Stratification of Indian Dance Forms Through Audio Signal

Sanghamitra Das, Suchibrota Dutta and Arijit Ghosal

Abstract Differentiating different Indian dance forms from an audio signal is a difficult problem. Though these dances are different from each other, still they can be broadly categorized into two categories—classical dance form as well as folk and tribal dance form. Previous works were mostly engaged to discriminate these two broad forms through facial expressions which are tough, complicated as well as expensive. The proposed approach relies on the audio signal which is easier as well as less expensive. It is observed that instruments associated with these two broad kinds of forms are quite different. To study this difference, chroma-based facets are used, which are capable of estimating the strength of all probable notes generated by different instruments. Also, it is seen that mel frequency cepstral coefficients (MFCC), being a very good aural facet, is overused. Classical dance forms generally have a lower energy level compared to other forms, and this is represented by short time energy (STE)-based facet. The two broad categories of dance forms bear a basic asymmetry between them, and this can be represented by skewness-based facet. The combination of chroma-based facets and skewness-based facet is projected as a substitute of MFCC having the same dimension. Support vector machine or SVM, neural network or NN and Naïve Bayes-based classifier are employed for classification of Indian dance forms. Classification accuracy obtained here is compared with classification accuracy of other preceding works done to echo the effectiveness of the projected set of features.

Keywords Classical folk and tribal dance · Classification · Chroma · Note · Short time energy · Skewness

S. Das · A. Ghosal
Department of Information Technology, St. Thomas' College of Engineering and Technology, Kolkata 700 023, WB, India

S. Dutta (✉)
Department of Information Technology and Mathematics, Royal Thimphu College, Thimphu, Bhutan

A. K. Das et al. (eds.), *Computational Intelligence in Pattern Recognition*, Advances in Intelligent Systems and Computing 1120,
https://doi.org/10.1007/978-981-15-2449-3_6

1 Introduction

India is witnessing a diverse and wide-ranging culture from ancient times. Also, India is constituted with different states, and each of these states has different languages, customs, eating habits, religion, etc. Resembling to all other aspects of culture, the dance forms of different states in India are also different. Still these dance forms can be classified into two broad categories—a) classical and b) folk and tribal dance. Now the repository of these dance forms, stored mostly in digital format, is of huge volume and also it is increasing every day. To store properly as well as to retrieve properly, an automatic classification or recognition of different dance forms is being required nowadays. This classification task must be supported by different audio features which are well capable of differentiating classical dance forms from folk and tribal dance forms. Hence, not audio features can be used in this work. They need to be chosen judiciously. Also at the same time, it is not possible to create a well-defined rule or set of rules for distinguishing classical dance forms from folk and tribal dance form, and the machine intelligence approach is required to classify these two categories. Differentiation of dance forms is tremendously inspiring and good research field to discover because this gets its purposes in different search engines, digital archive of dance forms, media services as well as in different intelligent human–computer interactive systems. Furthermore, improvement in the domain of signal processing as well as in data mining is another reason of fortification of research activities in the area of audio data handling which includes both audio data storing and retrieval. Discrimination of different dance forms into two broad categories—classical as well as folk and tribal—can be considered as elementary step towards managing digital archive of different dance forms. Though dance form discrimination is a very prospective research meadow, still awfully a small number of works have been done in this domain till now. These works are mostly dependent on face expression, body and hand movement which are little bit complex to implement; whereas, this effort intends to suggest some aural features which are of low-dimensional and in addition to computationally very simple to discriminate dance forms into the said two broad categories.

In this effort, concise explanation of earlier research works is discussed in Sect. 2. Advised scheme is portrayed in Sect. 3. Experimental outcomes along with evaluation with other previous works are discussed in Sect. 4 followed by conclusion segment.

2 Previous Works

Researchers have paid their efforts with different kinds of features to suggest a good facet set for working with difference dance form related activities. Dewan et al. [1] have used deep learning-based approach to classify Indian dance form. They have proposed a three-step deep learning pipeline in their scheme. Indian classical dance actions have been classified by Kishore et al. [2] using convolutional neural networks

(CNN)-based approach. Protopapadakis et al. [3] have tried to identify poses based on human skeletal data. Bharatanatyam mudra images have been classified by Anami et al. [4] by using artificial neural network. Indian classical dance is again discriminated with adaboost multiclass classifier by Kumar et al. [5]. Indian classical dance mudra has been classified by Kumar et al. [6] applying histogram of oriented (HOG) features and SVM classifier. Bhatt and Patalia [7] have classified folk dance song using MFCC and LPC. Protopapadakis et al. [8] have classified folk dance using kinect sensor.

Mohanty et al. [9] have used deep learning method for semantic understanding of Indian classical dance. A comparative analysis of representations of folk dances through video was done by Fotiadou et al. [10]. Devi et al. [11] have performed a survey to recognize dance gestures. Kitsikidis et al. [12] have followed unsupervised approach to partition dance sequences into multiple periods and motion patterns. They have used fused skeletal data of the dancer. Local spatiotemporal feature model on manifold to classify Indian classical dance has been adopted by Samanta and Chanda [13]. Hand gesture of Bharatanatyam has been recognized by Saha et al. [14]. Kapsouras et al. [15] have recognized Greek folk dances from video. They have used clustering and SVM in their approach. Chroma toolbox was designed for MATLAB by Müller and Ewert [16] to extract chroma-based audio features.

3 Proposed Methodology

Previous research activities indicate that earlier activities were mostly based on skeletal information, pose of body and video. Most of these approaches are based on processing of images. Hence, data collection becomes tough, complicated as well as expensive. Very fewer approaches are there which are based on audio signal. Most of these approaches have used mel frequency cepstral coefficients (MFCC) as facet. MFCC is an excellent aural facet but it is used in many approaches. Extreme usage of MFCC has formed an opening to find another aural facet which will act as a substitute of MFCC. The principal aim of this work is to suggest a low-dimensional acoustic facet which should not be based on MFCC but has discrimination property similar to MFCC.

Classical dance forms are mainly supported by percussion type instruments like tabla; whereas, folk and tribal dance form generally do not use percussion type of instruments. Notes constructed by two varieties of instruments are not same. The characteristics of notes are well observed by chroma-based facet.

Moreover, classical dance form bears less energy than folk and tribal dance form. This energy level variation can be measured through short time energy (STE). Also at the same time, it is observed that these two types of dance forms bear a basic asymmetry when it is performed. This asymmetry can be measured through skewness. Calculation of audio facets and categorization of aural signal into two different dance forms are elucidated in the subsequent subsections. The stairs of suggested approach are made clear in Fig. 1.

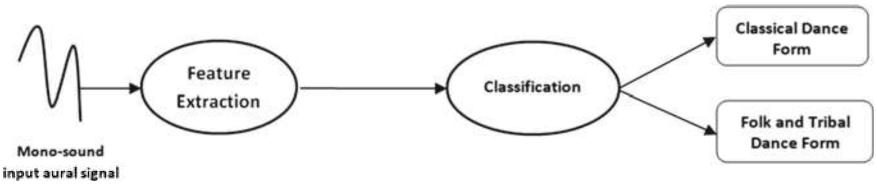


Fig. 1 General idea of suggested approach

3.1 Computation of Features

Classification job of a classifier gets easy if it is fed with an excellent feature set. Excellent feature set always retains inter-class distance high and intra-class distance low. If the facets are elected thoughtfully, then the facets set will be an excellent feature set. Also when the features are chosen, it is also taken care that the facet set has to be of low-dimensional to minimize computational expense.

3.1.1 Features Based on Chroma

Chroma facets represent tune where the complete spectrum of the entered audio signal is reflected through 12 distinctive semitones or chroma of melodic octave. Chroma-depended facets are symbolized as pitch category profiles too. These facets are considered as a very powerful tool for investigating music especially for instrumental melody. These facets are powerful as they are capable of capturing harmonic as well as melodic characteristics of music. When classical dance as well as folk and tribal dance are performed, it creates an impression to human mind that the sounds of these two categories are harmonically and melodically different from each other. The impressions created as the supporting instrumental music for these two categories are different. Generally, it is seen that classical dance form is supported by percussion type instrumental music; whereas, folk and tribal type dance form is supported by different types of instrumental music. It is known that notes can represent an instrumental music best, and characteristics of this note can be best represented by musical octave. Hence, chroma-based facets are the best facets to analyse instrumental music. Chroma values can be represented by

$$\{C, C\sharp, D, D\sharp, E, F, F\sharp, G, G\sharp, A, A\sharp, B\}$$

It signifies 12 pitch spelling traits which are exercised in western 12 note range representation. The set of chroma worths are recognized as a set of integers {1, 2, 3, …, 11, 12}. Here, 1 points out chroma C, 2 points out $C\sharp$, and lastly 12 indicates B. An anthology of all pitches having alike chroma is described as a pitch class. Chroma symbolizes slant of pitch revolving as this goes throughout the whirl. Two octave-linked pitches will carry identical slant in chroma encircle, and this circle is portrayed through Fig. 2. This connection can neither be echoed by linear pitch scale nor by pertaining mel. Benefit of pertaining chroma-based facets compared to mel-based facets like mel frequency cepstral coefficients (MFCC) is this.

This angle is carved up in 12 spots or pitch sets to investigate western tonal music. Strength or intensity of every pitch sets is referred to by chroma facets. So, there will be total 12 chroma facets corresponding to 12 pitch classes.

To extract chroma-based facets, input audio signal is divided into frames having 50% overlapped to avoid loss of any border line characteristics of a frame. Plots of chroma values with magnitude for classical dance form as well as for folk and tribal dance form are portrayed in Figs. 3 and 4 correspondingly.

Figures. 3 and 4 signify that chroma facets differ distinctly for different kinds of dance forms. Chromagram toolbox [16] is utilized to extract chroma facets.

Fig. 2 Chroma circle

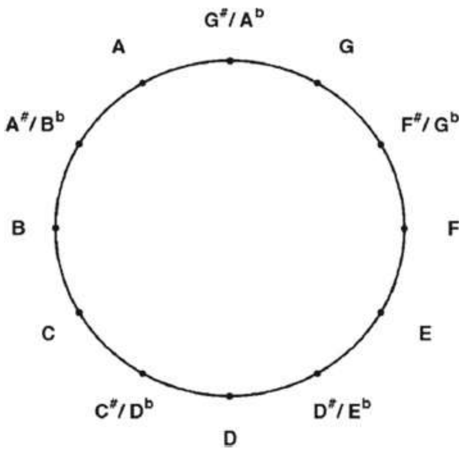
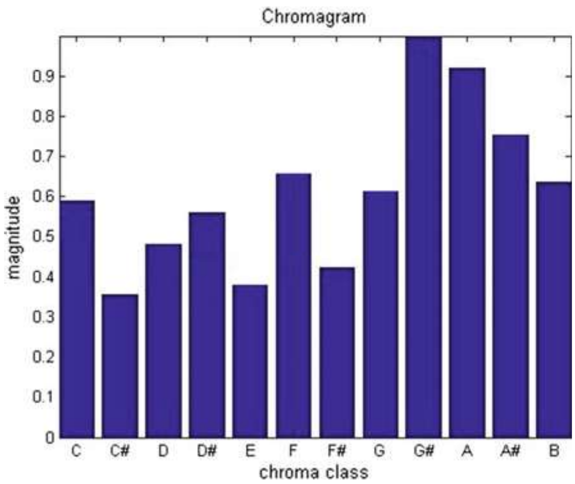


Fig. 3 Plot for chroma assessments aligned with scale for classical dance form



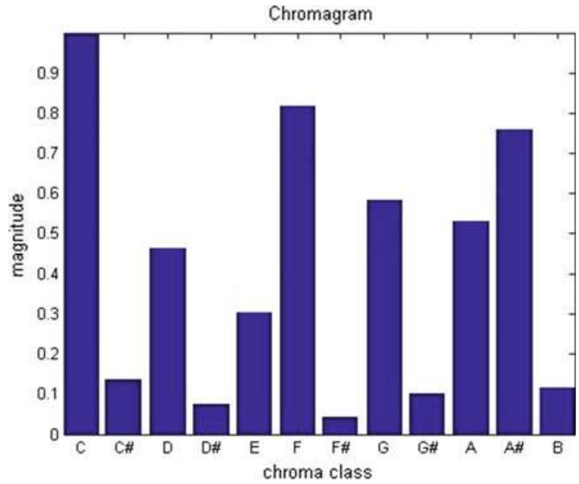


Fig. 4 Plot for chroma assessments aligned with scale for folk and tribal dance form

3.1.2 Feature Based on Short Time Energy (STE)

It is observed that energy level of classical dance form is reasonably different than that of folk and tribal dance form. To be acquainted with how energy level is diverging with respect to time for classical dance form as well as for folk and tribal dance form, STE is required to be calculated. STE is measured for any input aural signal divided into Q frames $\{x_i(m)$: $1 \leq i \leq Q\}$ by Eq. (1) as follows:

$$st_n = \frac{1}{\text{length}} * \sum_{m=0}^{n-1} [x_n(sa)]^2 \tag{1}$$

where st_n indicates energy corresponding to nth frame, length indicates length of the frame, and $x_n(sa)$ signifies sath sample in nth border. All edges are 50% overlapped to evade ignorance of some border line character for a frame.

To capture characteristics of overall energy level variation, only mean in addition to standard divergence of STE is considered.

3.1.3 Feature Based on Skewness

Third-order instant of the spectral circulation is represented as skewness. By nature, skewness is an arithmetical facet. It assesses unevenness existing for a normal circulation with reference to its mean pose. Classical dance form maintains a little quantity of unevenness in its normal circulation with reference to that of folk and tribal dance form. Supporting instrumentals for classical dance forms are different compared to that of folk and tribal dance form which leads towards an asymmetry between them. Skewness *ske* is quantified using the below mentioned Eq. (2):

$$ske = \frac{E(m - \mu)^3}{\sigma^3} \tag{2}$$

where mean of sampled data m is denoted by μ, standard divergence of m is denoted by σ, and anticipateed value of x is symbolized by $E(x)$.

3.2 Classification

Primary aim of this work is to differentiate classical dance forms from folk and tribal dance forms using a small dimensional facet set. Categorizing supremacy of the advised 15-dimensional facet set (12 chroma-based facets + 2 STE-based facets + 1 skewness-based facet) is tested by feeding it to support vector machine or SVM, Naïve Bayes and neural network or NN. All of these classifiers are very popular as well standard supervised classifier. Multilayer perceptron or MLP is employed aiming to put into service neural network or NN. An auditory data set has been structured consisting of 600 audio files—300 files for classical dance form and 300 files for folk and tribal dance form. For classification purpose, the entire aural data set is segregated into two equal subdivisions—teaching and examining. This stands for 150 files considered from both classical dance form and folk and tribal dance form from the aural data set effecting total 300 audio files both for training and testing. “Quadratic” kernel type is considered aiming to implement support vector machine (SVM). To implement Naïve Bayes classifier tenfold cross-validation is considered. Multilayer perceptron or MLP model is considered to put into operation neural network or NN having 15 neurons in the input layer. These 15 neurons indicate 15 facets. The MLP model has two neurons in the production level denoting two styles of categories—classical dance form as well as folk and tribal dance form. There are nine neurons in the hidden layer of MLP model.

4 Experimental Results

To execute this experimentation to discriminate classical dance forms from folk and tribal dance form, a custom aural data set has been put in order. This data set embraces 600 audio files—300 files for classical dance forms and 300 files for folk and tribal dance forms. The entire aural files of the data set are of monotype and also at the same time have 90s duration. In order to offer extensive variety in this data set, miscellaneous types of classical types of dance forms like Bharatanatyam, Kathak, Kathakali as well as varied types of folk and tribal dance forms like Bhangra and Bihu have been considered. These sound files are accumulated from recordings of CD, audio recordings of diverse live performances as well as from the Internet. To assess the feat of the recommended facet set in the day to day situation, a little of these aural files in the data set are noisy also for the said two categories.

Table 1 Classical dance form and folk and tribal dance form categorization accuracy

Classification scheme	Categorization accuracy (in %) for suggested facet set
SVM	95.50
Naïve Bayes	95.00
Neural network	94.50

Table 2 Classical dance form and folk and tribal dance form categorization accuracy (MFCC-based facet set)

Classification scheme	Categorization accuracy (in %) for MFCC-based facet set
SVM	94.00
Naïve Bayes	93.50
Neural network	92.50

All the aural files are broken into 50% overlie frames to evade loss of any margin nature of any frame. Half of the data set is used as training data set and rest is considered as testing data set for the used supervised type classifier. After classification, task training and testing data set has been overturned, and classification job is executed once again. Average of these classification tasks is judged as closing categorization outcome, and this is put in Table 1.

4.1 Relative Analysis

Classification performance of the suggested facet set is contrasted with MFCC as well as among other facet set proposed in some work done previously.

4.1.1 Comparative Classification Result with MFCC

As the facets utilized in this toil are advised as unconventional to MFCC, the same classification task is again carried out considering MFCC as facet set. This classification result is put in Table 2. This time also training and testing data set has been overturned, and average of two classification tasks is considered.

4.1.2 Comparative Classification Result with Other Work

Discrimination supremacy of suggested facet set is evaluated with precedent work. The identical aural data set employed in this work has been used to employ the scheme suggested by Dewan et al. [1]. They have used deep learning approach to classify Indian dance forms. The relative discrimination performance is put into Table 3, and

Table 3 Comparative presentation of intended toil with other work in % exactness

Instance method	Categorization accuracy (in %)
Dewan, et al. [1]	93.50
Proposed method (SVM-based)	95.50

since, this it is apparent that advocated facet set outperforms scheme suggested by Dewan et al. [1].

5 Conclusion

For discriminating classical dance form from folk and tribal dance form chroma-based aural facets, STE-based facet and skewness-based facet have been proposed here. The facet set exploited in this toil is advised as an option to MFCC as MFCC is overused in audio-related work. Experimental outcomes show that recommended facet set carries out better than not only MFCC but also better than other on hand facet set to separate classical dance form from folk and tribal dance form. The facet set advised in this effort is by nature short dimensional and at the alike instant computationally very simple and easy to exercise. In future, each of the categories may be further subdivided into several sub-categories.

References

1. Dewan, S., Agarwal, S., Singh, N.: A deep learning pipeline for Indian dance style classification. In Tenth International Conference on Machine Vision (ICMV 2017), vol. 10696, p. 1069611. International Society for Optics and Photonics, Apr. 2018
2. Kishore, P.V.V., Kumar, K.V.V., Kiran Kumar, E., Sastry, A.S.C.S., Teja Kiran, M., Anil Kumar, D., Prasad, M.V.D.: Indian classical dance action identification and classification with convolutional neural networks. Adv. Multimed. (2018)
3. Protopapadakis, E., Voulodimos, A., Doulamis, A., Camarinopoulos, S., Doulamis, N., Miaoulis, G.: Dance pose identification from motion capture data: a comparison of classifiers. Technol. **6**(1), 31 (2018)
4. Anami, B.S., Bhandage, V.A.: A comparative study of suitability of certain features in classification of Bharatanatyam mudra images using artificial neural network. Neural Process. Lett., 1–29 (2018)
5. Kumar, K.V.V., Kishore, P.V.V., Anil Kumar, D.: Indian classical dance classification with Adaboost multiclass classifier on multifeature fusion. Math. Probl. Eng. (2017)
6. Kumar, K.V.V., Kishore, P.V.V.: Indian classical dance mudra classification using HOG features and SVM classifier. Int. J. Electr. & Comput. Eng. (2088–8708) **7**(5) (2017)
7. Bhatt, M., Patalia, T.: Neural network based Indian folk dance song classification using MFCC and LPC. Int. J. Intell. Eng. Syst. **10**(3), 173–183 (2017)

8. Protopapadakis, E., Grammatikopoulou, A., Doulamis, A., Grammalidis, N.: Folk dance pattern recognition over depth images acquired via kinect sensor. 3D ARCH-3D Virtual Reconstruction and Visualization of Complex Architectures (2017)
9. Mohanty, A., Vaishnavi, P., Jana, P., Majumdar, A., Ahmed, A., Goswami, T., Sahay, R.R.: Nrityabodha: towards understanding Indian classical dance using a deep learning approach. Sig. Process. Image Commun. **47**, 529–548 (2016)
10. Fotiadou, E., Nikolaidis, N., Tefas, A.: A comparative study of representations for folk dances recognition in video. In 2016 24th European Signal Processing Conference (EUSIPCO), pp. 115–119. IEEE (2016)
11. Devi, M., Saharia, S., Bhattacharyya, D.K.: Dance gesture recognition: a survey. Int. J. Comput. Appl. **122**(5) (2015)
12. Kitsikidis, A., Boulgouris, N.V., Dimitropoulos, K., Grammalidis, N.: Unsupervised dance motion patterns classification from fused skeletal data using exemplar-based HMMs. Int. J. Herit. Digit. Era **4**(2), 209–220 (2015)
13. Samanta, S., Chanda, B.: Indian classical dance classification on manifold using Jensen-Bregman LogDet divergence. In 2014 22nd International Conference on Pattern Recognition, pp. 4507–4512. IEEE, Aug. 2014
14. Saha, S., Konar, A., Gupta, D., Ray, A., Sarkar, A., Chatterjee, P., Janarthanan, R.: Bharatanatyam hand gesture recognition using polygon representation. In Proceedings of the 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC), pp. 563–567. IEEE (2014)
15. Kapsouras, I., Karanikolos, S., Nikolaidis, N., Tefas, A.: Folk dance recognition using a bag of words approach and ISA/STIP features. In Proceedings of the 6th Balkan Conference in Informatics, pp. 71–74. ACM, Sept. 2013
16. Müller, M., Ewert, S.: Chroma toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR). hal-00727791, version 2-22 Oct 2012 (2011)