Rangsit Journal of Arts and Sciences, July-December 2011 RJAS Vol. 1 No. 2, pp. 159-168

DOI : 10.14456/rjas.2011.18

A new video similarity measurement for sports video classification

Prisana Mutchima[1]* and Parinya Sanguansat[2]

[1]Faculty of Information Technology, Rangsit University, Patumthani 12000, Thailand
E-mail: prisanut@hotmail.com
[2]Faculty of Engineering and Technology, Panyapiwat Institute of Management, Nonthaburi 11120, Thailand
E-mail: sanguansat@yahoo.com

*Corresponding author

Submitted 8 August 2011; accepted in final form 27 November 2011

Abstract

A key issue of video similarity measure is that most video data are huge files, resulting in time-consuming data processing. Therefore, reducing the dimensionality of the data becomes an essential. But data-dependent dimensionality reduction methods are not efficient. Furthermore, video data usually consists of a large number of frames which varies between different videos, making it difficult to compare their similarity. Therefore, this paper proposes a new framework to reduce the dimensionality of video data by Random Projection (RP) technique and fix dimension by distance space technique. In addition, Compressive Classification (CC) technique will be applied to classify videos. This technique works with a dimensionality reduction method that is data independent. Initially, all training videos frames are extracted by a color histogram based method. Next, all videos features are projected onto a low-dimensional subspace using a random projection. Then a clustering technique is performed to provide the centroids of each cluster, called reference vectors. These vectors are used as a set of basis to create new space, called distance space. For any sequence in distance space, the new feature is represented by the frequencies of similar frames compared with each reference vector. Finally, videos will be classified by the compressive classifier. Empirical evaluations of the results show that the proposed framework significantly outperforms other approaches in video classification.

Keywords**:** *video similarity, video classification, Random Projection, distance space, Compressive Classification*

1. Introduction

Video similarity measuring is the key issue in video classification, one of the vital steps in a Content-based Video Retrieval System (Cheung & Zakhor, 2003a). Furthermore, an efficient video similarity measure is an important operation in several multimedia information systems, owing to its wide applications in many areas such as news video broadcasting, advertising, and personal video archives (Blanken, Vries, Blok, & Feng, 2007; Calic, Campbell, Dasiopoulou, & Kompatsiaris, 2005; Hauptmann et al., 2002).

Many approaches have been attempted for video similarity measure and video classification. Following the literature review, one popular video representation technique is to represent each video sequence with frames (Man-Kwan & Suh-Yin, 1998; Ott, Lambert, Ionescu, & Coquin, 2007; Zhou, Zhou, & Shen, 2007) which contain all of the information of an image. In image comparison, various features such as color (Deng, Manjunath, Kenney, Moore, & Shin, 2001; Mojsilovic, Hu, & Soljanin, 2002; Nor Hazlyna et al, 2010; Zhang, Wenhui, & Yinan, 2009), texture (Chikkerur, Pankanti, Jea, Ratha, & Bolle, 2006; Suruliandi & Ramar, 2008) and shape (Huitao, 2005; Mazhar, Gader, & Wilson, 2009) were used in several approaches. Among these characteristics, color features are the most basic features, which are widely used and prove to be highly effective for image comparison (Cheung & Zakhor, 2003a; Cheung & Zakhor, 2003b; Ferman, Tekalp, & Mehrotra, 2002; Zhang, Wenhui, & Yinan, 2009). Therefore, this study focuses on the use of color features to compare the similarity of low-level visual features of images. The most common color descriptor used in the literature is the color histogram, which directly captures the probability distribution of the colors (Chakravarti & Meng, 2009; Xiong, Radhakrishnan, Divakaran, Rui, & Huang, 2005).

Recently, a technique for video similarity measure based on the percentage of visually similar frames between the two sequences has been proposed (Cheung & Zakhor, 2003a; Cheung & Zakhor, 2003b; Shen, Tao, Beng, & Zhou, 2005).

One commonly used technique for video similarity measure is Naïve Video Similarity (NVS) (Cheung & Zakhor, 2003a; Cheung & Zakhor, 2003b). This technique finds the total number of frames from each video sequence with at least one similar frame with the other sequence. Then, the ratio of these numbers will be computed to the total numbers of frames. After that, the threshold is used for comparing the difference between frames. The efficiency of such a technique depends on the effective selection of the optimal frame similarity threshold. Practically, it is rather difficult to identify the optimal frame similarity threshold because it often comes in an unpredictable pattern and has to be manually determined, resulting in time-consuming data processing. Moreover, the optimal threshold also depends on the training set. If the training set changes, the optimal threshold should also be changed, otherwise it will fail to categorize some test videos. Thus, Mutchima and Sanguansat (2010a) used expected value to average the distance of video frames instead of the threshold. Accordingly, they applied the L_1 metric to measure the distance in comparing the color histograms and averaged distance of video frames by expected value, i.e. harmonic mean, geometric mean, arithmetic mean and median. In addition, the nearest neighbor classifier was applied to classify videos. However, though this method is easy and convenient, it takes a great deal of time to process the data. This is because in distance comparison, each sampling frame of the test videos has to be compared with all the sampling frames of the training videos.

However, video data usually consists of a large number of frames which varies in different videos, making it difficult to compare their similarity. Mutchima and Sanguansat (2010b) used a technique called distance space which represented video frames with new feature vectors in the new feature space to fix the feature dimension of video data. Based on the idea that similar videos usually have a large number of similar frames, this technique used a clustering technique to identify centroids of frame similarity features, and used them as reference vectors. The observation videos were compared in terms of the distance between the observation video frames and reference vectors. The frequency of similar frames comparing to each reference vector was called new feature vectors, while the distance from each reference vector in the database to the observation sequence was called distance space. By representing the number of frames with the number of reference vectors in the new space, the dimensionality of the videos was fixed and easier to compare in terms of similarity.

Since video data are huge files, Mutchima and Sanguansat (2010b) utilized a Random Projection (RP) technique to reduce the dimensionality of video data. That is, all video features were projected onto a low-dimensional subspace using a random matrix whose columns have unit lengths without using any training sets. Compare this to other traditional dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) which require large memory for processing as these techniques have to refer to training video sequences which are usually large video files. The RP technique is data independent. Yet, the results are as good as other methods (Goela, Bebisa, & Nefianb, 2005).

In video classification, this study applies Compressive Classification (CC) to classify videos which are not dependent on data. CC originated a new paradigm in signal processing called "Compressive Sampling" or "Compressed Sensing" (CS) (Majumdar & Ward, 2010). CS combines dimensionality reduction with data acquisition by collecting a (random) lower dimensional projection of the original data instead of sampling it. The aim of CS is signal reconstruction from compressed samples. On the contrary, CC aims at directly classifying such compressed samples without the need to reconstruct the original signals. CC uses a Random Projection (RP) matrix for dimensionality reduction. The projection matrix is data independent. Compressive classifiers are data independent in the sense that they do not require retraining whenever new data are added.

The proposed approach can be briefly described: First, all frames of the training videos are extracted by a color histogram based method. Second, all features of videos are projected onto a low-dimensional subspace using a random projection. Third, a clustering technique is performed to provide the centroids of each cluster, called reference vectors. These vectors are used as a basis set to create a new space, called distance space. For any sequence in distance space, the new feature is represented by the frequencies of similar frames comparing with each reference vector. Finally, videos will be classified by a compressive classifier.

The remainder of this paper is organized in the following manner. In Section 2, the objective of the study is described. Materials and methods are proposed in Section 3. In Section 4, results are

described to demonstrate the performance of the proposed approach. Discussion is presented in Section 5. Finally, conclusions are proposed in Section 6.

2. Objectives

The objective of this study is to efficiently measure video similarity by a new framework to reduce the dimensionality of video data by a random projection (RP) technique and fix dimension by a distance space technique. In addition, a compressive classification (CC) will be applied to classify videos.

3. Materials and Methods

This process starts from preparing data in Section 3.1 and feature extraction in Section 3.2. Our proposed method consists of two steps as shown in Section 3.5 and 3.6. In Section 3.7, the compressive classifier is used to classify these extracted features.

3.1 Datasets

The data consist of 200 video sequences of TV sports programs, comprised of 10 sport genres, namely basketball, boxing, football, snooker, swimming, table tennis, tennis, beach volleyball, volleyball and wrestling. The datasets were divided into two groups, i.e. 100 training and 100 test video sequences. The number of frames of each video sequence is 30 frames per second in MPEG-2 format. The resolution of the datasets evaluation sequences is 480×720 pixels, and the length of each video is approximately 30 seconds.

3.2 Feature extraction

To reduce the dimensionality of video data, this study uses the feature extraction method. The original features are transformed into new sampling features. For image classification, the color histogram is widely used as an important color feature indicating the content of the image. Moreover, the advantage of using the color histogram is its robust ability for affine transformation, especially rotation and scaling of the image content (Xiaoling & Hongyan, 2009). Therefore, this study represented each video sequence with frames, and each individual frame in the video with the color histograms. In addition, to incorporate spatial information into the image features, the image was partitioned into four quadrants, with each quadrant having its own color histogram.

3.3 Naïve Video Similarity

Naïve Video Similarity (NVS) is a traditional technique to measure video similarity (Cheung & Zakhor, 2003a; Cheung & Zakhor, 2003b) by finding the total number of frames from each video sequence that has at least one visually similar frame with the other sequence, and then computing the ratio of this number to the overall total number of frames. Individual frames in a video are represented by high dimensional feature vectors from a metric space. In order to be robust against editing changes in the temporal domain, a video X is defined as a finite set of feature vectors and ignores any temporal ordering. The metric $d(x,y)$ measures the visual dissimilarity between frames x and y which are visually similar to each other if and only if $d(x,y) \leq \varepsilon$ for an $\varepsilon > 0$ independent of x and y, where ε is the frame similarity threshold.

This method uses the L_1 metric to measure the distance. It is defined by the sum of the absolute difference between each bin of the two histograms. This method denotes the L_1 metric between two feature vectors x and y as $d(x,y)$ as follows:

$$d(x,y) \equiv \sum_{i=1}^{4} d_q(x_i, y_i) \qquad (1)$$

where $$d_q(x,y) \equiv \sum_{i=1}^{n} \left\| x_i[j] - y_i[j] \right\| \qquad (2)$$

where x_i and y_i for $i \in \{1,2,3,4\}$ represent the quadrant color histograms from the two image feature vectors, n is the number of histogram bins and $\|\bullet\|$ is the L_1 metric. A small $d(\bullet,\bullet)$ value usually indicates visual similarity, except when two images share the same background color.

X and Y are two video sequences, represented as sets of feature vectors. The numbers of frames in video X that have at least one visually similar frame in Y is represented by, $\Psi_{(X,\,Y;\,\varepsilon)}$ where 1_A is the indicator function with $1_A = 1$ if A is not empty, and zero otherwise when x and y are two video frames, represented as feature vectors and ε is the frame similarity threshold. The Naïve Video Similarity between X and Y, $nvs(X,Y;\varepsilon)$, is defined as follows:

$$nvs(X,Y;\varepsilon) \equiv \frac{\Psi_{(x,y;\varepsilon)} + \Psi_{(y,x;\varepsilon)}}{|X| + |Y|} \qquad (3)$$

where

$$\Psi_{(x,y;\varepsilon)} = \sum\nolimits_{x \in X} 1_{\{y \in Y : d(x,y) \leq \varepsilon\}} \qquad (4)$$

and

$$\Psi_{(y,x;\varepsilon)} = \sum_{y \in Y} 1_{\{x \in X : d(y,x) \le \varepsilon\}} \quad (5)$$

where |•| denotes the cardinality of a set or the number of frames in a given video.

If every frame in video *X* has a similar match in *Y* and vice versa, *nvs(X, Y; ε)* = 1. If *X* and *Y* share no similar frames at all, *nvs(X, Y; ε)* = 0.

3.4 Expectation-based method

This method can measure the similarity of video efficiently by using an expected value to average the distance of video frames instead of the threshold (Mutchima & Sanguansat, 2010a). Each video sequence was represented with a frame and each frame was represented with the color histogram to help enhance feature reduction. After that, categorization was performed using the nearest neighbor classifier with the L_1 metric to measure distance by comparing each sampling frame of the training videos with all sampling frames of the test videos.

In this example, *X* and *Y* are two video sequences, represented as frames. The metric *d(x,y)* measures the visual similarity between frames *x* and *y*. They denote the distance metric between two feature vectors *x* and *y* as *d(x, y)*, as follows:

$$d(x_a, y_b) \equiv \sum_{i=1}^{4} \left\| x_i[a] - y_i[b] \right\| \quad (6)$$

where x_i and y_i for $i \in \{1, 2, 3, 4\}$ represent the quadrant color histograms from the two image feature vectors to merge the spatial information into the image features. Spatial information describes the physical location of objects and the relationship between objects. In this case, *a* is the a^{th} sampling frame of *X* and *b* is the b^{th} sampling frame of *Y*.

The measuring video similarity between two video sequences *X* and *Y*, *SIM*, is defined as:

$$SIM(X,Y) \equiv \min\{D(X,Y)\} \quad (7)$$

where

$$D(X,Y) \equiv \mathrm{E}[d(x_a, y_b)] \quad (8)$$

where E[•] is the expectation operator. The similarity between two video sequences can be measured at various intervals by changing the number of histogram bins and the expected value. The measuring video similarity is the comparison minimum of average of frame distance measures.

3.5 Random Projection

Random Projection (RP) has emerged as a powerful dimensionality reduction method. Its most important property is that it is a general data reduction method. In RP, the original high dimensional data is projected onto a low-dimensional subspace using a random matrix whose columns have unit length without using any training sets. Normally, the random matrix should be a normal distribution.

Traditional dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) require large memory for processing as the techniques have to refer to training video sequences which are usually large video files. RP is data independent. Yet, the results are as good as other methods (Deegalla & Bostrom, 2006; Gao, Li, & Katsaggelos, 2009; Goela, Bebisa, & Nefianb, 2005; Wu & Hu, 2008).

In random projection, the set of points of size *q* in original *s*-dimensional Euclidean space is projected to a *s*-dimensional ($p << q$) subspace through the origin, using a random $p \times q$ matrix *R* whose columns have unit lengths in order to achieve dimension reduction as follows:

$$W_{p \times s} = R_{p \times q} F_{q \times s} \quad (9)$$

where R_{pxq} is the random matrix, F_{qxs} is the original observations set of size *p* in *q*-dimension, and W_{pxs} is the projection in *s*-dimension subspace.

3.6 Distance space

This study applied a technique to create new feature space, called *distance space* which refers to distance from each reference vector in a database to the observation sequence (Mutchima & Sanguansat, 2010b).

The distance space process is described in Algorithm 1. First, all frames of the training videos, X_i, are extracted by the color histogram based method and are projected by random matrix. After that, the clustering technique is performed to provide the centroids of each cluster, called *reference vectors*, ξ_k using *k*-mean. Finally, the new feature vector, G_i, is represented by the frequencies of similar frame comparing with each reference vector.

3.7 Compressive Classification

Compressive Classification (CC) originated with a new paradigm in signal processing called "Compressive Sampling" or "Compressed Sensing" (CS) (Majumdar & Ward, 2010). CS combines dimensionality reduction with data acquisition by collecting a (random) lower dimensional projection of the original data instead of sampling it. CC refers to a new class of classification methods that are robust to data acquired using CS. Only a few properties are preserved by CS data acquisition, and compressive classifiers are designed to exploit these properties so that the recognition accuracy on data acquired by CS is approximately the same as that on data acquired by traditional sampling. There is a basic difference that separates CC from conventional classification methods. In conventional classification, data are acquired by traditional (Nyquist) sampling.

Algorithm 1 Distance Space Algorithm

Require: $X_{i=1...N}, \xi_{k=1...C}$

Ensure: $G_{i=1...N}$

1: All frames of the training videos, X_i, are extracted by color histogram based method and projected by random matrix.
2: Perform clustering in this feature spaces to keep centroids of each cluster as reference vectors, ξ_k .
3: **for** i = 1 to N **do**
4: $G_i \leftarrow 0$
5: **for** j = 1 to $|X_i|$ **do**
6: **for** k = 1 to C **do**
7: $D_k \leftarrow \|X_i[j] - \xi_k\|$
8: **end for**
9: $L \leftarrow \arg\min_k(D_k)$
10: $G_i[L] \leftarrow G_i[L]+1$
11: **end for**
12: $G_i \leftarrow G_i / |X_i|$
13: **end for**
14: **return** $G_{i=1...N}$

Once all data are obtained, a data-dependent dimensionality reduction technique is employed; data acquisition and dimensionality reduction are disjoint activities. CC operates on data acquired by a CS technique, where dimensionality reduction occurs simultaneously with data acquisition. Thus, CC works with a dimensionality reduction method that is data independent, whereas the dimensionality reduction techniques in traditional classifications are data dependent (e.g., PCA, LDA, etc.).

The aim of CS is signal reconstruction from compressed samples. On the contrary, CC aims at directly classifying such compressed samples without the need to reconstruct the original signals. CC uses a Random Projection (RP) matrix for dimensionality reduction. The projection matrix is data independent. Compressive classifiers are data independent in the sense that they do not require retraining whenever new data is added. Therefore, this study applied CC to classify videos.

The classification problem involves finding the identity of an unknown test sample given a set of training samples and their class labels. Compressive Classification (CC) addresses the case where compressive samples of the original signals are available instead of the original signal.

The Sparse Classifier (SC) is based on the assumption that the training samples of a particular class approximately form a linear basis for a new test sample belonging to the same class (Wright, Yang, Ganesh, Sastry, & Yi, 2009). If $v_{k,test}$ is the test sample belonging to the k^{th} class, then

$$v_{k,test} = \alpha_{k,1}v_{k,1} + \alpha_{k,2}v_{k,2} + ... + \alpha_{k,n_k}v_{k,n_k} + \varepsilon_k = \sum_{i=1}^{n_k} \alpha_{k,i}v_{k,i} + \varepsilon_k \quad (10)$$

where $v_{k,i}$ are the training samples of the k^{th} class, $\alpha_{x,i}$ is the weight corresponding weight and ε_k is the approximation error (assumed to be normally distributed).

Eq. (10) expresses the assumption in terms of the training samples of a single class. Alternatively, it can be expressed in terms of all the training samples such that

$$v_{k,test} = \alpha_{1,1} + ...\alpha_{k,1}v_{k,1} + ... + \alpha_{k,n_k}v_{k,n_k} + ... + \alpha_{C,n_c}v_{C,n_c} + \varepsilon = \sum_{i=1}^{n_1} \alpha_{1,i}v_{1,i} + ... \sum_{i=k}^{n_k} \alpha_{k,i}v_{k,i} + ... + \sum_{i=1}^{n_c} \alpha_{C,i}v_{C,i} + \varepsilon \quad (11)$$

where C is the total number of classes.

In matrix-vector notation, Eq.(11) can be expressed as

$$v_{k,test} = V\alpha + \varepsilon \quad (12)$$

where $V = [v_{1,1} | ... | v_{k,1} | ... | v_{k,n_k} | ... | v_{C,n_c}]$ and $\alpha = [\alpha_{1,1} ... \alpha_{k,1} ... \alpha_{k,n_k} ... \alpha_{C,n_c}]'$.

The linearity assumption coupled with Eq. (12) implies that the coefficient vector α should be nonzero only when it corresponds to the correct class of the test sample.

Based on this assumption, the sparse optimization problem is:

$$\min \|\alpha\|_0 \; subject\ to \; \|v_{k,test} - V\alpha\|_2 \le \eta \quad (13)$$

$\eta$is related to ε.

As previously mentioned, Eq. (13) is an NP-hard problem. Consequently, a convex relaxation to the NP-hard problem was made (Wright, et al., 2009), and the following problem was solved instead:

$$\min \|\alpha\|_1 \; subject\ to \; \|v_{k,test} - V\alpha\|_2 \le \eta. \quad (14)$$

The formulation of the sparse optimization problem as that in Eq. (14) is not ideal for this scenario, as it does not impose sparsity on the entire class as the assumption implies.

The Sparse Classification (SC) algorithm is as follows:

Algorithm 2 Sparse Classification Algorithm

Require: Test sample $v_{k,test}$, training matrix V, and error tolerance η.

Ensure: Estimated sparse weight α.

1: Solve the optimization problem expressed in Eq. (14).

2: For each class (i), repeat the following two steps.

a) Reconstruct a sample for each class by a linear combination of the training samples belonging to that class

using $v_{recon}(i) = \sum_{j=1}^{n_i} \alpha_{i,j} v_{i,j}$.

b) Find the error between the reconstructed sample and the given test sample by

$error(v_{test}, i) = \|v_{k,test} - v_{recon(i)}\|_2$.

3: Once the error for every class is obtained, choose the class having the minimum error as the class of the given test sample.

4. Results

In this study, the performance of the proposed method for video classification was evaluated against different criteria including feature dimension, number of sampling frames, number of histogram bins and number of reference vectors by running each criterion 10 times in order to identify the accuracy rate of each criterion in video classification.

4.1 Feature dimension

To evaluate the performance of the proposed method in video classification against feature dimensions, the study compared the accuracy rate in using different feature dimensions as the criterion. The experiments set the number of sampling frames as 10, the number of histogram bins as 18 and the number of reference vectors as 40, while varying the number of feature dimensions from 10 to 100. The results show 70 dimensions can achieve the highest accuracy of 95.50%, as shown in Table 1. Moreover, the boxplot of the proposed method in terms of the number of feature dimensions shows that the accuracy rate tends to increase when the number of feature dimensions increases, as plotted in Figure 1.

Table 1 Mean and standard deviation of feature dimension

Feature Dimension	Accuracy Rate	
	Mean (%)	S.D. (%)
10	90.00	4.83
20	93.50	2.42
30	94.20	2.30
40	95.00	2.94
50	94.00	2.16
60	95.30	2.36
70	***95.50***	***2.22***
80	95.30	2.63
90	94.30	3.13
100	95.40	0.97

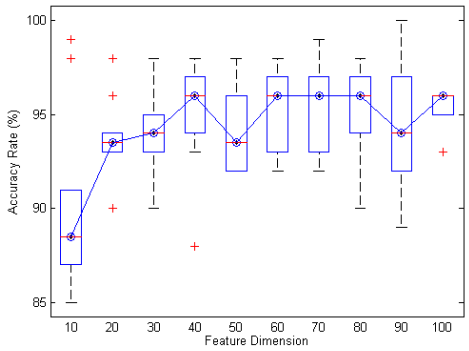

Figure 1 Boxplot of the accuracy rate in terms of the number of feature dimensions

4.2 Number of sampling frames

To evaluate the performance of the proposed method in video classification against the number of sampling frames, the study compares the accuracy rate in using different numbers of sampling frames. The experiments set the number of histogram bins as 18, the number of reference vectors as 40 and the feature dimension as 100, while varying the number of sampling frames from 10 to 20. The results show 18 sampling frames can achieve the highest accuracy of 96.40%, as shown in Table 2. However, the accuracy rate varies for different numbers of sampling frames as shown in the boxplot of the proposed method in Figure 2. Therefore, the results depend on the experiment.

Table 2 Mean and standard deviation of number of sampling frames

Number of Sampling Frames	Accuracy Rate	
	Mean (%)	S.D. (%)
10	95.40	0.97
11	94.60	0.84
12	95.60	1.90
13	95.60	1.71
14	96.30	1.16
15	96.00	1.94
16	96.30	1.49
17	94.80	2.20
18	***96.40***	***2.12***
19	93.30	2.21
20	95.00	1.76

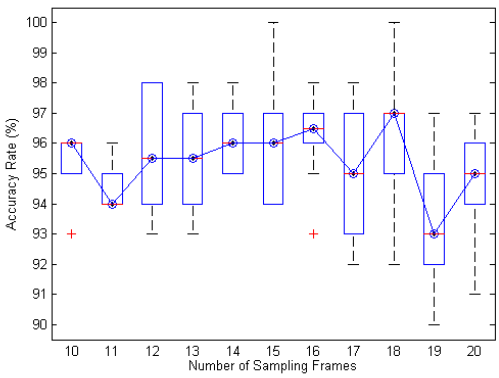

Figure 2 Boxplot of the accuracy rate in term of the number of sampling frames

4.3 Number of histogram bins

To evaluate the performance of the proposed method in video classification against the number of sampling frames, the study compared the accuracy rate in using a different number of histogram bins. The experiments set the number of sampling frames as 10, the number of reference vectors as 40 and the feature dimension as 100, while varying the number of histogram bins from 10 to 20. The results show 13 and 17 histogram bins can achieve the highest accuracy of 96.60%, as shown in Table 3. However, the accuracy rate varies for the different numbers of histogram bins as shown in the boxplot of the proposed method in Figure 3. Therefore, the results depend on the experiment.

Table 3 Mean and standard deviation of number of histogram bins

Number of Histogram Bins	Accuracy Rate	
	Mean (%)	S.D. (%)
10	96.40	1.58
11	94.50	1.65
12	93.70	2.45
13	***96.60***	***0.84***
14	96.00	1.25
15	96.50	1.08
16	96.30	1.57
17	***96.60***	***1.17***
18	95.40	0.97
19	95.00	2.00
20	95.70	1.77

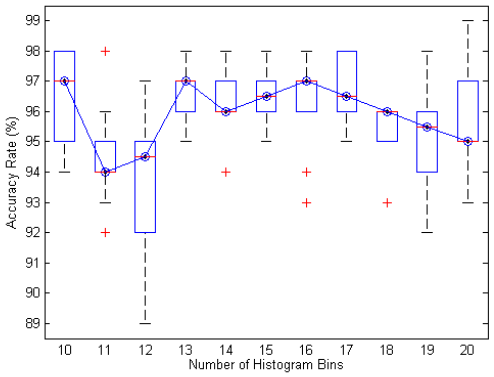

Figure 3 Boxplot of the accuracy rate in term of the number of number of histogram bins

4.4 Number of reference vectors

To evaluate the performance of the proposed method in video classification against the number of reference vectors, the study compared the accuracy rate in using a different number of reference vectors. The experiments set the number of sampling frames as 10, the number of histogram bins as 18 and the feature dimension as 100, while varying the number of reference vectors from 10 to 80. The results show 60 reference vectors can achieve the highest accuracy of 96.60%, as shown in Table 4. Moreover, the boxplot of the proposed

method in term of the number of reference vectors shows that the accuracy rate tends to increase when the number of reference vectors increase, as plotted in Figure 4.

Table 4 Mean and standard deviation of number of reference vectors

Number of Reference Vectors	Accuracy Rate	
	Mean (%)	S.D. (%)
10	85.40	4.38
20	92.70	3.74
30	94.90	2.51
40	95.40	0.97
50	95.40	2.88
60	***96.60***	***1.65***
70	96.50	1.18
80	96.00	1.83

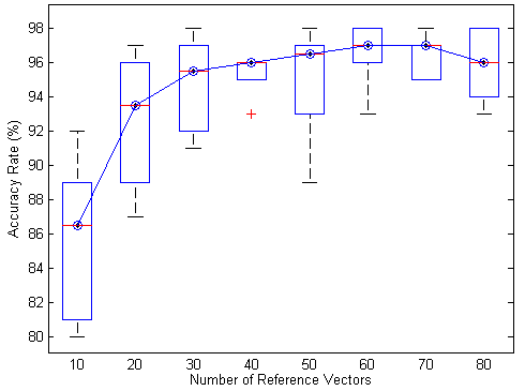

Figure 4 Boxplot of the accuracy rate in term of the number of number of reference vectors

4.5 Comparison with other methods

Comparing the efficiency of the NVS, the expectation-based and the proposed framework in video classification, the results show that the dimension of the proposed method is much smaller than other methods while the accuracy rate is comparable, as shown in Table 5.

Table 5 Accuracy rate comparison of NVS, expectation-based and the proposed method

Technique	Dimension	Accuracy Rate (%)
NVS Method	360,000	95.00
Expectation-based Method	360,000	97.00
Proposed Method	100	96.60

5. Discussion

Empirical evaluations of the results show that the proposed framework significantly outperforms other methods in video classification. Since a random matrix is used in the random projection method, it could be optimized by ℓ_0 minimization to increase the processing speed. However, one disadvantage of the proposed framework is that the optimal values of several criteria cannot be automatically specified and have to be changed if datasets are changed.

6. Conclusion

This paper proposes a new framework to enhance the performance in measuring video similarity and video classification. This framework applies the random projection (RP) technique to reduce the dimensionality of video data, and uses distance space techniques to fix video dimensions, followed by a compressive classifier to classify videos. This technique works with a dimensionality reduction method that is data independent. Moreover, when the number of dimension vectors becomes small, the classification process becomes quite fast. Thus, the proposed method can handle larger and longer videos. Comparing the efficiency of the NVS, the expectation-based and the proposed framework in video categorization, the results show that the dimension of the proposed framework is much smaller than other methods while the accuracy rate is comparable.

7. Acknowledgements

This work was assisted by Suan Dusit Rajabhat University through support with a scholarship and Rangsit University by providing the laboratory room for data processing. Additionally, invaluable recommendations and supervision from anonymous reviewers is greatly appreciated.

8. References

Blanken, H., Vries, A. P., Blok, H. E. & Feng, L. (2007). *Multimedia retrieval*. New York, USA: Springer-Verlag Berlin Heidelberg.

Calic, J., Campbell, N., Dasiopoulou, S. & Kompatsiaris, Y. (2005). A survey on multimodal video representation for semantic retrieval. In *The International Conference on Computer as a Tool (EUROCON 2005)*, Serbia, Montenegro, Belgrade, pp.135-138.

Chakravarti, R. & Meng, X. (2009). A study of color histogram based image retrieval. In *Sixth International Conference on Information Technology: New Generations (ITNG '09)*, Las Vegas, Nevada, USA, pp.1323-1328.

Cheung, S. S. & Zakhor, A. (2003a). Efficient video similarity measurement with video signature. In *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1), 59-74.

Cheung, S. S. & Zakhor, A. (2003b). Fast similarity search on video signatures. In *International Conference on Image Processing (ICIP 2003)*, Barcelona, Catalonia, Spain, pp.II - 1-4.

Chikkerur, S., Pankanti, S., Jea, A., Ratha, N. & Bolle, R. (2006). Fingerprint representation using localized texture features. In *18th International Conference on Pattern Recognition (ICPR 2006*), Hong Kong, China, pp.521-524.

Deegalla, S. & Bostrom, H. (2006). Reducing high-dimensional data by principal component analysis vs. random projection for nearest neighbor classification. In *5th International Conference on Machine Learning and Applications (ICMLA '06*), Orlando, Florida, USA, pp. 245-250.

Deng, Y., Manjunath, B. S., Kenney, C., Moore, M. S. & Shin, H. (2001). An efficient color representation for image retrieval. In *IEEE Transactions on Image Processing*, 10(1), 140-147.

Ferman, A.M., Tekalp, A.M. & Mehrotra, R. (2002). Robust color histogram descriptors for video segment retrieval and identification. In *IEEE Transactions on Image Processing*, 11(5), 497-508.

Gao, L., Li, Z. & Katsaggelos, A. K. (2009). A video retrieval algorithm using random projections. In *16th IEEE International Conference on Image Processing (ICIP 2009)*, Cairo, Egypt, pp.797-800.

Goela, N., Bebisa, G. & Nefianb, A. (2005). Face recognition experiments with random projection. In *Proceeding of the Biometric Technology for Human Identification II*, Orlando, FL, USA, pp.426-437.

Hauptmann, A., Yan, R., Qi, Y. Jin, R., Christel, M., Derthick, M., Chen, M.-Y, Baron, R., Lin, W. H. &.Ng, T.D (2002, November). Video classification and retrieval with the informedia digital video library system. In *Proceeding of the eleventh Text Retrieval Conference (TREC 2002*), Gaithersburg, Maryland, USA, pp. 119-127.

Huitao, L. (2005). Image-dependent shape coding and representation. In *IEEE Transactions on Circuits and Systems for Video Technology*, 15(3), 345-354.

Majumdar, A. & Ward, R. K. (2010). Robust classifiers for data reduced via random projections. In *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 40(5), 1359-1371. URL: http://ubc.academia.edu/AngshulMajumdar/Papers/699899/Robust_classifiers_for_data_reduced_via_random_projections

Man-Kwan, S. & Suh-Yin, L. (1998). Content-based video retrieval based on similarity of frame sequence. In *Proceedings of the International Workshop on Multi-Media Database Management Systems (IW-MMDBMS)*, Dayton, Ohio, pp.90-97.

Mazhar, R., Gader, P.D. & Wilson, J.N. (2009). Matching-pursuits dissimilarity measure for shape-based comparison and classification of high-dimensional data. In *IEEE Transactions on Fuzzy Systems*, 17(5), 1175-1188.

Mojsilovic, A., Hu, H. & Soljanin, E. (2002). Extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis. In *IEEE Transactions on Image Processing*, 11(11), 1238-1248.

Mutchima, P. & Sanguansat, P. (2010a). A new approach for measuring video similarity without threshold and its application in sports video categorization. In *First International Conference on.Pervasive Computing Signal Processing and Applications (PCSPA 2010)*, Harbin, China, pp.868-872.

Mutchima, P. & Sanguansat, P. (2010b). Video similarity measurement approach via dimensionality reduction with distance space and random projection: Application with sports video classification. In *International Symposium on Communications and Information*

Technologies (ISCIT2010), Tokyo, Japan, pp.430-434.

Nor Hazlyna, H, Mashor, M. Y., Mokhtar, N. R., Aimi Salihah, A. N., Hassan, R., Raof, R. A. A., & Osman, M. K. (2010, May 10-13). Comparison of acute leukemia Image segmentation using HSI and RGB color space. In *10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA)*, 749-752. doi: 10.1109/ISSPA.2010.5605410

Ott, L., Lambert, P., Ionescu, B., & Coquin, D. (2007). Animation movie abstraction: Key frame adaptative selection based on color histogram filtering. In *14th International Conference on Image Analysis and Processing Workshops (ICIAPW 2007*), Modena, Italy, pp. 206-211.

Shen, H., Tao, O., Beng C., & Zhou, X. (2005). Towards effective indexing for very large video sequence database. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data (SIGMOD 2005).* Baltimore, Maryland, USA: ACM, pp.730-741

Suruliandi, A., & Ramar, K. (2008). Local texture patterns - a univariate texture model for classification of images. In *16th International Conference on Advanced Computing and Communications (ADCOM 2008*), Bangalore, India, pp.32-39.

Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., and Yi, M. (2009). Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence*, 31(2), 210-227.

Wu, W., & Hu, J. (2008). Similarity search based on random projection for high frequency time series. In *IEEE Conference on Cybernetics and Intelligent Systems*, Chengdu, Chaina, pp. 388-393.

Xiong, Z., Radhakrishnan, R., Divakaran, A., Rui, Y., & Huang, T.S. (2005). *A unified framework for video summarization, browsing & retrieval: with applications to consumer and surveillance video*. Orlando, USA: Academic Press.

Zhang, Z., Wenhui, L., & Yinan, L. (2009). New color feature representation and matching technique for content-based image retrieval. In *International Conference on Multimedia Computing and Systems (ICMCS '09*), Montreal, Quebec, Canada, pp.118-122.

Zhou, X., Zhou, X., & Shen, H. T. (2007). Efficient similarity search by summarization in large video database. In *Proceedings of the eighteenth conference on Australasian database. Victoria: Australian Computer Society*, Ballarat, Victoria, Australia, pp .161-167.