International Journal of Engineering & Technology, 7 (1.1) (2018) 500-506

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Indian classical dance action identification using adaptive graph matching from unconstrained videos

K.V.V. Kumar[1], P.V.V. Kishore[1]*

[1]Biomechanics and Vision Computing Research Centre, Department of Electronics and Communication Engineering,
KLEF Deemed-to-be-University, Andhra Pradesh, India
**Corresponding author E-mail: pvvkishore@kluniversity.in*

Abstract

Extracting and recognizing complex human movements from unconstraint online video sequence is a challenging task. In this work the problem becomes complicated by the use of unconstraint video sequences belonging to Indian classical dance forms. A new segmentation model is developed using discrete wavelet transform and local binary pattern features for segmentation. We also explore multiple feature fusion models with early fusion and late fusion techniques for improving the classification process. The extracted features were represented as a graph and a novel adaptive graph matching algorithm is proposed. We test the algorithms on online dance videos and on an Indian classical dance dataset prepared in our lab. The algorithms were tested for accuracy and correctness in identifying the dance postures.

Keywords: *Indian Classical Dance Identification, Adaptive Graph Matching, Feature Fusion, Histogram of Oriented Features (HOG), Discrete Wavelet Transform (DWT), Local Binary Patterns (LBP).*

1. Introduction

Automatic human action recognition is a complicated problem for computer vision scientists, which involves mining and categorizing spatial patterns of human poses in videos. Human action is defined as a temporal variation of human body in a video sequence, which can be any action such as dance, running, jumping or simply walking. Automation encompasses mining the video sequences with computer vision algorithms for identifying similarities between actions. Last decade has seen a jump in online video creation and the need for algorithms that can search within the video sequence for a specific human pose or object of interest. The problem is to extract, identify a human pose and classify into labels based on trained human signature action models [1]. The objective of this work is to extract the signature of Indian classical dance poses from both online and offline videos given a specific dance pose sequence as input.

However, the constraints are video resolution, frame rate, background lighting, scene change rate and blurring to name a few. The analysis on online content is a complicated process as the most of the users end up uploading the videos with poor quality, which shows all the constraints as a hindrance in automation of video object segmentation and classification. Dance video sequences online are having a far many constraints for smooth extraction of human dance signatures. Automatic dance motion extraction is complicated due to complex poses and actions performed at different speeds in sink to music or vocal sounds. Fig. 1 shows a set of online and offline (lab captured) Indian classical dance videos for testing the proposed algorithm.

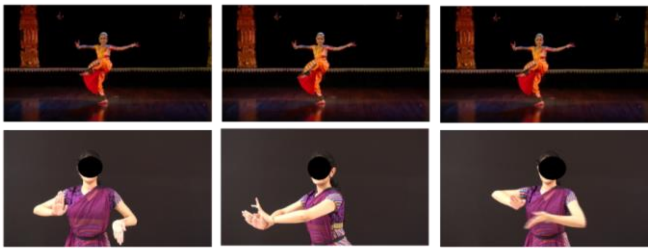


Fig. 1. Sample Online and Offline Dance datasets used in this work.

Indian classical dance forms are a set of complex body signatures produced from rotation, bending and twisting of fingers, hands and body along with their motion trajectory and spatial location. There are 8 different classical Indian dance forms; Bharatanatyam, Kathakali, Kathak, Kuchipudi, Odissi, Sattriya, Manipuri and Mohiniyattam [2]. Extracting these complex movements from online videos and classification requires a complex set of algorithms working in sequence. We propose to use silhouette detection and background elimination, human object extraction, local texture with shape reference model and 2D point cloud to represent the dancer pose as a graph. For recognition, an adaptive graph matching algorithm is proposed to classify query dance video based on the dance dataset.

The rest of the paper is organized into literature survey on the proposed techniques, theoretical background on the proposed models and experimental results. The proposed model is compared with SVM classifier already proposed by us in our previous work.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2. Literature Survey

Local information of the human in the video are the popular features for action segmentation and classification in recent times. This section focus on giving a current trend in human action recognition and how it is used in recent works for classifying dance performances. The human action recognition is subdivided into video object extraction, feature representation and pattern classification [3]. Based on these models, numerous visual illustrations have been proposed for discriminating human action based on shape templates in space – time [3], shape matching, interest points in 2D space time models [4] and representations using motion trajectories [5]. Impressively, dense trajectory based methods [6] have shown good results for action recognition by tracking sampled points through optical flow fields. Optical flow fields are based on pre-conditioned on brightness and object motion in a video [7].

In this work, human action recognition on Indian Classical Dance [8] videos is performed on recordings from both offline (controlled recording) and online (Live Performances, YouTube) data. Indian classical dance forms are practised from 5000 years worldwide. However, it is difficult for a dance lover to fully hold the content of the performance as it is made up of hand poses, body poses, leg movements, hands with respect to face and torso and finally facial expressions. All these movements should synchronize in precision with both vocal song and the corresponding music for various instruments. Aparna et al [9], highlights the difficulties in using state of the art pose estimation algorithms such as skeleton estimation [10] and pose estimation [11] fail to track the dancers moves in both offline and online videos. Samanta et al [12] used histogram of oriented optical flow (HOOF) features with sparse representations. Support vector machine classifier (SVM) classifies the Indian classical dance poses from KTH dataset with an accuracy of 86.67%. In our previous work [13], we approached the same problem with SVM classifier on dance videos and found that only multi class SVM's should be considered. Moreover, optical flow on online videos suffer at lot due to inconsistencies during capture and sharing process.

We propose to use adaptive graph matching (AGM) in this work to handle pose classification and human action recognition in Indian classical dance videos. This work on Indian classical dance pose recognition is based on graph matching (GM) in [14]- [16]. GM is attractive to solve pattern matching problems due to their invariance properties such as scaling, rotation and transformation in mathematical sense. In Adaptive Graph Matching (AGM), the graphs change constantly between consecutive video frames.

Classical GM methods used in [17] [18] and current research in GM [19] use tree search integrating heuristic estimation models to trim the search space and computation time. All these graph models use a fixed feature set to represent the objects in the image or video. With the feature set, the algorithms guaranteed to find an optimal solution are NP – hard, which are computationally intensive for large feature points. There are [20] [21] methods that are suboptimal and modelled as polynomial in nature resulting in non-optimal solutions. However, the approximate GM models suffer from struck at local minima and miss to approach an optimal solution.

This problem can be overcome by using weighted graph matching using eigen and linear programming approaches [22]. All these GM algorithms do not consider the structure of the graphs being matched. They start with an assumption that the matching graphs have equal number of points to be matched. In most instanced the points representing the graph are marked manually. The state of the art graph cuts used exceptionally and efficiently in complex image segmentation problems [23].

In this paper, we propose an Adaptive Graph Matching (AGM) based on localized multi point minimum distance metrics computation problem. The proposed AGM can effectively recover the query video frames from the dance dataset. The proposed method is compared with other GM models which are outperformed by a considerable margin.

3. Proposed Methodology

The proposed algorithm framework is shown in fig.2. An Indian Classical Dance (ICD) video library is created combining online and offline videos. Dancer identification, dancer extraction, local shape feature extraction and classifier are the modules of the system. Further feature fusion concept from [24] is also explored in this wok using Histogram of Oriented Gradients(HOG) features. HOG features are fused with local shape descriptor features and the fusion features are represented as graph. AGM algorithm explores the relativity between the query dance sequence and known dataset.

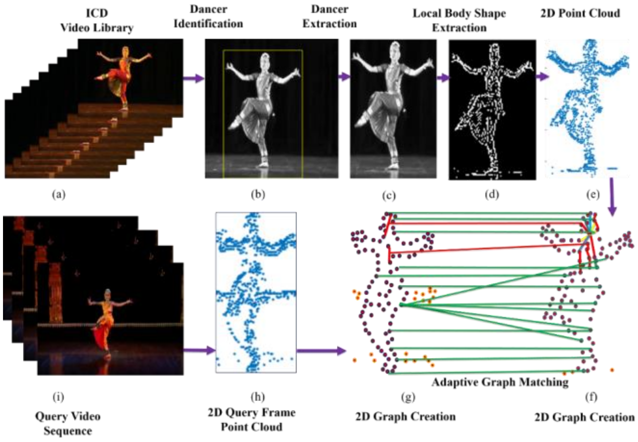


Fig. 2. Flow Diagram of the proposed Indian Classical Dance Recognition.

3.1. Dancer Identification

Most of the dance videos are poorly illuminated or fully brightened with too much background information during capture. Commercial video cameras have a frame rate of 30fps and dance movements are sometimes faster and at time slower which makes the object blurry. The objective is to extract moving dancer and segment it for further processing. This helps to prevent the algorithm form constantly upgrade the background information. The dancer identification module is based on one of the silhouette extraction methods proposed in [25]. To avoid background modelling and foreground extraction models, we propose to use the procedure shown in fig.2.

The dance video sequence $V(x,y,t) \subset \mathrm{R}^{+}$, with $(x,y) \subset \mathrm{Z}^{+}$ gives pixel location and $t \subset \mathrm{Z}^{+}$ is the frame number. Each frame in V is having RGB planes and is of size $N \times M \times 3$. This part of the module is only for motion segmentation and object extraction; color can be discarded. The frame V^{t} at t is mean filtered with mask defined by $m(x,y)$ with

$$V_m^t(x,y) = V^t(x,y) \otimes m(x,y) \tag{1}$$

The size of m is updated based on the frame size $N \times M$ for faster computations, where the object area is small compared to the background area. The $\otimes$ operator is linear convolution and the averaged frame is of same size as the input frame. The next step applies a Gaussian filter of μ mean and σ variance on the input frame V^{t}

$$V_g^t(x,y) = V^t(x,y) \otimes g(\mu,\sigma) \tag{2}$$

The size of the Gaussian mask is determined by the input video frame. Euclidian distance metric $S^{t}(x,y)$ between V_m^t and V_g^t gives the saliency map of the moving pixels in the frame

$$S^t(x,y) = \left\| V_g^t(x,y) - V_m^t(x,y) \right\|_2 \tag{3}$$

The second order normed distance map is shown in fig.3 which identifies the dancer's silhouette. However, to extract the dancer, a mask of this silhouette is used to determine the connected components in the object. Fig.3(d) shows the silhouette mask and connected component output is in fig.3(e).

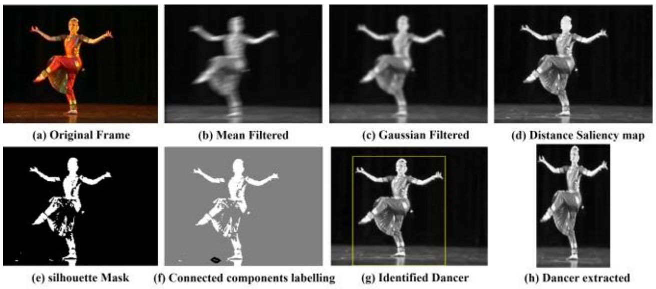

(a) Original Frame (b) Mean Filtered (c) Gaussian Filtered (d) Distance Saliency map (e) silhouette Mask (f) Connected components labelling (g) Identified Dancer (h) Dancer extracted

Fig. 3. Dancer Extraction Steps.

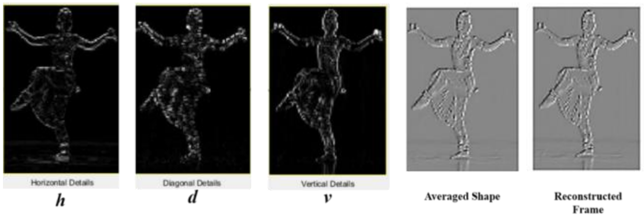
The centroid of the mask is mapped on the frame to crop out the moving dancer in the frame. The method is effective in all lighting conditions putting constraints on the input video frame size in selecting the masks used for mean and Gaussian filters. The boxed and extracted dancer from the video sequence is shown in fig.3. The extracted dancer is free from background variations in the video sequence. If a portion of background still appears at this stage can be nullified during the matching phase. Applying feature extraction on the extracted dancer allows for lesser computations as the background is almost eliminated.

3.2. Feature Extraction

From a dancer's perspective, to identify a dance type, body posture, hand shapes and their movements in space are the vital features. There are many shape descriptors available in literature for characterizing shape features [26]. Lighting, frame inconsistency, contrast, blurring and frame size are some of the critical factors that affect feature extraction algorithms.

3.2.1. Haar Wavelet Features – Global Shape Descriptor

For removing video frame noise during capture and to extract local shape information, we propose a hybrid algorithm with Discrete wavelet transform (DWT) [27] and Local Binary Patterns (LBP) [28]. The objective at this stage is to represent moving dancers shape with a set of wavelet coefficients. Here we propose to use Haar wavelet at level 1. At level 1, Haar wavelet decomposes the video frame V^t into 4 sub-bands. Fig.4. shows the 4 sub-bands at 2 levels. In the 1st level, the three sub-bands represent the shape information at three different orientations: Vertical v, Horizontal h and Diagonal d. Combining the three sub-bands and averaging the wavelet coefficients normalizes the large values.



h d v Averaged Shape Reconstructed Frame

Fig. 4. Harr Wavelet Sub-bands representing shape in three different orientations.

Fig.4. shows the reconstructed spatial domain frame producing the exact hand shapes. These shape features can be used as nodes and a graph can be constructed for recognition.

3.2.2. Thresholding

Apply threshold on the reconstructed ICD video frame V_r^t as

$$T^t = \sqrt{\frac{1}{NM}\sum_{j=1}^{M}\sum_{i=1}^{N}\left(V^t(j,i)\right)^2} \tag{4}$$

The binarized video frame B^t is

$$B^t = V_r^t > T^t \tag{5}$$

To extract the nodes for the graph, local pixel patterns provide exact shape representation.

3.2.3. Local Binary Patterns – Local Shape Models

LBP compares each pixel in a pre-defined neighbourhood to summarize the local structure of the image. For an image pixel $B^t(x,y) \in \Re^+$, where (x,y) gives the pixel position in the intensity image. The neighbourhoods of a pixel can vary from 3 pixels with radius $r=1$ or a neighbourhood of 12 pixels with $r=2.5$. The value of pixels using LBP for a centre pixel (x_c, y_c) is given by

$$L_S^t = LBP(x_c, y_c) = \sum_{j=1}^{P} B^t\left(g_p - g_c\right)2^P \tag{6}$$

$$B^t(x) = \begin{cases} 1 & \forall\, x \ge 0 \\ 0 & Otherwse \end{cases} \tag{7}$$

Where g_c is binary value of centre pixel at (x_c, y_c) and g_p is binary value around the neighbourhood of g_c. The value of P gives the number of pixels in the neighbourhood of g_c. The local shape descriptor L_S^t of the human dancers pose projects maximum number of points on to graph.

3.3. Graph Construction from Features

Fig.5. shows the extracted dancer represented with LBP features and Haar wavelet features. It also shows HOG features that will be used for feature fusion which may improve the quality of matching.

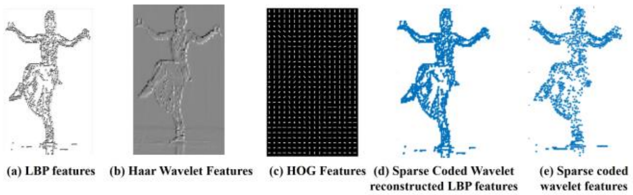

(a) LBP features (b) Haar Wavelet Features (c) HOG Features (d) Sparse Coded Wavelet reconstructed LBP features (e) Sparse coded wavelet features

Fig. 5. Various features of dancer.

Local shape features in fig.5 are used to construct a graph. Given a motion frame in a ICD video sequence V^t and successfully extracted local shape features L_S^t and transformed into a binary shape matrix B_S^t of ones and zeros using eq'n 5. A sparse representation of B_S^t eliminates all zeros and retains only ones and their locations in $M_S^t(x,y,w)$, where x,y are shape point locations and w is shape feature weight vector. Fig.5(d) and 5(e) shows a sparse representation for both wavelet reconstructed LBP (WR_LBP) and only Harr wavelet features (HWF) respectively. The points on the motion object are formed by extracting the location of the pixel and its feature value determines the shape of the dance pose. From these feature point locations and values a graph is constructed in this work.

Given a motion feature $M_S^t(x,y,w)$, let $G=(N,E)$ be its adaptive graph represented with N nodes in the set $(x_i,y_i)\forall\ i=1\ to\ N$ and E defines the edge set. A node in the graph represents a point selected in the feature space. The attributes of node are probability distribution of locations in the motion field of the non-rigid human object. For graph cuts based segmentation this probability distribution is computed by manually selecting the pixels in the motion object [29]. This work uses fully automated model for shape extraction.

3.4. Adaptive Graph Matching

Video or image frame similarity assessment is still an open-ended problem in computer science engineering. The graph matching (GM) articulates the similarity problem as solving matching between two graphs. Literature on GM points to applications addressing problems related to 2D shape matching [14], 3D shape matching [15], object classification [16], feature tracking, symmetry analysis and action recognition [24]. Compared to Random sample consensus RANSAC [30] and Iterative closest point (ICP) [31] algorithms, GM uses paired node relationships when matching structured objects such as human poses. Fig.6. exemplifies the regular GM and proposed model used in our work.

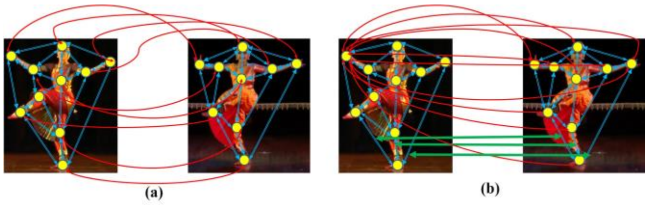


Fig. 6. Graph matching example: Matching two similar dance poses with 10 and 9 Node points (a) GM between pairs of nodes leaving non-paired nodes (b) Proposed GM, where one node is paired with all other nodes and conditional edge matching.

We propose to use adaptive graph matching (AGM) discussed in [32]. A node features in a graph $G=\{N,E,g,h\}$ are represented as node coordinates, $N=[n_1,n_2,....,n_k]\in\square^{d\times k}$. Where d is the dimension of the feature vector used for node creation with k nodes. The edge features $E=[e_1,e_2,...,e_{k-1}]\in\square^{d\times(k-1)}$, having $(k-1)$ edge features. Edges represent links between a pair of nodes. Here edges are computed as distance between the node pairs. The last two terms in G are paired as $\{g,h\}$ gives the graphs topology. The pair $\{g,h\}$ represent node – edge incidence matrices $g,h\in\{0,1\}^{d\times k}$, where $g_{ia}=1$ and $h_{ja}=1$ if the edge a stats from i^{th} node and ends at j^{th} node. In this work g and h are binary matrices representing edge connections.

Given two graphs, $G_1=\{N_1,E_1,g_1,h_1\}$ and $G_2=\{N_2,E_2,g_2,h_2\}$, we compute the affinity parameter $m_{i_1i_2}^N=\underset{\min}{\arg}\,d\left(N_{i_1}^1,N_{i_2}^2\right)$ measures the similarity between the i_1 node in G_1 and node i_2 in G_2. Where $'d'$ is the Euclidian distance between the nodes. Another parameter $m_{a_1a_2}^E=\underset{\min}{\arg}\,d\left(E_{a_1}^1,E_{a_2}^2\right)$ measures the matching similarity between the a_1^{th} edge in G_1 and a_2^{th} edge in G_2. The node and edge similarity measures are linear functions of the Euclidian distance defined as

$$m_{i_1i_2}^N=\left\|N_{i_1}^1-\Im\left(N_{i_2}^2\right)\right\|_2^2 \tag{8}$$

$$m_{a_1a_2}^E=\left\|\left(N_{i_1}^1-N_{j_1}^1\right)-\left(\Im\left(N_{i_1}^2-N_{j_1}^2\right)\right)\right\|_2^2 \tag{9}$$

The function $\Im(\square)$ is a geometrical transformation of G_1 with respect to G_2.

3.5. Adaptive Graph Matching on ICD Videos

The problem of GM is an act of finding similarity between them using a distance parameter. Nevertheless, for ICD videos which are influenced by stringent constraints, an exact matching is not possible. Automating graph construction problem from shape features poses a problem of unequal number of nodes and edges between graphs. The location of nodes in the two graphs representing the same feature is a non-occurring phenomenon. These two problems are addressed in this work. The solution for the two problems lies in the node mappings. Usually there is a one – to – one mapping between the graphs in conventional graph matching. We propose to use many – to – many mapping between nodes and edges which enables the matching to be more robust.

4. Experiments and Results

This section of the paper reports experimental results. The ICD database consists of 4 videos from 2 different dancers for 4 different songs of Bharatanatyam. Our database is labelled with the vocal words that are in the form of a song for a set of frames. The average length of the frames for a label is around 100. There are no repetitions in the database. For each dancer, the online Bharatanatyam video sequence for the vocal song 'Bho Shambho' consists of approximately 10,000 frames. Each vocal word divides into approximately 100 to 120 frames. The offline video dataset for a song on lord Ganesh is having approximately 6000 frames and are labelled every 100 frames. Hence, in our dataset for online videos we have 97 Labelled patterns from 4 online videos. For offline model, we have 59 labelled video samples from 4 dancers. The experiments in this work test the effectiveness of features for AGM. We use 8 combination of features along the proposed WR_LBP with adaptive graph matching.

4.1. Testing Pattern

Leave one out cross validation [33] model is used for testing the AGM model. Out of 4 videos, one video is used as a database and remaining 3 are used as a query video. Matching similarity matrix $M\in\mathrm{R}^{n_1n_2\times n_1n_2}$ decides on the accuracy of the proposed method. This is calculated on a frame to frame basis. However, to estimate the performance for the entire dance action sample, Matching accuracy is calculated as mean of matching similarity matrix over the sample sequence

$$A=\frac{1}{r^2}\sum_{t=1}^{100}\sum_{r}\left|M^t\right| \tag{10}$$

Where, t is number of frames in the sample sequence and r is the number of diagonal elements. The accuracy is in the range $A\in[0,1]$.

4.2. Experimentation

For experimentation, we propose to use 4 modules based on the datasets available to us. Experiment – I uses lab captured (offline) dance sequence matching with the same dancer videos for both training and testing. Exp – II uses two different dance performances with same set of poses with the offline video dataset. In Exp-III, we try to find similarity between same dancer with same set of poses and different dancers with same set of poses under different conditions is tested in Exp-IV with online video dataset. The performance of each experiment with the proposed AGM and the state of the art classification methods such as spectral graph matching (SGM) and Support vector machines (SVM) is tested exhaustively. Performance is measured based on matching accuracy calculated per frame with eq'n 10.

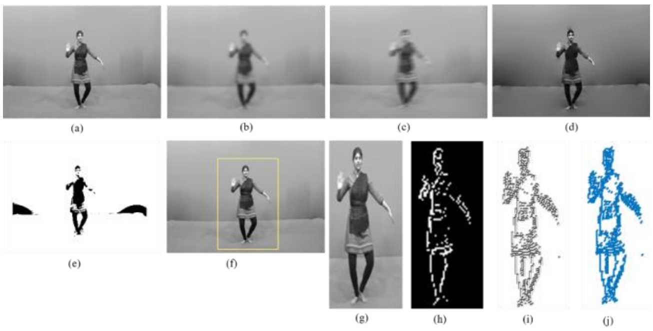
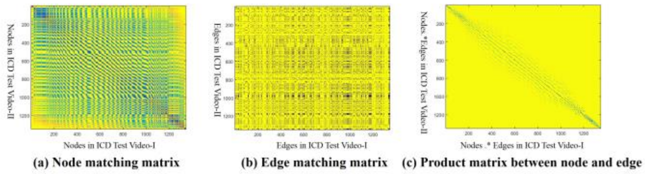


Fig. 7. (a) Gray Frame shot from ICD video, (b) Gaussian smoothing, (c) Averaging, (d) Saliency map, (e) Silhouete creation, (f) Dancer Identification, (g) Extracted Dancer, (h) Wavelet reconstructed features, (i) LBP Features from Wavelet and (j) Constructed graph.

Exp-I uses input videos from dance data set captured in the controlled environment. The dancer identification, feature extraction and graph representation for the dancer is shown in fig.7.

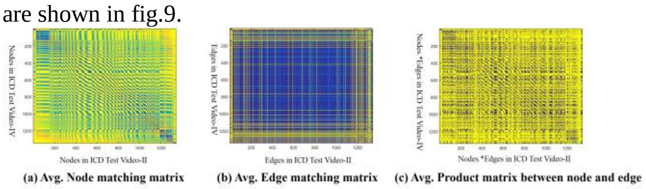
The average number of nodes in the graph are 1041 for the dance video used in exp-1. Two successive frames in the same video sequence will have almost same nodes. However, if there is a large change in the object shape, the number of nodes vary accordingly. We used a frame by frame AGM algorithm on a many to many node and edge matching model as described in eq'n 10. The matching matrix with node matching, edge matching and combined form is shown in fig.8 respectively.



(a) Node matching matrix (b) Edge matching matrix (c) Product matrix between node and edge

Fig. 8. Frame confusion matrix between same dance videos -I and II having the same dancer.

The average matching matrix for the entire video sequence having same dance action and the same dancer produces a near 98% matching appears like fig.8. In exp-2, we use videos-II and IV having same action sequence but performed by different dancers. Each action sequence is labelled into 100 frames and AGM is applied. The averaged node, edge and product matching matrices are shown in fig.9.



(a) Avg. Node matching matrix (b) Avg. Edge matching matrix (c) Avg. Product matrix between node and edge

Fig. 9. Averaged matching matrices for two different dancer videos with same action shapes.

The matching similarity achieved is around 90% calculated as number of frames correctly matched to total number of frames. There can be a lot of false matching that can be detected from fig.9. (c) which is due to small variations in the dancer's movements in the adjacent frames. This false matching is eliminated by diagonalizing the matching matrix. The average matching accuracy computed using eq'11 for offline videos with the proposed WR_LBP features and AGM is around 0.9 for 59 action poses in a performance. The online dance videos are tested using the proposed algorithm in exp-3 and exp-4. The matching similarity in exp-3 is 92% and in exp-4 is 77%. However, the AGM with proposed features produced matching accuracies of around 0.7.

For robustness testing of the features, we challenged our method WR_LBP with 7 other widely used feature modes: HOG, HWF, LBP, HWF+LBP, HWF+HOG, LBP+HOG and HWF+LBP+HOG.

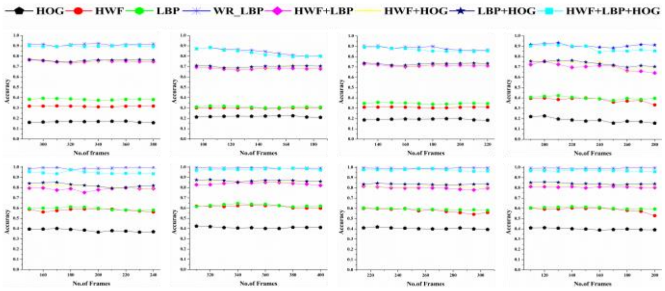


Fig. 10. Top row: performance plots of online dance videos with same train and test data. Bottom row: for offline videos.

The matching accuracies for offline video samples is always around 0.99, however for online video samples it dipped and oscillated between 0.75 - 0.9. The proposed features WR_LBP show a high accuracy on all the ICD data in both offline and online videos. Using AGM asks for good graph construction, which is achieved by WR_LBP features. From fig.10. we observe that the feature combination provides better matching accuracies compared to singleton features with AGM. Using different dataset for training and testing shows that WR_LBP features are good combination for graph construction. Fig.11. gives the plots of matching accuracies on a set of ICD action sequences with all features and AGM classification.

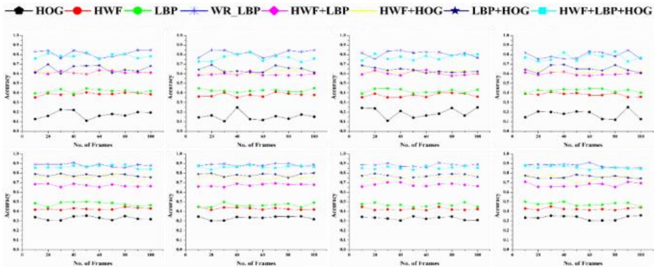


Fig. 11. Top row: performance plots of online dance videos with different train and test data. Bottom row: for offline videos.

The plots shows a 0.9 accuracy for offline videos and 0.73 – 0.86 variation for online videos with the proposed features which is better than the other features and feature combinations. The classifiers performance is tested against state – of – the art support vector machine (SVM) with multi class and Spectral graph matching (SGM) on online, offline ICD videos and compared in Table 1.

Table 1: Comparison of Recognition rates among different classifiers.

Dance Action	AGM		SGM		SVM	
	Onlin	Of-	Onlin	Of-	Onlin	Of-
Yeshoda	64.02	79.26	58.34	65.40	35.28	51.83
Yeshoda_Leg	65.31	88.46	57.41	79.92	37.59	41.43
Suthudu	77.41	99.01	60.37	79.99	44.62	44.86
Suthudu_Leg	78.92	91.46	51.53	75.46	44.09	45.80
Mudhugare	71.31	82.44	58.78	59.90	34.85	39.89
Mudhug-	67.52	67.69	54.76	61.38	38.76	49.73
Manikyam	69.28	69.75	52.79	55.49	34.82	49.80
Mani-	54.59	56.73	48.55	56.82	39.70	41.73
Mahimala	69.46	94.90	49.73	67.49	39.83	39.93
Mahimala_Leg	52.76	64.68	44.50	49.50	33.77	39.06
Kamsuni	70.57	91.55	60.81	89.42	41.06	55.67
Kalivajram	52.48	81.46	52.49	69.65	46.73	49.09
Diddarani	51.62	67.83	51.43	59.76	39.36	59.70
Diddarani_Leg	62.80	73.46	49.89	71.77	39.72	44.72
Devaki	50.21	67.74	52.43	69.49	37.86	57.93
Devaki_Leg	53.47	72.33	56.58	72.32	41.45	51.09
Antha_Intha	60.48	99.79	54.32	64.89	42.88	61.49
Antha_Madhe	63.74	99.85	60.19	76.93	38.74	54.06

From Table 1, the proposed AGM classifier results in good matching when compares to other two algorithms. Most of the reported results and our identification regarding SGP does not support large nodes on graph representation. The other problem is that the number of nodes change with respect to frames, dancer, background and contrast. The proposed AGM due to its many – to – many mapping is good at handling these 2D video problems. Multi class SVM is an excellent classifier, however recorded the lowest accuracy due to closely packed support vectors and the distance plane between samples is small. This results in mediocre classification to online videos. But it is better classifier for many image classification problems where the image object is captured qualitatively.

The proposed AGM classifier is slower and to make it faster, we executed the classifier on a NVDIA graphics processer. Reduction in background noise, efficient foreground extraction, reduction in number of nodes per frame can improve the matching and reliability of the algorithm.

5. Conclusion

Indian classical dance classification is a complex problem for machine vision research. The features representing the dancer should focus on the entire human body shapes. Hand and leg shape segmentation are critical part of a ICD. In this work, we proposed a fully automated ICD consisting of dancer identification, extraction, segmentation, feature representation and classification. Saliency based dancer identification and extraction helps in reducing the image space. Wavelet reconstructed local binary patterns are used for feature representation preserving local shape content of hands and legs. Adaptive graph matching with many – to -many distance calculation between two sets of dance video data is the classifier. Multiple experimentations on online and offline ICD video data is tested. Dance video data is labelled as per the vocal song sequence. The features and classifiers performance tests show that the proposed WR_LBP features and AGM classifier gives better classification accuracy respectively. AGM is accurate for unsymmetrical graphs such as produced in this work from features. More action features can be added for representing dancer graphs more realistically by elimination backgrounds and blurring artefacts.

References

[1] Poppe, Ronald. "A survey on vision-based human action recognition." Image and vision computing 28, no. 6 (2010): 976-990.

[2] Chakravorty, Pallabi. "Hegemony, dance and nation: The construction of the classical dance in India." South Asia: Journal of South Asian Studies 21, no. 2 (1998): 107-120.

[3] Rahmani, Hossein, Ajmal Mian, and Mubarak Shah. "Learning a deep model for human action recognition from novel viewpoints." IEEE Transactions on Pattern Analysis and Machine Intelligence (2017).

[4] Dawn, Debapratim Das, and Soharab Hossain Shaikh. "A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector." The Visual Computer 32, no. 3 (2016): 289-306.

[5] Wang, Heng, and Cordelia Schmid. "Action recognition with improved trajectories." In Proceedings of the IEEE International Conference on Computer Vision, pp. 3551-3558. 2013.

[6] Wang, Heng, Alexander Kläser, Cordelia Schmid, and Cheng-Lin Liu. "Action recognition by dense trajectories." In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp. 3169-3176. IEEE, 2011.

[7] Kishore, P. V. V., M. V. D. Prasad, D. Anil Kumar, and A. S. C. S. Sastry. "Optical Flow Hand Tracking and Active Contour Hand Shape Features for Continuous Sign Language Recognition with Artificial Neural Networks." In Advanced Computing (IACC), 2016 IEEE 6th International Conference on, pp. 346-351. IEEE, 2016.

[8] Vatsyayan, Kapila. Indian classical dance. Ministry of Information and Broadcasting, Government of India, 1992.

[9] Mohanty, Aparna, Pratik Vaishnavi, Prerana Jana, Anubhab Majumdar, Alfaz Ahmed, Trishita Goswami, and Rajiv R. Sahay. "Nrityabodha: Towards understanding Indian classical dance using a deep learning approach." Signal Processing: Image Communication 47 (2016): 529-548.

[10] Yang, Yi, and Deva Ramanan. "Articulated pose estimation with flexible mixtures-of-parts." In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp. 1385-1392. IEEE, 2011.

[11] Wang, Fang, and Yi Li. "Beyond physical connections: Tree models in human pose estimation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 596-603. 2013.

[12] Samanta, Soumitra, Pulak Purkait, and Bhabatosh Chanda. "Indian classical dance classification by learning dance pose bases." In Applications of Computer Vision (WACV), 2012 IEEE Workshop on, pp. 265-270. IEEE, 2012.

[13] K.V.V.Kumar, P.V.V.Kishore., "Indian Classical Dance Mudra Classification Using HOG Features and SVM Classifier" In Proceedings of International Conference on smart computing and information systems, Springer, India, 2017.

[14] Fischler, Martin A., and Robert A. Elschlager. "The representation and matching of pictorial structures." IEEE Transactions on computers 100, no. 1 (1973): 67-92.

[15] Kishore, P. V. V., Kumar, D. A., Sastry, A. S. C. S., & Kumar, E. K. (2018). Motionlets Matching with Adaptive Kernels for 3D Indian Sign Language Recognition. IEEE Sensors Journal, 1–1.

[16] Isenor, D. K., and Safwat G. Zaky. "Fingerprint identification using graph matching." Pattern Recognition 19, no. 2 (1986): 113-122.

[17] Sanfeliu, Alberto, and King-Sun Fu. "A distance measure between attributed relational graphs for pattern recognition." IEEE transactions on systems, man, and cybernetics 3 (1983): 353-362.

[18] Bunke, Horst, and Kim Shearer. "A graph distance metric based on the maximal common subgraph." Pattern recognition letters 19, no. 3 (1998): 255-259.

[19] Bougleux, Sébastien, Luc Brun, Vincenzo Carletti, Pasquale Foggia, Benoit Gaüzère, and Mario Vento. "Graph edit distance as a quadratic assignment problem." Pattern Recognition Letters 87 (2017): 38-46.

[20] Jiang, Bo, Jin Tang, Xiaochun Cao, and Bin Luo. "Lagrangian relaxation graph matching." Pattern Recognition 61 (2017): 255-265.

[21] Yang, Xu, Hong Qiao, and Zhi-Yong Liu. "Point correspondence by a new third order graph matching algorithm." Pattern Recognition 65 (2017): 108-118..

[22] Ye, Dong, Yujun Yang, Bholanath Mandal, and Douglas J. Klein. "Graph invertibility and median eigenvalues." Linear Algebra and its Applications 513 (2017): 304-323.

[23] Zheng, Qiang, Honglun Li, Baode Fan, Shuanhu Wu, Jindong Xu, and Zhulou Cao. "Modified localized multiplicative graph cuts based active contour model for object segmentation based on dynamic narrow band scheme." Biomedical Signal Processing and Control 33 (2017): 119-131.

[24] Patel, Chirag I., Sanjay Garg, Tanish Zaveri, Asim Banerjee, and Ripal Patel. "Human action recognition using fusion of features for unconstrained video sequences." Computers & Electrical Engineering (2016).

[25] Wang, Jin, Mary She, Saeid Nahavandi, and Abbas Kouzani. "A review of vision-based gait recognition methods for human identification." In Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on, pp. 320-327. IEEE, 2010.

[26] ping Tian, Dong. "A review on image feature extraction and representation techniques." International Journal of Multimedia and Ubiquitous Engineering 8, no. 4 (2013): 385-396.

[27] Yang, Mingqiang, Kidiyo Kpalma, and Joseph Ronsin. "A survey of shape feature extraction techniques." (2008): 43-90.

[28] Guo, Zhenhua, Lei Zhang, and David Zhang. "A completed modeling of local binary pattern operator for texture classification." IEEE Transactions on Image Processing 19, no. 6 (2010): 1657-1663.

[29] Sinop, Ali Kemal, and Leo Grady. "A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm." In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pp. 1-8. IEEE, 2007.

[30] Wang, Zeng-Fu, and Zhi-Gang Zheng. "A region based stereo matching algorithm using cooperative optimization." In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pp. 1-8. IEEE, 2008.

[31] Cheng, Shiyang, Ioannis Marras, Stefanos Zafeiriou, and Maja Pantic. "Statistical non-rigid ICP algorithm and its application to 3D face alignment." Image and Vision Computing 58 (2017): 3-12.

[32] Zhou, Feng, and Fernando De la Torre. "Factorized graph matching." IEEE transactions on pattern analysis and machine intelligence 38, no. 9 (2016): 1774-1789.

[33] Cawley, Gavin C., and Nicola LC Talbot. "Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers." Pattern Recognition 36, no. 11 (2003): 2585-2592.