

Edge Adaptive Gradient Action Descriptor and Kernel Discriminant Analysis for Human Action Recognition

K. Ruben Raju

Research Scholar, Dept. of CSE,
Shri JJT University, Rajasthan India.

Dr. Yogesh Kumar Sharma

Head & Associate Professor, Dept. of CSE,
Shri JJT University, Rajasthan India.

Dr. Birru Devender

Associate Professor, Dept. of CSE, Holy
Mary Institute of Technology & Science,
Hyderabad, India.

Abstract- Human Action Recognition is a challenging issue in the real time constraints where the action videos or images are contaminated with several side effects like noises, moving backgrounds, multiple views, hindered movements etc. Under these constraints, to recognize an action, we have developed a new Human Action Recognition system. Under this system, an edge efficient action descriptor called as Laplacian Histogram of Gradients is proposed through which the all possible movements of an action are extracted. Further, to ensure a perfect discrimination between different action descriptors, we have employed kernel discriminant analysis. The proposed recognition model is evaluated systematically on a standard action dataset, IXMAS. Experimental results prove that our method outperforms the existing methods in terms of recognition accuracy.

Keywords: Human action recognition, Laplacian Gradient, Histograms, Kernel Discriminant Analysis, Support vector Machine, Recognition Accuracy.

I. INTRODCUTION

Human Action Recognition (HAR) and analysis [1], is one of the most active topics in computer vision, which has drawn increasing attention due to its widespread applicability in various applications including Robotics, Human-computer Interactions, Behavior analysis [2], Content based Retrieval, Video Indexing [3], Gesture Recognition, sports video analysis [4], and Visual Surveillance [5] etc. The main objective of a HAR system is to identify actions in a video sequence under different situations like occlusion, cluttering and different lighting conditions. The main center of this system is the computational algorithms which understand the human actions. Similar to the Human Vision System (HVS), these computational algorithms ought to produce a label after the analysis of partial or entire action in the video sequence. Developing such algorithms is typically addressed in the computer vision research, which studies the computers to gain high level understanding regarding human actions from digital images and videos.

Recognizing human actions from a video is a challenging task in many practical applications. Basically, the HAR is accomplished under three phases such as Pre-processing, Feature extraction and classification. Under pre-

processing, the input action video is subjected to the some preliminary operations to extract the exact motion region from action frame or video. Next, in the feature extraction phase, the action is described in such a way the key features of an action are represented effectively. Finally in the classification, the derived action descriptor is given as input for classified to recognize the action present in it. Feature extraction plays a major role in the HAR and to effectively recognize the action the design of feature extractor must be much effective. Several feature extraction methods are developed in earlier but they have several disadvantages [6-9]. Moreover, they are also sensitive to different viewpoints and camera movements and had shown a poor performance in such circumstances.

In this paper, we have proposed a new method for HAR based on the Gradient features and Discriminant analysis. Initially, we represent the action with a newly proposed descriptor called as Laplacian Histogram Oriented Gradients (LHOG). Next, over the obtained feature set, we employ kernalized discriminant analysis (KDA) to reduce the dimensionality of feature set. The final feature set is processed through Support Vector Machine (SVM) to classify the action.

Remaining paper is organized as follows; Section II explores the Literature survey details. Section III explores the complete details of proposed action recognition framework. The details of simulation experiments are stipulated in section IV and the concluding remarks are stipulated in section V.

II. LITERATURE SURVEY

From the past decade, several approaches have been developed, proposing a variety of methods for Human action recognition. Among these methods, Histogram of Gradients (HoGs) is one of most effective action descriptor through which the local motion regions are represented. Inspired with HOG, I.C. Duta et al., [10] proposed Histograms of Motion Gradients (HMG) based on the spatial derivation, which captures the changes between two consecutive action frames. Further for feature extraction, this work employed Shape Difference Vector of Locally Aggregated Descriptors (SD-VLAD) which brings the complementary information using the shape information.

Further, Jin Wang et al., [11] employed Pyramid Histogram of Oriented Gradient (PHOG) and two state-space models such as Hidden Markov Model (HMM) and Conditional Random Field (CRF) to characterize human figures for action recognition.

Further, Bo Lin and Bin Fang [12] proposed Spatio-Temporal Pyramid Histogram of Gradients (SPHOG) which is based on the gradient changes between successive frames. To incorporate the local information distribution into VLAD, a Gaussian Kernel is implanted to measure the weighted distance histograms of local descriptors. Further, combining the Distance mean histogram of gradients with segmented block of mean image with normalization for generating action descriptor. Random forest algorithm is employed for classification. Next, considering the gradient of motion, V. Thanikachalam and K.K. Thyagarajan [13] proposed an action recognition based on Accumulated Motion Image (AMI) in which the histograms are built based on the energy distributions. After the evaluation of AMI, Discrete Fourier Transform (DFT) is employed and mean and variance are measured. Finally the Dynamic Time Wrapping (DTW) is employed for training. V. Tripathi et al.

[21] proposed two algorithms; they are image normalization with the help of block mean and Distance Mean Histogram of Gradients (DMH) for action recognition. Random forest algorithm is employed for classification.

The next problem in HAR is larger size of feature set. The large size feature vector creates an extra computational burden for classification algorithm due to multiple time comparison. To solve this some, some authors tried to reduce the size of feature vector through standard

dimensionality reduction algorithms like Independent component analysis (ICA) [14], Principal Component Analysis (PCA) [15], Linear Discriminant Analysis (LDA) [16]. Yuting et al., [16] employed LDA for open view action recognition through which a common discriminant subspace is obtained for every action class. However, the LDA achieve optimal space by projecting linearly separated instances which is a not practical scenario. Further some subsequent discriminant analysis method are proposed such as Robust Linear Discriminant Analysis (RLDA) [17], Independent Component based LDA (IC-LDA), and Regularized Discriminant Analysis (RDA) [18]. However, all methods are assumed that the features are linearly related and tried to reduce the dimensionality by deriving only linear discrimination

III. PROPOSED APPROACH

This section describes the details of prosed action recognition framework in detail. The architecture of proposed framework is shown in Figure.1. Accordingly, the proposed framework is carried out in three phases. (1) Feature extraction, (2) Dimensionality reduction and (3) Classification. The main contribution of this paper is done in the feature extraction phase by developing a new feature descriptor, called as LHOG. Next, at dimensionality reduction technique, we have focused to reduce the dimensions of feature vector, because it is very larger sized vector. To reduce the dimensions, we have employed KDA. Finally the obtained feature vector is fed to classification and at this phase, we have employed SVM to classify the actions.

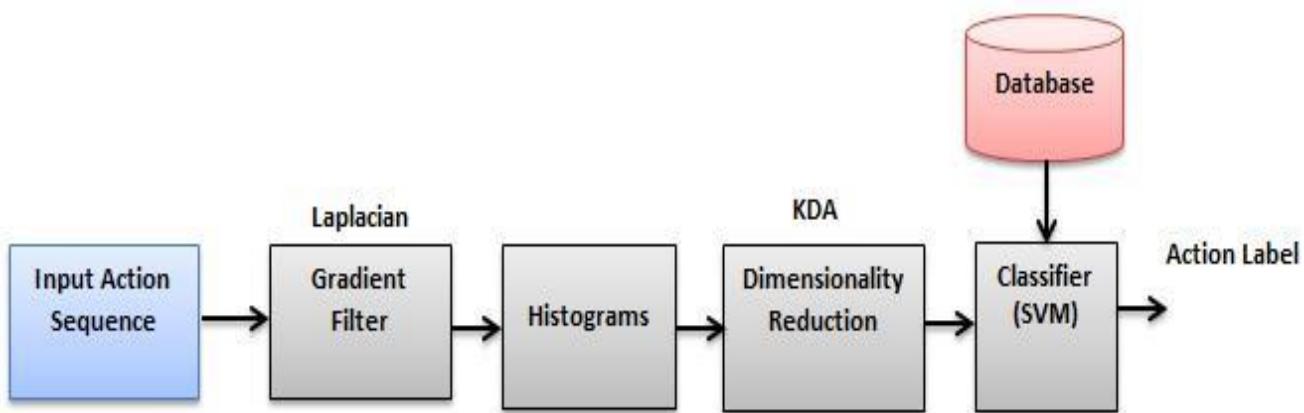


Figure.1 Block diagram of proposed recognition system

A. Feature Extraction

In this phase, we have focused to employ LHOG based feature extraction. In HAR, gradient features have more importance because the gradient of an image explores the fine details such as edges and sharp discontinuities. The main intention of gradient features is to explore the action

with respect to its direction of movement. Actually for a given action image, the pixel intensities vary with action movements and if such direction of movements are captured at the feature extraction phase, then the recognition system will become more effective. In this paper, for a pixel in the action image, to derive the direction of movements, we have

considered its neighbor pixels in horizontal and vertical directions. And a difference between the current pixel and its neighbor pixels gives information about the direction of movement. This flexibility can be gained through gradient operators.

Laplacian of gradient is one of the most powerful and effective gradient operator which derives the fine coarse details from the image. This fine detail helps in the detection of sharp discontinuities in the action image boundary. Figure.2 shows a simple example of the gradient features of an action image. With this inspiration we have adopted a one-dimensional Laplacian operator to capture the differences in an action image. There are two reasons behind this adoption of Laplacian operator for gradient features extraction.

- (1) Within an action image, the Laplacian operator can detect the fine details, highlight the edges, and also enhance the features with sharp discontinuities.
- (2) Laplacian is a second order derivative and hence it has a strong response towards the fine details than the first order derivatives, like gradient operator [19].

Due to these two reasons, the Laplacian operator over an action image can highlight regions of spontaneous changes in the pixel intensities and have it was used in several applications for blob and edge detection. For a given action image, the Laplacian operator will theoretically highlight the edges and boundaries.

Let's consider an action image A of size $M \times N$, where M is the row size and N is column size, first apply gradient operator ∇ on the action image, resulting in a first order gradient image, as

(1)

Where $\mathcal{L} = \nabla^2 = -\Delta$. Since the given input is a 2-D image, the gradient operator is employed in both

horizontal and vertical directions. Let the horizontal gradient

is and vertical gradient is , the overall gradient magnitude is computed as

$$= \sqrt{ \quad \quad \quad ^2 + \quad ^2} \quad \quad \quad (2)$$

Next, apply the gradient operator ∇ on the gradient image, resulting in a second order gradient image, as

$$= \nabla = \{-1, -2, -3, \dots\} \quad (3)$$

Where $\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} f_x \\ f_y \end{bmatrix}$, $f_x = \frac{\partial f}{\partial x}$, $f_y = \frac{\partial f}{\partial y}$. Here order gradient, f_x is the i th gradient feature in the and f_{i-1} is the $(i-1)$ th gradient feature of f . Since the second order gradient is also a 2-D object, the gradient operator is employed in both horizontal and vertical directions. Let the horizontal gradient is f_x and vertical gradient is f_y , the overall gradient magnitude is computed as

$$= \sqrt{2 + \frac{2}{2}} \quad (4)$$

The resultant is a second order derivative of an action image A. This is more helpful in the provision of sufficient discrimination between different actions. For example in the horizontal hand waving action, the movements are along horizontal direction and the gradient of such type of action highlights the edges along horizontal direction only. In such case the horizontal gradients such as

and have higher magnitudes compare the vertical gradients and . Similarly, for another hand waving action (upwards) in the KTH dataset, the movements are along vertical direction. In such case the vertical gradients such as and have higher magnitudes compare the vertical gradients and . Furthermore, the boundaries with sharp discontinuities are also enhanced giving a more clarity whether it is belongs to external edge or a part of

boundary.

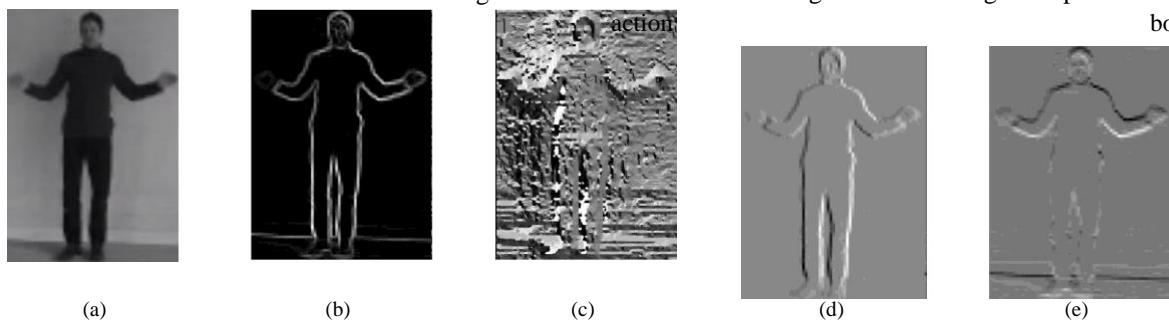


Figure.2 (a) Original Hand Wave Action image (b) Gradient Magnitude, (c) Gradient Direction, (d) Directional Gradient , and (e) Directional Gradient

Once the Gradients are measured, the final LHOG is obtained by the computation of Histograms. Generally, the histogram represents an image by discovering the occurrences of certain micro-patterns without local information. Hence to aggregate the local information to the action descriptor, we divide the action image into several blocks, $\{1, 2, \dots\}$ and measures histogram from

each block. Here, each grey level is considered as a bin and the occurrences are aggregated to create a histogram, as

$$() = \sum \quad (,) = \quad (5)$$

Where (\cdot, \cdot) denotes the pixel position in the block, \cdot is a grey level and (\cdot, \cdot) is a binary grey level of a pixel located at position (\cdot, \cdot) and G is an accumulation value. Next, the final LHOG is calculated by concatenating the Histograms of all blocks as;

$$\text{image is divided. } \otimes \text{ denotes the concatenation operation.} \quad (6)$$

Where N is the total number of blocks into which the action

Here the concatenation is accomplished in spatial fashion, and the obtained final LHOG plays an important role of in the representation of action image through its movement's directions.

B. Dimensionality Reduction

The dimensionality reduction is applied over the to reduce its dimensions. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are the two most popular dimensionality reduction techniques. PCA is an unsupervised and LDA is supervised methods. In these two methods, LDA have better performance compared to PCA because the Principal components obtained through PCA have high variance which won't give effective results in the recognition of actions, especially when the actions have similar trajectories like running and jogging.

In the case of supervised algorithms, LDA has perceived an excellent performance in the action recognition. In LDA, the optimal subspace is obtained by the optimization of fisher-rao's criterion which is defined as the ratio of within class scatter matrix to the between class scatter matrix. Mathematically the optimal subspace is defined as;

$$(7)$$

Where S_w is a within class scatter matrix and S_b is a between class scatter matrix, \mathbf{C} that are both are symmetric and positive definite matrices. The mathematical expressions for \mathbf{C} and \mathbf{B} are defined as;

$$(8)$$

And

$$(9)$$

Where n denotes the total number of samples in class i , \mathbf{X}_i is the mean of data in class i , \mathbf{m} is the mean of total class data and \mathbf{x}_j is the j th sample of class i .

LDA tries to maximize the separation between different classes and minimize the separation within the class simultaneously. However, LDA captures the linear spaced features only but not focused on the non-linear

techniques [V]. In KDA, the input data is mapped to the low dimensional feature space by non-linear mapping. In KDA, the within and between class scatter matrices are defined as;

$$= \mathbf{X}_i \quad \Sigma_i = (\mathbf{X}_i - \mathbf{C})(\mathbf{X}_i - \mathbf{C})^T \quad (10)$$

And

$$= \mathbf{X}_i \quad \Sigma_b = (\mathbf{X}_i - \mathbf{C})(\mathbf{X}_i - \mathbf{C})^T \quad (11)$$

Where n_i is the number of samples in the action class i ,

\mathbf{C}_i is the centroid of class i , \mathbf{C} is the global centroid, C is the number of classes, \mathbf{x}_i is a vector of specific class, and \mathbf{x}_i is the set of samples in the class i . In Eq.(10) Σ_i determines the scattering degree within the class of actions and is measured as the summation of covariance matrices of each class. Next, In the Eq.(11), Σ_b determines the scattering degree between the class of actions and is measured as the summation of covariance matrices of means of each class. Finally the optimal subspace is obtained as;

$$(12)$$

The major difference between LDA and KDA is the computation process of scattering matrices. In the LDA, the scattering matrix is measured through the computation of mean deviations. For within class, the class samples are discriminated by measuring the deviation of sample with the mean and in between class, the discrimination is computed by measuring the deviation of mean from overall mean of data. Unlike LDA, in KDA, the discrimination is computed based on centroids. For within class discrimination, initially one centroid is chosen and the samples in that class are discriminated by measuring the deviation of samples with centroid of that particular class. Next, for between classes, the discrimination is measured by computing the deviation of a centroid of particular class with overall centroid. This evaluation has one main advantage, i.e., it can measure the samples which are non-linearly related and this is the most realistic scenario in real time applications, because, all the action are not linearly related.

IV. SIMULATION RESULTS

To evaluate the performance of developed HAR system, we used a standard benchmark datasets, called as INRIA Xmas Motion Acquisition Sequences (IXMAS) dataset [20]. The simulation is accomplished through MATLAB software. Initially, we have discussed the details of datasets and then the results obtained after the deployment of proposed approach over them is discussed. Further a detailed comparative analysis is stipulated between proposed and conventional approaches.

A. Dataset Details

spaced features. Kernel Dimensionality Analysis (KDA) is a non-linear extraction of LDA which was used in this paper to obtain non-linear discriminant features through kernel

independent action recognition algorithms, including both cross-view and multi-view action recognition. This dataset consists of 12 action classes such as; *check watch* (C1), *cross arms* (C2), *scratch head* (C3), *sit down* (C4), *get up* (C5), *turn around* (C6), *walk* (C7), *wave* (C8), *punch* (C9), *kick* (C10), *point* (C11) and *pick up* (C12). Each action is performed three times and 12 different subjects are recorded with five cameras, four are fixed at four sides and one is fixe

IXMAS is a challenging dataset, acquired with multiple actors under multiple camera views. This dataset is more popular among the HAR methods for testing view

on the top. These five cameras captures five views such as left, right front back and top. The frame rate is 23 frames per second and the size of frame is 390×291 pixels. Figure 3 shows some samples of different actions under multiple views. Each row represents different action and each column represents different views.

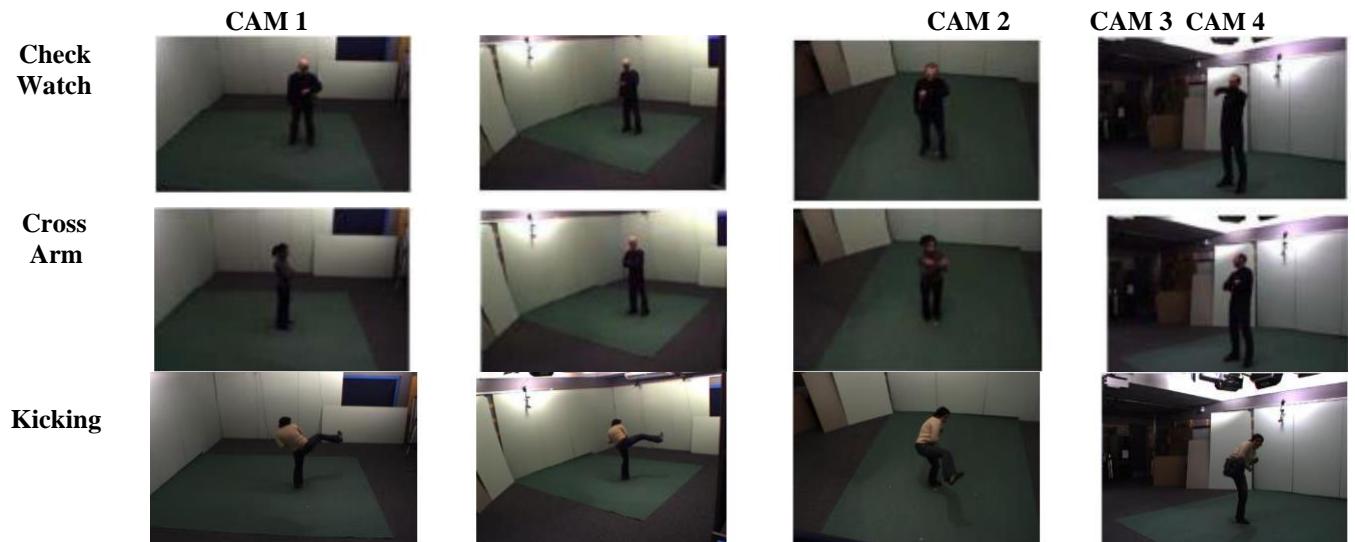


Figure.3 Some action samples of IXMAS dataset under multiple views

B. Results

The simulation is employed for four times, each time for one view. At each simulation, we have considered only one view for both training and testing. Each action is performed for three times; hence we have used the actions performed at two time instances for training and the actions performed at one time instance are used for testing. These combinations are changed and we have conducted the simulation for three times. For example, at the first phase simulation, the action performed at first and second time instance is used for training and the actions performed at third instance are used for testing. In the second phase simulation, the actions performed at first and third time instance is used for training and the actions performed at second instance are used for testing. In the last phase, the actions performed at second and third time instance is used for training and the actions performed at first instance are used for testing. At every simulation, for each action we have trained 200 frames/images and 100 frames/images are tested. Based on the recognized actions at every situation phase, the performance is measured through several

performance metrics such as Detection Rate or Recall, Precision, False Negative Rate (FNR), False Discovery Rate (FDR) and F-score. The average performance metrics obtained for View 1 are shown in Table 1.

As it can be seen from Table 1, for every action, the input instances considered for testing are 100. Out of 100, the total number of correctly recognized actions is highlighted with bold. For example, consider an action *check watch*, the total number of input instances those are correctly classified as *check watch* are 93 and among the remaining 8, 5 are recognized as *cross arms*, 1 is recognized as *scratch head* and 1 is recognized as *Punch*. Similarly, an action *Cross Arms*, the total number of input instances those are correctly classified as *Cross Arms* are 94 and among the remaining 6, 4 are recognized as *Check Watch*, and 2 are recognized as *Wave*. In this manner, the entire actions are classified and based on the obtained classification results; the performance is measured through several performance metrics. The evaluated performance metrics are shown in Table 2.

Table.1 Confusion matrix of actions of IXMAS under View 1

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	Total
C1	93	5	1	0	0	0	0	0	1	0	0	0	100
C2	4	94	0	0	0	0	0	2	0	0	0	0	100
C3	0	1	93	0	0	0	0	0	0	3	3	0	100
C4	0	0	0	90	4	0	0	0	0	0	0	6	100
C5	0	0	0	4	91	0	0	0	1	0	0	4	100
C6	0	0	0	3	3	89	3	0	2	0	0	0	100
C7	0	0	0	1	1	4	92	0	0	2	0	0	100
C8	0	3	2	0	0	0	0	93	0	0	2	0	100
C9	1	1	3	0	0	0	0	3	87	0	5	0	100
C10	1	4	2	0	0	0	3	1	0	89	0	0	100
C11	0	0	3	0	0	0	3	0	5	0	89	0	100
C12	1	0	0	5	4	0	0	0	0	0	0	90	100
Total	100	108	104	103	103	93	101	99	96	94	99	100	1200

Table.2 Average performance metrics for different actions of IXMAS dataset under View 1

Action/Metric	Recall (%)	Precision (%)	F-Score (%)	FNR (%)	FDR (%)
Check Watch	93.4574	93.5532	93.5053	6.5423	6.4468
Cross arms	93.6145	93.8454	93.6510	6.3855	6.1546
Scratch head	92.8741	92.9699	92.9220	7.1259	7.0301
Sit down	90.4785	90.5743	90.5264	9.5215	9.4257
Get up	91.0025	91.0983	91.0504	8.9975	8.9017
Turn around	89.3658	89.4616	89.4137	10.6342	10.538
Walk	92.4314	92.5272	92.4793	7.5686	7.4728
Wave	93.4647	93.5605	93.5126	6.5353	6.4395
Punch	87.4571	87.7785	87.6175	12.5429	12.221
Kick	88.7496	88.8954	88.8224	11.2504	11.104
Point	89.4963	90.1124	89.8033	10.5037	9.8876
Pick up	90.1247	90.2247	90.1747	9.8753	9.7753

Table.2 depicts the details of performance metrics evaluation under different actions. Here, all types of actions of the IXMAS dataset are processed for simulation. For an every action, the developed system displays a label to which it belongs. Based on the label, the correctly classified results are measured and they are called True Positives and the incorrectly classified results are called True Negatives. For example, if the action sequence of check watch is processed for testing and the system had displayed a label of Scratch Head, then it is counted under True negative. In this manner, for every action, the total number of positively and negatively classified results is measured. Based on those values, the performance metrics are measured. Similarly, the further metrics are also measured for every action. From the Table.2, we can notice that the maximum TPR (93.6145%) is achieved for Cross arms, while minimum TPR (87.4571%) is achieved for Punch action. Next, the maximum PPV (93.8454%) is achieved for is achieved for Cross arms, while minimum PPV (87.7785%) is achieved for Punch action. Next, the maximum F-Score (93.6510%) is achieved for is achieved for Cross arms, while minimum F-Score (87.6175%) is achieved for Punch action. Finally the maximum FNR (12.5429%) is achieved for Punch action while minimum FNR (6.3855%) is achieved for Cross arms action.

Figure.4 shows the comparison between proposed and several existing methods through accuracy at different views. From this figure, we can see that the minimum accuracy is obtained at View 5 and maximum accuracy is obtained at View 2. The major reason behind the less accuracy at view 5 is that the actions are captured with a camera fixed at the top position. In this position, some movements of the actions are hindered thereby the descriptor cannot represent that action effectively. Unlike, the actions captured through CAM 2 are in frontal view; hence the entire action movements are clearly visible thereby the proposed descriptor can represent the action perfectly. In the proposed LHOGL, we have employed Laplacian gradient which is a second order derivative and extracts the almost all edge regions.

The method in [15] considered the MHI as a feature descriptor and PCA for dimensionality reduction. However, the MHI descriptor reveals only motion features but not differentiates the necessary and unnecessary motions. In the action image, there exist backgrounds motions if any objects are there in the background and they are also considered as required features when MHI is employed as a feature descriptor. Hence for cluttered backgrounds, the MHI has limited performance. Next, the DMH [21] adopted histogram based descriptor based on the mean distance between segmented blocks in action images.

However, DMH never represents the edge regions at which the motion features are present. Mainly the motion features are present at the edge regions (hand, and legs) and the remaining body has smooth regions. For some actions which hinder the hand and leg movements, the DMH won't perform effectively. Further, the method in [16] employed only LDA for action recognition which has the problem of non-linearity constraints. For a non-linear data, the LDA won't perform effectively there by lessen the recognition accuracy.

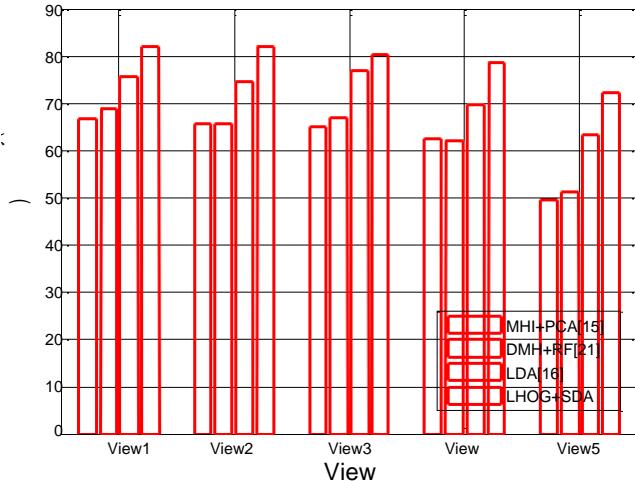


Figure.4 Accuracy comparison under different views

Since the proposed method adopted second order gradient operators, every action can be represented much effectively with their motion region. Further the SDA allows providing a sufficient discrimination between different actions with different views. Hence the proposed approach has higher accuracy under all views, as it was observed from figure.4. The accuracy of proposed LHOG+SVM at view 2 is observed as 82.06% while for existing methods, it is of 74.62%, 65.77% and 65.54% for LDA [16], DMH+RF [21], and MHI+PCA [15] respectively. Further accuracy of proposed LHOG+SVM at view 5 is observed as 72.29% while for existing methods, it is of 63.32%, 51.10% and 49.63% for LDA [16], DMH+RF [21], and MHI+PCA [15] respectively.

V. CONCLUSION

In this paper, we have developed a new HAR system to recognize the human actions from videos. The proposed method is focused on the edge based action representation through which action movements are described. The proposed action descriptor is based on the Laplacian gradient which has an efficient edge detection capability. Further the SDA is also successful in reducing the dimensionality of feature set. Simulation experiments conducted over IXMAS action dataset and the obtained results revealed the effectiveness at different views. On an average, the proposed method has gained an accuracy of 79.0360% while the accuracy of existing methods is noticed

as 61.8460%, 62.9100%, and 72.0620% for MHI+PCA, DMH+RF, and LDA respectively.

REFERENCES

- [1] J. K. Aggarwal and M. S. Ryoo. Human activity analysis: A review. *ACM Computing Surveys*, 43(3):16, 2011
- [2] Teddy Ko, "A survey on behavior analysis in video surveillance for homeland security applications", In: *Proc. 37th IEEE Applied Imagery Pattern Recognition Workshop*, Washington, DC, USA, pp. 1-8, 2008.
- [3] M. S. Ryoo, "Human activity prediction: Early recognition of ongoing activities from streaming videos", In: *Proc. of international Conf. on Computer Vision*, Barcelona, Spain, pp.1-5, 2011.
- [4] K. Soomro and A. R. Zamir, "Action recognition in realistic sports videos", *Advances in Computer Vision and Pattern Recognition*, Vol. 71. Cham, Switzerland: Springer, 2014, pp. 181-208.
- [5] T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications," in *Proc. 37th IEEE Appl. Imagery Pattern Recog. Workshop*, Washington, DC, 2008, pp. 1-8.
- [6] D. Weinland and E. Boyer, "Action recognition using exemplar-based embedding," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2008, pp. 1-7.
- [7] Ronald Poppe, "A survey on vision-based human action recognition", *Image and Vision Computing* 28 (2010) 976-990.
- [8] S. Sandhya Rani, G. Appa Rao Naidu, V. Usha Shree, "A Fine Grained research Over Human Action Recognition", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Volume-9 Issue-1, November 2019.
- [9] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local SVM approach," in *Proc. Int. Conf. Pattern Recognit.*, vol. 3, 2004, pp. 32-36.
- [10] Duta, I. C., Uijlings, J. R., Ionescu B, et al. Efficient Human Action recognition using Histograms of motion gradients and VLAD with descriptor shape information. *Multimed Tools Appl.* 76, 22445-22475, 2017.
- [11] Jin Wang et al. "Human action recognition based on Pyramid Histogram of Oriented Gradients", *IEEE International Conference on Systems, Man, and Cybernetics*, AK, USA, 2011.
- [12] Bo Lin and Bin Fang, "A new spatial-temporal histograms of gradients descriptor and HOG-VLAD encoding for human action recognition", *International Journal of Wavelets, Multi-resolution and Information Processing*, Vol. 17, No. 02, 2019.
- [13] V. Thanikachalam and K.K. Thyagarajan, "Human Action Recognition using Accumulated motion and gradient of motion from video", ICCCNT 2012.
- [14] Md. Zia Uddin, J.J. Lee, and T.-S. Kim, "Shape-Based Human Activity Recognition Using Independent Component Analysis and Hidden Markov Model", In: Nguyen N. T., Borzemski L., Grzech A., Ali M. (eds) *New frontiers in applied artificial intelligence. IEA/AIE 2008. Lecture notes I computer science*, vol. 5027, Springer, Berlin, Heidelberg.
- [15] M. A. Naeel, Abdelwahab, M. M., and El-Saban, M., "Multi-view human action recognition system employing 2DPCA", in *IEEE Workshop on Applications of Computer Vision (WACV)*, 2011, pp. 270-275.
- [16] Su, Y., Li, Y. & Liu, A., "open-view human action recognition based on Linear Discriminant Analysis", *Multimedia tools Appl.* 78, 767-782, 2019.
- [17] M. Guo and Z. Wang, "A feature extraction method for human action recognition using body-worn inertial sensors", *IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, itlay, 2015.
- [18] B. Mandal, How-Lung Eng, "Regularized Discriminant Analysis for Holistic Human Activity Recognition", *IEEE intelligent systems*, 2012.
- [19] R. Gonzalez and R. Woods, *Digital image processing*. Pearson/Prentice Hall, 2008.
- [20] D. Weinland, E. Boyer, and R. Ronfard, "Action recognition from arbitrary views using 3D exemplars," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1-7.
- [21] Vikas Tripathi, Durga prasad Gangodkar, Ankush Mittal, Vishnu Kanth, "Robust Action Recognition framework using Segmented Block and Distance Mean Histogram of Gradients Approach", *Procedia Computer Science*, Volume 115, 2017, Pages 493-500