International Journal of Engineering

Journal Homepage: www.ije.ir

Modal Frequencies Based Human Action Recognition Using Silhouettes and Simplicial Elements

O. Mishra[a], P. S. Kavimandan*[b], R. Kapoor[a]

[a] *Delhi Technological University, Department of Electronics & Communication, Bawana Road, Delhi, India*
[b] *IGDTUW, Department of Information Technology, Kashmere Gate, Delhi, India*

P A P E R I N F O

Paper history:
Received 30 June 2021
Received in revised form 25 September 2021
Accepted 16 October 2021

Keywords:
Finite Element Analysis
Simplicial Element
Displacement Matrix
Modal Frequenc
Support Vector Machine

A B S T R A C T

Human action recognition has been a pioneer research problem among researchers. This paper proposed a new local feature descriptor in terms of modal frequency using silhouette and simplicial elements of a silhouette with the help of Finite Element Analysis (FEA). This local descriptor represents the distinctive human poses in the form of modal frequency. These modal frequencies reduce the feature dimension and represent a wide range of poses of human action. These modal frequencies are subject to the stiffness matrix of the body that is associated with the displacement. The silhouettes of the human body are used for the analysis. These silhouettes are represented into simplicial elements. The modal frequencies of silhouettes are calculated using simplicial elements. These modal frequencies of the silhouette are used as the feature vectors that are given to the Radial Basis Function-Support Vector Machine (RBF-SVM) classifier. The challenging datasets Weizmann, KTH and IXMAS are used for validation of the proposed methodology.

doi: *10.5829/ije.2022.35.01a.05*

1. INTRODUCTION

Videos have become a very essential part of our life these days. We create, store and share these videos. The increasing number of accessible videos has also created the need to understand them. Several methods have been developed by researchers for human action recognition. These methodologies can be sectioned into two groups: global feature descriptors and local feature descriptors. Global features require the extraction of the human body whose action is to be recognized. The two-dimensional (2D) template matching technique is used [1-3]. In this technique, 2D silhouettes of the human body are extracted. These silhouettes are also used as space-time volumes [4, 5]. The main disadvantage of this methodology is that they require accurate background subtraction and motion of the pixels. Further global feature descriptors have major disadvantages that they give shape information but not motion information. That makes it weak in recognizing similar types of actions such as running and jogging where motion is involved. To avoid these problems researchers developed local feature descriptors. This methodology does not require background subtraction as they are established on the spatio-temporal points. Methods reported in literature [6-14] have used famous bag-of-words models. The main disadvantage of these methodologies is that they give only motion information but no information about the structure. Methodologies reported in literature [15-22] based on silhouette analysis have a major contribution to human action recognition. Kapoor et al. [15] have used average energy silhouette images whereas a hybrid classifier is used for action recognition by Mishra et al. [16]. Wu and Shao [17] used hybrid features which include both global and local features. The pose correlogram is used as a local feature and extended MHI is used as global feature. The three-dimensional histogram of the oriented gradient is used to represent the action video [18, 19]. Silhouette-based methodologies are also used in a human pose-based action recognition [23, 24].

The silhouette-based analysis is also used in deep learning-based methodologies [25-27] but these

*Corresponding Author Institutional Email: pronita012phd0115@igdtuw.ac.in (P. S. Kavimandan)

Please cite this article as: O. Mishra, P. S. Kavimandan, R. Kapoor, Modal Frequencies Based Human Action Recognition Using Silhouettes and Simplicial Elements, ***International Journal of Engineering, Transactions A: Basics***, Vol. 35, No. 01, (2022) 45-52

methodologies require large datasets. The methods based on global feature descriptors cannot give the motion information and the methodologies that are based on local descriptors cannot give shape information. Even methodology that used hybrid descriptors having both global and local descriptors require fusion problems. A feature descriptor based on the stiffness matrix discussed in literature [21] is capable of representing both shape as well as motion features. The main limitation of this methodology was that it increases the feature dimensionality and also requires large memory space. The proposed method overcomes these limitations and offers a new local feature descriptor embedding the information of both shape and motion with the help of modal frequency. Modal frequency plays an important role in the analysis of the shape of the structure against the deformation that occurred in the shape. The proposed methodology improved these issues in the following manner:

1. Modal frequencies of the human body silhouette reduce the feature dimensionality drastically when the reduced number of the mode of the frequency of the structure is selected. And thus, it takes less runtime as compared to literature [21].
2. Moreover different modes of frequencies of an action shape cover the wide range of the poses of human action. This increases the accuracy as compared to our previous work.
3. Every deformation in the shape has its unique modal frequency that represents the shape change because of motion. Therefore modal frequency is capable of representing shape and motion information. This makes the proposed method unique and more reliable as compared to other existing methods because it contains structural as well as motion information.

2. METHODOLOGY OF PROPOSED FRAMEWORK

The workflow diagram of the proposed methodology is shown in Figure 1. The first step for action recognition is to extract the silhouette from the action video. Then these silhouettes are represented in the form of simplicial elements using FEA. The displacement matrices of these simplicial elements are then found. The modal frequencies are extracted with the use of displacement matrices. Later on, feature vectors are formed in form of modal frequencies that are fed to the RBF-SVM classifier to recognize the action.

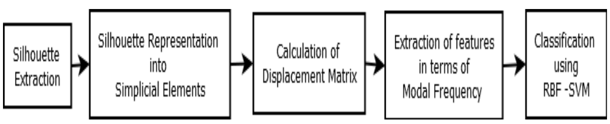

Figure 1. Workflow diagram of the proposed method

2. 1. Representation of Silhouette in Terms of Simplicial Elements

The silhouette extraction from the action video is a crucial step. Limitations such as background cluttering create a hurdle for background subtraction. The proposed methodology used the Gaussian mixture model-based background subtraction to extract the silhouette used in literature [28]. The advantage of this method is that it can deal with the problem occurring because of dynamic background and shadow. Figure 2 shows the frame of the action 'walk' and the background-subtracted image. The boundary of the silhouette is then found.

These Silhouettes are divided into different simplicial triangular elements using the Finite Element Method (Figure 3). We adopted the following steps for the representation of the simplicial elements: 1. The prominent points on the silhouette have been reported by Laptev and Lindeberg [29]. 2. These prominent points are used as the node vertices of the simplicial triangular elements.

When an actor performs actions, the prominent points (vertices of the simplicial elements) also get displaced in a unique pattern. The simplicial elements are represented by displacement matrices. The simplicial triangular element has three nodes $A(x_1,y_1)$, $B(x_2,y_2)$ and $C(x_3,y_3)$. The displacement vectors of the simplicial triangular elements given by $D = \{d_1, d_2, d_3, d_4, d_5, d_6\}^T$. Every node has a displacement in X-direction and Y-direction. Node A has displacement in X-direction is d_1 and in Y-direction is d_2, node B has displacement in X-direction is d_3 and in Y-direction is d_4 and for node C displacement in X-direction is d_5 and in Y-direction is d_6. The complete object with n elements can be represented globally by $G = \{D_1, D_2, D_3, \ldots, D_n\}^T$. Every movement of the silhouette results in the displacements of the vertices of the triangle. Figure 5 shows the small element having displacements a in X-direction and b in Y-direction.

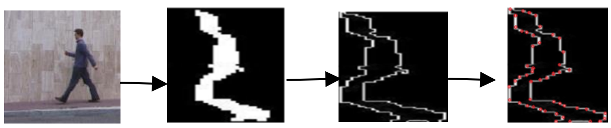

Figure 2. A frame of 'walk', background-subtracted image, external silhouette and the prominent points on the extracted silhouette

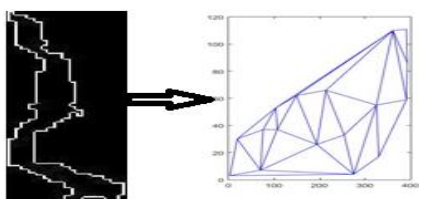

Figure 3. Representation of silhouette into simplicial elements

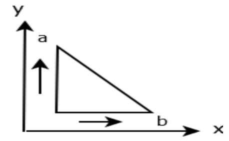

Figure 4. Displacement of simplicial element

Due to this displacement, a strain $\phi = \begin{bmatrix} \frac{\partial a}{\partial x} \\ \frac{\partial b}{\partial y} \\ \frac{\partial a}{\partial y} + \frac{\partial b}{\partial x} \end{bmatrix}$ is developed in X-direction, Y-direction and shear direction.

2. 2. Feature Extraction in Terms of Modal Frequency We can deduce the stiffness matrix of the single triangle with the use of displacement matrices. We call this stiffness matrix as a local stiffness matrix and it can be represented by Equation (1).

$$\mathbf{k_l} = \mathbf{D^T} \mathrm{D} \mathbf{t_e} \mathbf{S_e} \tag{1}$$

where D is displacement matrix subject to strain, t_e is the thickness and S_e is the area of the triangle and both are assumed constant. D is also called the strain displacement matrix which can be calculated by shape functions [21]. All the local stiffness matrices of triangle elements are united systematically to form the global stiffness matrix. We first set the degree of freedom of the structure. The value of the degree of freedom will start from one to D where D is the total degree of freedom. Following steps are involved in the formation of the global stiffness matrix:

The size of the global stiffness matrix K_g will be the total number of degree of freedom and it is initialized to zero.

We computed the local stiffness matrix k_l for every triangular element.

We added the local stiffness matrix k_l to the global stiffness matrix K_g by placing it properly.

Procedures 2 and 3 are repeated until the complete global stiffness matrix is constructed.

Once we get the global stiffness matrix we can calculate the modal frequency of the complete silhouette by calculating the Eigen values of the global stiffness matrix. The Eigen value of matrix K_g can be calculated by Equation (2).

$$|\mathbf{det}[\mathbf{K_g} - \boldsymbol{\lambda_i} \mathbf{I}]| = \mathbf{0} \tag{2}$$

where λ_i is Eigenvalue and I is an identity matrix.

The values of the λ_i give the information the how much variance in global stiffness matrix in their directions. The highest value of Eigen value will be the principal component. In the proposed methodology, we selected three principal components.

The corresponding modal frequency can be found in Equation (3).

$$\mathrm{F} = \sqrt{\lambda_i}/(2\pi) \tag{3}$$

With the help of the Eigen values, modal frequencies of the silhouette of the human body are extracted by Equation (3). These principal component based modal frequencies reduce the feature dimensionality drastically. We have selected the reduced number of the mode of the frequency of the structure i.e.3. The result of three modes of frequencies (mode 1, 16.61x10^{-5}Hz, mode 2, 28.58x10^{-5}Hz and mode 3, 36.80x10^{-5}Hz) applied on Weizmann datasets for action "Clapping" are given in Figure 5.

The silhouette extracted from the sequence of frames in a video is discretized into a number of simplicial elements. The global stiffness matrix that represents the local feature of the frame has the dimension depending upon the degree of freedom. For example, if the total number of nodes/vertices of a silhouette is n and if these nodes can move in X and Y directions, the degree of freedom of complete silhouette structure will be 2n. Thus, the dimension of the global stiffness matrix will be 2n x 2n. Since we are using Eigen values, the dimension of this structure will reduce to 2n. Thus, corresponding to the dimension of 2n, we will get 2n modal frequencies, out of which we have selected the most suitable number through cross-validation. These feature vectors are fed to the RBF-SVM classifier [21, 30] to recognize the action.

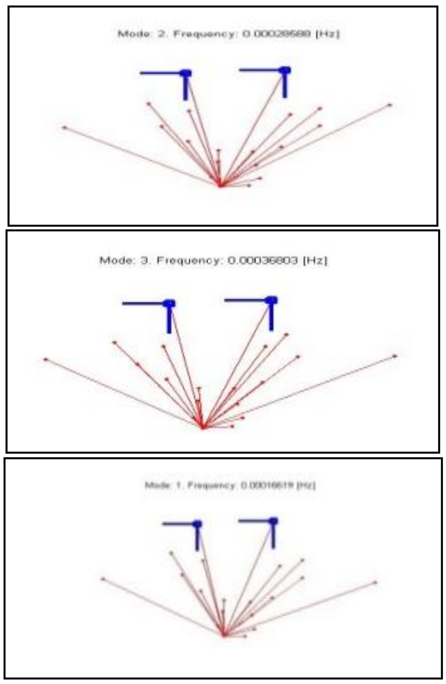

Figure 5. Three modal frequencies for action 'Clapping' of Weizmann dataset

3. EXPERIMENTAL RESULTS

To establish the authenticity of the proposed methodology, standard datasets like Weizmann, KTH and IXMAS are used. The experimentation on these datasets will also prove the robustness of the proposed methodology against background cluttering, execution rate and both inter and intra-class similarity. The proposed methodology has been developed on the following set-up:

Software: MATLAB R2015A

Hardware: Intel(R) Core (TM) i5-6200 CPU @2.30 GHz, 8 GB RAM, 64 bit Operating System

Accuracy is used as a performance evaluation parameter of the proposed methodology. For cross-validation, we used the Leave-One-Out strategy. The description of datasets is as follows:

Weizmann Dataset has 10 action classes; KTH Dataset has 6 action classes whereas IXMAS has 13 action classes. Sample frames of all three datasets are shown in Figure 6(a-c). Table 1 shows that when the total number of simplicial triangle elements in the silhouette is considered to be more than 20, it gives a superior result.

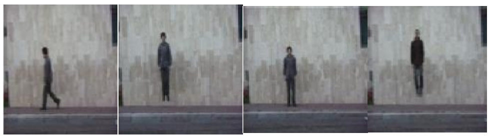

Figure 6(a). Weizmann dataset frames

Figure 6(b). KTH dataset frames

Figure 6(c). IXMAS dataset frames

TABLE 1. Parameter setting for no. of simplicial elements

Total no. of nodes	5	10	15	20	22	25
Accuracy (%)	0.67	0.83	0.89	0.93	0.94	0.94

TABLE 2. Parameter setting for no. of modes of frequency

No. of modes of frequency	1	2	3	5	7
Accuracy (%)	0.70	0.88	0.94	0.94	0.95

Moreover, any number higher than 22 do not yield significantly different results. Thus, we have selected 22 elements. We selected the reduced number of modes of frequency so that the feature vector dimension also gets reduced. Table 2 shows the result of different modes of frequency. It is clear from table 2 that three modes of frequencies gives good result and further increment in the number of modes do not show significant change. Thus, we have selected three modes of frequency for our proposed method. The graphs of modal frequencies versus the silhouettes in the frames are shown in Figure 7(a-f) where the X-axis represents the frames and the Y-axis represents a change in modal frequencies with the change in frames for different activities such as 'clapping', 'jumping', 'hand waving', 'hand waving (both hands)', 'skipping' and 'hopping'. Figure 7(a) shows 3 modes of frequency changes for the action 'clapping'.

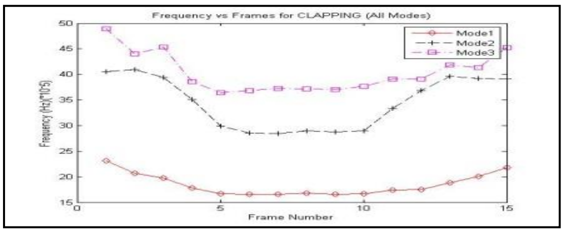

Figure 7(a). Modal Frequency for 'clapping'

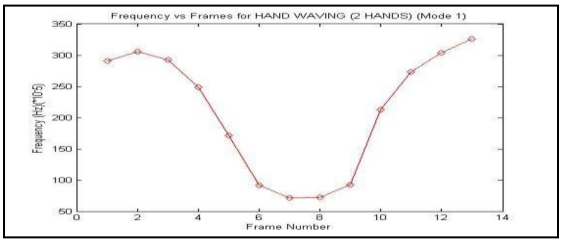

Figure 7(b). Modal Frequency for 'hand waving (both hands)'

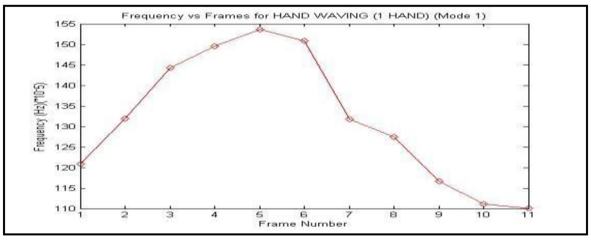

Figure 7(c). Modal Frequency for 'hand waving'

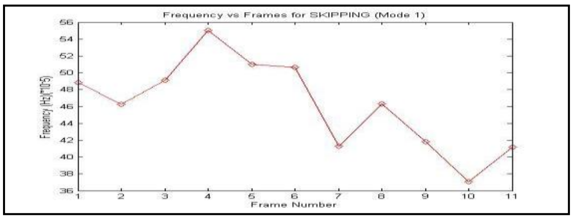

Figure 7(d). Modal Frequency for 'skipping'

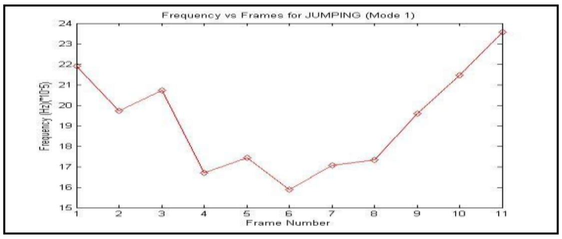

Figure 7(e). Modal Frequency for 'jumping'

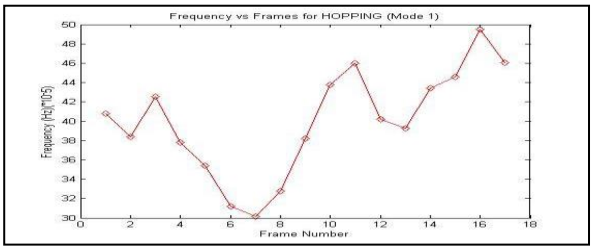

Figure 7(f). Modal Frequency for 'hopping'

3. 1. Comparison of the Proposed Algorithm with Different Methods on Standard Datasets

Weizmann dataset: Similar methodologies have been compared with the proposed methodology for the Weizmann dataset [12, 16, 17, 25, 26, 31] and the results are shown in Figure 8(a). The proposed method achieved an accuracy of 97.9%. All the methods mentioned in Figure 8(a) either retained shape information or motion information. The reason we achieved higher accuracy is that the proposed methodology has retained both shape and motion information through the modal frequencies of the silhouette structure.

KTH dataset: Figure 8(b) shows the comparison of the proposed methodology with other recent methodologies on KTH dataset [3, 12, 16, 18, 24, 25, 27, 31]. The proposed method achieved an accuracy of 96.2% for the KTH dataset.

IXMAS dataset: Figure 8(c) shows a comparison of the proposed methodology with other methodologies for IXMAS dataset [11, 16, 33]. The accuracy of the proposed method is 89.7% for IXMAS.

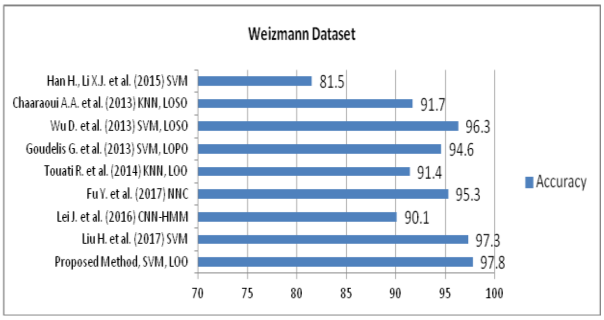

Figure 8(a). Comparison of the proposed methodology with other methods for Weizmann Dataset

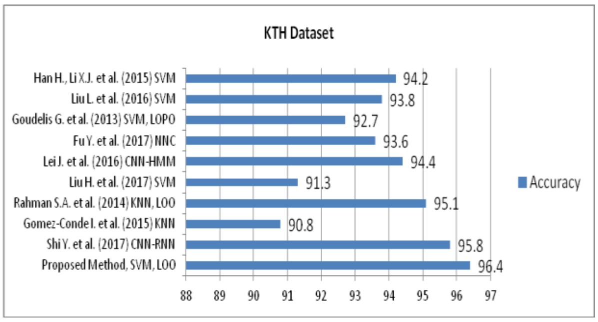

Figure 8(b). Comparison of the proposed methodology with other methods for KTH Dataset

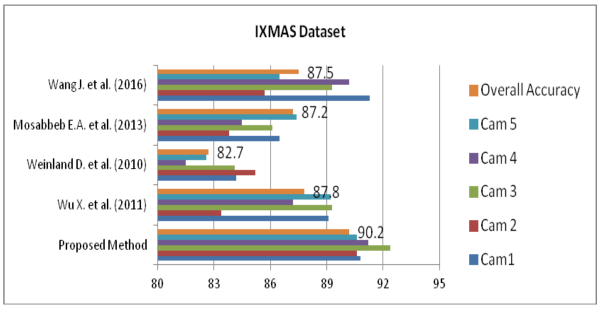

Figure 8(c). Comparison of the proposed methodology with other methods for IXMAS Data

Figure 9(a-c) show the confusion matrices of the proposed methodology applied on the Weizmann datasets, KTH dataset, Ballet dataset and IXMAS dataset, respectively. For Weizmann dataset, A1: running action is 6% confused with the action A2: walking, whereas walking action is recognized without any confusion. Similarly, A5: skipping action is 3% confused with A6: Jumping at a place whereas, action jumping at a place is 1% confused with skipping. All other actions are recognized correctly. For KTH dataset action A3: boxing is 4% confused with action A2: waving and actions A5: jogging and A6: running is 3% confused with each other. Actions such as applauding, waving and walking are correctly classified. Similarly, the confusion matrix on IXMAS datasets also shows good classification results.

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	0.94	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A_2	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A_3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A_4	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
A_5	0.00	0.00	0.00	0.00	0.97	0.03	0.00	0.00	0.00	0.00
A_6	0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.00	0.00	0.00
A_7	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
A_8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
A_9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
A_{10}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Figure 9(a). Confusion matrix for Weizmann Dataset (A1-Running, A2-Walking, A3-Jumping, A4-Jumping Jack, A5-Skipping, A6-Jumping at a place, A7-Side Jumping, A8-Bending, A9-Waving with one hand, A10-Waving with both hands)

	A_1	A_2	A_3	A_4	A_5	A_6
A_1	1.00	0.00	0.00	0.00	0.00	0.00
A_2	0.00	1.00	0.00	0.00	0.00	0.00
A_3	0.00	0.04	0.96	0.00	0.00	0.00
A_4	0.00	0.00	0.00	1.00	0.00	0.00
A_5	0.00	0.00	0.00	0.00	0.97	0.03
A_6	0.00	0.00	0.00	0.00	0.03	0.97

Figure 9(b). Confusion matrix for KTH Dataset (A1- Applauding, A2- Waving, A3- Boxing, A4- Walking, A5- Jogging, A6- Running)

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}
A_1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A_2	0.00	0.93	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A_3	0.00	0.00	0.96	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A_4	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A_5	0.00	0.00	0.03	0.00	0.95	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A_6	0.00	0.00	0.00	0.00	0.02	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A_7	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.04	0.00	0.00	0.00	0.00	0.00
A_8	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.91	0.00	0.00	0.00	0.00	0.00
A_9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
A_{10}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
A_{11}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	0.02	0.00
A_{12}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.95	0.00
A_{13}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Figure 9(c). Confusion matrix for IXMAS Dataset (A1- Walking, A2- Waving, A3- Punching, A4- Kicking, A5- Throwing, A6- Pointing, A7- Picking Up, A8- Getting Up, A9- Sitting Down, A10-Turning Around, A11-Folding arms, A12-Checking Watch, A13-Scratching Head)

In the proposed methodology, different modes of frequencies of an action shape cover the wide range of the pose of human action. It increases the accuracy also as compared to literature [21]. Table 3 shows the comparison between these two methodologies where the accuracy of the both is comparable for Weizmann and KTH dataset but the accuracy of the proposed methodology (93.2%) is clearly superior to the IXMAS dataset where the variation in the action pose is large.

To analyze the runtime of both methodologies, NVIDIA GPU with Parallel computing Toolbox is used. The total time taken by Kapoor et al. [21] is 2.92 s, whereas the proposed methodology having 3 modes of frequency of silhouette pose takes considerably less time 1.78 s.

TABLE 3. Comparison of the proposed methodology with similar methodologies

Datasets	Methodology applied by Kapoor et al. [15]	Proposed Method
Weizmann	97.8	98.1
KTH	96.4	97.2
IXMAS	90.2	93.2

4. CONCLUSION

This is a new method to recognize human action through finite element analysis. Local features are expressed in terms of the modal frequency of the action silhouette. This offers the uniqueness of this method as it can extract both shapes as well as motion information. It overcomes the drawback of other existing state-of-the-art methods based on local features, since they are not capable of extracting both shape and motion information together. Validation of the proposed method has been performed on different datasets of challenging environments. The proposed method demonstrates its supremacy over other existing methods for both less complex datasets like Weizmann and KTH as well as complex datasets like IXMAS. The feature descriptor used in the proposed method has demonstrated very good results, but the limitation of the proposed method is that it requires a sophisticated silhouette extraction technique.

5. REFERENCES

1. Bobick, F., Davis.J. W. "The recognition of human movement using temporal templates." ***IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI),*** Vol. 33, No. 6, (2001), 257-267[doi: 10.1109/34.910878].
2. SouvenirR., Babbs J. "Learning the viewpoint manifold for action recognition."IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'08),(2008), 1-7, doi: 10.1109/CVPR.2008.4587552.
3. Rahman S.A., Song I., Leung M.H.K., Lee I., Lee K. "Fast action recognition using negative space features." ***Expert Systems Applications,*** Vol. 41. No. 2, (2014), 574-587, https://doi.org/10.1016/j.eswa.2013.07.082
4. Gorelick L., Blank M., Shechtman E., Irani M., Basri R. "Action as space-time shapes." ***IEEE Transaction on Pattern Analysis and Machine Intelligence,*** Vol. 29, No. 12, (2017), 2247-2253, doi:10.1109/TPAMI.2007.70711
5. Grundmann M, Meier F., Essa I. "3D shape context and distance transform for action recognition."19thInternational Conference on Pattern Recognition (ICPR'08), Tampa, FL, (2008), 1-4[doi:10.1109/ICPR.2008.4761435]
6. Laptev I., Marszalek M., Schmid C., Rozenfeld B. "Learning realistic human actions from movies." IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, 1-8, doi: 10.1109/CVPR.2008.4587756
7. Wang Y., Mori G. "Human action recognition using semi-latent topic models." ***IEEE Transactions on Pattern Analysis and Machine Intelligence,*** Vol. 31, No. 10, (2009), 1762-1764, doi: 10.1109/TPAMI.2009.43
8. Wu X., Xu D., Duan L.,Luo J. "Action recognition using context and appearance distribution features." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, 2011, 489-496, doi:10.1109/CVPR.2011.5995624
9. Iosifidis A., Tefas A., Pitas I. "Discriminant bag of wordsbased representation for human action recognition." ***Pattern Recognition Letters,*** Vol. 49, No. 1, (2014), 185-192, https://doi.org/10.1016/j.patrec.2014.07.011

10. Liu L., Shao L., Li X., Lu K. "Learning spatio-temporal representations for action recognition: A genetic programming approach." ***IEEE Transactions on Cybernetics,*** Vol. 46, No. 1, (2016), 158-170, doi: 10.1109/TCYB.2015.2399172
11. Wang J., Zheng H., Gao J., Cen J. "Cross-view action recognition based on a statistical translation framework." ***IEEE Transactions on Circuits and Systems for Video Technology,*** Vol. 26, No. 8, (2016), 1461-1475, doi: 10.1109/TCSVT.2014.2382984.
12. Fu Y., Zhang T., Wang W. "Sparse coding-based space-time video representation for action recognition." ***Multimedia Tools and Applications,*** Vol. 76, No. 10, (2017), 12645-12658 https://doi.org/10.1007/s11042-016-3630-9
13. Gomez-Conde I., Olivieri D.N. A KPCA spatio-temporal differential geometric trajectory cloud classifier for recognizing human actions in a CBVR system," ***Expert Systems Applications,*** Vol. 42, No. 13, (2015), 5472-5490 https://doi.org/10.1016/j.eswa.2015.03.010.
14. Mishra O., Kapoor R., Tripathi M.M. "Human Action Recognition Using Modified Bag of Visual Word based on Spectral Perception." ***International Journal of Image, Graphics and Signal Processing,*** Vol. 11, No. 9, (2019), 34-43 https://doi.org/10.5815/ijigsp.2019.09.04.
15. Kapoor, R., Mishra, O. & Tripathi, M.M. "Anomaly detection in group activities based on fuzzy lattices using Schrödinger equation." ***Iran Journal of Computer Science***, Vol. 3, No. 2, (2020), 103-114. https://doi.org/10.1007/s42044-019-00045-y.
16. Mishra O., Kavimandan P.S., Tripathi M.M., Kapoor R., Yadav K. "Human Action Recognition Using a New Hybrid Descriptor."In: Harvey D., Kar H., Verma S., Bhadauria V. (eds) Advances in VLSI, Communication, and Signal Processing. Lecture Notes in Electrical Engineering, Vol 683, (2021) Springer, Singapore. https://doi.org/10.1007/978-981-15-6840-4_43
17. Wu D., Shao L. "Silhouette analysis-based action recognition via exploiting human poses." ***IEEE Transactions on Circuits and Systems for Video Technology,*** Vol. 23, No. 2, (2013), 236-243, doi: 10.1109/TCSVT.2012.2203731.
18. Touati R., Mignotte M. "MDS-based multi-axial dimensionality reduction model for human action recognition."Canadian Conference on Computer and Robot Vision, 2014, 262-267, doi: 10.1109/CRV.2014.42.
19. Weinland D., Özuysal M., Fua P. "Making action recognition robust to occlusions and viewpoint changes"In: Daniilidis K., Maragos P., Paragios N. (eds) Computer Vision - ECCV 2010, Lecture Notes in Computer Science, (2010), 6313, 635-648 https://doi.org/10.1007/978-3-642-15558-1_46.
20. Xia L.M., Huang J.X., Tan L.Z. "Human action recognition based on chaotic invariants", ***Journal of Central University,*** Vol. 20, No. 11, (2013), 3171-3179, https://doi.org/10.1007/s11771-013-1841-z
21. Kapoor R., Mishra O., Tripathi M.M. "Human action recognition using descriptor based on selective finite element analysis.", ***Journal of Electrical Engineering,*** Vol. 70, No. 6, 2019, 443-453, doi: https://doi.org/10.2478/jee-2019-0077.
22. Kavimandan, P.S., Kapoor R., Yadav K."Human Action Recognition using Prominent Camera," ***International Journal of Engineering, B: Applications,*** Vol. 34, No. 02, (2021), 427-432. doi: https://dx.doi.org/10.5829/ije.2021.34.02b.14
23. Chaaraoui A.A., Pérez P.C., Florez-Revuelta F."Silhouette-based human action recognition using sequences of key poses," ***Pattern Recognition Letters,*** Vol. 34, No. 15, (2013), 1799-1807. https://doi.org/10.1016/j.patrec.2013.01.021.
24. Goudelis G., Karpouzis K., Kollias S. "Exploring trace transform for robust human action recognition." ***Pattern Recognition,*** Vol. 46, No. 12, (2013), 3238-3248 https://doi.org/10.1016/j.patcog.2013.06.006.
25. Lei J., Li G., Zhang J.,Guo Q., Tu D. "Continuous action segmentation and recognition using hybrid convolutional neural network-hidden Markov model." ***IET Computer Vision***, Vol. 10, No. 6, (2016), 537-544, http://dx.doi.org/10.1049/iet-cvi.2015.0408.
26. Liu H., Shu N., Tang Q., Zhang W. "Computational model based on the neural network of visual cortex for human action recognition." ***IEEE Transactions on Neural Networks and Learning Systems,*** Vol. 29, No. 5, (2018), 1427-1440, doi: 10.1109/TNNLS.2017.2669522.
27. Shi Y., Tian Y.,Wang Y., Huang T. "Sequential deep trajectory descriptor for action recognition with three-stream CNN." ***IEEE Transactions on Multimedia,*** Vol. 19, No. 7, (2017), 1510-1520, doi: 10.1109/TMM.2017.2666540.
28. Dou J., Li J. "Robust human action recognition based on spatiotemporal descriptors and motion temporal templates." ***Optik***, Vol. 125, No. 7, (2014), 1891-1896, https://doi.org/10.1016/j.ijleo.2013.10.022.
29. Laptev I., Lindeberg T. "Space-time interest points." Proceedings Ninth IEEE International Conference on Computer Vision, Nice, France, (2003), 432-439, doi: 10.1109/ICCV.2003.1238378.
30. Chittora A., Mishra O., "Face Recognition Using RBF Kernel Based Support Vector Machine," ***International Journal of Future Computer and Communication,*** Vol. 1, No. 3, (2012), 280-283, doi: 10.7763/IJFCC.2012.V1.75
31. Han H., Li X.J. "Human action recognition with sparse geometric features."***The Imaging Science Journal,*** Vol. 63, No. 1, (2015), 45-53, doi: 10.1179/1743131X14Y.0000000091.
32. Liu L., Shao L., Li X. "Learning spatio-temporal representations for action recognition: A genetic programming approach." ***IEEE Transactions on Cybernetics,*** Vol. 46, No. 1, (2016), 158-170 [https://doi.org/10.1016/B978-0-12-818597-1.50072-2].
33. Mosabbeb E.A., RaahemifarK., Fathy M. "Multi-view human activity recognition in distributed camera sensor networks." ***Sensors,*** Vol. 13, No. 7, (2013), 8750-8770. https://dx.doi.org/10.3390%2Fs130708750
34. Hosseini M.S.,Ghaderi F., "A hybrid deep learning Architecture Using 3-D CNNs and CRUs for Human Action Recognition." ***International Journal of Engineering, B: Applications, Transactions B: Applications,*** Vol. 33, No. 5, (2020), 959-965. doi: 10.5829/IJE.2020.33.05B.29

Persian Abstract

چکیده

به رسمیت شناختن عملکرد انسان یک مشکل تحقیقاتی پیشگام در بین محققان بوده است. این مقاله یک توصیف کننده ویژه بومی جدید از نظر فرکانس مد با استفاده ازسایه و عناصر ساده یک شبه با کمک تجزیه و تحلیل عناصر محدود ارائه کرده است. این توصیف کننده بومی نشان دهنده حالت های متمایز انسان در قالب فرکانس مودال است. این فرکانس های مد بعد ویژگی را کاهش می دهد و طیف وسیعی از حالت های عملکرد انسان را نشان می دهد. این فرکانس های مودال تابع ماتریس سختی بدن است که با جابجایی همراه است. برای تجزیه و تحلیل از سایه های بدن انسان استفاده می شود. این سایه ها در عناصر ساده نشان داده شده اند. فرکانس مودال سایه ها با استفاده از عناصر ساده محاسبه می شود. این فرکانس های مد سایه به عنوان بردارهای ویژگی استفاده می شود که به طبقه بندی کننده شعاع تابع پشتیبانی ماشین بردار (RBF-SVM) داده می شود. مجموعه های چالش برانگیز Weizmann، KTH و IXMAS برای اعتباربخشی روش پیشنهادی استفاده می شوند.