

---

# On the Development Robust and Fast Algorithm of Action and Identity Recognition

Khawlah Hussein Alhamzah<sup>1</sup>, Tianjiang Wang<sup>2</sup>

<sup>1</sup>Computer Science Department, Basrah University, Basrah, Iraq

<sup>2</sup>School of Computer Science, HUST University, Wuhan, China

## Email address:

Khawlahussein@yahoo.com (K. H. Alhamzah)

## To cite this article:

Khawlah Hussein Alhamzah, Tianjiang Wang. On the Development Robust and Fast Algorithm of Action and Identity Recognition. *American Journal of Networks and Communications*. Vol. 4, No. 5, 2015, pp. 112-118. doi: 10.11648/j.ajnc.20150405.12

---

**Abstract:** Human action recognition and surveillance applications are playing a key important in the present days and took an increasing interest in modern. Since most previous methods strictly limited to action classification in different scenarios and not take attention to human identity that makes an action at the same time. We present a novel and fast algorithm to recognize action and identity in a single framework. We assumed one person makes one action in a video. To identify and training the owner of the video to the classifier, we proposed the watermark embedded as 2-D wavelet transform as binary image, which is contains identity information in the training video. We used these wavelet coefficients as identity descriptors. To represent feature motion representation, we used motion energy image (MEI) and motion history image (MHI) as temporal template of the human actions and Zernike moments to extract shape features of the action from MEI and MHI. In this research, a set of Zernike moment based feature vectors is proposed for human action recognition, which is capture the global properties of an object rather than the local ones. We have composed two different feature vectors by evaluating the variance values of lower order Zernike moments in the four-dimensional Zernike moment space with encouraging experimental results. It has discriminative information that is suitable for classification, especially on related actions, such as running and jogging, that is most previous researches fail to classify them even human vision HVS. Nearest neighbor classifier is used for action and identity categorization. The result of these experiments suggests that this method has a high recognition rate in both action and identity accuracy on KTH data sets.

**Keywords:** Action Recognition, Digital Watermarking, 2-D Wavelet Transform, MEI and MHI, 2-D Zernike Moments, Nearest Neighbor Classifier

---

## 1. Introduction

Human action recognition is an important field in computer vision. The security and surveillance are playing a key role in the present day and took an increasing interest in modern society, especially in light of the challenges facing the world. Therefore, biometrics, such as face, fingerprint and iris cannot be used unless the distance is close and if there is cooperation by the user [1-3]. Action recognition and identity from distance platforms such as surveillance camera is a difficult task. However, previous techniques that heavily relied on distance image face significant difficulties. On the other hand, object identity recognition and object tracking can be closely related problems. In a tracking scenario objects can be represented by their shapes and appearances. However, in tracking, objects are usually considered to have small displacements between observations; Therefore, most

tracking techniques such as Mean Shift and Kalman filter-based tracking make use of this information and search for the tracked objects within small spatial variation limits. Such techniques have proved their efficiency in continuous scenes where disappearances and clutters are minor. However, in the cases of long occlusions, tracking performance considerably decays and it even becomes totally inefficient when discontinuities are inherent in the video. Most of researches focus on gait recognition, which is define the style of people that they walk, but this is face difficulties, since people may change their walking depend on external situations. In this research, the feature vectors based Zernike moments are proposed for an action recognition system. From statistics view, we select four lower order Zernike moments with the highest variance values to form a set of moment feature vectors in the four dimensional Zernike moment space. To test the proposed algorithm, we used a set of KTH video

frames. The rest of the paper is organized as follows: In the next section, we present an overview of the related works. Section three; we describe the proposed method of single framework. Section four illustrates the results and discusses the experiments. Finally, section five presents the conclusions.

## 2. Related Works

Over the past two decades, many studies and searches of action recognition in different scenarios. Some of the current state-of-the-art solutions are either inaccurate or computationally intensive while others require human intervention. Several techniques have been developed in the recent years mainly based on the use of local descriptions of the video frames [1-3]. Following the success of SIFT [4] and HOG in image and scene recognition and classification, several methods of space-time extensions have been proposed. Videos are represented as a collection of descriptors or bag of visual features. Space-time descriptors represent the appearance and the motion features. Laptev et al. [5] have defined a descriptor as a concatenation of histograms of oriented 2D gradients and histograms of optical flow. In order to reduce the computation burden extensions of SURF have been presented in [6]. Scovanner et al. [7] Extended the SIFT to three-dimensional gradients normalizing 3D orientations bins. Klaser et al. [8] proposed to exploit 3D pixel gradients developing a technique based on Platonic solids. Finally Ballan et al. [9] developed an efficient descriptor, describing the spatial and temporal components and creating separated histograms of 3D gradient orientations. Moussa et al. [10] proposed an enhanced method for human action recognition by using fine tuning to reduce the number of interest point detectors that produced from SIFT algorithm, they get high accuracy of action recognition. Khawlah et al. [11] proposed identity as watermark with SIFT descriptors. However all of these descriptors are extremely high-dimensional, in terms of time and space complexity and often retain redundant information. Bobick and Davis [12] present a human action recognition algorithm with two temporal templates, named motion energy image (MEI) and motion history image (MHI). In the same time, researchers have exploited moments and invariant moments in pattern recognition [12]. Moments are scalar quantities used to characterize a function and to capture its significant features and they have been widely used for hundreds of years in statistics for description of the shape of a probability density function.

The first attempt by Hu [13], which is used moments for image processing and analysis of seven order moments, that is robust to scaling, transformation. The recent papers in [14] Khan et al. used Hu invariant moments and neural networks to train and recognize different actions, the accuracy is high. Besides Hu moment invariants are sensitive to noise, the recognition rate is descending sharply in the noisy environment. Okay et al. [15] used 3D Zernike moments, they obtain 3D volumetric data (voxel data), and used the surface information contained in the dataset. In [16] their model, four lower orders of Zernike moments with the highest variance

values are utilized to form a feature vector in a four dimensional Zernike moments to recognize a Chinese characters. In [16] their system used Hu invariant moments to identify aircraft by moment invariants. The Hu moments are more redundancy information in higher order of moments which are not suitable for recognition. In [18] their model used geometric Hu moments to drive invariant moments for gray images. Not much work has been reported in the literature of action and identity recognition. Zhou Zhang et al. [19] address the two problems for recognizing actions and identity. They propose an action and identity representation based on computing rich descriptors from ASIFT key point trajectories which capture more global spatial and temporal information. And they used the object categories depicted using a model developed in the statistical text literature: Latent Dirichlet Allocation (LDA), and combined action recognition and identification in a whole framework. They used trajectory extraction, which is may be not always useful for action recognition. Their method requires more computation and time for recognition and identification. They were separating two problems, at first; they classify action, and then second, assign the identity that belongs to the action. Their procedures take more time to identify the human that makes an action.

## 3. Methodology for Extracting Features of Action and Identity

We want to develop computer vision algorithms that go beyond learning one task at time, and improve the learning features in terms of speed and accuracy. In this paper, we show that our identity and action approach to human action recognition allows us to do a better job in the recognition task and provides many new abilities

### 3.1. Identify the Human by Using a Watermark

Based on the author's method [10], we develop the watermarked method, which the watermark can be embedded into each frame of the video. Digital watermarking [20] is the process of embedding or hiding digital information called watermark into a multimedia product, and then the embedded data can later be extracted or detected from the watermarked product. In this paper, we want to recognize the action from the video frames and also we want to know the owner of the video, therefore, the classifier are training for both action and hidden identity at the same time, without any computational cost in terms of time and space. So we present a single framework for recognize action and identity, as shown in Fig. 1a and Fig. 1b. We assumed that there is a single person performing only one action in the video like KTH dataset. The person identity embedded as a logo binary watermark at each video clip. This watermark based on 2-D wavelet transforms which decomposes a video frame into sub-frames, 3 details and 1 approximation. The 2-D wavelet transform separates the frequency band of a frame into a lower resolution approximation sub-band (LL) as well as horizontal (HL), vertical (LH) and diagonal (HH) detail components.

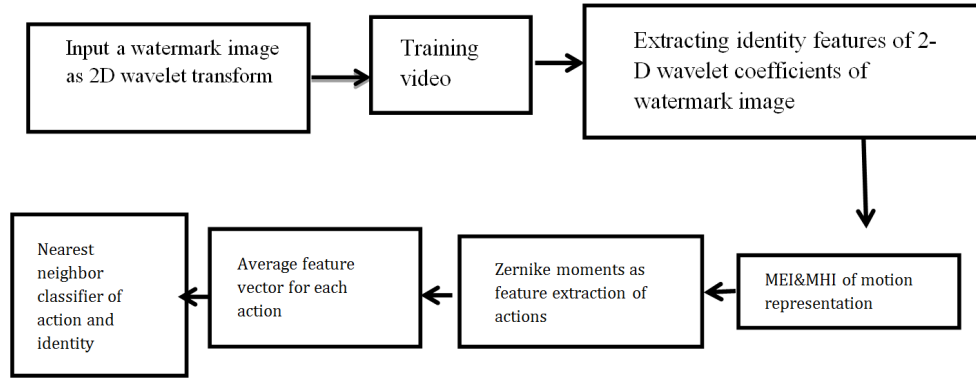


Figure 1a. Single framework of our proposed algorithm for action and hidden identity classification for training video.

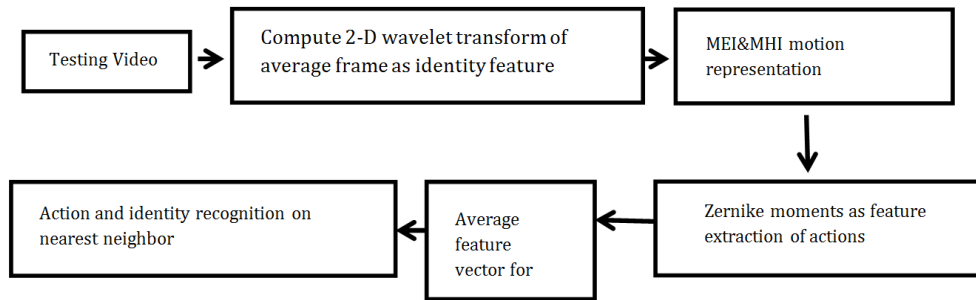


Figure 1b. Single framework of our proposed algorithm for action and identity recognition for testing video.

### 3.2. Embedding Procedure

We assumed the same procedure for embedding watermark in previous methods as follows [20, 21]:

*Step 1:* Convert the  $n \times n$  binary watermark logo into a vector  $W = \{w_1, w_2, \dots, w_{n \times n}\}$  of '0's and '1's.. Out of the four sub bands, only the three high resolution detail sub bands {LH, HL, HH} are selected.

*Step 2:* Pseudorandom sequence is used to embed the zero watermark bit in the selected sub-band as PN sequence of sub band matrix is generated for uniformly distributed.

*Step 3:* A watermark amplification factor  $K$  is embedding with the PN sequence in the selected DWT sub band. If we denote  $W_i$  as coefficients matrix of the selected sub band, then embedding is done according to the equations as follows:

If the watermark bit is 0

$$L_{i,uv} = W_{i,uv} + K \cdot PN_{uv} \text{ where } W \in \{LH, HL, HH\} \quad (1)$$

Otherwise,

$$L_{i,uv} = W_{i,uv} \quad (2)$$

At this step, the coefficients of wavelet transform are describing feature vector identity of the owner's human video that makes an action. Watermarking in the DWT domain can be split into the two procedures: embedding of the watermark and extraction of the watermark. More details about embedding watermark in video by using DWT can be found in [21]. The quality of the video is the same as original as shown in Fig. 2 and Fig.3 respectively.



Figure 2. Frames with watermark embedding as 2-D wavelet in the KTH dataset (not affect the quality).

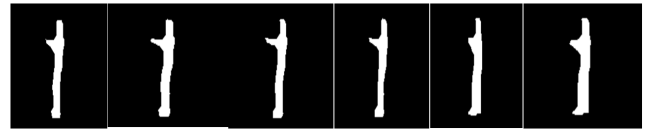


Figure 3. Example of frames that embedded a watermark of boxing action in KTH data set.

### 3.3. Temporal Templates of Human Action

There are many different types of methods to present human actions. These methods mainly can be divided into two kinds: model-based and appearance based. Appearance-based features are commonly used in action recognition since they can be easily and robustly extracted from video. We assumed motion energy image (MEI) and motion history image (MHI) in [2, 12] as the representation of human action.

#### A. Motion-Energy Images (MEI)

MEI is a cumulative motion image of a sequence of binary images. Let  $I(x, y, t)$  be an image sequence and let

$D(x, y, t)$  be a binary image sequence indicating regions of motion; the MEI  $E\tau(x, y, t)$  can be defined as [2, 12]:

$$E(\tau) = \bigcup_{i=0}^{\tau-1} D(x, y, t - i) \quad (3)$$

### B. Motion-History Images (MHI)

To describe how the image is moving, we form a MHI. MHI.  $H^t(x, y, t)$  can be defined as [2, 12]:

$$H^t(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H^t(x, y, t-1) - 1) & \text{otherwise} \end{cases} \quad (4)$$

Here  $\tau$  is time of action, presented by the length of sequence of binary images as shown in Fig.4.

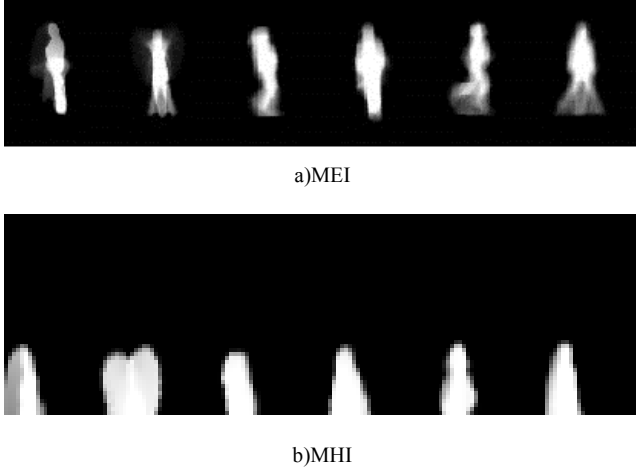


Figure 4. MEI and MHI in KTH dataset.

### 3.4. Extracted Features Using Zernike Moments

For action recognition we use moments, such that moments have been used in image processing and classification. Since Hu introduced them in his groundbreaking publication on moment invariants [22]. Hu used geometric moments and showed that they can be made to be translation and scale invariant. Since then more powerful moment techniques have been developed [23]. A notable example is Teague's work on Zernike Moments (ZM); he was the first to use the Zernike polynomials (ZP) as basis functions for the moments [24]. The use of ZP's as a basis function is theoretically beneficial because they are orthogonal polynomials which allows for maximum separation of data points, given that it reduces information redundancy between the moments, this is very useful in recognition task [25, 26]. Their orthogonal properties make them simpler to use during the reconstruction process as well. Furthermore, the magnitudes of ZM's are rotationally invariant, which is crucial for certain image processing applications, such as classifying shapes that are not aligned. Moments are common in statistics and physics, however, they are often referred to as variance or mean. Hu's moments [13] contain much redundant information about a character's shape. Nowadays, Zernike moments are becoming popular for shape recognition. Zernike moments have been also proposed in action recognition as holistic features in [28] to describe the human silhouettes. We used four lower orders of Zernike moments with the highest variance. Zernike polynomials [29] are a set of complex polynomials in which a complete orthogonal set is formed over the interior of the unit circle. The Zernike basis is a set of complete and orthogonal

functions on the unit disk defined as [16, 30]:

$$V_{nm}(x, y) = R_{nm}(\rho)e^{jm\theta} \quad (5)$$

Where  $\rho = \sqrt{x^2 + y^2}$ ,  $\theta = \tan^{-1}\left(\frac{y}{x}\right)$ ,  $n$  is non negative integer and  $m$  can be positive, negative, or zero

The radial polynomials of Zernike polynomials are:

$$R_{nm}(\rho) = \sum_{k=0}^{(n-m)/2} (-1)^k \frac{(n-k)!}{k! \left(\frac{n-m}{2}-k\right)! \left(\frac{n+m}{2}-k\right)!} \rho^{(n-2k)}, \quad (6)$$

From (3), we can find that  $V_{nm}$  is a complex number when  $m \neq 0$ . It makes Zernike moments different from other moments with real numbers only, we can use several different types of values from the same  $A_{nm}$ .

Where  $\rho$  is the radial distance  $0 < \rho \leq 1$ , and  $R_{nm}$  are the radial polynomials. Zernike polynomials have the property of being limited to a range of  $-1$  to  $+1$ , i.e.  $|Z_n^m(\rho, \varphi)| < 1$ .

The complex Zernike moments of order  $n$  with repetition  $m$ ,  $A_{nm}$ , is defined as [16]:

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x, y) V_{nm}^*(x, y) dx dy, \quad (7)$$

Where  $*$  represents complex conjugate. To compute the Zernike moments of an image, the center of the image is taken as the origin point of the unit circle and only pixels whose centers are located on the unit disk are used in computation. The Zernike moments have several advanced fundamental properties, such as being invariant to rotations and reflections. If an image is rotated  $\sigma$  degree, unlike geometric moments their invariants can be calculated independently to arbitrary high orders without having to recalculate low order invariants, the Zernike moments of the rotated image are:

$$A_{n,m}^\alpha = A_{n,m} e^{-jm\alpha}, \quad (8)$$

From Eq. (6), the magnitudes of  $A_{nm}$  are rotationally invariant.

These orthogonal properties also allow one to evaluate up to what order to calculate these moments to get a good descriptor for a given database. Another reason for using OG moments frequently mentioned in literature is for better image reconstruction [28, 29]. Therefore, the image reconstruction from OG moments can easily compute. We use these values of Zernike moments as action descriptors for the feature vectors, which is contain characteristic information about the contents of each frame in a numerical form. Furthermore, a framework is capable of recognizing the hidden features of identity features and recognizing action features for each defined human action.

## 4. Classification

The nearest neighbor is used in machine learning and pattern recognition. It obtains class membership for some testing feature descriptor based on its nearest neighbor from training feature descriptors in feature space. The testing is classified by a majority vote of its  $K$  nearest neighbors. The KNN [31] will go over all training samples to find the nearest

neighbor for testing sample, therefore, it takes long time, if the training samples are numerous, but in our research, the training data are small. In the KNN, three parameters are used. First parameter, K is set up to number of voting members. Second, distance metric type, we used Euclidean (squared difference). Third, the rule for selecting estimated class for testing sample is set up into: nearest neighbor. The KNN classifier is calculated the distances (d) between testing sample (x) and each training sample (m) provided from Eq. 6, where d is a distance metric, x is a testing sample, m are training samples,  $j = [1, 2, \dots, N]$ , N is a number of training samples [32]

$$D(x, m) = \arg \min_j \{d(x, m_j)\} \quad (9)$$

Where d is Euclidean distance in this research. To classify identity human, the 2-D wavelet coefficients of watermarked image used as feature identity vectors, of video frames of KTH dataset. We divide each video clips of each action into small clips of 0.5 second (13 frames) to reduce the high dimensions of the video, and minimize the complexity time for training. We suggest that it is enough to recognize action and identity, which is discriminate the shape of the motion of video frames as shown in Fig. 4. We split the video frames into training frames and testing frames. We compute real and imaginary parts of Zernike moments of video frames of all actions, and then perform average feature vector of each action. Nearest neighbor classifier used to classify action and identity. The frames are selected which embedded a watermark in video training to learn the features of videos belonging to different classes and labeling is done with action and identity in this step. For each class a number of frames having different styles of the same action performed by same person are taken for training.

## 5. Results and Discussion

In our algorithm, a one watermark embedded as binary image into frames, to identify the human identity. We assumed one person makes one action in each video clip. If we have more complex data sets such as multiple people do actions in same video, we can embedded each watermark for each human appears in the video, this is our future works. The binary human identity is of size 24 x 24; embedded in the training video. This requires a feature representation to describe the human identity as feature vectors of 2-D coefficients of wavelet watermarked image to use for identity classifier. We firstly calculate MEI and MHI of each action using the number of frame for each action as parameter. Secondly calculate the feature vectors of every MEI and MHI. To calculate the Zernike moments of a frame, the center of the frame is taken as the origin point of the unit circle and only pixels whose centers are located on the unit disk are used in computation [29], we calculated real part values, and magnitudes of 4 lowest orders Zernike moments and highest variance for the 6 actions of KTH dataset of feature vector of every MHI and MHI. The results are shown in Table 1, 2, and 3 respectively. Next, select label some actions with hidden

identity, using the constrained nearest neighbor to find minimum Euclidean distance similarity. Then use the same method to extract the feature vectors of test actions. Finally apply the nearest neighbor classifier to recognize test actions and identity. Since, the computation of moments requires that the image is of binary nature [16]. Thus, we not need any additional cost computation, because MHI and MEI temporal template representation provide binary representation of actions. For identity classifier, wavelet transformation provides both frequency and spatial description of an image. Temporal information retained in the transformation process. The confusion matrix in table 5, and table 6 respectively, show high accuracy of action and identity, which is better than [19]. Their work is similar to our paper, but different in methodology.

**Table 1.** Variance values of the 10 lowest orders of Zernike moments  $A_{nm}$  calculated by their magnitudes.

	N=0	N=1	N=2	N=3
m =0	0.00340		0.0083	
m =1		0.00037		
m =2			0.00021	
m =3				0.00043
	N=4	N=5		
m =0				
m =1		0.00073		
m =2	0.00072			
m =3		0.00071		
m =4	0.00176			
m =5		0.00083		

**Table 2.** Variance values of the 10 lowest orders of Zernike moments  $A_{nm}$  calculated by their real parts.

	N=0	N=1	N=2	N=3
m =0	0.00040		0.00216	
m =1		0.00042		
m =2			0.00206	
m =3				0.00435
	N=4	N=5		
m =0				
m =1		0.00135		
m =2	0.00211			
m =3		0.00159		
m =4	0.00173	0.00197		
m =5				

Based on our selection, the following two Zernike moment feature vectors are computed:

$$V_{real} [f1 = A02; f2 = A22; f3 = A33; f4 = A24]; \quad (10)$$

$$V_{magnitude} [f1 = A00; f2 = A02; f3 = A15; f4 = A55]; \quad (11)$$

**Table 3.** Selected four highest Zernike moment values from each part.

	F1	F2	F3	F4
Real part	A02	A22	A33	A24
Magnitude part	A00	A02	A15	A55

When the Zernike moment feature vectors  $V_{real}$  and  $V_{magnitude}$ , as defined in (7) and (8), are applied, we only need to consider the real part or magnitude of  $A_{nm}$ :

$$d(f_i, f_j) = \sqrt{(f_{1i} - f_{1j})^2 + (f_{2i} - f_{2j})^2 + (f_{3i} - f_{3j})^2 + (f_{4i} - f_{4j})^2} \quad (12)$$

Where  $i$  and  $j$  denote any pair of human actions of training and testing of video frames respectively. From the result of accuracy recognition in Table. 4. Khan et al. [14], and Moussa et al. [10], propose methods of action recognition with performing quite high recognition rate, because in [14], they used whole parts of seven Hu invariant moments, and in [10] used normalization after feature detection of interest points, this means more cost in terms of computation time, their method, time is 14.446, although the accuracy is high. While our method use four feature vectors, so our method performs faster than [10] even though calculating more feature vectors.

**Table 4.** Descriptor complexity (in time) comparison together with accuracy (the value results of other methods taken from their papers).

Method	Size	Computation time	Accuracy
Our method	54	0.0334 s	96%
Hu Moments [14]	-	-	97%
Moussa et al. [10]	-	14.446	97%
Pyramid Zernike 3D [11]	84	0.0300 s	91.30%
Gradient + PCA[33]	100	0.0060 s	81.17%
3D SIFT [31]	640	0.8210 s	82.60%
3D Gradn[32]	432	0.0400 s	90.38%
HOG 3D [7]	380	0.0020 s	91.40%
SURF 3D [6]	384	0.0005 s	84.26%

The difficulties of other similar related methods as in [19] are to classify between jogging and running, they used SIFT descriptor for identity representation and ASIFT descriptor for action representation in expensive computing, their method of action accuracy is 81.7 and for identity accuracy is 58.8. As experimental results of our method in recognition accuracy of action and identity are high and the computation time for extracted features 0.033, as shown in Table 4. Unlike the current state of the art approaches such that 3D Zernike, SIFT, HOG, MHI, etc., which is need more computation time to extract Spatio and temporal features. We expect that the property of discriminative Zernike moments and unique coefficients of wavelet values for identity features would increase the efficiency of the proposed action and identity recognition.

**Table 5.** Confusion matrix of action accuracy is 96.5%.

	walk	Jog	run	box	hand-wave	hand-clap
walk	98	2.0	0.0	0.0	0.0	0.0
jog	3.1	93	3.9	0.0	0.0	0.0
run	2.0	3.0	95	0.0	0.0	0.0
box	0.0	0.0	0.0	100	0.0	0.0
hand-wave	0.0	0.0	0.0	0.0	95	5.0
hand-clap	0.0	0.0	0.0	2.0	0.0	98

**Table 6.** Confusion matrix of identity accuracy is 89.7%.

	John	Wang	Ming	said	Hasan	Zane
John	90	5	5	0	0	0
Wang	0	88.4	7.2	4.4	0	0
Ming	0	0	93.4	0	3.5	3.1
Said	0	9.0	0	87.0	0	4.0
Hasan	5.0	5.0	0	0	90.0	0
Zane	3.0	0	0	5.0	2.6	89.4

## 6. Conclusion

In this paper, we have proposed a novel, robust and fast algorithm for action and identity recognition. We identify the human identity by embedding a watermark as binary image of 2-D wavelet transform in the video frames, then extract features identity as wavelet coefficients. To represent the motion features, we use temporal templates MEI and MHI representations. Zernike moments based feature extraction method which captures global features is proposed for human action description. Zernike moment has discriminative features than Hu moments, which are necessary to classify related actions such as jogging and running, which most previous methods fail to classify them. The advantage of using Zernike moments than Hu moments is orthogonal properties, so no redundancy of data, and robust to translation, rotation and scaling. The nearest neighbor classifier can be train to classify different actions of Zernike feature vectors and wavelet coefficients of human identity with higher accuracy in action and identity, without any expensive computation for identifying human identity. The experimental results show that the described system is simple and robust to few illumination changes while dealing with indoor and outdoor actions and only using very small number of labeled training data examples. There is a room to improve and extend this framework to recognize actions in more and complex data sets such that Hollywood data sets.

## Acknowledgements

We would like to thank the anonymous reviewers and editors for their constructive comments and suggestions that help to improve the quality of this manuscript.

## References

- [1] A. Wahi., S.S., P.P. A Comparative Study for Handwritten Tamil Character Recognition using Wavelet Transform and Zernike Moments. International Journal of Open Information Technologies ISSN: 2307-8162 vol. 2, no. 4, 2014.
- [2] Yanan Lu et al. A Human Action Recognition Method Based on Tchebichef Moment Invariants and Temporal Templates, 4th International Conference on Intelligent Human-Machine Systems and Cybernetics, 2012.
- [3] Omar O., Ramin M., M. Shah. Human Identity Recognition in Aerial Images. IEEE Conference on CVPR, CA, 2010.
- [4] Lowe, D.G.: Distinctive image features from scale-invariant key points. International Journal of Computer Vision 60(2), 2004
- [5] I. Laptev, M. M., C. Schmid, and B. Rozenfeld. Learning Realistic Human Actions from Movies. Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008.
- [6] Willems, G., Tuytelaars, T., Van Gool, L. An efficient dense and scale-invariant spatio-temporal interest point detector. In: Proc. of ECCV. (2008)

- [7] Scovanner, P., Ali, S., Shah, M.: "A 3-Dimensional SIFT Descriptor and its Application to Action Recognition". In: Proc. of ACM Multimedia. (2007)
- [8] A. Klaser, M. Marszałek, and C. Schmid. A spatio-temporal descriptor based on 3D-gradients. In: Proc. of BMVC, 2008.
- [9] Ballan, L., Bertini, M., Del Bimbo, A., Seidenari, L., Serra, G. Recognizing human actions by fusing spatio-temporal appearance and motion descriptors. In: Proc. Of ICIP. (2009)
- [10] Mona M. Moussa, Elsayed Hamayed, Magda B., Heba A. Enhanced method for human action recognition. Elsevier, Journal of Advanced Research, University of Cairo, 2013.
- [11] Khawlah Hussein Ali and T. Wang. Recognition of Human Action and Identification Based on SIFT and Watermark. Springer International Publishing Switzerland 2014.
- [12] Bobick, A.F., Davis, J.W.: The Recognition of Human Movement Using Temporal Templates. J. IEEE Trans. on Pattern Analysis and Machine Intelligence 25, 257–267, 2001.
- [13] H. Ming-Kuei. Visual pattern recognition by moment invariants. IRE Transactions on Information Theory, vol. 8, no. 2, pp. 179–187, 1962.
- [14] Y. D. Khan, Nabeel S., S. F. Adnan A., and M. K. Mahmood. An Efficient Algorithm for Recognition of Human Actions. Hindawi Publishing Corporation, the Scientific World Journal Volume 2014.
- [15] Okay Ank and A.Semih Bingo. Human Action Recognition Using 3D Zernike Moments. 978-1-4799-3866-7/14, IEEE. (2014)
- [16] T. Wang and Simon Liao. Chinese Character Recognition by Zernike Moments. Conference, ICALIP 2014 978-1-4799-3903, IEEE 2014.
- [17] S. A. Dudani, Kenneth J., and Robert B. Mcghee. Aircraft Identification by Moment Invariants. IEEE Transaction on computers, 2009
- [18] Y. Wang<sup>1</sup>, X. Wang, Bin Zhang. A Novel Form of Affine Moment Invariants of Grayscale Images. Elektro Technika, ISSN 1392-1215, VOL. 19, NO. 1, 2013
- [19] Zhuo Zhang, Jia Liu. Recognizing Human Action and Identity Based on Affine-SIFT. IEEE Symposium on Electrical & Electronics Engineering (EEESYM). 2012
- [20] Hai Tao <sup>1</sup>, Li Chongmin,, Jasni Mohamad Zain<sup>1</sup>, Ahmed N. Abdalla, Robust Image Watermarking Theories and Techniques: A Review, Malaysia, Vol. 12, February 2014.
- [21] Nataša Terzija, Markus R. Kerstin Luck, Walter G. Digital Image Watermarking Using Discrete Wavelet Transform Performance Comparison of Error Correction Codes. From Proceeding (364) Visualization, Imaging, and Image Processing, 2002.
- [22] Andrzej Sluzek. Shape Identification Using New Moment-based Descriptors. Conference on computer technology, 1994
- [23] S. S. Reddi, "Radial and Angular Moment Invariants for Image Identification", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. PAMI-3, no. 2, march 1981.
- [24] L. Shao, L. Ji, Y. Liu, and J. Zhang, "Human action segmentation and recognition via motion and shape analysis," *Pattern Recognition Letters*, vol. 33, no. 4, pp. 438–445, 2012.
- [25] Z. Jiang, Z. Lin, and L. Davis, "Recognizing human actions by learning and matching shape-motion prototype trees," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 533–547, 2012.
- [26] J. Flusser, B. Zitova, and T. Suk. Moments and Moment Invariants in Pattern Recognition. JohnWiley & Sons, New York, NY, USA, 2009.
- [27] S. Daniel Madan Raja, A. Shanmugam. Zernike Moments Based War Scene Classification Using ANN and SVM: A Comparative Study. Journal of Information & Computational Science 8: 2 (2011).
- [28] Sun, X., Chen, M., Hauptmann, A. Action recognition via local descriptors and holistic features. In: Proc. of Workshop on CVPR for Human communicative Behavior analysis (CVPR4HB). (2009).
- [29] Boiman, O., Shechtman, E., Irani, M. In defense of nearest-neighbor based image classification. In: Proc of. CVPR. (2008)
- [30] Margarita N. F., Lakhmi C. Jain, Editors. Computer Vision in Control Systems-2: Innovations in Practice. Springer International Publishing Switzerland 2015.
- [31] H. Jos'e Antonio Mart'in, M. Santos, and J. de Lope. Orthogonal variant moments features in image analysis. Information Sciences, vol. 180, no. 6, pp. 846–860, 2010.
- [32] Hejin Yuan and Cuiru Wang. A Human Action Recognition Algorithm Based on Semi-supervised Kmeans Clustering. Transactions on Edutainment VI, LNCS 6758, pp. 227–236, 2011.
- [33] Dollar, P., Rabaud, V., Cottrell, G., B., S. Behavior Recognition via Sparse Spatio-Temporal Features. In: Proc. of VSPETS. (2005).