ISSN 1512-1232

UDC Subject Classification **004.93 Pattern information processing**

OPTICAL FLOW BASED ETHIOPIAN TRADITIONAL DANCE VIDEO CLASSIFICATION SYSTEM

Andargie Mekonnen[1], Dr. Subramanian Karpaga Selvi[2], Dr Vajeravel Sampath Kumar[3], Bisratie Tesfaye[4]

[1]PG student Computer Engineering, Mekelle University (EIT-M), Ethiopia

E-mail: andumak@gmail.com

[2]Assistant Professor, Computer Engineering, Mekelle University (EIT-M), Ethiopia

E-mail:karpagaselvil@gmail.com

[3]Assitant Professor, Instrumentation and Control Engineering, Mekelle University (EIT-M), Ethiopia

E-mail:@gmail.com

[4]Head, Computer Engineering Chair, Mekelle University, Mekelle, Ethiopia

E-mail:bisratietesfaye@yahoo.com

Abstract

In this research the application of optical flow method on complex motion is studied. Optical flows are extracted in Ethiopian traditional dance video using Lucas-Kanade and Horn-schunck methods. The optical flows estimated from these methods have three dimensions which is complex for real-time application. In our study the magnitude of optical flow only considered to reduce complexity in training and to make it suitable for real time content based video classification. The effectiveness of the method is evaluated by artificial neural network (ANN) to classify videos with Ethiopian traditional dances. In the study the Lucas Kanade method outperformed by scoring 82.0% of total accuracy whereas Horn-Schunck method and the combination of Lucas-Kanade and Horn Schunck methods less performed by registering 74.6% and 52.8% of overall accuracy respectively.

Keywords: *Optical flow estimation, Pattern Recognition, Classification, Artificial Neural Network, Ethiopian traditional dance.*

1. Introduction

Traditional music helps young generation of the country to understand more about their country and be proud of their tradition and also motivates for further searching about the cultural aspects.

All traditional music is associated with unique traditional dance. As some source [1] indicated one of the earliest structured uses of dance is to tell the ancient stories through dance performance. Even before the production of written languages, dance was one of the methods of passing these stories down from generation to generation. People used it sometimes to show their feeling for their beloved one. It is also linked to the origin of “love making."

Nowadays traditional video clips database are growing very fast due to increase in internet bandwidth speeds, rapid development of digital communication technologies and of personal computer capacities [2]. This rapid growth of technology leading to country’s traditional video clip more pervasive in both corporate intranets and on the internet.

Ethiopia is a widely diverse country with over 80 unique rich ethnic, cultural, custom and linguistic groups [3]. The country is a musically traditional country, with extremely diverse and each of the country's ethnic groups being associated with unique traditional musical rhythm and dance style. Each ethnic group use music to transmit certain ideas, religious beliefs,

ISSN 1512-1232

historical events, ancient stories, emotions, thoughts and to express them through a peculiar way of body movements performed on a certain musical background. The extraordinary diversity of dances is a result of different cultures and people.

According to Timkehet et al. [3], Ethiopian traditional dances are not divided according to their function, but according to their uniqueness and individuality. Therefore, there are over 150 unique dance movements across the country.

Currently, Ethiopian traditional music's are available both in audio and video format in huge amount and one of the most accessible video on the internet. This huge availability of traditional video clips and a high number of viewers has generated a need for new techniques and automated video clip indexing mechanisms. Unique dance movements of particular tradition could be identified in video sequences by action recognition system and used for indexing

1.1 Human Action and Recognition

In computer vision, action recognition is used to interpret an action component of a video or image scenes. According to Turaga et al. [4], action recognition is the process of naming actions, usually in the simple form of an action verb, using sensory observations.

Action recognition from videos can be divided into two main paradigms based on the way video sequences analyzed [2]: Static and Dynamic Recognition. An approach using static representation analyzes the individual frames first and then combines the results into the sequence, whereas an approach using dynamic representation treats the entire sequence (or a fixed length of it) as its basic analysis unit.

According to Liang et al. [5], action recognition from video sequence has four basic operations: video segmentation (frame extraction), object detection and tracking, feature extraction and classification. In some action recognition studies, video segmentation, object detection and tracking are taken as video preprocessing tasks [6, 7, 8].

Video segmentation is the first step in action recognition system. A set of smaller video clips that best represent the action or activity are extracted. Object detection or people detector is the key to find the region of interest from a scene or frame. Feature extraction is used to extract the descriptive features that are used for representing the contents of the particular action. Classification of human Action based on frame representation is a process of assigning a frame or sequence of frames into predefined action labels. There are several machine learning algorithms used in action recognition such as, Support Vector Machine (SVM), ANN, K - Nearest Neigbour(KNN), Hidden Markov Model(HMM) and Linear discriminant analysis(LDA).

The rest of this paper organized as follow: Review of related work reveals the outcome of the intensive study made on selected research works in the area of action recognition and dance classification. Design of ETDV describes the proposed architectures for Ethiopian traditional dance video classification system. The implementation of each technique and parameters used during experimentation is discussed. Evaluations of the proposed method along the result achieved are reported in the result and analysis section. The final section contains concluding remarks and possible future directions.

2. Review of Related Work

There has been a significant amount of work done so far to solve the automatic action recognition problem for several purposes. Recent surveys [9, 10, 11] provide a detailed description of the literature in action recognition. In [9], the approaches are broadly classified as form based and motion based methods. In addition to this, there are methods that combine both form and motion features. The form based approaches are not good for classification of ETDV due to the fact that image data's on the internet that has similar form or shape with video frames will be retrieved and slow down video retrieving efficiency.

ISSN 1512-1232

Different approaches are introduced to perform action recognition by analyzing motion pattern of video frame pixels. Md. Atiqur et al. [12] find out an optical-flow based action recognition algorithm to cluster different human body movement. They demonstrate that, body motion of human can be recognized and identified using motion magnitude and direction of movement calculated from each video frame of human body. According to Md. Atiqur et al. [12], when action is performed between two different times there is a movement of pixels of a frame towards, different direction in accordance with motion of the actor, either vertically or horizontally that mainly indicate the unique feature of that move.

The system proposed by MD Atiqur et al. [12] start by extracting optical-flow of sequential video frames to detect the presence and direction of motion. The optical-flow also further processed for purification of motion detection using Random sample Consensus algorithm (RANSAC).

The classification accuracy of MD Atiqur et al. [12] system is demonstrated on applying Weizmann database by using Euclidean Distance and SVM classifier for classification; they perform 100% and 91.67% accuracy respectively. The authors admit that these results are achieved for simple movement of human body. They recommended that problem of complex motion is still open end and need to be addressed.

Heidmann et al. [13] investigate motion based situation recognition in group meeting. They analyzed videos taken from camera by, implying sociological model of individual behavior and group situations.

The evaluation of Heidmann et al. [13] system was performed as a two stage process. First, head and both hand feature detected and feature are extracted from the position data using the proposed method. Afterwards, the individual activity is recognized using HMM for individual participant. They test the trained HMM model with 23 minute video data to get 94% overall correct classification rate. They planned to include the recognition of group meeting in contrast of several group meetings using posture and gaze direction.

3. Design of Ethiopian Traditional Dance Video Classification System

In this study, the researcher addressed an interesting application of computer vision technique, namely the Optical-flow based classification of ETDV. As shown in figure1, the overall system design of Optical-flow based ETDV classification contains three main blocks; video preprocessing, feature extraction and classification.

3.1 Video Preprocessing

Preprocessing of videos is the first task in many action recognition researches. There is no standard dataset available for research purpose. For building and training the model and as well as for testing the system, the researcher prepared a dataset from the videos primarily downloaded from YouTube and Dire Tube video and SodereTube library. During preprocessing, the following three main tasks are done.

3.2 Video segmentation

A video is a sequence of frames and the temporal information from those images is important in defining an action. In our proposed framework, segmentation of input videos into smaller clips of two second length was done. The researcher considered four main reasons for inclusion of this step. The first reason is in this study, only traditional dances with a single actor or dancer are considered and the whole video clip doesn’t contain shots of only traditional dance. It is difficult to use the whole video clip directly during recognition. Secondly, smaller video clips are easier to work with unlike long video clips. The third reason is, optical flow feature extraction needs noise free frames for robust representation. The fourth reason is, video clip must consist sequential frames that have the same intensity for the convenient of optical flow estimation.

ISSN 1512-1232

Traditional video clip splitting was done with the rate of 15fps. For convenience the researcher converted all video formats to MPEG-2 video format. The main reason why the researcher used those rates is because most traditional dance of Ethiopia move is fast and performed locally and globally at the same time. So there is a need to select an appropriate frame per second for robust feature representation.

3.3 Video clip Selection and frame extraction

After video segmentation, the researcher manually selected 100 smaller video clips from all segmented clips. The reason for inclusion of video selection step is because all shots in a particular video clip don't contain traditional dance only or contains shot in the middle of the clip which doesn't describe a particular dance category. At the end, 100 video clips each contain 30 sequential frames (a total of 30*100=3,000) and extracted with 15fps are selected as a data set. Therefore, totally 3000 frames are a prepared for training and testing.

3.4 Gray Scale Conversion

Each frame of the data set is converted to gray scale frame in order to make it suitable for feature extraction technique. In addition to this, since Gray level representation of the frame pixel value is represented using 8 bits and conversion to Gray level can reduce 24 bit (number of bits required for representation of the color image) to 8 bit frame representation, it is important to reduce the complexity during feature extraction technique. On the other hand, according to [14] Gray scale Frames are enough for recognizing each action from a frame.

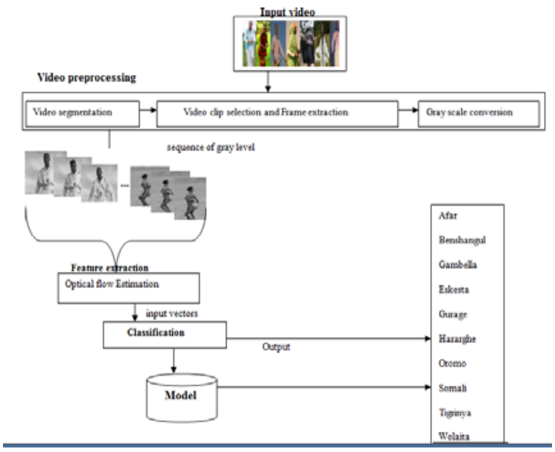

Figure 1: Block Diagram of ETDV classification system

3.5 Feature Extraction

After the conversion of each sequential frames of a video clip into a Gary level the next step is extracting a motion feature using optical flow method.

3.5.1 Optical flow estimation

Optical flow is an approximation of the local image motion based upon local derivatives in a given sequence of images [15]. In this study, differential optical flow is used to extract motion features. The basis of differential optical flow is the motion constraint equation that is represented as (1).

$$I(x,y,t) = I(x+\partial_x), y+\partial_y, t+\partial_t \qquad (1)$$

After simplification motion constraint equation is (2) or (3).

$$(I_x, I_y).(v_x, v_y) = -I_t \qquad (2)$$

$$I.\vec{v} = -I_t \qquad (3)$$

Where, $\nabla I = (I_x, I_y)$ the spatial gradient equation and $\vec{v} = (v_x, v_y)$ optical flow at pixel (x, y) at time t.

The motion constraint equation had two variables to be solved. In this paper Lucas and Kanade and Horn and Schunck optical flow algorithms are used to estimate optical flow.

a. Lucas and Kanade algorithm

A weighted least square (LS) fit of local first-order constraint equation is implemented to a constant model for $\vec{v}$ in each small spatial neighborhood Ω by minimizing [16];

$$\sum_{x,y\in\Omega} W^2(x,y)[\nabla I(x,y,t).\vec{v} + I_t(x,y,t)]^2 \qquad (4)$$

where $W^2(x,y)$ denotes a window function that gives more influence to constraint at the center of the neighborhood than those at the periphery. The solution of equation (4) is given by

$$\vec{v} = [A^T W^2 A]^{-1} A^T W^2 \vec{b} \qquad (5)$$

where, for N pixels (for a $n x n$) neighborhoods $N = n^2, (x_i, y_i) \in \Omega$ at a single time t:

$$A = [\nabla I(x_1, y_1), \dots, \nabla I(x_N, y_N)],$$
$$W = \mathrm{diag}[W(x_1, y_1), \dots, W(x_N, y_N)],$$
$$\vec{b} = -[I_t(x_1, y_1), \dots, I_t(x_N, y_N)] \qquad (6)$$

b. Horn and Schunck optical flow

Horn and Schunck [17] combined the gradient constraint of LS with a global smoothness term to constrain the estimated velocity field $\vec{v} = (v_x, v_y)$, minimizing:

$$\int_D (\nabla I.\vec{v} + I_t)^2 + \lambda^2\left[\left(\frac{\partial_{v_x}}{\partial_x}\right)^2 + \left(\frac{\partial_{v_x}}{\partial_y}\right)^2 + \left(\frac{\partial_{v_y}}{\partial_x}\right)^2 + \left(\frac{\partial_{v_y}}{\partial_y}\right)^2\right] dx\,dy \qquad (7)$$

Defined over a domain D (the image), where the magnitude of λ reflects the relative smoothness of the smoothness term.

3.6 Classification

A number of pattern classification techniques have been used for the recognition of patterns [18]. In this study, ANN classifiers are used to classify Ethiopian traditional dance video.

3.6.1 Artificial Neural Network

ANNs are highly distributed interconnections of adaptive nonlinear processing elements. In other words, they are large set of interconnected neurons, which execute in parallel to perform the task of learning. Hence, ANN resembles human brain in two respects. The first property is that knowledge is acquired by the network through a learning process. The other is interneuron connection strengths known as weights are used to store the knowledge, i.e., the weights on the connections encode the knowledge of a network.

These networks are inspired by the concept of the biological nervous system, and have proved to be robust in dealing with the ambiguous data and the kind of problems that require the interpolation of large amounts of data. Instead of sequentially performing a program of instructions, neural networks explore many hypotheses simultaneously using massive parallelism. Neural networks have the potential for solving problems in which some inputs and corresponding output values are known, but the relationship between the inputs and outputs is not well understood or is difficult to translate into a mathematical function [18, 19].

ISSN 1512-1232

4. Results and analysis

In this section, an evaluation is done based on classification accuracy of the proposed method of Ethiopian traditional dance video classification system. The system only focuses on recognizing traditional dance performed by single dancer or actor. Motion features are represented as magnitude of the optical-flow extracted between two time sequences frames independently from each sequential frame. Each optical-flow is represented by 80 by 80 squared matrix forms. Both unsupervised and supervised back propagation ANN was implemented during classification of each traditional dance video.

4.1 Data Set Preparation

Since previously there have been no data set for similar study, the researcher prepared new data set used in this study. 158 traditional video clips are mainly collected from three main video websites; YouTube, Dire tube and SodereTube. The dance video clips are cut into 2 second small video clips to select traditional dance only. Two second video is optimum length that hold right traditional dance in most traditional dance video display on the screen. The researcher selected the most descriptive video clips to use in the system. From the selected clips, most of them has different format. Since video splitting tool we have used support format like MPEG video format, conversion of this file to MPEG was done using Prism Video File Converter.

After format conversion, input videos are split into smaller clips of generally around 30 frames if they are in ; video clip which contains most descriptive sequential frames contains a single dancer or actor and recorded with fixed camera. Then each clip is segmented using Virtual Dub tool with 15fps rate. We used 15fps rate due to its representative property. Figures show frames extracted using 15fps.

Figure 2: Video frames taken from Gurage traditional dance video using 15fps

At the end of segmentation, the next task to be done is cropping each frame to get the same person to background proportion between each sequential frame since the background in consecutive frames is used as reference frame to estimate optical-flow of two consecutive frames. In order to reduce the computational cost, all cropped frames are finally resized to 80 by 80 pixels and converted to gray scale frames.

4.2 Optical flow extraction

As mentioned in section 3 optical flow of a video is calculated between two sequential frames. For instance, for frames $x_1, x_2, x_3, \dots, x_n$ that happened $t_1, t_2, t_3, \dots, t_n$ time, then the optical flow is calculated between x1-x2, x2-x3… xn-x1. We used vision.OpticalFlow (parameter,value); built in matlab function to call the optical flow object and step (opticalFlow, xn-1, xn) to calculate the optical flow between xn-1,and xn frames. In addition to these functions there are parameters and properties used to extract optical flow from two consecutive frames. Table1 shows the parameters and their values of optical flow algorithm.

Table 1: general parameters of optical flow algorithm

Parmeter	Values
Method	Horn-Schunck | Lucas-Kanade.
ReferenceFrameSource	Input port
OutputValue	Magnitude-squared

ISSN 1512-1232

a. Optical flow using Lucas-Kanade algorithm

As described in section3.3.2 Lucas–Kanade algorithm is the most latest and robust gradient based optical flow estimation technique. In addition to this, according to [20], the algorithm is good in estimating pixel displacement of a moving object. To calculate the optical flow using Lucas-Kanade we selected the method properties of as Lucas-Kanade and the set of parameter values of optical flow estimation algorithm as listed in table 1.

Table 2: parameters of the Lucas –Kanade optical flow algorithm

Parameter	**Value**
Method	'Lucas-Kanade'
TemporalGradientFilter	'Difference filter [-1 1]'
NoiseReductionThreshold:	0.0039

After using the algorithm, we calculated the squared magnitude of optical flow represented by an 80 by 80 square matrix. Figure 3 shows some images that have taken from after Lucas-Kanade optical flow estimation.

(a)

(b)

Figure 3: optical flow magnitude square calculated using (a) Lucas-Kanade (b) Horn-Schunck algorithm

b. Optical flow using Horn and Schunck algorithm

As describe in 3.3.1optical flow of a traditional dance video is calculated using Horn and Schunck algorithm from sequential frames. The results of Horn and Schunck algorithm is a 80-by-80 magnitude squared matrix. This squared matrix is then reshaped in order to use as training and testing dataset for classifiers. As an example, optical flow calculated using Horn and Schunck algorithm are shown in figure 3(b).

Table 3: parameters used for Horn-Schunck optical flow estimation

Parameter	**Value**
Method	'Horn-Schunck'
Smoothness	1
IterationTerminationCondition:	'Maximum iteration count'
MaximumIterationCount	10

4.3 Classification

In this study, ANN which is a popular and classical technique for classification is used. A classification usually involves with training dataset $\{x_k, y_k\} \in \{-1,1\}$ and testing dataset, where x_k contains the training features and y_k is the class labels. The goal a classifier is to produce a model which predicts the class labels based on the given feature values in the testing set. Since every machine learning algorithms have two phases (i.e. training and testing), the data set must be split into two, training and testing data set.

ISSN 1512-1232

4.3.3 Training Neural Network Classifier

Neural networks with multiple hidden layers can be useful for solving classification problems with complex data, such as images. Each layer can learn features at a different level of abstraction. However, training neural networks with multiple hidden layers can be difficult in practice. One way to effectively train a neural network with multiple hidden layers is to train each layer individually. It can be achieved by training a special type of network known as auto encoder for each desired hidden layer. First, individual hidden layers were trained in an unsupervised fashion using auto encoders. Next step was to create a softmax layer and to train it from the vectors of second auto encoders in a supervised fashion using labels for the training data.

Three separate components of a deep neural network have been trained in isolation. Then these layers are joined together to form a multilayer neural network. Neural network is created, and then configured the settings, and copy the weights and biases from the auto encoders and softmax layer. The results for the deep neural network could have been improved by performing back propagation on the whole multilayer network. The network was fine-tuned by retraining it on the training data in a supervised fashion. After this the final neural network model has been developed for ETVD classification as shown in figure 4.

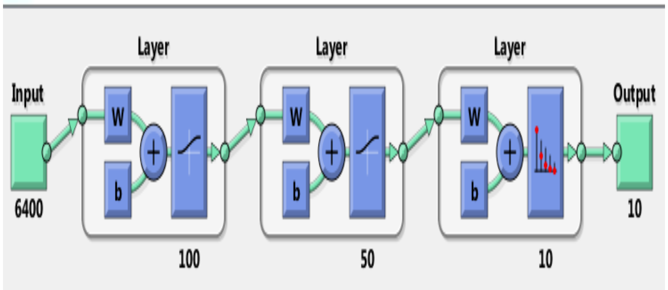

Figure 4: structure of ANN final model

The original vectors in the training data had 6400 dimensions. After passing them through the first auto encoder, this was reduced to 100 dimensions. After using the second auto encoder, this was reduced again to 50 dimensions. The final softmax layer is trained to classify these 50 dimensional vectors into ten different traditional dance video classes. The experiment was conducted under three scenarios as shown below.

4.3.2 Testing ANN Classifier

After creating a model, the performance of the proposed system is evaluated. Similar to training, testing was done in three phases based on the method of feature extraction. On the first phase, testing was done using features extracted using Lucas-Kanade algorithm; the second one is using optical flow estimated by Horn-Schunck algorithm; on the third using optical flow that calculated by both Lucas-Kanade and Horn-Schunck .Confusion matrix for performance evaluation of three phases that are taken from experiment is shown from figure 5 to figure 7 respectively.

Phase 1: testing using optical flow estimated by Lucas-Kanade algorithm

- Number of Frames used during training: 2700
- Number of Frames used during Testing: 300

In Figure 5, classification rate for each category is shown with black diagonal box and for each category with high misclassification rate is shown with dark gray box. Overall classification rate is 82.0%. This is achieved for motion features extracted from Lucas- Kanade method.

ISSN 1512-1232

	Afar	Benshangul	Gambella	Eskesta	Gurage	Hararghe	Oromo	Somali	Tigrinya	Wolaita
Afar	70.0	13.0	0.0	0.0	7.0	0.0	3.0	3.0	0.0	0.0
Benshangul	0.0	80.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0
Gambella	0.0	0.0	100	0.0	10.0	0.0	7.0	0.0	0.0	0.0
Eskesta	0.0	0.0	0.0	100	7.0	0.0	3.0	0.0	0.0	0.0
Gurage	0.0	0.0	0.0	0.0	30.0	0.0	3.0	3.0	0.0	0.0
Hararghe	0.0	1.0	0.0	7.0	20.0	100	7.0	0.0	0.0	0.0
Oromo	0.0	0.0	0.0	3.0	0.0	0.0	60.0	10.0	0.0	0.0
Somali	0.0	3.0	0.0	0.0	7.0	0.0	3.0	80.0	0.0	0.0
Tigrinya	0.0	0.0	0.0	0.0	20.0	0.0	3.0	0.0	100	0.0
Wolaita	0.0	3.0	0.0	7.0	0.0	0.0	7.0	3.0	7.0	100

Figure 5: Confusion Matrix for Classification results of ANN using optical flow estimated using Lucas-Kanade algorithm.

Phase 2: testing using optical flow estimated by Horn-Schunck algorithm:

- Number of Frames used during training: 2700
- Number of Frames used during Testing: 300

	Afar	Benshangul	Gambella	Eskesta	Gurage	Hararghe	Oromo	Somali	Tigrinya	Wolaita
Afar	83.0	27.0	0.0	0.0	3.0	0.0	7.0	10.0	0.0	0.0
Benshangul	3.0	40.0	0.0	0.0	0.0	0.0	0.0	13.0	0.0	0.0
Gambella	0.0	3.0	100	0.0	3.0	0.0	3.0	0.0	0.0	7.0
Eskesta	0.0	7.0	0.0	70.0	0.0	0.0	10.0	0.0	0.0	0.0
Gurage	0.0	0.0	0.0	0.0	37.0	0.0	7.0	0.0	3.0	3.0
Hararghe	0.0	0.0	0.0	0.0	17.0	100	0.0	0.0	0.0	0.0
Oromo	0.0	0.0	0.0	0.0	7.0	0.0	63.0	0.0	0.0	7.0
Somali	13.0	17.0	0.0	30.0	3.0	0.0	0.0	73.0	0.0	0.0
Tigrinya	0.0	3.0	0.0	0.0	23.0	0.0	7.0	0.0	97.0	0.0
Wolaita	0.0	3.0	0.0	0.0	7.0	0.0	3.0	3.0	0.0	83.0

Figure 6: Confusion Matrix for Classification results of ANN using optical flow estimated using Horn-Schunck algorithm:

The Confusion matrix in figure 6 is derived from the ANN model that trained and tested with optical flow extracted from Horn Schunck method. It shows correct classification rate for each category with black diagonal box and misclassification is highlighted with dark gray box. Overall classification accuracy is 74.6%.

Phase 3: testing using optical flow estimated by both Lucas-Kanade and Horn-Schunck algorithm

- Number of Frames used during training: 5400
- Number of Frames used during Testing: 600

	Afar	Benshangul	Gambella	Eskesta	Gurage	Hararghe	Oromo	Somali	Tigrinya	Wolaita
Afar	50.0	35.0	42.0	2.0	35.0	17.0	12.0	7.0	15.0	5.0
Benshangul	0.0	52.0	2.0	0.0	0.0	0.0	3.0	10.0	0.0	20.0
Gambella	0.0	0.0	52.0	0.0	5.0	0.0	7.0	3.0	2.0	0.0
Eskesta	0.0	030	0.0	47.0	0.0	0.0	5.0	0.0	0.0	0.0
Gurage	0.0	0.0	0.0	0.0	10.0	0.0	3.0	3.0	0.0	0.0
Hararghe	0.0	0.0	2.0	0.0	35.0	83.0	18.0	0.0	12.0	2.0
Oromo	0.0	0.0	0.0	22.0	0.0	0.0	37.0	2.0	3.0	13.0
Somali	50.0	3.0	0.0	30.0	0.0	0.0	5.0	73.0	2.0	3.0
Tigrinya	0.0	0.0	0.0	0.0	12.0	0.0	7.0	0.0	67.0	0.0
Wolaita	0.0	0.0	3.0	0.0	3.0	0.0	3.0	2.0	0.0	57.0

ISSN 1512-1232

Figure 7: Confusion Matrix for Classification results of ANN using of optical flow estimated using by both Lucas-Kanade and Horn-Schunck algorithm:

The Confusion matrix in figure 7 is derived from the ANN model that trained and tested with optical flow extracted from both Lucas- Kanade and Horn Schunck method. It shows correct classification rate for each category with black diagonal box and misclassification is highlighted with dark gray box. Overall classification accuracy is 52.8%.

5. Conclusion and recommendation

5.1. Conclusion

Classification was done using back propagation ANN in unsupervised and supervised way. Training and testing were done in three phases based on the type of optical flow algorithm used.

The performance of the system at the end is evaluated using back propagation ANN algorithm in three phases. The result of each classifier is, 82.0%, 74.6%, and 52.8% using Lucas-Kanade, Horn-Schunck and both Lucas-Kanade and Horn-Schunck methods respectively. This result implies that Ethiopian traditional dance video classification using Lucas-Kanade method produces a high performance as compared to classification using both Lucas-Kanade and Horn-Schunck methods. This is mainly due most Ethiopian traditional dances focus on both on local and global motions and the capability of Lucas-Kanade method to handle both motions. The low accuracy by both Lucas-Kanade and Horn-Schunck methods shows that the inter similarity between feature extracted by Lucas-Kanade from one class with optical flow extracted by Horn-Schunck method from another class.

On the other hand, for some category, there is a high correlation between traditional dance styles. For instance, from the confusion matrix result traditional dance of Gurage highly similar with Hararghe and Tigrinya.

In this study, in addition to well-known challenges of Action recognition from video, there are challenges that are unique to traditional dance video classification. Since we used the optical flow of a dancer for recognizing a particular traditional dance, factors such as traditional Clothes, traditional women’s hair style and traditional objects often increased the percentage of misclassification.

5.2. Recommendation

This work introduces our steps for classification of the traditional dance from a video sequence. This highly applicable topic still in its infancy and much is left to be done. This work is extensible in many ways, such as:

- The overall accuracy of the best algorithm is 80.0% which is not that satisfactory for real time application so there is a need to other optical flow algorithms more robust than it.
- We only use the magnitude of optical flow, scale and direction of optical flow can be tested.
- During dataset preparation we convert each video clip into sequential frames to calculate optical flow; video clips can be used to calculate optical flow.

REFERENCE

[1] Jimmy Dunn.”To Dance in Ancient Egypt". Available: http://www.touregypt.net/featurestories/dance.htm [Accessed:January 14, 2015].

[2] Md. Atiqur Rahman Ahad. “Computer Vision and Action Recognition: A Guide for Image Processing and Computer Vision Community for Action Understanding”. Paris, Atlantis of Paris, 2011.

[3] Timkehet Teffera."THE ROLE OF TRADITIONAL MUSIC AMONG EAST AFRICAN SOCIETIES: THE CASE OF SELECTED AEROPHONES". in: Proceedings from the 16th

ISSN 1512-1232

International meeting on on Folk Musical Instruments , Martin Luther University Halle-Wittenberg, Germany, 2006

[4] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 18, No 6,pp. 1473-1488, 2008.

[5] Liang Wang, Li Cheng, Guoying Zhao. “Machine Learning for Human Motion Analysis: a theory and Practices”. United States of America, IGI Global, 2010.

[6] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," in Computer Vision Proceedings Ninth IEEE International Conference on, 2003, pp. 726-733.

[7] H. Jhuang, T. Serre, L. Wolf, and T. Poggio, "A biologically inspired system for action recognition," in Computer Vision, IEEE 11th International Conference on, 2007, pp. 1-8.

[8] K. Schindler and L. Van Gool, "Combining densely sampled form and motion for human action recognition," in Pattern Recognition, ed: Springer, 2008, pp. 122- 131.

[9] D. M. Gavrila, "The visual analysis of human movement: A survey," Computer vision and image understanding, vol. 73, No 4, pp. 82-98, 1999.

[10] R. Poppe, "A survey on vision-based human action recognition," Image and vision computing, vol. 28, No 8,pp. 976-990, 2010.

[11] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition," Computer vision and image understanding, vol. 115, No 8, pp. 224-241, 2011.

[12] Upal Mahbub,Hafiz Imtiaz,and Md. Atiqur Rahman Ahad, “An Optical Flow-based Action Recognition Algorithm”, in Upal Mahbub, 2014.

[13] Julia Moehrmann ; Xin Wang and Gunther Heidemann"Motion based situation recognition in group meetings", Proc. SPIE 7538,Image Processing: Machine Vision Applications III, 75380N(January 28, 2010);doi:10.1117/12.838963; http://dx.doi.org/10.1117/12.838963.

[14] M. Nixon and A. S. Aguado, “Feature extraction & image processing:” Access Online via Elsevier, 2008.

[15] E.P. Simoncelli, “Design of multi-dimensional derivative filters”,IEEE Int. Conf. Image Processing”, Vol.1, pp.790-793,1994.

[16] B.D. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision”, DARPA Image Understanding Workshop, 1981, pp.121–130 (see also IJCAI’81, pp.674–679).

[17] B.K.P. Horn and B.G. Schunck,“Determining Optical Flow”, Artificial Intelligence, 17, 1981, pp.185–204.

[18] Tinku Acharya and Ajoy K. Ray, Image Processing Principles and Applications, Jhon Wiley, 2005.

[19] Coolen, A. C. C., R. Kühn and P. Sollich:theory of Neural information processing, Systems, Oxford University Press, 2005.

[20] Othman, M.; Abdulrahim, K. ; Zulkifli, Z. ,”Analysis of Motion Detection of Breast Tumor Based on Tissue Elasticity From B Mode Ultrasound Images Using Gradient Method Optical Flow Algorithm”, in Artificial Intelligence, Modelling and Simulation (AIMS), IEEE 1st International Conference on, 2013, pp.278-283.

Article received: 2015-12-18