Anahtar Nokta Takibi ile Videolarda İnsan Hareketi Tanıma
Human Action Recognition in Videos Using Keypoint Tracking

Yunus Emre Kara, Lale Akarun

Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi, İstanbul
yunus.kara@boun.edu.tr, akarun@boun.edu.tr

ÖZETÇE

Bu çalışmada, bilgisayarlı görme temelli insan hareketi tanıma için yeni bir sistem sunulmaktadır. Önerilen sistem girdi olarak videoları kullanmaktadır. Yaklaşım, hareketin konumuna, ölçek seviyelerine, kişinin görünümüne, kendini örtmeler de dahil olmak üzere kısmi örtmelere ve bir takım görüş açısı değişikliklerine karşı değişimsizdir. Süre değişimlerine karşı gürbüzdür. Anahtar noktalar zaman boyunca takip edilmekte ve takip edilen anahtar noktaların gezingeleri videodaki insan hareketini yorumlamak için kullanılmaktadır. Ardından, videolardan öznitelikler çıkarılmaktadır. Gezingeyi tanımlamak için bir grup öznitelik önerilmektedir. Gezingeler, bu öznitelikler kullanılarak öbeklenmektedir. Öbeklenen gezingeler bir imge dizisini tanımlamak için kullanılmaktadır. İmge dizisi tanımlayıcıları, gezinge öbeklerinin düzgelenmiş histogramlarıdır. Son aşamada, önerilen sistem, imge dizilerinin tanımlayıcılarını bir güdümlü öğrenme yönteminde kullanır.

ABSTRACT

In this study, a new system for computer vision-based recognition of human actions is presented. The proposed system uses videos as input. The approach is invariant of the location of the action and zoom levels, the appearance of the person, partial occlusions including self-occlusions and some viewpoint changes. It is robust against temporal length variations. Keypoints are tracked through time and the trajectories of tracked keypoints are used for interpreting the human action in the video. Then, features from videos are extracted. A group of features for describing a trajectory are proposed. Trajectories are clustered using these trajectory features. The clustered trajectories are used for describing an image sequence. Image sequence descriptors are the normalized histograms of the clusters of trajectories. At the final stage, the proposed system uses the descriptors of the image sequences in a supervised learning approach.

1. GİRİŞ

Bilgisayarla görme temelli insan hareketi tanıma videolara hareket etiketleri atama işlemidir. İnsan hareketi tanıma, güvenlik, gözetim, destekli yaşam ve eğlence gibi bir çok alanda uygulaması olan çok aktif bir araştırma konusudur. Erişim kontrolü, kişi tanıma, olağandışılık sezimi ve insan bilgisayar etkileşimi insan hareketi analizinden fayda sağlayan bazı alanlardır.

İnsan hareketi tanıma, zor bir problemdir. Farklı kişiler arasında hareketlerin uygulanması farklı olabilir. Örneğin, bazı insanlar daha hızlı bazıları daha yavaş yürüyebilir. İnsanların adım uzunlukları farklılık gösterebilir. İnsanların vücut yapıları, görünüşleri farklılık gösterebilir. Bazı durumlarda ana harekete bazı yan hareketler eşlik edebilir. Örneğin, yürürken bir engelden kaçınma yürüme hareketine gürültü olarak eklenebilir. Bunlara ek olarak ışıklandırma değişimleri, gölgeler, örtmeler de zorlayıcı durumlar yaratır. Karmaşık veya hareketli arkaplanlar insan tespit etmeyi zorlaştırır. Zamana göre hareketleri bölütlemek ayrı bir zorluktur. Ayrıca değişik yakınlık seviyeleri ve hareketin sahnedeki konumu tanıma etkileyen başka etmenlerdir.

İnsan hareketi tanıma sorununu çözmek için genel olarak iki yaklaşım kullanılmıştır. Global yaklaşımlarda, öncelikle kişinin konumu tespit edilir. Daha sonra ilgi bölgesi bir bütün olarak tanımlanır. Yerel yaklaşımlarda, resimlerdeki çeşitli yamalar ayrı ayrı kodlanır. Bu yamalar genellikle ilgi noktalarının uzamsal komşuluklarından çıkarılır. Gösterim bu yamaların birleşik bilgileri kullanılarak yapılır.

İnsan hareketi takibi ve hareket tanıma alanlarında bir çok yazın taraması yapılmıştır [1]. Bobick ve Davis [2], hareketleri tanımak için hareket enerji imgeleri ve geçmiş hareket imgeleri adında iki yaklaşım önermiştir. Efros ve diğerleri [3], optik akış ölçümleri hesaplayarak hareket tanımada kullanmışlardır. Gorelick ve diğerleri [4], peş peşe gelen çerçevelerdeki insan silüetlerini yığınlayarak uzay zaman hacimleri yapmış ve bu hacimleri hareket tanımada kullanmışlardır. Laptev [5] Harris ilgi noktası sezinleyicisini üç boyuta aktaran uzay zaman ilgi noktası (STIP) adlı yeni yöntem önermiştir. Willems ve diğerleri [6], belirgin noktaları bulmak için toplamsal(integral) videoları kullanmıştır. Sun ve diğerleri [7], SIFT tanımlayıcılarını bularak ilgi noktalarını takip etmiştir.

Daha önceden bahsedilen zorlukların bir kısmının üstesinden gelmek için yerel bir yaklaşım öneriyoruz. Yerel yaklaşımlar görüş açısı değişikliklerine, kişinin görünüşüne ve kısmi örtmelere karşı dayanıklıdır. Kişinin konumlandırılması ve arkaplan çıkarımı gerektirmez. Yerel anahtar noktaları takip edip onların zaman içindeki gezingelerinden anlam çıkarıyoruz. Ek olarak, gezingeleri zamana karşı düzgeleyip hareketin uygulanma hızından bağımsız hale getiriyoruz. Ayrıca, gezingeleri uzamsal konuma göre düzgelememiz değişik yaklaşım seviyeleri ve hareketin sahne üzerindeki konumuna karşı değişimsizlik sağlıyor. Yöntemimizin başarımını, birini kendi topladığımız üç farklı veri kümesi üzerinde sınıyoruz. Veri kümelerinde, her

videoda tek uygulayıcı tarafından gerçekleştirilmiş tek hareket bulunmaktadır.

Bölüm 2'de anahtar nokta takip etme yöntemi anlatılmaktadır. Bölüm 3'te gezingelerden öznitelik özütlenmesi ve bu özniteliklerin videoların tanımlanmasında kullanılması anlatılmaktadır. Bölüm 4'te deneyler ve sonuçlar anlatılmaktadır.

2. ANAHTAR NOKTA TAKİPÇİSİ

Bu kısımda, videolar üzerinde anahtar noktaları takip etme yöntemi anlatılmaktadır. İlk olarak, bir anahtar nokta sezimi yöntemi kullanılarak video çerçevesi üzerindeki anahtar noktalar tespit edilmektedir. Daha sonra bu anahtar noktaların yerel komşuluk yamalarından, bir anahtar nokta tanımlama yöntemi kullanılarak tanımlayıcı vektörler özütlenmektedir. Bu tanımlayıcı vektörlerin yardımıyla, peş peşe gelen çerçevelerdeki anahtar noktalar eşleştirilmekte ve bunların zaman içindeki gezingeleri bulunmaktadır. Her çerçevede bu işlemlerin yapılmasını takiben bazı gürültülü gezingeler elenmekte ve bazı gezingeler de depolanmaktadır. Şekil 1'de yöntemimizle bulunan örnek gezingeler göstermektedir.

Şekil 1: Örnek gezingeler

2.1. Anahtar Nokta Bulma ve Tanımlama

Anahtar noktalar, resimler üzerinde belirli özelliklere sahip noktalardır. Anahtar noktaların yinelenebilirlik özelliği taşıması beklenir. Yinelenebilirlik, aynı noktanın farklı görüş açısı ve ışık koşullarında tespit edilebilmesidir. Anahtar noktalar genel olarak köşeler, t-kavşakları gibi alanlarda bulunur. Yazında anahtar nokta tespit etmek için çeşitli yöntemler önerilmiştir.

SURF(Speeded Up Robust Features) [8] gürbüz bir anahtar nokta bulma ve tanımlama yöntemidir. Standart hali SIFT'ten daha hızlıdır ve yazarları tarafından değişik resim dönüşümlerinde SIFT'ten daha gürbüz olduğu belirtilir. SURF algoritması ile bulunan anahtar noktaların tanımlayıcı vektörleri 64 elemanlıdır. Bu çalışmada anahtar noktalar SURF yöntemi kullanılarak tespit edilmektedir. Anahtar noktaların tespit edilmesini takiben yine SURF yöntemiyle bu noktaların tanımlayıcıları özütlenmektedir. Yeni bulunan her nokta ya daha önceki gezingelerden birine dahil edilir ya da yeni bir gezingenin başlangıcı olur.

2.2. Anahtar Nokta-Gezinge Eşleştirme

Anahtar noktaların bulunması ve tanımlayıcı vektörlerinin özütlenmesini, noktaları mevcut gezingelere eşleştirme işlemi izler.

Öncelikle mevcut her gezinge için sırasıyla, o gezingeye eklenmiş son anahtar noktanın uzamsal komşuluğunda bulunan yeni tespit edilmiş anahtar noktaların kümesi seçilir. Bu kümedeki anahtar noktaların tanımlayıcı vektörleriyle gezingenin tanımlayıcı vektörü arasında Öklid uzaklığı hesaplanır. Bu hesaplanan uzaklıklardan en küçük olan uzaklığın ikinci en küçük olan uzaklığa oranı hesaplanır. Eğer bu oran yeteri kadar küçükse, en küçük uzaklığı veren anahtar nokta ilgili gezingeye eklenmek üzere aday olarak işaretlenir.

Daha sonra benzer şekilde, bu işaretlenen anahtar noktanın uzamsal komşuluğundaki gezingeler kümesi bulunur. Bu gezingelerin tanımlayıcıları ile anahtar noktanın tanımlayıcısı arasındaki uzaklıklar hesaplanır. Benzer bir oran kontrolünden sonra işaretlenen gezinge başlangıçtaki gezingeyle aynıysa, anahtar nokta ile eşleştirilir.

2.3. Gezinge Güncelleme

Her gezinge için yapılan eşleştirme işleminden sonra bazı gezingelerle hiçbir nokta eşleştirilememiş olabilir. Benzer bir şekilde bazı noktalar hiçbir gezingeyle eşleştirilememiş de olabilir. Gezingelerle eşleştirilememiş noktalar bir sonraki çerçeveden itibaren kullanılmak üzere yeni gezingelerin başlangıcı olurlar. Eşleştirilmesi yapılan gezingelerin sonuna ise eşleştirilen anahtar noktalar eklenir.

2.4. Eleme ve Saklama

Her çerçevede yapılan işlemlerin sonuncuları eleme ve saklama işlemleridir. Eleme işlemi belirlenen kurallar dahilinde işi biten gezingeleri arama uzayından çıkartma işlemidir. Değersiz bulunan gezingeler her çerçeveden sonra elenip arama uzayından çıkarılır, bazı gezingeler ise beğenilip daha sonra hareket yorumlamada kullanılmak üzere yedeklenip arama uzayından çıkarılır. Geri kalan gezingeler ise sonraki gelen çerçevedeki anahtar noktaları eşleştirebilmek için arama uzayında bırakılır.

Gezinge eleme işlemi üç aşamadan oluşur. Birinci aşamada, elenecek gezingelere karar verilir. İkinci aşamada, geri kalan gezingeler arasından uzun süredir güncellenmeyen ama kullanım için uygun olanları yedeklenir. Üçüncü ve son aşamada ise, elenecek gezingeler ve yedeklenen gezingeler arama listesinden çıkarılır.

Bir gezingenin, elenmeye aday olması için gezingeye eklenmiş nokta sayısının gerekenden az olması veya gezingedeki noktaları içeren en küçük çevreleyen kutunun köşegen uzunluğunun istenenden kısa olması gerekir. Bu uzunluğun kısa olması gezingenin hareketsiz bir gezinge olduğu anlamına gelir.

1. Bir gezinge bu koşulları sağlıyor ve gezingenin ilk görülmesinden itibaren yeteri kadar zaman geçtiyse elenmesine karar verilir.
2. Bir gezinge bu koşulları sağlamıyor ve gezingenin son görülmesinden itibaren yeteri kadar zaman geçtiyse daha sonra kullanılmak üzere yedeklenir.
3. Gezingenin elenmesine karar verilse de, kullanım için uygun görülüp yedeklense de arama listesinden çıkarılır.

3. VİDEOLARDAN ÖZNİTELİK ÖZÜTLENMESİ

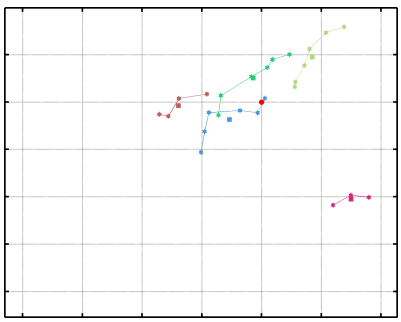
Bulunan gezingelerin sayısı, farklı videolar arasında değişiklik gösterebilir. Yöntemimizde bir videoya bakarak o videodaki olay anlaşılmaya çalışıldığı için farklı videoları birbirleriyle karşılaştırabilecek ortak bir anlayışa ihtiyaç duyulur. Bu kısımda öncelikle gezingeleri zamana ve uzamsal konuma karşı düzgeleme yöntemimiz anlatılmaktadır. Sonrasında ise gezingelerden öznitelik özütlenmesi anlatılmaktadır. Son olarak bu öznitelikler kullanılarak video tanımlayıcılarının özütlenmesi anlatılmaktadır.

3.1. Zamana Karşı Düzgeleme

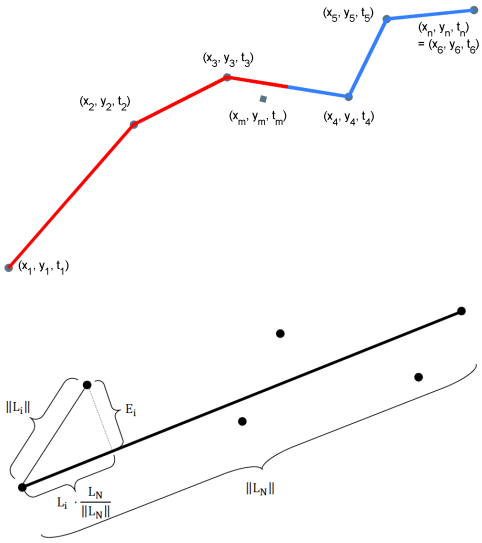
Bir videoda bulunan gezingelerin zamana karşı düzgelenmesi, aynı hareketin farklı hızlardaki çeşitlerine karşı değişimsiz olmasını sağlar. Zamana karşı düzgeleme işleminde, öncelikle gezingeleri oluşturan anahtar noktaların zaman damgalarından videonun başlangıç zaman damgası çıkarılır. Daha sonra elde edilen yeni zaman damgası videonun uzunluğuna bölünür. Böylece video dahilindeki bütün zaman damgaları sıfır ile bir arasına çekilmiş olur. Bu işlem, kararı videonun kare sayısından bağımsız hale getirir.

3.2. Uzamsal Konuma Karşı Düzgeleme

Aynı hareket, farklı videolarda çerçevenin farklı konumlarına düşecek şekilde yapılmış olabilir. Ayrıca farklı yakınlık seviyelerinde hareketin ölçeği farklı görülür. Yöntemimizi konumdan ve yakınlık seviyesinden bağımsız kılmak için uzamsal konuma karşı düzgeleme işlemi yaparız. Bu işlemde öncelikle videodaki bütün gezingeleri oluşturan anahtar noktalar seçilir. Bu noktaların ortalaması ve her iki yöndeki standart sapmaları bulunur. Çerçevenin koordinat sisteminin başnoktası ortalama noktaya ötelenir. Eksen ölçekleri ise ilgili standart sapmaya göre ölçeklenir. Daha sonra anahtar noktaların koordinatları bu yeni koordinat sistemine göre güncellenir. Güncelleme işlemi, anahtar noktaların her iki koordinatından ilgili merkez koordinatının çıkarılması ve daha sonra bu sonucun ilgili standart sapmaya bölünmesi ile yapılır. Şekil 2 uzamsal konuma karşı düzgelenmiş koordinat sistemi örneği göstermektedir. Yeni koordinat sisteminin merkezi kırmızı daire ile gösterilmiştir.



Şekil 2: Uzamsal konuma karşı düzgelenmiş koordinatlar



Şekil 3: Gezinge öznitelikleri

3.3. Gezingelerden Öznitelik Özütlenmesi

Gezingeleri tanımlamak için 10 adet öznitelik kullanılmaktadır. Gezingenin ilk ve son noktalarının düzgelenmiş zaman damgaları ilk iki özniteliği oluşturmaktadır.

Gezingenin düzgelenmiş anahtar nokta koordinatlarının ortalamaları üçüncü ve dördüncü özniteliklerdir.

Gezinge zaman kümsesinde ikiye bölündüğünde oluşan parçaların uzamsal uzunlukları beşinci ve altıncı özniteliklerdir. Gezingenin son noktasından ilk noktası çıkarılınca oluşan vektörün bileşenleri yedinci ve sekizinci öznitelikleri oluşturmaktadır.

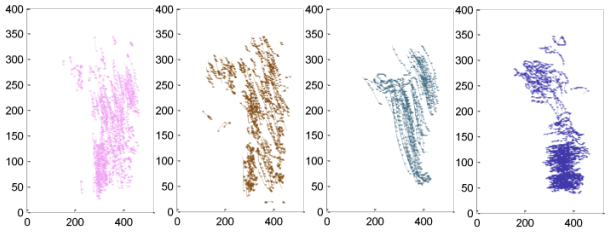
Son olarak gezingenin ilk ve son noktasından geçen doğru bulunur. Diğer noktaların bu doğruya olan uzaklıkları hesaplanır. Bu uzaklıklardan en büyüğü, ilgili noktanın doğruya olan tarafına göre -1 veya +1 ile çarpılır. Bu çarpım dokuzuncu özniteliği oluşturur. En büyük uzaklığı veren bu noktanın doğru üzerindeki izdüşümü bulunur bu izdüşümün ilk noktaya olan uzaklığı bu doğrunun toplam uzunluğuna bölünür. Bu oran onuncu ve son öznitelik olarak kullanılır. Şekil 3 öznitelikleri göstermektedir.

3.4. Gezinge Çoklu Kümesi

Gezingeler öncelikle alt gezingelere bölünür. Alt gezingelere bölme işlemi herbir gezinge için daha önceden belirlenmiş bir kayan pencere genişliğinde yapılır. Bu işlem sonucunda gezingelerden birbirleriyle örtüşen daha kısa alt gezingeler çıkar.

Daha sonra bu alt gezingelerden öznitelik özütlenir. Bu öznitelik vektörleri t-istatistiği kullanılarak düzgelenir. Düzgelenmiş öznitelik vektörleri k-ortalama algoritması kullanılarak öbeklenir. Vektörlerin en yakın olduğu öbek merkezleri bulunur.

Bir video için bu işlemler yapıldıktan sonra video içinde her bir öbeğe düşen alt gezinge sayısı bulunur. Bu sayılar toplam alt gezinge sayısına bölünüp düzgelenir. Bu düzgelenmiş sayıların peş peşe eklenmesiyle oluşturulan k boyutlu vektör videonun tanımlayıcısı olarak kullanılır.



Şekil 4: Örnek gezinge öbekleri

4. SONUÇLAR

Elde edilen video tanımlayıcıları bir güdümlü sınıflandırma yöntemi kullanılarak sınıflandırılır. Bu çalışmada Destek Vektör Makineleri (SVM - Support Vector Machines) kullanılmaktadır. Histogram karşılaştırma için uygunluğundan dolayı χ^2 çekirdeği (kernel) ile bire-karşı-bir (1 vs 1) çok-sınıflı SVM tercih edilmiştir.

Yöntemin başarımı üç farklı veri kümesi üzerinde sınanmıştır. Bu veri kümeleri; KTH İnsan Hareketi veri kümesi [12], Rochester Üniversitesi Günlük Yaşam Aktiviteleri (URADL) veri kümesi [9] ve tarafımızca toplanmış olan WeCare [13] veri kümesidir. WeCare yaşlı bakım sistemleri odaklı yeni bir çok kipli veri kümesidir. Veri kümesinin asıl amacı insan düşmelerinin saptanmasıdır ve düşme hareketiyle karıştırılabilecek yere yatma, zıplama, koltuğa düşme gibi bazı başka hareketler de veri kümesine dahil edilmiştir.

KTH ve WeCare veri kümeleri eğitim, sağlama ve sınama kümeleri olmak üzere üçe ayrılmıştır. URADL veri kümesi ise Messing'in düzeneğine uygun olması için eğitim ve sağlama kümelerine ayrılmıştır. KTH ve WeCare veri kümelerinde 10 katlı çapraz sağlama, URADL kümesinde ise 5 katlı çapraz sağlama yapılmıştır. Verilen sonuçlar, çapraz sağlama sonucunda bulunan en uygun parametreler kullanılarak sınama kümesinde alınan başarımdır. Seçilen parametreler, veri kümelerine göre özelleşmiştir. Alt gezinge uzunlukları 4, SVM masraf parametresi 1.6, k-ortalama parametresi KTH, URADL ve WeCare kümeleri için sırasıyla 6000, 1000 ve 2000'dir.

Önerilen yöntem yazındaki yöntemlerle kıyaslanabilir başarımdadır. KTH veri kümesinde yüzde 87,25 ve URADL veri kümesinde yüzde 88 hatasızlık başarımına sahiptir. WeCare veri kümesinde ise hatasızlığı yüzde 98,75'tir. Ancak, URADL veri kümesinde Messing'in artırılmış hız geçmişi yöntemi [9] hareketin çervede üzerindeki konumuna ve yüzün pozisyonuna bağımlı iken önerilen yöntem kamera konumuna karşı dayanıklıdır.

Tablo 1: Sonuçlar

Veri	Yöntem	Başarım
KTH	Önerilen yöntem	%87,25
	Schuldt [12]	%71,72
	Niebles [11]	%81,5
	Laptev [10]	%91,8
	Messing [9]	%74
URADL	Önerilen yöntem	%88
	Laptev [10]	%59
	Messing (Hız geçmişi) [9]	%63
	Messing (Gizli hız geçmişi) [9]	%67
	Messing (Artırılmış hız geçmişi) [9]	%89
WeCare	Önerilen yöntem	%98,75

5. TEŞEKKÜR

Bu çalışma, 108E161 ve 108E207 numaralı projeler kapsamında TÜBİTAK tarafından desteklenmiştir.

6. KAYNAKÇA

[1] Poppe R., "A survey on vision-based human action recognition", Image and Vision Computing. 2010;28(6):976-990.

[2] Bobick, A. and J. Davis, "The recognition of human movement using temporal templates", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 3, pp. 257-267, Mar. 2001.

[3] Efros, A. A., A. C. Berg, G. Mori and J. Malik, "Recognizing action at a distance", Proceedings Ninth IEEE International Conference on Computer Vision, Vol. 2, No. October, pp. 726-733, 2003.

[4] Gorelick, L., M. Blank, E. Shechtman, M. Irani and R. Basri, "Actions as spacetime shapes.", IEEE transactions on pattern analysis and machine intelligence, Vol. 29, No. 12, pp. 2247-2253, Dec. 2007.

[5] Laptev, I., "On Space-Time Interest Points", International Journal of Computer Vision, Vol. 64, No. 2-3, pp. 107-123, Sep. 2005.

[6] Willems, G., T. Tuytelaars and L. Van Gool, "An efficient dense and scale-invariant spatio-temporal interest point detector", Computer Vision ECCV 2008 , pp. 650-663, 2008.

[7] Sun J, Wu X, Yan S, et al. "Hierarchical spatio-temporal context modeling for action recognition", Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. 2009:2004-2011.

[8] Bay H, Ess A, Tuytelaars T, Gool LV. "Speeded-Up Robust Features (SURF)". Computer Vision and Image Understanding. 2008;110(3):346-359.

[9] Messing, R., C. Pal and H. Kautz, "Activity recognition using the velocity histories of tracked keypoints", IEEE 12th International Conference on Computer Vision, pp. 104-111, Sep. 2009.

[10] Laptev, I., M. Marszalek, C. Schmid and B. Rozenfeld, "Learning realistic human actions from movies", Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pp. 1-8, IEEE, 2008.

[11] Niebles, J. C., H. Wang and L. Fei-Fei, "Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words", International Journal of Computer Vision, Vol. 79, No. 3, pp. 299-318, Mar. 2008.

[12] Schuldt, C., I. Laptev and B. Caputo, "Recognizing Human Actions: A Local SVM Approach", Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., Vol. 3, pp. 32-36 Vol.3, IEEE, 2004.

[13] Alemdar HO, Kara YE, Ozen MO, et al. A robust multimodal fall detection method for ambient assisted living applications. In: 2010 IEEE 18th Signal Processing and Communications Applications Conference. IEEE; 2010:204-207.