Activity Recognition in Egocentric video using SVM, kNN and Combined SVMkNN Classifiers

K.P.Sanal Kumar [1], Dr.R.Bhavani [2]

[1]Research Scholar, Dept. of CSE, Annamalai University

sanalprabha@yahoo.co.in

[2]Professor, Dept. of CSE, Annamalai University.

Abstract: Egocentric vision is a unique perspective in computer vision which is human centric. The recognition of egocentric actions is a challenging task which helps in assisting elderly people, disabled patients and so on. In this work, life logging activity videos are taken as input. There are 2 categories, first one is the top level and second one is second level. Here, the recognition is done using the features like Histogram of Oriented Gradients (HOG), Motion Boundary Histogram (MBH) and Trajectory. The features are fused together and it acts as a single feature. The extracted features are reduced using Principal Component Analysis (PCA). The features that are reduced are provided as input to the classifiers like Support Vector Machine (SVM), k nearest neighbor (kNN) and combined Support Vector Machine (SVM) and k Nearest Neighbor (kNN) (combined SVMkNN). These classifiers are evaluated and the combined SVMkNN provided better results than other classifiers in the literature.

Keywords: Egocentric, Histogram of Oriented Gradients, Life Logging Activity, Motion Boundary Histogram, Trajectory.

1. Introduction

The conventional method of activity recognition includes examining the behavior of a person from one or more cameras [1]. A significant improvement over the recent years in identifying actions has been identified using first person camera. However, major challenges in this field need subtle movements/gestures. This is due to occlusions and distractions from actual action image regions. Another method is implemented by using egocentric cameras with which the actions are obtained and analyzed from the user's perspective.

Egocentric video analysis for user activities has attracted attention in the recent development technology. These videos offer a helping hand for the disabled or elderly people. Lifelogging is a method to record some portions of the life of a person. Here, the videos taken are for activities like drinking, eating, house work etc. The recording is done automatically using wearable devices. Many applications of Lifelogging activity are behavior analysis, lifestyle analysis, health monitoring and so on.

In this research work, HOG (Histogram of oriented gradients), Motion Boundary Histogram (MBH) and Trajectory are extracted. Here, HOG gives the static information; MBH gives the motion information and Trajectory capture the local motion information of the video.

For experiments, lifelogging egocentric activity dataset is used which contains 5 top level categories like food, house work, office work, motion and social interaction. Food contains 2 second level categories like eat and drink. House work contains 1 second level categories like house work. Office work contains 4 second level categories like read, write, use internet and watch videos. Motion contains 4 second level categories like walk straight, walk back and forth, walk up and down and running. Social Interaction contains 2 second level categories like talk on the phone and talk to people.

2. Related Works

[2] outlines the advancement in FPV video analysis highlighting the most commonly used features, methods, challenges and opportunities within the field. [3] provides the adaptability to acquire the interactions without considering the number of individuals included and their level of acquaintance in context with a variable level of social contribution. [4] proposed a novel method for solving the body part and motion identification problem. [5] designed a new algorithm, Ensemble Actions EMT (EA-EMT), which utilized the initial environment model as a library of state transition functions and implements a variety of prediction to assemble and adjust a modified model. [6] presents an approach that invents the inherent relevant information from structured hand labeling for hand recognition and hand part labeling. In [7], office environment is taken egocentric activity recognition. Here, motion descriptors are extracted which are combined with user eye movements. In [8], First Person Videos are temporally segmented into twelve hierarchical classes. Here, the problem of FPV ADL analysis is implemented for multi-task learning framework. [9] implemented the HOG feature and support vector machine (SVM) was applied to train an action classifier using HOG features. [10] Implemented the Probabilistic Neural network (PNN) classifier for classifying the actions of supplied and PCA for dimensionality reduction. [11] utilized a Hessian detector, HOG and Histograms of Optical Flow (HOF) descriptors along with a Bag-of-Features (BoF) representation to perform activity recognition on simple and realistic databases. [12] The Motion Boundary Histogram (MBH) is the most recent appealing method to suppress the constant movement by considering the flow gradient. It is powerful to some extent to the presence of camera motion, yet it does not explicitly handle the camera movement. [13] proposed the motion boundary histograms (MBH) descriptor for human detection by calculating derivatives independently for the horizontal and vertical components of the optical flow. The descriptor encodes the relative movement between pixels. [14] recognizing activities of individuals by SVM multi-class classifier whose structure is determined by a clustering procedure.

3. Proposed Work

Egocentric video is used to track the daily activities of people. In this system there are two phases. First one is Training Phase and the next one is Testing Phase. Figure 1 depicts the flow diagram for the proposed model.

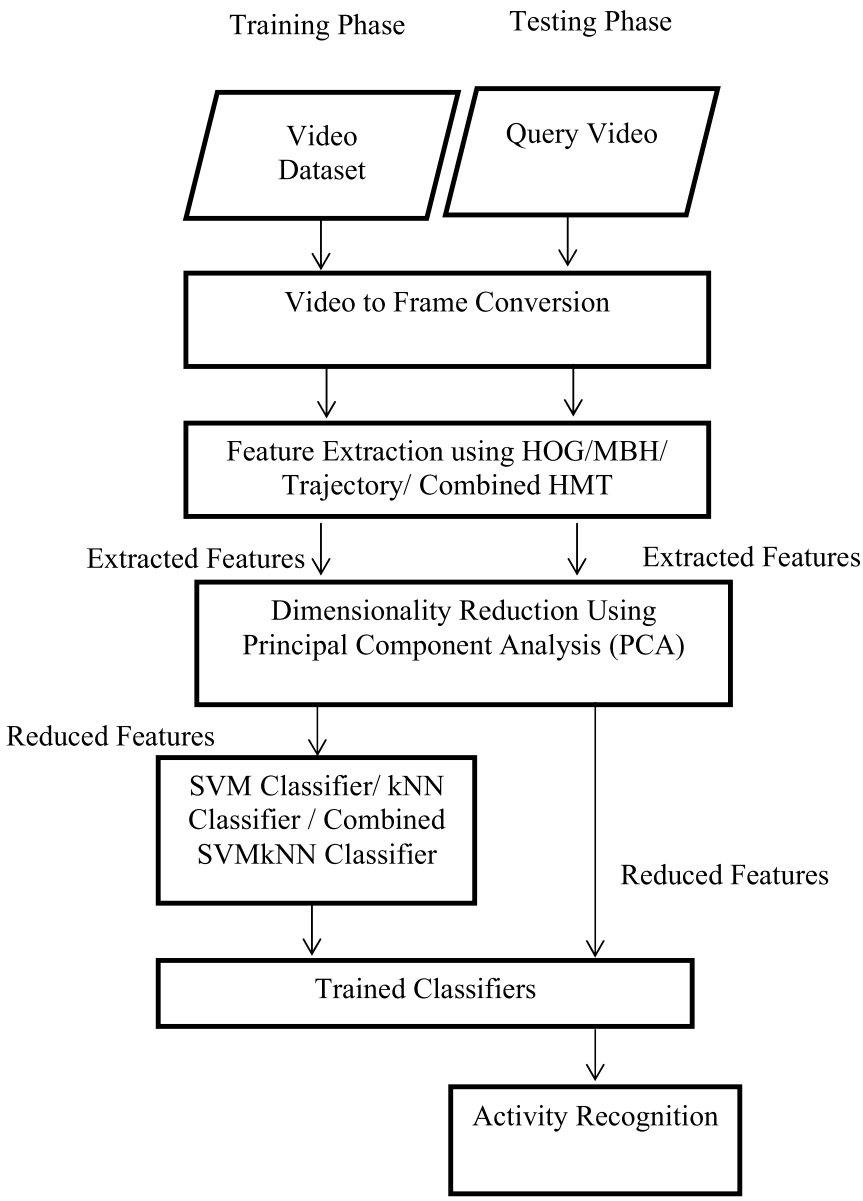

Fig. 1: Block diagram for the proposed model

The system implemented here has the following steps in training phase.

1. The input video is fed where the feature extraction is applied.
2. The features like HOG (Histogram of oriented gradients), Motion Boundary Histogram (MBH) and Trajectory are extracted. The extracted features are fused together to form combined HMT (HOG, MBH, Trajectory).
3. The extracted features are provided as input to Principal component Analysis (PCA) which reduces the feature dimensionality.

4. The reduced features are fed to the classifiers like Support Vector Machine (SVM), k Nearest Neighbor (kNN) and combined Support Vector Machine and k Nearest Neighbor (combined SVMkNN).
5. The activity recognition is obtained as output.

The above steps are repeated with the testing phase where the input is a query video.

4. Feature Extraction

4.1 Histogram of Oriented Gradients (HOG)

This feature is local shape information portrayed by the distribution of gradients or edge directions. This generally focuses on static appearance information [15].

The input frame is chosen where the normalization is applied on both color values and gamma correction is applied. The next step is to apply gradient filter for finding the gradient values. The window is divided into adjacent, non-overlapping cells of size C×C pixels (C = 8). In each cell, a histogram of the gradient orientations is calculated which is binned into B bins.
Calculate the weights by using bilinear interpolation. These are subjected to normalization by concatenating overlapping blocks. The normalized block features are combined into a single feature vector which is the HOG. Initially the videos are divided into frames and then the feature extraction technique is applied. Here, the size of the HOG feature is 8102.

4.2 Motion Boundary Histogram (MBH)

The optical flow's horizontal and vertical components are defined separately using two scalar maps, of the motion components [13]. Histograms of oriented gradients are calculated for each of the two optical flow component images, similar to still images. Due to low differences, the data about differences in motion boundaries is retained whereas the constant motion information is rejected. This leads to the cancelation of most of the impacts of camera motion. Spatial derivatives are calculated for each and orientation is quantized into histograms. The magnitude used here is for weighting. The output of this process is a pair of horizontal (MBHx) and vertical (MBHy) descriptors. Here, the size of the MBH feature is 2050.

4.3 Trajectory

Here, feature points are followed on each spatial scale independently. For each frame I_t its dense optical flow field $\omega_t = (u_t, v_t)$ is calculated with respect to the next frame I_{t+1} where u_t and v_t are the horizontal and vertical components of the optical flow. Let a point $P_t = (x_t, y_t)$ in frame I_t, its position in frame I_{t+1} is smoothed by implementing a median filter on ω_t

$$P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (M * \omega_t)\big|_{(x_t, y_t)} \tag{1}$$

where M is the median filtering kernel [16]. The size of the median filter kernel M is 3 × 3 pixels. As the median filter is more powerful to outliers, it enhances trajectories for points at motion boundaries. To extract dense optical flow fields, the algorithm includes a translation motion model between neighborhoods of two consecutive frames [17]. Polynomial expansion is incorporated to approximate intensities of pixel in the neighborhood. Points of subsequent frames are fused to form trajectories: $(P, P_{t+1}, P_{t+2} \cdots)$.

For each frame, if no tracked point is found in its neighborhood, a new point is sampled and added to the tracking process so that a dense coverage of trajectories is identified. The shape of a trajectory encodes local motion patterns. Let L be a trajectory of length, its shape can be given by a sequence $(\Delta P, \ldots \Delta P_{t+L-1} \ldots)$of displacement vectors$(\Delta P_t = ((P_{t+1} - P_t) = (x_{t+1} - x_t, y_{t+1} - y_t)$ The vector obtained as output, is normalized by the sum of displacement vector magnitudes:

$$T = \frac{(\Delta P_t, \ldots \Delta P_{t+L-1})}{\sum_{j=t}^{t+L-1} \|(\Delta P_j)\|} \tag{2}$$

Here, the size of the trajectory feature is 902.

4.4 *Combination of features (Combined HMT)*

In this article, the features are concatenated into a single feature vector i.e., the HOG, MBH and trajectory are fused together. The HOG feature vector contains 8102 features, MBH has 2050 features and trajectory feature has 902 features. All these features are combined together into 11054 features.

5. Dimensionality Reduction using Principal Component Analysis (Pca)

The obtained feature vectors are of large size. This large dimension results in a very slow computation. As a result, the dimensions are reduced using Principle Component Analysis.

In statistics, Principal Components Analysis (PCA) [18] is a technique which is implemented to decrease the dimensionality of a high dimensional data set. It is a linear transformation that chooses a new coordinate system for the data set. The new coordinate system is a representation of the directions where the variance of the data is high. The PCA can be utilised for reducing dimensionality in a data set while retaining the attributes of the data set that contribute most to its variance, by retaining lower-order principal components and ignoring higher-order ones.

Applying Principal Component Analysis

Algorithm for PCA

Step 1: Generate the column vectors from the input feature vectors.

Step 2. Calculate the covariance matrix of the two column vectors that is formed in step 1.

Step 3. The variance of each column vector is obtained from the diagonal elements of the 2×2 covariance vector with itself, respectively.

Step 4. Compute the Eigen values and the Eigenvectors of the covariance matrix [19].

Initially the covariance matrix is calculated and then the eigen vectors are derived. After the eigenvectors are found from the covariance matrix, rank them by their eigenvalue, in the descending order that results in arranging the components in the order of significance. To decrease the dimensionality of the data-set, ignore the components of lower significance [20]. Here, the feature size is reduced to 800.

6. Classifiers

6.1 Support Vector Machine

In machine learning, support vector machines are models that are associated with learning algorithms for analyzing and classifying data. An SVM algorithm creates a model that allocates examples to one category or another. This model maps the examples which divide into the separate categories [21]. New examples are trained to assign into categories already divided and they are predicted. In addition to performing linear classification, SVMs can also perform a non-linear classification. Here, multiclass SVM is used.

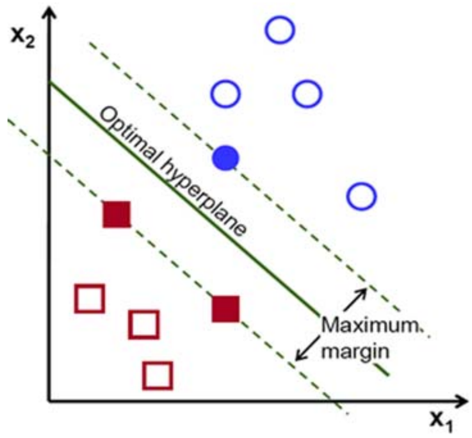

Fig. 2: Support Vector Machine

Figure 2 shows the separation of examples using optimal hyperplane which is cited from [22]. The operation of SVM is based on the hyperplane that gives largest minimum distance to the training examples. The input feature taken is 800 after implementing dimensionality reduction method (PCA). The output is activity recognition for 5 top level categories and 13 second level categories.

6.2 kNN Classifier

The kNN is k-Nearest Neighbor which is supervised algorithm. Here, the result is classified based on majority of k-Nearest Neighbor category [23]. This algorithm classifies a new entity based on characteristics and training samples. This algorithm uses neighborhood for classification as the prediction value of the new example. Figure 3 depicts the kNN classifier which separates the objects into different classes. This is cited from [24].

Steps in kNN classifier algorithm:

1. Assign a value for k
2. Compute the distance between the test object and every object in the set of training objects.
3. Select the closest training object with respect to the test object.
4. Select the class with maximum number of matched objects.
5. Repeat until a same class is obtained.

In our research work, Euclidean distance function is used

$$d_E(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}, d_A(x,y)=\sum_{i=1}^{n}|x_i-y_i|$$

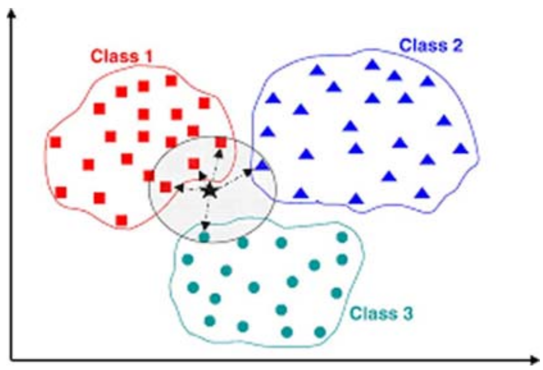

Fig. 3: kNN classifier for a vector where k value is 5

In this research work, the input feature taken is 800 after implementing dimensionality reduction method (PCA). The output is activity recognition for 5 top level categories and 13 second level categories.

6.3 COMBINED SVM kNN CLASSIFIER (COMBINED SVM kNN)

Support vector machine (SVM) is a supervised learning method for classification. In which, a hyperplane is created through which data is separated from each other. k nearest neighbor algorithm is basically a machine learning algorithm. Support vector machine (SVM) basically utilizes support vectors to create a hyperplane. Hyperplane is used to separate normal and abnormal data. kNN algorithm is used to find new data added to training data set. Support vector machine (SVM) and k nearest neighbor (kNN) algorithms are combined together to evaluate false positive rate is known as Combined Support Vector k Nearest Neighbor (Combined SVMkNN) algorithm. Here, in SVM classifier, the Guassian kernel is defined as

IOP Conf. Series: Materials Science and Engineering **225** (2017) 012226 doi:10.1088/1757-899X/225/1/012226

$$K(x,\acute{x}) = exp\left(\frac{-\|x-\acute{x}\|^2}{2\sigma^2}\right) \tag{3}$$

$||x - \acute{x}||^2$ is squared euclidean distance and σ is a measure of expansion.

Algorithm for combined SVMkNN classifier

Step 1: Select data from different class;

Step 2: Classify the data by using SVM classifier and kNN classifier.
Step 3: Calculate False Positive Rate (FPR) for both the classifiers
Step 4: If FPR (SVM) is greater than FPR (kNN)
then apply kNN for clustering the data.
Step 5: If new data added to data set then update dataset;
Repeat the above steps for all data in the data set.

These two algorithm works together in Combined SVMkNN algorithm in which, support vector machine (SVM) uses training data set to learn something from data set. If any new is added to its dataset. It is updated by k nearest neighbor (kNN) algorithm. In this research work, the input feature taken is 800 after implementing dimensionality reduction method (PCA). The output is activity recognition for 5 top level categories and 13 second level categories.

7. Performance Metrics

Accuracy is defined as the ratio between the summation of true positive rate and true negative rate and the total population.

Accuracy = Tp+Tn / (Tp+Tn+Fp+Fn)
(4)

where,

Tp is the number of data classified correctly as positive class

Tn is the number of data classified correctly as negative class

Fp is the number of data classified wrongly as positive class

Fn is the number of data classified wrongly as negative class

8. Experimental Results

In our work, the lifelogging activity videos are taken which are classified into two levels. They are 5 top level categories and 13 second level categories. The top level categories are motion, social interaction, office work, food and house work. The second level categories are walk straight, walk back and forth, walk up and down, running, talk on the phone, talk to

IOP Conf. Series: Materials Science and Engineering **225** (2017) 012226 doi:10.1088/1757-899X/225/1/012226

people, watch videos, use internet, write, read, eat, drink and housework. Table 1 shows the accuracy of SVM, kNN and Combined SVMkNN classifiers for top level.

Table 1: Accuracy for top level categories

Features / Classifiers	Histogram of Oriented Gradients	Motion Boundary Histogram	Trajectory	Combined
SVM	79.37	83.45	85.68	86.31
kNN	72.19	76.26	78.48	80.11
Combined SVMkNN	82.34	85.79	86.73	88.47

In the above table, Combined SVMkNN provides 82.34% for HOG, 85.79% for MBH, 86.73% for trajectory and 88.47% for combined of accuracy for top level categories which are higher than other classifiers. Therefore Combined SVMkNN classifier performs better than other classifiers. Table 2 depicts the accuracy of SVM, kNN and Combined SVMkNN classifiers for second level.

Table 2: Accuracy for second level categories

Features / Classifiers	Histogram of Oriented Gradients	Motion Boundary Histogram	Trajectory	Combined
SVM	70.85	79.46	74.92	77.89
kNN	68.21	76.17	70.35	73.43
Combined SVMkNN	72.48	80.53	76.17	80.76

In the above table, Combined SVMkNN provides 72.48% for HOG, 80.53% for MBH, 76.17% for trajectory and 80.76% for combined of accuracy for second level categories which are higher than other classifiers. Therefore Combined SVMkNN classifier performs better than other classifiers.

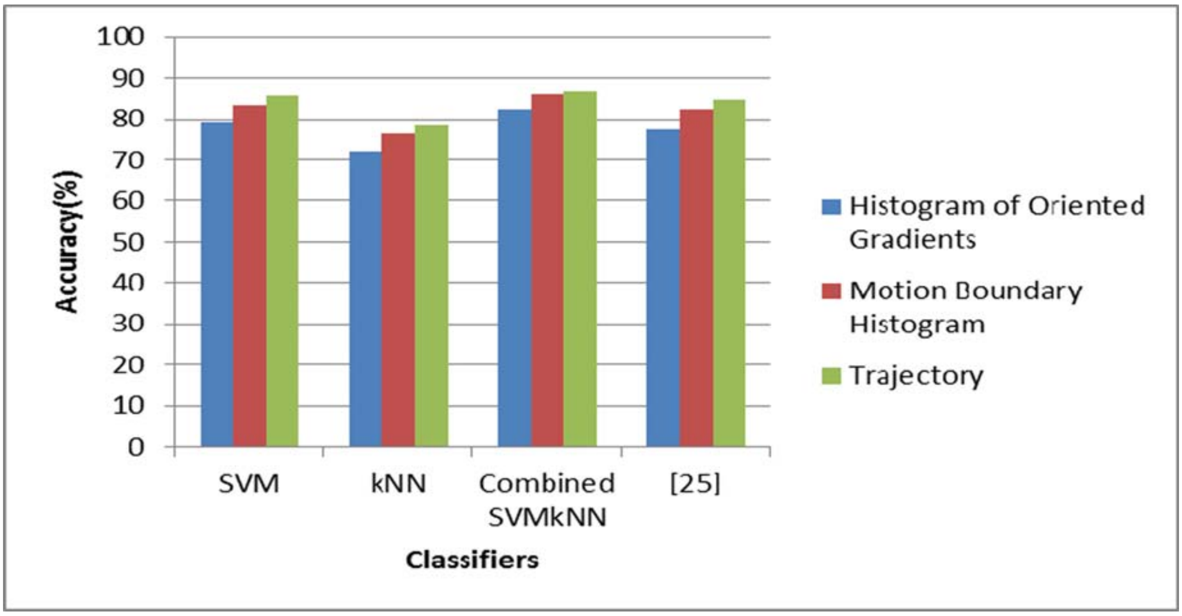

Fig. 4: chart for comparison of classifiers with literature for top level category

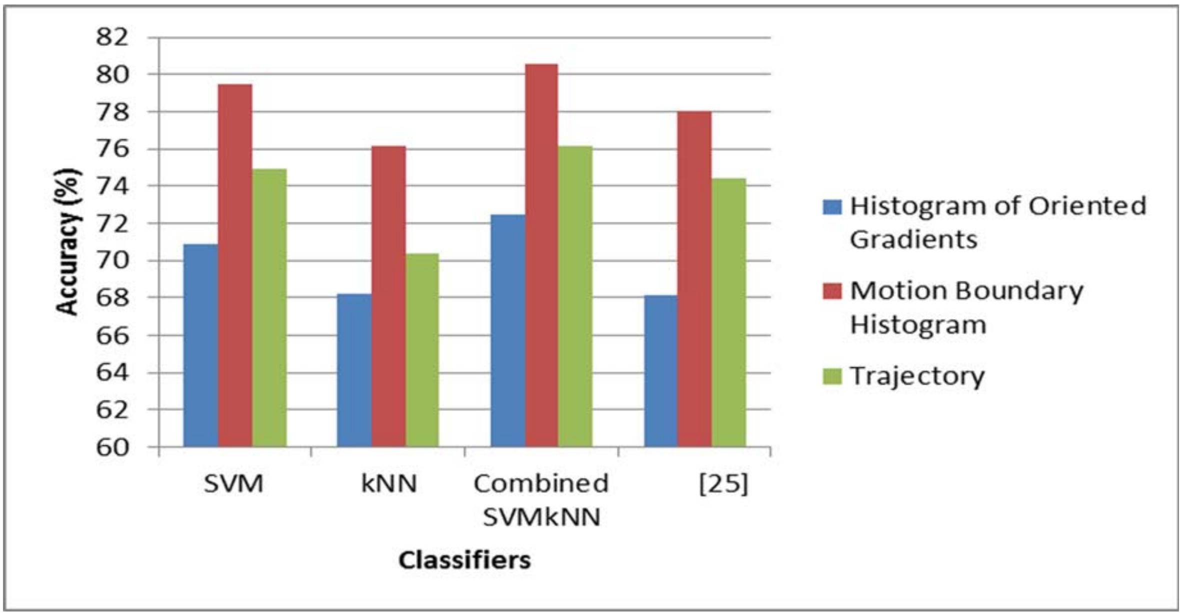

Fig. 5: chart for comparison of classifiers with literature for second level category

From figures 4 and 5, Combined SVMkNN performed the best for both categories than literature methods. Therefore, Combined SVMkNN is better than other classifiers.

9. Conclusion

In this research work, Combined SVMkNN classifier provided better results. SVM provides advantages like it is accurate and robust even when the training sample has some bias. It delivers unique solution since the optimality problem is convex. kNN classifier is a simple classifier. These advantages are combined together to form Combined SVMkNN classifier which provides better results than other classifiers in literature.

References

[1] Poppe, R., “ A survey on vision-based human action recognition”, Image and Vision Computing, Elsevier, 2010.
[2] Alejandro Betancourt, Pietro Morerio, Carlo S. Regazzoni, and Matthias Rauterberg, “The Evolution of First Person Vision Methods:A Survey”, IEEE Transactions On Circuits And Systems For Video Technology, vol. 25, no. 5, May 2015.
[3] Stefano Alletto, Giuseppe Serra, Simone Calderara, Rita Cucchiara, “Understanding social relationships in egocentric vision”, Pattern Recognition, Elsevier, vol.48, pp.4082–4096, 2015.
[4] Sihao Ding, Qiang Zhai , Ying Li , Junda Zhu , YuanF.Zheng , Dong Xuan, “Simultaneous body part and motion identification for human-following robots”, Pattern Recognition, Elsevier, vol.50, pp.118–130, 2016.
[5] Zinovi Rabinovich, Nicholas R. Jennings Robotics and Autonomous Systems, “A hybrid controller based on the egocentric perceptual principle”, Robotics and Autonomous Systems, Elsevier, vol.58, no.9, pp.1039-1048.
[6] XiaolongZhu , XuhuiJia,Kwan-YeeK.Wong, “Structured forests for pixel-level hand detection and hand part labeling”, Computer Vision and Image Understanding, Elsevier, vol.141, pp. 95–107,2016.
[7] K. Ogaki, K. M. Kitani, Y. Sugano, and Y. Sato, “Coupling eye-motion and ego-motion features for first-person activity recognition,” in CVPR Workshop on Egocentric Vision, 2012.
[8] Kishor B. Bhangale and R. U. Shekokar, “Human Body Detection in static Images Using HOG & Piecewise Linear SVM”, International Journal of Innovative Research & Development, vol.3 no.6, June 2014.
[9] C. P. Huang, C. H. Hsieh, K. T. Lai and W. Y. Huang, “Human Action Recognition Using Histogram of Oriented Gradient of Motion History Image”, 2011 First International Conference on Instrumentation, Measurement, Computer, Communication and Control, Beijing, 2011, pp.353-356.
[10] A. Jeyanthi Suresh, P.Asha, “Human Action Recognition in Video using Histogram of Oriented Gradient (HOG) Features and Probabilistic Neural Network (PNN)”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 7, July 2016.
[11] Tenika P. Whytock , Alexander Belyaev , Neil M. Robertson, “Improving Robustness and Precision in GEI + HOG Action Recognition”, in 9th International Symposium, ISVC 2013, Rethymnon, Crete, Greece, July 29-31, 2013. Proceedings, Part I pp 119-128.
[12] H.Wang, A. Klaser, C. Schmid, and C.-L. Liu, “Action recognition by dense trajectories”, In CVPR, Jun. 2011.
[13] N. Dalal, B. Triggs, and C. Schmid, “Human detection using oriented histograms of flow and appearance”, in European Conference on Computer Vision, 2006.

[14] Huimin Qian , Yaobin Mao, Wenbo Xiang, Zhiquan Wang, "Recognition of human activities using SVM multi-class classifier", Pattern Recognition Letters, Elsevier, vol. 31, no.2, pp.100–111, 2010.
[15] Kishor B. Bhangale and R. U. Shekokar, "Human Body Detection in static Images Using HOG & Piecewise Linear SVM", International Journal of Innovative Research & Development, vol.3 no.6, June 2014.
[16] Heng Wang, Alexander Kl¨aser, Cordelia Schmid, Cheng-Lin Liu. Dense trajectories and motion boundary descriptors for action recognition. [Research Report] RR-8050, 2012.
[17] N. Sundaram, T. Brox, and K. Keutzer, "Dense point trajectories by GPU-accelerated large displacement optical flow," in European Conference on Computer Vision, 2010.
[18] Alireza Sarveniazi, "An Actual Survey of Dimensionality Reduction" American Journal of Computational Mathematics, vol.4, pp.55-72, 2014.
[19] S.Anu H Nair, P.Aruna, "Image Fusion Techniques For Iris And Palmprint Biometric System", Journal of Theoretical and Applied Information Technology, Vol. 61 No.3, , pp.476-485, March 2014.
[20] Arun Garg, Richika, "A Novel Approach for Brain Tumor Detection Using Support Vector Machine, K-Means and PCA Algorithm", International Journal of Computer Science and Mobile Computing, vol.4, no.8, pp. 457-474, August- 2015,.
[21] M.Suresha, N.A. Shilpa and B.Soumya, "Apples Grading based on SVM Classifier", in National Conference on Advanced Computing and Communications, April 2012.
[22] OpenCV- Introduction to Support Vector Machines http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html Accessed: 2016-07-22.
[23] K.P.Sanal Kumar, R. Bhavani, "Analysis of SVM and kNN Classifiers For Egocentric Activity Recognition", in Proceedings of the International Conference on Informatics and Analytics, Pondicherry, India — August 25 - 26, 2016, ACM.
[24] Wanpracha Art Chaovalitwongse, Ya-Ju Fan, Rajesh C. Sachdeo, "On the Time Series K-Nearest Neighbor Classification of Abnormal Brain Activity", IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, vol.37, no.6 November 2007.
[25] Sibo Song, Vijay Chandrasekhar, Ngai-Man Cheung, Sanath Narayan, Liyuan Li, Joo-Hwee Lim , "Activity Recognition in Egocentric Life-Logging Videos", vol.9010 Lecture Notes in Computer Science pp 445-458, Springer 2014.

Mr. K.P.Sanal Kumar received the M.C.A., degree from Noorul Islam College of Engineering, Kumarakoil, Tamil Nadu in 2000. He received the M.Tech degree in Computer Science and Engineering from Bharath University, Chennai in the year 2005. He has been with Annamalai University, since 2008. He is pursuing his Ph.D in Computer Science and Engineering at Annamalai

University. He has published 10 papers in international conferences and journals. He is a life member of CSI and ISTE. His research interest includes Digital Image Processing, Pattern Recognition and Artificial Intelligence.

Dr. R. Bhavani received the B. E degree in Computer Science and Engineering from Regional Engineering College, Trichy in 1989. She received the M.E degree in Computer Science and Engineering from Regional Engineering College, Trichy in the year 1992. She worked at Mookambigai College of Engineering, Keeranur from 1990 to 1994.She has been with Annamalai University, since 1994. She obtained her Ph.D degree in Computer Science and Engineering at Annamalai University, in the year 2007. She published 40 papers in international conferences and journals. Her research interest includes Image processing, Image Segmentation, Image Compression, Image Classification, Stenography, Pattern Classification, Medical Imaging, Content Based Image Retrieval and Software metrics. She is an author of a book and has published 3 book chapters as well.