisliği Bölümü, Hacettepe Üniversitesi, Ankara, Türkiye
{basal,aykut,erkut}@cs.hacettepe.edu.tr

Özetçe —Günümüzde kullanılan yeni kamera teknolojileri ve İnternet sayesinde milyonlarca videoya istenildiği zaman, istenilen yerden ulaşmak mümkün hale gelmiştir. Bu videoların önemli bir bölümünü, dinamik sahnelerde çeşitli eylemlerde bulunan insanlar ve araçlar gibi aktörler içeren gözetleme amaçlı videolar oluşturmaktadır. Bu çalışmanın amacı, videolardaki kalabalık insan hareketlerinin analizidir. Daha detaylı bir şekilde belirtmek gerekirse; çalışmamızda bir video sekansı boyunca hareket halinde olan insanlar takip edilmekte; ve daha sonra takip edilen bu insan kalabalıkları yol benzerliği üzerinden Baskın Kümeler yöntemi kullanılarak öbeklenmektedir. Bu öbeleme sonucu üzerinden ayrıca kalabalıkların sahne içinde ne kadar uyum içinde hareket ettikleri sayısal bir değer olarak da ölçülebilmektedir.

Anahtar Kelimeler—*insan kalabalıkları, çizge tabanlı öbeleme, baskın kümeler.*

Abstract—Due to recent advances in new camera technologies and the Internet, millions of videos can be easily accessed from any place at any time. A significant amount of these videos are for surveillance, and include actors such as humans and vehicles performing different actions in dynamic scenes. The goal of this study is to analyze human crowd motions in videos. More specifically, moving humans are tracked throughout a video sequence, and the collective crowd motions are then clustered using path similarities via the Dominant Sets method. Moreover, this clustering result can be used to predict the coherency of the motion as a scalar value.

Keywords—*human crowds, graph based clustering, dominant sets.*

I. GİRİŞ

Çevremizde gözlemlediğimiz en ilgi çekici olaylardan biri de kalabalık hareket davranışlarıdır. En küçüğünden, en büyüğüne kadar bu hareket davranışları, birçok disiplinden araştırmacıların ilgisini çekmektedir. Bu davranışları oluşturan sebeplerden birisi de bireylerin kendi arasındaki davranış etkileşimidir. Son yıllarda videolardaki dinamik sahneler içindeki farklı hareket örüntülerinin bulunması konusu gittikçe önem kazanmaktadır. Bu problemin çözümü özellikle dinamik sahnelerdeki kalabalık analizi, şehir yollarındaki trafik analizi vb. konular için önemli bir motivasyon oluşturmaktadır.

Bu bildiride ele alınan öncelikli konu belirli bir sahnedeki kalabalık grupların hareket örüntülerinin tespit edilmesidir. Bu çalışmaların temelini hareket izleme yöntemleri oluşturmaktadır. Bu noktada karşımıza belirli

zorluklar çıkmaktadır; bu sorunlardan bazıları arka plandaki beklenmeyen kamera hareketleri, kesişen kalabalık hareketleri ve benzeri sorunlardır. Bu kapsamda, genel olarak gerçekleştirilen şu üç ana aşamadan bahsedilebilir;

- 1) Video dosyasının çerçevelere ayrılıp, her bir çerçeve için çeşitli noktaların belirlenip, bu noktaların çerçeveler boyunca takip edilmesi,
- 2) İlgili nokta gruplarından, belirli bir benzerlik metriği kullanılarak bir benzerlik matrisinin oluşturulması (Tek bir çerçeve üzerindeki (önerilen yöntem kapsamında) veya çıkartılan belirli bir iz nesnesi üzerinden [7]),
- 3) Farklı öbeleme yöntemlerinin ilgili benzerlik matrisi üzerinde uygulanıp, hareket gruplarının çıkarılması.

Bu alandaki çalışmalar genellikle, ilk aşamadaki her bir çerçeve için ilgili nokta gruplarını çıkarırken, Optik Akış yöntemi [1],[2],[3] veya KLT (Lucas-Kanade izleyici) yöntemini [4],[5],[6],[8] kullanmaktadırlar. İkinci aşamadaki nokta gruplarından benzerlik matrisini elde etmek için, çalışmalar genellikle her bir ikili noktanın arasındaki benzerlik değerlerini yakalayabilmek için hem hız hem de pozisyon bilgilerini kullanan bir benzerlik ölçütü kullanmaktadırlar [4],[5],[6]. Üçüncü aşama için elde edilen benzerlik matrisi üzerine genellikle Hiyerarşik Toplu Öbeleme [1] veya Bağlı Bileşen Arama [5],[6] tabanlı yöntemler kullanılmıştır. Öbeleme sonucu ayrıca, sahnelerdeki hareket dağılımını gösteren bir topluluk değeri hesaplanmıştır. [6]

Bu çalışma kapsamında, belirli bir veri kümesi üzerinde hareket halinde olan insanlar için takip noktaları çıkartılmış, bu noktalar Baskın Kümeler yöntemi kullanılarak öbeklenmiştir. Ayrıca ilgili öbeleme sonucu üzerinden de, çıkartılan takip noktalarının (hareketlerin) birbirleriyle ne kadar uyumlu olduğunu gösteren sayısal bir değer hesaplanmıştır. Baskın Kümeler yöntemi, çıkartılan nokta grupları arasındaki aykırı noktaları da (gürültü) etkin bir şekilde ayıklamaktadır. Bu yöntemin başka bir avantajı, noktalar üzerindeki grup sayısını otomatik olarak çıkartabilmesidir. Ayrıca altı video sekansı üzerinde, önerilen yöntem ile başka bir yöntem (Collective Merging [6]) görsel olarak karşılaştırılmıştır.

II. ÖNERİLEN YÖNTEM

KLT'nin, optik akış yönteminden farklı olarak herhangi bir çerçeve üzerindeki her piksel üzerinde işlem yapmak yerine, kenar belirleme yöntemi ile belirli koşulları sağlayan



Şekil 1: Farklı veri kümelerinden ((Collective Motion Database [6]) ve [10]) örnek videolar ve nokta takip sonuçları.

daha az sayıda önemli olduğu düşünülen piksel için ilgili noktanın çıkarılması işlemini gerçekleştirmektedir. Bu çalışma kapsamında ilgili noktaların elde edilmesi işlemi için KLT yöntemi kullanılmıştır. Bu yöntem işleyiş aşamasında üç tane varsayım kullanılmaktadır;

- 1) Ardışık çerçevedeki noktaların yansımaları birbirleriyle benzerlik göstermelidir (Parlaklık Tutarlılığı).
- 2) Ardışık çerçevedeki noktaların konum farklarının belirli bir değerin altında olması gerekir (Küçük Hareket).
- 3) Ardışık çerçevedeki noktaların hareket yönlerinin birbiriyle benzer olması gerekir (Hareket Uyumluluğu).

Bu yöntem ile ilgili potansiyel problemlerden birisi istenmeyen kamera hareketidir. Bu hareketten dolayı bazen arka planda olduğu halde, seçilen noktaların yanında (ilgili şekilde farklı renklerdeki çizgiler olarak gösterilmiştir), ayrıca algılanan gereksiz noktalar seçilebilmektedir (Şekil 1'de ikinci resim).

Bu çalışma kapsamında kullanılan benzerlik ölçütü, çalışmamıza motivasyonu oluşturan çalışmadaki gibi [6], hız ve pozisyon bilgilerini içerecek şekilde iki aşamadan oluşmaktadır. Bir video çerçevesinde elde edilen N nokta içerisinde herhangi iki (i ve j olmak üzere) nokta için hız vektörü bileşenleri arasındaki benzerlik aşağıdaki denklemle hesaplanmaktadır:

$$B(i, j) = \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{\|\mathbf{v}_i\| \|\mathbf{v}_j\|} \quad (1)$$

Denklem(1) ile elde edilen $(-1, 1)$ arasındaki değerler üzerinden de aşağıdaki $(0, 1)$ arasında değişen benzerlik değerleri oluşturulmaktadır:

$$W_V(i, j) = 1 - \frac{\arccos(B(i, j))}{\pi} \quad (2)$$

Bu benzerlik matrisi W_V ($N \times N$ boyutunda), noktaların hız vektörlerinin birbirleri arasındaki yön benzerliğini temsil etmektedir.

Noktalar arasındaki pozisyon benzerliği, için de:

$$W_X(i, j) = \begin{cases} 1 & j \in K_i \text{ veya } i \in K_j \\ \exp\left(-\frac{\|X_i - X_j\|^2}{2\sigma^2}\right) & \text{aksi durumda} \end{cases} \quad (3)$$

denklemleri kullanılmıştır. Burada K_i veya K_j , i veya j noktasının, uzaklığına bağlı, kendisine en benzer K komşu noktasını temsil etmektedir. Buradaki σ parametresi, noktaların birbirlerine olan uzaklıklarının ortalaması alınarak otomatik olarak belirlenmektedir. Bu işlemden sonra;

$$W = W_X \odot W_V \quad (4)$$

denklemleri kullanılarak matrisler arasında nokta bazlı çarpım işlemi uygulanmıştır. Elde edilen W matrisi üzerinde noktalar arasındaki komşuluk ilişkilerini daha iyi bir şekilde yansıtabilmesi amacıyla, l -yol benzerliği yöntemi kullanılmıştır;

$$A = W^l \quad (5)$$

Baskın Küme (İng. Dominant Set) [9] kavramı çizge teorisindeki azami klik (maximal clique) kavramının kenar ağırlıklı çizgelere genelleştirilmesi olarak düşünülebilir. Elemanları arasındaki ikili benzerliklere göre oluşturulan yönsüz, kenar ağırlıklı ve kendine döngü içermeyen (köşegen üzerindeki elemanların sıfırlanması) bir çizge, $G = (V, E, w)$ olarak ifade edildiği varsayılırsa burada V düğüm kümesi, $E \subseteq V \times V$ kenar kümesini, $w : E \rightarrow R_+$ ise pozitif ağırlık fonksiyonunu göstermektedir. Böylece $n \times n$ boyutlu G çizgesi, simetrik bir benzerlik matrisi $A = [a_{ij}]$ ile gösterilebilir.

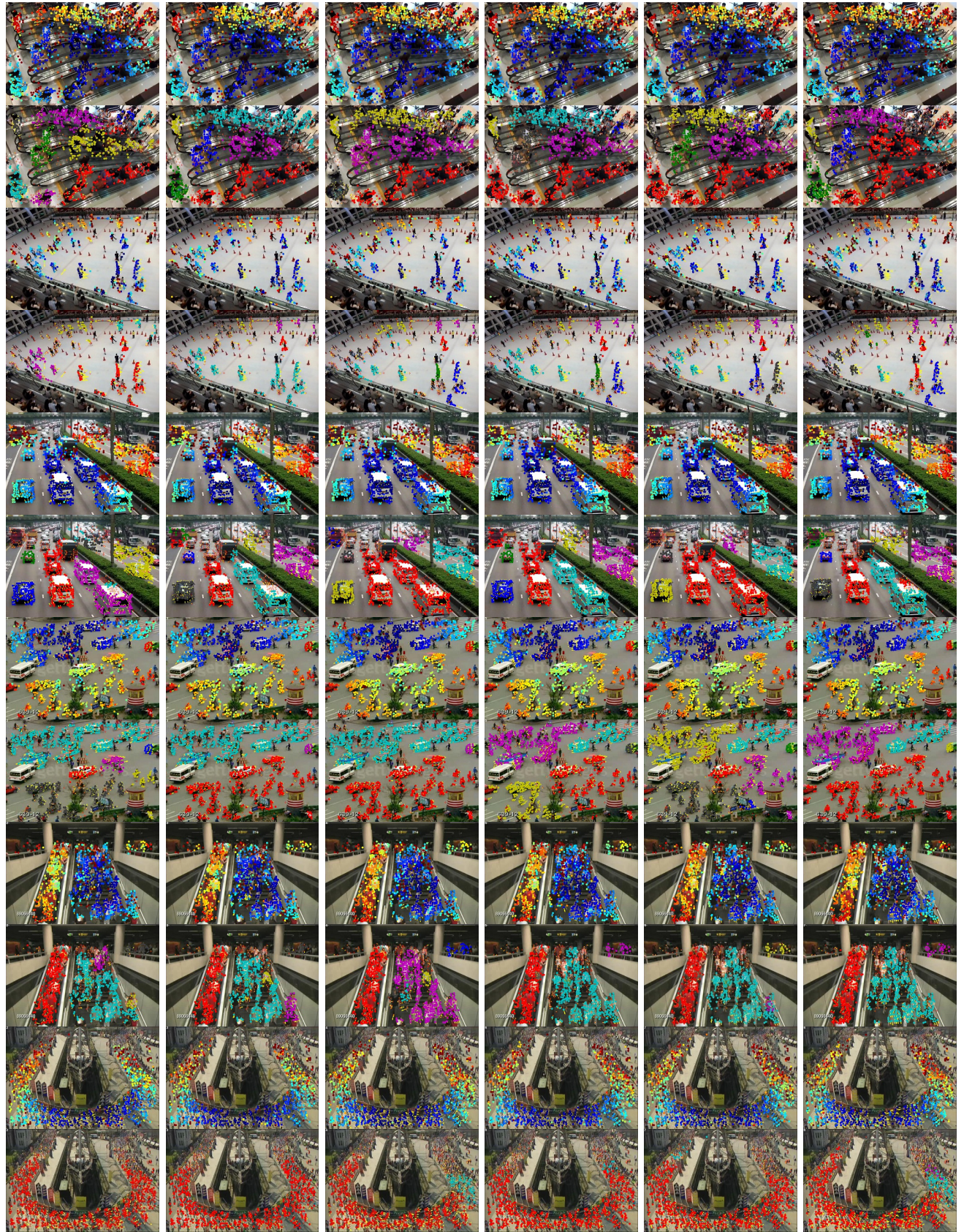
$$a_{ij} = \begin{cases} w(i, j) & (i, j) \in E \\ 0 & \text{aksi durumda} \end{cases} \quad (6)$$

Bunun dışında $S \subseteq V$ boş olmayan bir alt küme ve $i \in V$ iken, i . elemanın S kümesine göre ortalama ağırlık hesaplaması;

$$awdeg_S(i) = \frac{1}{|S|} \sum_{j \in S} a_{ij} \quad (7)$$

Bunun dışında her $j \notin S$ için j ve i arasındaki benzerliği i 'nin S kümesi içindeki $(\phi_S(i, j) = a_{ij} - awdeg_S(i))$ benzerliğine göre hesaplanabilir. ϕ_S 'ye göre bir düğümün S öbeği içindeki uyumluluğu;

$$w_S(i) = \begin{cases} 1 & |S| = 1 \\ \sum_{j \in S \setminus \{i\}} \phi_{S \setminus \{i\}}(j, i) w_{S \setminus \{i\}}(j) & \text{aksi durumda} \end{cases} \quad (8)$$



Şekil 2: Collective Merging ile önerilen yöntemin Collective Motion Database içerisindeki altı video sekansı üzerinden (Her sekans için iki satır halinde, üst satır önerilen yöntem, alt satır [6] olmak üzere) karşılaştırılması. Her iki yöntem için de her bir grup farklı bir renk tonu ile gösterilmiştir. Önerilen yöntem için kahverengiye yakın renkler, hızı sıfır olan noktaları göstermektedir. Ayrıca her bir video için sırasıyla (üstten alta) yönelim dağılımının çeşitliliği ile görece ters orantılı olarak artan veya azalan genel uyumluluk değerleri (0.397, 0.363, 0.425, 0.420, 0.415, 0.479) hesaplanmıştır. Collective Merging yönteminin hesapladığı topluluk değerleri sırasıyla (0.373, 0.174, 0.512, 0.463, 0.478, 0.431) şeklindedir. Son olarak, Collective Merging kapsamında, sekanslar için deneklerden toplanmış ortalama referans değerleri (1.50, 0.80, 1.90, 1.00, 1.40, 1.90) şeklindedir.

fonksiyonu ile ölçülebilir. $S \subseteq V$ kümesinin baskın bir küme olabilmesi için S 'nin $\forall i \in S, w_S(i) > 0$ ve $\forall i \notin S, w_{S \cup \{i\}}(i) < 0$ koşullarını (bir grup içindeki elemanların yüksek benzerliği ve farklı gruplardan elemanların benzemez olması) sağlamalıdır.

[9] nolu çalışmada;

$$f(x) = \mathbf{x}^T \mathbf{A} \mathbf{x}, \quad \mathbf{x} \in \Delta \quad (\text{Uyumluluk}) \quad (9)$$

fonksiyonunu optimize eden vektörün pozitif elemanlarının oluşturduğu kümenin, $\sigma(\mathbf{x}) = \{i \in V | x_i > 0\}$, bir baskın kümeye denk olduğu ve bu amaçla Çoğalma dinamiği kullanılabilirliği gösterilmiştir.

Denklem (9)'da A, G çizelgesinin komşuluk matrisini, $\Delta = \{x \in \mathbb{R}^n | x \geq 0 \text{ ve } e^T \mathbf{x} = 1\}$, e tamamı birlerden oluşan vektör olmak üzere \mathbb{R}^n deki yapıyı göstermektedir.

Bu koşullar üzerinde yöntem, her seferinde belirli sayıda noktayı bir gruba atayıp (baskın bir grup), geriye kalan atanmamış diğer noktalar üzerinden tekrar yeni bir benzerlik matrisi oluşturup aynı işlemi tekrarlamaktadır. Buna, soyma stratejisi denilmektedir. Baskın Kümeler yönteminin avantajları her gruba özel bir ait olma değeri (uyumluluk) çıkartması ve veri içinde olabilecek aykırı noktalara ve gürültülere karşı esnek olmasıdır.

III. DENEYSSEL SONUÇLAR

Bu çalışma kapsamında, iki yöntemin görsel kıyaslaması için Collective Motion Database kullanılmıştır. K komşuluk değeri 20 olarak kullanılmıştır. Collective Merging ile ilgili görsel sonuçlar Şekil 2'de gösterilmiştir. Sonuçlarda görüldüğü gibi Baskın Kümeler yöntemi, koyu kahverengi renkli noktaları aykırı nokta olarak hesaplamıştır. Her bir video sekansı için ayrıca genel bir uyumluluk değeri ;

$$\text{Genel Uyumluluk} = \frac{1}{n} \sum_{g \in G} \mathbf{x}_g^T \mathbf{A} \mathbf{x}_g \times |\sigma(\mathbf{x}_g)| \quad (10)$$

(10) denkleminde göre hesaplanmıştır. Burada;

- 1) g , o anki grubu,
- 2) G , hesaplanan tüm grupları,
- 3) $\mathbf{x}_g^T \mathbf{A} \mathbf{x}_g$, o anki grup için belirlenen uyumluluk değerini, (Denklem (9))
- 4) $|\sigma(\mathbf{x}_g)|$, o anki gruba atanan nokta sayısını,
- 5) n , bütün nokta sayısını,

belirtmektedir.

Görsel performans olarak Baskın Kümeler yöntemi, Collective Merging yöntemine göre yakın sonuçlar vermiştir. Sadece son videoda bariz şekilde, Collective Merging yönteminin performans olarak öne çıktığı görülmektedir. Ayrıca Şekil 2'de verilen genel uyumluluk değerleri incelendiğinde, bu değerlerin ilgili video sekanslarındaki yönelim hareketlerindeki çeşitlilik arttıkça, görece azaldığı görülmektedir.

Örnek verecek olursak, üstten ikinci videodaki hareket grupları diğer videolardaki hareket gruplarına göre daha fazla dağılım gösterdiği için bu videonun genel uyumluluk değeri, diğer videolara göre daha düşük olarak bulunmuştur. Aynı şekilde alttan birinci videodaki hareket grupları diğer videolardaki hareket gruplarına göre daha düşük dağılım (toplu bir yarım daire hareketi) gösterdiği için bu videonun genel uyumluluk değeri, diğer videolara göre daha yüksek olarak bulunmuştur.

Ayrıca ilgili şekildeki son iki sekans incelendiğinde, üst

sekansın alt sekansa (yarım daire hareketi) göre daha yüksek bir yönelim dağılımı gösterdiği gözlemlenebilir. Bu gözlemi destekleyecek bir şekilde, ilgili iki sekans için yöntemimizde hesaplanan genel uyumluluk değerleri, Collective Merging çalışması kapsamında insan deneklerden toplanan ortalama referans değerlerine paralel olarak üst sekans ve alt sekans için sırasıyla artmaktadır. Fakat Collective Merging çalışması kapsamındaki topluluk değerleri referans değerlerin aksine, üst sekans ve alt sekans için sırasıyla azalmaktadır. Buna göre genel uyumluluk değerinin videolardaki kalabalık yönelim dağılımını daha iyi bir şekilde temsil ettiği söylenebilir.

IV. SONUÇ

Bu çalışma kapsamında farklı bir gruplama yöntemi ve görece farklı benzerlik ölçütleri ile dinamik sahnelerdeki kalabalık insan hareket grupları çıkartılmış ve önerilen yöntem, Collective Merging yöntemi ile genelde yakın bir performans sergilemiş ve video sekansları üzerindeki aykırı noktaları belirlemiştir.

Daha sonraki çalışmalarda farklı benzerlik ölçütleri ile noktalar arasındaki komşuluk ilişkilerini daha iyi bir şekilde yansıtmaya öngörülen Seyrek Altuzay Öbeleme (İng. Sparse Subspace Clustering) [11] yönteminin, Baskın Kümeler yöntemi ile bir arada kullanılması planlanmaktadır. Sparse Subspace Clustering yöntemi bu ilişkiyi, noktaların yaşadığı altuzayları hesaba katarak, özgün bir eniyileme problemini çözerek temsil etmektedir. Ayrıca ilgili veri kümesi içindeki videolardaki, çeşitli yönelim dağılımına bağlı olarak, Baskın Kümeler yönteminin hesapladığı uyumluluk değerleri ile [6] çalışmasındaki topluluk değerleri arasındaki ilişkinin analizi planlanmaktadır. Bu analizin amacı, iki değer arasında doğrudan bir eşitlik aramaktan ziyade, yönelim dağılımlarını farklı olarak temsil eden bu iki değer arasında karşılaştırılmasıdır.

KAYNAKÇA

- [1] M. Hu, S. Ali, M. Shah "Learning Motion Patterns in Crowded Scenes Using Motion Flow Field." ICPR 2008: 1-5.
- [2] H. Wang, A. Kläser, C. Schmid, C. L. Liu "Action Recognition by Dense Trajectories." CVPR 2011: 3169-3176.
- [3] M. Jain, H. Jégou, P. Bouthemy "Better Exploiting Motion for Better Action Recognition." CVPR 2013: 2555 - 2562.
- [4] P. Matikainen, M. Hebert, R. Sukthankar "Trajectons: Action recognition through the motion analysis of tracked features." Workshop on VOEC, ICCV, Sep.2009.
- [5] B. Zhou, X. Tang, and X. Wang "Coherent Filtering :Detecting Coherent Motions from Crowd Clutters." ECCV 2012.
- [6] B. Zhou, X. Tang and X. Wang "Measuring Crowd Collectiveness." CVPR 2013.
- [7] A. Abrams, J. Tucek, J. Little, N. Jacobs, R.Pless "LOST: Longterm Observation of Scenes (with Tracks)" WACV 2012.
- [8] C. Tomasi, T. Kanade "Detection and Tracking of Point Features." IJCV 1991.
- [9] M. Pavan, M. Pelillo "Dominant Sets and Pairwise Clustering." TPAMI 2007: 29(1): 167-172.
- [10] V. Mahadevan, N. Vasconcelos "Spatiotemporal Saliency in Dynamic Scenes." TPAMI 2010.
- [11] E. Elhamifar, R. Vidal "Sparse Subspace Clustering: Algorithm, Theory, and Applications." TPAMI 2013: 35:(1): 2765 - 2781.