

Comparison Analysis On Supervised Learning Based Solutions For Sports Video Categorization

Min Xu, Mira Park, Suhui Luo, and Jesse S. Jin

School of Design, Communication and IT

University of Newcastle, Callaghan 2308, NSW, AUSTRALIA

Abstract—Due to the wide viewer-ship and high commercial potentials, recently, sports video analysis attracts extensive research efforts. One of the main tasks in sports video analysis is to identify sports genres i.e. sports video categorization. Most of the existing work focus on mapping content-based features to sports genres by using supervised learning methods. Moreover, video data sets seeks efficient data reduction methods due to the large size and noisy data. It lacks comparison analysis on the implementation and performance of these methods. In this paper, the research is carried out by using four dominant machine learning algorithms, namely Decision Tree, Support Vector Machine, K Nearest Neighbor and Naive Bayesian, and comparing their performance on a high dimensional feature set which selected by some feature selection tools such as Correlation-based Feature Selection (CFS), Principal Components Analysis (PCA) and Relief. Experimental results shows that Support Vector Machine (SVM) and k-NN are not sensitive to reduction of training sets. Moreover, three different feature reduction methods perform very differently with respect to four different tools.

I. INTRODUCTION

With substantial progress in the Internet and multimedia technology, searching and indexing for multimedia content becomes more and more necessary and urgent. Recently, image and video categorization has been a focus of image and video content search, more and more strategies involving learning a supervised model are emerging. When there are clearly identified categories, as well as, large domain-representative training data, learning can be effectively employed to construct a model of the domain. The researchers work on the semantic image or video classification have typically used color, texture, objects etc. as features for mapping to high level concepts by learning through K Nearest Neighbor [1], Rule-based systems [2], Linear discriminant analysis [3], Vector quantization [4], Decision trees [5] and Support vector machine [1].

Nowadays, as a large portion of commercial videos, sports video attracts more and more research efforts due to the wide viewer-ship and high commercial potentials. Domain-specific knowledge is more or less used for sports analysis, which depends on sports genres. Therefore, sports categorization become more and more important and also the pre-step for further sports analysis [6], [7], [8], [9], [10], [11]. In [8], KNN is used to train edge features to sports genre. [9] used motion and color features to learn Hidden Markov Models (HMM) for sports video categorization. Support Vector Machines is used in [11].

Although machine learning algorithms are widely used in sports video categorization, it lacks comparison analysis

on the performance of those algorithms, when particularly used for video data. Normally, video data set has a large size and requires high dimensional features to represent the characteristics. In order to improve learning performance and avoid very much time consuming, data reduction methods are widely used in Multimedia field to reduce feature dimensions as well as to find the efficient features. Therefore, comparison analysis on the combination of machine learning and data reduction is necessary.

In this paper, we look into several aspects of video classification using decision tools over a domain of four classes, which is a good platform for comparison, especially because most of the tools that have been used in this field. The challenges inherent in the development of a video categorization system include among other: 1) Forming close association between descriptor space and meaningful classes; 2) Performing automatic meaningful dimension evaluation on only a few relevant dimensions; 3) Dealing with large dimensionality of the descriptor space in an effective and efficient manner; 4) Providing co-existent labels to multimedia content.

The rest of the paper is organized as follows. In Section I-A, the framework of video categorization is introduced. Section II presents the features used in this research. Features are related to color, edge, shape, texture and so on. In Section III, three popular dimension reduction methods are introduced and the dimension reduction results are compared. Section IV briefly introduced four decision tools. Experiments are designed to compare performance of each decision tool respect to reduced features.

A. Proposed Video Categorization Framework

The aim of this paper is to test and analyze the generalization performance of four different decision tools in the domain of video categorization. The following is our system structure. The system consists of basically two phases: feature selection and decision tools comparison. Figure 1 shows the framework proposed in this paper.

We extract frames from four kinds of videos which are basketball, soccer, table tennis and swimming on given intervals. Features are extracted from these frames.

II. FEATURE EXTRACTION

Key frames are extracted from video shots. Furthermore, features are extracted from these key frames. The MPEG-7 standard formally aims to create a standard for describing

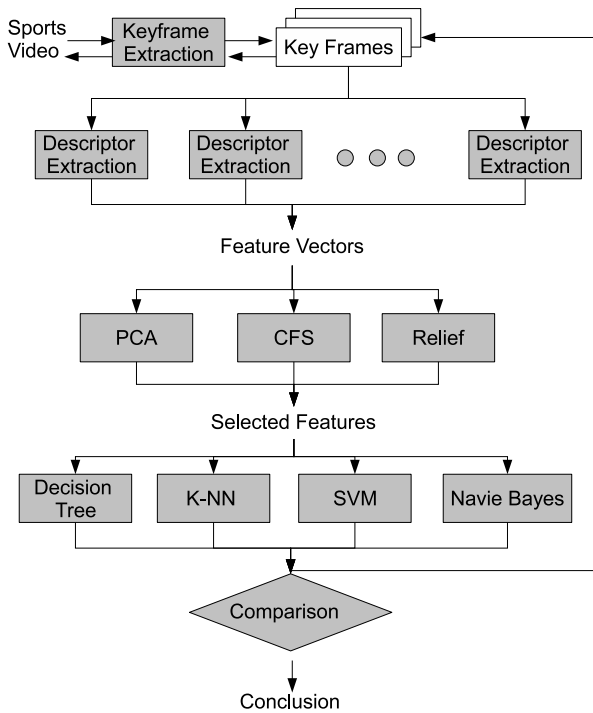


Fig. 1. The Framework of Comparison Analysis on Video Categorization

the multimedia content data that will support some degree of interpretation of the information's meaning. Instead of feature analysis, the focus of this research is the comparison analysis on the performance of feature reduction and machine learning algorithms achieved in sports video categorization. Therefore, we simply use 6 of the MPEG-7 descriptors, which are briefly described as follows. After feature extraction, a set of features forms a feature vector to represent the characteristics of the key-frame as well as the whole video shot.

Color-Structure Descriptor The Color Structure descriptor is a color feature descriptor that captures both color content (similar to a color histogram) and information about the structure of this content. Its main functionality is image-to-image matching and its intended use is for still-image retrieval.

Color Layout This descriptor specifies the spatial distribution of colors for high-speed retrieval and browsing. It targets not only image-to-image matching and video-clip-to-video-clip matching, but also layout-based retrieval for color, such as sketch-to-image matching which is not supported by other color descriptors. This descriptor can be applied either to a whole image or to any part of an image.

Edge Histogram The edge histogram descriptor represents the spatial distribution of five types of edges, namely four directional edges and one non-directional edge. Since edges play an important role for image perception, it can retrieve images with similar semantic meaning. Thus, it primarily targets image-to-image matching (by example or by sketch), especially for natural images with non-uniform edge distribution.

Region-Based Shape The shape of an object may consist of

either a single region or a set of regions as well as some holes in the object. Since the region-based shape descriptor makes use of all pixels constituting the shape within a frame, it can describe any shapes, i.e. not only a simple shape with a single connected region but also a complex shape that consists of holes in the object or several disjoint regions. The Region-based shape descriptor not only can describe such diverse shapes efficiently in a single descriptor, but also is also robust to minor deformation along the boundary of the object.

Homogenous Texture Descriptors Homogeneous texture has emerged as an important visual primitive for searching and browsing through large collections of similar looking patterns. An image can be considered as a mosaic of homogeneous textures so that these texture features associated with the regions can be used to index the image data.

Texture Browsing The Texture Browsing Descriptor is useful for representing homogeneous texture for browsing type applications, and requires only 12 bits (maximum). It provides a perceptual characterization of texture, similar to a human characterization, in terms of regularity, coarseness and directionality.

III. DIMENSION REDUCTION

High-dimensional data sets present many mathematical challenges and are bound to achieve high accuracy of categorization. Dimension reduction is the process of reducing the number of random variables under consideration in order to improve categorization performance and reduce computational complexity.

A. Dimension Reduction algorithms

In this research, we select three popular dimension reduction algorithms for comparison analysis, which are briefly introduced as follows.

Correlation-based Feature Selection (CFS) Like the majority of feature selection programs, CFS uses a search algorithm along with a function to evaluate the merit of feature subsets. The heuristic by which CFS measures the "goodness" of feature subsets takes into account the usefulness of individual features for predicting the class label along with the level of intercorrelation among them the hypothesis on which the heuristic is based can be stated: good feature subsets contain features highly correlated with the class, yet uncorrelated with each other. More details are in [12].

Principal Components Analysis (PCA) is a multivariate analysis technique commonly used to reduce the dimensionality of a dataset with a large number of interdependent variables. This reduction is achieved by finding a set of N_t orthogonal vectors in the input space of dimension N_c , with N_t, N_c , which accounts for as much as possible of the data variance. More details are in [13].

Relief is a feature weight based algorithm which evaluates each attribute by its ability to distinguish among instances that are near each other. The key idea of Relief is to estimate the quality of features according to how well their values distinguish between the instances of the same and different classes that are near each other. More details are in [14].

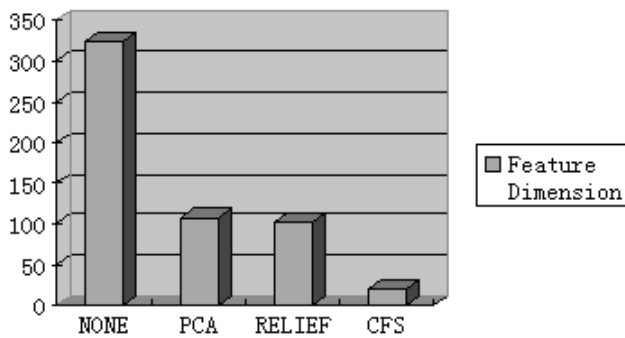


Fig. 2. Comparison of Feature Selection Results

B. Dimension Reduction Results

The original data have 323 dimensions. CFS, PCA and Relief are used to select features individually. Figure 2 shows the dimension reduction results for comparison. For each method, the number of reduced dimensions performs the best classification results.

Different algorithms remain feature dimensions by different rules. As shown in Figure 2, CFS remains the least dimensions, which will mostly reduce the computational complexity for further categorization. To know which dimension reduction method selects most efficient and effective feature set, we need to do more experiments of categorization on selected feature sets by using difference decision tools (See IV-B).

IV. VIDEO CATEGORIZATION

In this work, to identify the sports types for the sports videos is actually a video categorization task. Supervised machine learning methods are widely used to achieve the goal of categorization and classification.

A. Decision Tools

In the following, four decision tools are chosen due to the popularity and the different characteristics.

Decision Trees are supervised machine learning methods frequently used for automated knowledge acquisition. Decision trees are powerful and popular tools for classification [15] because of the following reasons: 1) Decision trees are able to generate understandable rules. 2) Decision trees perform classification without requiring much computation. 3) Decision trees provide a clear indication of which fields are most important for prediction or classification.

Support Vector Machine is a powerful method for finding high dimensional decision surfaces in upper dimensional pattern space. A Support Vector Machine (SVM) is used in our experiments. Choosing a SVM classifier for feature selection was motivated by the successful story of SVM classification for high dimension features [16]. The Radial Basis function is chosen to be a kernel function defined as: $K(x, y) = \exp(-\|x - y\|^2/c)$, with $c = 0.5$.

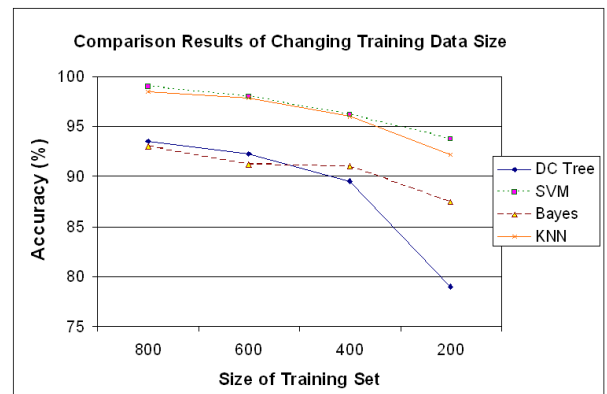


Fig. 3. Comparison Results of Changing Training Data Size

Naive Bayesian classification is based on Bayesian theorem and exhibited high accuracy and speed when applied to large databases and is particularly suited when the dimensionality of the inputs is high. Naive Bayesian classifier makes class conditional independence assumption. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods [17].

K Nearest Neighbor classifiers are based on learning by analogy. It is instance-based or lazy learners in that they store all of the training samples and do not build a classifier until a new sample needs to be classified [17]. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors.

B. Comparison Analysis on the Performance of Decision Tools

Normally, the accuracy of categorization relies on two main factors, which are enough size of the training data and the efficient features for training. To compare the performance of the different decision tools, we set experiments for two groups as follows. 1) Decision tools are used for categorization by changing training data size. 2) Decision tools work on different features which are selected by different feature selection tools. The experimental data include four categories, which are basketball, soccer, table tennis and swimming. Key frames are extracted from each video shot. Totally, 1200 frames are used for experiments.

Comparison on Changing Training Data Size We use two third of the data for training and one third for testing. To compare the performances of four classifiers, we change training data size for four experiment groups. The results are shown in Figure 3. It seems that SVM give the best classification accuracy for a varying training data size and K-NN also show the similar performance. From Figure 3, we can see that larger training data size may bring higher accuracy because it helps classification model generalization. The accuracy of decision tree drop down greatly indicates that it is not good choice when rare class classification and small training set.

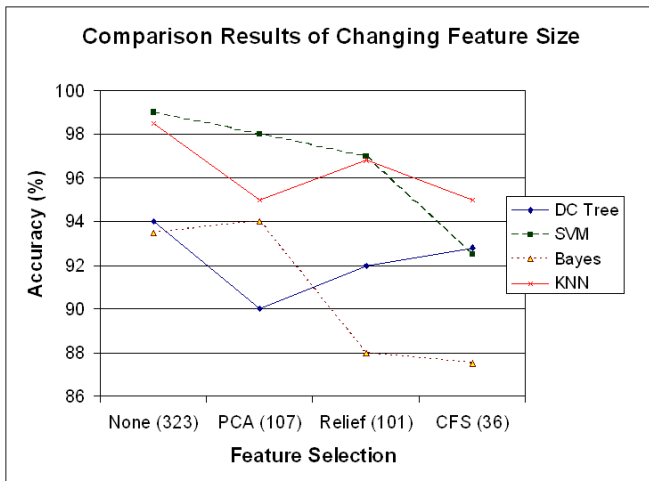


Fig. 4. Comparison Results of Changing Feature Size

Comparison on Changing Feature Size We let four classifiers work on different feature set which selected by feature selection tools, which are PCA, CFS and Relief. The results are shown in Figure 4. As shown in Figure 4, varying the feature size affect the performance of categorization. The performance of Naive Bayesian and SVM seem to drop consistent while the feature size reducing. Three feature selection methods perform very differently with the classification in use. 1) PCA works well for Naive Bayesian and SVM, which reduce the feature size from 323 dimensions to 107 dimension. For Naive Bayesian, using PCA even get a higher accuracy comparing to non feature selection. And, for SVM, using PCA just drops little accuracy. 2) For KNN, SVM and decision tree, relief reduce the feature size to 101 dimensions with little loss in accuracy. 3) CFS works well for decision tree. From decision tools' points of view, some findings might be as follows: 1) SVM prefers PCA and Relief; 2) KNN performs well when using Relief for feature selection; 3) For decision tree, CFS is the best choice; 4) PCA is the unique choice for Naive Bayesian.

C. Discussion

From our experimentation it seems that SVM gives the best trade off in terms of training data size and accuracy. Moreover, SVM achieve the best accuracy even under conditions of varying feature size, which also seems to maintain reasonable amount of consistency in terms of accuracy.

The relationship between feature selection method and decision tools can be concluded as following. PCA: Bayesian network > SVM >> (k-NN, DC tree) CFS: DC tree > k-NN >> (Bayesian network, SVM) Relief: (k-NN, SVM, DC tree) >> Bayesian network Where "A > B" means the feature selection method fit more for A than B, and "A >> B" means there is a big difference between their accuracy.

V. CONCLUSIONS

In this paper, we have analyzed the performance of 4 decision tools, namely (1) Decision Trees (2) SVM (3) Naive

Bayesian (4) k Nearest Neighbor for sports video categorization. Our results indicate that SVM and k-NN could be good choices.

We also compare three different feature selection methods with respect to four different decision tools. It is found that 1) CFS and Relief perform better if we employ decision tree as a classification tool. 2) PCA and Relief can be used to reduce size of feature sets to around 1/3 of the original and make little difference in the accuracy for SVM For k-NN. CFS is a choice with acceptable accuracy drop but apparently decrease of feature numbers. And the last, PCA is the only eligible way to reduce feature size for Bayesian network.

This work shows an example of comparison analysis on the performance and effects of decision tools and feature selection as well as the relationship between decision tools and feature selection. More decision tools and various feature selection should be included in future.

REFERENCES

- [1] E. C. Yiu, *Image Classification Using Color Cues and Texture Orientation*. Department of EECS, MIT: Master's thesis, 1996.
- [2] M. N. Gorkhani and R. W. Picard, "Texture orientation for sorting photos," *the Proceedings of the 12th International Conference on Computer Vision and Image Processing*, vol. 1, pp. 459–464, October 1994.
- [3] A. Vailaya and A. K. Jain, "Reject option for vq-based bayesian classification," *the Proceedings of the 15th International Conference on Pattern Recognition*, vol. 2, pp. 2048–2051, 2000.
- [4] D. Forsyth, J. Malik, M. Fleck, H. Greenspan, T. Leung, S. Belongie, C. Carson, and C. Bregler, "Finding pictures of objects in large collections of images," *the Proceedings of International Workshop on Object Recognition for Computer Vision*, 1996.
- [5] M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," *the Proceedings of the 21st Australian Computer Science Conference*, pp. 181–191, 1998.
- [6] S. Takagi, S. Hattori, K. Yokoyama, A. Kodate, and H. Tominaga, "Sports video categorizing method using camera motion parameters," *the Proceedings of IEEE International Conference on Multimedia and Expo (ICME'03)*, vol. 2, pp. II – 461–4, July 2003.
- [7] S. H. Jang, M. Y. Song, and H. Cho, "Semantic classification of sports news video using color and motion features," *the Proceedings of IEEE International Conference on Hybrid Information Technology (ICHIT'06)*, vol. 99, no. 7, pp. 1–100, January 2006.
- [8] Y. Yuan and C. Wan, "The application of edge feature in automatic sports genre classification," *the Proceedings of IEEE Conference on Cybernetics and Intelligent Systems*, vol. 99, no. 7, pp. 1–100, December 2004.
- [9] X. Gibert, H. Li, and D. Doermann, "Sports video classification using hmms," *the Proceedings of IEEE*, vol. 99, no. 7, pp. 1–100, January 1999.
- [10] J. Wang, C. Xu, and E. Chng, "Automatic sports video genre classification using pseudo-2d-hmm," *the Proceedings of IEEE International Conference on Pattern Recognition*, vol. 99, no. 7, pp. 1–100, January 2006.
- [11] X. Yuan, W. Lai, T. Mei, X. S. Hua, X. Q. Wu, and S. Li, "Automatic video genre categorization using hierarchical svm," *the Proceedings of IEEE*, vol. 99, no. 7, pp. 1–100, January 1999.
- [12] E. E. Ghiselli, *Theory of Psychological Measurement*, 1964.
- [13] G. Deco and D. Obradovic, "An information-theoretic approach to neural computing," 1996.
- [14] K. Kira and L. A. a. Rendell, "The feature selection problem: Traditional methods and a new algorithm," *the proceeding of AAAI*, pp. 129–134, 1992.
- [15] T. M. Mitchell, *Machine learning*. WCB/McGraw.
- [16] V. Vapnik, *Statistical Learning Theory*. Wiley, 1998.
- [17] J. Han, *Data Mining Concepts and Techniques*. Morgan Kaufmann, 2001.