Research Article

Indian Classical Dance Classification with Adaboost Multiclass Classifier on Multifeature Fusion

K. V. V. Kumar, P. V. V. Kishore, and D. Anil Kumar

Department of Electronics and Communications Engineering, KL University, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India

Correspondence should be addressed to P. V. V. Kishore; pvvkishore@kluniversity.in

Received 1 June 2017; Revised 27 July 2017; Accepted 17 August 2017; Published 26 September 2017

Academic Editor: Daniel Zaldivar

Copyright © 2017 K. V. V. Kumar et al. 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.

Extracting and recognizing complex human movements from unconstrained online video sequence is an interesting task. In this paper the complicated problem from the class is approached using unconstrained video sequences belonging to Indian classical dance forms. A new segmentation model is developed using discrete wavelet transform and local binary pattern (LBP) features for segmentation. A 2D point cloud is created from the local human shape changes in subsequent video frames. The classifier is fed with 5 types of features calculated from Zernike moments, Hu moments, shape signature, LBP features, and Haar features. We also explore multiple feature fusion models with early fusion during segmentation stage and late fusion after segmentation for improving the classification process. The extracted features input the Adaboost multiclass classifier with labels from the corresponding song (tala). We test the classifier 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.

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 dancing, running, jumping, or simply walking. Automation encompasses mining the video sequences with computer algorithms for identifying similarities between actions in the unknown query dataset with that of the known dataset. 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 sync to music or vocal sounds. Figure 1 shows a set of online and offline (lab captured) Indian classical dance videos for testing the proposed algorithm.

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 [2–4]. 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. Five features are calculated that represent the exact

FIGURE 1: Online and offline dance datasets used in this work and the video constraints.

shape of the dancer in the video sequence. For recognition, a multiclass multilabel Adaboost 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 and Graph Matching (GM) classifier already proposed by us in our previous work.

2. Literature Survey

Local information of the human in the video is the popular features for action segmentation and classification in recent times. This section focuses 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 [5, 6]. Based on these models, numerous visual illustrations have been proposed for discriminating human action based on shape templates in space-time [5], shape matching [6], interest points in 2D space time models [7], and representations using motion trajectories [8]. Impressively, dense trajectory based methods [9] have shown good results for action recognition by tracking sampled points through optical flow fields. Optical flow fields are based on preconditioned on brightness and object motion

in a video [10]. The algorithms assume uniform brightness variations and object motions in consecutive frames which produce excellent results under minimum constrained video recording. Minimum constrained video recordings are having uniform brightness, less blurring, fixed camera angle, and high contrast between object and background. Finding such a video happens in a movie or in a lab setup. Hence, these approaches need to be robust in estimating human actions, which is still an open-ended problem on real time videos. Data driven methods with multiple feature fusion [11] with artificial intelligence models [12] are currently being explored with the increase in computing power.

In this work, human action recognition on Indian Classical Dance [13] 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. Apart from these complications, the dancer wears complicated dresses with nice makeup and at times during performance the backgrounds are changing

FIGURE 2: Flow diagram of the proposed process for Indian classical dance recognition. (a) Training datasets, (b) detected dance object, (c) extracted, (d) DWT and LBP features, (e) feature points, (f) query dance video, (g) query feature points, and (h) multifeature extraction and classification.

depending on the story which truly makes this an open-ended problem. Mohanty et al. [14] highlight the difficulties in using state-of-the-art pose estimation algorithms such as skeleton estimation [15] and pose estimation [16] which fail to track the dancers moves in both offline and online videos. The author in [14] proposes using deep learning based convolutional neural networks (CNNs) and shows they perform well in estimating the correct pose of dancers on both 3D Kinect dance poses and online videos. CNNs require large training data for a specific class of inputs which makes them computationally slow for a video datasets that change for every 2 frames. In real time, there is no Kinect [17] like effects and hence 2D video analysis must be refined in accuracy for identifying poses of Indian dance forms. Samanta et al. [18] used histogram of oriented optical flow (HOOOF) 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 [19], we approached the same problem with SVM classifier on dance videos and found that only multiclass SVMs should be considered. Moreover, optical flow on online videos suffers a lot due to inconsistencies during capture and sharing process.

In [20], Samanta and Chanda proposed a video descriptor manifold on ICD YouTube videos and KTH dataset with nonlinear SVM classifier and recorded recognition accuracies in the range 70 to 95%. The other works proposed also used SVM classifier with simple image processing models on images of dancers for pose estimation which can be found in [21–23]. In [24], authors used Kinect sensor to capture

leg poses in Indian classical dance forms and classification is initiated with SVM classifier for a set of 40 poses. Kinect sensor produces skeleton data of the human body pose and fails to reproduce data related to fingers which are important in classifying a dance pose.

The objective is to select features that represent a sign and are easily distinguishable in closely related sign words and are computationally efficient. The attributes for a selfie sign language recognizer chosen are shape signature [25] for hand and head shapes, Hu moments [26] for hand orientations, hand–head distance, and hand position vectors for tracking. The chosen attributes perfectly characterize a sign in Indian sign language.

Classifying at faster rate on a huge dataset is a complicated problem. Adaboost [27] classifier is fast and efficient algorithm for large datasets [28]. Inspired by [29, 30], the feature matrix is labelled and inputted to Adaboost classifier for training and testing. The performance indicators are recall-precision curves and execution time on mobile and are recorded to check the robustness of the algorithm and feasibility of implementing more efficiently.

In this paper, we propose an multiclass multilabel Adaboost (MCMLA) based classification problem on multidimensional feature vector. We show that this can be used to match large unconstrained dance features which are automatically extracted from video datasets. The feature representation of video objects depends on the efficiency of video segmentation algorithms. As illustrated in Figure 2, the proposed Adaboost can effectively recover the query video

frames from the dance dataset, by shape–texture observation model defined by discrete wavelet transform (DWT) and local binary patterns (LBP).

In summary, our MCMLA algorithm on online and offline Indian classical dance videos combines the representational flexibility and trivial computations. We perform experiments on two different datasets of Indian classical dance Bharatanatyam and Kuchipudi created from online downloads and offline controlled lab capture. The proposed method is compared with other GM models which are outperformed by a considerable margin in speed.

3. Proposed Methodology

The proposed algorithm framework is shown in Figure 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 [31] is also explored in this work using 5 feature types, Zernike moments, Hu moments, shape signature, LBP features, and Haar features. Adaboost algorithm explores the relativity between the query dance sequence and known dataset.

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 30 fps and dance movements are sometimes faster and at times 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 from constantly upgrading the background information and modelling the object characteristics in real time. The dancer identification module is based on one of the silhouette extraction methods proposed in [32]. A significant indication in determining dancers motion for extraction lies in the temporal changes in the dancer's silhouette during performance. To avoid background modelling and foreground extraction models, we propose using the following procedure.

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

$$\mathbf{V}_m^t(\mathbf{x}, \mathbf{y}) = \mathbf{V}^t(\mathbf{x}, \mathbf{y}) \otimes \mathbf{m}(\mathbf{x}, \mathbf{y}). \quad (1)$$

The size of $\mathbf{m}$ is updated based on the frame size $\mathbf{N} \times \mathbf{M}$ 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 the same size as the input frame. The next step applies a Gaussian filter of μ mean and σ variance on the input frame $\mathbf{V}^t$:

$$\mathbf{V}_g^t(\mathbf{x}, \mathbf{y}) = \mathbf{V}^t(\mathbf{x}, \mathbf{y}) \otimes \mathbf{g}(\mu, \sigma). \quad (2)$$

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

$$\mathbf{S}^t(\mathbf{x}, \mathbf{y}) = \|\mathbf{V}_g^t(\mathbf{x}, \mathbf{y}) - \mathbf{V}_m^t(\mathbf{x}, \mathbf{y})\|_2. \quad (3)$$

The second order normed distance map is shown in Figure 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. Figure 3(d) shows the silhouette mask and connected component output is in Figure 3(e).

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 Figures 3(g) and 3(h), respectively. The extracted dancer is free from background variations in the video sequence. If a portion of background still appears at this stage it can be nullified during the matching phase. Applying feature extraction on the extracted dancer allows for lesser computations as the background is almost eliminated and leads to good matching accuracy.

3.2. Feature Extraction. Which features can help recognize and classify dance correctly is the question. From a dancer's perspective, to identify a dance type, body posture, hand shapes, and their movements in space are the vital features. Feature extraction phase explores the methodology in extracting these features. There are many shape descriptors available in literature for characterizing shape features [33]. Lighting, frame inconsistency, contrast, blurring, and frame size are some of the critical factors that affect feature extraction algorithms. In addition, the dancer velocity during performance has instincts for a faster shape extractor.

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) [34] and Local Binary Patterns (LBP) [35]. The objective at this stage is to represent moving dancers shape with a set of wavelet coefficients. Here we propose using Haar wavelet at level 1. At level 1, Haar wavelet decomposes the video frame $\mathbf{V}^t$ into 4 subbands. Figure 4 shows the 4 subbands at 2 levels. At 1st level we have 4 subbands and at 2nd level have 8 subbands. In the 1st level, the three subbands represent the shape information at three different orientations: Vertical $\mathbf{v}$, Horizontal $\mathbf{h}$, and Diagonal $\mathbf{d}$. Combining the three subbands and averaging the wavelet coefficients normalizes the large values.

$$\mathbf{W}_s^t = \frac{\mathbf{h} + \mathbf{v} + \mathbf{d}}{3}. \quad (4)$$

The averaged shape harr wavelet coefficients $\mathbf{W}_s^t$, along with $\{\mathbf{h}, \mathbf{v}, \mathbf{d}\}$ subband coefficients are reconstructed to spatial domain. Figure 4 shows the reconstructed spatial domain

FIGURE 3: Dancer extraction. (a) Original frame, (b) mean filtered, (c) Gaussian filtered, (d) distance saliency map, (e) silhouette mask, (f) connected components labelling, (g) identified dancer, and (h) dancer extracted.

FIGURE 4: Harr wavelet subbands representing shape in three different orientations.

frame producing the exact hand shapes. These shape features can be used as nodes and a graph can be constructed for recognition. However, background noise is still a major concern at this stage. Local pixel information becomes vital in selection of nodes that exactly represent the graph.

3.2.2. Thresholding. Apply threshold on the reconstructed ICD video frame $\mathbf{V}_r^t$ as

$$\mathbf{T}^t = \sqrt{\frac{1}{NM} \sum_{j=1}^M \sum_{i=1}^N (\mathbf{V}^t(j, i))^2}. \quad (5)$$

The binarized video frame $\mathbf{B}^t$ is

$$\mathbf{B}^t = \mathbf{V}_r^t > \mathbf{T}^t. \quad (6)$$

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

3.2.3. Local Binary Patterns and Local Shape Models. LBP compares each pixel in a predefined neighbourhood to summarize the local structure of the image. For an image pixel $\mathbf{B}^t(\mathbf{x}, \mathbf{y}) \in \mathcal{R}^+$, $(\mathbf{x}, \mathbf{y})$ gives the pixel position in the intensity image. The neighbourhood of a pixel can vary from

FIGURE 5: (a) Original dancer in online video frame, (b) LBP features from reconstructed and thresholded wavelet coefficients, (c) Haar wavelet features, (d) sparse coded wavelet reconstructed LBP features, and (e) sparse coded wavelet features. (f)–(j) Same as (a)–(e) for offline lab captured video frame.

3 pixels with radius $r = 1$ or a neighbourhood of 12 pixels with $r = 2.5$. The value of pixels using LBP code for a centre pixel (x_c, y_c) is given by

$$\mathbf{L}_s^t = \text{LBP}(x_c, y_c) = \sum_{j=1}^P \mathbf{B}^t(\mathbf{g}_p - \mathbf{g}_c) 2^P \quad (7)$$

$$\mathbf{B}^t(\mathbf{x}) = \begin{cases} 1 & \forall \mathbf{x} \geq 0 \\ 0 & \text{Otherwise,} \end{cases}$$

where $\mathbf{g}_c$ is binary value of centre pixel at (x_c, y_c) and $\mathbf{g}_p$ is binary value around the neighbourhood of $\mathbf{g}_c$. The value of P gives the number pixels in the neighbourhood of $\mathbf{g}_c$. The local shape descriptor $\mathbf{L}_s^t$ of the human dancers pose projects maximum number of points on to graph.

3.3. Multifeatures: Zernike, Hu Moments, Shape Signature, LBP, and Haar. Figure 5 shows the extracted dancer represented with LBP features and Haar wavelet features. Local shape features in Figure 5(a) are used to construct a graph. Given a motion frame in a ICD video sequence $\mathbf{V}^t$ successfully extracted local shape features $\mathbf{L}_s^t$ and transformed into a binary shape matrix $\mathbf{B}_s^t$ of ones and zeros using (5). A sparse representation of $\mathbf{B}_s^t$ eliminates all zeros and retains

only ones and their locations in $\mathbf{M}_s^t(\mathbf{x}, \mathbf{y}, \mathbf{w})$, where $\mathbf{x}, \mathbf{y}$ are shape point locations and $\mathbf{w}$ is shape feature weight vector. Figures 5(d) and 5(e) show 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.

Haar and LBP features are enough to label the dancers in the frames and put them under the classifier. This is early fusion of features at the segmentation stage. The fusion operator is principle component analysis. PCA of wavelet features and LBP features is concatenated using the following expression:

$$\mathbf{E}_{fu} = \text{PCA}(\mathbf{W}_s^t) \cup \text{PCA}(\mathbf{L}_s^t). \quad (8)$$

However, 3 more features are proposed in this work that can effectively represent shape features of the dancer. Dancers in Indian classical dance videos have a large motion vector field and Haar and LBP depend on variations in lighting, camera movement, and background. To counterbalance camera movements, we propose using Zernike Moments (ZM)

to represent dancer in each frame on the 2D point cloud extracted from WR_LBP vectors.

Dancer body orientations provide rotation invariant feature of a dance movement in a dancing space. However, these movements are incorrectly classified if there are unavoidable sudden camera vibrations. Moments $\mathbf{M}_{pq}$ project 2D points in dance segments $\mathbf{B}_s^t(\mathbf{x}, \mathbf{y})$ on to basis $\mathbf{x}^p \mathbf{y}^q$ which results in a piecewise continuous linear function in the spatial plane. Moments and moment functions are used as pattern shape features in number of applications [36–38]. Geometric moments are usually defined as

$$\mathbf{M}_{pq} = \iint_{\mathbb{R}} \mathbf{x}^p \mathbf{y}^q \mathbf{B}_s^t(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y}, \quad (9)$$

where $\mathbf{M}_{pq}$ is the $(p+q)$ th order moment of $\mathbf{B}_s^t > 0$. However, geometric moments do not exhibit any invariance properties such as translation, rotation, and scaling. Teague [39] proposed ZM to recover image from moments representing the image in terms of orthogonal polynomials. ZM are used in pattern recognition applications [40, 41].

ZM project the 2D point cloud $\mathbf{B}_s^t(\mathbf{x}, \mathbf{y})$ to a set of orthogonal polynomials, called Zernike polynomials defined as a complete set

$$\mathbf{Z}_{nm}(\rho, \theta) = \mathbf{R}_{nm}(\rho) \mathbf{e}^{jm\theta}, \quad (10)$$

where $\mathbf{R}_{nm}(\rho)$ are real valued radial polynomials defined in [39] and ρ is moment magnitude and θ is angle.

The orthogonality of ZM is modelled as

$$\begin{aligned} & \iint_{\substack{0 \leq \rho \leq 1 \\ 0 \leq \theta \leq 2\pi}} \mathbf{Z}_{nm}^*(\rho, \theta) \mathbf{Z}_{n'm'}(\rho, \theta) \rho \, d\rho \, d\theta \\ &= \frac{\pi}{n+1} \delta_{nn'} \delta_{mm'}, \end{aligned} \quad (11)$$

where “*” indicates complex conjugate and $\delta_{nn'}$ should satisfy

$$\delta_{nn'} = \begin{cases} 1, & \mathbf{n} = \mathbf{n}' \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

ZM of order $\mathbf{n}$ with repetition $\mathbf{m}$ for continuous feature set $\mathbf{B}_s^t(\mathbf{x}, \mathbf{y})$ per frame $\mathbf{t}$ over a unit disk is defined as

$$\mathbf{A}_{nm} = \frac{n+1}{\pi} \iint_{\text{Unit Disk}} \mathbf{Z}_{nm}^*(\mathbf{x}, \mathbf{y}) \mathbf{B}_s^t(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y}. \quad (13)$$

If θ_0 is the rotational angle, with original ZM and rotated ZM as $\mathbf{A}_{nm}$ and $\mathbf{A}_{nm}^R$, respectively, we have

$$\begin{aligned} |\mathbf{A}_{nm}^R| &= |\mathbf{A}_{nm} \mathbf{e}^{-jm\theta_0}| = |\mathbf{A}_{nm}|; \\ \Theta_{nm}^R &= \Theta_{nm} - m\theta_0, \end{aligned} \quad (14)$$

where $|\mathbf{A}_{nm}|$ and Θ_{nm} represent magnitude and phase, respectively. In (14) the magnitude remains constant while the image rotates, whereas the phase changes with image rotation. Hence, most of the applications use ZM magnitude as a feature vector for pattern classification. In this work, we propose ZM magnitude on 2D shape point cloud as invariant feature representing the small variations in camera movement that occur in online video capture of the dance performance. To represent changes in the dancer dimensions which happen nonlinearly, we propose using a nonlinear function defined over geometric moments.

Hu moments in [42] are nonorthogonal centralized moments that are scale, translation, and rotation invariant. Human dancers come in all shapes and sizes and the features describing them may change with dancer. Modelling invariance in dancer shape features from the 2D point segments of Figure 5(e) or Figure 5(j) is done with Hu moments. Hu moments are derived for 2D normalized central moments M_{pq}^H using algebraic invariants:

$$\begin{aligned} \mathbf{h}_1 &= M_{2,0} + M_{0,2}, \\ \mathbf{h}_2 &= M_{2,0} - M_{0,2} - 4M_{1,1}^2, \\ \mathbf{h}_3 &= (M_{3,0} - 3M_{1,2})^2 + (3M_{2,1} - M_{3,0})^2, \\ \mathbf{h}_4 &= (M_{3,0} + M_{1,2})^2 + (M_{2,1} + M_{0,3})^2, \\ \mathbf{h}_5 &= (M_{3,0} - 3M_{1,2})(M_{3,0} + M_{1,2}) \\ &\quad \cdot [(M_{1,2} + M_{3,0})^2 - (M_{2,1} + M_{0,3})^2] \\ &\quad + (3M_{2,1} - M_{3,0})(M_{2,1} + M_{0,3}) \\ &\quad \cdot [3((M_{3,0} + M_{1,2})^2 - (M_{3,0} + M_{1,2})^2)], \\ \mathbf{h}_6 &= (M_{2,0} - M_{0,2})[(M_{3,0} + M_{1,2})^2 - (M_{2,1} + M_{0,3})^2] \\ &\quad + 4M_{1,1} + (M_{3,0} + M_{1,2})(M_{0,3} + M_{2,1}), \\ \mathbf{h}_7 &= (3M_{2,1} - M_{3,0})(M_{3,0} + M_{1,2}) \\ &\quad \cdot [(M_{1,2} + M_{3,0})^2 - (M_{2,1} + M_{0,3})^2] \\ &\quad - (M_{3,0} - 3M_{1,2})(M_{2,1} + M_{0,3}) \\ &\quad \cdot [3((M_{3,0} + M_{1,2})^2 - (M_{2,1} + M_{0,3})^2)]. \end{aligned} \quad (15)$$

These 7 moments are calculated for every extracted dance shape in each frame.

Every pixel is represented with a shape feature [43] or a descriptor [44] in each frame to model dancer shape for classification. However, to reduce the computations on the feature vector during recognition, we propose using shape signatures in [25] to represent dancer shapes. In [25], shapes are represented on a multiscale model based on integral kernels. A shape descriptor or signature is an integral

invariant at various scales forming a shape signature function on dancer shapes $B_s^t(x, y)$

$$S_s(\Gamma) = \frac{\int_x \int_y B_s^t(x, y) \cdot (1 - G_\rho * B_s^t(x, y)) dy dx}{\int_x \int_y B_s^t(x, y) dx dy}, \quad (16)$$

where G_ρ is a Gaussian kernel in 2D with $\gamma > 0$ and zero mean. Also,

$$\gamma = \Gamma \left(\int_x \int_y F_s(x, y) \right)^{1/2}. \quad (17)$$

Integrating shape feature values over shape 2D spatial domain results in a shape signature value normalized by area of the shape to achieve scale invariance. The range of shape signature is $[0, 1]$ based on value of scale γ . For low values of scale, it is 0 and for large scales it approaches 1.

3.4. Feature Matrix Construction. From the 5 features, a complete feature matrix is constructed per dance frame. Early fusion is performed with Haar and LBP features with PCA at the end of dancer segmentation and are labelled with dance vocal words. This feature matrix is named as EFDF (early fused dance features). It is a 2D feature matrix of size $256 \times 256 \times 2$. Max pooling is performed to reduce the feature set to $1 \times 256 \times 2$ per frame. Further, late fusion model is introduced with more robust set of features in the Zernike moments, Hu moments, and shape signatures (SS) on 2D shape point cloud of the dancer in the video frame. These late features are mixed with Haar and LBP features to create a multifeature matrix per frame. We calculated 5 ZM, 7 Hu Moments, 1 SS, max pooled, and thresholded LBP and Haar features limited to 30 features per frame. The final feature matrix in late fusion strategy is $5 \times 7 \times 1 \times 30 \times 30$ per frame. These features are carefully labelled with vocal words representing the dance form in the video frame. For a 25-frame dance word “swami ra ra,” we have 73×25 feature matrix. This feature matrix or a set of matrices are inputted to MCMLAB classifier.

3.5. Dance Classifier: Adaboost Multiclass Multilabel. Boosting based classifications [29, 30, 45] find very precise hypothesis from a set of weak hypotheses. Here hypothesis is a classification rule. Set of weak hypotheses are simple rules that generate a predictable classification. Let $T = [(f_1, L_1), (f_2, L_2), (f_3, L_3), \dots, (f_v, L_v)]$ be a set of training examples at an instance f_i on i th frame in feature space f with labels L_i on label space L . The algorithm accepts the training samples T along with some class distribution $D = \{1, \dots, m\} \in \mathbb{R}$ represented as weak learners. On the input, the weak learner computes a weak hypothesis H . Generally, $H : f \rightarrow \mathbb{R}$. The interpretation for classification is based on $\text{sign}\{H(f)\} = \{+1, -1\} \rightarrow \{f_i\}$ for a binary classifier. $|H(f_i)|$ gives prediction confidence.

The key to boosting is to use the weak learner to produce a very precise prediction rule by repeatedly addressing the weak learner on different distribution of training examples. In this work, a multiclass version of Adaboost is used having a set of strings as class labels. The problem is modelled as given

T and size of final strong classifier C . The Adaboost initializes the distribution function as

$$D_1(i, l) = \frac{1}{vm} \quad \forall i = 1, \dots, v, l = 1, \dots, m, \quad (18)$$

where $l = |L|$. For $c = 1, \dots, C$, we select a weak classifier $H_c : T \times L \rightarrow [-1, 1]$ with distribution D_c , to maximize the absolute value of

$$a_c = \sum_{i,c} D_c(i, l) L_i(l) H_c(f_i, l). \quad (19)$$

We choose the biasing value α_c as

$$\alpha_c = \frac{1}{2} \ln \left(\frac{1 + a_c}{1 - a_c} \right) \quad (20)$$

and update the distribution function as

$$D_{c+1}(i, l) = \frac{D_c(i, l) e^{-\alpha_c L_i(l) H_c(f_i, l)}}{N_c}, \quad (21)$$

where N_c is normalization factor to keep the distribution as probability density function. The final output strong classifier is

$$H(f, l) = \text{sign} \left(\sum_c (\alpha_c H_c(f, l)) \right). \quad (22)$$

For the multiclass problem $c_1, \dots, c_k$, we use the real valued 2D Look-Up Table model in [41], which is defined as

$$H_{\text{LUT}}(f, L) = \sum_{i=1}^n \sum_{j=1}^k (2P_i^{(j)} - 1) B_n^{i,j}(f, l), \quad (23)$$

where $P_i^{(j)} = P(f \in C_i | \text{frame} \in \text{sign})$ and $B_n^{i,j}(f, l) = \{1, f \in l; 0, \text{otherwise}\}$. From this weak hypothesis, through training a strong hypothesis is generated to recognize sign labels $Z_i = H(f_i)$.

4. Experimentation and Results

A set of 4 experiments are initiated to test the robustness of the proposed multifeature fusion with Adaboost classifier. Our Indian classical dance datasets consist of performances on “Bharatanatyam” and “Kuchipudi” from online YouTube videos and offline dance videos in controlled environment at KL University, cams department studio. We have created 4 dance videos from 5 dancers for 2 songs in two different dance styles. Similar online YouTube downloaded dance performances are also collected. Matching frames in each dataset are shown in Figure 6.

In exp-1 and exp-3, we use offline and online dataset of same dancer video for training and testing with early fusion and late fusion of multiple features 28 words in the dance sequence. Each mudra pose is coordinated with vocal manually by labelling each set. Variations in number of frames per label are nullified and normalized to 15 Key frames per dance pose across all video data. Exp-2 and Exp-4 are conducted based on different training and test

FIGURE 6: (a) Online Indian classical dance datasets from YouTube. (b) Offline dance video dataset created at KL University in a controlled lab environment.

video set from the online and offline dataset individually. These experiments test features in their early and late fusion models with Adaboost classifier. In the next phase, we test the classifier based on accuracy and efficiency in classifying dance gestures.

We use three performance evaluators for validating the results. They are precision–recall curves, percentage recognition rate, and the computation time per sign. For a strong hypothesis $\mathbf{H}$ resulted from Adaboost training and testing for an input feature $\mathbf{f}_i$ on a trained distribution $\mathbf{D}$ with $\mathbf{Z}_i = \mathbf{H}(\mathbf{f}_i)$ predicted labels. The following metrics in [46] are

$$\text{Precision}(\mathbf{H}, \mathbf{D}) = \frac{1}{|\mathbf{D}|} \sum_{i=1}^{|\mathbf{D}|} \frac{|\mathbf{L}_i \cap \mathbf{Z}_i|}{|\mathbf{Z}_i|}, \quad (24)$$

$$\text{Recall}(\mathbf{H}, \mathbf{D}) = \frac{1}{|\mathbf{D}|} \sum_{i=1}^{|\mathbf{D}|} \frac{|\mathbf{L}_i \cap \mathbf{Z}_i|}{|\mathbf{L}_i|}, \quad (25)$$

$$\text{Recognition} = \frac{1}{|\mathbf{D}|} \sum_{i=1}^{|\mathbf{D}|} \frac{|\mathbf{L}_i \cap \mathbf{Z}_i|}{|\mathbf{L}_i \cup \mathbf{Z}_i|}. \quad (26)$$

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 are shown in Figure 7. Saliency maps from average and Gaussian distance metric create Silhouette, which identifies the dancer in the video frame. The dimensions of the bounding box extract the dancer. Then the dancers features are extracted with segmentation as early features. Haar wavelet at level 1 is averaged in high frequency components to remove background and IDWT is performed to recover the global shapes on the dancer. Applying LBP on the resulting IDWT dance frame captures local shape information. At this stage, we perform Adaboost multiclass multilabel classification with the PCA based Haar and LBP feature fusion.

Early fused Haar-LBP features of offline dance video of a dancer are used to train the Adaboost classifier. For the same dance video shot with slight variations is provided as query dance video for same set of labels. The resulting confusion matrix from early fusion of features on same training and testing set is shown in Figure 8.

The normalized values in the confusion matrix are calculated using (26). We did plot the confusion matrix for the song sequence “Siva Shambhu” with first 28 word labels. The average recognition on total 116 labels from two dances in “kuchipudi” from a dancer is 0.96. From other dancer’s videos, the Adaboost classifier averaged around 0.955. Exp-1 is repeated with all the same parameters with late fusion with Zernike moments, Hu Moments, and liner shape signature along with Haar and LBP features. Late feature matrix is a 73×25 matrix per label. For our 116-label dance sequence we have a $73 \times 25 \times 116$ dimension feature vector. Figure 9 shows the results of the classifier in the form of confusion matrix (showing 28 labels) from (26).

The results show an average of 0.99 for all dance videos in the dataset. Clearly, multiple features and late fusion have increased the ability of the classifier to recognize dance poses correctly. False matching is much less in this experiment as the datasets used for training and testing are similar, that is, same dancer and same performance.

In Exp-2, the performance is the same, meaning same labels for training but the dancer performing the dance will be different. However, testing with a different query video having different dancer results in lower recognition rates compared to the previous experiment. Here also, the simulations are performed for early feature fusion and late fusion. Confusion matrices from the two simulations are shown in Figures 10 and 11, respectively. The average recognition dropped due to noncoherence between dance moves with respect to body

FIGURE 7: (a) Offline frame shot from ICD video, (b) Gaussian smoothing, (c) averaging, (d) saliency map, (e) silhouette creation, (f) dancer identification, (g) extracted dancer, (h) wavelet reconstructed features, (i) LBP features from wavelet, and (j) constructed 2D point shape cloud.

FIGURE 8: Exp-1 early fusion confusion matrix with same offline dancer in training and test video.

shapes, pose shapes, and movement speeds between the two dances.

Early feature fusion in exp-2 provided a recognition 0.81 averaged over a set of 4 video samples of the same song. However, late feature fusion with more versatile feature

representing a dance frame produced a recognition rate of 0.91. Exp-3 trains the Adaboost classifier with online dance video content and tests with the same set for early fusion and late feature fusion. The initial shape extraction from online dance video content is shown in Figure 12.

FIGURE 9: Exp-1 late fusion confusion matrix with same offline dancer in training and test video.

FIGURE 10: Exp-2 early fusion confusion matrix with different offline dancer in training and test video.

FIGURE 11: Exp-2 late fusion confusion matrix with different offline dancer in training and test video.

FIGURE 12: (a) Gray frame shot from ICD video, (b) Gaussian smoothing, (c) averaging, (d) saliency map, (e) silhouette creation, (f) dancer Identification, (g) extracted Dancer, (h) wavelet reconstructed features, (i) LBP Features from Wavelet, and (j) constructed 2D Point Shape Cloud.

FIGURE 13: Exp-3 early fusion confusion matrix with same online dancer in training and test video.

The confusion matrices for early fusion and late fusion are presented in Figures 13 and 14 for exp-3, online dance videos, respectively.

The recognition rate decreased for online dance videos as these are unconstrained in real time capture. Problems such as vibrant video backgrounding, poor or over lighting, and motion blurring in online videos made the feature extraction complicated. However, the proposed algorithm with early feature fusion has resulted in an average recognition rate of 0.84 for one test video under these circumstances. The average recognition rate for 4 test samples in exp-3 is 0.83. The problems in online videos are effectively handled by increasing the number of features to represent a dance pose. Figure 14 gives the confusion matrix with late fusion features. The average recognition rate is 0.93 for 4 test videos.

Similarly, we obtained average recognition of 0.68 with early feature fusion and 0.82 with late fusion, respectively. The confusion matrices are shown in Figures 15 and 16.

To summarize, the results obtained with the proposed early fusion and late feature fusion on offline and online dance videos are presented in Table 1. The dataset consists of 4 online and offline videos. For each video, we computed the average recognition rate obtained by taking mean of all recognition rates per dance pose.

From Table 1, the performance of the Adaboost classifier on different online and offline dance videos can be recorded. This data can be interpreted to understand the ability of features to uniquely model dancers pose in a real time dance video to measure the dancers' performance. Harr and LBP

together model global and local dance shapes, respectively. However, these features have constraints on motion blurring, scale variations, brightness, and contrast variation in the video frames. Apart from these image variations, we have camera vibrations, dancer shape variations, and video backgrounds to restrict these feature vectors from uniquely modelling a dancer in the video frame. To compensate for these variations, additional features in the form of shape signatures, Zernike moments, and Hu moments are proposed to represent a dance pose. Shape signature models the dancer pose uniquely with an integral kernel having Gaussian characteristics. The shape signature is calculated on the shape curve on a set of frames which matches similar dance pose at a different location in a video sequence. Similarly, ZM are linear orthogonal representation of video data that handle dancer's movements in spatial domain. Hu moments handle camera vibrations and other nonlinear movements of the dancer in the video frames.

Early fusion and late fusion concepts are introduced to understand which set of features can correctly and uniquely model a dance pose irrespective of the constraints during video capture. Simulations show that late fusion and more number of features are necessary for training Adaboost classifier. Same dancer videos resulted in better recognition rates compared to different dancer in both offline and online dance videos. We also tested the classifier with HOG (Histogram of oriented Gradients), SIFT (Scale Invariant Feature Transform), and SURF (Speeded Up Robust Features) with MCMSAB classifier. Training vector is made of 50 best

FIGURE 14: Exp-3 late fusion confusion matrix with same online dancer in training and test video.

FIGURE 15: Exp-4 early fusion confusion matrix with different online dancer in training and test video.

TABLE 1: Summary of mean recognition rates for 4 offline and online dance performances for the song “Siva Shamboo” in kuchipudi dance.

Training dance video	Early PCA fusion on with Haar and LBP				Late fusion with Haar, LBP, ZM, HuM and SS			
	Exp-1	Exp-2	Exp-3	Exp-4	Exp-1	Exp-2	Exp-3	Exp-4
Offline video-1	0.95	0.88			0.99	0.92		
Offline video-2	0.93	0.82			0.99	0.91		
Offline video-3	0.94	0.78			0.98	0.85		
Offline video-4	0.87	0.74			0.92	0.81		
Online video-1			0.92	0.80			0.98	0.85
Online video-2			0.93	0.78			0.97	0.82
Online video-3			0.92	0.69			0.99	0.77
Online video-4			0.81	0.66			0.88	0.71

FIGURE 16: Exp-4 late fusion confusion matrix with different online dancer in training and test video.

features and the same number is used for testing. The average recognition rates were 0.84 for HOG, 0.82 for SIFT, and 0.8 for SURF in Exp-1 where same training and query dance video is used. But for different dancer videos the recognition dropped to 0.67, 0.65, and 0.59 in case of exp-2 for HOG, SIFT, and SURF, respectively. Similar results were seen in Exp-3 and Exp-4 for online dance videos. However, in Exp-4, for different dance video sets for training and testing have drastically reduced the recognition rate of the classifier by 50%. The drop-in classifier performance can be attributed to the poor feature extraction due a large variation in the video frames even though they have same dance pose.

HOG, SIFT, and SURF features are extracted from the original gray scale video frame. To improve their performance the algorithms are applied on the extracted dancer

from our Haar-LBP sparse segmentation module. On the segmented and extracted dancer, the average recognition rate from HOG is 0.89, SOFT is 0.91, and SURF is 0.82 for exp-1. In exp-2 these values were again reduced by a factor of 18%. Exp-3 and Exp-4 showed an improvement of 30% increase in average recognition rate for all three feature extraction models. However, after consecutive testing and measurements these state-of-the art features reported less overall average recognition rates compared to the proposed late fusion features that form a complete set to model a dance video sequence.

The set of 5 late fused features with 73 attributes per frame performed well with MCMLAB classifier compared to early fusion and other feature extraction models; it is time to test for different classifiers. We applied the same

FIGURE 17: Comparison of classifiers for dance pose recognition.

late fusion features to adaptive graph matching (AGM) and Support Vector machine classifier. In AGM, each node is modelled with 73 features and the distance between the features is models as edge. For each dance pose video, a one-to-one matching on both node and edge feature is calculated. Recognition rate is calculated using (24). Multiclass SVM is the most trusted classifier for character recognition. Hence we apply the late fused features to a SVM classifier and each dance pose is measured.

Three classifier are compared here with late fusion features and the other Adaboost with early fusion features taking it to 4 classifiers. All these classifiers are compared for recognition rate and efficiency. We plot average recognition for the 4 classification models averaged across 4 offline and online videos in Figure 17.

From Figure 17, the Adaboost with late feature fusion with 5 features outperforms the AGM and SVM. However, AGM comes close to Adaboost classifier and sometimes is better than MCMLAB on late fusion features for dance pose recognition. Nevertheless, AGM is far slower than MCMLAB algorithm. SVM came last in the comparison due to inefficiency in defining initial support vectors for classification from the fused feature vectors. MCMLAB with early fusion features is better than SVM classifier. MCMLAB is better, if the example set is uniquely defined by the feature set and the proposed features are the best choice for Indian classical dance pose recognition.

Equations (25) and (26) are used to calculate precision and recall with late and early fusion on Adaboost for 2 offline and 2 online video sets. Similarly, the same datasets are used for AGM and SVM classifiers. One video will be used for same feature train-test model and other is a different train-test model. The average precision and recall values are plotted as a graph in Figure 18.

The ability to recall precisely the same label as the query video as a performance measure is plotted in Figure 18. It shows the same result as above, making MCMLAB as the

FIGURE 18: Classifier performance measurements.

FIGURE 19: Comparisons with other models of ICD with the proposed early and late fusion models with Adaboost classifier.

classifier for dance pose recognition with Haar, LBP, ZM, Hu moments, and SS giving satisfactory outcomes in robust comparisons.

Figure 19 shows the percentage recognition obtained from methods in [20, 21, 23, 24] along with the proposed early and late feature fusion models. The recognition percentage is averaged over the entire dataset. The plots in Figure 19 highlight the use of multiple features for various representations of moving objects in a dance video for accurate classification and recall.

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 is a critical part of a ICD. In this work, we proposed a fully automated ICD consisting of dancer identification, extraction, segmentation, and 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. Two fusion models are proposed for feature fusion. Early fusion at the segmentation stage with PCA based Haar wavelet and LBP is used and late fusion using the Zernike moments, Hu moments, and shape signatures is used with Haar and LBP is proposed. Multiclass multilabel Adaboost on features of early fusion and late fusion between two sets of dance video data is the classifier. Multiple experiments on online and offline ICD video data are tested. Dance video data is labelled as per the vocal song sequence. The early and late features and classifiers performance tests show that the proposed late fusion features and multiclass multilabel Adaboost classifier give better classification accuracy and seed compared to AGM and SVM. More action features can be added for representing dancer more realistically by elimination backgrounds and blurring artefacts to improve the efficiency of the classifier.

Conflicts of Interest

The authors declare that they have no conflicts of interest related to this research in any form.

References

- [1] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, 2010.
- [2] J. G. Lochtefeld, *The Illustrated Encyclopaedia of Hinduism*, vol. 1, The Rosen Publishing Group, 2002.
- [3] P. Chakravorty, "Hegemony, dance and nation: the construction of the classical dance in India," *South Asia: Journal of South Asia Studies*, vol. 21, no. 2, pp. 107–120, 1998.
- [4] A. Sinha, *Let's Know Dances of India*, Star Publications, 2006.
- [5] H. Rahmani, A. Mian, and M. Shah, "Learning a deep model for human action recognition from novel viewpoints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1.
- [6] E. Rodolà, S. R. Bulò, T. Windheuser, M. Vestner, and D. Cremers, "Dense non-rigid shape correspondence using random forests," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pp. 4177–4184, Columbus, OH, USA, June 2014.
- [7] D. Das Dawn and S. H. Shaikh, "A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector," *Visual Computer*, vol. 32, no. 3, pp. 289–306, 2016.
- [8] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 3551–3558, Sydney, Australia, December 2013.
- [9] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu, "Action recognition by dense trajectories," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 3169–3176, June 2011.
- [10] P. V. V. Kishore, M. V. D. Prasad, D. A. 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 *Proceedings of the 6th International Advanced Computing Conference, IACC 2016*, pp. 346–351, Bhimavaram, India, February 2016.
- [11] A. Jalal, Y. Kim, Y. Kim, S. Kamal, and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features," *Pattern Recognition*, vol. 61, pp. 295–308, 2017.
- [12] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013.
- [13] K. Vatsyayan, Indian classical dance. Ministry of Information and Broadcasting, Government of India, 1992.
- [14] A. Mohanty, P. Vaishnavi, P. Jana et al., "Nrityabodha: towards understanding indian classical dance using a deep learning approach," *Signal Processing: Image Communication*, vol. 47, pp. 529–548, 2016.
- [15] Y. Yang and D. Ramanan, "Articulated pose estimation with flexible mixtures-of-parts," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011*, pp. 1385–1392, Colorado Springs, Colo, USA, June 2011.
- [16] F. Wang and Y. Li, "Beyond physical connections: tree models in human pose estimation," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2013*, pp. 596–603, Portland, Ore, USA, June 2013.
- [17] S. Saha, S. Ghosh, A. Konar, and A. K. Nagar, "Gesture recognition from Indian classical dance using kinect sensor," in *Proceedings of the 5th International Conference on Computational Intelligence, Communication Systems, and Networks, CICSyN 2013*, pp. 3–8, Madrid, Spain, June 2013.
- [18] S. Samanta, P. Purkait, and B. Chanda, "Indian Classical Dance classification by learning dance pose bases," in *Proceedings of the 2012 IEEE Workshop on the Applications of Computer Vision, WACV 2012*, pp. 265–270, Breckenridge, Colo, USA, January 2012.
- [19] K. V. V. Kumar and P. V. V. Kishore, "Indian Classical Dance Mudra Classification Using HOG Features and SVM Classifier," in *Proceedings of the In Proceedings of International Conference on smart computing and information systems*, India, 2017.
- [20] S. Samanta and B. Chanda, "Indian classical dance classification on manifold using jensen-bregman logdet divergence," in *Proceedings of the 22nd International Conference on Pattern Recognition, ICPR 2014*, pp. 4507–4512, Stockholm, Sweden, August 2014.
- [21] S. Mozarkar and C. S. Warnekar, "Recognizing bharatnatyam mudra using principles of gesture recognition gesture recognition," *International Journal of Computer Science and Network*, vol. 2, no. 2, pp. 46–52, 2013.
- [22] M. Devi and S. Saharia, "A two-level classification scheme for single-hand gestures of Sattriya dance," in *Proceedings of the 2016 International Conference on Accessibility to Digital World (ICADW)*, pp. 193–196, Guwahati, India, December 2016.
- [23] Shubhangi and U. S. Tiwary, "Classification of Indian classical dance forms," *Lecture Notes in Computer Science (including*

- subseries *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), vol. 10127, pp. 67–80, 2017.
- [24] S. Saha, S. Ghosh, A. Konar, and R. Janarthanan, "A study on leg posture recognition from Indian classical dance using Kinect sensor," in *Proceedings of the 2013 International Conference on Human Computer Interactions, ICHCI 2013*, Chennai, India, August 2013.
 - [25] B.-W. Hong and S. Soatto, "Shape matching using multiscale integral invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 1, pp. 151–160, 2015.
 - [26] D. Dahmani and S. Larabi, "User-independent system for sign language finger spelling recognition," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 1240–1250, 2014.
 - [27] G. Rätsch, T. Onoda, and K. R. Müller, "Soft margins for AdaBoost," *Machine Learning*, vol. 42, no. 3, pp. 287–320, 2001.
 - [28] B. Wu, H. Ai, C. Huang, and S. Lao, "Fast rotation invariant multi-view face detection based on real adaboost," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '04)*, pp. 79–84, May 2004.
 - [29] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class AdaBoost," *Statistics and Its Interface*, vol. 2, no. 3, pp. 349–360, 2009.
 - [30] C. Qi, Z. Zhou, Y. Sun, H. Song, L. Hu, and Q. Wang, "Feature selection and multiple kernel boosting framework based on PSO with mutation mechanism for hyperspectral classification," *Neurocomputing*, vol. 220, pp. 181–190, 2017.
 - [31] C. I. Patel, S. Garg, T. Zaveri, A. Banerjee, and R. Patel, "Human action recognition using fusion of features for unconstrained video sequences," *Computers and Electrical Engineering*, 2015.
 - [32] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A review of vision-based gait recognition methods for human identification," in *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, DICTA 2010*, pp. 320–327, Sydney, Australia, December 2010.
 - [33] D. P. Tian, "A review on image feature extraction and representation techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 385–395, 2013.
 - [34] M. Yang, K. Kpalma, and R. Joseph, "A survey of shape feature extraction techniques. (2008): 43–90.
 - [35] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.
 - [36] E. G. Karakasis, A. Amanatiadis, A. Gasteratos, and S. A. Chatzichristofis, "Image moment invariants as local features for content based image retrieval using the Bag-of-Visual-Words model," *Pattern Recognition Letters*, vol. 55, pp. 22–27, 2015.
 - [37] G. A. Papakostas, D. E. Koulouriotis, E. G. Karakasis, and V. D. Tourassis, "Moment-based local binary patterns: a novel descriptor for invariant pattern recognition applications," *Neurocomputing*, vol. 99, pp. 358–371, 2013.
 - [38] S. Nigam and A. Khare, "Integration of moment invariants and uniform local binary patterns for human activity recognition in video sequences," *Multimedia Tools and Applications*, vol. 75, no. 24, pp. 17303–17332, 2016.
 - [39] M. R. Teague, "Image analysis via the general theory of moments," *Journal of the Optical Society of America*, vol. 70, no. 8, pp. 920–930, 1980.
 - [40] D. Zhang and G. Lu, "Content-based shape retrieval using different shape descriptors: A comparative study," in *Proceedings of the 2001 IEEE International Conference on Multimedia and Expo, ICME 2001*, pp. 1139–1142, jpn, August 2001.
 - [41] M. Khare, R. K. Srivastava, and A. Khare, "Object tracking using combination of daubechies complex wavelet transform and Zernike moment," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 1247–1290, 2017.
 - [42] M. K. Hu, "Visual pattern recognition by moment invariant," *IRE Transactions of Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.
 - [43] P. V. V. Kishore, P. E. Rajesh Kumar, K. Kumar, and S. R. C. Kishore, "Video audio interface for recognizing gestures of indian sign," *International Journal of Image Processing (IJIP)*, vol. 5, no. 4, 479 pages, 2011.
 - [44] P. V. V. Kishore, M. V. D. Prasad, C. R. Prasad, and R. Rahul, "4-Camera model for sign language recognition using elliptical fourier descriptors and ANN," in *Proceedings of the 4th International Conference on Signal Processing and Communication Engineering Systems, SPACES 2015 - In Association with IEEE*, pp. 34–38, ind, January 2015.
 - [45] B. Wu, H. Ai, C. Huang, and S. Lao, "Fast rotation invariant multi-view face detection based on real adaboost," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '04)*, pp. 79–84, May 2004.
 - [46] S. Godbole and S. Sarawagi, "Discriminative methods for multi-labeled classification," in *Advances in Knowledge Discovery and Data Mining*, vol. 3056 of *Lecture Notes in Computer Science*, pp. 22–30, Springer, Berlin, Germany, 2004.

Hindawi

Submit your manuscripts at
<https://www.hindawi.com>