Video Sequence Matching Based On Multiclass SVM Classifiers and RGB Component Relations

Meera R. Nath

Student, Mahatma Gandhi University

Abstract: ***The concept of RGB feature extraction and multiclass SVM Classifiers can be used to locate an input video in a database of videos. The proposed system aims at providing matching video for the query video if present in the database. The proposed system helps in the area of copyright protection of videos and also for reducing the storage redundancy. The system uses the concept of both the SVM (Support Vector Machine) classifiers and RGB components. Firstly, each key frame of the video is split into non overlapping blocks. The order of average intensity of the red, green and blue components of each pixel of the non overlapping block and then of the key frame is determined. There are six cases which shows the relation of red, green and blue components which in turn is used to create the feature set of the colour frame and then video. The feature of the database videos are saved and query video is found using RGB intensity variations. The video sequence matching operation is performed using the multiclass SVM classifiers. Multiclass SVM can be used to classify data into multiple classes. In Multiclass SVM, two functions are present. First one is svmtrain( ) which trains the Support Vector Machine using the training data. Second, svmclassify( ) which is used to classify new data using this trained Support Vector Machine. Here svmtrain( ) trains the feature set of the database videos and create model for each of them. svmclassify( ) classifies the query video to any of the database video's class using the trained model. Thus with the help of these two functions the query video could be located in the video database if present. The proposed system provides more accuracy and fastness in determining the matching video. The computational complexity and storage complexity is very low and space complexity is satisfactory for the proposed system.***

Keywords: Content Based Technology, Support Vector Machine, Colour Histogram

1. Introduction

Today there is a large availability of multimedia and internet to common people. The advancement of new techniques enables the users to handle vast amount of digital contents. Combinations of different digital contents are referred to as multimedia. The contents of multimedia include combination of images, text, audio, video, animations. Electronic devices which is used to store and experience multimedia content are called multimedia devices. Multimedia can be recorded, played, displayed, or accessed using these devices. Interactive multimedia is also called rich media.

Now it is very common to observe the copies on internet in vast amount. This causes the problems related to storage redundancy and copyright infringement. Copyright infringement is the use of exclusive works under copyright without permission from the copyright holder. It includes the right to reproduce, distribute, display or perform the copyrighted or the derived work. The main motive for engaging in copyright infringement is pricing, unavailability, usefulness, user friendliness and anonymity. The copies of same multimedia contents also cause problems due to memory wastage.

There are mainly two methods for detecting the copies. They are watermark-based technique and content-based technology. In watermark based techniques, watermarks are introduced into the original copies[2]. This may be some invisible signal or a visible one which helps for the ease of detection of illegal copies. These are done before distribution so this technique is classified as an active technology .But watermark techniques have many limitations. First, watermarks can easily be destroyed by malicious users [3]. The watermarks can be damaged when undergone some media processing such as geometric distortion and severe compression. The most important problem is that if the original image is not watermarked, then it is not possible to detect the other copies. In such cases, the method would fail at detecting copies.

In content-based technology no extra information is added to the original copies for detection. Here based on the content of the original copy a unique signature is created. Different copy detection algorithms are used to assign a unique fingerprint using features of the original content. These features could be compared with other ones to detect copies.

Here the proposed system aims at locating copies of an input video from a database of videos [1]. The copy means the transformed version of the original video. The transformations include blurring, geometric distortion, contrast enhancement, noise contamination and re-encoding. An important issue for video sequence matching is the robustness of the video feature against the above-mentioned operations. Robustness is the amount that the feature of video changes after the content preserving operations.

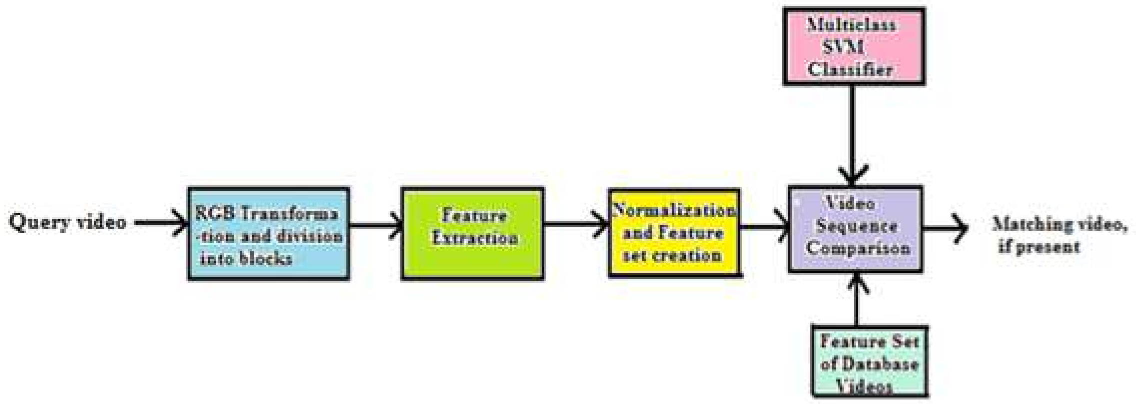

Figure 1: Architecture of video sequence matching based on multiclass SVM classifiers and RGB component relations

2. Related Works

Many image copy detection [4],[5] methods have been proposed but they cannot be used for video sequence matching because of high computational and storage complexity. Today, lots of effective methods are available for the detection of matching videos. In [6], each video key frame is first divided into 3×3 sub images, and then the average block intensity is calculated using the ordinal measurement [7] which is considered as a fingerprint for videos sequence matching. To improve the performance of the original scheme [6] for letter-box and pillar-box operations it is modified in [8]. The temporal measurement [9] can achieve better performance for the shifting and insertion of pattern when compared with spatial ordinal measurements [6][8]. In [10], Principal component analysis (PCA) based approach was introduced. For the fast and robust detection of near duplicate videos [12], MPEG developed video signature descriptors. The robustness of the above mentioned global features against rotation and flipping operations is still poor. SIFT [13] and CS-LBP [14] which can handle many challenging distortions are used to detect sequence copies but their computational and storage complexity is very high [11]. In [15],[16] some trajectory-based techniques are proposed which focus on temporal distortions, such as frame deletion and insertion but they are expensive for the matching phase, since the trajectories must be aligned first. Methods from video hashing [17] can also be employed for the purpose of video sequence matching.

3. System Architecture

The system architecture is given in fig. 1.The query video is first divided into N key frames then each key frame is transformed into RGB channels and divided into blocks. For each key frame the intensity of RGB component is found according to six colour relation. This is then normalized and the feature set for a video is created. Video sequence comparison compares the feature set of database video and query video using multiclass SVM Classifier for fast and accurate comparison. The matched video is displayed as output.

4. Proposed System

The proposed system aims at providing matching video for the query video if present in the database. The system uses the concept of both the SVM (Support Vector Machine) classifiers and RGB components. This section provides both of them. The concept of RGB component is based on the order of average intensity of the red, green and blue components of each pixel in each of the colour frame [1]. Using this, six cases can be formed which in turn is used to create the feature set of the colour frame and then video. The feature of the database videos are saved and query video is found using RGB intensity variations. The video sequence matching operation is performed using the SVM classifiers.SVM classifiers are learning models with learning algorithms which classify data into one category or the other [18].SVM classifiers analyze data and recognize patterns so they are used for classification and regression analysis [19]. They perform both linear and non linear classification. Here multiclass SVM classifiers are used. Multiclass SVM can be used to classify data into multiple classes. Multiclass SVM uses svmtrain and svmclassify functions for training the model and for classifying the new data.

A. Feature set creation of video

A video frame can be represented using different colour models e.g., RGB, cyan, magenta, and yellow (CMY), cyan, magenta, yellow, and black (CMYK), and YCbCr . These color models are used in different applications.RGB Model is selected here for the feature generation of video frame. It is because of its easy transformation from and to other color models and its robustness against most content-preserving operations.

B. Relation of RGB components of a colour frame

The average intensities of the RGB components of a pixel at the coordinates (a,b) within the video key frame or a colour image of size w × h can be represented using the tuple of numbers (Rab, Gab, Bab).

Figure 2: Four typical videos. (a) Lighthouse (b) Fish. (c). Car Racing (d) A Cartoon

The relation of the red, green and blue components of a colour frame would satisfy one of the following six cases:

case 1 : Rab ≥ Gab ≥ Bab ;
case 2 : Rab ≥ Bab ≥ Gab
case 3: Gab ≥ Rab ≥ Bab ;
case 4 : Gab ≥ Bab ≥ Rab
case 5 : Bab ≥ Rab ≥ Gab ;
case 6 : Bab ≥ Gab ≥ Rab

where 1 ≤ a ≤ w, 1 ≤ b ≤ h.

When these six cases are plotted as histogram, it is called as colour histogram. It is one of the most commonly used method to represent a colour image or a video frame [21]. An image and its corresponding colour histogram is given inn fig.3. The colour histogram is very fast to compute and is flexible in terms of storage when compared with other shape based or texture-based features.

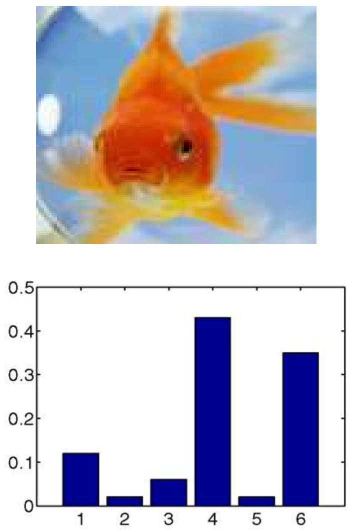

Figure 3: An image and its corresponding colour histogram

Features derived from the colour histogram have been is used in video retrieval, segmentation, and identification. Since the colour histogram does not use any spatial information, it is expected to provide better robustness against most content preserving operations. The ability of the feature to distinguish different video clips are very promising for the proposed method.

C. Feature Extraction Process

The process of feature extraction includes the following three steps [1].They are:

1) RGB Transformation and Division into Blocks

Each key frame of the input video is first transformed into RGB channels. Then it is divided into non overlapping blocks. Average intensities of RGB components of each pixels in these blocks are calculated to reduce the noise like operations. Let the size of key frame or image is w × h.. When it is divided into block of size b × b , an image of size m × n is obtained where m = [w/b], n = [h/b], and [x] is the nearest integer to x..The size of the block affects the performance of the algorithm.

2) Feature Extraction

For feature extraction, the non overlapping blocks are used. The average intensity of RGB components for each of the pixels of the block is calculated. The percentages of pixels belonging to each of the six cases are considered and six normalized real values are obtained for each image or video frame finally. The pixels with same value for RGB components are removed from this process to reduce the special cases.

3) Normalization and Feature set creation of Video

Let Pi denote the set of those pixels whose three color components satisfy the ith case of the six RGB component relations, where ˜Rab, ˜Gab, and ˜Bab represent the three color components of the lower resolution image of size m × n, where 1 ≤ a ≤ m, 1 ≤ b ≤ n and 1 ≤ i ≤ 6

Pi = {(˜Rab, ˜Gab, ˜Bab)|˜Rab, ˜Gab, ˜Bab s.t. case #i}.

The feature obtained for an image undergoes normalization which can be described as,

$$\text{Feature}\,(\,\text{i}) = \frac{|\,\text{Pi}\,|}{\sum_{i=1}^{6} |\,\text{Pi}\,|}$$

where |Pi| is the cardinality of set Pi

D. Robustness

Robustness refers to the amount that the feature of a video will change after some common content-preserving operations. Some of the operations are illustrated in Fig. These operations do not affect the feature extracted which is already proved in existing system [1].

Some of the operations are:

1) **Noise-Like Contamination**: These include includes blurring and adding noise, which can change the intensity of RGB Components of individual pixels. However, such effects can be effectively decreased through averaging.

2) **Scaling, Rotation, and Flipping**: Scaling operation will change the spatial resolution of the video, but will not change the feature of original video [1]. For pure rotation, redundant pixels with the same values for the three channels will be removed in the feature extraction.

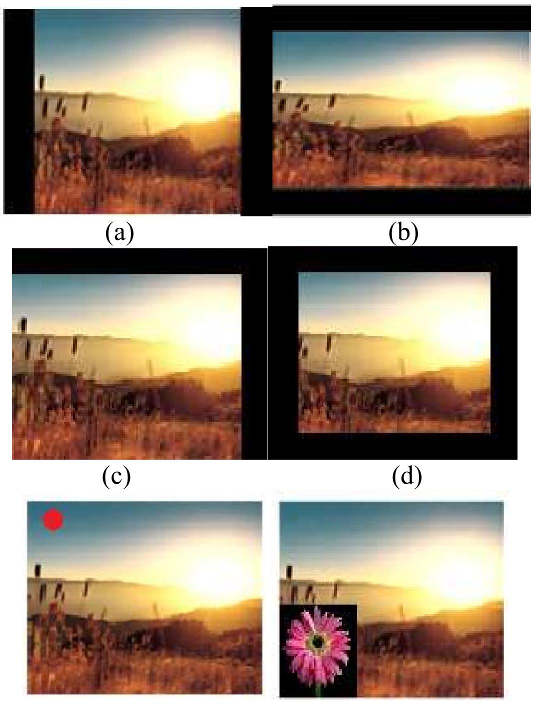

Figure 4: Illustration of six geometric operations

Thus, there is no change for the feature. In flipping (vertical or horizontal), the values of the pixels are not changed. The operations only change the pixel positions. So the feature will be preserved.

3) **Letter-Box and Pillar-Box**: As illustrated in Fig. 4(a) and (b), letter-box and pillar-box operations occur in the video when black bars are placed on the sides of the video. This is a commonly used operation for modifying the aspect ratio of the video. For both of these operations, only the black pixels are added. Since such added pixels have the value of 0 for their red, green, and blue components, they will be removed in feature extraction, thus the feature will be preserved.

4) **Cropping and Shifting**: As illustrated in Fig. 4(c) and (d), cropping and shifting can be modeled by replacing the original image region with black pixels.

5) **Insertion of Pattern and Picture in Picture**: As illustrated in Fig. 4(e) and (f) these two operations can be modeled by the replacement of an original image region with a given pattern or picture.

6) **Contrast Enhancement**: Contrast enhancement like histogram equalization and gamma correction, is also a commonly used operation for image manipulation. Unlike grey-scale images, color images contain color information for each pixel.

E. SVM Classifiers

Support Vector Machine is a non-linear and non-parametric classification technique, which shows good results in different fields. SVMs are a new technique suitable for binary classification tasks. SVM is related to non-parametric applied statistics, neural networks and machine learning.

Linear SVM

In The linear svm, the data can be classified into two groups. Here SVM chooses the best hyperplane to separate the two classes. The data point of one class is separated from the other using this hyperplane. When the margin between the two classes is largest, it is the best hyperplane. The data points that are closest to the separating hyperplane are called support vectors. Maximizing the margin is good because it implies that only support vectors are important i.e. other training examples are ignorable and empirically it works very well.

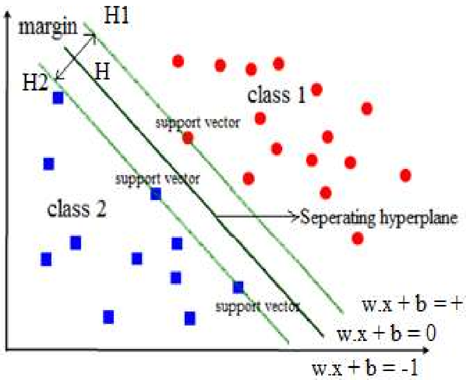

Figure 5: A separating hyperplane H which separates two classes

The main factor of linear SVM is that the classifier is the hyperplane and the training points are the support vectors. The support vectors define the hyperplane. The fig. 5 illustrates these definitions, with Red colour indicating data points of class1, and Blue colour indicating data points of class 2. All hyperplanes in the region are represnted by w ,a vector and b,a constant. using the algebraic equation for a hyperplane,The hyperplane H can be expressed as w•x+b=0. The linear SVM finds a hyperplane f(x)=sign(w•x+b), that classify the data correctly.The hyperplane H can be defined such that:

$x_i \cdot w + b \geq +1$ when $y_i = +1$
$x_i \cdot w + b \leq -1$ when $y_i = -1$

where Data: $<x_i, y_i>$, $i=1,..,l$, $x_i \in R^d$ and $y_i \in \{-1,+1\}$

H1 and H2 are the planes and the Support Vectors are the points on the planes H1 and H2
H1: $x_i \cdot w + b = +1$
H2: $x_i \cdot w + b = -1$

Using the distance formula, the distance between H and H1 is $|w \cdot x+b|/||w||=1/||w||$ Therefore,the distance between H1 and H2 is: $2/||w||$.The main goal of linear SVM is that to maximize the margin, i.e to minimize $\frac{1}{2}w^t w$ with the condition that no datapoints lie between H1 and H2 and also correctly classify all training data $x_i \bullet w+b \geq +1$ when $y_i =+1$; $x_i \bullet w+b \leq -1$ when $y_i =-1$ which can be combined into $y_i(x_i \bullet w) \geq 1$.For that Quadratic Optimization Problem is formulated and solved for w and b.

The solution has the form:
$w =\Sigma\alpha_i y_i x_i$ and $b= y_k- w^T x_k$

for any x_k such that $\alpha_k \neq 0$

Each non-zero α_i indicates that corresponding $\mathbf{x_i}$ is a support vector. Then the classifying function will have the form:

$f(x) = \Sigma\alpha_i y_i x_i^T x + b$

Soft Margin

If the case is not perfectly separable then the margin is called soft. This means that some error had occurred in the classification of data which should be minimized. Here a non negative slack variable, ξi is introduced. To allow misclassification of difficult or noisy examples Slack variables ξi can be added. In most cases ξi =0, that means data is correctly classified. In the case of a positive ξi the data point i is misclassified. The criterion for calculating w and b is that all misclassifications have to be minimized that is,

$\Phi(w) =½ w^T w + C \Sigma\xi i$ is minimized and for all $\{(\mathbf{x_i}, y_i)\}$

By imposing the constraint that no data point should be within the margin except some classification errors, SVMs require that either

$x_i \bullet w+b \geq 1- \xi$
or
$x_i \bullet w+b \geq 1+ \xi$

which can be summarized with:

$y_i (x_i \bullet w+b) \geq 1- \xi_i$

for all i=1,2,3,…n

Non-linear SVMs

In non-linear SVMs the classification problem is solved by transforming the input space using Φ into a feature space of a higher dimension, where it is easier to find a separating hyperplane. It finds a hyperplane which separates the data's in the feature space and classify data points in that space .It does not need to represent the space explicitly but defines a kernel function.

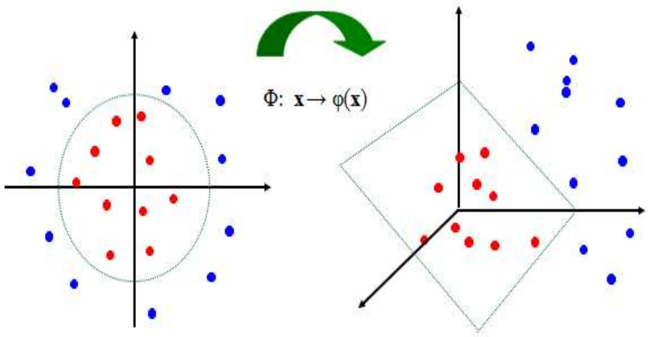

Figure 6: Transformation of an input space to feature space

The role of the dot product in the feature space is played by the kernel function. Fig. 6 illustrates the transformation of an input space to a high dimensional feature space by kernel function φ linear classifier depends on dot product between vectors K(xi,xj)=xiTxj.. If every data point is mapped into high-dimensional space with the use of some transformation Φ: x → φ(x), the dot product will be,

K(xi,xj)= φ(xi) Tφ(xj)

A kernel function is a function which corresponds to an inner product in a feature space. The non-linear SVM is thus a linear combination, but with new variables, which are derived through a kernel transformation .for nonlinear SVM, the solution is
f(x) = ΣαiyiK(xi, xj)+ b

Multiclass SVM

SVM classifiers are mainly used in the case of classification between two classes. They can also be used to handle multiple classes. Then they are called Multiclass Support Vector Machines. Multiclass SVM is used here for video sequence matching. This enables the matching operation to become more fast and accurate.

In multiclass SVM, there are two functions called Svmtrain ( ) and Svmclassify ( ). Svmtrain ( ) trains a Support Vector Machine using the training data. Svmclassify ( ) is used to classify new data using this trained Support Vector Machine.

F. Matching Operation

The comparison of two video feature set occurs in this section. The comparison is done to check whether two of them are equal or not. In the matching operation, the feature set of the database video which is already saved is compared with the feature set of the input or query video. The feature set of the video is made using the RGB component relations. The main component in the matching operation is multiclass SVM Classifier.

```
function [result] = multisvm (TrainingData,
 Group,TestData)
numClasses=length(Group);
%build models
for k=1:numClasses
models(k) = svmtrain(TrainingData,Group(k));
end
```

```
%classify test cases
for j=1:size(TestSet,1)
for k=1:numClasses
 if(svmclassify(models(k),TestData))
break;
end
end
result(j) = k;
end
```

In multiclass SVM Classifier there are two functions called Svmtrain( ) and Svmclassify( ). Svmtrain( ) trains a Support Vector Machine using the training data. Svmclassify ( ) is used to classify new data using this trained Support Vector Machine.

The above given code is of multiclass SVM. The trainingData implies the database videos feature set and the TestData implies the query video feature set. Group corresponds to number of database videos. The function svmtrain( ) trains the support vector machine with the feature set of database videos. For each of the database video a model is created. The function svmclassify( ) use this model which consist of different data about the database videos. It uses the model information to classify the input video to any of the database video. The classification of the input video to any one of the database video means they are matching. In this way the matching operation takes place. This provides more accuracy to the comparison of two videos. The proposed system is also very fast in performing video sequence matching.

5. Conclusion

A very promising feature for video sequence matching based on multiclass SVM Classifiers and the intensity of RGB component is introduced. The feature extracted from the video using the invariance of RGB component is robust against most content preserving operations. It is already proved in the existing system. Multiclass SVM Classifiers does the matching operation between the two videos with greater accuracy and fastness than the existing system. It trains the SVM classifier with the feature set of the database video using svmtrain function. Thus a model is created for each of the database videos which consist of information about them. Then the query video is classified to one of these models which shows the similarity between the corresponding videos. Thereby, the matching video could be found with greater accuracy. The proposed system is also very fast in performing the video sequence matching operation. The extensive results showed the effectiveness of the proposed method compared with existing works .The space complexity is also satisfactory.

6. Future Enhancement

The feature of the video is determined using the RGB relation. For more accuracy of the feature calculation, temporal or spatial constraints could also be added or its integrated form could be used to enhance the feature calculation. The classification could also be extended to multiple query videos which is classified to multiple database videos.

References

[1] Yanqiang Lei, Weiqi Luo, Yuangen Wang, and Jiwu Huang, "Video Sequence Matching Based on the Invariance of Color Correlation," IEEE Transactions On Circuits And Systems For Video Technology, Vol. 22, No. 9, September 2012
[2] F. Hartung and M. Kutter, "Multimedia watermarking techniques," Proc. IEEE, vol. 87, no.7, pp. 1079–1107, Jul. 1999.
[3] Y. Li and R. H. Deng, "Publicly verifiable ownership protection for relational databases," in Proc. ACM Symp. Inform. Comput. Commun. Security, 2006, pp. 78–89.
[4] C. Kim, "Content-based image copy detection," Signal Process.: Image Commun., vol. 18, no. 3, pp. 169–184, Mar. 2003.
[5] J.-H. Hsiao, C.-S. Chen, L.-F. Chien, and M.-S. Chen, "A new approach to image copy detection based on extended feature sets," IEEE Trans. Image Process., vol. 16, no. 8, pp.2069–2079, Aug. 2007.
[6] R. Mohan, "Video sequence matching," in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., vol. 6. May 1998, pp. 3697–3700.
[7] D. N. Bhat and S. K. Nayar, "Ordinal measures for image correspondence,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 4, pp.415–423, Apr. 1998.
[8] C. Kim and B. Vasudev, "Spatiotemporal sequence matching for efficient video copy detection," IEEE Trans. Circuits Syst. Video Technol., vol. 15,no. 1, pp. 127–132, Jan. 2005.
[9] L.Chen and F. Stentiford, "Video sequence matching based on temporal ordinal measurement," Patter Recognit. Lett., vol. 29, no. 13, pp. 1824– 1831, Oct. 2008.
[10] L. Gao, Z. Li, and A. Katsaggelos, "An efficient video indexing and retrieval algorithm using the luminance field trajectory modeling," IEEE Trans. Circuits Syst.Video Technol., vol. 19, no 10, pp. 1566–1570, Oct. 2009.
[11] L. Shang, L. Yang, F. Wang, K.-P. Chan, and X.-S. Hua, "Real-time large scale near-
[12] duplicate web video retrieval," in Proc. ACM Int. Conf. Multimedia, 2010, pp. 531–540.
[13] P. Brasnett, S. Paschalakis, and M. Bober, "Recent developments on standardization of MPEG-7 visual signature tools," in Proc. IEEE Int. Conf. Multimedia Expo., Sep.2010, pp.1347–1352.
[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vision, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[15] M. Heikkil¨a, M. Pietik¨ainen, and C. Schmid, "Description of interest regions with local binary patterns," Pattern Recognit., vol. 42, no. 3, pp. 425–436, Mar. 2009.
[16] Y. Caspi, D. Simakov, and M. Irani, "Feature-based sequence-to sequence matching," Int. J.Comput. Vision, vol. 68, no. 1, pp. 53–64, Mar. 2006.
[17] B. Liu, Z. Li, M. Wang, and A. K. Katsaggelos, "In-sequence video duplicate detection with fast point-to-line matching," in Proc. IEEE Int. Conf. Image Process., Sep. 2010, pp. 1037–1040.

[18] S. Lee and C. D. Yoo, "Robust video fingerprinting for content-based video identification, "IEEE Trans. Circuits Syst. Video Technol., vol. 18,no. 7, pp. 983–988, Jul. 2008.
[19] Vojislav Kecman ,"Support Vector Machines Basics"School Of Engineering Report 616,School Of Engineering,The University Of Auckland, April, 2004
[20] Laura Auria and Rouslan A. Moro, "Support Vector Machines (SVM) as a Technique for Solvency Analysis," Berlin, August 2008
[21] Christopher J.C. Burges," A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery 2, 121-167, 1998
[22] M. J. Swain and D. H. Ballard, "Color indexing," Int. J. Comput. Vision, vol. 7, no. 1, pp. 11–32, Jun. 1991