Asian Journal of Information Technology 11 (4): 144-150, 2012
ISSN: 1682-3915

A Robust Video Copy Detection System Using TIRI-DCT Fingerprints and DWT Statistical Features

[1]S. Devi and [2]N. Vishwanath
[1]Department of Computer Science, PET Engineering College
[2]Department of Computer Science and Engineering, Cape Institute of Technology,
Vallioor, Tirunelveli District, Tamil Nadu, India

Abstract: A video copy detection system that is based on content fingerprinting can be used for video indexing and copyright applications. Most of the video copy detection algorithms proposed so far focus mostly on coping with signal distortions introduced by different encoding parameters; however, these algorithms do not cope well with display format conversions. They may rely on a fingerprint extraction algorithm followed by a fast approximate search algorithm. The fingerprint extraction algorithm extracts compact content-based signatures from special images constructed from the video. Each such image represents a short segment of the video and contains temporal as well as spatial information about the video segment. These images are denoted by temporally informative representative images. To find whether a query video is copied from a video in a video database, the fingerprints of all the videos in the database are extracted and stored in advance. The search algorithm searches the stored fingerprints to find close enough matches for the fingerprints of the query video. The content based fingerprint extraction process does not work to get a better level search. Further in an enhancement of this concept of fingerprints this system handles the TIRI-DCT and DWT features to detect the copyright information. This study proposes a novel sequence matching technique to detect copies of a video clip. If a video copy detection technique is to be effective, it needs to be robust to the many digitization and encoding processes that give rise to several distortions including changes in brightness, color, frame format as well as different blocky artifacts. It also handles a new nonmetric distance measure to find the similarity between the query and a database video fingerprint and it is proposed to achieve accurate duplicate detection. Then the performance of the TIRI DCT and the co-efficient is compared. The proposed method has been extensively tested and the results show that the proposed scheme is effective in detecting copies which has been subjected to wide range of modifications.

Key words: Content-based fingerprints, feature extraction, copyright protection, video copy detection, video matching, clusters, fingerprints

INTRODUCTION

Growing broadcasting of digital video content on different media brings the search of copies in large video databases to a new critical issue. Copying of video increases with the rapid development of multimedia technologies and media streaming, copyrighted materials become videos and can be easily copied, stored and distributed over the Internet. As video is the most complex type of digital media, it has so far received the least attention regarding copyright management. Because videos are available in different formats, it is more efficient to base the copy detection process on the content of the video rather than its name, description or binary representation. This situation, aside from enabling users to access information easily, causes huge piracy issues. One possible solution to identify copyrighted media is watermarking.

Digital watermarking: Digital watermarking (Langelaar *et al*., 2000) was proposed for copyright protection and fingerprinting. The basic idea is to embed information into the signal of the media (audio, video or photo). Some watermarks are visible (e.g., text or logo of the producer or broadcaster) while others are hidden in the signal which cannot be perceived by human eye. Watermarking technique is not designed to be used for video retrieval by querying with a sample video clip. As

Corresponding Author: S. Devi, Department of Computer Science, PET Engineering College, Vallioor, Tirunelveli District, Tamil Nadu, India

a disadvantage, watermarks are generally fragile to visual transformations (e.g., re-encoding, change of the resolution/bit rate).

Content-Based Copy Detection (CBCD): Content-Based Copy Detection (CBCD) is introduced as an alternative or in fact, a complementary research field to watermarking approach (Joly *et al*., 2007; Sivic and Zisserman, 2003). Video copy detection is a challenging problem in computer vision due to the following reasons. First of all, the problem domain is exceptionally wide. Depending on the purpose of a video copy detection system, different solutions can be applied. For example, a simple frame-based color histogram similarity approach could be enough for detecting exact duplicates of video segments or identifying commercial breaks. On the other hand, matching news stories across different channels (camera viewpoints) is a totally different problem and will probably require interesting point matching techniques. Therefore, no general solution can be proposed to video copy detection problem. Secondly, the problem space is extremely large which often requires real-time solutions. For in the case of YouTube, the system needs to process 20 h of uploaded video content per second to find an exact or near-duplicate segment of a copyrighted material.

The main idea of CBCD is that the media visually contains enough information for detecting copies. Therefore, the problem of content-based (Su *et al*., 2009; Xiaohong and Jinhua, 2008; Krishnan *et al*., 2007) copy detection is considered as video similarity detection by using the visual similarities of video clips. Multimedia fingerprinting (also known as robust hashing) has been recently proposed for this purpose (Swaminathan *et al*., 2006; Krishnan *et al*., 2007).

Fingerprints: Fingerprinting is an important tool for automated multimedia identification. It involves computing a short and compact identifier that capturesrobust and distinct properties called a fingerprint which can be used for identifying the multimedia. A fingerprint is a content-based (Su *et al*., 2009; Cheung and Zakhor, 2003; Xiaohong and Jinhua, 2008; Liu *et al*., 2007; Kekre and Thepade, 2009a, b; Krishnan *et al*., 2007) signature derived from a video (or other form of a multimedia asset) so that it specifically represents the video or asset. Video fingerprinting is a technique in which software extracts characteristic components of a video file. The characteristics can be both visual as well as audio. The fingerprints are highly compressed files and can be stored in databases for comparison. To find a copy of a 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.

In general, video segment identification concerns two challenging problems: representation and searching, namely how to select appropriate features uniquely and robustly describe the video content and how to accelerate the search process based on the extracted features. Among many video representation techniques, key frame based shot representation is the most popular one and has extensive applications in video browsing, indexing and retrieval. However when applied to video segment identification, the traditional key-frame based representation has some drawbacks. First of all, the performance of keyframe based shot representation strongly depends on the accuracy of shot segmentation algorithm and the appropriate selection of key frame to characterize the video content. Since, fingerprinting is one of the important tool for the identification of multimedia as the properties of fingerprint it should be robust to the content-preserving distortions present in a video. It should also be discriminant, easy to compute, compact and easy to search (Kekre and Thepade, 2009b) for in a large database. The proposed approach is shown in Fig. 1.

In some applications such as copyright protection, the fingerprinting system should also be secure. Robustness of a fingerprint requires that it changes as little as possible when the corresponding video is subjected to content-preserving operations, i.e., operations that do not affect the perceptual content of the video. Content-preserving attacks (distortions) are changes that are made to the video unintentionally or intentionally by users of video-sharing websites. These changes can include format changes, signal processing operations, changes in brightness/contrast, added noise, rotation, cropping, logo insertion, compression, etc. The fingerprints should also be discriminant to ensure that two perceptually different videos have distinguishable fingerprints. Because a change in the content can be considered as an extreme distortion of the video, there is a trade-off between robustness and discrimination. As a fingerprinting algorithm becomes more robust to distortions, it becomes less sensitive to changes in the content, i.e., it has less discrimination ability. The fingerprint should also be easy to compute.

More specifically for online applications, a fingerprinting algorithm should be able to extract the signatures as the video is being uploaded. A computationally demanding algorithm is not suitable for

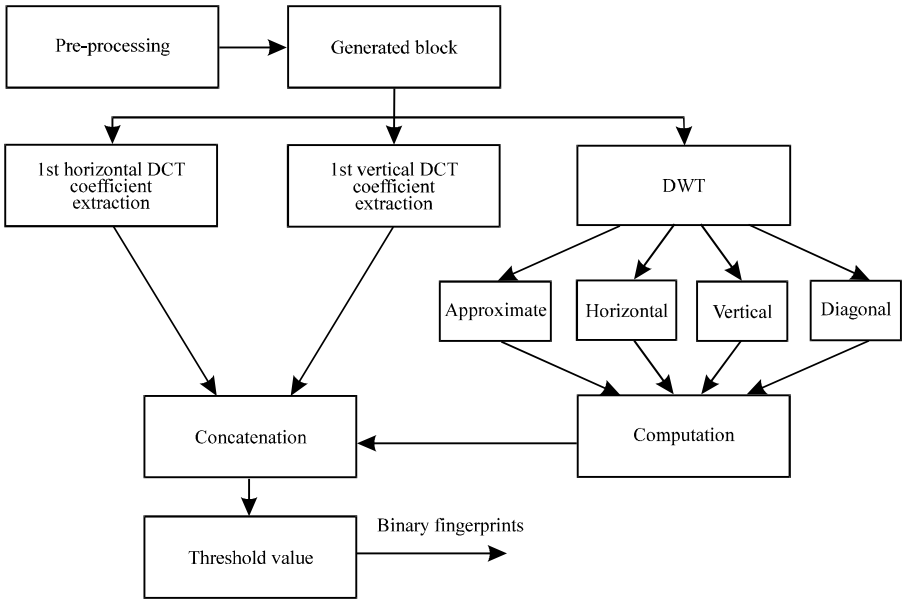

Fig. 1: Schematic of TIRI-DCT and DWT algorithm

online applications where thousands of videos need to be examined simultaneously in order to find possible copyright infringements. For the same reason, fingerprints should be compact as well. If a fingerprint is not compact, finding a match for it in a very large database can become a time-consuming process. It should be noted that the compactness of a fingerprint does not guarantee that it can be easily matched or found in a large database. The fingerprint structure should be designed to allow utilization of fast approximate search algorithms. For some applications, the fingerprinting system should be secure so as to prevent an adversary from tampering with it. Security is specifically important for copy-detection applications. The more secure a multimedia fingerprinting algorithm is the more difficult it is for an adversary to generate similar fingerprints for different videos and thus manipulate the copy-detection system. For indexing applications however, security of a fingerprinting system does not pose a problem.

EXTRACTING ROBUST AND DISCRIMINANT FINGERPRINTS

This study deals with the extraction of fingerprints.

A preprocessing: Preprocessing is the first step which is performed because the copies of the same video with different frame sizes and frame rates usually exist in the same video database. A fingerprinting algorithm should be robust to changes in the frame size as well as the frame rate. First the input video is processed to change the rate and then change the pixel size of the video. These resized

Fig. 2: a) Resized image; b) Resized gray image

video (Fig. 2a) is converted to gray scale (Fig. 2b) and again divided into short segments. Final step in the preprocessing approach of this method is to generate the fingerprints from the segments obtained from the previous steps. Thus, fingerprints are generated by this preprocessing step.

Fingerprint extraction using TIRI-DCT: In the first step a video is given as input and in the initial stage the query image is resized and changed to a standard format for the reason of efficiency of the fingerprints. Features are derived by applying a 2D-DCT on overlapping blocks of size from each TIRI. The first horizontal and the first vertical DCT coefficients are then extracted from each block. The same feature extraction process is done using the TIRI-DCT transform for the capture of the temporal information (Esmaeili and Ward, 2010; Esmaeili *et al*., 2011; Malekesmaeili *et al*., 2009; Coskun *et al*., 2006) in a video. In this TIRI Transform Method there are various different weight factors such as constant, linear and exponential which generates the images from the video. The value of the features from all the blocks is concatenated to form the feature vector. Each feature is then compared to a threshold and a binary fingerprint is generated. For TIRI-DCT (Esmaeili *et al*., 2011) all features are in the same frequency range and binarization based on a common threshold.

TIRI-DCT and DWT algorithms: The proposed approach based on TIRIs (Esmaeili *et al*., 2011) and DWT. The TIRI-DCT and DWT involves the following steps:

- For each segment S generate a TIRI block using $\omega_k = \gamma^k$
- Segment each TIRI into overlapping blocks of size 2w×2w
  $B^{i,j} = \{l`_{x,y} \mid x \in iw \pm w, y\ jw \pm w\}$ where $i \in \{0, 1, 2, .. W/w-1\}$ and $i \in \{0, 1, 2, .. W/w-1\}$
- Extract the two DCT coefficients such as horizontal and vertical DCT from each block. The first vertical frequency can be found by $\alpha_{i,j} = V^T B^{i,j} 1$. Similarly the first vertical frequency can be found by $\beta_{i,j} = 1^T B^{i,j} V$ the α and β are found for $B^{i,j}$
- Apply DWT to the TIRI block for further processing so as to extracting the statistical features:

$$n\_\theta = \frac{n\_\theta}{(\text{height x width} - n\_\theta)}$$

- The outcome of the DWT is four block which represents the approximation, horizontal, vertical and diagonal coefficients of an image
- One digital signature is extracted from directionality estimation
- All these coefficients are concatenated to form a single vector say r
- The median value m for each elements r is calculated
- Using the below formula generate the binary hash b from r

$$b_k = \begin{cases} 1, r_k \geq m \\ 0, r_k < m \end{cases}$$

Statistical feature extraction using DWT: To augment the efficiency of the system and to gather more features about the video, statistical features (Joly *et al*., 2004) are extracted using DWT. Daubechies wavelet transform is one of the DWT used to extract the statistical features. With the help of this energy in a image is brought into one single side. For each and every 50% of horizontal and vertical DCT overlapped block DWT is applied. After this processing there results four separate blocks which represents the approximation, horizontal edge, vertical edge and diagonal of the image. Here, the concatenation of horizontal and vertical coefficients is done to find the mean value. By finding the mean, standard deviation, variance and entropy for the four blocks, eight features are obtained as the output. Then, the TIRI image is given as the input and it is also processed to get a single output which is known as directional feature. So, totally nine features are extracted, eight from the statistical and one from the directionality feature. All the nine features are included and clustered (Cheung and Zakhor, 2004) to get the output result. Based upon the below steps the statistical feature n_θ is can be extracted.

If (above threshold image (i. j) = = 1.0)

$$\theta(i, j) = a\tan\left(\frac{c_x(i, j)}{c_y(i, j)}\right)$$

Else:

$$\theta\ (i, j) = 0$$
$$n\_\theta = n\_\theta + 1$$

And:

$$n\_\theta = n\_\theta + \theta\ (i, j)$$

$$n\_\theta = \frac{n\_\theta}{(\text{height x width} - n\_\theta)}$$

MATCHING OF FINGERPRINTS WITHIN A VIDEO DATABASE

This system is mainly designed, to determine whether a query video is an attacked version of a video which is already available in the database or a new version Its fingerprint is first extracted and is then searched in the fingerprint database (previously added in the video database based on the data that is available from the previous video) for the closest fingerprint to the extracted query fingerprint. The signature is extracted from the test media and is compared to the original media signature to

determine whether the test video is a duplicate copy. It should be mentioned that in copy detection, the problem is to determine if a specific query video is a pirated version of a video in the database. On the other hand, the problem of finding all copies of a video in a database is called copy retrieval and requires a different approach which is not the concern of this system. Fingerprints of two different copies of the same video content are similar but not necessarily identical. This obviously leads to find a close match of the query in the fingerprint database and not an exact match.

Inverted-file-based similarity search: An inverted file based similarity search is one of the most commonly used image retrieval process. This search method is based on the idea that for two fingerprints which are similar enough to be considered as matches, the probability of an exact match between smaller sub-blocks of those fingerprints is high (Oostveen *et al*., 2002). Each fingerprint is divided into small non-overlapping blocks of bits and are called as small blocks words. Words are then used to create an inverted file from the fingerprints of database. The horizontal dimension of this table refers to the position of a word inside a fingerprint and the vertical direction corresponds to possible values of the word. To generate a table, start the process with the first word of each fingerprint and add the index of the fingerprint to the entry in the first column corresponding to the value of this word. It is imperative to continue this process for all the words in each fingerprint and all the columns in the inverted file table. To find a query fingerprint in the database, first the fingerprint is divided into words. The query is then compared to all the fingerprints that start with the same word. The indices of these fingerprints are found from the corresponding entry in the first column of the inverted file table. The Hamming distance between these fingerprints and the query is then calculated. If a finger print has a Hamming distance, less than some predefined threshold, it will be identified as a match and if no such match is found, the procedure is iterated for the fingerprints that have exactly the same second word as the query's second word.

Clustering based search: In this study, another similarity search algorithm for binary fingerprints is proposed. The main idea is to use clustering to reduce the number of queries that are examined within the database. By assigning each fingerprint to one and only one cluster, the fingerprints in the database will be clustered into non-overlapping groups. To do so, a centroids chosen for each cluster, termed the cluster head. A fingerprint will be assigned to cluster if it is closest to this cluster's head. To determine if a query fingerprint matches a fingerprint in the database, the cluster head closest to the query is found. All the fingerprints belonging to this cluster are then searched to find a match, i.e., the one which has the minimum Hamming distance from the query. If a match is not found, the cluster that is the second closest to the query is examined. This process continues until a match is found or the farthest cluster is examined. In the latter case, the query is declared to be out of the database. The cluster heads should be chosen such that a small change in the fingerprint does not result in the fingerprint being assigned to another cluster. Each bit of the cluster head can be replicated m times and the Hamming distance between the expanded bit version of all the cluster heads and the fingerprint is calculated. The cluster head closest to the fingerprint is then assigned to that fingerprint.

EXPERIMENTAL RESULTS

The present approach can be integrated with the state-of the- art coding schemes to extract the features of videos and for efficient searching. In the experiments, two transforms are used such as DCT and DWT. Before applying the two algorithms for efficiency the size and color are changed to a standard format. The coefficients are calculated using both the algorithms so that the proposed approach is better when comparing with the other algorithms. It is shown experimentally that TIRI-DCT has a high average true positive rate of 99.05% and a low average false positive rate of 0.98%. Researchers also propose two fast approximate search algorithms: The Inverted-File-Based Method which is a generalization of an existing search method and another method based on a novel clustering-based approach.

Performance analysis of the fingerprint extraction algorithm: Usually, the performances are evaluated, in order to get the most accurate results and to study about its performance. Here, in this study the comparison is made on the performance of TIRI-DCT with DWT and TIRI-DCT when an exhaustive search is used for searching the database. Table 1 shows the results of

Table 1: Comparing TIRI-DCT with TIRI-DCT and DWT when exhaustive search is used

	TPR (%)		FPR (%)		F-Score	
Various attacks	TIRI-DCT	TIRI-DCT and DWT	TIRI-DCT	TIRI-DCT and DWT	TIRI-DCT	TIRI-DCT and DWT
Brightness	99.4000	99.4200	0.6000	0.67	0.9800	0.9900
Noise	99.2400	99.2600	0.6600	0.67	0.9800	0.9900
Contrast	99.0200	99.1000	0.7500	0.80	0.9800	0.9900
Frame loss	98.3900	98.4500	0.4200	0.50	0.9700	0.9800
Average	99.0125	99.0575	0.6075	0.66	0.9775	0.9875

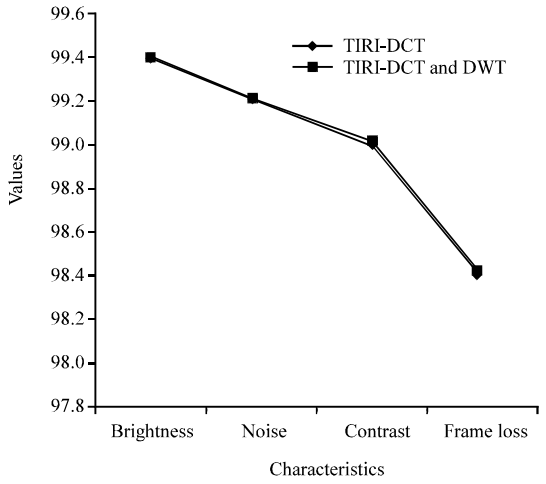

Fig. 3: TPR % comparison chart for TIRI-DCT, TIRI-DCT and DWT

Table 2: Hash Extraction time (in msec) for different segment sizes

	Length of the segment in msec		
Methods	1	2	3
3D-DCT	3.3	4.6	6.0
TIRI-DCT	1.4	1.5	1.6
TIRI-DCT and DWT	1.6	1.7	1.9

applying TIRI-DCT with DWT and TIRI-DCT DCT to the test database. Attacks were mounted independently on the videos to generate the queries. For each attack parameter, five equally spaced values were chosen from the corresponding range. Table 2 shows the average True Positive Rate (TPR), False Positive Rate (FPR). Table 1 shows that both TIRI-DCT and 3D-DCT have an average F-score of 0.99. So both algorithms have a very good performance on average but as it can be seen from the table that TIRI-DCT with DWT maintains this high performance for all the attacks while the performance of TIRI is degraded for some of the attacks.

Figure 3-5 shows that both TIRI-DCT with DWT, TIRI-DCT algorithms are robust to noise addition, changes in brightness/contrast (Table 1). The performance (for TIRI-DCT) reported above is achieved using short fingerprint lengths of 512 bits. By increasing the fingerprint length to 1024 bits, researchers were able to decrease the FPR on average by about 90% without affecting the TPR, resulting in a 0.9% increase in the F-score. One more factor is also to be consider while analyzing the performance, i.e.,) speed. When the length of the fingerprints increases automatically the performance will be lost. However, its quite general that larger fingerprints result in a decrease in detection speed as they require more computation in calculating the Hamming distances between the fingerprints.

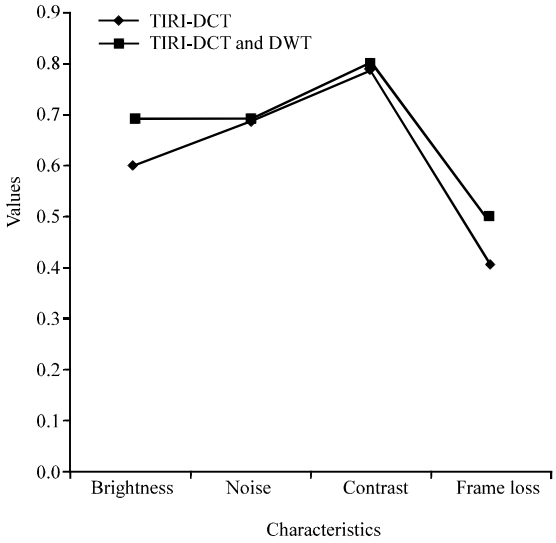

Fig. 4: FPR % comparison chart for TIRI-DCT, TIRI-DCT and DWT

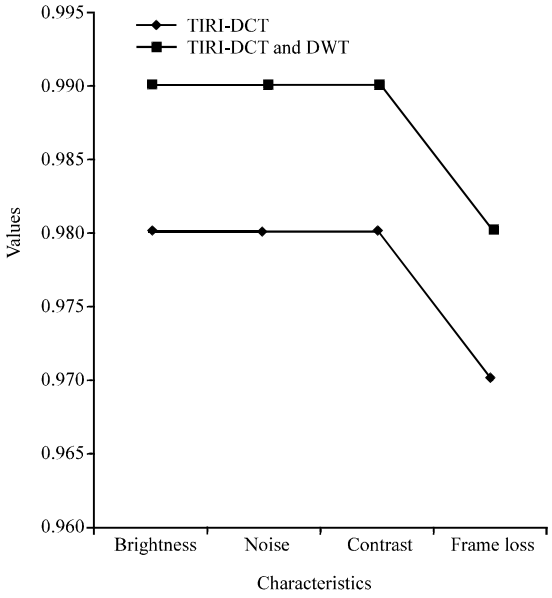

Fig. 5: F-score comparison chart for TIRI-DCT, TIRI-DCT and DWT

CONCLUSION

A video copy detection system is an emerging research area that has received a considerable amount of attention in recent years. The main goal of this video copy detection system is to find whether a query video is pattern matched with the videos in the database. This system uses strengths of TIRI-DCT, DWT algorithms for finger-print generation of a particular video and fast

search methods for efficient match of finger-prints within a large database. The contribution of this system include extracting compact signatures from TIRI image constructed from the video.

To detect whether the query video is pirated video or not the finger-prints of all the videos in the database are extracted and stored in advance. The search algorithm searches the stored fingerprints to find close enough matches for the finger-prints of the query video. All the techniques are tested and compared within the same framework by evaluating their robustness under single and mixed image transformations as well as for different lengths of video segments. In addition, these signatures require small storage and are easy to compute and compare. The algorithm is simple to implement and is more computationally efficient than previous algorithms in literature. The proposed system can be used for video indexing and copyright applications.

As part of future research, try to apply more advanced hashing methods to TIRIs and analyze their performance. One may also plan to study the performance of the algorithms in the presence of other common attacks that have not been studied such as cropping and logo insertion. In addition to that one can also try to extend this project which supports various types of videos that may be of different size and that may run on different time slots. Much more attention can be give while extracting the features such as corner, edges, etc. Different algorithms can be applied in extracting the features of a video and different searching techniques can be used so the systems acts much more efficient.

REFERENCES

Cheung, S.C. and A. Zakhor, 2004. Fast similarity search and clustering of video sequences on the world-wide-web. IEEE Trans. Multimedia, 7: 524-537.

Cheung, S.S. and A. Zakhor, 2003. Efficient video similarity measurement with video signature. IEEE Trans. Circuits Syst. Video Technol., 13: 59-74.

Coskun, B., B. Sankur and N. Memon, 2006. Spatiotemporal transform based video hashing. IEEE Trans. Multimedia, 8: 1190-1208.

Esmaeili, M.M. and R.K. Ward, 2010. Robust video hashing based on temporally informative representative images. Proceedings of the IEEE International Conference on Consumer Electronics, January 9-13, 2010, Las Vegas, NV, USA., pp: 179-180.

Esmaeili, M.M., M. Fatourechi and R.K. Ward, 2011. A robust and fast video copy detection system using content-based fingerprinting. IEEE Trans. Inf. Forensics Secur., 6: 213-222.

Joly, A., C. Frelicot and O. Buisson, 2004. Feature statistical retrieval applied to content-based copy identification. Proc. Int. Conf. Image Proc., 1: 681-684.

Joly, A., O. Buisson and C. Frelicot, 2007. Content-based copy retrieval using distortion-based probabilistic similarity search. IEEE Trans. Multimedia, 9: 293-306.

Kekre, H.B. and S.D. Thepade, 2009a. Improving the performance of image retrieval using partial coefficients of transformed image. Int. J. Inf. Retrieval, 2: 72-79.

Kekre, H.B. and S.D. Thepade, 2009b. Rendering futuristic image retrieval system. Proceedings of the National Conference EC2IT-2009, March 20-21, 2009, Mumbai, India.

Krishnan, N., M.S. Banu and C.C. Christiyana, 2007. Content based image retrieval using dominant colour identification based on foreground objects. Proceeding of the International Conference on Computational Intelligence and Multimedia Applications, December 13-15, 2007, M.S. University Washington, pp: 190-194.

Langelaar, G., I. Setyawan and R.L. Langedijk, 2000. Watermarking digital image and video data. A state-of-the-art overview. IEEE Signal Proc. Magazine, 17: 20-46.

Liu, P., K. Jia, Z. Wang and Z. Lv, 2007. A new and effective image retrieval method based on combined features. Proceeding of the 4th International Conference on Image and Graphics, Auguast 22-24, 2007, Beijing University of Technology, Beijing, pp: 786-790.

Malekesmaeili, M., M. Fatourechi and R.K. Ward, 2009. Video copy detection using temporally informative representative images. Proceeding of the International Conference Machine Learning Application, December 13-15, 2009, University of British Columbia, Vancouver, BC, Canada, pp: 69-74.

Oostveen, J., T. Kalker and J. Haitsma, 2002. Feature extraction and a database strategy for video fingerprinting. Recent Adv. Visual Inf. Syst., 2314: 117-128.

Sivic, J. and A. Zisserman, 2003. Video Google: A Text Retrieval Approach to Object Matching in Videos. IEEE Computer Society, Washington, Pages: 1470.

Su, X., T. Huang and W. Gao, 2009. Robust video fingerprinting based on visual attention regions. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, April 19-24, 2009, Taipei, Taiwan, pp: 1525-1528.

Swaminathan, A., M. Yinian and W. Min, 2006. Robust and secure image hashing. IEEE Trans. Inform. Forensics Secur., 102: 215-230.

Xiaohong, Y. and X. Jinhua, 2008. The related techniques of content-based image retrieval. Int. Sympos. Comput. Sci. Comput. Technol., 1: 154-158.