Appearance and Motion Information based Human Activity recognition

Mahmoud Al-Faris[1]***, John Chiverton***[1]***, Linda Yang***[1]***, David Ndzi***[2]

1. School of Engineering, University of Portsmouth, Portsmouth, UK, PO1 3DJ,
Email: mahmoud.al-faris@port.ac.uk, john.chiverton@port.ac.uk, linda.yang@port.ac.uk
2. School of Engineering, University of the West of Scotland, Scotland, UK
Email: david.ndzi@uws.ac.uk

Keywords: Motion history image, Optical flow

Abstract

Activity recognition is an essential objective of a smart building system which responds to what is happening in a scene. In this paper, a view invariant activity recognition system is proposed to recognise human actions. Selection of applicable features is made and solutions are proposed to deal with probable challenges including differing views on actions and directionality issues. This paper explores a number of features that can be utilised in action recognition systems and chooses suitable features to mitigate the challenges properly. Motion History Image (MHI) based on historical appearance information is used in combination with local motion vectors which are computed through each iteration sequence of the MHI information using an optical flow algorithm. A multi-view dataset (MuHaVi) and a single view dataset (Weizmann) are used to demonstrate and validate the proposed method. Our method, can detect a wide range of actions in multi-view scenarios and shows competitive performance in comparison with state-of-the-art action classification techniques.

1 Introduction

Activity recognition can be considered an essential component of an intelligent system. It will enable the system to respond to what is happening in a scene. Also, it gives the opportunity to build an interactive system which can react to gestural commands, instruct and correct a user learning or interact with people [1]. Moreover, human activity recognition (HAR) is very important in numerous application, for example, video-surveillance, human-PC interaction, video analysis, etc [2], [3].

Different techniques and systems have been used to recognise human activities such as sensors, cameras and mobile phones. Some researchers such as Najafi et al. and Xiong et al. [4], [5], have used multi-modal sensors and RFID devices in sensor-enabled ubiquitous environments detection, recognition and tracking. In some studies, audio is combined with the video to detect the action [6]. A combination of Hidden Markov Models (HMM) with audio was used to determine the actions. The main disadvantage of using audio recordings is the location of the recording devices and the surrounding noise that can affect the results. In some other research, smart cameras and sensors are used for activity monitoring and object recognition [7]. These devices are not cost effective and they are often impractical to be deployed outside laboratories. Therefore, many researchers use cameras as input devices because the same systems can be applied to other domains.

Many researchers have used computer vision based activity recognition using different algorithms and techniques. Motion History Image (MHI) is one of the methods that were used in a HAR system to provide historical information in terms of the appearance of an action combined into one single template. MHI conventionally represents the whole cycle of an action in multiple frame sequences converted into a 2D gray-scaled image [8], [9], [10], [11], [12]. However, each foreground pixel is given by a fixed intensity value with more brightness for recent pixels. In this paper, optical flow is used to present motion characteristics of each pixel based on the historical appearance information. Optical flow can explicitly give horizontal and vertical motion vectors of each individual pixel in the image for two successive frames. But, it can be calculated based on each iteration of MHI sequences to provide motion vectors information for the whole cycle of an action. Furthermore, this paper provides a view independent HAR model. Multi-view (MuHavi) datasets have been used in the training to improve the overall view invariance of the developed system which can help avoid using multiple cameras or other complex processes. The experiments show that inferring these features can provide a rich analysis of human appearance and motion; helping it to be more sensitive in term of action direction. Our method provides competitive results based on MuHaVi and Weizmann datasets.

2 Related Work

Motion history image is a kind of temporal template matching that makes a space-time shape in a video. MHI provides information of motion by weighting sum of past successive frames and the weights descend as time decay. One of the advantages of the MHI representation is that a range of times may be encoded in a single frame, and in this way, it is possible for the MHI to span the time scales of human gestures [1], [10], [12]. Many researchers have used MHI for activity recognition due to its simplicity and low cost of computation.

Researchers Bobick and Davis in [13] proposed a recognition system that decomposed motion-based recognition by describing the location of motion and then describing how the object is moving. Whereas Rosales et al. in [14], used a trajectory guided recognition method which was used to track a person using an extended Kalman filter and then used motion history image for action recognition.

A temporal-template approach was proposed by Bobick and Davis in [15] to construct motion energy images (MEIs) and motion history images (MHIs) for the determination of motion in terms of when and where. Then, a set of moment invariants were computed for recognition. Tian et al. in [16], used motion history image which was proposed for action recognition in combination with a Harris corner detector and a local HOG descriptor which were used to form a representation of an action.

However, MHI represents the whole cycle of an action converted into a 2D gray-scaled image. By using optical flow, motion characteristics of each pixel of an action can be provided. Optical flow has been used by many researchers in human activity recognition such as Ahad et al. in [17], where the authors combined optical flow with human body shape information and represented it by a set of HMMs to estimate human body posture. Also, Chathuramali et al. used a spatio-temporal feature descriptor consisting of optical flow and silhouette information as used in [18] to help resolve problems of human activity recognition. Furthermore, Vrigkas et al. used optical flow features, which were clustered using k-means to build a hierarchical template tree representation for each action in a video sequence as shown in [19].

Our goal is to combine all of these features together and consider using one of them as a foundation stone to obtain historical motion information and apply a feature learning process to learning different view datasets to obtain a single and multi-view action recognition model that could be implemented in a simple way in the training and testing stages to handle missing views of an action. Therefore, there is no need to have all camera views available during the training stage.

3 Methodology

The framework design of the proposed method based actions recognition system is consisting of features extraction, feature vector generation and classification stages as shown in Fig. 1.

3.1 Features Extraction

3.1.1 Motion History Image. Motion history image is a form of temporal template matching that makes a space-time shape in a video. Human actions can take different periods with a cyclic nature. In this paper, we assume that daily actions of a human are cyclic and can take a duration of not more than 50 frames to complete. For the MHI image, the most recent image is brighter than the past ones. This can be written mathematically taking $B(x, y, t)$ as a binary image and if there exists movement at time t and at location (x, y) results in $B(x, y, t) = 1$. Then MHI at time t could be computed as:

$$M_t = \begin{cases} \tau & \text{if } B(x, y, t) = 1 \\ \max(0, M_{t-1}(x, y) - 1) & \text{otherwise,} \end{cases} \quad (1)$$

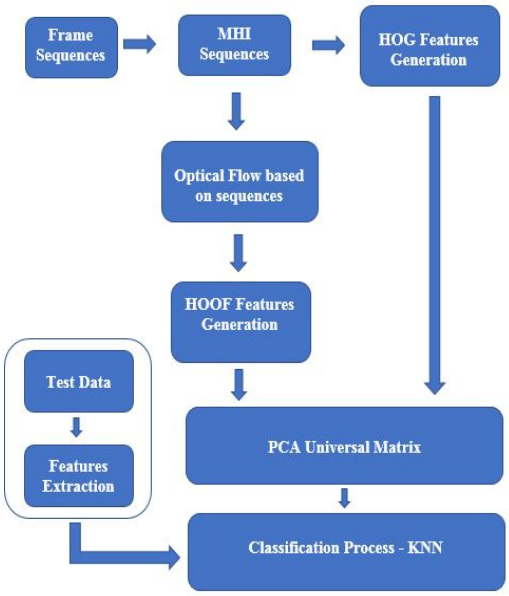


Figure 1: Framework of the proposed method.

where $B(x, y, t)$ is a binary image resulted from frame subtraction and τ is the number of frame sequences used to compute MHI template. In research literature, the value of τ is chosen either to be equal to the total number of frames in video sequences or a fix value as in our case. The resulting MHI can be represented by grayscale image such as the one in Fig. 2 that demonstrates the motion sequence of an individual. The brighter or the higher the value of the pixels indicates the most recent foreground [9], [12], [20].

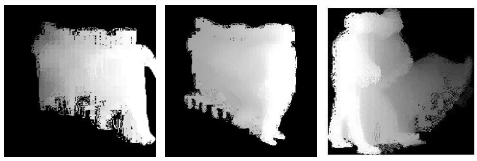


Figure 2: Motion History Image

Furthermore, to provide scale and location invariant representation, MHIs would be centred and scaled to a fixed size. Also, a normalisation stage is added to achieve illumination and contrast invariant representation which is computed as follows:

$$M_n = \frac{M_\tau}{\sum_{x=1}^{R} \sum_{y=1}^{c} M_\tau(x, y)}. \quad (2)$$

where R represents the total number of rows, C represents the total number of columns in M_τ.

3.1.2 Optical flow. Optical flow estimation here, computes motion vectors through the motion history image sequences to form motion vectors sequence too. Using MHI with a duration of τ as a base level for the optical flow estimation model will support the recognition of multiple actions with different view scenarios. A result of this is it will also increase the accuracy rate of the action recognition.

This paper uses layer-wise optical flow estimation model as in

[21] to compute optical flow of an object's motion. All optical flow methods are based on the colour constancy and small motion assumptions, for brightness constancy, it assumes:

$$f(x, y, t) = f(x + dx, y + dy, t + dt). \quad (3)$$

A first order Taylor expansion can be done for small motion assumption:

$$f(x, y, t) = f(x, y, t) + \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial t}dt. \quad (4)$$

Thereafter, the constraint equations on the optical flow are:

$$f_x d_x + f_y d_y + f_t d_t = 0, \quad (5)$$

$$f_x u + f_y v + f_t t = 0. \quad (6)$$

The layer-wise optical flow model can be explained by taking I_1 and I_2 frames with the visible mask V_1 and V_2 of a layer, also let u_1,v_1 be the flow field from I_1 to I_2, and u_2,v_2 the flow field from I_2 to I_1. A series of steps are implemented for estimating layer-wise optical flow. Firstly, match the two frames and the visible layer masks by designing a data term as follows:

$$E^{(1)}_{data} = \int G * V_1(x, y)|I_1(x + u_1, y + v_1) - I_2(x, y)|, \quad (7)$$

where G is a Gaussian filter. Notice that the data term $E^{(2)}_{data}$ for u_2,v_2 can be computed in the same way. Secondly, taking advantage of smoothness as follows:

$$E^{(1)}_{smooth} = \int (|\nabla u_1|^2 + |\nabla v_1|^2)^\rho, \quad (8)$$

where $\rho \in [0.5, 1]$. The last step is the symmetric matching:

$$E^{(1)}_{sym} = \int |u_1(x, y)+u_2(x+u_1, y+v_1)|+|v_1(x, y)+v_2(x+u_1, y+v_1)|. \quad (9)$$

Finally, the objective function of estimation model is the sum of mentioned terms:

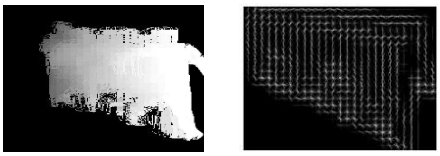
$$E(u_1, v_1, u_2, v_2) = \sum_{i=1}^{2} E^{(i)}_{data} + \alpha E^{(i)}_{smooth} + \beta E^{(i)}_{sym}, \quad (10)$$

where u and v are the flow vectors from frames I_1 to I_2 and I_2 to I_1. α and β are user variables that used to handle different elasticities, a larger α or β results in a smoother flow field.

3.2 Feature vector generation

After the feature extraction processes, histogram of oriented gradients and histogram of oriented optical flow are then used to generate a feature vector.

3.2.1 Histogram of Oriented Gradients. Histogram of oriented gradient descriptor [22] is widely used in computer vision based people detection and human activity recognition. In this paper, historical appearance and motion information is considered as "saliency masks", on which, HOG descriptor is computed as shown in Fig. 3. As a result, a significant description and representation of these information achieves more accurate recognition. HOG computation can be explained with the following steps:



(a) (b)

Figure 3: Histogram of oriented gradients, (a) a template of motion history image (b) oriented bins visualisation.

- Gradient computation: filtering is performed to compute the horizontal and vertical gradients using the following kernel in a vertical and horizontal direction: $K_x = [-1\ 0\ 1]$ and $K_y = [1\ 0\ -1]^{\mathrm{T}}$.
  Later, the magnitude G and orientation θ of the gradient can be calculated using the vertical and horizontal gradients (I_x,I_y) that calculated before: $|G| = \sqrt{(x^2 + y^2)}$ and $\theta = \tan^{-1}\left(\frac{I_x}{I_y}\right)$.
- Orientation binning: the histogram of a cell is computed using 9 orientation bins on the interval of $\theta \in [0°, 180°]$. 8x8 pixels is the size of non-overlapping cells that's computed over the gradient image. The corresponding bin of each pixel is found based on the orientation and related magnitude of the pixel.
- Descriptor blocks: cells are grouped into larger blocks and processing applied to achieve an illumination invariant representation. There are two kinds of block geometry which are: rectangular and circular HOG. In this paper, the rectangular geometry is used with a different block size consisting of 4x4, and 8x8 cells to provide different results.

3.2.2 Histogram of Oriented Optical Flow. We use histogram of oriented optical flow as in [23] to show the distribution of optical flow and to avoid the effects of the background noise and scale changes. Each flow vector is being binned and weighted based on its angle and magnitude. Hence, all optical flow vectors $v = (x\ y)^{\mathrm{T}}$ that resulted from optical flow estimation model with θ direction that ranging between: $-\frac{\pi}{2} + \pi\frac{b-1}{B} \leq \theta < -\frac{\pi}{2} + \pi\frac{b}{2}$. These will contribute by $\sqrt{(x^2 + y^2)}$ to the sum in bin b, where b is ranging between $1 \leq b \leq B$, where B is total number of bins. In this paper, we use different number of bins to show its effect on the accuracy of recognition model. We notice that 100 bins based histogram can achieve significant action recognition results.

3.2.3 Universal components matrix using PCA. Principal Components Analysis (PCA) is a powerful analysis method that is used to identify patterns in data, and the data is projected in such a way that helps to highlight their similarities and differences. Moreover, once newly derived features are computed from the data, the principal components can be used to project the computed feature vector to a basis with a reduced dimensionality or shorter feature space.

In this paper, the PCA technique is used to provide a universal components matrix that is adopted to reduce the feature dimensionality producing a suitable data for the classification process.

3.3 Classification using K-Nearest Neighbour

K-Nearest Neighbour (KNN) classifier is used here as a simple and effective way to perform classification. This approach is considered as one of the efficient supervised learning classifiers that achieves relatively accurate classification rates [24]. This classifier does not require complicated assumptions in terms of the training and similar. When the test data is fed to the classifier, the classifier will look for the similarity in the sections of labelled training data, and later, the label with the highest percent of similarity is assigned to the test data point. The classifier uses Euclidean metric to measure the distance between two points of data as follows:

$$D(\mathbf{p}, \mathbf{q}) = \left(\sum_{i=1}^{n} (p_i - q_i)^2 \right)^{\frac{1}{2}}, \qquad (11)$$

where $\mathbf{p}$ and $\mathbf{q}$ are the description vectors and n is the length of the vectors. A comparison between test and train data is run depending on the minimum distance of them, the label will be assigned to the corresponding test data.

4 Datasets

In this paper two datasets are used for validation. MuHaVi dataset [25] is a multiple camera human activity video data that contains selected action sequences of 14 actions performed by 2 actors using two viewing directions to capture the actions. Moreover, this dataset contains 136 sequences with annotated silhouette sets that is known as MuHaVi-14. In addition, some of these 14 actions can be merged into 8 actions to form MuHaVi-8. In addition, Weizmann dataset is used in this paper as a single view dataset which contains 10 actions performed by 9 people [26].

5 Experiments and Results

In this paper two types of cross-validation scheme are used to validate the proposed method. The first one is leave-one-actor-out (LOAO) and the second is leave-one-sequence-out (LOSO). The same characteristics of the methodology section are used in all experiments. The results of the experiments are discussed below based on the type of validation scheme.

5.1 Leave One Actor Out scheme (LOAO)

In this experiment, the classification process consists of training the classifier on sequences of one actor and then testing on the sequences of the second actor. The average accuracy is calculated by alternatively testing both actors. The result shows that the proposed approach is more robust against the variability of actors' style of doing an action as the training is performed on one actor and testing on another one and vice versa. This experiment achieves accuracy rates of 86.93% and 95.7% for MuHAVi-14 and MuHAVi-8 datasets respectively. The confusion matrix of the recognition model is shown in the tables 1 and 2 in terms of MuHAVi -14 and MuHAVi-8 datasets.

	collapseleft	collapseright	guardtokick	guardtopunch	kickright	punchright	Run-right	Run-left	Standup-left	Standup-right	Turnback-left	Turnback-right	Walk-right	Walk-left
collapseleft	100	0	0	0	0	0	0	0	0	0	0	0	0	0
collapseright	25.0	75.0	0	0	0	0	0	0	0	0	0	0	0	0
guardtokick	6.3	0	25.0	43.8	6.3	18.8	0	0	0	0	0	0	0	0
guardtopunch	6.3	6.3	31.3	56.3	0	0	0	0	0	0	0	0	0	0
kickright	0	0	0	12.5	87.5	0	0	0	0	0	0	0	0	0
punchright	0	0	0	0	0	87.5	0	0	0	0	0	12.5	0	0
Run-right	0	0	0	0	0	0	100	0	0	0	0	0	0	0
Run-left	0	0	0	0	0	0	0	100	0	0	0	0	0	0
Standup-left	0	0	0	0	0	0	0	0	75.0	25.0	0	0	0	0
Standup-right	0	0	0	0	0	0	0	0	0	100	0	0	0	0
Turnback-left	0	0	0	0	0	0	0	0	0	0	100	0	0	0
Turnback-right	0	0	0	0	0	0	0	0	0	0	0	100	0	0
Walk-right	0	0	0	0	0	0	0	0	0	0	0	0	100	0
Walk-left	0	0	0	0	0	0	0	0	0	0	0	0	0	100

Table 1: Confusion matrix of LOAO scheme of Muhavi-14

	Collapse	Guard	Kick	Punch	Run	StandUp	TurnBack	Walk
Collapse	100	0	0	0	0	0	0	0
Guard	12.5	78.1	0	9.4	0	0	0	0
Kick	12.5	0	87.5	0	0	0	0	0
Punch	0	0	0	100	0	0	0	0
Run	0	0	0	0	100	0	0	0
StandUp	0	0	0	0	0	100	0	0
TurnBack	0	0	0	0	0	0	100	0
Walk	0	0	0	0	0	0	0	100

Table 2: Confusion matrix of LOAO scheme of Muhavi-8

Table 1 shows the confusion matrix of MuHaVi-14 dataset classification, it is clear that the majority of misclassifications are between "GaurdToKick" action and "GaurdToPunch" due to the similarity between these two actions. Aside from a little misdirection of "StandUpLeft" and "CollapseRight" actions, it is obvious that this method being sensitive to the direction of an action's motion; therefore, it can easily discriminate between actions of opposite directions e.g. "Walking" and "Turningback". Moreover, table 2 shows the confusion matrix in terms of MuHaVi-8 dataset classification. From this, it can be seen that the accuracy rate of the recognition increases due to the merge of multi-direction actions in to a single action. Our proposed method outperforms similar state-of-the-art approaches as indicated in table 3.

Paper	$MuHaVi-14$	$MuHaVi-8$
Singh et al. [25]	61	76.47
Orrite et al. [27]	75	85.9
Cheema et al. [28]	75.53	83.08
Fiza et al. [9]	81.6	92.3
Ours	**86.93**	**95.7**

Table 3: LOAO scheme: Comparison between our method and state of art approaches

Our proposed method outperforms similar state-of-the-art approaches [9], [25], [27], [28]. However, [9] has used the total number of frames of an action video in MHI computation which provides a suitable platform to classify an action. Whereas our method proposed 50 frames per action to be flexible with any sudden change of an action, also, this will give an opportunity to learn an action if less information might be available.

5.2 Leave One Sequence Out scheme (LOSO)

In this kind of validation, the classifier deals with sequences rather than actors. The classifier is being trained on whole sequences except one which is left for testing, in our case, 135 sequences for training and 1 for testing respectively. The experiments show the effectiveness of the proposed method, achieving promising accuracy rates of 95.1% and 97.7% for Muhavi-14 and Muhavi-8 respectively, as shown in tables 4 and 5.

	collapseleft	collapseright	guardtokick	guardtopunch	kickright	punchright	Run-right	Run-left	Standup-left	Standup-right	Turnback-left	Turnback-right	Walk-right	Walk-left
collapseleft	100	0	0	0	0	0	0	0	0	0	0	0	0	0
collapseright	25.0	75.0	0	0	0	0	0	0	0	0	0	0	0	0
guardtokick	0	0	87.5	0	12.5	0	0	0	0	0	0	0	0	0
guardtopunch	0	0	37.5	62.5	0	0	0	0	0	0	0	0	0	0
kickright	0	0	0	0	100	0	0	0	0	0	0	0	0	0
punchright	0	0	0	0	0	100	0	0	0	0	0	0	0	0
Run-right	0	0	0	0	0	0	100	0	0	0	0	0	0	0
Run-left	0	0	0	0	0	0	0	100	0	0	0	0	0	0
Standup-left	0	0	0	0	0	0	0	0	100	0	0	0	0	0
Standup-right	0	0	0	0	0	0	0	0	0	100	0	0	0	0
Turnback-left	0	0	0	0	0	0	0	0	0	0	100	0	0	0
Turnback-right	0	0	0	0	0	0	0	0	0	0	0	100	0	0
Walk-right	0	0	0	0	0	0	0	0	0	0	0	0	100	0
Walk-left	0	0	0	0	0	0	0	0	0	0	0	0	0	100

Table 4: Confusion matrix of LOSO scheme of Muhavi-14

Insignificant misclassification can be noticed in both groups, for instance, 'Turnback' action is being misclassified with 'Walk' action in Muhavi-8. Whereas 'Guardtokick' action is being misclassified with 'Guardtopunch' in MuHavi-14 due to the high similarity between the two actions. It is worth to notice that the proposed method is more robust and effective in the second group (Muhavi-8) due to the multiple primitive actions that the first group consists. Our proposed approach outperforms similar state-of-the-art approaches as shown in table 6.

Moreover, in this paper, a comparison of results as shown in table 7, is provided showing the effects of different parameter values such as the number of bins that are used in HOOF descriptor, the block size of the HOG descriptor and the number of nearest neighbours that are used in the KNN classifier.

A KNN classifier is used with different nearest neighbours values ranging from 1 to 10. This shows different accuracy

	Collapse	Guard	Kick	Punch	Run	StandUp	TurnBack	Walk
Collapse	100	0	0	0	0	0	0	0
Guard	0	96.9	3.1	0	0	0	0	0
Kick	0	0	100	0	0	0	0	0
Punch	0	0	0	100	0	0	0	0
Run	0	0	0	0	100	0	0	0
StandUp	0	0	0	0	0	100	0	0
TurnBack	0	0	0	0	0	0	83.3	16.7
Walk	0	0	0	0	0	0	0	100

Table 5: Confusion matrix of LOSO scheme of Muhavi-8

Paper	$MuHaVi-14$	$MuHaVi-8$
Cheema et al. [28]	86.03	95.58
Singh et al. [25]	82.35	97.8
Fiza et al. [9]	92.6	98.3
Ours	**95.1**	**97.7**

Table 6: LOSO scheme: our method and state-of-the-art comparisons

Bins of HOOF	100		150	
HOG Blocks	4x4	8x8	4x4	8x8
KNN (1)	82.3%	83.5%	81%	80.1%
KNN (4)	79.9%	86.93%	77.6%	81.8%

Table 7: Different feature selection parameter characteristics demonstrating different results.

rates of recognition. The most accurate results are achieved with (1) and (4) nearest neighbours, whereas other values provide lower recognition rates. In addition, 4x4 and 8x8 cell sizes are used in terms of the HOG descriptor and (100,150) bins in terms of the HOOF descriptor.

In addition, we did different experiments with numerous classifiers as shown in table 8 helping to demonstrate the effectiveness and robustness of different classifiers in terms of application to our problem area.

Classifier	$MuHaVi-14$	$MuHaVi-8$
Naive Bayes	69.91	82.42
SVM	75.70	86.80
Feed-forward NN	82.60	91.71
KNN	86.93	95.70

Table 8: Results comparison of different classifiers in terms of Muhavi-14 and Muhavi-8.

Furthermore, this method has been evaluated with a single view dataset (Weizmann), and it gives a robust result (97%) with the whole 10 actions of the dataset. It misclassified the 'Skip' action with the 'Run' action due to the huge similarity of the two actions. This method can achieve 100% accuracy with the Weizmann dataset when 'Skip' action is neglected or added to the 'Run' action as shown in table 9.

Paper	*No.of Action*	*Accuracy*
Venugopal T. et al. [29]	9	91
C. Li et al. [8]	9	97.53
Ahsan et al. [30]	10 (9)	94.26 (97.5)
Niebles et al. [31]	9	72.8
Ours	**10 (9)**	**97 (100)**

Table 9: Performance comparison between our method and others in terms of the Weizmann dataset.

6 Conclusion

A view-invariant human activity recognition system based on historical appearance and motion information in the spatial domain and the transform domain is presented. The proposed method achieved a high accuracy rate in action recognition in addition to minimum storage requirements in comparison with the most recent reported methods. Different algorithms have been used in this system to provide a suitable, robust and rich feature extraction process in addition to a feature vector generation stage, in which, HOG and HOOF descriptors are used.

Experiments are performed on MuHavi dataset with different scenarios such as MuHaVi-8 and MuHAVi-14 datasets. In addition, Weizmann dataset is also used in the experiments for validation purposes. The results confirmed that our approach provides significant results compared to similar state-of-the-art approaches by providing a view-independent recognition system with competitive accuracy rates.

References

[1] H. Meng, N. Pears, and C. Bailey, "Motion information combination for fast human action recognition." in *VISAPP (2)*, 2007, pp. 21–28.

[2] X. Zhang, Z. Miao, and L. Wan, "Human action categories using motion descriptors," in *Image Processing (ICIP), 2012 19th IEEE Int. Conf.* IEEE, 2012, pp. 1381–1384.

[3] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, "A fast statistical approach for human activity recognition," 2012.

[4] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C. J. Bula, and P. Robert, "Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 711–723, 2003.

[5] J. Xiong, B.-C. Seet, and J. Symonds, "Human activity inference for ubiquitous rfid-based applications," in *Ubiquitous, Autonomic and Trusted Computing, UIC-ATC'09*. IEEE, 2009, pp. 304–309.

[6] B. U. Töreyin, Y. Dedeoğlu, and A. E. Çetin, "HMM based falling person detection using both audio and video," in *Int. Workshop Human-Comp. Interaction*. Springer, 2005, pp. 211–220.

[7] A. Williams, D. Xie, S. Ou, R. Grupen, A. Hanson, and E. Riseman, "Distributed smart cameras for aging in place," DTIC Document, Tech. Rep., 2006.

[8] C. Li, Y. Liu, J. Wang, and H. Wang, "Combining localized oriented rectangles and motion history image for human action recognition," in *Computational Intelligence and Design (ISCID), 7th Int. Symposium*, vol. 2. IEEE, 2014, pp. 53–56.

[9] F. Murtaza, M. H. Yousaf, and S. A. Velastin, "Multi-view human action recognition using histograms of oriented gradients (hog) description of motion history images (mhis)," in *Frontiers of Information Technology (FIT), 2015 13th Int. Conf.* IEEE, 2015, pp. 297–302.

[10] M. Ahmad and M. Z. Hossain, "Sei and shi representations for human movement recognition," in *Comp. and Information Technology, 2008. ICCIT 2008. 11th Int. Conf.* IEEE, 2008, pp. 521–526.

[11] M. Hassan, T. Ahmad, N. Liaqat, A. Farooq, S. A. Ali *et al.*, "A review on human actions recognition using vision based techniques," *Journal of Image and Graphics*, vol. 2, no. 1, pp. 28–32, 2014.

[12] M. A. R. Ahad, J. K. Tan, H. Kim, and S. Ishikawa, "Motion history image: its variants and applications," *Machine Vision and Applications*, vol. 23, no. 2, pp. 255–281, 2012.

[13] A. Bobick and J. Davis, "An appearance-based representation of action," in *Pattern Recognition, 1996., Proc. of the 13th Int. Conf.*, vol. 1. IEEE, 1996, pp. 307–312.

[14] R. Rosales and S. Sclaroff, "3d trajectory recovery for tracking multiple objects and trajectory guided recognition of actions," in *Comp. Vis. Patt. Recog., 1999. IEEE Comp. Society Conf.*, vol. 2. IEEE, 1999, pp. 117–123.

[15] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. Patt. anal. mach. intell.*, vol. 23, no. 3, pp. 257–267, 2001.

[16] Y. Tian, L. Cao, Z. Liu, and Z. Zhang, "Hierarchical filtered motion for action recognition in crowded videos," *IEEE Trans. Systems, Man, and Cyb., Part C (Apps. and Reviews)*, vol. 42, no. 3, pp. 313–323, 2012.

[17] M. A. R. Ahad, J. Tan, H. Kim, and S. Ishikawa, "Human activity recognition: Various paradigms," in *Control, Automation and Systems, 2008. ICCAS 2008. Int. Conf.* IEEE, 2008, pp. 1896–1901.

[18] K. M. Chathuramali and R. Rodrigo, "Faster human activity recognition with svm," in *Advances in ICT for Emerging Regions (ICTer), 2012 Int. Conf.* IEEE, 2012, pp. 197–203.

[19] M. Vrigkas, C. Nikou, and I. A. Kakadiaris, "A review of human activity recognition methods," *Frontiers in Robotics and AI*, vol. 2, p. 28, 2015.

[20] Z. Z. Htike, S. Egerton, and K. Y. Chow, "Real-time human activity recognition using external and internal spatial features," in *Intelligent Environments (IE), 2010 Sixth Int. Conf.* IEEE, 2010, pp. 52–57.

[21] C. Liu, W. T. Freeman, E. H. Adelson, and Y. Weiss, "Human-assisted motion annotation," in *Comp. Vis. Patt. Recog., 2008. CVPR 2008. IEEE Conf.* IEEE, 2008, pp. 1–8.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Comp. Vis. Patt. Recog., 2005. CVPR 2005. IEEE Comp. Society Conf.*, vol. 1. IEEE, 2005, pp. 886–893.

[23] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *Comp. Vis. Patt. Recog., 2009. CVPR 2009. IEEE Conf.* IEEE, 2009, pp. 1932–1939.

[24] N. Bhatia *et al.*, "Survey of nearest neighbor techniques," *arXiv preprint arXiv:1007.0085*, 2010.

[25] S. Singh, S. A. Velastin, and H. Ragheb, "Muhavi: A multi-camera human action video dataset for the evaluation of action recognition methods," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE Int. Conf.* IEEE, 2010, pp. 48–55.

[26] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in *10th IEEE Int. Conf. Comp. Vision (ICCV'05)*, 2005, pp. 1395–1402.

[27] C. Orrite, M. Rodriguez, E. Herrero, G. Rogez, and S. A. Velastin, "Automatic segmentation and recognition of human actions in monocular sequences," in *Pattern Recognition (ICPR), 2014 22nd Int. Conf.* IEEE, 2014, pp. 4218–4223.

[28] S. Cheema, A. Eweiwi, C. Thurau, and C. Bauckhage, "Action recognition by learning discriminative key poses," in *Comp. Vision Workshops (ICCV Workshops), 2011 IEEE Int. Conf.* IEEE, 2011, pp. 1302–1309.

[29] T. Thanikachalam and K. Thyagharajan, "Human action recognition using motion history image and correlation filter," *Int J Appl Eng Res*, vol. 10, pp. 361–363, 2015.

[30] S. M. M. Ahsan, J. K. Tan, H. Kim, and S. Ishikawa, "Histogram of spatio temporal local binary patterns for human action recognition," in *SCIS, 2014 Joint 7th ISIS, 15th Int. Symp.* IEEE, 2014, pp. 1007–1011.

[31] J. C. Niebles and L. Fei-Fei, "A hierarchical model of shape and appearance for human action classification," in *Comp. Vis. Patt. Recog., 2007. CVPR'07.* IEEE, 2007, pp. 1–8.