Video Genre Classifiaction Using MARS based on Factor Analysis Method

Hnin Ei Latt

Faculty of Information and Communication Technology

University of Technology (Yatanarpon Cyber City)

Pyin Oo Lwin, Myanmar

hnineilatt@gmail.com

Nu War

Faculty of Information and Communication Technology

University of Technology (Yatanarpon Cyber City)

Pyin Oo Lwin, Myanmar

nuwar81@gmail.com

Abstract *:*

Video genre classification, video contents retrieval and semantics research are very attractive for many researchers in video processing and analysis domain. Many researchers try to propose structure or frameworks to classify the video genre that's integrating many algorithms using low and high level features. Features generally include both useful and useless information that are difficult to separate. In this paper, video genre classification is proposed by using only the audio channel. A decomposition model is based on multivariate adaptive regression splines to separate useful and useless components and the genre identification is performed on these low-level acoustic features such as MFCC and timbral textual features. Factor Analysis is proposed to reduce feature dimension in this system. For comparison, it is implemented the three feature dimension methods such as Principal Component Analysis, Stochastic Proximity Embedding algorithm, Stochastic Neighbor Embedding . MARS method is used as classifier, to classify the genre type,. Experiments are conducted on a corpus composed from Cartoon, Sport, News, Dahma and Music.

Keywords-component; *Principal Component Analysis, Stochastic Proximity Embedding algorithm, Stochastic Neighbor Embedding, Factor Analysis*

I. Introduction

Recently, advances in technology have supported the storage of large amount of data. Therefore, fast information retrieval became a hot challenge. One of the most interesting applications concerns with social network users, which have looked forward to meet people that share common preferences, and also to discover such as new videos music, cartoons and so on. Most of the research on video classification has the intent of classifying an entire video, while some authors have focused on classifying segments of video such as identifying violent [1] or scary [2] scenes in a movie or distinguishing between different news segments within an entire news broadcast [3]. For the purpose of video classification, features are drawn from three modalities: text, audio, and visual. Most of the proposed approaches rely on image analysis. In [4], many works are motivated by the critical need of efficient tools for structuring audiovisual databases these last years. In [2], they investigate higher level analysis like tracking of audiovisual events. Audio-based approaches were explored by automatic transcription of speech contents, or by low level audio stream analysis.

Acoustic-space characterization is presented by using statistic classifier like gaussian mixture model (GMM), neural nets or support vector machines (SVM) on cepstral domain features [5, 6, 7]. Various kinds of acoustic features have been evaluated in the field of video genre identification. In [8, 7, 5], time-domain audio features are proposed like zero crossing rates or energy distributions. Therefore, low-level approaches present a better robustness to the highly variable and unexpected conditions that may be encountered on videos. In the cepstral domain, one of the main difficulties in genre identification is due to the diversity of the acoustic patterns that may be produced by each video genre. In this paper, this problem is aim to address in the field of identifying video genres by applying multivariate adaptive regression splines. Video genre classification framework is focused on by using an audio-only method.

In the next section an overview of the presented system is provided first. The architecture of the system and the basic underlying concepts are explained. Secondly, feature extraction methods are explained. Thirdly, feature reduction method is described. Fourthly, the multivariate adaptive regression splines algorithm is described. Finally, the experimental results are also shown in figure1,2,3,4 and 5.

II. SYSTEM ARCHITECTURE

The overall procedure to extract audio file from an video clip has shown in Figure.1. Firstly, base audio features are extracted from the audio signal. The MFCC, zero crossing rate, short time energy, spectral flux, spectral centroid, spectral rolloff, noise frame ratio and silence ratio are used as the base audio features in this paper. Secondly, Factor Analysis method is used to reduce the feature dimension. And then, classifier such as multivariate adaptive regression splines develop the model for each genre types by using the base audio features set. The next step is an efficient mechanism for classifying genre in the database and measuring their performance. The details of the proposed audio fingerprint are explained in figure 1.

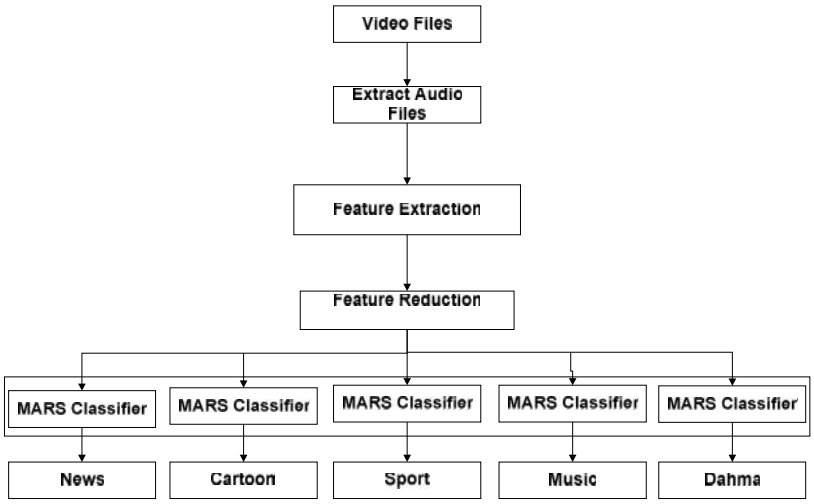

Figure 1. *Overview of system architecture*

III. FEATURE EXTRACTION

Many of the audio-based features are chosen to approximate the human perception of sound. In this frame work uses low-level acoustic features that are both time-domain features and frequency-domain features. The

timbral textual features are calculated from the given audio signal. Timbral textual features are those used to differentiate mixture of sounds based on their instrumental compositions when the melody and the pitch components are similar. The use of timbral textural features originates from speech recognition. Extracting timbral features require preprocessing of the sound signals. The signals are divided into statistically stationary frames, usually by applying a window function at fixed intervals. The application of a window function removes the so-called "edge effects." Popular window functions including the Hamming window function. Short-term fourier transform feature is a set of features related to timbral textures and is also captured using MFCC. It consists of Spectral Centroid, Spectral Rolloff, Spectral Flux and Low Energy, Zero Crossings and then computes the mean for all five. In the time-domain, Zero crossing rate (ZCR) is the number of signal amplitude sign changes in the current frame. Higher frequencies result in higher zero crossing rates. Speech normally has a higher variability of the ZCR than in music. If the loudness and ZCR are both below thresholds, then this frame may represent silence. The silence ratio is the proportion of a frame with amplitude values below some threshold. Speech normally has a higher silence ratio than music. News has a higher silence ratio than commercials. In the frequency-domain, the energy distribution (short time energy) is the signal distribution across frequency components. The frequency centroid, which approximates brightness, is the midpoint of the spectral energy distribution and provides a measure of where the frequency components are concentrated. Normally brightness is higher in music than in speech, whose frequency is normally below 7 kHz. Bandwidth is a measure of the frequency range of a signal. Some types of sounds have more narrow frequency ranges than others. Speech typically has a lower bandwidth than music. The fundamental frequency is the lowest frequency in a sample and approximates pitch, which is a subjective measure. Mel-frequency cepstral coefficients (MFCC) are produced by taking the logarithm of the spectral components and then placing them into bins based upon the Mel frequency scale, which is perception-based. So, there are a total of twenty features. In the next section, it is described the feature dimension reduction method in the proposed system.

IV. FACTOR ANALYSIS

In the proposed system, factor analysis is used to reduce the feature dimension. Factor analysis is a multivariate method used for data reduction purposes. To describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors, it is very useful statistical method. Factor analysis searches for joint variations in response to unobserved latent variables. The observed variables are modeled as linear combinations of the potential factors and error. To reduce the set of variables in a dataset, the information gained about the interdependencies between observed variables can be used later. This method is computationally equivalent to low rank approximation of the matrix of observed variables. Factor analysis originated in psychometrics, and is used in behavioral sciences, social sciences, marketing, product management, operations research, and other applied sciences that deal with large quantities of data. It is related to principal component analysis (PCA), but the two are not identical.

The basic idea is to represent a set of variables by a smaller number of variables. In this case they are called factors. These factors can be thought of as underlying constructs that cannot be measured by a single variable. This model can be written as follows.

$$Yj = sj1F1 + sj2F2 + \ldots + sjmFm + ej \quad (1)$$

where the sj s are the factor loadings (or scores) for variable j and ej is the part of variable Yj that cannot be 'explained' by the factors. If p variables (Y1,Y2, . . . ,Yp) is need to measure on a sample of n subjects, then variable j can be written as a linear combination of m factors (F1, F2,…, Fm) where, as explained above m < p. In this method, the three major steps are calculation the initial factor loadings, factor rotation, and calculation of factor scores.

For calculating the initial factor loadings, the two most common methods are principal component method and principal axis factoring. In this system, a principal axis factoring method tries to find the lowest number of factors which can account for the variability in the original variables that is associated with these factors (this is in contrast to the principal components method which looks for a set of factors which can account for the total variability in the original variables).

The second step is factor rotation. In this step, the initial factor loadings have been calculated, the factors are rotated. This is done to find factors that are easier to interpret. If there are groups of variables which subgroups of variables that are strongly interrelated and then the rotation is done to try to make variables within a subgroup score as highly (positively or negatively) as possible on one particular factor while, at the same time, ensuring that the loadings for these variables on the remaining factors are as low as possible. In other words, the

object of the rotation is to try to ensure that all variables have high loadings only on one factor. There are two types of rotation method, orthogonal and oblique rotation. In orthogonal rotation the rotated factors will remain uncorrelated whereas in oblique rotation the resulting factors will be correlated. There are a number of different methods of rotation of each type. In [10], the most common orthogonal method is called varimax rotation.

In the final step, it is calculated the factor scores. When calculating the final factor scores (the values of the m factors, (F1, F2,…, Fm), for each observation), a decision needs to be made as to how many factors to include. This is done using the scree plot of the eigenvalues. This will indicate whether there is an obvious cut-off between large and small eigenvalues. Classification method is explained in the next section.

V. MULTIVARIATE ADAPTIVE REGRESSION SPLINES

In this video genre classification step, MARS is used as a classifier. Analyses were performed using multivariate adaptive regression splines, a technique that uses piece-wise linear segments to describe non-linear relationships between audio features and video genre. The theory of multivariate adaptive regression splines (MARS) was developed by Jerome Friedman [9] in 1991. Let z be the dependent response, which can be continuous or binary, and let $Y = (Y_1, ..., Y_n) \in D \in \Re^n$ be the set of potential predictive covariates. Then the system assume that the data are generated from an unknown "true" model. In case of a continuous response this would be $z = f(Y_1, Y_2, ..., Y_n) + e$ (2)

The distribution of the error e is member of the exponential family [1]. f is approximated by applying functions, which include interactions of at most second order. That means that use the model

$$f(Y) = g_0 + \sum_{j1} g_{j1}(Y_{j1}) + \sum_{j1<j2} g_{j1,j2}(Y_{j1}, Y_{j2}) + e \quad (3)$$

whereas $e \sim N(0, \pm^2)$ with error variance $\pm^2 \in \{0.25.4\}$. Linear splines and their tensor products are used to model the function g(.). A one-dimensional spline can be written as

$$g(y) = b_{-1} + b_0 y + \sum_{k=1}^{K} b_k (y - t_k)_+ \quad \backslash \quad (4)$$

and the knot tk in the range of the observed values of Y. For this reason the function g is situated in a linear space with the K + 2 basis functions. Thus the following model results: $g_0 = \ _0$, $g_{j_1}(Y_{j_1}) = \sum_{i=1}^{M} \ _i^{j_i} B_i^{j_i}(Y_{j_i})$ and then $g_{j_1,j_2}(Y_{j_1}, Y_{j_2}) = \sum_{i=1}^{M} \ _i^{j_1 j_2} B_i^{j_1 j_2}(Y_{j_1}, Y_{j_2})$ because the interaction gj1,j2 is modeled by means of tensor product splines as $g_{12}(y_1, y_2) = g_1(y_1) \times g_2(y_2)$ .The M represent the number of basis functions in the model and the Bs represent spline basis functions as described above and the βs are coefficients. In this approach the coefficients are estimated by using the Least Squares method. Now the coefficient matrix can be written as $\hat{} = (Y^{*T} Y^*)^{-1} Y^{*T} Z.$ Y is represented as the design matrix of the selected basis functions, and Z represents the response vector. Instead of Yj, MARS uses a collection of new predictors in the form of piecewise linear basis functions are as $\{(Y_j - t)_+, (t - Y_j)_+\}, \quad j = 1,..., n, \quad t \in \{y_{1j},.., y_{Nj}\}$. After that, the generalized cross-validation criterion is used to measure the degree of fit or lack of accuracy of the model :

$$GCV(M) = \frac{\frac{1}{N}\sum_{i=1}^{N}[z_i - \hat{f}_M(y_i)]^2}{[1 - \frac{d.M}{N}]^2} \quad (5)$$

whereas $\hat{f}$ denotes the fitted values of the current MARS model and d denotes the penalizing parameter. The numerator is the common residual sum of squares, which is penalized by the denominator, which accounts for the increasing variance in the case of increasing model complexity. A smaller d generates a larger model with more basis functions, a larger d creates a smaller model with less basis functions.

According to the table.1, Forward Process Stage is that the stepwise addition process basis functions are added until the maximal allowed model size is reached. The largest model generally overfits the data. Then Backward pruning Stage -the stepwise deletion process- is that all 'unnecessary' basis functions are removed again until a final model is obtained which is best considering the GCV that is the one with the minimum GCV. In the first step of the addition process a constant model is fitted. Subsequently the number of candidate basis

functions depends on the number of possible knots per predictor variable. To keep the procedure fast, the results robust the number of possible knots per predictor and also the possible candidates per step are limited. To determine the number of potential knots of a specific covariate an order statistic is computed and a subset of it is then chosen as potential knots. Commonly these are about 20 knots per predictor, at most every third value is chosen yet. In the first iteration – after the fit of the constant model – a linear basis function on one of the predictor variables is fitted. The second iteration is accounted for both linear basis functions on another covariate and basis functions with knots of the covariate already in the model.The model to choose in every step during the forward process is the one out of all possible models which minimizes the GCV. In the backward process one basis function is deleted per step and the GCV is computed for the reduced model. The model which yields the smallest increase of GCV becomes the new one.

VI. EXPERIMENT I

To validate the efficiency of the proposed system, the five categories are selected that are commonly targeted by video genre classification tasks: News, Cartoon, Sport, Music and Dahma. The system is composed of 1096 videos with duration about 5 minutes. Among them, 203, 235, 215, 236, and 212 clips are ‘News’, ’Dahma’, ’Cartoon’, ’Sport’ and ‘Music’ respectively. Using the cross validation method, three classification schemes are implemented on Matlab platform. For comparison, SVM, KNN and MARS classifiers are proposed in this paper. In figure 1, each model of five genre type by using MARS method is explored. The accuracy results of the proposed system is shown in figure 2.

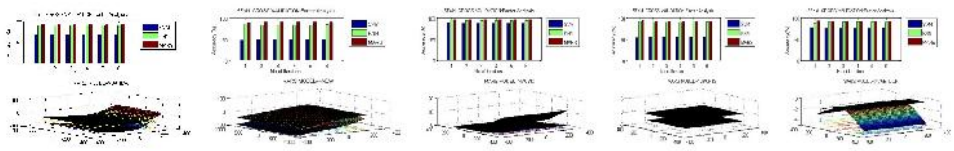

Figure 2. Classification Accuracy and MARS model with Factor Analysis

VII. EXPERIMENT II

For feature dimension reduction method, three methods are demonstrated for comparison: (1) PCA (2) Stochastic Proximity Embedding algorithm (3) Stochastic Neighbor Embedding.

a. Stochastic Proximity Embedding algorithm

Stochastic proximity embedding applies a self-organizing scheme that attempts to bring each individual stress (dij - rij)2 rapidly to zero. The method starts with an initial configuration iteratively refines it by repeatedly selecting two points at random, and adjusting their coordinates so that their Euclidean distance on the map dij matches more closely their corresponding proximity rij.

The correction is proportional to the disparity $\frac{|r_{ij}-d_{ij}|}{d_{ij}}$, where λ is a learning rate parameter that decreases during the course of the refinement to avoid oscillatory behavior. The implemented algorithm is as follows:

Step 1. Initialize the coordinates xi. Select an initial learning rate λ.

Step 2. Select a pair of points, i and j, at random and compute their distance $d_{ij} = \|x_i - x_j\|$. If $d_{ij} \neq r_{ij}$, update the coordinates xi and xj by:

$$x_i \leftarrow x_i + \frac{\lambda}{2}\frac{r_{ij} - d_{ij}}{d_{ij} + \varepsilon}(x_i - x_j) \quad (6)$$

and

$$x_j \leftarrow x_j + \frac{\lambda}{2}\frac{r_{ij} - d_{ij}}{d_{ij} + \varepsilon}(x_j - x_i) \quad (7)$$

where is a small number to avoid division by zero.

Step 3 . Repeat Step 2 for a prescribed number of steps S.

Step 4 Decrease the learning rate $\lambda$by prescribed decrement $\delta\lambda$.

Step 5 . Repeat Step 2– Step 4 for a prescribed number of cycles C.

In figure 3, the classification accuracy is described for five genre types. This results is based on the database from the experiment I.

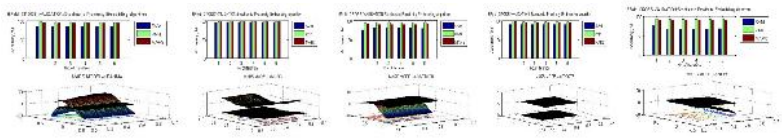

Figure 3. Classification Accuracy and MARS model with Stochastic Proximity Embedding algorithm

b. Stochastic Neighbor Embedding

Denote by xt $\mathbb{R}^D$ an object showed by a D-dimensional vector. The vector $\vec{x}$ $\mathbb{R}^{DN}$ is a long vector that is constructed by stacking {x1, . . . , xN } in a single column. The image of xt is explained by y_t $\mathbb{R}^d$ (d D) and the vector $\vec{y}$ $\mathbb{R}^{dN}$ is constructed in a similar manner. The original algorithm is implemented below.

Step 1. Select neighbors by ϵ neighborhoods or nearest neighbors.

Step 2. Compute the probability, pij , that xi would pick xj as its neighbor:

$p_{ij} = \frac{exp\left(-d_{ij}^2\right)}{\sum_{k \neq i} exp\left(-d_{ik}^2\right)}$ where dij2 are dissimilarities between two objects xi and xj in the high-dimensional space and σi is a Gaussian kernel width usually set by hand. The dissimilarities are computed by the scaled Euclidean distance $d_{ij}^2 = \frac{\|x_i - x_j\|^2}{2\sigma_i^2}$

In the lower-dimensional space, the induced probability qij (with a fixed variance) that the image yi pick yj as its neighbor, is described by

$$q_{ij} = \frac{exp\left(-\|y_i - y_j\|^2\right)}{\sum_{k \neq i} exp(\|y_i - y_k\|^2)} \quad (8)$$

Step 3. The aim of the embedding is to match pij and qij as well as possible. This is achieved by minimizing a cost function which is a sum of Kullback-Leibler divergences between pij and qij for each object. The cost function is given by

$$J = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (9)$$

Step 4. The set of images, ~ y in the lower-dimensional space, are updated by a gradient-descent method which has the form

$$\vec{y}^{(k+1)} = \vec{y}(k) - \eta(k) \nabla J(k) \quad (10)$$

Where η(k) is a learning rate and the gradient J is given by

$$\nabla J = \left[\left(\frac{\partial J}{\partial y_1}\right)^T, \ldots, \left(\frac{\partial J}{\partial y_N}\right)^T\right]^T \quad (11)$$

$$\frac{\partial J}{\partial y_i} = 2 \sum_j (y_i - y_j)(p_{ij} - q_{ij} + p_{ji} - q_{ji}) \quad (12)$$

The accuracy results of three classifiers with Stochastic Neighbor Embedding are shown in figure 4. It is tested with the database from the experiment I.

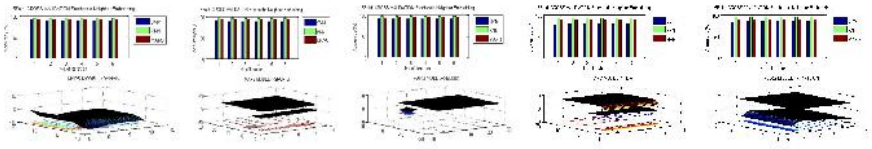

Figure 4. Classification Accuracy and MARS model with Stochastic Neighbor Embedding algorithm

c. Principal Component Analysis

Principal component analysis (PCA) is a classical statistical method. This linear transform has been widely used in data analysis and compression. Principal component analysis is based on the statistical representation of a random variable. A random vector population x has supported, where

$$x = (x_1, \dots, x_n)^2$$

and the mean of that population is denoted by

$$\mu_2 = E\{x\}$$

and the covariance matrix of the same data set is

$$C_x = E\{(x - \mu_x)(x - \mu_x)^T\} \tag{13}$$

The components of C_x, denoted by C_{ij}, represent the covariances between the random variable components x_i and x_j. The component C_{ji} is the variance of the component x_i. The variance of a component indicates the spread of the component values around its mean value. If two components x_i and x_j of the data are uncorrelated, their covariance is zero ($C_{ij} = C_{ji} = 0$). The covariance matrix is, by definition, always symmetric.

From a sample of vectors $x_1, \dots, x_M$, calculating the sample mean and the sample covariance matrix as the estimates of the mean and the covariance matrix.

From a symmetric matrix such as the covariance matrix, determining an orthogonal basis by finding its eigenvalues and eigenvectors. The eigenvectors e_i and the corresponding eigenvalues λ_i are the solutions of the equation

$$C_x e_i = \lambda_i e_i, i = 1, \dots, n \tag{14}$$

Assuming for simplicity that the λ_i are distinct. These values can be found, for example, by finding the solutions of the characteristic equation

$$|C_x - \lambda I| = 0 \tag{15}$$

where the I is the identity matrix having the same order than C_x and the |.| denotes the determinant of the matrix. If the data vector has n components, the characteristic equation becomes of order n. This is easy to solve only if n is small. Solving eigenvalues and corresponding eigenvectors is a non-trivial task, and many methods exist. One way to solve the eigenvalue problem is to use a neural solution to the problem . The data is fed as the input, and the network converges to the wanted solution.

By ordering the eigenvectors in the order of descending eigenvalues (largest first), one can create an ordered orthogonal basis with the first eigenvector having the direction of largest variance of the data. In this way, finding directions in which the data set has the most significant amounts of energy.

Suppose one has a data set of which the sample mean and the covariance matrix have been calculated. Let A be a matrix consisting of eigenvectors of the covariance matrix as the row vectors.

By transforming a data vector x, $y = A(x - \mu_x)$

which is a point in the orthogonal coordinate system defined by the eigenvectors. Components of y can be seen as the coordinates in the orthogonal base. Reconstructing the original data vector x from y by

$$x = A^T y + \mu_x \tag{16}$$

using the property of an orthogonal matrix $A^{-1} = A^T$.The A^T is the transpose of a matrix A. The original vector x was projected on the coordinate axes defined by the orthogonal basis. The original vector was then reconstructed by a linear combination of the orthogonal basis vectors.

Instead of using all the eigenvectors of the covariance matrix, representing the data in terms of only a few basis vectors of the orthogonal basis. If we denote the matrix having the K first eigenvectors as rows by A_K, creating a similar transformation as seen above

$$y = A_K(x - \mu_x) \tag{17}$$

$$x = A_K{}^T y + \mu_x \tag{18}$$

This means that we project the original data vector on the coordinate axes having the dimension K and transforming the vector back by a linear combination of the basis vectors. This minimizes the mean-square error between the data and this representation with given number of eigenvectors.

The classification accuracy is tested by using the database of the experiment I. Using this PCA method, the results are shown in figure 5.

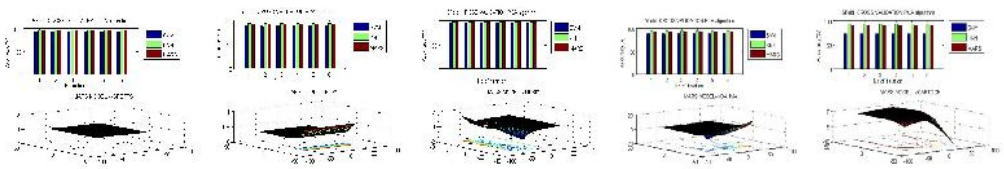

Figure 5. *Classification Accuracy and MARS model with PCA algorithm*

TABLE I. THE COMPARISON OF PRECISIONS OF THE THREE CLASSIFICATION SCHEMES COMBINED WITH THE FOUR DIMENSION REDUCTION METHODS IN FIVE CLASSIFICATION TASKS

Classifiers	**SVM**				**KNN**				**MARS**			
Feature Reduction method	***FA***	***SPE***	***SNE***	***PCA***	***FA***	***SPE***	***SNE***	***PCA***	***FA***	***SPE***	***SNE***	***PCA***
News	44.03	68.85	78.56	86.57	85.15	92.08	92.34	92.21	91.63	91.00	88.32	90.82
Cartoon	76.19	79.55	85.23	72.03	90.11	94.22	94.09	93.73	91.63	90.51	89.70	89.48
Sport	55.07	88.85	88.09	92.17	89.57	96.46	96.24	95.97	92.08	94.72	93.64	95.74
Music	87.51	95.07	90.24	94.85	94.22	96.64	96.86	96.64	96.10	96.10	94.36	97.31
Dahma	67.87	84.65	89.30	88.81	90.42	94.58	94.09	93.28	93.46	93.33	92.48	92.61

In the experiment, SVM-based classification, KNN-based classification and MARS-based classification are tested with the four feature reduction methods for the five genre types. According to the accuracy result from Table.1, for music video genre classification, MARS with PCA is the highest accuracy. In the News, Cartoon, Dahma and Sport video genre classification, KNN with Stochastic Neighbor Embedding is highest accuracy rate but computation time of stochastic reduction method is longer than that of other reduction methods. However, using only factor analysis, the average accuracy rate of MARS, KNN and SVM are 92.95, 89.89, and 66.13 respectively. Thus, MARS classifier is better accuracy than KNN and SVM based on Factor Analysis.

VIII. CONCLUSION

Among these five genres, Music is the best and News is hard for all three classifiers. From the experimental results, Factor Analysis-based MARS video genre classifier obtains the classification rate of about 92.95% significantly better than Factor Analysis-based SVM and Factor Analysis-based KNN. The experimental evaluation of this proposed system confirms the good performance of video classification system except News but it is still reasonable results. Further experiments on larger volume of audio, audio-visual features will be tested in this framework. Integrating genetic algorithms, with neural networks are possible research directions in further improving the classification accuracies. As this studies mainly use demographic variables as independent variables, future studies may aim at collecting more important variables to improve the classification accuracies.

ACKNOWLEDGMENT (HEADING 5)

My Sincere thanks to my supervisor Dr. Nu War, for providing me an opportunity to do my research work. I express my thanks to my Institution namely University of Technology (Yatanarpon Cyber City) for providing me with a good environment and facilities like Internet, books, computers and all that as my source to complete this research work. My heart-felt thanks to my family, friends and colleagues who have helped me for the completion of this work.

REFERENCES

1. J. Nam, M. Alghoniemy, and A. H. Tewfik, "Audio-visual content-based violent scene characterization," in International Conference on Image Processing (ICIP '98), vol. 1, 1998, pp. 353–357.

2. S. Moncriefi, S. Venkatesh, and C. Dorai, "Horror film genre typing and scene labeling via audio analysis," in Multimedia and Expo, 2003, 2003.

3. W. Zhu, C. Toklu, and S.-P. Liou, "Automatic news video segmentation and categorization based on closed-captioned text," in Multimedia and Expo, ICME, 2001. pp. 829–832

4. D. Brezeale and D. J. Cook, "Automatic video classification : A survey of the literature," in Systems, Man, and Cybernetics, 2008.

5. M. Roach, L.-Q. Xu, and J. Mason, "Classification of non-edited broadcast video using holistic low-level features," in (IWDC'2002), 2002.

6. R. Jasinschi and J. Louie, "Automatic tv program genre classification based on audio patterns," in Euromicro Conference, 2001, 2001.

7. L.-Q. Xu and Y. Li, "Video classification using spatial-temporal features and pca," in Multimedia and Expo, (ICME '03), 2003.

8. M. Roach and J. Mason, "Classification of video genre using audio," in European Conference on Speech Communication and Technology, 2001.

9. Friedman, J.H., "Multivariate adaptive regression splines". Ann. Stat. 19, 1–141 (with discussion) 1991.

10. Manly, B.F.J., "Multivariate Statistical Methods", A primer, Third edition, Chapman and Hall, 2005.