Video Action Classification: A New Approach combining Spatio-temporal Krawtchouk Moments and Laplacian Eigenmaps

Imen Lassoued
Team of research SIIVA
RIADI laboratory
University of Tunis El manar
High Institute of computer science (ISI)
Email: lassoued.imen@yahoo.fr

Ezzeddine Zagrouba
Team of research SIIVA
RIADI laboratory
University of Tunis El manar
High Institute of computer science (ISI)
Email: ezzeddine.zagrouba@fsm.rnu.tn

Youssef Chahir
Team of research IMAGE
GREYC LABORATORY
universite de caen, 14000 Caen, France
Email: chahir@info.unicaen.fr

Abstract—Action classification and recognition is a challenging research area that has significant applications in computer vision domain including robotics, video surveillance, human-computer interaction and multimedia retrieval. Action classification domain uses a large variety of approaches. This paper proposes a new approach for video actions classification based on extension of Krawtchouk moments in spatio-temporal domain . In fact, Krawtchouk moments have interesting properties for describing structural and temporal information of a time varying video sequence.The proposed approach is composed of three main steps. First, the original video is transformed into a spatio-temporal volume of images. Then, silhouettes of human in movement are extracted from these images to define a 3D shape. In the third step, higher order spatio-temporal Krawtchouk moments are applied to the obtained 3D shapes and Laplacian eigenmaps is used to achieve dimension reduction for different moments vectors . Finally, we use SVM algorithm and computed descriptors to classify actions in videos. This new approach has been validated on the two video datasets Weizmann and KTH. Experimental results show a good classification rate compared to other approaches using different descriptors.

I. INTRODUCTION

Advancing automatic human-action classification methods is of relevance to both scientific and industrial communities. Applications of action classification algorithms include surveillance, video indexing and summarization, etc. However, despite significant efforts by the computer-vision community, human-action classification is still an open problem. Indeed, many works were interested by events or actions in videos and use it as an index for classification and retrieval. Action classification consists to classify action of human detected in a given video. In this paper, a new approach of video action classification using extension of Krawtchouk moments into spatio-temporal domain is presented. These moments are computed for each spatio-temporal silhouettes volumes witch describe a human action. the remainder of this paper is organized as follows . Section I will present an overview of existing methods for action classification and recognition. In the second section, the different steps of the proposed method based on spatio-temporal krawtchouk moments and Laplacien eigenmmaps will be described. Section III will be dedicated to experimental results and analysis. In the last section, conclusion and perspectives will be given .

II. RELATED WORKS

Human action classification and recognition have been studied extensively in recent years and several techniques have been developed for this purpose.These techniques can be organized in two main categories.The first class is based on global image descriptors while the second class uses silhouette descriptors

A. Global image descriptors based methods

There are many existing works in action recognition which are based on global features of images such as optical flow, gradient histogram and intensity. Zelnik-Manor and Irani [1] used marginal histograms of spatio-temporal gradients at multiple temporal scales to cluster video events. Wu [2] developed an algorithm to automatically extract figure-centric stabilized images from the tracking system. He also proposed to use Decomposed Image Gradients (DIG) which can be computed by decomposing the image gradients into four channels to classify the person's actions. Efros and al[3] have proposed a descriptor based on blurred optical flow measurements and have applied it to recognize actions on ballet, tennis and football datasets. Dollar and al[4] have characterized behaviors through spatio-temporal feature points, in which a behavior was described in terms of the types and locations of feature points present. Song and al[5] have represented an action by 40 curves derived from the tracking results of five body parts using a cardboard people model. Positions, angles and velocities of body parts have been used to establish motion descriptors by Yacoob and al [6]. Ali and Aggarwal [10] have used angles inclinations of the torso, the lower and upper parts of legs as features to recognize activity. For this class of methods, the recognition results depend greatly on the recording conditions. Features tracking is complex due to the

large variability in the shape and articulation of human body.In particular, perfect limb tracking is not yet well solved.

B. silhouettes descriptors based methods

Many researches in action classification are oriented to human silhouette-based methods [11,12]. In fact, human actions can be characterized as motion of human silhouettes sequence over time . Gorelick et al[13] consider human actions as a three-dimension shape induced by silhouettes in the space-time volume and use properties of the Poisson equation solution to extract space-time features such as local space-time saliency, action dynamics, shape structure and orientation. Guo and al[14],have considered action as a temporal sequence of local shape-deformations of object silhouette centroid. Each action is represented by the empirical covariance matrix of 13-dimensional normalized geometric feature vectors that capture the shape of the silhouette tunnel. Kellokumpu et al. [15] proposed a human activity recognition method from sequences of postures. Sminchisescu et al [16] recognized human motions by discriminative conditional random field (CRF) and maximum entropy Markov models(MEMM).they used an image descriptors composed by shape context and pairwise edge features extracted on the silhouette. Recent vision techniques use frequently human silhouette extraction from videos especially in the imaging setting with fixed cameras. Our approach belongs to this class of method which is based on moving silhouette

III. PROPOSED APPROACH

Researchers have exploited invariant moments in pattern recognition [17]. Moments which are scalar quantities used to characterize a function and to capture its significant features. Moments have been widely used in statistics for description of shapes. Literature[17,19]show that Zernike and Krawtchouk moments are the most used in shape representation and recognition. Many studies [19]show that recognition rate is better by using krawtchouk moments than Zernike moments. some works used Zernike moments to describe human silhouettes and video features for action classification [18,20]. In the proposed approach, our contribution is propose a novel descriptor extending 3D Krawtchouk moments computed for space-time silhouettes volume to describe and classify actions in videos.

Architecture of the proposed approach is composed of several steps represented by Fig.1. First, human silhouettes are extracted in the space-time volumes. Then, we compute action descriptor using Krawtchouk moments and Laplacian eigenmaps. In the last step, different videos are classified by SVM algorithm.

A. Space-time volume extraction

To extract silhouettes from a video sequence, we applied a foreground segmentation using the Codebook background subtraction method described in [21,22]. This method has been shown to be robust in handling both foreground camouflage and shadows. This is achieved by separating intensity and chromaticity in the background model. Moreover, the used
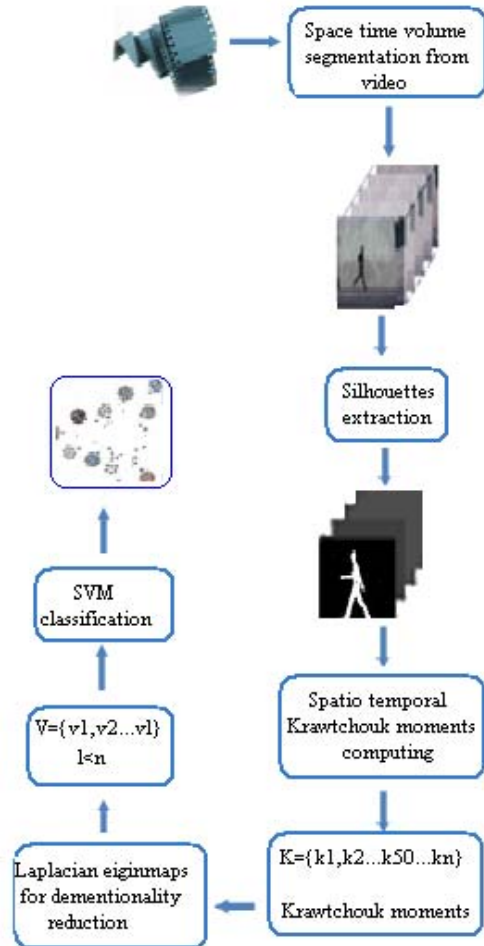

Fig. 1.   General architecture for the proposed approach

background model is multi modal and multi layered which is considered as advantage to model moving backgrounds. To obtain a good background subtraction quality over time, it is essential to update the background model.Fihl and al[21]propose two different update mechanisms to handle rapid and gradual changes. We can use a diverse set of input sequences from both indoor and outdoor scenes by using this robust background subtraction method.

B. Action descriptor computing

We propose a novel descriptor extended of a 3D Krawtchouk moments computed for space-time silhouette volumes extracted from videos in previous step. We first describe the formulation of the Krawtchouk moments in three dimensions, and then introduce the generalization to the space-temporal domain.

1) 3D Krawtchouk moments : The Krawtchouk polynomials play an important role in coding, graph and group theories [27,28,29], they form an important family of orthogonal polynomials. We define the n-th order Krawtchouk polynomial in

the variable x as described in [27].

$$K_n(x; p_x, N_x) = \sum_{j=0}^{n} (-1)^j (p_x - 1)^{n-j} \left(\begin{array}{c} N_x - x \\ n - j \end{array} \right) \left(\begin{array}{c} x \\ j \end{array} \right)$$

(1)

Where $\quad 0 \leq n \leq N_x; p_x \in (0, 1)$

$$\left(\begin{array}{c} x \\ j \end{array} \right) = \begin{cases} \frac{x(x-1)...(x-j+1)}{j!} & \text{if} \quad j \geq 1 \\ 1 & \text{if} \quad j = 0 \end{cases}$$

(2)

The set $S = \{K_n(x; p_x, N_x)\}_{n=0}^{N}$ has $N_x + 1$ Krawtchouk polynomials witch satisfy a discrete orthogonality relation of the form :

$$\sum_{x=0}^{N} w(x; p_x, N_x) K_i(x; p_x, N_x) K_j(x; p_x, N_x) = \Psi(i, j)$$

(3)

Where

$$\Psi(i, j) = h(i; p_x, N_x) \delta_{ij}$$

(4)

Where $i, j = 1, ...N_x$ and $w(x; p_x, N_x)$ is a weight function in x and h is a function depending on i:

$$w(x; p_x, N_x) = \left(\begin{array}{c} N_x \\ x \end{array} \right) p_x (1 - p_x)^{N_x - x}$$

(5)

δ_{ij} is the Kronecker delta function and

$$h(i; p_x, N_x) = \frac{1}{\left(\begin{array}{c} N_x \\ i \end{array} \right)} \left(\frac{1 - p_x}{p_x} \right)^i$$

(6)

$K_0(x; p_x, N_x)$ is the constant polynomial 1 and $K_1(x; p_x, N_x) = N_x(p_x - 1) - p_x x$. For simplicity, we note $K_n(x) \equiv K_n(x; p_x, N_x)$, then we can obtain the following recurrence relation:

$$K_{i+1}(x) = \frac{1}{p_x(N_x - i)} * \Phi(x)$$

(7)

where

$$\Phi(x) = ([p_x(N_x - i) + i(1 - p_x) - x] K_i(x) - i(1 - p_x) K_{i-1}(x))$$

(8)

2) Spatio-temporal Krawtchouk moments : In this work, we are interested in the orthonormality condition, so we define the orthonormal Krawtchouk polynomial by

$$\tilde{K}_n(x; p_x, N_x) = \sqrt{\frac{w(x; p_x, N_x)}{h(n; p_x, N_x)}} K_n(x; p_x, N_x)$$

(9)

Krawtchouk moments, unlike Zernike and Legendre moment, belong to class of discrete orthogonal moments.we

describe our extension of the Krawtchouk moments in (2d+t).Their use discriminate the original behaviours and minimize the computational time and complexity of the classifier. 2d+t Krawtchouk moment of order (n+m+1) of the function f are obtained from the weighted Krawtchouk polynomials as follows:

$$\varsigma_{nml}^{\tilde{v}} = \sum_{x=0}^{N_x} \sum_{y=0}^{N_y} \sum_{t=0}^{N_t} \tilde{K_{n,m,l}} f(x, y, t)$$

(10)

where

$$\tilde{K_{n,m,l}} = \tilde{K}_n(x; p_x, N_x) \tilde{K}_m(y; p_y, N_y) \tilde{K}_l(t; p_t, N_t)$$

(11)

the projection of any 2d+t function f(x,y,t) can be completely reconstructed using the Krawtchouk moments as follows:

$$f(x, y, t) = \sum_{x=0}^{N_x} \sum_{y=0}^{N_y} \sum_{t=0}^{N_t} \tilde{K_{n,m,l}} \tilde{\varsigma_{nml}}$$

(12)

We used these moments as a spatio-temporal feature of any video expressed as a function f(x,y,t). We recall that f(x,y,t) is a discrete binary volume. The parameters N_x, N_y and N_t are in correlation with the dimensions of the binary volume in video. Each video can be expressed by a weighted undirected graph G = (V,E,w), where V ={$v_1, v_2, ..., v_n$ } is the set of all voxels (2d+1) of the video. f(v) is represented by a triplet $\{x, y, t\} \in R^3$ and the ε-neighborhood of v is defined by

$$N_\varepsilon(v) = \{v \in V, f(u) = (\acute{x}, \acute{y}, \acute{t})\}$$

(13)

Where

$$|x - \acute{x}| \leq \varepsilon_x , |y - \acute{y}| \leq \varepsilon_y, |t - \acute{t}| \leq \varepsilon_t$$

Many references [31,32] show that lower orders of Krawtchouk moments allow capturing low spatial frequencies features of different images in video. The lower orders of Krawtchouk moments store informations for a specific region-of-interest in images while higher orders are used to capture high spatial frequencies features. Padam and al [19] and B.Bayrakta and al [30] show that high orders of Krawtchouk moment give more precise information about silhouette reconstruction and classification. That's why, it is interesting to compute higher orders Krawtchouk moments to try to describe more precisely changes in silhouettes poses in extracted volumes. Thus, our descriptor vector f(v) of each

video become :

$$f(v) = \begin{cases} \tilde{\varsigma_{nml}^v} & \text{if } n+m+l \leq s \end{cases} \qquad (14)$$

The next step after computing descriptor vector f(v) for each space time volume is to use Laplacian eigenmaps to find a low-dimensional data representation for each video by preserving local properties of descriptor vector f(v).

C. Laplacian Eigenmaps

Laplacian Eigenmaps (LE) method allow to reduce dimensionality of a given dataset[7]. As with any data reduction method, the problem is that given a set of $x_1,...,x_M$ of M points in R^l ,find a set of point $y_1,...,y_m$ in $R^n (n << l)$ where yi represents xi. Let $y = (y_1,y_2,y_3...y_m)^T$, this objective is to minimize the following function: Π_{ij}

$$\Pi_{ij} = \sum_{ij} W_{ij}(y_i - yj)^2 \qquad (15)$$

Where W_{ij} is defined as:

$$W_{ij} = \begin{cases} \exp \frac{-||x_i-x_j||^2}{2\sigma} & \text{if } i\,\&j\ are\ connected \\ 0 & otherwise \end{cases}$$
$$(16)$$

This weight function W_{ij} ensures that points close to each other assigned a large weight while points further apart assigned a smaller weight. Since this function decreases exponentially, points mapped further apart incur a heavier penalty [7-8]. The LE algorithm has the following steps:

- First step: The Euclidean distance matrix is computed $||x_i - x_j||^2$, then n nearest neighbours are connected. i.e. if node j is among the n nearest neighbours of i then nodes i and j are connected.
- Second step: Weight matrix is computed according to eq.(16).
- Third step: for the connected component, the generalized eigenvalues and eigenvectors are computed.

$$Lf = \lambda Df \qquad (17)$$

Where λ is the eigenvalue and f is the corresponding eigenvector With

$$D_{ii} = \sum_j W_{ij} \qquad (18)$$

Laplacien Matrix

$$L = D - W \qquad (19)$$

- Forth Step: The eigenvectors are sorted according to their eigenvalues:

$$0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq ...\lambda_m \qquad (20)$$

- Final Step: The mapping of x_i into the lower m dimension space is then given by $(f_1(i),...,f_m(i))$. Ignoring the first eigenvector f_0 which corresponds to the eigenvalue 0. There are two parameters to be determined in the Laplacian eigenmaps dimension reduction algorithm, namely n and σ. It is reported in [8, 9] that choosing n=12 and σ =1 give a better cluster separation. Finally, Laplacian eigenmaps dimension reduction was limited to four dimensions .

D. SVM classification

To classify actions in videos, many works are proposed using several attributes such as neurones networks [25], GMM[26], etc. The most used classifier is SVM because it demonstrates efficient results. SVM is a powerful classifier used successfully in many pattern recognition problems. Consider the problem of separating the set of training data $(x_1,y_1),\quad (x_2,y_2),...(x_m,y_m)$ into different classes where $x_i \in R^N$, is a feature vector and y_i are the different classes. Support Vector Machines are designed for binary classification. When dealing with several classes, as in object recognition and image classification, one needs an appropriate multiclass method. Different possibilities has been proposed in[23]. We choose the algorithm named 'one against the others' witch construct n hyperplanes and n is the number of classes. Each hyperplane separates one class from the other classes. This references [23,24] detail the different steps of computing the one against the others SVM multiclass classifier.

IV. EXPERIMENTAL RESULTS

A. Experimentation process

The proposed approach was evaluated on two publicly available benchmark datasets Weizmann and KTH. Experimentation begins by segmenting each video of the dataset in the space-time volume. Then, the associated silhouettes are extracted. Space-time Krawtchouk moments are computed for each silhouette volumes to characterize video. Some preliminary experimentation have been performed and they show that order 100 of Krawtchouk moment was sufficient to have a good description of silhouette volumes. After that, Laplacian eigenmaps was applied for a computed space time Krawtchouk moment to reduce dimension of vectors. In the experiment, we have selected (1/3) of the silhouette series from each action category to form the training set for SVM machine and the remaining videos are used as a test set. The outputs of SVM classifier give the corresponding classes of actions.

B. Experimental results on Weismann dataset

Weizmann dataset is composed by 81 low resolution videos (180 * 144) where 9 persons perform 9 different actions (running, bending, waving with one hand, jumping in place,Jack, jumping, walking, skip, and waving with two hands). Each video clip contains one subject performing a single action. Illustrative examples for each action are shown in Fig.2.

Experimental results using spatio-temporal Krawtchouk moments are shown by the confusion matrix in Fig.3.

Fig. 2. Actions of Weizmann dataset.

Fig.4 presents the confsion matrix using spatio-temporal Krawtchouk moments with Laplacian Eigenmaps. The designations used in the Fig.3 ane Fig.4 are: a1=" walk ", a2=" run ", a3="skip", a4="jack", a5="jump", a6="jump in place", a7="wave with one hand", a8= "wave with two hands", and a9="bend".these results show that spatio-temporal Krawtchouk moments is effectively for actions characterization,we have a good experimental results with a mean accuracy of 90%. In addition, the use of laplacian eigenmaps to reduce dimensionality of the representative vectors improves the classification rate and the mean accuracy becomes 91.55% . Some confusion are observed in the classification results and this can be caused by similarities found between some actions of the dataset. Fig.5 shows experimental results on weizmann dataset for several methods of action classification. The comparison shows that our approach achieves accuracy at 91,55%. So, it is considered among the best methods. This results show the good efficiency of Krawtchouk moment to characterize global space-time shapes.

	$a1$	$a2$	$a3$	$a4$	$a5$	$a6$	$a7$	$a8$	$a9$
$a1$	94	4	2	0	0	0	0	0	0
$a2$	6	94	0	0	0	0	0	0	0
$a3$	0	0	80	3	17	0	0	0	0
$a4$	0	0	0	91	0	0	0	3	6
$a5$	10	4	0	0	86	0	0	0	0
$a6$	0	0	3	0	0	93	0	4	0
$a7$	0	0	0	0	9	0	89	2	0
$a8$	0	0	11	0	0	0	0	89	0
$a9$	0	0	0	0	2	0	0	0	98

Fig. 3. Confusion matrix of Weizmann dataset using Krawtchouk moments.

C. Experimental results on KTH dataset

KTH action dataset is composed by six actions (walking, running, jogging, hand-clapping, hand-waving, boxing) performed several times by 25 actors under four different

	$a1$	$a2$	$a3$	$a4$	$a5$	$a6$	$a7$	$a8$	$a9$
$a1$	97	3	0	0	0	0	0	0	0
$a2$	5	95	0	0	0	0	0	0	0
$a3$	0	0	83	3	14	0	0	0	0
$a4$	0	0	0	90	0	0	0	4	6
$a5$	8	4	0	0	88	0	0	0	0
$a6$	0	0	3	0	0	93	0	4	0
$a7$	0	0	0	0	4	0	92	4	0
$a8$	0	0	12	0	0	0	0	88	0
$a9$	0	0	0	0	2	0	0	0	98

Fig. 4. Confusion matrix of Weizmann dataset using Krawtchouk moments and Laplacian Eigenmaps.

$Method$	$Accuracy(\%)$
$Gorelick \& al\,[13]$	$97,5\%$
$Proposed\,approach$	91.55
$lassoued \& al\,[20]$	90.4%
$Dhillon \& al\,[35]$	88.55%
$Dollar \& al\,[4]$	86.7%
$Niebles \& al\,[33]$	72.8%
$L.Zelnik-Manor \& al\,[36]$	58.91%

Fig. 5. Comparison between weizmann accuracy of different classification methods

scenarios of illumination, appearance and scale change. The dataset contains 2391 video sequences with resolution of 160 120 pixels. Actions of KTH data set are shown in Fig.6.

Fig. 6. Actions of the KTH dataset

	$A1$	$A2$	$A3$	$A4$	$A5$	$A6$
$A1$	100	0	0	0	0	0
$A2$	0	90	8	0	0	2
$A3$	0	5	95	0	0	0
$A4$	0	0	0	78	14	8
$A5$	0	0	0	4	92	4
$A6$	0	0	0	6	4	90

Fig. 7. Confusion matrix of KTH dataset using Krawtchouk moments only

Experimental results using spatio-temporal Krawtchouk moments are shown by the confusion matrix in Fig.7. Fig.8 presents the confsion matrix using spatio-temporal

	$A1$	$A2$	$A3$	$A4$	$A5$	$A6$
$A1$	100	0	0	0	0	0
$A2$	0	93	5	0	0	2
$A3$	0	5	95	0	0	0
$A4$	0	0	0	80	12	8
$A5$	0	0	0	5	95	0
$A6$	0	0	0	0	0	100

Fig. 8. Confusion matrix of KTH dataset using krawtchouk moment and Laplacian Eigenmaps

$Method$	$Accuracy(\%)$
$Kim\,\&\,al$ [34]	95.3%
$Proposed\,approach$	93.83
$Costantini\,\&\,al$ [18]	91.3%
$lassoued\,\&\,al$ [20]	88.7%
$Dollar\,\&\,al$ [4]	81.17%
$Niebles\,\&\,al$ [33]	81.50%
$Schuldt\,\&\,al$ [32]	71.72%

Fig. 9. Comparison between KTH accuracy of different classification methods

Krawtchouk moments with Laplacian Eigenmaps. The following designations: A1=" boxing ", A2=" hand clapping ", A3="hand waving," A4="jogging," A5="running", A6="walking". Results show that, there is many confusions between running and jogging. This can be explained by a quite similar video sequences even naked eye cannot fully distinguish the difference between these actions.

Fig.9 shows experimental results on KTH dataset for several methods of action classification. The presented methods are based on Zernike moments[20,18], histogram representations of the local space-time interest points[4,32]and PLSA[33]. The comparison shows that our approach achieves accuracy at $93,33\%$. So, it is considered among the best methods. This results show the good efficiency of Krawtchouk moment to characterize global space-time shapes.

V. Conclusion

In this paper, we have developed a new approach for video actions classification and characterization based on a compact descriptor for spatio-temporal silhouettes. This descriptor uses a new instance of Krawtchouk moment in spatio-temporal domain to characterize actions of persons. The approach was tested on the KTH and Weizmann datasets and experimental results show a good classification rate compared to other classification methods in the literature. Our future work will deal with evaluation of this method on more realistic datasets. We will also integrate and evaluate the performance of this work in the framework of video indexing and retrieval.

References

[1] L.Zelnik-Manor, and M.Irani, "Event-Based Analysis of Video", Computer Vision and Pattern Recognition, Computer Vision and Pattern Recognition, Proceedings of the 2001 IEEE Computer Society Conference vol 2, pp.123-130, 2001.

[2] X.Wu, "Templated-based Action Recognition: Classifying Hockey Players Movement", Master's thesis, The University of British Columbia, 2005.

[3] A.Efros,C.Breg,G.Mori,and J.Malik, "Recognizing Action at a Distance", Computer Vision, 2003. Proceedings. Ninth IEEE International Conference,vol.2,pp 726 - 733 . 2003.

[4] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features" presented at the Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005.

[5] Y. Song, L. Goncalves, and P. Perona, "Unsupervised learning of human motion," IEEE Transaction on Pattern Analyses and Machine Intelligence, vol. 25,pp. 814-827, 2003.

[6] Y. Yacoob and M. Black, "Parameterized modeling and recognition of activities," Computer Vision on Image Understand", vol. 73,p. 232-247, 1999.

[7] M.Belkin, P.Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation". Neural Computation.pp 1373-96. 2003.

[8] Y. Ghanbari, L. Spence, and P. Papamichalis, A Graph-Laplacian-Based Feature Extraction Algorithm for Neural Spike Sorting, Proc. of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, Minnesota, Sep 2009, pp. 3142-3145.

[9] Y. Ghanbari , P.Papamichalis, L.Spence, "Graph-Spectrum-Based Neural Spike Features for Stereotrodes and Tetrodes". IEEE International Conference on Acoustics, Speech, and Signal Processing. 598-601.2010.

[10] A. Ali and J. Aggarwal, "Segmentation and recognition of continuous human activity," in Procedure of Intelligent Workshop on Detection and Recognition of Events in Video, pp. 28-35, 2001.

[11] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," IEEE Transaction on Pattern Analyses and Machine Intelligence, vol. 23, pp 257-267, 2001.

[12] D. Weinland, R. Ronfard, and E. Boyer, "Motion history volumes for free viewpoint action recognition" presented at the IEEE Workshop Modeling People and Human Interaction, pp 87- 89, 2005.

[13] L.Gorelick, " al Actions as Space-Time Shapes" , Pattern Analysis and Machine Intelligence, IEEE Transactions, vol 29,pp.2247 - 2253,2007 .

[14] K.Guo, P.Ishwar, J.Konrad, " Action Recognition in Video by Covariance Matching of Silhouette Tunnels" , XXII Brazilian Symposium on Computer Graphics and Image Processing, pp.299-306, 2009.

[15] V. Kellokumpu, M. Pietikainen, and J. Heikkila, "Human activity recognition using sequences of postures" presented at the IAPR Conference on Machine Vision Applications, 2005.

[16] C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas, "Conditional models for contextual human motion recognition," in Proc. Int. Conf. Computer Vision, vol. 2, pp 1808-1815, 2005.

[17] J.Flusser, B.Zitova, T.Suk,"Moments and Moment Invariants in Pattern Recognition". Wiley Publishing (2009).

[18] L.Costantini,S.Lorenzo, S.Giuseppe, C.Licia , and B.Alberto Del, "Space-time Zernike Moments and Pyramid Kernel Descriptors for Action Classification" , ICIAP, July 21, 2011

[19] S.P.Priyal, P.K.Bora,"A study on static hand gesture recognition using moments", International Conference on Signal Processing and Communications (SPCOM),septembre 02 2010.

[20] I.Lassoued, E.Zagrouba ,and Y.Chahir, "An Efficient Approach for Video Action Classification Based on 3D Zernike Moments", Future Information Technology;Communications in Computer and Information Science, Volume 185, pp 196-205 juin, 2011.

[21] P.Fihl, R.Corlin, S.Park, T.Moeslund, and M.Trivedi, "Tracking of Individuals in Very Long Video Sequences", In Symposium on Visual Computing, Lake Tahoe, Nevada,USA, November 6-8 2006.

[22] K.Kim, T.Chalidabhongse, D.Harwood, and L.Davis, "Real-time Foreground-Background Segmentation using CodebookModel", Real-time Imaging, vol 11,pp 167-256, 2005.

[23] M.Pontil and A.Verri, Support vector machines for 3-d object recognition, in Pattern Analysis and Machine Intelligence, june 1998, vol. 20

[24] V.Blanz, B.Scholkopf, H.Bulthofi, C.Burges, V.Vapnik and T.Vetter, Comparison of view-based object recognition algo- rithms using realistic 3d models, in Artificial Neural Networks ICANN, Berlin, 1996, pp. 251256.

[25] M.Baccouche, F. Mamalet, C.Wolf, C.Garcia and A.Baskurt, " Une approche neuronale pour la classification d'actions de sport par la prise en compte du contenu visuel et du mouvement dominant ", In Compression et Reprsentation des Signaux Audiovisuels (CORESA), pp.25-30, 2010.

[26] L. Cao,Y. Tian, Z. Liuc, B. Yao, Z. Zhang and S. Thomas Huang, "ACTION DETECTION USING MULTIPLE SPATIAL-TEMPORAL INTEREST POINT FEATURES",IEEE International Conference on Multimedia and Expo (ICME), pp. 340 - 345, July 2010 .

[27] P. T. Yap, R. Paramesran and S. H. Ong, Image Analysis by Krawtchouk Moments,. IEEE Transactions on Image Processing, Vol. 12, No. 11, pp. 13671376, November 2003

[28] W. Gautschi, "Orthogonal Polynomials: Computation and Approximation", Journal of Computational and Applied Mathematics, vol. 178 pp. 215234, 2005.

[29] F. J. MacWilliams, N. J. A. Sloane, The Theory of Error-Correcting Codes, North Holland, Amsterdam, 2003.

[30] B.Bayraktar, T. Bernas, J. Paul Robinson, B. Rajwa, "A numerical recipe for accurate image reconstruction from discrete orthogonal moments", Pattern Recognition, vol 40(2), pp. 659-669, 2007.

[31] A.Saradha,S.Annaduria ,"A Hybrid Feature Extraction Approach for Face Recognition Systems" , ICGST-GVIP Journal, Volume (5), Issue (5), May 2005.

[32] J. D.Shutler, & al.,"Statistical gait recognition via temporal moments", Proc SSIAI, Austin,Texas. 2000.

[33] J.C. Niebles, H. Wang, and L. Fei-Fei. "Unsupervised learning of human action categories using spatial-temporal words", In BMVC, 2006.

[34] K.Tae-Kyun, S .Wong,R.Cipolla,"Tensor Canonical Correlation Analysis for Action Classification",IEEE Conference on Computer Vision and Pattern Recognition, pp 1-8, juillet 2007.

[35] L. Zelnik-Manor and M. Irani, Event-Based Analysis of Video, Computer Vision and Pattern Recognition, pp. 123-130, Sept. 2001.

[36] C. H. Dhillon, P. S. Nowozin, S. Lampert, "Combining appearance and motion for human action classification in videos", Computer Vision and Pattern Recognition Workshops, pp. 22 29 (2009)