A Survey on Automated Human Action Recognition Using Multi-View Feature

Ashwini S ¹, Varalatchoumy M ²

¹PG-Scholar, Department of Computer Science & Engineering, Cambridge Institute of Technology, Bangalore,
²Associate Professor, Department of Computer Science & Engineering, Cambridge Institute of Technology, Bangalore

Abstract— Recognizing the human action plays a significant role in surveillance cameras. Usually cameras are situated at distant place and convey actions in form of signals at one particular place. This paper presents a framework for recognizing a sequence of actions based on multi-view video data. To depict various actions activities performed in various perspectives, view-invariant feature is being used. The features of multi-view are extracted from various temporal scales, which are demonstrated using global spatial-temporal distribution. The proposed system performs is designed to work on cross tested datasets wherein the system doesn’t require retraining for same scenario that occurs multiple times.

Keywords— Human Action, view-invariant feature, multi-view data, spatial-temporal distribution, cross tested datasets.

I. INTRODUCTION

The measure of video is constantly growing, not just about TV and movie footage but also with the revolution in personal video recording for upload to sites such as YouTube or Google videos [3]. Human Action Recognition has gained high attention towards practical vitality in the wake of increased and growing security concerns. Recognition of actions is one of the most successful applications in pattern recognition and in various applications, such as, human–computer interaction, human activities analysis and real-time surveillance systems. The goal of human action recognition is to find the actions being performed in a video sequence under different complications such as cluttering, occlusion, and change of lighting conditions [5]. This growth becomes the need for automatic video analysis and the recognition of events.

Human-computer interaction is a vital application in real life in which signals play a significant role in human-computer interaction to empower better communication [13]. The majority of the ongoing activity action work tests an activity grouping manually before it can be recognized in a film. However, it is not useful to physically set the start and completion of an activity succession of the film beforehand. The current techniques for action recognition frequently tests an action sequence of arrangement, before it is being perceived in a film. Nonetheless, it is not practical that setting the start and end of an action grouping of the film already. Therefore, a practical action framework needs to isolate numerous activities at a picture arrangement so human actions can be executed as various subjects, such as, size, position, movement and attire, which are as yet a difficult issue for a few reasons, such as, illumination, occlusion, shadow, camera movement or other changes.

The current methodologies don’t require any specific variable for data handling and it explicitly expresses spatio-temporal information at multiple temporal scales. Human action strategies regularly accept the action, caught under confined and implied conditions [21]. In case of multi-view features, moving camera can obscure position in fundamental view varieties [26], the actions can have all earmarks of being not the same from various edges, and along with this feature moving
camera could influence the action by fusing dynamic view changes. Therefore, an action framework ought to be vigorous against environment conditions and view-point changes in action sequence.

In existing system the action is captured from a static viewpoint without any camera movement. However, the patterns of human actions seem to be different from different angles. Some of the approaches train a single classifier for all viewpoints or a set of classifiers where each classifier deals with one viewpoint and the performance depends on the extracted features and the trained classifiers. The main disadvantage of existing system is more sensitive to noise, hence to overcome the challenges of noise in human action recognitions many approaches are being evolved. The proposed approach is able to capture local and global temporal information, and it is capable to capture sequence of motions and occlusion at low computational cost. The features of multiview are extracted from various temporal scales, which are demonstrated using global spatial-temporal distribution.

II. LITERATURE SURVEY

Videos from the real world are fairly complicated, especially when taken in uncontrolled environments with camera motion [14][16]. Usually in non-static camera, the motion field in an action region is often contaminated by background and camera movements, suppressing these features and extracting valid foreground information is a difficult problem in unconstrained field[17]. In existing system, action recognition techniques were based on optical and spatio-temporal shape template. Some popular spatio-temporal features are extracted from video sequences that is commonly extended from counterparts of 2D image domain and are demonstrated to achieve relatively good results for action recognition [7][15].

The simple features from compound images are grouped as spatial and temporal features which increases system complexity. The boundaries of the segmented human body are more sensitive to noise, which are extracted from the flow templates. Yang et al.[1] proposed a novel on representing articulated human actions, gestures and facial expressions. These actions were recognized using examples known as one shot learning. The proposed system can automatically discover sequence of actions using raw optical flow, learns statistical distributions of primitives, and because of discriminative nature in this optical flow, very competitive results are obtained using the simplest recognition and classification schemes[18][19]. The action recognition is demonstrated using KTH dataset containing noises such as low resolution, zooming, and camera movement. KTH dataset is the largest human sequence action dataset in video with different scenarios and the most popular dataset at time. Gilbert et al.[2] has stated few approaches for action recognition, the problems on shared aspects, includes the necessity of handling significant class variation, occlusions, viewpoint, illumination, and scale changes. The features such as presence of background clutter as been overcome and object recognition represents an object as a bag of visual words via a histogram. Another solution to recognize the actions within video sequences is to use a mined hierarchical grouping at simple corners.

In this paper, the proposed system uses Nearest Neighbor Classifier, widely used in action recognition, which calculates the absolute distance between the testing vector and the training vector. To reduce the quantity of the feature vectors another method used is Gaussian Mixture Model classifiers, to model the training data and speed up recognition time. Li Liu et al. [4] proposed a model for human actions recognition is based on key pose representation wherein each frame of action sequence is treated as a pose. In multi-resolution analysis of image processing, the pose is described using Extensive Pyramidal Features (EPFs), which is composed of the Gabor, Gaussian, and wavelet Laplacian pyramids. Konstantinos et al.[8] proposed a unified framework for the interrelated concepts of action spotting, the spatiotemporal detection and localization of human actions in video. A compact local descriptor of video dynamics in the context of action spotting and recognition is introduced based on visual space-time oriented energy measurements. This descriptor can efficiently compute raw image data and thereby forgoes the problems typically associated with
flow-based features. The proposed approach for action spotting and recognition are evaluated on a wide variety of standard datasets such as KTH, UCF sports and Hollywood2 evaluation datasets. Guha and Ward et al.[11] have divided the video sequence into small subsequences. On the first frame of each subsequence they find a 2D interest points and calculate the different order moments of the pixels in the neighborhood of those points. The points are called as local motion pattern. Each video is represented using sparse linear combination of STIP features. Here UCF sports data is compared to state-of-the-art methods and it achieves inferior performance on complex datasets. Samanta et al.[12] presents objects, human figure and background possess some kind of regularity in smoothness, surface curvature and texture. The characteristics of scene and light are reflected in image through spatial consistency. Unlike traditional Gaussian derivative or Gabor filter analysis, this method detects interest points based on 3D facet model on space-time domain without explicit noise removal. Yuanbo Chen et al[20] proposes a new spatio-temporal interest point detector based on flow vorticity, which not only suppress the effects of camera motion but also provide prominent spatio-temporal interest points around key positions. To separate foreground and background motions from camera is a critical problem in human action recognition under unconstrained environments. The detected points from interest point detector are local dense, and adopts the frame with randomly selected size for feature extraction. In proposed system scale invariance is achieved without invoking multiple spatiotemporal-scales, thereby reducing complexity of the system. Jiang et al.[23] proposed a shape-motion prototype-based approach, representing an action as a sequence of prototypes for efficient and flexible action matching in long video sequences. The prototype-based approach discovers the information about human body configuration, but relies on object segmentation and tracking, which is typically difficult and time consuming. The motion-based approaches extract optical flow features for recognition, but rely on segmentation of foreground for reducing effects on background flows. The space time feature-based approaches either characterize actions using global space-time 3D volumes or compactly using sparse space-time interest points. Patil et al.[24] proposed the most common method for recognizing human action is based on human visual interactions. The problem with authentication system is based on fingerprint, voice, and iris. Face recognition is more significant in recognition, not just because of its ability in potential applications, yet in addition to the capability of solving classification problems like object recognition. The identity of face is achieved by comparing a query face image against a template face image by one-to-one matching whose identity is being claimed on face verification. Chaitra B H et al.[25] proposed an approach to recognize human actions using two dimensional discrete cosine transforms (2DDCT) and self organize map (SOM) Neural Network as classifier. The framework utilizes two-dimensional discrete cosine transform (2D-DCT) for image compression and self organizing map (SOM) neural network for recognition purpose. The action images are compressed for feature processing, using two dimensional discrete cosine transform (2D-DCT). At the point when 2D-DCT is applied, high-coefficients in an image are discarded. Later 2D-IDCT is applied to regenerate the compressed image, which is blurred due to loss of quality and also smaller in size. A novel represents a human action primitive that captures the spatial layout, shape, temporal extent, as well as the motion flow of a primitive. The proposed approach applies view-invariant features to address multi-view action recognition from different perspectives.

III. PROPOSED METHODOLOGY

The proposed system is divided into two sub divisions such as offline training and online testing (fig. 3.1). In offline testing, the actions are being trained to the system based on the movement of object. Image sequence in these trained data is resized and fed to classifiers for detecting and segmenting object without the noise such as camera movements, zoom, shadows etc., further the actions performed by the person are shown in the bounding boxes. The first set of global and holistic features is concerned with the shape of the foreground object. The object is segmented from the detected foreground area. Spatial information, such as human action information, can be preserved by the
detected points of interest, and stored in databases. In online testing the data becomes available in sequential order, and then using the histogram range of the database, the dimension of the feature vector is reduced. Finally on comparing with database the results show which action is being present in the test data.

In offline training, the labeled data is used to train machine on different data sets. The data obtained, is further fed for extracting silhouettes and describing features on point of interest. Finally, the obtained feature is compared with database to produce human action.

In Feature extraction (fig. 3.2), extracting the feature is based on information on moving objects and points of interest. Locating moving objects is processed by prewitt edge detector, which segments the object based on movement, zoom, brightness etc. segmentation of objects varies the intensity in pixels. In surveillance camera, even background image is being considered hence Gaussian mixture model (GMM) is used to obtain silhouette by background subtraction.

In Gaussian mixture model (fig. 3.3) the Low pass filter is used to reduce the noise from frames. A temporal difference is applied to extract background regions then background image is constructed by GMM and silhouette is obtained by background subtraction. The edge detector is used to detect the location of moving object from foreground image.
Points of interest are local spatio-temporal features obtained from frames either from salient or descriptive action. The actions performed are usually shown in boundary called Bounding Boxes. These Bounding Boxes are used especially to locate boundary around the hands. Gabor filters are used to filter region interest based on various orientations. The Gabor filter is a linear filter which is widely used in visual recognition systems and provides accurate description of spatial aspects in simple receptive fields. Gabor filters are employed to extract orientation information, in which temporal domain filter can detect variation in intensity where as in spatial domain 2D Gabor filter is used as Gaussian kernel function modulated in wave form. The points of interest is obtained after convolving the bounding boxes with filters the human action can be represented using local and global properties trained in data sets.

Feature description (fig. 3.4) consists of two features vectors namely box features and clouds feature described by moving object location and point of interest. Box feature is concerned with shape and action obtained from global and holistic features of foreground object. Box feature measures the ratio of object’s height and width along with speed of objects. These features are obtained by Prewitt Edge Detector. Cloud feature are scale dependent, features are extracted from point of interest on different scales. The data about human body proportions ratio is used. A height of person is estimated via minimum and maximum value of static cloud video sequence. To have an equal grid of all frames in video sequence, the threshold value is fixed based on viewing point. The empirical cumulative distribution function reduces feature space dimension, makes system less sensitive to noises and invariant duration for each action sequence. It processes images in order to adjust the contrast of image by modify the intensity distribution. The system separates the empirical cumulative distribution function into portion images. The histogram range [40] from all data sets is fed to quantization to reduce noise and space dimension. The histogram range obtained from points of interest compressing a range of value to single quantum value.

In online testing, the first two stages are similar to offline training. Then, using the histogram range from database, the dimension of the feature vector is reduced. Thus, the results show which action is present in the test data.
In Feature Reduction and Classification (fig. 3.5) the histogram range of the training database transforms the testing data to a feature vector. The Nearest Neighbor’s Classifiers (NNC) is a non-parametric method utilized for classification and regression. The input consists of the training examples in the feature space. The output depends on whether NNC uses classification or regression. Too large input is considered as redundant data, the input data is transformed into set of features these features are called as Feature extraction. Feature extraction is performed on raw data prior to applying $k$-NN algorithm on the transformed data in feature space. Another classifier used is Gaussian mixture model Classifiers (GMMC). Gaussian model is stated as a probabilistic model that accepts all the data points are generated from a finite number of Gaussian distributions with obscure dimension. To diminish the amount of the feature vectors an alternative technique is being utilized namely GMMC. It is demonstrated that training data will accelerate the recognition time and utilize $k$ Gaussian functions to model each feature of the feature vectors in the database. Nearest Mean Classifier (NMC) [26] is the mean value of the feature vectors of the same action and the same view. The NMC uses minimum distance between the testing vector and training vectors. An absolute distance is chosen for the recognition decision. Therefore, NMC is more suitable for real-time recognition and has a better recognition rate.

![Figure 3.5: Feature Reduction and Classification](image)

**A. Model Validation**

**KTH Action Dataset:**

KTH dataset utilized for action recognition which is comprised of 25 actors who performs 6 actions like running, walking, boxing, jogging, handclapping, and hand waving in four distinctive scenarios. Using a static camera, in the homogenous background, a total of 600 video clips in the dataset and each video only contains one person performing a single action. The sequences are taken with a frame size of $160 \times 120$.

**Weizmann Action Dataset:**

This dataset consisted of 9 actions such as bending; running; walking, skipping, place jumping, side movement, jumping forward, two hand waving, and one hand waving that were performed by total 9 subjects. This dataset includes 90 video clips with an average of 15 frames per clip where, the frame size is $144 \times 180$.

**The Multi-camera Human Action Video Data (MuHAVi) datasets:**

This dataset consisted of selected action sequences for the evaluation of action recognition methods. The creators collected a large body of human action video (MuHAVi) [27] data using 8 cameras.
There are 17 action classes as performed by 14 actors. Each actor performs each action several times in the action zone highlighted using white tapes on the scene floor. As on-screen characters were beginners the pioneer needed to interfere with the entertainers now and again and approach them to re-try the activity for consistency. There are 8 CCTV Schwann cameras located at 4 sides and 4 corners of a rectangular platform.

IV. EXPECTED EXPERIMENT RESULTS

We first evaluate the performance of the proposed approach on the three challenging action datasets namely KTH Action Dataset, Weizmann Action Dataset and Multi-camera Human Action Video Data (MuHAVi) datasets. All of the testing inputs are uncompress ed AVI video files. Secondly, evaluating the performance of separation of descriptors (table 4.1) and combination of descriptors used are Nearest Neighbor’s Classifiers (NNC), Gaussian Mixture Model Classifiers (GMMC), and Nearest Mean Classifier (NMC) to yield the histogram represents for actions.

Table 4.1: comparing all 3 datasets with 3 classifiers

| Classifiers          | Weizmann dataset |
|----------------------|------------------|
| Proposed approach    | Nearest Mean Classifier (NMC) 95.56% |
| G.Goudelis et al[30] | RBF and SVM 95.45% |
| JA Nasiri et al[31]  | SVM 95.56% |
| J.Jiang et al[32]    | GMRM 95.10% |
| S.Zhang et al[33]    | RBF 88.1% |

Table 4.2: KTH Action Dataset

| Classifiers | KTH | WEIZMANN | MuHAVi |
|-------------|-----|----------|--------|
| NNC         | 89.31% | 87.78% | 92.5% |
| GMMC        | 90.2% | 91.1% | 93.2% |
| NMC         | 90.5% | 95.5% | 97% |

Recognition performance of all datasets proposed using various feature vectors and classifiers is shown in Table 4.2, Table 4.3 and Table 4.4

Table 4.3: Weizmann Action Dataset

| Classifiers       | Kth dataset |
|-------------------|-------------|
| Proposed approach | Nearest Mean Classifier (NMC) 90.58% |
| A Gilbert et al[28]| SVM and boosting 94.5% |
| Gilbert et al[2]  | Data mined 89.9% |
| Nowozin et al [29]| Subseq boost SVM 87.04% |
| Suhuldt et al [3] | SVM split 71.71% |
| Kel et al[10]     | Vol boost 94.50% |
Table 4.4: MuHAVi Action Dataset

| Classifiers                  | MuHAVi |
|-----------------------------|--------|
| Nearest Mean Classifier (NMC) | 97.5%  |
| MHI and HOG                 | 96.26  |
| HOG                         | 92.36  |
| STBPM                       | 93.75  |

A. Subject Invariance

Subject invariance (Table 4.5) is evaluated by cross validation known as Leave-One-Out Cross Validation (LOOCV), in which estimate is more accurate of out-of-sample accuracy and more efficient to use data, as every observation is used for both training and testing. It selects a group of clips from a single subject in a dataset as the testing data, and the rest of the clips are the training data. The repeated progress ensures that each group of clips in the dataset is used once as the testing data. For the KTH dataset, the training set contains 24 subjects. For the WEIZMANN dataset; the training set contains 8 subjects. For the MuHAVi dataset, the clips of 6 subjects were used for training and the clips belonging to the remaining subjects are used for validation.

Table 4.5: Comparative results on KTH, Weizmann, and MuHAVi datasets for subject invariance

|                  | KTH     | WEIZMANN | MuHAVi  |
|------------------|---------|----------|---------|
| Proposed system  | 90.58%  | 95.56%   | 97.5%   |
| S Zang et al[33] | 98.30%  | 95.80%   | 96.40%  |
| S.Gong et al[37] | 93.17%  | 96.66%   | 91.78%  |
| Niebles et al[6] | 83.3%   | 90%      | -       |
| Savarese et al[38]| 86.83%   | -        | -       |
| Beipping et al[39]| 90.23%   | -        | -       |

B. View Invariance

View Invariant in visual object recognition is the ability to recognize visual objects despite the vastly different images that each object can project View Invariant (Table 4.6) is evaluated as a group of clips from a single view in a dataset is employed as the training data and the remaining clips are the frames, of each action as the testing data. Five actions out of 17 in the MuHAVi dataset were chosen as the experimental data similar to the subject invariance evaluation. Then, one of the eight views in the MuHAVi dataset is utilized in training and the other view is utilized in testing. This procedure is repeated for all 8 views and the resulting recognition rates are then averaged. This was repeated so that each group of clips in this dataset is used once as the training data.

Table 4.6: Comparative results on MuHAVi datasets for View invariance

|                  | MuHAVi  |
|------------------|---------|
| Proposed system  | 81.43%  |
| A.Eweiwi et al[10]| 77.50%  |
| S.Gong et al[37] | 72.8%   |

V. DISCUSSION

This paper presents an approach that is capable to capture sequence of motions and occlusions at a low computational cost. The features of multi-view action recognition are extracted from various temporal scales, which are demonstrated using global spatial-temporal distribution. There is no need to re-train the system if the scenario changes. The experiments on the KTH and WEIZ- MANN...
datasets demonstrate that using subject-invariant features obtained by extracting the features from interest points. The experiments on the WEIZ- MANN and MuHAVi datasets demonstrate that using view-invariant features obtained by extracting holistic features from clouds of interest points is highly discriminative for recognizing actions under different view changes. The view-invariant features are obtained by extracting features from different temporal scale clouds, which are modeled on the explicit global, spatial and temporal distribution of interest points.

REFERENCES

I. Y. Yang, I. Saleemi, and M. Shah, “Discovering motion primitives for unsupervised grouping and one-shot learning of human actions, gestures, and expressions,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 7, pp. 1635–1648, Jul. 2013.
II. A. Gilbert, J. Illingworth, and R. Bowden, “Action recognition using mined hierarchical compound features,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 5, pp. 883–897, May 2011.
III. C. Schuldt, I. Laptev, and B. Caputo, “Recognizing Human Actions: A Local SVM Approach,” Proc. Int’l Conf. Pattern Recognition, vol. 3, pp. 32-36, 2004. P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, “Behavior Recognition via Sparse Spatio-Temporal Features,” Proc. 14th Int’l Conf. Computer Comm. and Networks, pp. 65-72, 2005.
IV. L. Liu, L. Shao, X. Zhen, and X. Li, “Learning discriminative key poses for action recognition,” IEEE Trans. Cybern., vol. 43, no. 6, pp. 1860–1870, Dec. 2013.
V. C. Thurau and V. Hlavác, “Pose primitive based human action recognition in videos or still images,” in Proc. IEEE Conf. Comput. Vis. Pattern Recogn. Anchorage, AK, 2008, pp. 1–8.
VI. J. Niebles and L. Fei-Fei, “A hierarchical model of shape and appearance for human action classification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recogn., Minneapolis, MN, 2007, pp. 1–8.
VII. Piotr Dollár, Vincent Rabaud, Garrison Cottrell, and Serge Belongie. Behavior recognition via sparse spatio-temporal features. In VS-PETS, October 2005.
VIII. Konstantinos G. Derpanis, Member, IEEE, Mikhail Sizintsev, Member, IEEE, Kevin J. Cannons, and Richard P. Wildes, Member, IEEE “Action Spotting and Recognition Based on a Spatiotemporal Orientation Analysis”
IX. Geert Willems, Tinne Tuytelaars, and Luc Van Gool. Behavior recognition via sparse spatio-temporal features. In VS-PETS, October 2005.
X. Piotr Dollar, Vincent Rabaud, Garrison Cottrell, and Serge Belongie. Behavior recognition via sparse spatio-temporal features. In VS-PETS, October 2005.
XI. Tanaya Guha and Rabab Kreidieh Ward. “Learning sparse representations for human action recognition.” IEEE Trans. Pattern Analysis and Machine Intelligence, 34(8):1576–1588, 2012.
XII. M. S. Ryoo, “Human activity prediction: Early recognition of ongoing activities from streaming videos,” in 2011 International Conference on Computer Vision, 2011, no. Iccv, pp. 1036–1043.
XIII. W. S. Lasceki, Y. C. Song, H. Kautz, and J. P. Bigham, “Real-time crowd labeling for deployable activity recognition,” in Proceedings of the 2013 conference on Computer supported cooperative work - CSCW ’13, 2013, p. 1203.
XIV. P. Scovanner, S. Ali, and M. Shah, “A 3-Dimensional Sift Descriptor and Its Application to Action Recognition,” Proc. Conf. Multimedia, pp. 357-360, 2007.
XV. K. Mikolajczyk and H. Uemura, “Action Recognition with Motion Appearance Vocabulary Forest,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.
XVI. I. Junejo, E. Dexter, I. Laptev, and P. Perez, “View-Independent Action Recognition from Temporal Self-Similarities,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 33, no. 1, pp. 172-185, Jan. 2011.
XVII. P. Parameswaran and R. Chellappa, “View Invariance for Human Action Recognition,” Int’l J. Computer Vision, vol. 66, no. 1, pp. 83101, 2006.
XVIII. R. Souvenir and J. Babbs, “Learning the Viewpoint Manifold for Action Recognition,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.
XIX. Y. Chen, Z. Li, X. Guo, Y. Zhao, and A. Cai, “A spatio-temporal interest point detector based on vorticity for action recognition,” in Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW), Jul. 2013, pp. 1–6.
XX. P. Parameswaran and R. Chellappa, “View Invariance for Human Action Recognition,” Int’l J. Computer Vision, vol. 66, no. 1, pp. 83101, 2006.
XXI. R. Souvenir and J. Babbs, “Learning the Viewpoint Manifold for Action Recognition,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2008.
XXIII. Z. Jiang, Z. Lin, and L. S. Davis, “Recognizing human actions by learning and matching shape-motion prototype trees,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 3, pp. 533–547, Mar. 2012.

XXIV. Shailaja A Patil and Dr. P. J. Deore Department of Electronics & Telecommunication Engineering, R. C. Patel Institute of Technology, Dist: Maharashtra “Face Recognition: A Survey” Informatics Engineering, an International Journal (IEIJ), Vol.1, No.1, December 2013

XXV. Chaitra B H PG Student Department of CSE, RVCE Anupama H S Assistant Professor Department of CSE, RVCE Bangalore Cauvery N K Professor and Head Department of ISE, RVCE Bangalore” Human Action Recognition using Image Processing and Artificial Neural Networks” International Journal of Computer Applications (0975 – 8887) Volume 80 – No.9, October 2013.

XXVI. Piew Datta and Dennis Kibler Department of Information and Computer Science University of California “Symbolic Nearest Mean Classifiers”

XXVII. Sanchit Singh, Sergio A Velastin, DIRC, Kingston University Kingston upon Thames, UK Hossein Ragheb ISBE, University of Manchester Manchester, UK “MuHAVi: A Multicamera Human Action Video Dataset for the Evaluation of Action Recognition Methods” Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance.

XXVIII. A. Gilbert, J. Illingworth, and R. Bowden, “Scale invariant action recognition using compound features mined from dense spatio-temporal corners,” in Proc. Eur. Conf. Comput. Vis., 2008, pp. 222–233.

XXIX. Y. Sheikh, M. Sheikh, and M. Shah, “Exploring the space of a human action,” in Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV), vol. 1. Oct. 2005, pp. 144–149.

XXX. G Goudelis, K Karpouzis, S Kollias, Exploring trace transform for robust human action recognition. Pattern Recogn. 46(12), 3238–3248 (2013)

XXXI. JA Nasiri, NM Charkari, K Mozafari, “Energy-based model of least squares twin support vector machines for human action recognition.” Signal Process. 104, 248–257 (2014)

XXXII. J Jiang et al., “Human action recognition via compressive-sensing-based dimensionality reduction.” Optic-International Journal for Light and ElectronOptics 126(9), 882–887 (2015)

XXXIII. S Zhang, W Zhang, Y Li, in Proceedings of 2016 Chinese Intelligent Systems Conference. Human action recognition based on multifeature fusion (Springer, 2016)

XXXIV. F Murtaza, MH Yousaf, SA Velastin, in Frontiers of Information Technology (FIT), 2015 13th International Conference on. Multi-view human action recognition using histograms of oriented gradients (HOG) description of motion history images (MHIs) (IEEE, 2015)

XXXV. F Murtaza, MH Yousaf, SA Velastin, “Multi-view human action recognition using 2D motion templates based on MHIs and their HOG description”. IET Comput. Vis. 10(7), 758–767 (2016)

XXXVI. S Maity, D Bhattacharjee, A Chakrabarti, “A novel approach for human action recognition from silhouette images.” IETE J. Res. 63(2), 160–117 (2017)

XXXVII. S. Singh, S.A. Velastin, and H. Ragheb, “MuHAVi: A multi camera human action video dataset for the evaluation of action recognition methods,” in Proc. 17th IEEE Int. Conf. Adv. Video Signal Based Survell. (AVSS), Sep. 2010, pp. 48–55.

XXXVIII. A. Ali and J. K. Aggarwal, “Segmentation and recognition of continuous human activity,” in Proc. IEEE Work shop IEEE Detection Recognit. Events Video, Jul. 2001, pp. 28–35.

XXXIX. H Beiping, Z Wen, “Fast human detection using motion detection and histogram of oriented gradients.” JCP 6(8), 1597–1604 (2011)

XL. Navneet Dalal, Bill Triggs, Cordelia Schmid “Human Detection Using Oriented Histograms of Flow and Appearance” HAL Id: inria-00548587 https://hal.inria.fr/inria-00548587 Submitted on 20 Dec 2010.

XLI. Steffen bickel and tobias scheffer department of computer science Humboldt university “multi-view clustering” proceedings of fourth IEEE international conference on data mining.