Recognizing Human Actions Based on Silhouette Energy Image and Global Motion Description

Mohiuddin Ahmad* and Seong-Whan Lee
Department of Computer Science and Engineering, Korea University, Korea
{mohi, swlee}@image.korea.ac.kr

Abstract

In this paper, we propose a spatio-temporal silhouette representation, called silhouette energy image (SEI), and variability models, to characterize motion and shape properties for automatic recognition of human actions in daily life. To address the variability in the recognition of human actions, several parameters, such as anthropometry of the person, speed of the action, phase (starting and ending state of an action), camera observations (distance from camera, slanting motion, and rotation of human body), and view variations are proposed. We construct the variability models based on SEI and the variability parameters. The global shape-based motions express the spatio-temporal properties of SEI and variability models. Our construction of the optimal model for each action and view is based on the support vectors of motion descriptions of combined action models. We recognize different daily human actions of different styles successfully in the indoor and outdoor environment. Our experimental results show that the proposed method of human action recognition is robust, flexible and efficient.

1. Introduction

Recognition of human actions from multiple views by the classification of image sequences has the applications in video surveillance and monitoring, human-computer interactions, model-based compressions, video retrieval in various situations. Typical situations include scenes with moving or clutter backgrounds, stationary or non-stationary camera, scale variation, starting and ending state variation, individual variations in appearance and cloths of people, changes in light and view-point, and so on. These situations make the human action recognition a challenging task.

The standard approach for human action recognition is to extract a set of features from each image sequence frame and use these features to train classifiers and to perform recognition. Therefore, it is important to consider the appropriateness and robustness of features of action recognition in varying environment. Actually, there is no rigid syntax and well-defined structure for human action recognition available. Moreover, there are several sources of variability that can affect human action recognition, such as variation in speed, viewpoint, size and shape of performer, phase change of action, scaling of persons, and so on. In addition, the motion of the human body is non-rigid in nature. These characteristics make human action recognition a sophisticated task. Considering the above circumstances, we consider some issues that affect the development of models of actions and classifications, which are as follows: (1) An action can be characterized by the local motion of human body parts, (ii) an action can be illustrated by the silhouette image sequence of the human body, which can be regarded as global motion flow, (iii) the trajectory of an action from different viewing directions is different; some of the body parts (part of hand, lower part of leg, part of body, etc) are occluded due to view changes, and (iv) human actions depend on several variabilities, such as anthropometry, method of performing the action, speed, phase variation (starting and ending time of the action), and camera view variations such as zooming, tilting, and rotating.

Among various features, the motion of the body parts and human body shape play the most significant roles for recognition. Motion based features can represent the approximation of the moving direction of the human body and human action can be effectively characterized by motion rather than other cues, such as color, depth, and spatial features. In the motion-based approach, the motion information of the human such as optic flows, affine variation, filters, gradients, spatial-temporal words, and motion blobs are used for recognizing actions. Motion-based action recognition had been performed by several researchers; a few of them are [3, 11, 12, 14]. However, motion-based techniques are not always robust in capturing velocity when motions of the actions are similar for the same body parts. On the other hand, the human body silhouette represents

*Current address : Khulna University of Engg. and Tech., Khulna-9203, Bangladesh, Email:ahmad@eee.kuet.ac.bd

978-1-4244-2154-1/08/$25.00

the pose of the human body at any instant in time, and a series of body silhouette images can be used to recognize human action correctly, regardless of the speed of movement. Different descriptors of shape information of motion regions such as points, boxes, silhouettes, and blobs are used for recognizing or classifying actions. Several researchers performed action recognition using shapes or silhouettes, such as [1, 2]. Bobick and Davis [1] proposed the motion energy image (MEI) and motion history image (MHI) for human movement representation and recognition and were constructed from the cumulative binary motion images. We propose silhouette energy image (SEI) which gives shape information with global motion information but MEI and MHI give only motion information.

In addition of shape and motion, several variabilities that occurred frequently is also responsible for human action recognition. Sheikh and Shah [15] explicitly identified three sources of variability in action recognition, such as viewpoint, execution rate, and anthropometry of actors and they used the 3D space with thirteen anatomical landmarks for each image. In contrast to their work, we explicitly define and employ the anthropometry variation, camera observations (zooming of a person, slanting body, and rotation of human body), phase variation, speed variations and multiple views variation of the action. Related works have typically concentrated on the variability in viewpoint [13] by deriving view invariant features or proposing a view invariant algorithm.

During the action recognition of persons, we utilize the global shape motion features in addition to several variabilities for recognizing the periodic as well as non-periodic or single occurrence actions. The global shape motions are extracted from geometric shape of models. Therefore, based on the combined information of global motion, sources of variabilities, and multiple views, human action recognition is more robust and flexible. We propose to recognize several actions of humans in the daily life from multiple views learning of local and global motion features using the multiclass support vector machine (MCSVM).

In our system, we assume that silhouettes of an image sequence are correctly captured. From the silhouette image sequence, we estimate the temporal boundary (i.e. period or duration) of each action. Depending on the temporal boundary, an action model (i.e. SEI) is constructed by the silhouette image sequence. Using the variability parameters and the SEI, variability models (adaptable models) are generated. The models are characterized by various geometric shape motions. We learn an action for multiple views global motion descriptors by using a MCSVM, and generate SVM models for specified actions. For recognizing actions, we classify (using the similarity of features) descriptions using SVM models. The actions modeling and classification in this work involve both the Korea university full body gesture database (FBGDB) [4] and the KTH database (KTHDB) [10]. Of particular interest is the detection method, which we use for the recognition of several daily actions of elderly people for human-robot interaction (HRI) or similar applications.

This paper is organized as follows: Section 2 presents action representation and variability generation in our system. Section 3 discusses global motion descriptors of combined models. Section 4 presents experimental results and discussions of the selected approaches. Finally, conclusions are drawn in Section 5.

2. SEI and variability models

2.1. Silhouette energy image

Human action is the movement of humans for performing a task within a short period of time. The action may be simple or complex depending on the number of body limbs involved in the action. Many actions performed by humans have cyclic nature and they show periodicity of short duration. Besides, many actions show single occurrence or non-periodic with time frame of specific length (i.e duration). We have considered human actions daily performed which are almost cyclic in nature, either multiple cyclic (period=nT) or periodic actions (different types of walking, running, jogging, etc), and single occurrence (duration=p) or non-periodic actions (bowing, raising the hand, sitting on the floor, etc). Under the above circumstances, it is possible to transform a human action in the spatio-temporal space or $3D$ space, into a $2D$ spatial space, where the $2D$ space contains temporal information. Let us assume $x_t = f(x, y, t)$ is the silhouette image in a sequence at time t, which includes an action under a duration or a period. Therefore, SEI (SEI = $S(x, y)$) is defined by Eq. (1):

$$S(x,y) = \frac{1}{t_e - t_s} \sum_{t_s}^{t_e} x_t, \begin{cases} t_e - t_s = nT \\ t_e - t_s = p \end{cases} \quad (1)$$

Here, t_s and t_e are the starting and ending states of an action. Since the average $2D$ image stores the global motion distribution and orientation of the silhouette images, we can designate this as a SEI. The number of frames in the action depends on the person, time, and type of action. Since, we use the average of the time sequence silhouette images; the normalized variation affects are very low. Figure 1 shows the sample silhouette images with the SEI of the "bowing" action along with the variation of motion. This representation shows the shape as well as motion changes of an action.

The SEI represents an action model (AT), due to the following reason: (1) The energy of a pixel at every point is a result of an action formation. (2) Each silhouette represents the unit energy of a human action at any instance. (3) It determines the energy distribution of an action.

Figure 1. Human action model representation. (a) Silhouette images of bowing action. (b) Silhouette energy image (SEI) as the action model. (c) Variation of motion.

2.2. Variability action models

To consider the diversity of modeling (learning) and classifying actions, we consider multiple variability models (VTs). *Firstly*, we consider the anthropometry variation of an action which represents the width and height variation of a person. Due to these variations, human action recognition should adapt anthropometry. We can define basic eight sets of anthropometric variation by Eq. (2).

$$S(x,y)|_a = \{S(x\pm a, y), S(x, y\pm b), S(x\pm a, y\pm b)\} \quad (2)$$

where a and b is the anthropometric variation parameters. This can be done by resizing the human body using bilinear interpolation method. We can generate a set of anthropometric images which are shown in Fig. 2(a).

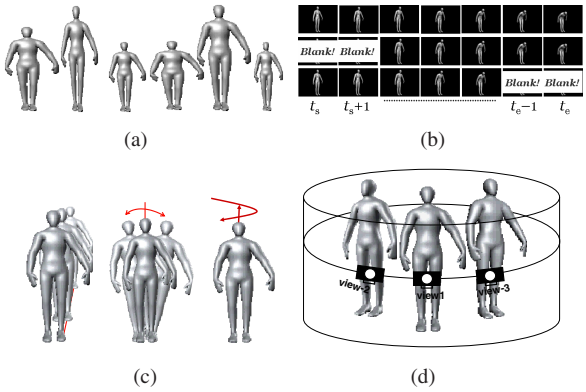

Figure 2. Variabilities of human actions. (a) Anthropometry variation images with different body width and height. (b) Phase variation. Top row: complete action. Middle & bottom row: incomplete action. "*Blank!*" refers to some frames missing at start or end. (c) Camera observations. Left: person's distance from camera (scaling of a person). Middle: slanting position of human body. Right: Human body rotation around upward axis. (d) Multiple views variation.

Secondly, an action can be performed at a different speed (number of frames), which is the number of shape images in an input sequence. By considering temporal transformation, we can adopt an action at different speeds. Two factors are considered for speed variation. These include (i) change of the number of frames and (ii) pixel variations. Following these factors, the speed variability images are modeled by Eq. (3). The expression does not rigorously follow the speed variation, but it approximates the variation of speed of an action.

$$S(x,y)|_s = \begin{cases} S(x,y)(\frac{N}{N+n})\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\pi\sigma^2}}|p > N \\ S(x,y)(\frac{N}{N-n})\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\pi\sigma^2}}|p < N \end{cases} \quad (3)$$

where n is a small time unit and $n \ll N$. Here, p refers to the period of an action and N is the required time for performing an action.

Thirdly, the variable 'phase variation' refers to an action occurred at different starting and ending state. The starting and ending phase of an action depends on persons, time, style, and so on. For example, in the 'bowing' action, a person bends the waist at different angles from the reference position, i.e. from a standing position. Therefore, we can express the phase variability models at starting (ϕ_s) and ending (ϕ_e) by Eq. (4).

$$S(x,y)|_\tau = \begin{cases} \frac{1}{p-\phi_s}\sum_{t_s+\phi_s}^{t_e} x_t, \ \phi_s \text{ varies} \\ \frac{1}{p-\phi_e}\sum_{t_s}^{t_e-\phi_e} x_t, \ \phi_e \text{ varies} \end{cases} \quad (4)$$

In this definition, the parameters ϕ_s and ϕ_e represent the starting and ending phase variation from start and end. Due to phase variation, the starting and ending state of an action changes because of few frames blank (which we can consider incomplete actions). An illustrating situation of phase variation is shown in Fig. 2(b).

Fourthly, at the time of performing an action, the position, orientation, scaling of the persons, and viewpoints can be changed. Therefore, we have considered three kinds of camera parameters variation and they include (1) distance from camera - it refers to the varying scale of the persons body position from camera, (2) tilting motion or slanting motion - human body may in slanting position when a human performs an action, (3) human body rotation - body rotation during the action. The parameters (1) and (2) are modeled by using affine transforms. The parameter (3) variation is modeled by projection geometry. We use affine transformation to simulate a planar shape that undergoes 2D rotation, translation, and scaling. Suppose, a point $\mathbf{x} = (x, y)$ in the coordinate system of shape is affine transformed to a point $\mathbf{x}_a = (x_a, y_a)$ in the imaging plane's coordinate system, then variability models $S(x,y)|_c = S(x_a, y_b)$ from the camera observations are given by Eq. (5).

$$S(x,y)|_c = S\Big(\begin{bmatrix} d+s_x & s_y - r \\ r+s_y & d-s_x \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}\Big) \quad (5)$$

where, d, r, t_x, t_y, s_x, and s_y represent the dilation (scaling or divergence), rotation, translation along x-axis, trans-

lation along y-axis, shear component along-x, and shear-component along-y, respectively. By modeling the coefficient parameters, the diverse representation and learning of actions can be achieved. *Finally*, we consider that an action can be seen from multiple views. Figure 2(c) and (d) illustrate the camera observations and multiple views variation of an action.

3. Global geometric shape motion descriptions

We define adaptable action models or combined models (AAT) as the combination of action models (ATs) and variability models (VTs). Now, we described the geometric shape motions by $\{s_g, s_z, s_x, s_y, s_h, s_p^k, s_o^i\}$. The notations are defined in the following subsections.

3.1. Geometric moments

Moments and function of moments have been utilized as pattern feature in pattern recognition applications. Such features capture global information about the image and do not require close boundaries as required by Fourier descriptors. Hu [6] introduced seven nonlinear functions, h_i, where $i = 1, 2, \cdots, 7$ defined on regular moments using central moments that are translation, scale, and rotation invariant. We use $s_g = \{h_1, h_2, h_3, h_4\}$ as the geometric moment features.

3.2. Zernike moments

The geometric moment shows highly inaccurate results when the image is noisy. Zernike polynomials provide very useful moment kernels, present native rotational invariance and are far more robust to noise. The magnitude of Zernike moments has been treated as global motion descriptors because they are rotation invariant. The 2D Zernike moment of the image intensity function $S(\rho, \theta)$ with order n and repetition m is given in [9].

$$Z_{nm} = \frac{n+1}{\lambda_N} \int_0^{2\pi} \int_0^1 R_{nm}(\rho) e^{-jm\theta} S(\rho, \theta) \rho d\rho d\theta \quad (6)$$

where, ρ = length of vector from origin, θ = angle between vector ρ and x-axis in ccw direction, $0 \leq \rho \leq 1$, λ_N is a normalization factor, and $R_{nm}(\rho)$ is radial polynomial. To achieve scale and translation uniformity, the image function $f(x, y)$ is transformed into $g(x, y)$, where $g(x, y) = f(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y})$ [9] and hence the Zernike moment be $\hat{Z}_{nm}$ with $(\bar{x}, \bar{y})$ being the centroid (COM) of $f(x, y)$ and a is a predetermined value. We use $s_z = \{\hat{Z}_{20}, \hat{Z}_{22}, \hat{Z}_{31}\}$ as Zernike moment features.

3.3. Shape motion distribution

We mentioned that the action model describes the global motion over the duration. For any action, the mean absolute deviation $\mathbf{s}(x, y) = (s_x, s_y)$ of a pixel (x, y) relative to the COM is used for motion description.

$$\mathbf{s}(x, y) = \frac{\sum\sum_{(x,y) \in S(x,y) \geq Th} (\mathbf{x} - \bar{\mathbf{x}}) S(x, y)}{\sum\sum_{(x,y) \in S(x,y) \geq Th} S(x, y)} \quad (7)$$

where, Th represents the threshold value. With this motion description, we can distinguish between actions where more body parts are involved in motion (e.g. sitting on floor, getting down on the floor, lying down on the floor, etc), and an action concentrated in a smaller area where only small parts of the body move (e.g. sitting on a chair, bowing, etc). Another important feature for describing global motion is the mean intensity of motion, s_h of a pixel (x, y). This represents the average absolute height or elevation of motion distribution and is expressed as Eq. (8).

$$s_h = \frac{\sum\sum_{(x,y) \in S(x,y) \geq 0} S(x, y)}{\sum\sum_{(x,y) \in S(x,y) \geq 0} \max S(x, y)} \quad (8)$$

A large value of v_e indicates very intense motion of the human body parts and a small value indicates minimal motion.

3.4. Partial shape motion distribution

It should be noted that the global motion distribution from the center of motion is different for each action, i.e. it results in a different partial global motion distribution for different kinds of geometry. Each partial distribution (layer) has its own characteristics. According to human body shape, either elliptical, quadrant, or block configuration of motion distribution may be suitable. For any kind of geometry, the motion over successive regions are given by the Eq. (9).

$$s_p^k = \frac{1}{n \in B(k)} \sum\sum_{(x,y) \in B(k)} S(x, y) \quad (9)$$

where, $k = 1, 2, \cdots, B$ and B is the number of blocks or quadrants, n is the number of pixels in a block, and p represents ellipse, quadrant, or block. For unique representation of motion, we consider three kinds of partial motion distribution, $s_p^k = \{s_e^k, s_q^k, s_b^k\}$.

3.5. Shape motion orientation

The 2D orientation (direction of major axis, or minor axis) of the motion distribution for every action is different. The global motion orientation is obtained from the eigenvalues λ_i, of the covariance matrix of the action models. Therefore, the relative differences in magnitude of the eigenvalues are an indication of the elongation of the image (action model). We consider the projection of major and minor axis orientation, $s_o^i = \{proj(\lambda_1), proj(\lambda_2)\}$ as global shape features for ATs and VTs.

4. Experimental results and discussion

4.1. Databases

4.1.1 FBGDB

The FBGDB [4] contains 14 representative full body actions in the daily life of 20 performers. In the database, all the performers are elderly persons. The database consists of 2D video data and silhouette data taken at three views: Front view ($v1$), left-side or -45^0 view ($v2$), and right-side or $+45^0$ view ($v3$). The sample images are shown in Fig. 3, where the symbols represent the actions.

4.1.2 KTHDB

The KTHDB is one of the largest databases with sequences of human actions taken over different scenarios [10]. The database contains six types of human actions, performed several times by 25 subjects in four scenarios: outdoors ($s1$), outdoors with scale variation ($s2$), outdoor with different cloths ($s3$), and indoor ($s4$). The sample images are shown in Fig. 4, where the symbols represent the actions.

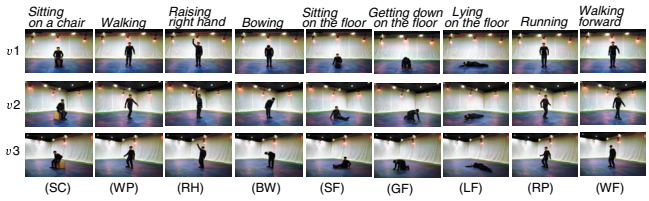

Figure 3. Example images of FBGDB in three views.

Figure 4. Example images of KTHDB in four scenarios.

4.2. Classification results

Support vector machines (SVM) have high generalization capabilities in many tasks, especially in terms of object recognition. In applying MCSVM [5, 16], the training data of the global shape motions of the image sequences are divided into defined classes manually. The MCSVM predicts the class label information for arbitrary actions. Before applying MCSVM, we normalized each component.

	SC	WP	RH	BW	SF	GF	LF	RF	WF
$v1$	1.00	0.57	0.86	1.00	0.71	0.57	1.00	0.57	0.86
$v2$	0.86	0.43	1.00	0.86	0.86	0.86	1.00	0.86	0.86
$v3$	1.00	0.86	1.00	1.00	0.86	1.00	1.00	0.57	0.71
vA	0.95	0.57	0.95	1.00	0.76	0.76	0.90	0.76	0.76

Table 1. CRRs of each action and each view of FBGDB.

Scenario	BP	HC	HW	JP	RP	WP
$s1$	1.00	0.97	0.97	0.91	0.74	0.83
$s2$	1.00	0.89	0.94	0.83	0.61	0.78
$s3$	1.00	0.97	0.98	0.74	0.81	0.81
$s4$	0.95	0.94	0.93	0.81	0.67	0.80

Table 2. Recognition of actions of KTHDB.

Table 1 shows the action recognition results of FBGDB using MCSVM where we use global motions for each view. We use 9 subjects, 9 actions, and 4 views variation for testing (vA represents arbitrary view), & 7 subjects, 9 actions, and 4 view are used for training. The correct recognition rate (CRR) for any action is calculated as follows:

$$CRR(\%) = (N_c/N_a) \times 100 \tag{10}$$

where, N_c is the total number of correct recognition sequences while N_a is the number of total action sequences. As can be seen, there is a clear separation among different kinds of actions. The overall CRRs of $v1$, $v2$, $v3$, and vA are 79.34, 84.12, 90.47 and 82.53 respectively, of FBGDB. We use AATs to evaluate the performance.

We also have tested our approach by using the KTHDB, since it is one of the largest human action databases and several researchers used this database. We have tested 8 subjects, 6 actions, and 4 scenarios and each scenario contains 2 or 3 action sequences and 7 subjects, 4 actions, and 4 scenarios are used for training. Table 2 shows the recognition of each action in various scenarios for global shape motions. The CRRs of $s1$, $s2$, $s3$, $s4$, sA are 90.33, 84.17, 88.50, 89.3, 88.50 respectively for KTHDB, where sA is arbitrary scenarios.

It is important to mention that in some cases, the motion of the elderly persons is similar. In our method, it is shown that by using the 2D action model with variability selection, the action recognition is more robust, since we use the natural actions of humans, with emphasis on elderly persons (FBGDB). The movement of elderly person's is significantly different than that of young people. For example, the speed and style of walking and running of elderly people are very similar. We test the system performance without generating adaptable models, and we make a comparison of performance among AT, VT, and AAT. As an example, the performance (in CRR) of AT, VT, and AATs are 80%, 84.33%, and 90.33% respectively. The performance of AAT is significantly better than AT. Moreover, we present

Input features	No.	CRR
Geom. & Zernike moments (s_g, s_z)	7	0.753
GM dist. & orien. (s_x, s_y, s_h, s_o^i)	5	0.743
Quadrant motion (s_q^k)	4	0.815
Elliptical motion distribution (s_e^k)	8	0.802
Partial block-motion (s_b^k)	20	0.817
Overall ($s_g, s_z, s_x, s_y, s_h, s_p^k, s_o^i$)	44	0.903

Table 3. Performance of each kind of feature set ($s1$ scenario).

the performance of each feature set on test samples which is shown in Table 3. Different feature sets contribute in different proportions. The partial global motion distributions show the good performance. The combined features considerably improve the performance and characterize the shape geometry in multiple-view points.

Method	CRR	Scenario
Dollár et al. [3]	81.17	$s1$+$s2$+$s3$+$s4$
Jiang et al. [7]	84.43	$s1$+$s2$+$s3$+$s4$
Ke et al. [8]	62.96	$s1$+$s2$+$s3$+$s4$
Niebles et al. [12]	81.50	$s1$+$s2$+$s3$+$s4$
Schüldt et al. [14]	71.72	$s1$+$s2$+$s3$+$s4$
Our method	88.50	$s1$+$s2$+$s3$+$s4$
Schüldt et al. [14]	62.33	$s2$
Our method	84.17	$s2$

Table 4. Comparison results of the action recognition.

We compare our works with some state-of-art action recognition approaches by using the same database and similar test sequences but different methods. For example, we compare our method with [3, 7, 8, 12, 14] using KTHDB. Our results by global shape motions flow are compared with their results by spatio-temporal filters, volumetric features, spatio-temporal words, and local space time features. The overall comparison of different methods is listed in Table 4. Compared to the mentioned researches, our approach yields best recognition results.

5. Conclusions and further research

This paper proposed a novel method for multiple views human action recognition using the SEI with variable action models. The variabilities provided a more natural and robust environment for human action recognition, using an advanced human-machine interface due to consideration of multiple factors. Moreover, by adapting the variability, incomplete actions and partial occluded actions were recognized successfully. With adding multiple global motion descriptions, the action recognition becomes sparse and flexible and it can be adapted to practical applications of human movement, human action recognition, and so on. Due to 2D representation of human actions, the current limitation of our experiment is to compare the direction recognition. Our future work will include the precise detection and recognition of actions with direction indication, improvement of all modules including adaptability.

Acknowledgements

This research was supported by the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Knowledge Economy of Korea.

References

[1] A. F. Bobick and J. W. Davis. The recognition of human movement using temporal templates. *IEEE Trans. on PAMI*, 23(3):257–267, 2001.

[2] S. Carlsson and J. Sullivan. Action recognition by shape matching to key frames. In *IEEE Workshop on Models vs. Exemplars in CV*, pages 263–270, 2002.

[3] P. Dollár, G. C. V. Rabaud, and S. Belongie. Behavior recognition via sparse spatio-temporal filters. In *IEEE Workshop VS-PETS*, pages 65–72, 2005.

[4] FBGDB. http://gesturedb.korea.ac.kr/.

[5] C.-W. Hsu and C.-J. Lin. A comparison of methods for multiclass support vector machines. *IEEE Trans. on NN*, 13(2):415–425, 2002.

[6] M.-K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, (8):179–187, 1962.

[7] H. Jiang, M. S. Drew, and Z. N. Li. Successive convex matching for action detection. In *CVPR*, volume 2, pages 1646–1653, 2006.

[8] Y. Ke, R. Sukthankar, and M. Hebert. Efficient visual event detection using volumetric features. In *ICCV*, pages 166–173, 2005.

[9] A. Khotanzad and Y. H. Hong. Invariant image recognition by zernike moments. *IEEE Trans. on PAMI*, 12(5):489–497, 1990.

[10] KTHDB. http://www.nada.kth.se/cvap/actions/.

[11] O. Masoud and N. Papanikolopoulos. Recognizing human activities. In *AVSBS*, pages 157–162, 2003.

[12] J. C. Niebles, H. Wang, and L. Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. In *BMVC*, volume 3, pages 1249–1258, 2006.

[13] V. Parameswaran and R. Chellappa. View invariants for human action recognition. In *CVPR*, volume 2, pages 613–619, 2003.

[14] C. Schüldt, I. Laptev, and B. Caputo. Recognizing human actions: a local svm approach. In *ICPR*, volume 3, pages 32–36, 2004.

[15] Y. Sheikh, M. Shah, and M. Shah. Exploring the space of a human action. In *ICCV*, pages 144–149, 2005.

[16] J. Westons and C. Wtkins. Support vector machines for multiclass pattern recognition. In *European Symposium on ANN*, pages 219–224, 1999.