Space-time Templates based Features for Patient Activity Recognition

Muhammad Adeel Abbas[1], Fiza Murtaza[2,3], Muhammad Obaid Ullah[1], and Muhammad Haroon Yousaf[2]
[1]University of Engineering and Technology, Department of Electrical Engineering, Pakistan
[2]University of Engineering and Technology, Department of Computer Engineering, Pakistan
[3]Sino-Pak Center for Artificial Intelligence (SPCAI), PAF-IAST, Pakistan

Abstract: *Human activity recognition has been the popular area of research among the computer vision researchers. The proposed work is focused on patient activity recognition in hospital room environment. We have investigated the optimum supportive features for the patient activity recognition problem. Exploiting the strength of space-time template approaches for activity analysis, Motion-Density Image (MDI) is proposed for patient's activities when used supportively with Motion-History Image (MHI). The final feature vector is created by combining the description of MHI and MDI using Motion Orientation Histograms (MOH) and then applying Linear Discriminant Analysis (LDA) for dimensionality reduction. The LDA technique not only reduced the complexity cost required for classification but also played vital role to get best recognition results by increasing between-class separation and decreasing the with-in class variance. To validate the proposed approach, we recorded a video dataset containing 8 activities of patients performed in hospital room environment under indoor conditions. We have successfully validated the results of the proposed approach on our dataset by training the SVM classifier and achieved 97.9% average recognition accuracy.*

Keywords: *Human activity recognition, motion templates, patient monitoring, LDA.*

Received September 3, 2019; accepted July 14, 2020
https://doi.org/10.34028/18/4/2

1. Introduction

The smart city projects are creating more secure and safe environments, to provide their residents with a better lifestyle, by installing a massive amount of surveillance and monitoring cameras at almost every type of public place. Health care systems are under keen observation by developed countries, as the population aging of the large baby boomer cohorts is foreseen to be exacerbating health care demand [19]. The upcoming era demands improvements in existing health care systems and needs new methods to incorporate based on cutting edge technology. Hospital staff has to navigate patient's room occasionally to gather information to care them properly, which is a time-consuming task. This issue becomes more challenging for the elderly patients admitted to the hospitals than the patients of other ages. Elderly population above 65 years of age is growing, and their ratio to the remaining population can reach 35% in 2030 [2]. Researchers are working for the development of such systems as incorporated in [16] for smart hospital applications. Among such systems, an important aspect is to monitor and understand the patient's activities in hospital room environment. The successful recognition of the patient's activity will play an essential part in disease recovery and patient safety. One such system, proposed Elmezain and Al-Hamadi [9], was to prevent pneumonia by detecting out of bed activities of patients.

The proposed work is mainly demonstrating a computer vision-based approach for Patients' Activity Recognition (PAR) in hospital room environment. PAR can be useful in assistive living, smart city, smart hospitals, and surveillance applications. The main objective of this work is the recognition of patient's daily life activities in hospital room environment. The survey in [6] covers existing video databases for human action/activity recognition, but no standard dataset is available for validation of patients' activities. Therefore, in this work we recorded a video dataset containing eight activities of patients performed in hospital room environment under indoor conditions. Activities were selected from the patient's routine activities keeping in view their health care problems. Eight activities, that can lead the patient towards dangerous condition, were captured in hospital room environment. Eight volunteers of different ages were involved in the preparation of the dataset. This research adds a contribution to the field of healthcare in hospital environments. More details about the PAR dataset can be seen in section 4.1.

The motivation behind the proposed work is to develop a PAR system to achieve considerable recognition accuracy but with less processing cost and complication. Secondly, our main purpose is to investigate the simple method which can give us the significant recognition results on PAR dataset. In this work, we used temporal templates for the representation of action videos. The reason to choose

the temporal templates technique is their ability to reduce complexity in processing the videos, i.e., they represent a whole video sequence with a single image containing dominant posture and motion information of full activity. We proposed a view based temporal template named as Motion Density Image (MDI). Results showed that supportively using the features from the Motion-History Image (MHI) and MDI leads to high recognition accuracy.

In MHI the motion is overwritten because of self-occlusion by the body parts [1]. If an activity contains its necessary atomic actions in opposite directions (e.g., Backward_Fall, Faint, etc.), then motion history information of first atomic action is overwritten by the latter one. Thus, self-occlusion, during the activity of the moving body, overwrites the prior information. To solve this problem, we proposed MDI which, when used supportively with MHI, gives the improved recognition results.

The rest of the paper is organized as follows: section 2 describes the related work. Section 3 represents our proposed method in detail. The experimental results and the dataset details are discussed in section 4. Finally, the conclusions are drawn in section 5.

2. Related Work

For monitoring daily activities in a healthcare facility, researchers have recently introduced lots of wearable non-wearable devices. Wearable devices are serving adequately but with some constraints, e.g., wearing them all the time (for activity monitoring devices) and to put them at specific body location (for vital measurement devices). They also have some drawbacks: like a battery problem, the user forgot to wear it or lose its consciousness [13]. Non-Wearable devices, e.g., camera-based consumer electronics device installed at the room can overcome these drawbacks and remove the constraints required for wearable devices. This work proposes a novel approach for PAR which can be useful in the development of such non-wearable consumer device.

Human fall detection and activity recognition systems have been widely studied by the researchers [4, 9, 10, 17, 20]. Their main focus is to detect the falls and non-fall activities, but they are not further categorizing them into more specific activities. There is a need to classify the emergency activities in the healthcare application more specifically. PAR will assist the nursing staff in detecting any prompt action requirement [3].

Appearance-based representation of human action in videos was pioneered in [4]. In this work, the authors proposed Motion Energy Images (MEI) as a method of capturing the spatial information of motion in an action video. MEI was a binary-image localizing the spatial region where motion has occurred. Another view-based template MHI is proposed in [5]. MHI is the scalar-valued version of MEI, keeping track of the motion history (how motion has occurred) in an action video. MEI and MHI are used together for representation of human actions in [5, 7]. A new method for recognizing movements in real time scenarios is proposed in [8], in which motion was represented by MHI. This technique acquires multiple overlapping Motion Orientation Histograms (MOH) from MHI template using a hierarchical overlapping partitioning window. A detailed analysis of variants and applications of MHI approaches are reviewed in [1].

3. Proposed Methodology

The general framework for our proposed PAR system is illustrated in Figure 1. The videos for all activities are pre-processed to prepare them for segmentation to get the binary silhouettes. After silhouette extraction, a visual representation of the input activity is acquired in the form of scalar-valued images, i.e., MDI and MHI. MOH is then calculated from MHI and MDI locally (under the hierarchical window partitions). The MOH from both MHI and MDI is combined to form the Supportive Feature Vector (SFV). Linear Discriminant Analysis (LDA) is applied on SFV to increase the interclass separation which results in improved recognition rate. The dimensionality reduction ability of LDA also decreases the processing complexity in recognition step. The resultant feature vectors, from training data, are then used to train multiclass SVM classifier for recognition. For validation of the results, query activity sequence is processed with the same steps, and supportive feature vector representation of MOH is formed which is then predicted through the trained classifier models. The description of steps included in PAR framework is explained in the following subsections.

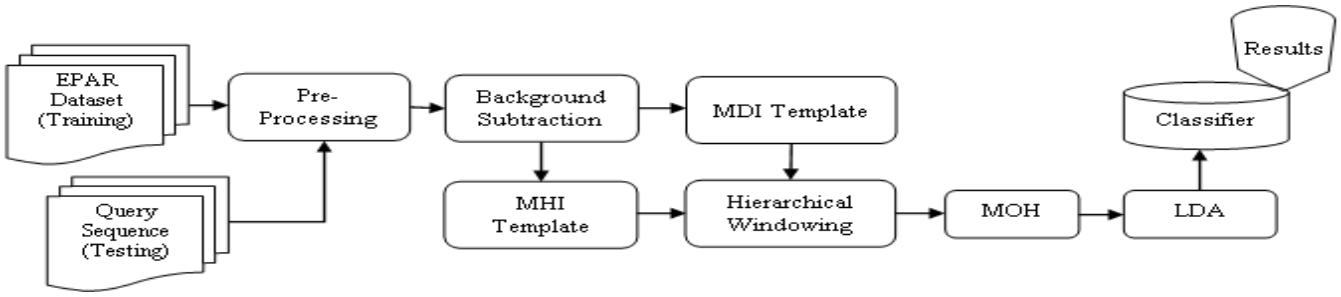


Figure 1. General framework for patient activity recognition system.

3.1. Video Pre-Processing

Raw videos in the dataset were contaminated with background noise and dynamic illumination changes. Gaussian kernel averaging is used to remove the background noise and inter-frame averaging is performed to get rid of dynamic illumination problem. Whole videos are filtered frame by frame with isotropic Gaussian smoothing kernels (i.e., kernel masks with same standard deviation along both axes). The coefficients of this isotropic Gaussian kernel are the results of 2D Gaussian function given as:

$$k(x,y) = \exp\left[\frac{-(x^2+y^2)}{2\sigma^2}\right] \quad (1)$$

Where x and y are the values of the coordinates and σ is the standard deviation. We have used the kernel size of 7×7 and σ=3. The reference background frame Rb is acquired by applying Gaussian smoothing followed by the inter-frame averaging of all background frames V as:

$$Rb(x,y) = \frac{\sum_{t=1}^{n} V(x,y,t)}{n} \quad (2)$$

Where t is the frame number and n is the total number of frames in the background video sequence.

Each video frame is subtracted from reference background frame, *RB*, to suppress the unwanted background information. The binary representation of the foreground object (human), i.e., *B*, is then acquired by applying Otsu's threshold on the background subtracted scalar image.

$$B(x,y,t) = \begin{cases} 1 & if\ |Rb(x,y) - f(x,y,t)| > T_{(otsu)} \\ 0 & otherwise \end{cases} \quad (3)$$

3.2. Space-Time Motion Templates

Once the silhouettes $B(x,y,t)$ are extracted, next step is to represent them in discriminant representation. We use MHI and MDI to find the motion templates.

3.2.1. Motion-History Image

MHI is the spatiotemporal representation of the video frames, and it converts the activity video into a single scalar valued image. It represents *how* and *where* motion a pixel undergoes during activity. It paints the history of motion to signify the sequence of poses during activity visually. Given the foreground segmented binary silhouette sequence $B(x,y,t)$, the 2D MHI template having dimensions equal to the input frame, is given by:

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

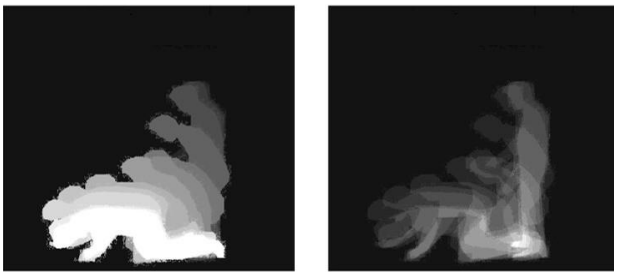
Where the condition *if* $B(x,y,t)$=1 checks whether there exists the motion in the new coming frame t, if yes then it assigns τ to the moving pixels of the current time-stamp (frame number), otherwise retain the previously generated MHI.

3.2.2. Motion-Density Image

To represent how much (rather than how and where) motion a pixel undergoes during activity, we form a MDI, which cope with the problem of self-occlusion. MDI so represents all frame of an action sequence using a single scalar image. It paints the amount of motion to give an in-depth insight into the sequence of poses by informing about self-occluded poses during activity. The value of the pixel depicts its exposure to the motion. Given the foreground segmented binary silhouette sequence $B(x,y,t)$ the 2D MDI template, having dimensions equal to the input frame, is given by:

$$MDI(x,y) = \sum_{t=1}^{n} f(t).B(x,y,t) \quad (5)$$

Where $f(t)$ is the weight function that can be used to give the higher weights to the most recent frames e.g., $f(t)$=t or $f(t)$ =t^2. We have chosen the constant weight function $f(t) = 1\ ,\ \forall\ t$. MHI template for Forward_Fall activity is shown in Figure 2-a) whereas the MDI template is shown in Figure 2-b).



a) MHI template for Forward_Fall Activity.

b) MDI template for Forward_Fall Activity.

Figure 2. Temporal templates.

3.3. Feature Extraction

Feature extraction produces a more abstract output from motion templates that will assist in developing the machine learning algorithms. In feature extraction step, specific attributes of the MHI and MDI are detected for further processing that can lead to a discriminant representation of activities. The MHI and MDI present motion information in such a way that direction of motion is being perceived. From Figure 2 it can be observed that MHI and MDI templates are visually encoding the human movements. This directional movement information is the distinctive characteristic of MHI and MDI and serves to distinguish activities of different activity classes [11]. Gradients were computed to grab this useful information as explained below.

3.3.1. Gradient Computation

We find the bounding boxes containing motion information for each of MHI and MDI templates. Then we crop the MHI and MDI templates to fix size and calculate the gradients on the cropped templates. We cropped the images because non-moving (zero valued) pixels will corrupt the motion information and thus influence the recognition results. Only interior pixels

having motion information are involved in the computation of gradients. This process makes this approach computationally more efficient and hence more attractive for real-time applications.

To capture the direction of motion, gradients were computed in both *x* and *y* directions using the following Sobel gradient kernels.

$$F_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \; F_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{6}$$

Convolving the obtained motion templates with the above kernels gives gradient values for each pixel in both *x* and *y*-direction. The convolution of MDI and MHI template with Sobel gradient kernels is given below:

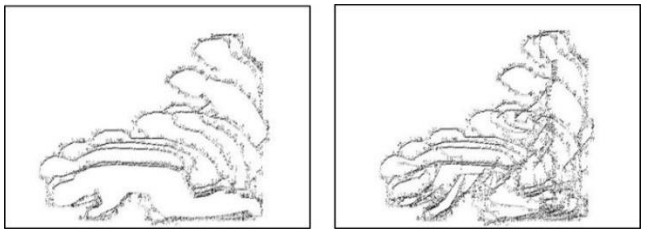
$$G_x(x,y) = M(x,y) * F_x \tag{7}$$

$$G_y(x,y) = M(x,y) * F_y \tag{8}$$

Where $M(x,y)=MHI(x,y)$ or $M(x,y)=MDI(x,y)$. After the calculations of gradients, G_x and G_y, the orientation at each pixel of the template can be effectively computed as:

$$\theta(x,y) = \tan^{-1}\frac{G_y(x,y)}{G_x(x,y)} \tag{9}$$

The Figures 3-a) and Figure 3-b) show calculated orientations for MHI and MDI templates of Figure 2-a) and Figure 2-b) respectively. Figure 3 depicts that the orientation of MDI is encoding the self-occluded motion information in a better way than the orientation of MHI. These orientations provide the directional information of each patient's activity which acts as a strong feature for classification. The range of orientation values $\theta(x,y)$ at each pixel was [-180°, 180°]. For better performance, we find the unsigned orientations as in [16, 17]. Therefore, the angles less than 0° are summed up with 360° to get their unsigned equivalent.



a) MHI. b) MDI.

Figure 3. Motion orientations calculated.

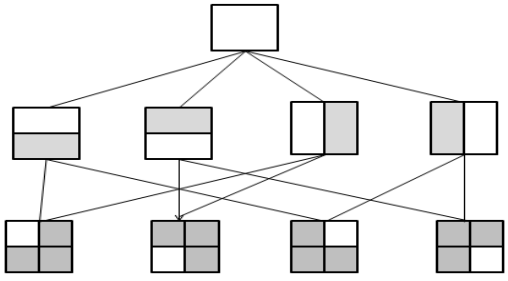


Figure 4. Hierarchical overlapping windows for extracting local features. White mask represents the included region.

3.3.2. Hierarchical Window Partitioning

Aforementioned discussion clarifies the feature extraction stage of our proposed approach by finding the gradient orientation globally from MHI and MDI templates. Though the gradient orientations computed globally are admirably describing the motion information but still cannot characterize the different region around the body. For example, the movement of the head is much more than the other body parts. In this work, we have chosen hierarchical window partitioning [13] to pay attention to the different region of the template separately. It uses a set of eight overlapping hierarchical windows, as shown in Figure 4, to characterize the human movements locally. These windows have overlapping regions and cover the whole template; and features are extracted from each window separately to get a concise representation of the activity [14].

To apply hierarchical window partitioning on templates, the centroid of the template along its bounding box is calculated thus localizing the person and the size of activity region. Defining the centre point of the window yields translation invariance in X-Y axis and the bounding box helps to achieve scale invariance during recognition [7].

3.4. Motion Orientation Histograms

A compact distinguishable form of extracted features is the requirement of any recognition system which is the result of feature description step. Feature description is used to describe the extracted features in a more efficient way which is used for the final classification step. For each window, the gradient orientations are extracted from MHI and MDI (as described in section 3.3) and then described using MOH. The range of resultant orientation values is [0°-360°]. Histograms with 12 bins (each of size 30°) are collected under each partitioned window. The quantization of orientation values to 30° bin size was mainly required to speed up the recognition step.

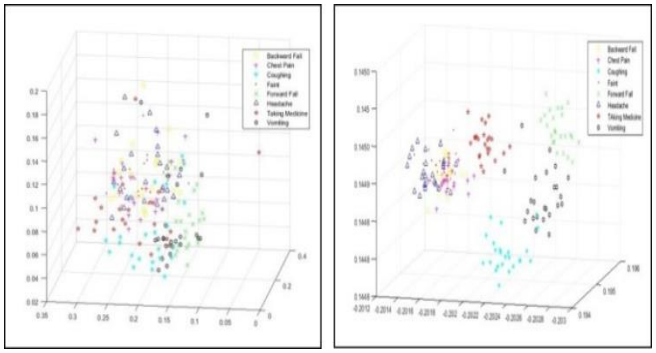
To handle the scenario of different sized people or changes in location of depth, normalization of the histograms is needed with respect to some measure of motion or size of the person. A better method, as given by [8], is adapted to normalize each window, i.e., by dividing the MOH with the sum of the total motion orientation pixels in overall template's histogram.

$$\mathrm{MOH}_n = \frac{\mathrm{MOH}}{\sum_{i=1}^{12}\mathrm{MOH}} \tag{10}$$

The feature vector is then formed by concatenating the resultant normalized motion orientation histograms, MOH_n, of each partitioned window. To improve the accuracy and to reduce the complexity of the recognition step, dimensionality reduction using LDA is performed to get the final feature vector.

3.5. Linear Discriminant Analysis (LDA)

To increase the inter-class separation and minimize the intra-class separation [18], LDA is applied on feature vectors (Figure 5). LDA is also capable of reducing the dimensionality of the feature space that allows us to decrease processing complexity in recognition step.



a) Before LDA Implementation. b) After LDA Implementation (a).

Figure 5. Feature space representation using 3D scatter plots.

The inter-class and intra-class scatter matrices are computed as:

$$S_B = \frac{1}{N}\sum_{i=0}^{C} n_i(\mu_i - \mu)(\mu_i - \mu)^T, \quad (11)$$

$$S_W = \frac{1}{N}\sum_{i=1}^{C}\sum_{x\in X_i}(x - \mu_i)(x - \mu_i)^T \quad (12)$$

Where S_B represents inter-class scatter matrix and S_W is the intra-class scatter matrix. Moreover, N is the total number of data samples, n_i is the number of data samples from ith class, C is the total number of classes; x is a vector from specified class, and X_i is the set of data samples of ith class. The global centroid for feature space is μ, whereas μ_i is the centroid of ith class. The degree of scattering for inter-classes is represented by S_B, the covariance matrix of means. The degree of scattering for intra-classes is represented by S_W, the summation of covariance matrices of each class. Solving the optimization criteria yields optimal discrimination projection matrix O_{opt} in projection space as in [12, 18].

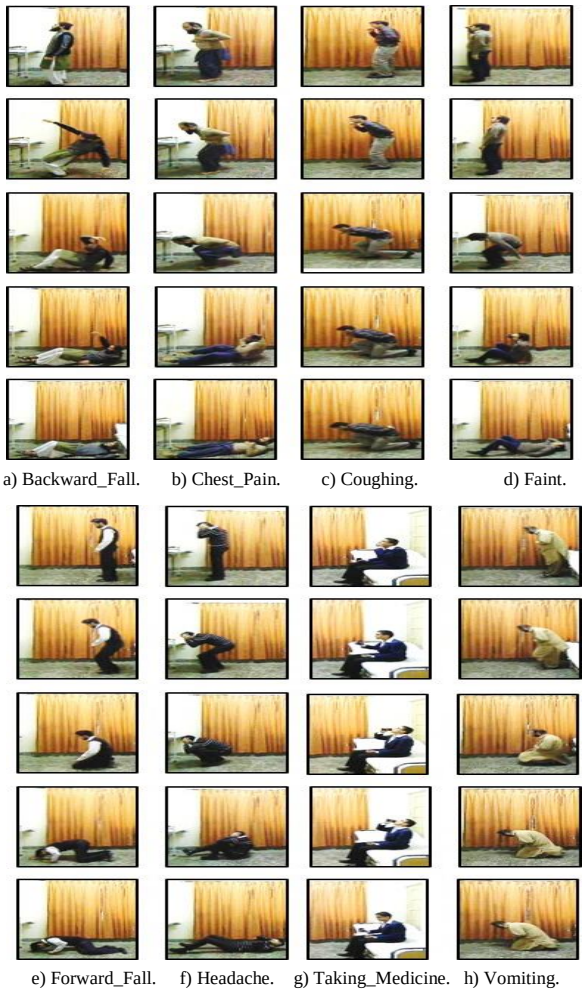
$$O_{opt} = \arg\max_{o} \frac{O^T S_B O}{O^T S_W O} \quad (13)$$

The above optimization problem O_{opt} can be solved as generalized Eigen value problem that results in *C-1* largest Eigen values as in [19]. The goal is to seek a set of feature vectors O_i that maximizes the expression in Equation (13). This leads to the symmetric eigenvector equation [17]:

$$S_B O_i = \lambda_i S_W O_i \quad (14)$$

Where S_w is a non-singular matrix, i.e., its inverse exists. So, the above equation can be written as:

$$S_W^{-1} S_B O_i = \lambda_i O_i \quad (15)$$



a) Backward_Fall. b) Chest_Pain. c) Coughing. d) Faint.

e) Forward_Fall. f) Headache. g) Taking_Medicine. h) Vomiting.

Figure 6. Selective key frames from eight activities.

In this work, the LDA implementation on both MHI and MDI features reduces the feature spaces to (1x7), resulting in the supportive feature vector with dimensions (1x14).

4. Experiments and Results

4.1. Dataset

PAR dataset is used to evaluate our proposed algorithm. We recorded PAR dataset in a hospital room. The living area of this place is covered by the view of a single camera installed. The dataset includes eight activities performed by eight volunteers, of different age groups, with three repetitions per activity. Eight activities for patients were selected to capture the dataset. Selected activities are those which need emergency medical help and can lead the patient to a dangerous condition. These activities are: Backward_Fall, Chest_Pain, Coughing, Faint, Forward_Fall, Headache, Taking_Medicine, and Vomiting. In Figure 6-a) to Figure 6-h), selective keyframes from raw activity videos are given in eight columns. These keyframes demonstrate a visual idea about PAR dataset. The dataset contains total 192 (8×8×8) videos with an average length of 3 to 4 seconds and total 25248 of frames (24 video sequences per activity). The dataset constraints are: the patient is

admitted in the hospital room; dataset includes indoor environment activities only under the artificial light condition. The main problem faced while working with this dataset was the changing illumination in the background which was fixed using the inter-frame averaging.

4.2. Experimental Settings

In section 3, we described the overall approach for silhouette extraction and feature description. All activity sequences were first pre-processed to get the silhouette sequences. From this silhouette information, each activity is represented by MHI and MDI templates of resolution 320×240. For eight overlapping windows, 12 bin MOH is computed which resulted in a 1×96 dimensional feature vector for each MHI and MDI, subsequently creating a Supportive Feature Vector (SFV) with a dimension of 1×192.

In classification step, we analyzed the accuracy using SVM [20]. The SVM is trained using multiclass ECOC (error-correcting output codes model) with one-vs-one strategy and polynomial as the kernel function for classification. Different types of cross-validation schemes are used for performance evaluation including: Leave-One-Out (LOO) and k-fold cross-validation scheme with *k*=3, 4 and 5. All the experiments are performed using MATLAB R2015a with Intel Core i7 3.60 GHz, 8GB RAM, 64-bit operating system.

4.3. Results and Discussion

In this section, we present the results on the PAR dataset and discuss which feature combination result in better recognition accuracy using SVM classifier. We also illustrate the significance of the LDA by applying the classifier before and after LDA on all feature combinations with different validation schemes.

4.3.1. Action Recognition without LDA

In this case, we find results of the proposed method without applying LDA on following feature vectors:

fv1 = MOH features from MHI
fv2 = MOH features from MDI
fv3= SFV, i.e., the concatenation of fv1, fv2

First, we present the results on fv1, fv2, and fv3 before applying LDA. Afterwards, we will show the effect of applying LDA on these feature descriptors. In Table 1 we presented the results of the fv1, fv2 and fv3 using different cross-validation schemes. The results in the Table 1 show that fv3 i.e., SFV gives the best results for all validations schemes. Accuracy rate for MOH description of MDI (fv2) is higher than MHI feature descriptor (fv1) for all cross-validation scheme. Similarly, fv3 got highest accuracy rate as compared to fv1 and fv2 for all cross-validation schemes which shows that the combination of both MHI and MDI lead to better results. The proposed method got highest accuracy rates of 76.9 % for fv3 using 5-fold cross-validation scheme.

Table 1. Average accuracy rate (in %) using SVM before applying LDA.

	Cross-Validation	**Accuracy (%) - 3-Fold cross-validation**			**Accuracy (%) - 4-Fold cross-validation**			**Accuracy (%) - 5-Fold cross-validation**			**Accuracy (%) - Leave-one-out**		
S. No	**Activity Name**	**fv1**	**fv2**	**fv3**	**fv1**	**fv2**	**fv3**	**fv1**	**fv2**	**fv3**	**fv1**	**fv2**	**fv3**
1	Backward_Fall	54.2	62.5	62.5	66.7	16.7	66.7	100	60.0	80.0	25.0	41.7	70.8
2	Chest_Pain	66.7	45.8	45.8	83.3	66.7	83.3	20.0	60.0	40.0	8.33	37.5	62.5
3	Coughing	79.2	75.0	75.0	100	83.3	83.3	100	80.0	80.0	70.8	75.0	75.0
4	Faint	58.3	66.7	66.7	16.7	66.7	50.0	20.0	20.0	60.0	16.7	66.7	62.5
5	Forward_Fall	91.7	95.8	95.8	83.3	100	83.3	100	100	100	50.0	91.7	87.5
6	Headache	45.8	54.2	54.2	33.3	66.7	33.3	50.0	50.0	75.0	37.5	45.8	70.8
7	Taking_Medicine	91.7	95.8	95.8	100	100	100	80.0	100	100	83.3	95.8	91.7
8	Vomiting	83.3	83.3	83.3	66.7	6.7	83.3	80.0	100	80.0	29.2	79.2	75.0
	Average accuracy	71.4	72.4	72.4	68.7	70.8	72.9	68.7	71.2	**76.9**	40.1	66.7	74.5

4.3.2. Action Recognition with LDA

LDA is implemented on MOH feature descriptors of MHI, MDI, and their concatenation to get dimensionality of 1×7, 1×7, and 1×14 respectively. Multi-class SVM is trained using one-vs-one strategy and a polynomial kernel function. This experiment was performed using 5-fold cross-validation technique because in a previous experiment this scheme outperformed other cross-validation schemes. After applying LDA on fv1, fv2, and fv3, we attained average accuracies of 70.83, 75.0% and 97.91% respectively. These results are summarized in Table 2. LDA outperforms in recognition results over the simple MOH descriptors due to its ability to increase between-class variance and decrease the within-class variation.

Table 2. Average accuracy rate using LDA with 5-fold cross-validation scheme.

S. No.	**Activity Name**	**fv1**	**fv2**	**fv3**
1	Backward_Fall	66.67	66.67	100.0
2	Chest_Pain	50.0	50.0	100.0
3	Coughing	100.0	100	100.0
4	Faint	66.67	66.67	100.0
5	Forward_Fall	50.0	83.33	100.0
6	Headache	50.0	66.67	100.0
7	Taking_Medicine	100.0	100.0	100.0
8	Vomiting	83.33	66.67	83.33
	Average accuracy	70.83	75.00	**97.91**

Using fv3 feature descriptor after LDA, the corresponding results are shown in Figure 7 in the form of confusion matrix. In this case, only Vomiting activity is confused with the coughing activity due to their highly similar postures.

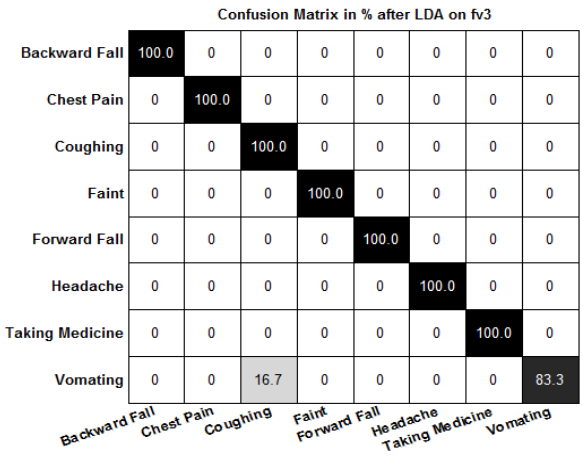


Figure 7. Confusion matrix (in %) after LDA on fv3.

4.3.3. Comparison

Considering the motion templates-based features, we now compare our results with those of [15]. This comparison is provided in Table 3. Histogram of Oriented Gradients (HOG) descriptors are extracted from MHI in [17]. We compare our results with those of [17] using 3, 4- and 5-fold cross-validation schemes and using leave one out validation scheme. In Table 3, we presented the average accuracy rates of our proposed approach using LDA on fv3 features. Results in Table 3 show that supportively using the features from the MHI and MDI leads to high recognition accuracy as compared to HOG-MHI [17] for all validation techniques. Our proposed method outperforms the MHI-HOG [15] due to two reasons. First, in MHI the motion is overwritten because of self-occlusion by the body parts. If an activity contains its necessary atomic actions in opposite directions (e.g., Backward_Fall, Faint, etc.), then motion history information of first atomic action is overwritten by the latter one. Thus, self-occlusion, during the activity of the moving body, overwrites the prior information. Secondly, in HOG the gradient orientations are computed globally which cannot characterize the different region around the body. For example, the movement of the head is much more than that of other body parts. In contrast, we proposed MOH, to describe the MDI and MHI, which uses hierarchical window partitioning to pay attention to the different region of the template individually.

Table 3. Comparison results of our proposed approach with using LDA of fv3 features.

Cross-validation scheme	Accuracy (%)- Our (LDA-fv3)	Accuracy (%)- HOG-MHI [17]
Leave-one-out	81.8	77.6
3-fold	85.9	75.5
4-fold	95.0	72.4
5-fold	**97.9**	73.7

As we discussed in Introduction (section 1) that no standard dataset is available for validation of patients' activities and we recorded a video dataset in hospital room environment under indoor conditions. Therefore, we are unable to perform the direct comparison with other similar methods as those are evaluated on other human action recognition datasets.

5. Conclusions

In this work, we proposed a computer vision-based technique for recognizing patient activities. A video dataset containing eight activities for patients under hospital room environment has been recorded. MDI as a view-based template for representation of activity in a video sequence has been proposed. The proposed representation is used along with the MHI representation to get a supportive feature vector. LDA is implemented on MOH based description of MHI, MDI and their concatenation. The proposed method obtained highest accuracy rate of 97.91% using multiclass SVM which shows that MHI and MDI when used together result in much better performance than when used individually.

For future work, results can be improved by introducing a weighting scheme to separate out MHIs that are most discriminant among actions. Moreover, instead of MHI/MOH, feature extraction on RGB videos can be introduced since the extraction of silhouettes is a problem because of illumination changes and moving background.

References

[1] Ahad M., Tan J., Kim H., and Ishikaw S., "Motion History Image: its Variants and Applications," *Machine Vision Applications*, vol. 23, pp. 255-281, 2012.

[2] Amin M., Zhang Y., Ahmad F., and Ho K., "Radar Signal Processing for Elderly Fall Detection, The Future in-Home Monitoring," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 71-80 2016.

[3] Anderson D., Keller J., Skubic M., Chen X., and He Z., "Recognizing Falls from Silhouettes," *in Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society*, New York City, pp. 6388-6391, 2006.

[4] Bobick A. and Davis J., "An Appearance-Based Representation of Action," *in Proceedings of 13th International Conference on Pattern Recognition*, Vienna, pp. 307-312, 1996.

[5] Bobick A. and Davis J., "The Recognition of Human Movement Using Temporal Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257-267, 2001.

[6] Chaquet J., Carmona E., and Fernández-Caballero A., "A Survey of Video Datasets for Human Action and Activity Recognition," *Computer Vision and Image Understanding, Elsevier*, vol. 117, no. 6, pp. 633-659, 2013.

[7] Davis J. and Bobick A., "The Representation and

Recognition of Human Movements Using Temporal Templates," *in Proceedings of Computer Vision and Pattern Recognition*, San Juan, pp. 928-934, 1997.

[8] Davis J., "Recognizing Movement Using Motion Histograms," Technical Report, MIT Media Lab, USA. Perceptual Computing Section Tech, 1998.

[9] Elmezain M. and Al-Hamadi A., "Vision-Based Human Activity Recognition Using LDCRFs," *The International Arab Journal of Information Technology*, vol. 15, no. 3, pp. 389-395, 2018.

[10] Gnanavel R., Anjana P., Nappinnai K., and Sahari N., "Smart Home System Using A Wireless Sensor Network for Elderly Care," *in Proceeding of 2nd International Conference on Science Technology Engineering and Management*, Chennai, pp. 51-55, 2016.

[11] Hsu C. and Lin C., "A Comparison of Methods for Multiclass Support Machines," *IEEE Transaction on Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.

[12] Khan Z. and Sohn W., "Abnormal Human Activity Recognition System Based on R-Transform and Kernel Discriminant Technique for Elderly Home Care," *IEEE Transaction on Consumer Electronics*, vol. 57, no. 4, pp. 1843-1850, 2011.

[13] Liu L. and Mehrotra S., "Detecting Out-of-Bed Activities to Prevent Pneumonia for The Hospitalized Patient Using Microsoft Kinect V2," *in Proceeding of IEEE 1st Conference on Connected Health: Applications, Systems, and Engineering Technologies*, Washington, pp. 364-365, 2016.

[14] Murtaza F., Yousaf M., and Velastin S., "Multi-View Human Action Recognition using Oriented Gradients (HOG) Description of Motion Images (MHIs)," *in Proceeding of 13th International Conference on Frontiers of Information Technology*, Islamabad, pp. 297-302, 2015.

[15] Murtaza F., Yousaf M., and Velastin S., "Multi-View Human Action Recognition Using 2D Motion Templates Based on Mhis and Their HOG Description," *IET Computer Vision*, vol. 10, no.7, pp. 758-767, 2016.

[16] Sánchez D., Tentori M., and Favela J., "Activity Recognition for the Smart Hospital," *IEEE Intelligent Systems*, vol. 23, no. 2, pp. 50-57, 2008.

[17] Wang L., Hsiao Y., Xie X., and Lee S., "An Outdoor Intelligent Healthcare Monitoring Device for the Elderly," *IEEE Transactions on Consumer Electronics*, vol. 62, no. 2, pp. 128-135, 2016.

[18] Webb A., *Linear Discriminant Analysis in Statistical Pattern Recognition*, John Wiley and Sons, 2002.

[19] Wister A. and Speechley M., "Inherent Tensions Between Population Aging and Health Care Systems: What Might the Canadian Health Care System Look Like in Twenty Years?," *Population Ageing*, vol. 8, pp. 227-243, 2015.

[20] Yu M., Yu Y., Rhuma A., Naqvi S., Wang L., and Chambers J., "An Online One-Class Support Vector Machine-Based Person-Specific Fall Detection System for Monitoring an Elderly Individual in a Room Environment," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 6, pp. 1002-1014, 2013.

Muhammad Adeel Abbas received the B.S. degree in Electronics Engineering from International Islamic University Islamabad, Pakistan in 2014 and received the M.S. degree in Electronics Engineering from University of Engineering and Technology Taxila, Pakistan in 2017. His research interests are in image processing and human activity recognition.

Fiza Murtaza is currently serving at Sino-Pak Center for Artificial Intelligence (SPCAI), PAF-IAST, Pakistan. She got her BS degree in Electrical (Computer) Engineering from COMSATS Institute of Information Technology, Abbottabad, Pakistan in 2013. She completed her MS in Computer Engineering with specialization in Signal and Image Processing from the University of Engineering and Technology (UET), Taxila, Pakistan, in 2015. She completed her PhD. in Computer Engineering with specialization in Computer Vision from UET, Taxila, Pakistan. Her PhD research title is "Temporal human action detection in long and untrimmed videos". Her area of research includes Artificial Intelligence, Computer Vision, Machine Learning, Human Action Recognition and Temporal Human Action Detection.

Muhammad Obaid Ullah received the B.Sc. (Hons.) and M.Sc. degrees in Electrical engineering from the University of Engineering and Technology (UET) at Taxila, Pakistan, in 2000 and 2006, respectively. He received the Ph.D. degree in Electrical engineering from the University of Manchester, Manchester, U.K., in 2012. He is currently working as a Professor with the University of Engineering and Technology, at Taxila, Pakistan. His current research interests include applied signal processing, sensor networks, and machine learning.

Muhammad Haroon Yousaf is currently serving as a Professor in Computer Engineering Department, University of Engineering and Technology Taxila, Pakistan. His research interests are image processing/computer vision. He is heading Swarm Robotics Lab under National Centre for Robotics and Automation. He has published more than fifty papers in International Conferences and Journals. He has been the recipient of Best University Teacher Award (2012-2013) given by Higher Education Commission (HEC) of Pakistan. He is serving as a TPC member for many conferences and reviewer for renowned journals. He is the Senior Member of IEEE and Professional Member of Pakistan Engineering Council.