

# KARMAŞIK İNSAN HAREKETLERİNİN SIRASAL PİRAMİT EŞLEŞMESİ İLE TANINMASI

## RECOGNITION OF COMPLEX HUMAN ACTIVITIES BY USING SEQUENTIAL PYRAMID MATCHING

Aytaç Çavent

Bilgisayar Mühendisliği Bölümü  
Hacettepe Üniversitesi  
aytaccavent@cs.hacettepe.edu.tr

Nazlı İkizler Cinbiş

Bilgisayar Mühendisliği Bölümü  
Hacettepe Üniversitesi  
nazli@cs.hacettepe.edu.tr

### ÖZETÇE

*İnsan hareketlerinin bilgisayarlar tarafından tanınması konusu son yıllarda gittikçe artan bir ilgiye sahip olmaktadır. Çalışmaların çoğu yürümek, koşmak ve sarılmak gibi basit ve tekrarlı yapısı olan hareketler üzerinde yoğunlaşmıştır, fakat günlük yaşantımız çok çeşitli ve tanınması daha zor olan karmaşık eylemleri içermektedir. Uzay-zaman ilgi noktaları ve kelime kümeleri yöntemlerinin basit hareket tanıma alanında iyi performans sergilediği son çalışmalardan gözlenmektedir. Karmaşık hareketlerin daha ayrıntılı şekilde temsil edilmeleri gerekmektedir. Bu çalışmada, uzay-zaman ilgi noktaları ve kelime kümeleri yönteminin sırasal histogram piramidi kullanımı ile zorlu bir veri kümesi üzerindeki başarıyı sunacağız. Test sonuçlarına göre, önerilen yönteminin karmaşık insan hareketlerini tanımada başarıyı artırdığı görülmektedir.*

### ABSTRACT

*Human action recognition is a topic of increasing interest in recent years. Most of the work is focused on actions that have simple, periodic structure such as walking, running and hugging, but our everyday life contains very different types of actions with challenging problems. Space-time interest points and the bag of words approach have been shown a good performance on action recognition. For more complex activities, finer representations must be employed. In this article, we present the performance of the space-time interest points and bag of words approach by using sequential histogram pyramid on a challenging dataset. According to the test results, it is shown that in complex activity recognition the proposed approach increases recognition performance.*

### 1. GİRİŞ

İnsan hareketleri tanıma yöntemlerinin, insan bilgisayar etkileşimi, güvenlik, spor, otomatik video gruplama gibi çok farklı alanlarda uygulamaları bulunmaktadır. Tüm bu senaryoların önemli zorlukları vardır: karmaşık arka planlar, ilgilendiğimiz cismin kısmen ya da tamamen kapanması, çekim ölçeğindeki değişimler, kıyafet, ışık ve görünümdeki değişimler, hareketin farklı kişilerce farklı şekillerde gerçekleştirilmesi gibi. Bu problemler farklı şekillerde ele alınmalı ve farklı adaptasyon teknikleri gerektirmektedir. Hareketlerdeki çeşitliliğin tamamının bir seferde ele alınamayacak kadar fazla olması, hareketlerin ana olarak üç farklı gruba bölünmesini gerektirmiştir. Literatürde, çok farklı

şekilde gruplamalar yapılsa da kendi aralarında hareketin süresi ve tekrarlı olup olmadığı gibi ortak özellikleri bulunmaktadır. Bu iki özelliği kullanarak hareketleri aşağıdaki gibi gruplandırabilmekteyiz:

- Kısa ve tekrarsız eylemler (ör. sarılmak, içmek)
- Basit ve tekrarlı eylemler (ör. yürümek, koşmak)
- Kısa-tekrarsız ve basit-tekrarlı eylemlerden oluşan karmaşık eylemler.

Hareketler hem zaman hem de uzay boyutunda gerçekleştirildiğinden, basit bir hareket tanıma sistemi için hem zamansal hem de uzaysal öznitelikler kullanılmalıdır. Uzay-zaman ilgi noktaları (STIP[1]), uzay boyutunda yerel olarak yüksek değişime sahip noktaları ve zaman boyutunda da değişken harekete sahip noktaları tespit etmektedir. Tespit edilen noktalar hareketi belirtmesi açısından anlamlı değişime sahip olmaktadır.

Görsel verilerin sınıflandırılması, iyi bir temsil yöntemi gerektirmektedir. HOG (histogram of gradient orientations) ve HOF (histogram of optical flow) betimleyicilerinin yerel görüntü ve hareketin tariflenmesinde oldukça iyi oldukları bilinmektedir[2]. Bu betimleyiciler ilgililenen noktanın komşularına göre yönelimlerini tarif etmektedir. Projemizde STIP yöntemi ile bulunan ilgi noktaları HOG/HOF betimleyicileri ile temsil edilmiştir.

Videoolar içerisinde bazı tekrarlı uzaysal ve zamansal noktaları içerebilmektedir ve bu noktaların kelime kümeleri (bag of words) ilgili videonun bir temsili olarak kullanılabilir. Bu noktaların belirlenmesi kümeleme yöntemlerinin kullanımı ile mümkündür. Bu amaçla, tespit edilen ilgi noktalarında HOG/HOF betimleyicileri üzerinde k-means kümeleme işlemi uygulayarak görsel video kelimeleri bulunmaktadır.

Bulunan görsel kelimelerin histogramı, ilgili videonun küçük boyutlu ve anlamlı noktaları içeren temsili olmaktadır. Basit hareketlere ait videolar genelde tek tip hareket içerdiği için tek bir histogram ile ifade edilebilmektedir. Fakat, birden çok hareketten oluşan karmaşık insan hareketlerini tanımak için, videoların zamansal boyutta daha detaylı olarak incelenmesi gerekmektedir. Bu amaçla, bu çalışmamızda, sırasal piramit histogram eşleşmesi yöntemini kullanmaktayız. Önerilen sistemde, karmaşık hareketler, zamansal boyutta birden çok alt alanın ardışık sıralı histogramları olarak ifade edilmektedir. Karmaşık hareketleri tanımak için, bu sıralı piramit tanımlayıcıları ile destek vektör makineleri (SVM) kullanarak sınıflandırmaktayız.

Deneyisel çalışmalar, Niebles vd.'nin Olimpiyat Oyunları veri kümesi([3]) üzerinde gerçekleştirilmiştir. Bu veri kümesinde 16 çeşit spor dalındaki karmaşık hareketler

YouTube video kaynağı üzerinden toplanmıştır. Bu veri setinde bulunan videolar laboratuvar ortamında çekilmediği için karmaşık arka planlar, kıyafet ve görünümde değişiklikler, başka cisimler tarafından kapanan nesnelere, ışıkta değişimler ve hareketli kamera gibi pek çok güçlükler içermektedir.

## 2. İLGİLİ ÇALIŞMALAR

Literatürde insan hareketi tanıma ile ilgili olarak farklı yöntemler uygulanmıştır. Bu yöntemleri genel olarak üç grupta toplayabiliriz. Birinci grup çalışmalar, hareket tabanlı analiz yöntemini benimsemiştir ve hareket kümelerini (motion clusters) zamansal boyutta kullanarak hareketi tanıma çalışmıştır. Örneğin, Bobick & Davis [4] zaman boyutunda şekil ve hareket içeren şablonlar kullanıp eşleştirme yapmıştır. Little & Boyd[5] yürüme tarzından insan tanıma amaçlı kullanmıştır. Temel hareket öğelerinin zamansal boyutta gösterdiği dizilimin hareketi oluşturduğu düşünülmüştür. İkinci grupta Saklı Markov Modelleri (HMM), Bayes Modeli ve Sonlu Durumlu Makineler kullanılarak hareketlerin yine zamansal boyuttaki dinamik yapısı öğrenilmeye çalışılmıştır. Üçüncü ve güncel olarak uygulanan grupta ise yerel öznitelik noktaları sınıflandırıcılar ile kullanılarak hareket tanıma çalışılmaktadır (ör. Laptev [1], Marszalek and Laptev[6]). Shechtman and Irani [7] denetimsiz öğrenme yöntemini kullanmıştır. Bizim çalışmamız bu üçüncü gruba dahildir, fakat farklı olarak teksele hareketleri değil, karmaşık insan hareketlerini tanımayı amaçlamaktadır.

## 3. ÖNERİLEN YÖNTEM

Karmaşık hareketleri tanıma için önerdiğimiz yöntem şu adımlardan oluşmaktadır:

1. Uzay-zaman ilgi noktalarının bulunması ve bu noktaların HOG/HOF betimleyicileri ile temsil edilmesi
2. K-means kümeleme algoritması ile görsel kelimelerin bulunması ve histogramların elde edilmesi.
3. Her videonun sıralı piramit eşleşmesi (sequential pyramid matching) yöntemi ile ifade edilmesi
4. Her karmaşık hareket sınıfı için bir SVM sınıflandırıcısının oluşturulması

Yukarıdaki adımların her birinde probleme ait olan zorlukların bir kısmı çözülebilmektedir. Örneğin, birinci adımda yerel uzay-zaman noktalarının kullanılması karmaşık arka plan içeren ya da hareketin başka cisimler tarafından görülmesinin engellendiği durumlara karşı önemli bir adaptasyon sağlamaktadır. İkinci adımda kümeleme yönteminin kullanılmasıyla işlenmesi gereken veri miktarı azaltılarak, problem kolay yönetilebilir hale getirilmektedir. Sıralı piramit eşleşmesi yöntemi, karmaşık hareketlerin zamansal boyutta birden çok küçük hareketten oluşmasını gözönüne alarak, daha etkili bir eşleştirme sunmaktadır.

### 3.1. İlgi Noktalarının Bulunması

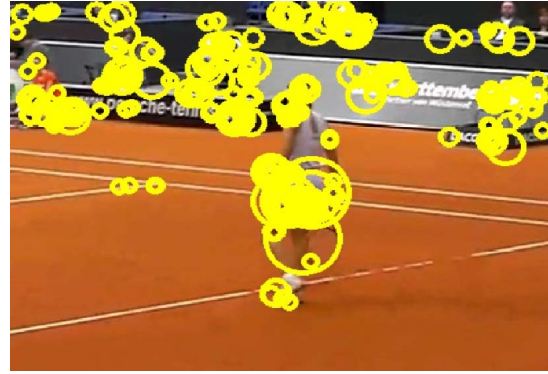
Bilgisayarlı görü alanında yerel ilgi noktalarının kullanılmasının tanıma ve eşleştirmede iyi performans gösterdiği bilinmektedir. İlgi noktalarının yerelliği, karmaşık arka plan ve engellemelerin olduğu durumlarda tanımayı kolaylaştırmaktadır.

Uzay-zaman ilgi noktalarının bulunması Laptev[1] tarafından Harris[8] köşe bulucusunun üzerine zaman boyutunun eklenmesi ile önerilmiştir. Videodaki her nokta için farklı zamansal ve uzaysal ölçeklerde yönelimler

(gradients) bulunmaktadır. Bu yönelimlerden Harris Köşe Bulucu fonksiyonun maksimum olduğu noktalar ilgi noktası olarak belirlenmektedir. Zamansal ve uzaysal ölçeklerin otomatik olarak seçimini [1]'de belirlendiği şekilde yapmaktayız.

Yerel hareket ve görüntünün bilgisayar ortamında modellenmesi için HOG(gradyen histogramları - histogram of oriented gradients) ve HOF(optik akış histogramları - histogram of optical flow) betimleyicilerini, tespit edilen her ilgi noktasını tanımlamak için kullandık. HOG betimleyicisi 72 bileşenden oluşmaktadır ve yerel şekil görünümünü belirtir. Video ilgi noktaları etrafındaki alan  $3 \times 3 \times 2$ 'lik alt bölümden oluşan parçalara bölünerek, her bir alt parça için 4 bileşenli yönelimlerin histogramı hesaplanmaktadır. HOF betimleyicisi ise 90 bileşenden oluşmaktadır yerel hareketi belirtir. Benzer şekilde, video ilgi noktaları etrafındaki alan  $3 \times 3 \times 2$ 'lik alt bölümden oluşan parçalara bölünerek, her bir alt parça için 5 bileşenli hareket yönelimlerin histogramı hesaplanmaktadır.

Tespit edilen ilgi noktalarına ait bir örnek Şekil 1'de verilmiştir. Burda görüldüğü üzere, kamera hareketi nedeniyle, uzay-zaman ilgi noktaları, videonun sadece önplanından değil, arka planından da çıkabilmektedir. Bu durum, gürültü oluşmasına neden olmaktadır.



Şekil 1: Örnek İlgi Noktası Tespitleri

### 3.2. Görsel Kelimelerin Bulunması

İlgi noktaları eğitim aşamasında kullanılan videolar ile test aşamasında kullanılacak olan videolar arası benzerliğin bulunması açısından önemli olduğundan sistemin başarımında önemli bir yere sahiptir. Videoların işlenmesi sonucunda yüksek sayıda ilgi noktası çıkmaktadır (bizim veri kümemizde toplamda yaklaşık 3 milyon adet ilgi noktası çıkmıştır). Bu ilgi noktalarını metin erişim sistemlerinde yapıldığı gibi sık görülen öğelerini (görsel kelimeleri) tespit edip, videoları bu öğelerin bulunma sıklığı ile modellemek etkin bir yaklaşımdır.

Sistemimizde kullanılan görsel kelimelerin sayısını daha önceki hareket tanıma uygulamalarındaki başarımına güvenerek 500 olarak belirledik ve bu görsel kelimelerin tespiti için K-means algoritmasını kullandık. Verilerin tamamının K-means algoritması ile kümelemesi bellek gereksinimi açısından problem oluşturduğundan, verileri altörneklem ile indirgeyerek, rastgele seçilmiş %25'ini kümeleme işleminde kullandık. En düşük hata oranına sahip küme noktalarını bulmak için k-means algoritmasını 200 iterasyonla çalıştırdık, ancak hiç bir adımda tespit edilen hata değerinin sabitlenmediğini gördük. Bu durum bulduğumuz ilgi noktalarının tam anlamı ile belirli gruplar

etrafında toplanmadığını ve bazı gürültü kümelerinin varlığını göstermektedir. Bunun en büyük nedeni, Şekil 1'deki örnekte olduğu gibi çıkarılan ilgi noktalarının gürültü miktarıdır.

Bu aşamanın sonunda her bir video için tespit edilen tüm ilgi noktaları (1) nolu kosinüs benzerlik formülüne göre, kendisine en fazla benzeyen görsel kelime kümesi cinsinden ifade edilmiştir. Burda  $X$  ilgi noktası vektörünü,  $C$  ise bulunan görsel kelime kümelerinin orta noktasına karşılık gelmektedir.

$$\text{benzerlik} = \cos(\Theta) = \frac{X \cdot C}{\|X\| \|C\|} \quad (1)$$

### 3.3. Sıralı Piramit Eşleşmesi

Videoolar, en basit şekilde bulunan görsel kelimelerin tek bir histogramı ile temsil edilip tanıma işleminde kullanılabilir. Ancak karmaşık hareketler, birden fazla hareketin belirli sırada dizilimi ile gerçekleştiğinden, zamansal sırayı gözönüne almak gerekmektedir. Bu amaçla, Lazebnik tarafından gerçekleştirilen [9] 2 boyutlu resimler üzerinde tanımlanmış uzamsal piramit eşleştirme (spatial pyramid matching) yönteminden esinlenen sıralı piramit eşleştirme (sequential pyramid matching) yönteminden faydalandık [10]. Video, zaman boyutunda bölünerek, tek bir histogram yerine farklı zaman dilimlerinden ve ölçeklerden elde edilen histogram piramidi ile ifade edilmektedir. Bu histogram piramidleri, ağırlıklandırılarak ya da ağırlıklandırılmadan eşleştirilebilir. Bu amaçla, videoları zaman boyutunda farklı seviyelerde alt parçalara bölerek her bir alt parça için bir görsel kelime histogramı hesaplanmaktadır ve her alt parça, detay düzeyine göre ağırlıklandırılmaktadır.

Birinci seviyede videonun bütününden 1/8 ağırlığı ile histogram hesaplanmıştır. İkinci seviyede video iki eşit parçaya bölünüp yine 1/8 ağırlığı ile üçüncü seviyede dört parçaya bölünüp 1/4 ağırlığı ile ve son seviyede de video 8 eşit parçaya bölünüp 1/2 ağırlığı ile histogramlar hesaplanmıştır. Bu şekilde toplam 15 tane histogram bir videoyu temsil etmektedir.

Yukarıda anlatılan ağırlıklandırılmış histogram piramidi yöntemine ek olarak videoları yine ağırlıklandırılmış olarak;

- tek bir histogram,
- 2 adet, herbiri videonun bir yarısını temsil eden, histogramın birleşimi,
- 4 adet, herbiri videonun bir çeyreğini temsil eden, histogramın birleşimi
- ağırlıklandırılmamış histogram piramidi,

olmak üzere toplamda beş ayrı yöntem oluşturduk. Bu aşamanın sonunda elde edilen histogramlar ile sınıflandırma işlemi için gerekli olan anlamlı veri sağlanmış olmaktadır.

### 3.4. Sınıflandırma

Sınıflandırma işleminde, elde edilen beş farklı histogram kümeleri üzerinde hareket modelleri öğrenilmektedir. Bu amaçla, destek vektör makineleri (support vector machines-SVM) kullanılmıştır. SVM'lerin bilgisayarlı görü alanındaki en başarılı sınıflandırma yöntemlerinden olduğu bilinmektedir.

SVM'lerde farklı çekirdek (*kernel*) seçiminin sınıflandırmaya etkisi oldukça fazladır. Yöntemimizde video tanımlayıcıları histogramlar olarak ifade edildiği ve genel olarak ortamda ilgi noktalarından kaynaklanan gürültü mevcut olduğu için, çekirdek fonksiyonu olarak histogram keşişim

çekirdeği (*histogram intersection kernel-HIK*)[11] kullanmak daha uygundur. Histogram keşişim çekirdek fonksiyonu  $k(x,z)$ ,  $x$  ve  $z$  vektörleri için (2) nolu formül ile ifade edilmektedir:

$$k(x,z) = \sum_{i=1}^n \min(x(i), z(i)) \quad (2)$$

## 4. DENEYLER

Yaptığımız çalışmayı değerlendirmek için Niebles vd.'nin Olimpiyat Oyunları veri kümesini [3] kullandık. Bu veri kümesinde 16 çeşit spor dalına ait videolar YouTube video kaynağı üzerinden toplanmıştır. Bu veri kümesinde basit ve tekrarlı hareketler yerine karmaşık hareketler bulunmaktadır. Örneğin uzun atlama videolarında önce bir atlet ayakta durmak, sonra koşmakta, zıplamakta, yere inmekte ve yeniden ayağa kalkmaktadır. Çalışmamızda olimpiyat oyunları veri kümesinin her sınıf için 15 adet rastgele seçilmiş ve toplamda da 240 videodan oluşan alt kümesi üzerinde çalıştık.

SVM ikili bir sınıflandırma yöntemi olduğu ve veri kümemizde 16 sınıf olduğu için sınıflandırma öğrenmesi bire karşı hepsi (one-against-all) yöntem ile yapılarak 16 ayrı SVM eğitilmiştir. Sınıflandırma başarımını ölçmek için tek çıkarımlı çapraz doğrulama (leave-one-out cross validation) yöntemi kullanılmıştır.

Tablo 1: Yöntemlerin doğruluk değerleri

Yöntem	SVM Çekirdek Türü	Doğruluk
1 histogram	LK	%44.58
1 histogram	HIK	%55.83
2 histogram	HIK	%55.00
4 histogram	HIK	%54.58
Ağırlıklandırılmış histogram piramidi	HIK	%55.00
Ağırlıklandırılmamış histogram piramidi	HIK	<b>%57.08</b>

Önerilen yöntemin başarımını test etmek amacıyla, öncelikle bir taban çizgisi yöntemi olarak tek histogramlardan oluşan ve SVM çekirdeği olarak lineer çekirdek (LK) kullanan yöntem denenmiştir. Bu yöntemde her bir video tek bir histogram ile ifade edilmiştir. Elde edilen doğruluk değerleri Tablo 1'de verilmiştir.

Tablo 1'den görüleceği üzere, lineer çekirdek ve tek histogramlı yöntemin karmaşık hareket tanımadaki başarımı nispeten düşüktür. Buna oranla, histogram keşişim çekirdeği (HIK) kullanımının daha başarılı sonuçlar verdiğini görmekteyiz (%11.25'lik başarımlı artış). Bunun en önemli nedeni, gürültülü ilgi noktalarının histogram keşişim yöntemi kullanılarak sınıflandırmaya etkisinin azaltılmasıdır.

Yöntemlerin geneline bakıldığında ise, en iyi sonucu ağırlıklandırılmamış histogram piramidi yönteminin verdiği görülmektedir. Buna göre, sıralı histogramların birarada değerlendirilmesinin başarıma katkı sağladığı, fakat farklı bölümlerin farklı ağırlıklarla değerlendirilmesine gerek olmadığı görülmektedir.

Tablo 2: Ağırlıklandırılmamış histogram piramidi yönteminin doğruluk değerleri

GERÇEKLER ↓	BULUNANLAR ↑	Basketball Layup	Bowling	Clean and Jerk	Discus Throw	Diving Platform	Diving Springboard	Hammer Throw	High Jump	Javelin Throw	Long jump	Pole Vault	Shot Put	Snatch	Tennis Serve	Triple Jump	Vault	Doğruluk Oranı
Basketball Layup		6	2	1	0	0	0	0	1	1	0	0	3	0	1	0	0	%40.00
Bowling		0	10	0	0	0	3	0	1	1	0	0	0	0	0	0	0	%66.67
Clean and Jerk		1	0	6	1	0	0	0	0	0	0	1	1	3	2	0	0	%40.00
Discus Throw		0	2	0	4	0	1	4	0	0	0	0	1	3	0	0	0	%26.67
Diving Platform		0	0	0	0	13	0	0	0	0	0	0	2	0	0	0	0	%86.67
Diving Springboard		1	1	1	0	2	6	0	0	1	0	0	1	0	0	2	0	%40.00
Hammer Throw		0	0	0	1	0	0	12	0	0	0	0	1	0	1	0	0	%80.00
High Jump		3	0	0	0	0	0	0	9	0	0	2	1	0	0	0	0	%60.00
Javelin Throw		1	0	0	0	0	0	0	1	8	0	1	1	0	0	1	2	%53.33
Long jump		0	1	0	0	0	0	0	0	0	12	0	0	0	0	2	0	%80.00
Pole Vault		1	0	0	0	1	0	0	4	2	1	4	0	0	0	0	2	%26.67
Shot Put		0	1	1	0	0	0	3	0	0	0	0	9	0	1	0	0	%60.00
Snatch		0	0	1	3	0	2	0	0	0	0	0	0	9	0	0	0	%60.00
Tennis Serve		0	1	1	1	0	1	1	0	0	0	0	1	1	8	0	0	%53.33
Triple Jump		0	0	0	0	0	0	0	0	1	3	1	0	0	1	9	0	%60.00
Vault		0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	12	%80.00

En başarılı yöntem olan ağırlıklandırılmamış histogram piramidi için sınıflar arası karışım(confusion) matrisi Tablo 2’de verilmiştir. En çok karışım, Disk atma(Discus Throw) ve Çekiç atma(Hammer Throw) ile Yüksek atlama(High Jump) ve Sırkla yüksek atlama(Pole Vault) hareket sınıfları arasında görülmektedir. Bu hareketlerin insan vücudu açısından yapısı düşünüldüğünde, aralarındaki görsel farkların çok az olduğu görülecektir ve buna bağlı olarak sınıflandırıcıların yanlışma oranının artması doğal olarak nitelendirilebilir.

## 5. SONUÇLAR

Bu çalışmada uzay-zaman ilgi noktalarının ve kelime kümeleri yönteminin, karmaşık insan hareketlerinin tanınmasında sıralı piramit eşleşmesi ile birlikte kullanıldığında başarılı olduğunu incelenmiştir. Çalışmamızın sonucunda, ağırlıklandırılmamış sıralı histogramların, histogram kesişim çekirdek fonksiyonu kullanılarak SVM sınıflandırmasının daha başarılı sonuç verdiğini gördük. Yöntemimizi oldukça zorlu bir karmaşık insan hareketleri veri kümesinde denedik. Bu gerçek dünya video veri kümesinde uzay-zaman noktası tespitinin kamera hareketinden oldukça fazla etkilendiği gözlenmiştir. Kamera hareketinden kaynaklı olarak bulunan ilgi noktaları sistemin başarımını düşürdüğünden, uzay-zaman noktası tespit algoritmasının kamera hareketinden etkilenmeyecek şekilde geliştirilmesi yöntemin iyileştirilmesi açısından anlamlı olacaktır.

## 6. KAYNAKÇA

- [1] Laptev, I., “On Space-Time Interest Points”, *IJCV* 64 (2005) 107–123G.
- [2] Heng Wang, Muhammad Muneeb Ullah, Alexander Kläser, Ivan Laptev, Cordelia Schmid, “Evaluation of local spatio-

temporal features for action recognition”, *University of Central Florida, U.S.A, 2009.*

- [3] Juan Carlos Niebles, Chih-Wei Chen and Li Fei-Fei. “Modeling Temporal Structure of Decomposable Motion Segments for Activity Classification.”, *ECCV’10 Proceedings of the 11th European conference on Computer vision: Part II.*
- [4] A. F. Bobick and J. W. Davis, “The recognition of human movement using temporal templates,”, *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 257–267, 2001.*
- [5] J. L. Little and J. E. Boyd, “Recognizing people by their gait: The shape of motion”, *Videre, vol. 1, no. 2, pp. 1–32, 1998.*
- [6] Marszalek, M., Laptev, I., Schmid, C. “Actions in context”, *: CVPR, IEEE (2009) 2929–293J.*
- [7] E. Shechtman and M. Irani, “Space-time behavior based correlation,”, *in International Conference on Computer Vision and Pattern Recognition, 2005.*
- [8] I. Laptev and T. Lindeberg, “Space-time interest points”, *In ICCV, 2003.*
- [9] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories”, *In CVPR, 2006.*
- [10] Yong Hong Tian, Menglin Jiang, Luntian Mou, Xiaoyu Rang, and Tiejun Huang, “A multimodal video copy detection approach with sequential pyramid matching”, *ICIP 2011.*
- [11] Subhransu Maji and Alexander C. Berg and Jitendra Malik, “Classification Using Intersection Kernel Support Vector Machines is efficient”, *In Proceedings, CVPR 2008, Anchorage, Alaska.*