International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 8
August 2017

Academic Science

Content Based Video Retrieval using Neural Network

Priya Singh
PG student: Department of EXTC
Thakur College of Engineering and Technology
Mumbai, India

Sanjeev Ghosh
Associate Professor: Department of EXTC
Thakur College of Engineering and Technology
Mumbai, India

ABSTRACT

A content based video retrieval system that is based on content fingerprinting and artificial neural networkbased classification is proposed.A video fingerprint is obtained from a segment of video content. The video fingerprinting methods obtain unique features of a video that differentiates each video clip from other videos.The system extracts features using a fingerprint extraction algorithm followed by fingerprint matching using Neural network. Firstly, the Fingerprint Extraction algorithm is employed which extracts a fingerprint through the features from the image content of video. These images are represented as Temporally Informative Representative Images (TIRI). Then, the second step is to find the presence of videos in a video database having content similar to that of query video. Multi Layer Feed Forward (MLF) Neural networkthat uses Backpropagation Algorithm for training is used for video retrieval. On input of query video the videos having similar content in database are retrieved and displayed

Keywords

Fingerprint, TIRI-DCT, feature extraction, ANN, MLF

1. INTRODUCTION

In today's world, due to the rapid advancements in digital devices, Internet infrastructures, and Web technologies the capture, storage, uploads and delivery of videos has become effortless. Duplicate contents in video frustrate the user. Globally accepted indexing technique and video retrieval are not well defined or available. Most of the multimedia search systems depend on available contextual information or metadata in text form. These problems motivate us to present video mining from multimedia warehouse using multimodal features.

For those using applications like digital libraries, publications, education, broadcasting and entertainment and multimedia systems such applications are useful only when video retrieval systems are efficient enough to retrieve videos and all other important information from large database as quick as possible. However, it is extremely difficult for the available web search engines to search for video over the web so new methodologies are required that are capable of manipulating the video information according to the content.Most of the web based video retrieval systems work by indexing and searching videos based on texts associated with them but this technique does not perform well because the texts do not contain sufficient information of the videos [1]. Since video retrieval is not effective using conventional query-by-text retrieval technique, Content Based Video Retrieval (CBVR) is considered as one of the best practical solutions for better retrieval quality [2].

Content Based Video Retrieval (CBVR) has been increasingly used to describe the process of retrieving desired videos from a large collection on the basis of features that are extracted from the videos. The extracted features are used to index, classify and retrieve desired and relevant videos while filtering out undesired ones. This is leading the area of CBVR into a direction promising to create more effective video search engines in future [3].

The traditional way of searching for a part of video based on search criteria, lacks expressiveness and precision. The user starts with searching the video database for a video document that contains the specified search criteria. The process results in making a reference to the matching video document. Users view the video documents sequentially to locate the required clips. This approach is time consuming, imprecise and

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 8
August 2017

inefficient in applications with a vast amount of video data.so to overcome these limitations technologies are needed for video document to support content based searching and retrieval of video information.

In this paper a technique called video fingerprinting is used that extracts features in the form of fingerprints. A fingerprint is a content-based signature derived from a video (or other form of a multimedia asset) so that it specifically represents the video or asset. To find a video similar to query video in a video database, one can search for a close match of its fingerprint in the corresponding fingerprint database (extracted from the videos in the database). Closeness of two fingerprints represents a similarity between the corresponding videos; two perceptually different videos should have different fingerprints. Neural network is used here for video retrieval.

2. STRUCTURE OF FINGERPRINTING SYSTEM

Fig. 1 shows the overall structure of this fingerprinting system. Content-Based video retrieval system retrieves the video by comparing the fingerprint of the query video with the fingerprints of the database videos. To find videos similar to query video in a video database, one can search for a close match of its fingerprint in the corresponding fingerprint database (extracted from the videos in the database). Closeness of two fingerprints represents a similarity between the corresponding videos.

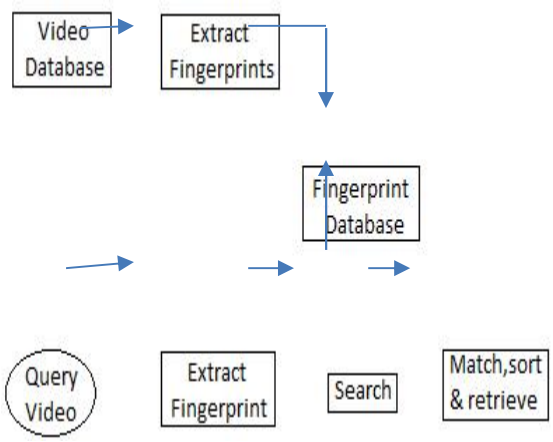


Fig 1. Overall Structure of Fingerprinting System

3. PROPERTIES OF FINGERPRINTS

Ideally, the design of a video fingerprint should have the following characteristics.

3.1 Robustness

Robustness of a fingerprint requires that it changes as little as possible when the corresponding video is subjected to content-preserving distortions.

3.2 Discriminant

Different video content should have distinct video fingerprint.

3.3 Easy to compute

The fingerprint should be easy to compute. A fingerprinting algorithm should be able to extract the signatures as the video is being uploaded for online applications

3.4 Compact

It is a time consuming process to find a match for a fingerprint that is not compact in a very large database

3.5 Secure

The fingerprinting system should be secured, so as to prevent an adversary from tampering with it.

3.6 Low complexity

The algorithm for extracting video fingerprints should have low computational complexity so that a video fingerprint can be computed fast.

4. TRADITIONAL VIDEO FINGERPRINTING EXTRACTION ALGORITHMS

4.1 Color-space-based fingerprints

The first feature extraction methods used for video fingerprinting includes Color-space-based fingerprints.According to Lienhart [4]depending on the size of the color region the pixels belong to, the color coherence vector (CCV) differentiates between pixels of the same color. Lienhart et al and Sanchez et al have been tested for the domain of TV commercials and are susceptible to color variations.First disadvantage of color-space-based fingerprint is that they are not applicable to black and white videos. Another drawback of color features is that the color features change with different video formats.

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 8
August 2017

4.2 Temporal fingerprints

In order to overcome the drawback of color-space-based fingerprints, new video fingerprint extraction algorithm is developed that can be applied to the luminance (the gray level) value of the frames. According to Shivkumar and Indyk [5], first, shots are formed by segmenting a video sequence. Then, a temporal signature is obtained from the duration of each shot, and the sequence of concatenated shot durations form the fingerprint of the video. From adjacent frames of a videoTemporal signatures are computed. Temporal fingerprints are extracted from the characteristics of a video sequence over time. Since sufficient discriminant temporal information is not present in short video clips these features do not work well with short video clips but they perform well for long sequences of long videos.

4.3 Spatial fingerprints

A video image is converted into YUV color space in Spatial fingerprint algorithm, keeping the luminance (Y) and discarding the chrominance components (U, V). Spatial fingerprints are the features that are derived from each frame or from a key frame. They are widely used for both image and video fingerprinting. Spatial fingerprints can be further subdivided into local and global fingerprints. Local fingerprints usually represent local information around some interest points within a frame like edges, corners, etc, while Global fingerprints focus on the global properties of a frame or a subsection of it like image histograms. One limitation of spatial fingerprints is that they are unable to capture the video's temporal information, which is an important discriminating factor, therefore Spatio-temporal fingerprints are developed.

4.4 Spatio-temporal fingerprints

Better performance is expected from Spatio-temporal fingerprints that contain both spatial and temporal information about the video thanfingerprints that use only spatial or temporal fingerprints. Some spatio-temporal algorithms consider a video as a three-dimensional (3-D) matrix and extract 3-D transform-based features[6][7].

5. PROPOSED TIRI-DCT SYSTEM

There are some disadvantages of existing fingerprint extraction systems.They are as follows

- Applying 3-D transform to a video is a computationally demanding process.
- The computational bottleneck is the search time in the matching process rather than the fingerprint extraction time.

Temporally Informative Representative Images-Discrete Cosine Transform (TIRI-DCT) system overcomes these drawbacks. As a Temporally informative representative Images (TIRI) contains spatial and temporal information of a short segment of a video sequence, the spatial feature extracted from a TIRI would also contain temporal information. An efficient fingerprinting algorithm known as Temporally Informative Representative Images-Discrete Cosine Transform (TIRI-DCT) is an improved version of 3D-DCT and is based on TIRIs.

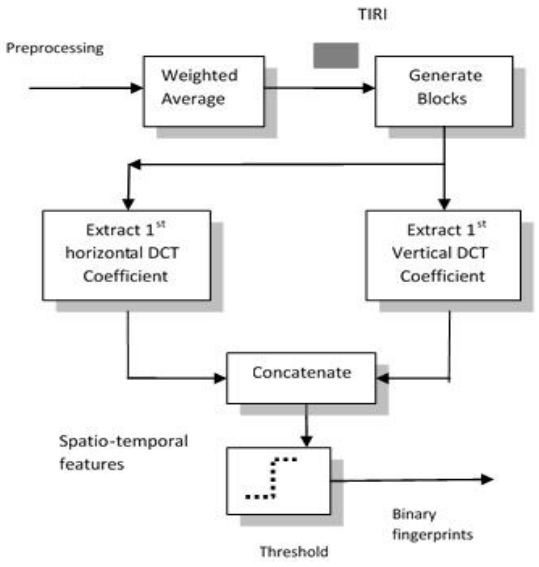


Fig 2. Schematic of TIRI-DCT Algorithm

Fig. 2 shows the block diagram of TIRI-DCT which is based on temporally informative representative images (TIRIs). In TIRI-DCT the input video signal is processedbefore extracting the fingerprints.Copies of the same video with different frame sizes and frame rates usually exist in the same video database. As a result, a fingerprinting algorithm should be robust to changes in the frame size as well as the frame rate. Down-sampling[8] can increase the robustness of a fingerprinting algorithm to these changes. This down-sampling

process provides the fingerprinting algorithm with inputs of fixed size (W X H) pixels and fixed rate (F frames/second). After preprocessing, the video frames are divided into overlapping segments of fixed-length, each containing J frames. The fingerprinting algorithms are applied to these segments. Overlapping reduces the sensitivity of the fingerprints to the synchronization problem which is called as time shift. As TIRI-DCT transform algorithm captures the temporal information in a video using the feature extraction process. TIRI-DCT Algorithm includes following steps

Step 1: Generate TIRIs from each segment of J frames after preprocessing of input video .TIRIs are generated using $W_k = \gamma^k$

Step 2: Segment each TIRI into overlapping blocks of size $2\omega \times 2\omega$, using

$$B^{i,j} = \{l'_{x.y} \,|\, x \in i\omega \pm \omega, y \in j\omega \pm \omega\} \ldots\ldots\ldots\ldots\ldots (1)$$

Where $i \in \{1,1,2,\ldots,W/\omega - 1\}$ and

$j \in \{0,1,2,\ldots,H/\omega - 1\}$

When indexes are outside of boundary then TIRI image is padded with 0's.

Step 3: Extract DCT coefficient from each TIRI block. These are first horizontal and first vertical coefficients adjacent to the DC coefficient. First vertical frequency $\alpha_{i,j}$ can be found for $B^{i,j}$ as

$$\alpha_{i,j} = v^T B^{i,j} 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where

$v = [\cos(0.5\pi/2\omega), \cos(1.5\pi/2\omega), \ldots\ldots\ldots \cos(1 - 0.5\pi/2\omega)]^T$

And 1 is column vector of all ones. Similarly first horizontal frequency $\beta_{i,j}$ can be found for $B^{i,j}$ as

$$\beta_{i,j} = 1^T B^{i,j} v \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Step 4: Concatenate all coefficients to form feature vector f.

Step 5: Find median m, using all elements of f.

Step 6: Generate binary hash h, using f

$$h_k = \begin{cases} 1, & f_k \geq m \\ 0, & f_k < m \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

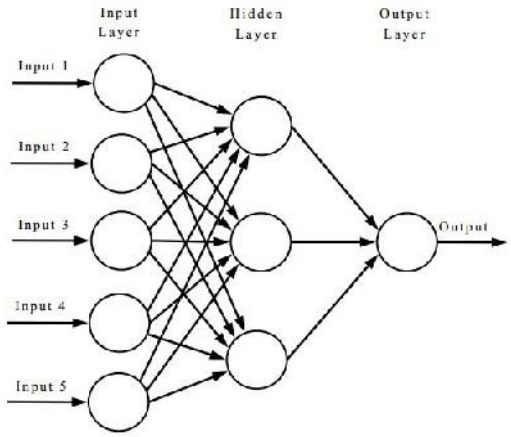
6. ARTIFICIAL NEURAL NETWORK BASED CLASSIFIER

Artificial Neural Network (ANN) based classifier is used based on the computational simplicity. Neural network is a computing system which is made up of a number of simple, highly interconnected processing elements that process information by their dynamic state response to external inputs. Neural networks are typically organized in layers. Layers are made up of a number of interconnected nodes which contain an activation function.The input layer presents patterns to the network and communicates to one or more hidden layers where the actual processing is done through a systemof weighted connections. The hidden layers then link to an output layer where the answer is output.

In this proposed system, a feed forward multilayer network is used and makes use of back propagation algorithm for training. ANNs contain some form of learning rule which modifies the weights of the connections according to the input patterns that it is presented with. In short, ANNs learn by example. Learning is a supervised process that occurs each time the network is presented with a new input pattern through a forward activation flow of outputs, and the backwards error propagation of weight adjustments. More simply, when a neural network is initially presented with a pattern it makes a random guess as to what it might be. It then sees how far its answer was from the actual one and makes an appropriate adjustment to its connection weights.

Neural network analysis often requires a large number of individual runs to determine the best solution. Once a neural network is 'trained' to a satisfactory level it may be used as an analytical tool on other data. To do this, the user no longer specifies any training runs and instead allows the network to work in forward propagation mode only. New inputs are presented to the input pattern where they filter into and are processed by the middle layers as though training were taking place, however, at this point the output is retained and no back propagation occurs. The output of a forward propagation run is the predicted model for the data which can then be used for further analysis and interpretation.

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 8
August 2017



.Fig 5. Typical Feed Forward Network Composed of Three Layers

7. RESULT

The proposed system is evaluated using several kinds of videos. The system was experimented using a database of videos against 1 query video. In order to evaluate the quality of the proposed system, recall and precision values of the retrieved results are used. Precision is the fraction of retrieved videos that are relevant whereas recall is the fraction of relevant videos that are retrieved.

Precision=

$$\frac{\text{Number of relevant videos that are retrieved}}{\text{Total number of videos that are retrieved}}$$

Recall=

$$\frac{\text{Number of relevant videos that are retrieved}}{\text{Total number of relevant videos in database}}$$

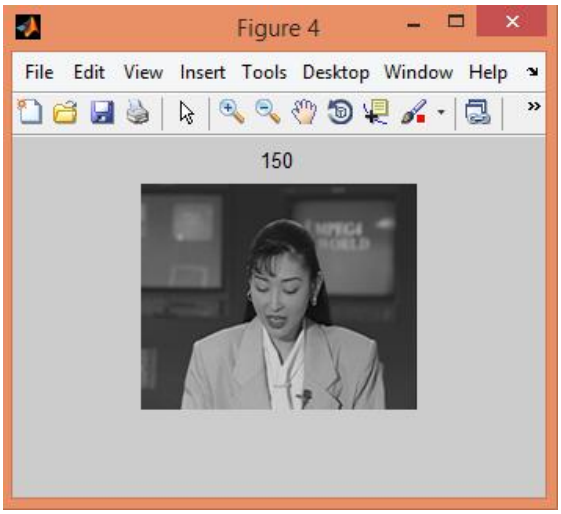


Fig 6. Query video

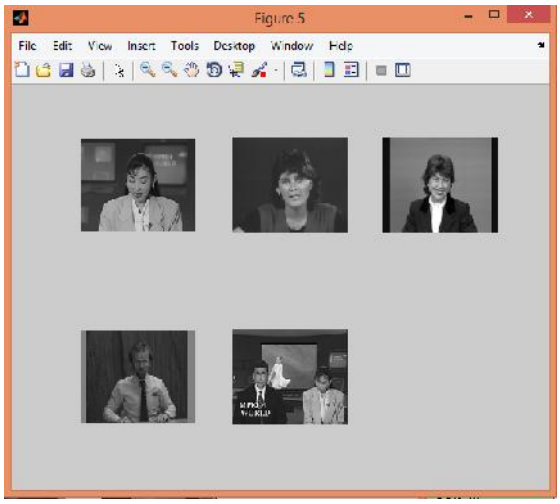


Fig 7. Retrieved videos

In the proposed system the database is divided into 7 different class. On input of a query video from a particular class the neural network searches the entire database for retrieving the videos that are similar in content. A threshold value of 1×10^{-1} is set. Query video is then compared with each and every video in the database. If the MSE obtained is within the threshold value then that particular video gets retrieved. From the above result it can be seen that on input of news video all the videos that belonged to news class were retrieved.

There are 7 different classes each containing several videos. One video from each class is taken as a sample. The precision and recall values obtained for these sample videos are as templated below.

Table 1. Precision and recall values of videos

Video	Precision(%)	Recall(%)
1	100	100
2	0	0
3	100	100
4	100	100
5	100	100
6	100	100
7	0	0

From the above table it can be seen that the videos retrieved on input of query video are either completely relevant or completely irrelevant. On testing, the system resulted in 71.42% precision and 71.42% recall in average.

8. CONCLUSION

In this paper, we briefly review the need and significance of video retrieval systems and explain their basic building stages. The first step is feature extraction, here which is extracted using video fingerprinting using TIRI-DCT algorithm. Feature matching is then performed using neural network. For video retrieval a feed forward multilayer network is used and makes use of back propagation algorithm for training. The system was then tested on a database of videos and an average precision of 78.04% and recall of 78.04% was achieved. The proposed system outperforms the existing video retrieval system.

REFERENCES

[1] Liang-Hua Chen, Kuo-Hao Chin, Hong-Yuan Liao, "An integrated approach to video retrieval", Proceedings of the nineteenth conference on Australasian database- Volume 75, 49–55, 2008

[2] Ja-Hwung Su, Yu-Ting Huang, Hsin-Ho Yeh, Vincent S. Tseng , "Expert Systems with Applications", 37, pg 5068-5085, 2010

[3] P. Indyk, G. Iyengar, and N. Shivakumar. Finding pirated video sequences on the internet, Technical report, Stanford University, 1999.

[4] B. Coskun, B. Sankur, and N. Memon, Spatiotemporal transform based video hashing, IEEE Trans. Multimedia, vol. 8, no. 6, Dec. 2006

[5] Radhakrishnan and C. Bauer, Robust video fingerprints based on Subspace embedding,in Proc. ICASSP, Apr. 2008

[6] Mani Malek Esmaeili, Mehrdad Fatourechi and Rabab Kreidieh Ward ,Robust and Fast Video Copy Detection System Using Content Based Fingerprinting, IEEE Trans On Information Forensics And Security,Vol.6 No.1,March 2011

[7] J. Oostveen, T. Kalker, and J. Haitsma, Feature extraction and a database strategy for video fingerprinting, in Proc. Int. Conf. Recent Advances in Visual Information Systems (VISUAL), London, U.K., 2002, Springer-Verlag

[8] V.S. Tseng, J-H Su, J.-H. Huang, & C-J. Chen,"Integrated mining of visual features, speech features and frequent patterns for semantic video annotation." IEEE Transactions on Multimedia, 10(1), 2008

[9] Virga, P., Duygulu, P., "Systematic evaluation of machine translation methods for image and video annotation", In Proceedings of the fourth international conference on image and video retrieval (pp. 487–496), Singapore, 2005.

[10] T.N.Shanmugham, Priya Rajendran, "An Enhanced Content-Based Video Retrieval System Based on Query Clip", International Journal of Research and Reviews in Applied Sciences, Volume 1, Issue 3, 2009

[11] Xiang Bai, Bo Wang, Cong Yao, Wenyu Liu, ZhuowenTu, "Co-Transduction for Shape Retrieval", IEEE transactions on Image Processing, vol. 21, no. 5, May 2012

[12] SalahuddinA., NaqviA., MurtazaK.,AkhtarJ., "Content Based Video Retrieval Using Particle Swarm Optimization" , Proc. 10th International Conference on Frontiers of Information Technology (FIT), DOI: 10.1109/FIT.2012.23, pp. 79 – 83, 2012

[13] B. V. Patel, A. V. Deorankar, B. B. Meshram, "Content Based Video Retrieval using Entropy, Edge Detection, Black and White Color Features", Proc. 2nd International Conference on Computer Engineering and Technology (ICCET), 2010, vol.6, pp.V6-272 - V6-276

[14] S.Padmakala,Dr.G.S.AnandhaMala, M.Shalini, "An Effective Content Based Video Retrieval Utilizing Texture, Color and Optimal Key Frame Features", International Conference on Image Information Processing (ICIIP 2011), 978-1-61284-861-7/11, 2011

[15] Di Zhong, Shih-Fu Chang, "An Integrated Approach for Content-Based Video Object Segmentation and Retrieval", IEEE transaction on Circuitsand Systems for Video Technology, vol. 9, no. 8, 1999