Journal of
Mathematics and Informatics

Depth Motion Maps and Log-Gabor Features Based Human Action Recognition Using Support Vector Machine

Biplab Madhu, Md. Zahidul Islam **and** *Lasker Ershad Ali*[*]

Mathematics Discipline, Science Engineering and Technology School
Khulna University, Khulna-9208, Bangladesh.
[*]Corresponding author. Email: ershad@math.ku.ac.bd

Abstract. In this paper, an efficient scheme is introduced for the human action recognition from depth video sequence. In particular, we have proposed an effective feature extraction technique which is based on Depth Motion Maps (DMMs) and Log-Gabor wavelet. Firstly, three Depth Motion Maps (DMMs) are projected from front, side and top views to accumulate DMMs features. Then, the Log-Gabor filter is employed on DMMs to attain complex wavelet features illustration. In Log-Gabor feature, we have evaluated two parts of a depth sequence called real and imaginary information for front, side and top sequence of DMMs projections. After that, we have adopted principal component analysis to reduce the dimension of the feature matrix. Finally, support vector machine (SVM) is utilized for action classification. The experimental results on the public Microsoft Research Action3D dataset suggest that our proposed DMM-Log-Gabor features based action recognition method is efficiently achieved the best performance on human action recognition.

Keywords: Human action recognition, depth motion maps, Log-Gabor feature, principal component analysis, support vector machine.

AMS Mathematics Subject Classification (2010): 68T10

1. Introduction
Nowadays, computer is a machine which is very necessary and important in many fields. People who done a work practically which is now done by computer in minimum time with high accuracy. Since the improvement in the field of human computer communication is enriched, so it is a big research area to exhibit more human-computer processes. Human actions are not only the behavior of human body-parts but also it is the implementation of human objects, motions and thoughts. In addition, the term action recognition means the identification of an activity or action by using a process that investigates the video data to learn about the actions completed and practices. It attained knowledge to further identify the similar actions established by Aggarwal et al. [1]. This research is inspired by progressive use of human action recognition which is a significant zone in computer vision due to its many applications such as intelligent video

surveillance [3], human computer interaction [2], detect dangerous events, security surveillance, entry/exit control, video retrieval, analysis of sports events, healthcare system, gestures recognition, monitor people living alone, robotics, telemedicine, content-based video search, video game etc. Understanding human action can help to find illegal events such as cheating, snatching, thefts, violent actions, etc. It is a significant challenge to identify human actions from RGB video for a long time. For the reason that the appearance of depth sensors such as the Microsoft Action, the inconvenience of human action recognition has been diminished as it can track the human motion. However, due to the several characteristics of human behaviors, it is still tough to extract the feature from the depth of the motion of the action to identify actual human actions. Specifically, there are numerous difficulties to be solved in action recognition, such as sudden pauses, nonlinear stretching, repetitions and to extract suitable feature.

In this paper, we have described how to identify human actions from depth image sequence. However, some of the prevailing methods immediately operate on point of depth maps where the calculation is very large. An example of depth map sequence is shown in Figure 1. In this work we suggest an efficient method to recognize human actions by extracting Log-Gabor features from depth motion maps. Firstly, for every depth video, all its video frames are projected onto three orthogonal Cartesian planes to make the projected maps equivalent to three projection views (front, side and top). For each projection view, the collection of absolute differences between consecutive projected maps forms the matching DMMs (i.e., DMM_f, DMM_s, DMM_t) [4]. The depth maps sequences are separated into segments and then the Log-Gabor filters are applied to every segment to extract features. The produced features in a DMM are presented as a feature vector. Since there are numerous DMMs for a depth video, the feature-level orientation method is examined using Gabor function to learn machines. In order to analyze our introduced method, we perform it on a benchmark depth action dataset known Microsoft Research Action3D dataset and we compare our recognition accuracy with other methods.

The other parts of our paper are arranged as follows. In section 2, some related works have been discussed. In section 3, proposed methodology has been deliberated. In section 4, Feature extraction and classification has been stated. Also, in section 5, Experimental setup has been deliberated. In section 6, Results and discussion have been narrated in brief. Again, in section 8, Conclusions have been alluded. Finally, equivalent helping sources have been alluded as references.

Figure 1: Depth map sequence for the *High throw* action

2. Related work

Action means the assortment of well-organized attitudes of human. Besides, for action recognition two key issues are measured such as the illustration of human behavior features, and also the model of human movement. Although, based on depth images, various illustrations have been established for action recognition, like hyper-surface normal [5,6], cloud points [7,8,9], skeleton joints [10,11] and depth motion maps (DMMs) [12] etc. The remaining depth-based action recognition ways uses depth maps or skeletal joint's position for human actions identification. Our work relies on DMMs since they will effectively capture the behavior and clues of human actions. Yang et al. [12] introduced DMMs in their work, they projected depth maps onto three orthogonal planes and accumulated global actions through entire video sequences to create DMMs, then histogram of oriented gradient (HOG) features were calculated from DMMs as the illustration of the sequences. Liu et al. [15] signifies every depth frame using as a bag of 3D points on the human silhouette, and employs HMM to model the dynamics of action. Chen et al. [13] changed the DMMs by neglecting the threshold for real-time action recognition, and they castoff local binary patterns (LBP) to illustrate texture information of DMMs [14]. Wang et al. [16] investigated random occupancy pattern (ROP) features were extracted from depth sequence using a weighted sampling theme. A distributed writing method was implemented to encipher ROP features and the features were shown as robust to occlusion. More recently, Wang et al. [17] designed three DMMs and every DMM served as an input to a Deep Convolutional Neural Network (D-CNN) for classification. In all of above works, DMMs were designed from the whole video sequence and also the temporal information of an action. So, it is tough to recognize two actions with likeactions but opposite temporal orders, like actions "stand up" and "sit down".

In fact, DMMs cannot accommodate to behavior speed variations that ends up intra-class variations. Vemulapalli et al. [18] planned a body part-based skeleton illustration was projected to model the space geometry connection between body essentials. After that human actions were sculptured as curves employing a Lie group. Log-Gabor wavelets model pretty well the receptive field profiles of cortical simple cells. The Log-Gabor wavelet illustration, captures salient visual characteristics like spatial localization, orientation selectivity, spatial frequency characteristic etc. Lades et al. [19] founded the utilization of Log-Gabor wavelets for face recognition by means of the Dynamic Link Architecture (DLA) framework. Further, The DLA starts by calculating the Gabor jets, so it achieves a flexible template comparison between the resulting image decompositions victimization graph-matching. Wiskott et al. [22] have enlarged on DLA when they established a Gabor wavelet based elastic bunch graph corresponding method to label and recognize human faces. Based on the 2-DGabor wavelet representation and also the labeled elastic graph matching, Lyons et al. [20, 21] planned an algorithm for two-class classification of gender, race, and facial feature. The algorithm contains two steps such as registration of a grid with the face discrimination either elastic labeled graph matching [19,22] or manual description of 34 points on every face image [20]. Donato et al. [23] have recently shown through examines that the Gabor wavelet illustration provides performance than alternative techniques for classifying facial actions.

Biplab Madhu, Md. Zahidul Islam and Lasker Ershad Ali

In this paper, we have a tendency to propose an efficient method to handle the above problems. At first, a completely unique descriptor called as Depth Motion Maps (DMMs) is proposed to capture the sequential material of actions. The DMMs are designed over multi-size sequential windows, therefore they encode more specifics of motion and shape information. Secondly, Log-Gabor features evaluation the information of DMMs are extracted to further improvement of the discriminative capacity of our descriptors. So, we have investigated the DMM-Log Gabor as feature. Thirdly, after diminishing the dimensions by Principle Component Analysis (PCA), the final illustrations are classified by Support vector machine (SVM). The proposed technique cannot only represent the motion and shape information, but also take sequential order and action speed differences into attention. Our method is estimated on Microsoft-Action3D datasets and achieves better performance over the existing approaches.

3. Methodology

An enormous investigation has been showed within the fields as well as Human action recognition employing a variety of methods. So, here we have suggested a new procedure using DMM-Log Gabor feature extraction for human action recognition. This procedure has enriched the accuracy of action recognition compared to some previous investigators. In Figure 2 the details of our proposed method have been declared. Also, the process has been discussed step by step in the following illustrations.
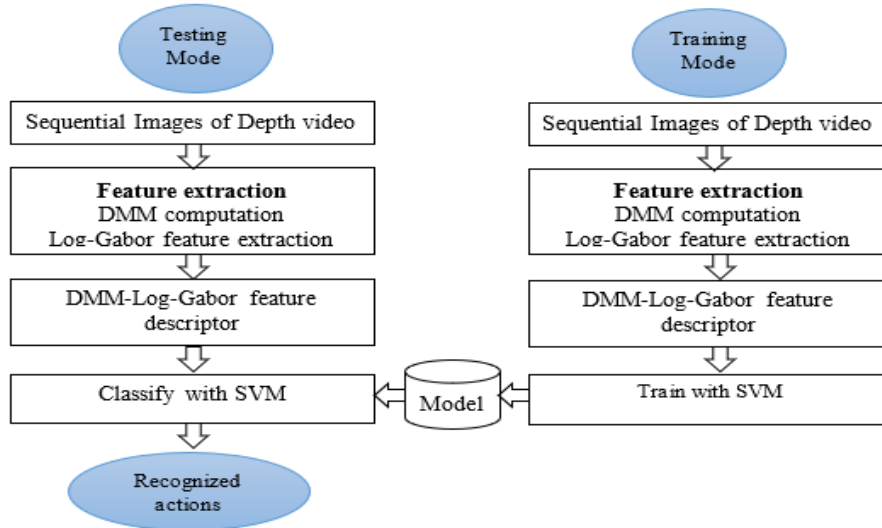
Figure 2: Proposed frame works

From depth map sequences, we have discussed about feature extraction and classification techniques. Generally, feature extraction may be a little variety in appearance, background, perspective, and activity performing. For this reason, the descriptor must be sufficiently free to take into attention influential classification of human action. Then we should examine these features by suitable classifier for classification of human action. The feature extraction and classification approaches are discussed below.

Depth Motion Maps and Log-Gabor Features Based Human Action Recognition Using Support Vector Machine

3.1. Depth motion maps as feature extraction

Depth map can be accustomed to capture the 3D structure and shape information. Yang et al. [24] planned to project depth frames onto three orthogonal Cartesian planes for the determination of describing the motion of an action. Because of its computational simplicity, the same method in [24] is adopted in this work while the process to obtain DMMs is modified. More specifically, each 3D depth frame is castoff to create three 2D projected maps corresponding to front, side, and top views, denoted by $map_{\{f,s,t\}}$. For a point (x, y, z) in a depth frame denoting the depth value in a right-handed coordinate system, the pixel value in three projected maps is showed by z, x and y respectively. Different from [24], for every projected map, the motion energy is calculated here as the absolute difference between two consecutive maps without thresholding. For a depth video sequence with N frames, $DMM_{\{f,s,t\}}$ is attained by loading the motion energy across the total depth video sequence as follows:

$$DMM_{\{f,s,t\}} = \sum_{j=a}^{b} \left| map_{\{f,s,t\}}^{j+1} - map_{\{f,s,t\}}^{j} \right|, \tag{1}$$

where j represents the frame index; $map_{\{f,s,t\}}^{j}$ is the projected map of jth frame under projection view $\{f,s,t\}$; $a \in \{2,...,N\}$ and $b \in \{2,...,N\}$ denote the starting frame and the end frame index. It should be noted that not all the frames in a depth video sequence are used to generate DMMs. This point is discussed further in the experimental setup section. A bounding box is then set to extract the non-zero region as the foreground in each DMM.

Let the foreground extracted DMM be denoted by $DMM_{\{f,s,t\}}$ hereafter. An example of $DMM_{\{f,s,t\}}$ generated from the *High through* sequences is shown in Figure3. DMMs from the three projection views effectively capture the characteristics of the motion in a distinguishable system. So, the reason here for using DMMs as feature descriptors for action recognition. Since $DMM_{\{f,s,t\}}$ of different action video sequences may have different sizes, bi-cubic interpolation is used to resize all $DMM_{\{f,s,t\}}$ under the same projection view to a fixed size in order to reduce the intra-class variability, for example due to different subject heights.

3.2. Log-Gabor feature representation

Gabor wavelet is a wavelet which is invented by Dennis Gabor. Gabor wavelet is introduced to image analysis due to their biological relevance and computational characteristics [27]. The Gabor wavelets, whose kernels are almost identical to the 2-D specific field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domain. The Log-Gabor wavelet illustration of an image has been achieving a great popularity. It is first proposed by Field D. [26]. It is also defined in

the frequency domain because of the uniqueness in the log function at origin. The Log-Gabor function is defined as following:

$$G(f) = \exp\left(\frac{-(\log(f / f_0))^2}{2(\log(\sigma / f_0))^2}\right) \tag{2}$$
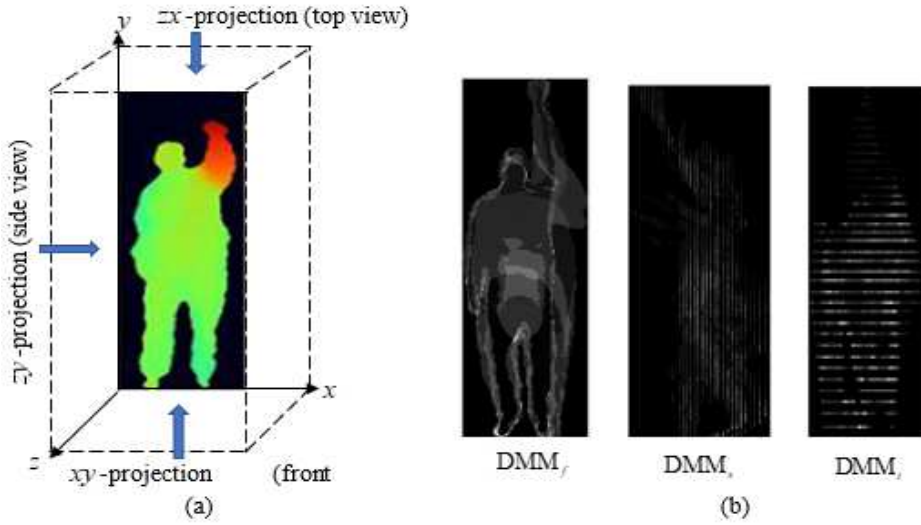
Figure 3: (a) Projection views of depth video frame and (b) $DMM_{\{f,s,t\}}$ generated from High throw action video sequence.

where f_0 and σ are the parameters of the Log-Gabor filter. The parameter f_0 will give the center frequency of the Log-Gabor filter and σ will affects the bandwidth of the Log-Gabor filter. To maintain the similar shape when the frequency is changed. For this, the ratio σ / f_0 should remain constant. Log-Gabor filter have a null DC component and which can be optimized to produce filters with minimal spatial extent. Log-Gabor filter can be divided into two mechanisms: radial and angular filters. So, Log-Gabor filter is defined as follows:

$$G(f, \theta) = \exp\left(\frac{-(\log(f / f_0))^2}{2(\log(\sigma_f / f_0))^2}\right)\exp\left(\frac{-(\theta - \theta_0)^2}{2(\sigma_\theta)^2}\right) \tag{3}$$

where f_0 is the center frequency, θ_0 is the orientation angle, σ_f determine the width parameter for the frequency, σ_θ determine the parameter of orientation.

3.3. Support vector machine classification
Support Vector Machine (SVM) is a classification and regression forecast tool that practices to maximize projecting accuracy while automatically avoiding over-fit to the data. It is a supervised learning algorithm which is also known as Support vector network. Vladimir N. Vapnik and Alexey Ya. Chervonenkis invented the original SVM

algorithm in 1963. Depending on the nature of the data, such a separation might be linear or non-linear.

Let us consider a linear classifier (or, hyperplane) $f(x) = w^T x + b$

where w represents weight vector, x is the input feature vector and b represents the position of the hyperplane. Here,
i)   if the input vector is 2-dimensional, the linear equation will represent a straight line.
ii) if the input vector is 3-dimensional, the linear equation will represent a plane.
iii) if input vector more than 3-dimension, the linear equation will represent a hyperplane.

The SVM algorithm is to find an optimal hyperplane for classification of two classes. Assume that the equation of hyper plane is $w^T x + b = 0$. The distance between $w.x + b = +1$ and $w.x + b = -1$ is the margin of this hyperplane. By applying the distance rule between two straight lines, we get the margin, $m = \dfrac{2}{\|w\|}$
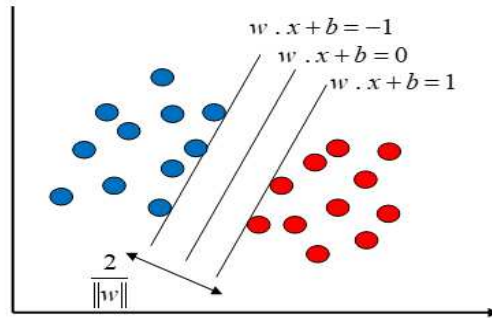
Figure 4: Maximum margin of hyperplane for SVM trained with samples

For non-linear classifier SVM use kernel function to separate the data points. Obtain a nonlinear SVM regression model by replacing the dot product $x_1^T . x_2$ with a nonlinear kernel function $K(x_1, x_2) = <\varphi(x_1), \varphi(x_2)>$ where, $\varphi(x)$ is a transformation that maps x to a high-dimensional space. Popular kernel functions are

i)   Linear kernel function: $K(x_j, x_k) = x_j^T x_k$

ii) Gaussian kernel function: $K(x_j, x_k) = \exp(-\|x_j - x_k\|^2)$

iii) Polynomial kernel function: $K(x_j, x_k) = (1 + x_j^T x_k)^q$

iv) Radial basis kernel function: $K(x_j, x_k) = \exp(-\dfrac{\|x_j - x_k\|^2}{2\sigma^2})$

where x_i and x_j are support vector where support vector is the input vectors that just touch the boundary of the margin. Simply, support vectors are the data points that lie closest to the decision surface (or hyperplane). DMM-Log Gabor featured based human action recognition we have applied these kernels for better classification.

4. Experimental setups

In this section, it is explained however our methodology was applied to the general public domain Microsoft Research (MSR) Action3D dataset [15] with the depth map sequences captured by associate an RGBD camera. Our method is then compared with the prevailing strategies. The MSR-Action3D dataset includes 20 actions performed by 10 subjects. The sort of 20 human actions are: *high wave, horizontal wave, hammer, hand catch, forward punch, high throw, draw x, draw tick, draw circle, hand clap, two hand wave, side boxing, bend, forward kick, side kick, jogging, tennis swing, tennis serve, golf swing, and pickup throw*. Every subject achieved every action 2 or 3 times. Each subject performed the same action differently. As a result, the dataset incorporated the intra-class variation. For instance, the speed of performing an action varied with different subjects. The resolution of every depth map was 320×240. To facilitate a fair comparison, the same experimental settings is done in [15,29-31]were considered. The actions were divided into three subsets (AS1, AS2, AS3) which are listed in Table 1. For every action subset, three totally different tests were performed. In our method for Test One, 1/3 of the illustrations were used as training examples and the others as test examples. In Test Two, 2/3 of the examples were used as training examples and the others as test examples. In Cross Subject Test (or Test Three), half of the themes were used as training and the others as test themes. In the investigational system described in [15], in Test One (or Two), for every action and every subject, the first (or first two) action sequences were castoff for training, while in Cross Subject Test, subjects 1, 3, 5, 7, 9 (if existed) were castoff for training. Noting that the examples or themes used for training and testing were static, they are discussed to as Fixed Tests here.

Table 1: Three action subsets and test used for Microsoft-Action3D dataset

Action set 1 (AS1)	Action set 2 (AS2)	Action set 3 (AS3)
Horizontal wave (2)	High wave (1)	High throw (6)
Hammer (3)	Hand catch (4)	Forward kick (14)
Forward punch (5)	Draw X (7)	Side kick (15)
High throw (6)	Draw tick (8)	Jogging (16)
Hand clap (10)	Draw circle (9)	Tennis swing (17)
Bend (13)	Two hand waves (11)	Tennis serve (18)
Tennis serve (18)	Forward kick (14)	Golf swing (19)
Pickup throw (20)	Side boxing (12)	Pickup throw (20)

Another research was showed by arbitrarily selecting training examples or training theme equivalent to the three tests. On the other hand, the action sequences of every theme for every action were arbitrarily selected to assist as training examples in Test One and Test Two. For Cross Subject Test, half of the themes were arbitrarily selected for training and the rest castoff for testing. These tests are denoted to as Random Tests here.

For every depth video sequence, the principal five frames and the last five frames were detached and also the enduring frames were castoff to create $DMM_{\{f,s,t\}}$. The goal of this frame elimination was two-fold. Initially at the starts and also the finish, the subjects were generally at a stand-still position with solely slight body actions, which did not pay to the motion appearances of the video sequences. Second, in our method of making DMMs, small activities at the beginning and the end lead to in a stand-still body

shape with large pixel values along the edges which paid to a large quantity of renewal error. Therefore, the primary and terminal frame elimination was done to eliminate no motion condition. Other frame choice systems may be castoff here to attain the same. To get three fixed sizes for $DMM_{\{f,s,t\}}$, the sizes of the DMMs of all the examples (training and test samples) were create under every projection view. For the training feature set and the test feature set, principal component analysis (PCA) was applied to diminish the dimensionality. The PCA transform matrix was intended using the training feature set and then applied to the test feature set. This dimensionality reduction step provided computational effectiveness for the classification.

4.1. Evaluation of DMM-Log Gabor
We first analyze the effect of DMM normalization size for recognizing performances. As elaborate in Section 4.1, we normalize the three depth motion maps to a fixed size. Then we evaluate the effect of Log-Gabor filter to extract new feature. For applying Log-Gabor feature every image transform into a real and imaginary image label. Since DMM is measured from the front, side and top. As for action set AS1 the size of front, side and top are respectively 100×50, 100×82, 82×47. While for action set cross subject AS1CrSub the size of front, side and top are respectively same action set AS1. For action set AS2 the size of front, side and top are respectively 102×51, 103×67, 67×51. While for action set cross subject AS2CrSub the size of front, side and top are respectively same action set AS2.And for action AS3 the size of front, side and top are respectively 104×53, 104×84, 84×53. While for action set cross subject AS3CrSub the size of front, side and top are respectively same action set AS3.Although lower resolutions are able to reduce computational cost in computing DMM, we extract Log-Gabor filter for each depth motion maps from front, side and top. So, for each video sequence is converted to real and imaginary parts. We examine every depth image for real, imaginary and real-imaginary with respectively. The difference of computation time between different and Log-Gabor is differed. The following experimental results are based on this process. After Log-Gabor filtering the real, imaginary and real-imaginary depth sequence is shown in Figure5, In Cross Subject Test, different subjects perform actions with great variations but the number of subjects is limited, which results in considerable intra-class variations. Furthermore, some actions in AS2 are quite similar, e.g. Draw X, Draw Tick, and Draw Circle, which generates small inter-class variations. The performances on cross subject test might be enhanced by accumulating in more subjects.

4.2. Parameter selection
In Log-Gabor wavelet feature, an essential parameter is wavelength λ which monitoring the related effect of the orientation term in the optimization step. Many methods have been presented in the literature for discovery an ideal value for this filtered constraint. To discovery an ideal accuracy, a set of ethics were examined. For optimal appreciation rates with different ethics of λ was examined for action set AS1, AS2, AS3 and Fixed Cross Subject Test. For optimal action recognition we change the parameter value such scale parameter υ, minimum wavelength λ was also achieved with the different set of morals. For individual value of υ and λ the testing was recurrent 5 times.
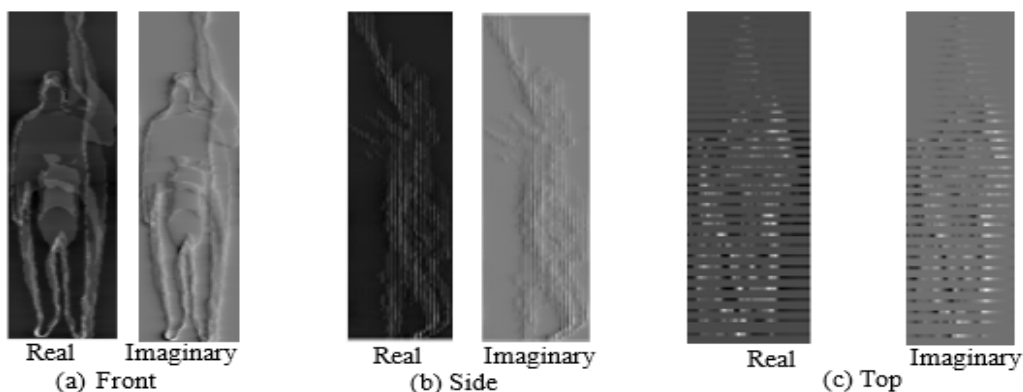
Figure 5: Real and imaginary sequence of (a) Front view (b) Side view and (c) Top view

In our experiment, in test one test one for action set AS1, we choose the parameter value number of scale parameter $\upsilon = 4$, minimum wavelength $\lambda = 3$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 93.4%. In test two, we choose the parameter value number of scale parameter $\upsilon = 6$, minimum wavelength $\lambda = 0.92$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 92%. In test three, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 95.4%. In test four, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=0.2 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 94%. In test five, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=1.2 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 96.7%.

For action set AS2, in test one, we choose the parameter value number of scale parameter $\upsilon = 4$, minimum wavelength $\lambda = 3$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 90.1%. In test two, we choose the parameter value number of scale parameter $\upsilon = 6$, minimum wavelength $\lambda = 0.92$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 94.2%. In test three, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 93.4%. In test four, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=0.2 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 87.5%. In test five, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=1.2 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 97.4%.

Then for action set AS3, in test one, we choose the parameter value number of scale parameter $\upsilon = 4$, minimum wavelength $\lambda = 3$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 94.3%. In test two, we choose the parameter value number of scale parameter $\upsilon = 6$, minimum wavelength $\lambda = 0.92$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 96.6%. In test three, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 97.3%. In test four, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=0.2 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 95.3%. In test five, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=1.2 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 96.0%.

In our experiment for cross subject test in test one, we choose the parameter value number of scale parameter $\upsilon = 4$, minimum wavelength $\lambda = 3$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 86.2%. In test two, we choose the parameter value number of scale parameter $\upsilon = 6$, minimum wavelength $\lambda = 0.92$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 89.2%. In test three, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=1.7 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 88.6%. In test four, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=0.2 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 87%. In test five, we choose the parameter value number of scale parameter $\upsilon = 5$, minimum wavelength $\lambda = 3$, multiplication time=1.2 and sigma on frequency $\sigma = 0.15$. For these parameters our method recognizes the action 90.1%.

The optimal recognition rates are found at $\upsilon = 4, \lambda = 3$. As a result, in all the experiments reported here, the value of $\upsilon = 4, \upsilon = 4, \sigma = 0.15$ were thus chosen. Above discussed results are shown in figure 6.

5. Results and discussions
In this section, we have discussed the recognition result and how we have got these results by our proposed method. So, the relatively recognition result and comparison our proposed method with existing method are discussed below.
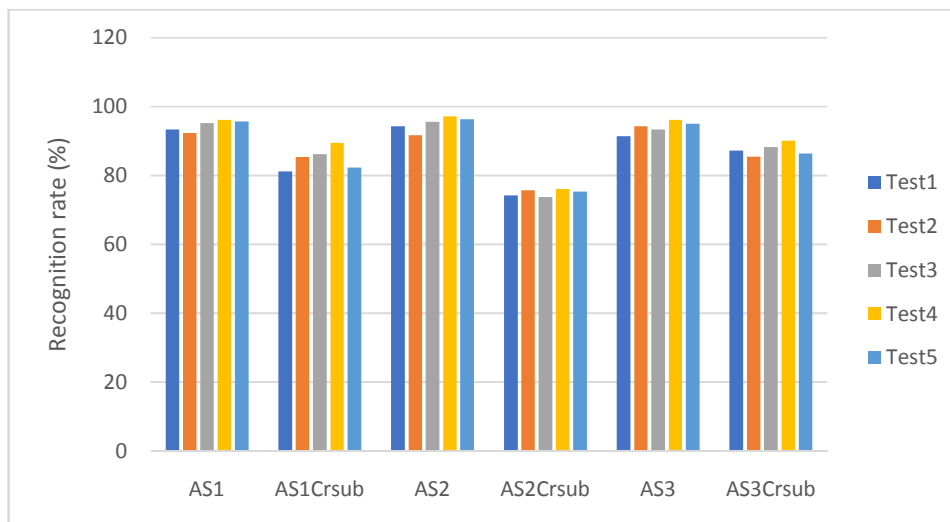
Figure 6: The recognition rates (%) of DMM-Log-Gabor with different parameter values under a variety of test sets

5.1. Recognition results

In this approach, we have proposed an easy algorithm using a novel feature descriptor, because the proposed algorithm would not be useful in practical applications if it needs tuning lots of parameters. Therefore, we want to find our smart assignments for the parameters within the algorithm and then fix these parameters for all test cases. For instance, in the step of Log-Gabor wavelet, how to control the DMM size, the number of action levels and the assortment of the wavelength? In order to improve the computational efficiency in the classification step, Principle Component analysis (PCA) is adopted to reduce the dimension of the Log-Gabor feature vector. The PCA transform matrix is calculated using the training feature set and then applied to the test feature set.

In our experiments, we resized DMM to three different sizes of front, side, top and then utilized Log-Gabor filter for action set AS1 are respectively 100×50, 100×82, 82×47, Then we transformed it into Log-Gabor filter. Every DMMs sequence divided into two images after Log-Gabor feature apply. The front image is featured into real part and imaginary part. Then along real and imaginary is featured into a new featured as total front feature. We used real, imaginary and total featured individually for recognition. After feature extraction of front, side and top we used SVM for training and testing for classification. Since Log-Gabor is a wavelet feature, it has some parameter. In Log-Gabor feature we use the number of scales, wavelength, multiplication and ratio between window width and wavelength which are discussed in (4.2).

Firstly, we adopted our method for real part of front, side and top. In real part of these images, we recognized the Action 96.0% whereas for cross subject test the recognition action is 85%. These action recognition rates are evaluated by the average recognition rate of AS1, AS2 and AS3. Secondly, we adopted our method for imaginary part of front, side and top. In imaginary part of these images, we recognized the Action 97.5% whereas for cross subject test the recognition action is 78%. These action recognition rates are evaluated by the average recognition rate of AS1, AS2 and AS3.

Finally, we adopted our method for total part of front, side and top. In total part of these images, we recognized the Action 96.3% whereas for cross subject test the recognition action is 85.1%. These action recognition rates are evaluated by the average recognition rate of AS1, AS2 and AS3. The overall recognition rate is 92.8%.

Table 2: Recognition rates (%) of our method compared to the others method on MSR Action3D dataset

|	Liu et al. [15]	Xia et al. [29]	Yang et al. [30]	Ours
Test one				
AS1	89.5	**98.5**	94.7	96.0
AS2	89.0	96.7	95.4	**97.2**
AS3	96.3	93.5	**97.3**	96.1
Average	91.6	96.2	95.8	**96.4**
Test two				
AS1	93.4	**98.6**	97,3	97.8
AS2	92.9	97.2	**98.7**	94.6
AS3	96.3	94.9	97.3	**98.7**
Average	94.2	97.2	**97.8**	97.0
CST				
AS1	**92.9**	88.0	74.5	89.1
AS2	71.9	**85.5**	76.1	76.2
AS3	79.2	63.6	**96.4**	90.2
Average	74.7	79.0	82.3	**85.1**

5.2. Comparison with state-of-the-art methods
In this study we compare our DMM-Log Gabor technique with some related existing techniques on the Microsoft Action3D dataset. The comparison results are shown in Table 2. The highest average recognition rate for each test case is highlighted in bold. The average recognition rates of our method outperform all the other methods listed in Table 2 for each test case except the method proposed by Yang et al. [30]. We have to point out that Yang et al. used several parameters for the proposed learning, and the values of all parameters were not fixed. So, it's very clear that we should governance the experiment on each range for many times to find a suitable corresponding value of the parameter. However, if we don't use fixed values for all parameters our method is also capable of showing expected higher performance, like the method reported in [15]. To suggest a suitable and faster algorithm with higher recognition accuracies, we fixed the parameter values.

This paper presents a simple method for higher recognition accuracies and uses parameter tuning. Three confusion matrices for the cross-subject test are shown in Figure 8. The corresponding number of each action is used to understand the classification accuracy and error. From the confusion matrices, it can be easily understood the fitness of the proposed algorithm with the actions. In every test, more than half of the recognition accuracy has arrived at100 percent. The lowest recognition accuracy rate also arrived at 91 percent. Although the introduced technique is based on DMM-Log Gabor, our

Biplab Madhu, Md. Zahidul Islam and Lasker Ershad Ali

recognition rate is better than [15,29] which is shown in Figure 7. It can be seen that the DMM-LogGabor descriptor shows higher state of being more acc.
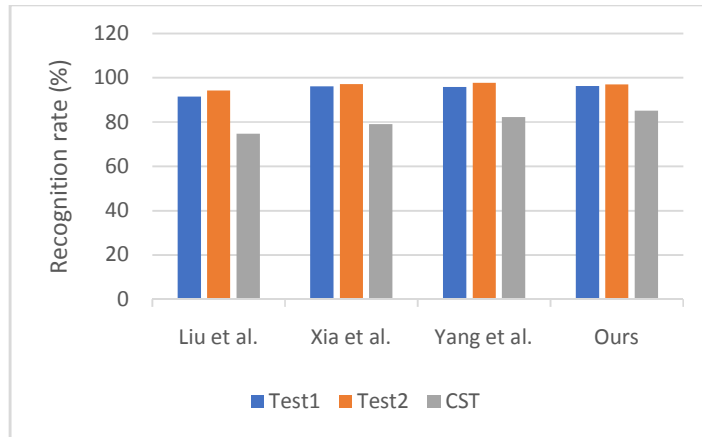
Figure 7: Recognition rates (%) of comparison with the state-of-the-art methods

(a)

(b)

(c)
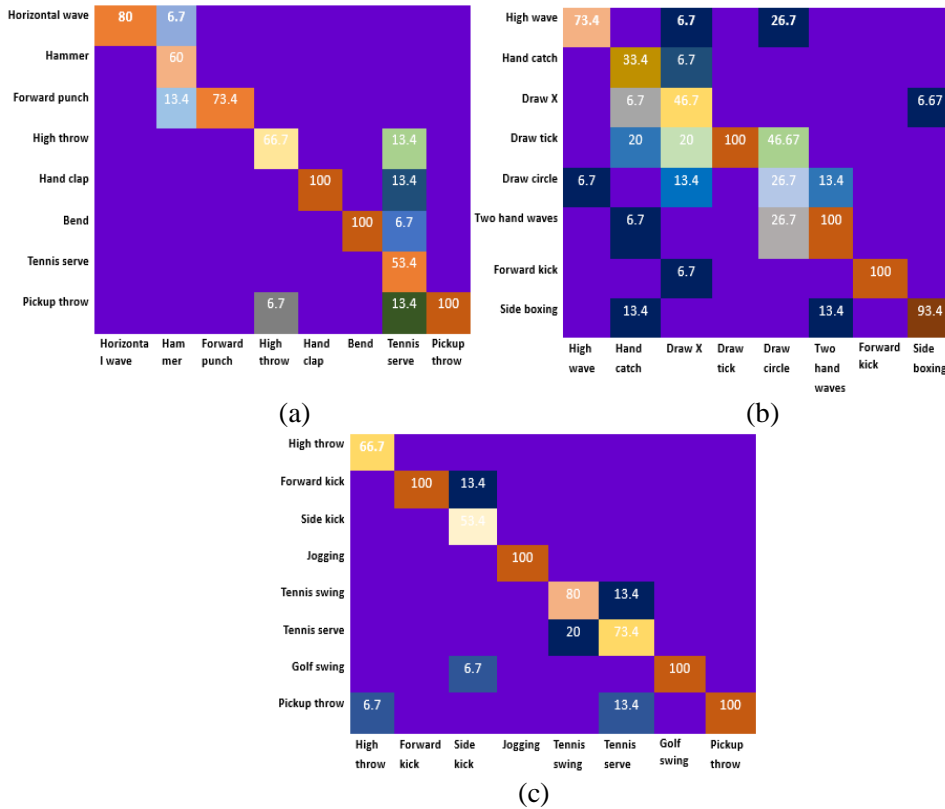
Figure 8: Confusion matrix of proposed method: (a)AS1(CST), (b)AS2(CST), (c) AS3(CST).

86

6. Conclusion

We have investigated an efficient technique for human action recognition and established a novel feature descriptor known as DMM-Log-Gabor feature extraction technique in this research. For recognizing human action, support vector machine is adopted as the basic classifier in this approach. The experimental results on MSR Action3D dataset reveal that our approach considerably outperforms of the present progressive techniques. Since, Log-Gabor descriptor transform an image in two part called real and imaginary. We have investigated our method both real and imaginary part of front, side and top views. After that we have adopted the real and imaginary features with one descriptor called DMM-Log Gabor. Then PCA is employed to reduce the dimension of data set. Our mentioned algorithm improves the average recognition rate of human action than Liu et al. [15] and Xia et al. [29]. Therefore, it is shown by the empirical results that DMM-Log-Gabor feature descriptor is better than the above-alluded existing recognition technique.

REFERENCES

1. J.K.Aggarwal, M.S.Ryoo and J.K.Aggarwal, Semantic representation and recognition of continued and recursive human activities, *International Journal of Computer Vision*, 82(1) (2009) 1-24.
2. I.Rodomagoulakis, N.Kardaris and V.Pitsikalis, Multimodal human action recognition in assistive human-robot interaction, in *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP) (2006) 2702–2706.
3. W.Lin, M.Sun, R.Poovendran and Z.Zhang, Activity recognition using a combination of category components and local models for video surveillance, IEEE transactions on circuits and systems for video technology, 18(8) (2008) 1128–1139.
4. C.Chen, K.Liu and N.Kehtarnavaz, Real-time human action recognition based on depth motion maps, *Journal of Real-Time Image Processing,* 2013.
5. X.Yang and Y.Tian, Super normal vector for activity recognition using depth sequences, in *IEEE Conference on Computer Vision and Pattern Recognition*, (2014) 804– 811.
6. O.Oreifej and Z.Liu, Hon: Histogram of oriented 4Dnormals for activity recognition from depth sequences, in *IEEE Conference on Computer Vision and Pattern Recognition*, (2013) 716–723.
7. W.Li, Z.Zhang and Z.Liu, Action recognition based on a bag of 3d points, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, (2010) 9–14.
8. J.Wang, Z.Liu, J.Chorowski, Z.Chen and Y.Wu, Robust 3d action recognition with random occupancy patterns, in *ECCV*, (2012) 872–885.
9. L.Xia and J.K.Aggarwal, Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* (2013) 2834–2841.
10. X.Yang and Y.Tian, Eigenjoints-based action recognition using naive-bayes-nearest-neighbor, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, (2012) 14–19.

11. M.Zanfir, M.Leordeanu and C.Sminchisescu, The moving pose: An efficient 3d kinematics descriptor for low-latency action recognition and detection, in *IEEE International Conference on Computer Vision*, (2013) 2752–2759.

12. X.Yang, C.Zhang and Y.Tian, Recognizing actions using depth motion maps-based histograms of oriented gradients, in *ACM International Conference on Multimedia,* (2012) 1057–1060.

13. C.Chen, K.Liu and N.Kehtarnavaz, Real-time human action recognition based on depth motion maps, *Journal of Real-Time Image Processing,* (2013) 1–9.

14. C.Chen, R.Jafari and N.Kehtarnavaz, Action recognition from depth sequences using depth motion maps based local binary patterns, *IEEE Winter Conference on Applications of Computer Vision*, (2015) 1092–1099.

15. W.Li, Z.Zhang and Z.Liu, Action recognition based on a bag of 3D points, *IEEE Workshop on CVPR for Human Communicative Behavior Analysis*.

16. J.Wang, Z.Liu, J.Chorowski, Z.Chen and Y.Wu, Robust 3D action recognition with random occupancy patterns, In *ECCV*, (2012) 872-885.

17. P.Wang, W.Li, Z.Gao, C.Tang, P.Zhang, and J.Ogunbona, Conv-nets-based action recognition from depth maps through virtual cameras and pseudo coloring, *ACM International conference on Multimedia,* (2015) 1119–1122.

18. R.Vemulapalli, F.Arrate and R.Challappa, Human action recognition by representing 3D skeletons as points in a Lie group, *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

19. M.Lades, J.C.Vorbruggen, J.Buhmann, J.Lange, C.Von-der Malsburg, R.P.Wurtz and W.Konen, Distortion invariant object recognition in the dynamic link architecture, *IEEE Transaction on Comput.,* 42 (1993) 300–311.

20. M.J.Lyons, J.Budynek, A.Plante and S.Akamatsu, Classifying facial attributes using a 2-D Gabor wavelet representation and discriminant analysis, in Proc. 4th *IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2000.

21. M.J.Lyons, J.Budynek and S.Akamatsu, Automatic classification of single facial images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21 (1999) 1357–1362.

22. L.Wiskott, J.M.Fellous, N.Kruger and C.Von-der Malsburg, Face recognition by elastic bunch graph matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (1997) 775–779.

23. G.M.Donato, S.Bartlett, J.C.Hager, P.Ekman and J.T.Sejnowski, Classifying facial actions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21 (1999) 974–989.

24. X.Yang, C.Zhang and Y.Tian, Recognizing actions using depth motion maps-based histograms of oriented gradients, *Proceedings of ACM International Conference on Multimedia*, (2012) 1057–1060.

25. J.G.Daugman, Complete discrete 2-D Log-Gabor transforms by neural networks for image analysis and compression, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36 (1988) 1169–1179.

26. D.Field, Relations between the statistics of natural images and the response properties of cortical cells, *Journal of Optical Society for America*, 4(12) (1987) 2379–2394.

27. J.Jones and L.Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, *J. Neurophys*, (1987) 1233–1258.

28. C.Liu and H.Wechsler, Evolutionary pursuit and its application to face recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (2000) 570–582.

29. L.Xia, C.Chen and J.K.Aggarwal, View invariant human action recognition using histograms of 3D joints, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops,* (2012) 20–27.

30. X.Yang and Y.Tian, Eigen joints-based action recognition using Naive-Bayes-Nearest-Neighbor, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (2012) 14–19.

31. A.Vieira, E.Nascimento, G.Oliveira, Z.Liu and M.Campos, Stop: Space-time occupancy patterns for 3D action recognition from depth map sequences, *Ibero American Congress on Pattern Recognition*, (2012) 252–259.