FRAME SAMPLING ASSISTED DEPTH MOTION MAPS FOR DEPTH BASED HUMAN ACTION RECOGNITION

Sivanagi Reddy Kalli[1], K. Mohanram[2], S.Jagadeesh[3]
[1,2,3] *Professor, Dept., of ECE, Sridevi Women's Engineering College,Hyderabad,*
[1]*ksivanagireddy@swec.ac.in.*,[2]*kmohanram@swec.ac.in.*,[3]*sjagadeesh@swec.ac.in.*

Abstract: The discovery of depth sensors has brought new opportunities in the Human Action Research by providing depth image data. Compared to the conventional RGB image data, the depth image data has additional benefits like color, illumination invariant, and provides clues about the shape of body. Inspired with these benefits, we present a new Human Action Recognition model from depth images. For a given action video, the consideration of an entire frames constitutes less detailed information about the shape and movements of body. Hence we have proposed a new method called Frame Sampling to reduce the frame count and chooses only key frames. After key frames extraction, they are processed through Depth Motion Map for action representation followed by Support Vector Machine for classification. The developed model is evaluated on a standard public dataset captured by depth cameras. The experimental results demonstrate the superior performance compared with state-of-art methods.

Keyword: new dataset; artificial intelligence; sensor; natural language processing; fake detector; machine learning.

I. INTRODUCTION

In recent years, Human Action Recognition (HAR) has been attracted a lot of research attention in the community of computer vision. HAR is necessary for several applications that demand for public safety, people's behavior, Visual Surveillance, Virtual Reality, Human-Computer Interaction (HCI), Video Indexing, etc. [1, 2]. The conventional methods developed for HAR are mainly focused on Red Green Blue (RGB) colored data. However the major problem with such data is less recognition performance due to the limited available information about the shape [3] and poses [4] of the body. Even though the RGB based HAR methods have gained a better performance in particular contexts, they have limited performance under varying illuminations. Moreover even they are failed to recognize human actions under some challenging scenarios like cluttered backgrounds, and occlusions.

Recently, the development of low-cost depth sensors (ex. Washington, Redmond, and Microsoft Kinect) has directed the HAR research into another direction. Because, the data acquired through these low-cost depth sensors is illumination invariant, color invariant and also provides clues about the shape and pose of human body performing an action. Compared to RGB cameras, the Kinect sensors provide structural information of the body which has a significant impact on the HAR by simplifying the variations in intra-actions and removing noises and cluttered backgrounds. Figure.1 and Figure.2 shows an example of RGB colored action video and Depth action video respectively. Hence, researchers have put a lot of effort to the depth based HAR and developed several feature extraction techniques like Depth Motion Maps [5], Depth Cuboid Similarity feature (DCSF) [6], Super Normal Vector (SNV) [7], and Histogram of oriented 4D normal (HON4D)[8].

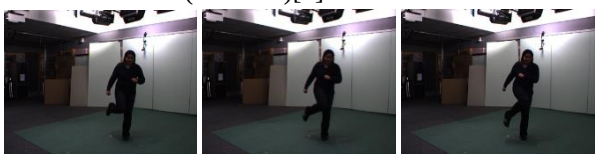


Figure.1 RGB colored action video

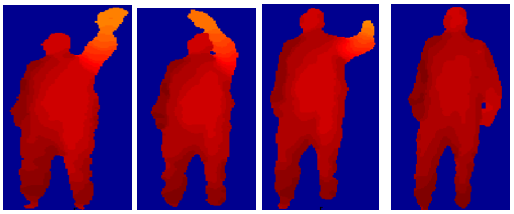


Figure.2 Depth action video

Even though an extensive research is carried out over depth data based HAR, still there exists a room for further research [9]. In the case DMM computations, for a given action video, the entire frames are processed by which the obtained action map won't have a detailed motion information. Moreover, the computational burden is also high due to this process. In the given action video, only few frames exists which have more information about the action present in it, and if we are able to extract only that frames the additional computational burden will get reduced. As well as the obtained map have clear information about the shape and motion of body. Based on this perspective, we have developed a new depth data assisted HAR model, called as Frame Sampled DMM (FSDMM). Initially, the FSDMM extracts the key frames through a new process called frame sampling and then computes DMM. For

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

classification, we have employed Support Vector Machine (SVM) algorithm.

Rest of the paper is organized as follows; Section II summaries the literature survey details. After that, the complete details of the proposed HAR model are explained in section III. Section IV explains the details of experimental results. Finally, the concluding remarks are represented in section V.

II. LITERATURE SURVEY

Several action recognition models are constructed by considering the depth data from action videos. A detailed and summarized survey over different HAR methods is described in this section.

Initially, X. Yang et al. [5] projected the depth maps onto three orthogonal planes and accumulate the global activity through the entire video sequence to obtain the DMM. Then "Histogram of Oriented Gradients (HOGs)" is generated through these DMMs for action representation. O Omar et al. [8] developed a new descriptor through the histograms which can capture the features like spatial coordinates, depth, and time in 4D space, distributed over the orientations of different surface normals. To obtain this histogram, the 4D projectors are created at first for quantization of 4D space and then focused over the representation of possible 4D normal directions. Next, a SNV [7] is constructed through the low-level polynomials, obtained through the grouping of super surface normal in depth sequence. Finally these SNVs are fed to the action recognition model. To increase the robustness, another 4D descriptor called as Random Occupancy Pattern (ROP) is proposed by J. Wang et al. [10] for HAR. The ROP is a sparse coding method which can have robustness with occlusions and noise.

With an inspiration of the success of Spatio-Temporal Interest Points (STIPs) [11, 14] in RGB videos based HAR, L. Xia L and J. Agarwal [6] proposed a new filtering method called DSTIP to extract the STIPs through which the noise is suppressed. Further, a novel feature called, "Depth Cuboid Similarity Feature (DCSF)" to represent the depth of the local 3D cuboid located around the DSTIPs with an appropriate size. Next, in the method proposed by D. Kim et al. [12], initially the front view depth action image is processed for side view generation. Next, the both side and front views of action image are into two descriptors, namely "Depth Motion Appearance (DMA)" and "Depth Motion History (DMH)". SVM classifier is accomplished here for action classification. Another approach based on multi-view projection is proposed by Chen at al. [13] in which the initial projection of the original depth action image is accomplished over three Cartesian planes. Next, the DMM is obtained by the accumulation of absolute difference between two successive depth maps

Further, Chen et al., [15] proposed to construct Local Binary Patterns (LBPs) followed by Fisher Kernel Vectors after the computation of DMMs of three orthogonal views. Kernel Based Extreme Learning Machine (KELM) is employed for action classification. This method also had shown its contribution at fusion level by applying two levels of fusions such as feature level fusion and decision level fusion. Further, the same author, Chen et al., [16] generated the DMMs at segment level. Each segment is characterized by DMM followed by LBP to extract the location and rotation invariant information. In the final stage, Fisher kernel is accomplished to generate a compact feature vector for every action and ELM is applied for action classification. Next, an extension of DMM called as D^2MM is proposed by Sandhya rani et al., [17] to represent an action video by removing the fake moving pixels due to cluttered background and body shaking movements. Though the DMM have an excellent perform in depth data assisted action recognition, the consideration of an entire frames for DMM construction lasts temporal information.

III. PROPOSED ACTION RECOGNITION

A. Overview

The main aim of the proposed action recognition model is to achieve an efficient recognition performance with less computational burden. For this purpose, we have introduced a new technique called frame subsampling to extract only eh key frames of an action video. For a given action video, initially this model derives only key frames through frame sampling and then computes DMM. For classification purpose, we have employed Support Vector Machine (SVM) algorithm.

B. Frame Sampling

DMMs are basically used to determine the motion and shape information of a depth action video. Generally the DMM is obtained by the accumulation of difference between the frames of an action sequence. Actually the DMM is constructed by considering all the frames into the process. However, this process may not capture more and detailed temporal information. Hence to capture more and effective motion information, we have employed Frame Sampling in which only few frames are selected called as key frames. Different with multi-level temporal sampling (MTS) [18] which is based on the motion energy of key frames, the Frame Sampling is a simple frame selection process [19] based on the detection of motion features.

Consider a depth action image sequence $F = (F_1, F_2, F_3, \ldots, F_T)$, where T is total number of frames, F_i is the frame at i^{th} time instant. Define a new variable D_t as difference image sequence, obtained by the computation of difference between successive frames, as

$$D_t = \begin{cases} \sqrt{(F_t - F_{t-1})^2} & if\ t > 1 \\ F_1 & if\ t = 1 \end{cases} \quad (1)$$

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Where *t* is the time index. In the above expression.(1), the D_t is measured as the difference between successive frames when the time index is greater than 1. On the other hand, at the time index t=1, the D_t is simply the first frame which denotes the reference difference image. For time index above 1, the difference image $D_t \in R^{M\times N}$ is obtained by the square of pixel-by-pixel difference of two successive frames F_t and F_{t-1}.

After the computation of difference images, the sum of all difference values are stored in another variable and let it be S_t, measured as

$$S_t = \sum_{m=1}^{M}\sum_{n=1}^{N} D_t(m,n) \quad (2)$$

Where $D_t(m,n)$ is the difference value of two pixels at the location (m,n) in the two successive frames F_t and F_{t-1}. The $S_t \in R^T$ is a Colum array and the size of T. Each value in the variable S_t denotes a cumulated difference of two successive frames F_t and F_{t-1}. Next, we apply normalization of S_t as follows;

$$S_{min} = \min_t(S_t) \quad (3)$$

$$S_{max} = \max_t(S_t) \quad (4)$$

$$S'_t = \frac{S_t - S_{min}}{S_{max} - S_{min}} \quad (5)$$

Where S'_t is the normalized summation of successive frames difference, S_{min} is the minimum difference and S_{max} is the maximum difference. The range of S'_t is lies in between 0 and 1, where 0 denotes no difference and 1 denotes large difference. Based on the obtained normalized values, we select the key frames as;

$$Key\ Frames_t = S'_t > \tau \quad (6)$$

Where τ is the threshold values and we have approximated it as $\tau = 0.3$, means 30% of frames are discarded. In the normalized vector S'_t, we can observe different values ranging from 0 to 1. Among these values, we have to choose only the values which have a marginal difference and that difference is decided based on the threshold τ. Here we have approximated threshold τ as 0.3 means the successive frames must be differed at least by 30%.

C. DMM

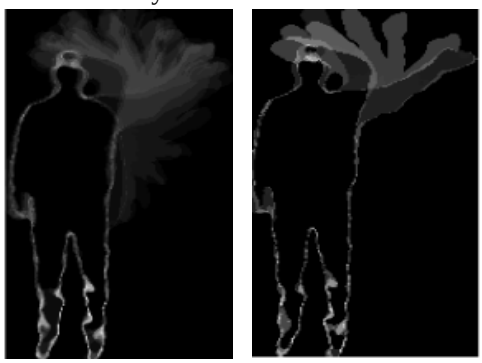
Once the key frames are extracted from frame sampling, we compute DMM. DMM is initially introduced by Yang et al. [5], which exploits the motion information from a depth sequence. DMM gives a visual perception of human activity and it is generated by the accumulation of motion energy throughout the entire depth sequence. Depth maps contain additional depth coordinates along with Cartesian coordinates which are generally provided by color images. Due to the presence of additional depth coordinates, depth maps are more informative than the normal color images. DMM based action representation transforms the action recognition issue from 3D to 2D and then applies for HAR. Particularly, the DMMs are constructed by the accumulation of energy after projecting the depth frames over an orthogonal Cartesian plane. Basically, the main intention of DMM is to signify the shape and motion of an action.

Unlike the threshold-based DMM generation [5], Chen Liu et al. [13] evaluated the DMM based on the accumulation of motion energy of the absolute difference between consecutive frames. This method preserves the motion information more effectively and hence our method also considered it as a base for DMM evaluation. According to Chen Liu et al. [13], the DMM evaluation is formulated as follows;

$$DMM = \sum_{t=0}^{N-2} |D(i,j,t) - D(i,j,t-1)| \quad (7)$$

Where $D(i,j,t)$ is a pixel value at the position (i,j) of a depth frame at the instant of *t* and $D(i,j,t-1)$ is a pixel value at the position (i,j) of a depth frame at the instant of *t-1*, where *t* varies from 0 to N-1.

DMM can acquire the shape and motion cues of a depth action image more effectively, results in a discriminative map that gives more discrimination between different actions. This discrimination is provided through the spatial distribution of energy. Figure.2 shows the DMMs derived by considering the entire frame and the key frames.



(a) (b)

Figure.2 (a) DMM with entire frames and (b) DMM with key frames

IV. SIMULATION EXPERIMENTS

In this section, we investigate our developed HAR model on a publicly available benchmark data set, Microsoft Research (MSR) Action 3D dataset [20]. We employ MATLAB software for the simulation of the proposed approach. In this section initially, the details of the "MSR Action 3D dataset" are explored. Next, the obtained simulation results are described. Finally, a comparative analysis is described between the proposed and conventional approaches through the obtained results.

A. MSR Action 3D dataset

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

The "MSR action 3D dataset" is an action dataset constructed through 20 different actions such as forward kick, side kick, jogging, throw, pick up, tennis serve, golf-swing, high arm wave, hammer, draw X, forward punch, draw circle, two hand wave, draw tick, side boxing, bend, hand clap, high throw, hand catch, horizontal arm wave. Every action is performed by ten actors and totally three times. This is a very challenging dataset which has a lot of speed variations and also the actions have more similarities. A single viewing point is only used at which the actions are faced with a frontal view with the camera while capturing. Some depth action sequences of this dataset are shown in Figure.3.

Under the simulation experiments, totally three types of experiments are conducted by dividing the entire action set into three different sets. Half of the subjects are used for training and remaining half are used for testing. At every set, the training was done with the subjects 1, 3, 5, 7 and 9 and the testing is done through the subjects 2, 4, 6, 8, and 10. The three action sets are shown in Table.1. After the simulation of Subset 1, Subset 2 and Subset 3, the obtained confusion matrices are shown in table.2, Table.3 and table.4 respectively.

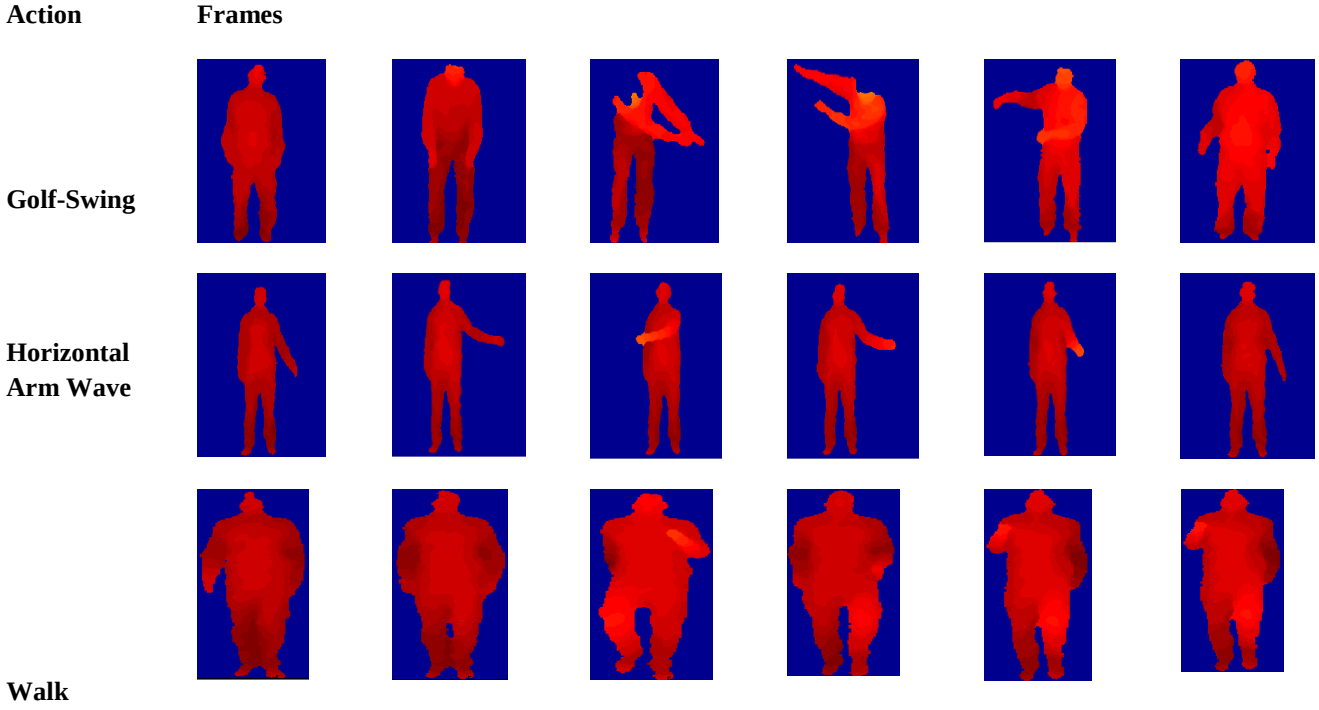


Figure.3 Some action samples of MSR action 3D dataset

Table.1. subsets of MSR Action 3D dataset

Subset 1	**Subset 2**	**Subset 3**
Hand Clap (HCP)	Draw Circle (DCL)	Forward Kick (FK)
High Throw (HT)	Draw Cross (DCS)	Jogging (JG)
Horizontal Arm wave (HAWH)	Draw Tick (DTK)	Tennis Serve (TSR)
Tennis Serve (TSR)	Hand Catch (HCT)	Pick-up & Throw (PUT)
Forward punch (FP)	High Arm Wave (HAWV)	Golf Swing (GS)
Hammer (HM)	Side-boxing (SB)	Tennis Swing (TSG)
Bend (BD)	Forward Kick (FK)	Side Kick (SK)
Pick-up & Throw (PUT)	Two-hand Wave (THW)	High Throw (HT)

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

Table.2. Confusion matrix of Subset 1

	HCP	HT	HAWH	TSR	FP	HM	BD	PUT
HCP	**0.95**	0.00	0.05	0.00	0.00	0.00	0.00	0.00
HT	0.00	**0.93**	0.00	0.07	0.00	0.00	0.00	0.00
HAWH	0.05	0.00	**0.95**	0.00	0.00	0.00	0.00	0.00
TSR	0.00	0.10	0.00	**0.90**	0.00	0.00	0.00	0.00
FP	0.04	0.00	0.00	0.00	**0.96**	0.00	0.00	0.00
HM	0.00	0.10	0.00	0.00	0.00	**0.90**	0.00	0.00
BD	0.00	0.00	0.00	0.058	0.00	0.00	**0.91**	0.032
PUT	0.00	0.05	0.00	0.05	0.00	0.00	0.05	**0.85**

Table.3. Confusion matrix of Subset 2

	DCL	DCS	DTK	HCT	HAWV	SB	FK	THW
DCL	**0.78**	0.11	0.11	0.00	0.00	0.00	0.00	0.00
DCS	0.10	**0.72**	0.18	0.00	0.00	0.00	0.00	0.00
DTK	0.00	0.14	**0.86**	0.00	0.00	0.00	0.00	0.00
HCT	0.00	0.00	0.00	**0.94**	0.00	0.06	0.00	0.00
HAWV	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
SB	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
FK	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00
THW	0.00	0.00	0.00	0.10	0.00	0.00	0.00	**0.90**

Table.4. Confusion matrix of Subset 3

	FK	JG	TSR	PUT	GS	TSG	SK	HT
FK	**0.95**	0.00	0.00	0.05	0.00	0.00	0.00	0.00
JG	0.00	**0.92**	0.00	0.00	0.00	0.00	0.08	0.00
TSR	0.00	0.00	**0.94**	0.06	0.00	0.00	0.00	0.00
PUT	0.00	0.00	0.00	**0.90**	0.00	0.00	0.00	0.10
GS	0.00	0.00	0.00	0.00	**0.92**	0.08	0.00	0.00
TSG	0.00	0.00	0.00	0.00	0.09	**0.91**	0.00	0.00
SK	0.05	0.00	0.00	0.00	0.00	0.00	**0.95**	0.00
HT	0.00	0.00	0.00	0.06	0.00	0.00	0.00	**0.94**

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

The confusion matrix of proposed approach after the simulation of Subset 1 is shown in Table.2. In this table, the maximum recognition rate (96%) is achieved for Forward Punch action and minimum recognition rate achieved for Pick-up & Throw (85%). Since the FP action is much deviated in their motion characteristics and hence it was gained maximum recognition rate. Further, the Pick-up & Throw action has similar characteristics with High Throw; the recognition rate is reduced for both actions. In this simulation, the proposed system has gained an average recognition rate of 91.87%.

Next, the confusion matrix of proposed approach after the simulation of Subset 2 is shown in Table.3. In this table, the maximum recognition rate (100%) is achieved for three actions such as High Arm Wave, Side boxing and Forward Kick and minimum recognition rate achieved for Draw Cross (72%). Here the three actions namely, Draw Circle, Draw Cross and Draw Tick have almost similar characteristics and hence the recognition rates are less. To perform all these actions, the hand follows a similar movement and thus the system confuses, resulting in a less recognition rate. Further it can also be observed that the false positive rate of all these actions is significantly high. For example, when a depth frame of draw tick is given for testing, the system has classified it as draw cross which results in increased false positives. Further, in the case of two-hand wave action, only 90% actions are classified as it is and remaining is classified as Hand Catch, because they have similar motion characteristics. Finally, in this simulation, the proposed system has gained an average recognition rate of 90%.

The confusion matrix of proposed approach after the simulation of Subset 3 is shown in Table.4. In this table, the maximum recognition rate (95%) is achieved for two actions such as Side Kick and Forward Kick and minimum recognition rate achieved for Pick up & Throw (90%). Due to the similar motion characteristics of Pick up & Throw and High Throw, it has achieved a less recognition rate. Among the given Pick up & Throw test input sequences, only 90% of actions are recognized correctly and remaining is recognized as High throw. Furthermore, due to the unique characteristics of Forward kick and Jogging, they also achieved an efficient recognition rate and it is of 95% and 92% respectively. Finally in this simulation, the proposed system has achieved an average recognition rate of 92.87%.

In the further simulation, we have done a comparative analysis between proposed and several conventional approaches through recognition accuracy. Under the conventional approaches we have considered DMM + HOG [5], Random Occupancy Patterns [10], HON4D [8] and DMA+DMH+HOG [12], and DCSF [6].

Table.5 Comparison of recognition accuracy

Method	**Accuracy (%)**
DMM + HOG [5]	85.52
DSTIP [6]	89.30
HON4D [8]	88.89
Random Occupancy Patterns [10]	86.50
DMA+DMH+HOG [12]	90.45
FSDMM	**91.58**

Table.4 represents the accuracy comparison of proposed and conventional approaches. It can be seen from the Table.5, the accuracy of the proposed system is higher compared to the conventional approaches. The most nearby method for the proposed model is DMA+DMH+HOG [12] and it achieved an accuracy of 90.45%. However, in this approach, the DMM evaluation didn't consider the frame sampling by which the recognition system will get confuse due to the overlapping of action movements. By discovering the key frames before DMM evaluation, the proposed model has gained better accuracy when compared to DMA+DMH+HOG. Furthermore, the one more conventional approach, i.e., DMM + HOG [5] is a basic approach for DMM based action recognition. In this method, the DMM is evaluated based on the thresholding by which the obtained DMM won't have significant information which is more helpful in the action recognition. The main drawback of this approach is information loss due to thresholding. Though the DSTIP method suppresses the noises effectively, they can't capture the real movements of body in depth action sequences.

V. CONCLUSION

In this paper, we have proposed a new HAR model based on depth image sequences. The proposed recognition model considers depth action image sequences as an input and recognizes the action present in that video. Under the recognition process, DMM is employed for action representation. Before the evaluation of DMM, the input video is subjected to frame sampling to derive only the important frames. The importance is measured based on relation between successive frames in the action video. By reducing the frame count, we have gained improved recognition accuracy. The main reason behind this improvement is perfect and clear discrimination between the body movements of different actions. Moreover, we also reduced computational time require d for action representation. From the simulation results, we have observed that the average accuracy of FSDMM is 91.58% while the conventional DMM is 85.52%.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

REFERENCES

[1]. Dikmen, M., Ning, H, Lin, D. J, "Surveillance event detection", TRECVID - 2008.

[2]. R. A. Seger, M. M. Wanderley, & A. L. Koerich, (2014), "Automatic detection of musicians ancillary gestures based on video analysis", Expert Systems with Applications, 41 (4), 2098–2106.

[3]. Liu J, Ali S, and Shah M., "Recognizing human actions using multiple features", In: IEEE conference on computer vision and pattern recognition, Anchorage, Alaska, USA, 23–28 June 2008, pp. 1–8.

[4]. Raptis M and Sigal L., "Poselet key-framing: a model for human activity recognition", In: Proceedings of the IEEE conference on Computer vision and pattern recognition, Portland, Oregon, USA, 23–28 June 2013, pp. 2650–2657.

[5]. Yang X, Zhang C, and Tian Y., "Recognizing actions using depth motion maps-based histograms of oriented gradients", In: Proceedings of the 20th ACM international conference on Multimedia, Nara, Japan, 29 October–2 November 2012, pp. 1057–1060.

[6]. Xia L and Aggarwal J. Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Portland, OR, USA, 23–28 June 2013, pp. 2834–2841.

[7]. Yang X and Tian Y. Super normal vector for activity recognition using depth sequences. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, USA, 23–28 June 2014, pp. 804–811.

[8]. Oreifej O and Liu Z. HON4D: histogram of oriented 4D normals for activity recognition from depth sequences. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Portland, Oregon, USA, 23–28 June 2013, pp. 716–723.

[9]. S. Escalera, V. Athitsos, and I. Guyon, (2017). "Challenges in multi-modal gesture recognition", In Gesture recognition (pp. 1–60). Cham: Springer.

[10]. Wang J, Liu Z, Chorowski J, Chen Z, and Wu Y, "Robust 3d action recognition with random occupancy patterns," in Computer vision–ECCV, Springer, pp. 872–885, 2012.

[11]. Yanshan Ali, Rongjie Xia, Q. Huang, W. Xie, and X. Li, "Survey of Spatio-Temporal Interest Point Detection Algorithms in Video", IEEE Access, Vol.5, 2017, pp.10323-10331.

[12]. Kim D, Yun W. H, Yoon H. S, and Jaehong H. S, "Action recognition with depth maps using hog descriptors of multi-view motion," in proc., of 8th International Conference on Mobile Ubiquitous Computing, Systems, Services, and Technologies, UBICOMM, pp. 2308–4278, 2014.

[13]. C. Chen, K. Liu, and N. Kehtarnavaz, "Real-time human action recognition based on depth motion maps". Journal of Real-time Image Processing, 12 (1), 155–163, 2016.

[14]. Y. Zhu, W. Chen, G. Guo, Evaluating spatiotemporal interest point features for depth-based action recognition, Image and Vision Computing, 32 (8) (2014) 453–464.

[15]. C. Chen, R. Jafari, and N. Kehtarnavaz, ``Action recognition from depth sequences using depth motion maps-based local binary patterns," in Proc. WACV, Jan. 2015, pp. 1092_1099.

[16]. [16] Chen, C., Liu, M. , Zhang, B. , Han, J. , Jiang, J. , & Liu, H. (2016). 3d action recognition using multi-temporal depth motion maps and fisher vector. In IJCAI (pp. 3331–3337).

[17]. S. Sandhya Rani, Dr. G. Appa Rao Naidu, Dr. V. Usha Shree, "D2MM-CNN: Difference Depth Motion Map and Convolutional Neural Networks for Human Action Recognition", International Journal of Advanced Science and Technology, Vol. 28, No. 15, (2019), pp. 747-763.

[18]. R. Azad, M. Asadi Aghbolaghi, S. Kasaei, and S. Escalera, "Dynamic 3D hand gesture recognition by learning weighted depth motion maps", IEEE Trans. Circuits Syst. Video Technol., to be published.

[19]. A. Sobral, T. Bouwmans, and E. H. Zahzah, "Comparison of matrix completion algorithms for background initialization in videos", in Proc. Int. Conf. Image Anal. Process., Vol. 9281, 2015, pp. 510-518.

[20]. W. Li, Z. Zhang, Z. Liu, "Action recognition based on a bag of 3D points". In Proc., of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA, pp. 9–14, 2010

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731