PERCEPTUAL VIDEO HASHING BASED ON WALSH HADAMARD TRANSFORM WITH APPLICATION TO INDEXING AND RETRIEVAL OF NEAR IDENTICAL VIDEOS

Asha kiran M U[1], Sandeep R[2]

[1]*Assistant Professor, Department of Electronics and Communication Engineering, Dr.T.Thimmaiah Institute of Technology, Karnataka, India*

[2]*Associate Professor, Department of Electronics and Electrical Engineering, Vidyavardhaka College of Engineering, Karnataka, India*

Abstract

Advancement in technology enables us to access videos at all times and even modify them with available media editing tools. This leads to fraudulency of the video and even the author copyrights are intruded. Hence, there is a need to identify the forgery of the video and exploitation of author copyrights. Many techniques have been described in the literature to overcome these problems; they are Cryptographic hashing, Watermarking and Perceptual hashing. Our main concern here is towards perceptual video hashing. Perceptual video hashing captures the perceptual gist of the video into a concise string called the perceptual hash. To counteract malicious modification and exploitation of copyrights of the video, obtained perceptual hash is employed. The hash is also used for video indexing and retrieval applications. The proposed method uses Walsh Hadamard Transform for the engendering of perceptual hash. Hash generated is used for authentication and video retrieval and indexing method. The algorithms performance is weighted on the basis of ROC curves and recall versus precision curves.

Keywords: *Perceptual hash, Authenticity, Copyright and Walsh Hadamard Transform.*

----------***----------

1. INTRODUCTION

In recent times technology has improved to such an extent that videos can be accessed even in remote places and at odd times .This has increased the probability of the videos being modified. Content preserving modifications and content changing modifications (table1) are one of the prominent ones which pose a threat to copyrights of the videos and the originality of the videos. To overcome fraudulency of the video and to protect its copyrights, conventional cryptographic technique [1], or watermarking techniques can be used. Nevertheless these techniques also possess disadvantages. Cryptographic technique is susceptible to trivial changes in input and the encryption of large volume of database is prolonged whereas in watermarking technique the embedded watermark causes distortion in the video. When a video is transmitted, it undergoes changes like compression, addition of noise and other content preserving changes, this does not change the perceptual gist of the video therefore the video can yet be identified. In case of cryptographic hashing a minimal change in the input leads to the complete change of the hash value even though the perceptual content of the video remains same. Considering these issues, cryptographic hash functions are not used. Hence it is necessary to obtain hash functions that sustain any insignificant modifications and interpret content changing manipulations. This gives us way towards a technique called perceptual video hashing, where the hash generated is based on the perceptual content of the video. Unlike watermarking [2], [3] perceptual video hashing does not distort the original video or is not sensitive to single bit change like conventional cryptographic techniques. Accordingly, perceptual video hashing can be used for authentication [4], video indexing [5], confidentiality [6], video Integrity and near-identical detection of the video [7]. The perceptual video hashing extracts the perceptually invariant features from the input video data and generates a fixed compact hash vector. Similar hash values are generated for perceptually near identical videos and different hash values for perceptually different videos. The generated hash function must be robust for content preserving attacks and sensitive for content changing attacks.

Let $\mathbf{h}$ represent the perceptual hash of a video with length l_h . Let H denotes the hashing function which maps the video V to a short string based on its perception along with a secret key K . Let $d_H(\cdot,\cdot)$ be the Hamming distance (HD) between the hashes, used to measure the perceptual similarity between the videos. The perceptual hash is denoted as $\mathbf{h} = H(V, K)$. The perceptual hash function must satisfy the following properties.

1. Non-invertibility: The video must not be recovered from the hash vector. Mathematically,

$$V \mapsto H(V,K) \qquad (1)$$

2. Conciseness: The size of the hash is always less than the size of the video. Mathematically,

$$Size\left(H\left(V,K\right)\right) \square\ Size\left(V\right) \quad (2)$$

3. Visual robustness: With very high probability, the normalized hamming distance (NHD) measured between the hashes of the perceptually near-identical videos with same secret key should be close to zero. Mathematically,

$$P_r\left(d_H\left(H\left(V,K\right),H\left(V_{nid},K\right)\right)\approx 0\right)\approx 1 \quad (3)$$

4. Diffusion: With very high probability, the NHD measured between the hashes of perceptually different videos with same secret key should be close to length of the hash. Mathematically,

$$P_r\left(d_H\left(H\left(V,K\right),H\left(V_{diff},K\right)\right)\approx l_h\right)\approx 1 \quad (4)$$

5. Confusion: With very high probability, the NHD measured between the hashes of perceptually near-identical videos with different secret keys should be close to length of the hash. Mathematically,

$$P_r\left(d_H\left(H\left(V,K_1\right),H\left(V,K_2\right)\right)\approx l_h\right)\approx 1 \quad (5)$$

Table -1: List of content-preserving attacks and content-changing attacks

Content non-preserving attacks	Content changing attacks
Transmission errors	Detach frame objects, emplacement change of frame elements, Appending new objects, Modifying frame symptomatic such as colour, texture; Variation in environment such as day, time; Changes of light conditions for a frame and adding it to the video.
Noise	Frame rate change
Compression and Quantization	Frame deletion
Scaling	Adding logo
Rotation	Adding subtitles
Cropping	
Resolution changes	
Colour Inversions	
Contrast modification	
Brightness changes	

A number of techniques exist to obtain the perceptual hash of a video, among them we shall discuss about Tucker decomposition based perceptual video hashing algorithm proposed by Sandeep R et al [14] and Low rank tensor approximation (LRTA) based perceptual video hashing algorithm proposed by Li M, Mong V [13].

The authors Li and Monga developed perceptual hashing algorithm by modeling video as third order tensors [12, 13]. A tensor depicts the inherent mathematical composition of the video and facilitates better utilization of the intrinsic spatio-temporal redundancy. After modeling video as tensor, specific number of overlapping sub cubes was selected such that they cover the entire video. Later each sub cube was subjected to rank-r parallel factor analysis (PARAFAC) [9] which yields three vectors and these three vectors are arithmetically averaged to obtain the hash. The obtained hash exhibits robustness against compression, contrast enhancement, blurring, AWGN, frame rotation, frame cropping, frame rate change and random frame deletion but shows poor performance towards content changing attacks.

Sandeep R et al [14] developed Tucker decomposition [16, 17] based perceptual video hashing algorithm. Here video is modeled as a third order tensor and discrete numbers of 3D sub blocks are selected such that they cover the entire video by employing a secret key. Each block is subjected to tucker decomposition where PARAFAC model is a special case. The model provides flexibility to select different number of factors along each mode during decomposition of the multi-way data array, by providing better analysis. The algorithm exhibits robustness against rotation attack, frame dropping attack, compression, blurring attack, noise addition, brightness modification, contrast modification, cropping, frame rate change and changing the spatial resolution but shows poor performance towards malicious attacks.

Though the described algorithms provide efficient results towards content preserving changes and content changing attacks, Walsh Hadamard Transform [8] is employed here due to its mathematical simplicity. Walsh Hadamard Transform is used to decorrelate the image data and it has good energy compaction where the energy of an image is stored in few coefficients. The transforms computation speed is also very high.

2. PROPOSED ALGORITHM

The proposed algorithm aims at generating robust hash based on perceptual content of the video for authentication and; video indexing and retrieval applications. The block diagram of Walsh Hadamard Transform based video hashing algorithm is shown below

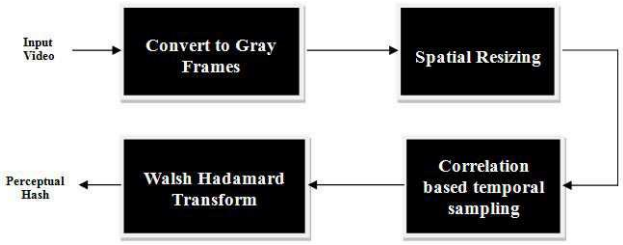

Fig -1: Block diagram of Walsh Hadamard Transform based video hashing algorithm

2.1 Convert to Gray Frames

The colour image is converted to grayscale image which consists of shades of gray. The conversion from colour to grayscale makes the algorithm robust against colour variations and it also reduces the storage space required. The colour to gray conversion of an image is the conversion from RGB to YIQ space [8] which is given by

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (6)$$

Where Y is the luminance channel representing light intensity and contains 93 percent of the grayscale information, I is the in-phase and Q is the quadrature are the chrominance channels representing colour details and contains 7 percent of colour information. In the proposed algorithm the input colour video is converted into sequence of gray frames as shown in figure below,

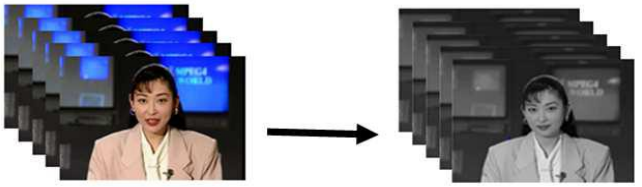

Fig -2: Colour to gray conversion

2.2 Spatial Resizing

Spatial Resizing enables the algorithm to be robust against changes such as frame resizing. Spatial resizing helps in generation of standard hash. In the proposed algorithm the frames of the gray converted video is resized to width=64 and height=64 as shown in the figure below,

Fig -3: Spatial Resizing

2.3 Temporal Sampling based Correlation

Correlation provides the extent of similarity between signals or images. Mathematically, correlation between two matrices is given by

$$\mathrm{corr}(\mathbf{A},\mathbf{B}) = \frac{\sum_{e=0}^{N-1}\sum_{f=0}^{N-1}\mathbf{A}(e,f)*\mathbf{B}(e,f)}{\sqrt{\sum_{e=0}^{N-1}\sum_{f=0}^{N-1}\mathbf{A}(e,f)*\mathbf{A}(e,f)*\sum_{e=0}^{N-1}\sum_{f=0}^{N-1}\mathbf{B}(e,f)*\mathbf{B}(e,f)}} \quad (7)$$

In the proposed algorithm, the first frame of video is compared with the second frame of video, if the two frames are highly correlated, the first frame is neglected and the second frame is compared with the third frame and so on. If the first and second frame is not correlated then the second frame is neglected and second frame is compared with third frame and so on. After applying correlation to the entire video, frames which are uncorrelated with each other under specified threshold are retained. The correlation operation reduces the mathematical complexity of the algorithm by reducing the number of frames for operation. The figure below illustrates the concept

Fig -4: Correlation operation

In the above figure before correlation the video consisted of 300 frames but after correlation only a single frame retained which means the other 299 frames are highly correlated.

2.4 Walsh Hadamard Transform

Walsh-Hadamard transform [8] is adapted due to its mathematical simplicity in various video and image processing applications such as filtering, data compression etc. The basis image of Hadamard transform consists of only ±1. The Hadamard transform matrices $\mathbf{H}_n$ of size $N \times N$ is generated by the core matrix,

$$\mathbf{H}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Therefore,

$$\mathbf{H}_n = \mathbf{H}_{n-1} \otimes \mathbf{H}_1$$

$$\mathbf{H}_n = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{H}_{n-1} & \mathbf{H}_{n-1} \\ \mathbf{H}_{n-1} & -\mathbf{H}_{n-1} \end{bmatrix}$$

Where $n = \log_2 N$

1-Dimensional Walsh Hadamard transform is given by

$$v(k) = \frac{1}{\sqrt{N}}\sum_{m=0}^{N-1} u(m)(-1)^{b(k,m)} \qquad 0 \le k \le N-1 \quad (8)$$

Inverse transform of 1-Dimensional Walsh Hadamard transform is given by

$$u(m) = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} v(k)(-1)^{b(k,m)} \qquad 0 \le m \le N-1 \quad (9)$$

Where $b(k,m)=\sum_{i=0}^{n-1}k_i m_i \qquad k_i m_i = 0,1$

1-Dimensional Walsh Hadamard transform and Inverse transform of Walsh Hadamard transform in matrix form are given by

$$\mathbf{V}=\mathbf{H}\mathbf{U} \tag{10}$$

$$\mathbf{U}=\mathbf{H}\mathbf{V} \tag{11}$$

2-Dimensional Walsh Hadamard transform [15] is given by

$$v(k,l)=\frac{1}{\sqrt{2}}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}u(m,n)(-1)^{\sum_{i=0}^{p-1}\left[b_i(k)b_i(m)+b_i(l)b_i n\right]} \quad 0\le k,l\le N-1 \tag{12}$$

$$u(m,n)=\frac{1}{\sqrt{2}}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}v(k,l)(-1)^{\sum_{i=0}^{p-1}\left[b_i(k)b_i(m)+b_i(l)b_i n\right]} \quad 0\le m,n\le N-1 \tag{13}$$

Matrix form of 2-Dimensional Walsh Hadamard transform and Inverse 2-Dimensional Walsh Hadamard transform

$$\mathbf{V}=\mathbf{H}\mathbf{U}\mathbf{H}^{\mathrm{T}} \tag{14}$$

$$\mathbf{U}=\mathbf{H}\mathbf{V}\mathbf{H}^{\mathrm{T}} \tag{15}$$

2.4.1 Properties of Walsh Hadamard Transform [8]

1. Orthogonality and Symmetric property:

$$\mathbf{H}=\mathbf{H}^{*}=\mathbf{H}^{\mathrm{T}}=\mathbf{H}^{-1} \tag{15}$$

The above equation indicates the transform is real, symmetric and orthogonal.

2. High computation speed:
The Hadamard Transform contains only ±1. Hence multiplications are not required during computation. And also the number of addition and subtraction operations can be reduced from N to $N\log_2 N$ as $\mathbf{H}_n$ can be expressed as a product of *n* sparse matrices as shown in below equation,

$$\mathbf{H}=\mathbf{H}_n=\tilde{\mathbf{H}}^{n} \tag{16}$$

3. The Hadamard transform coefficient's natural occurring order is equal to the bit reversed gray code representation of its sequence.
4. Hadamard transform exhibits good energy compaction for highly correlated images.

In the proposed algorithm, Walsh Hadamard Transform is given by Eq. (14), is applied to the obtained decorrrelated frames of the video with Walsh Hadamard matrix 64×64. The coefficients obtained after applying Walsh Hadamard Transform to decorrelated frames represent the perceptual hash of the input video.

3. EXPERIMENTAL RESULTS

The Receiver Operating Characteristics (ROC) [10] curve for the proposed algorithm is obtained under various individual and multiple image processing attacks and is compared with ROC curves of Low Rank Tensor Factorisation Algorithm (LRTA) and Tucker decomposition algorithm to compute the performance of the intended algorithm. ROC is a plot of miss probabilities versus false alarm probabilities. Miss probability is the measure of similar video being not predicted as same. False alarm probability is the measure of different video being classified as same video. A video database of 224 videos [18, 19] is taken for testing of the proposed algorithm and it is made to undergo certain content preserving manipulations and content changing manipulations such as frame rotation, blurring, adding noise, brightness changes, contrast changes, cropping of frame, compression, spatial resolution change, changes in frame rates and malicious changes by inserting a logo to measure the capability of algorithm under these attacks. The ROC curves are generated using MATLAB software for the attacks mentioned and compared with LRTA and Tucker Decomposition algorithms ROC curves.

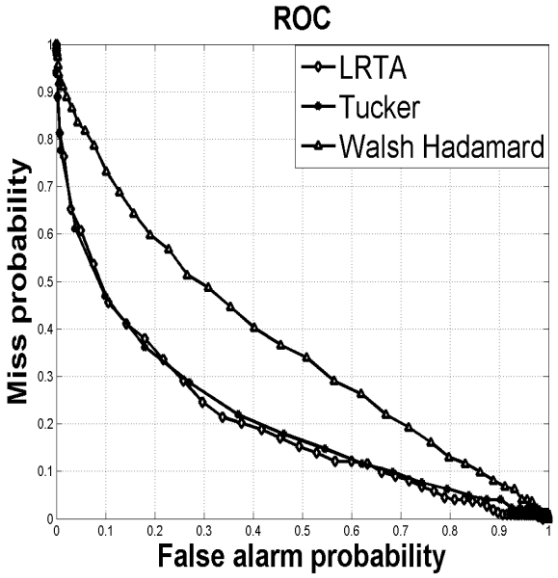

Fig -a: Brightness changes

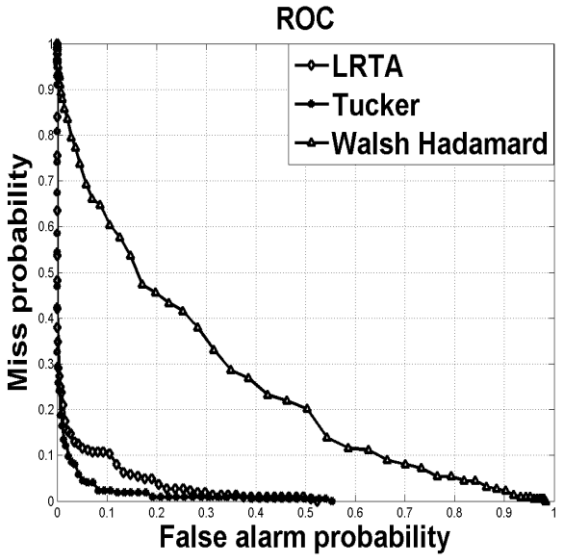

Fig -b: Blurring attack

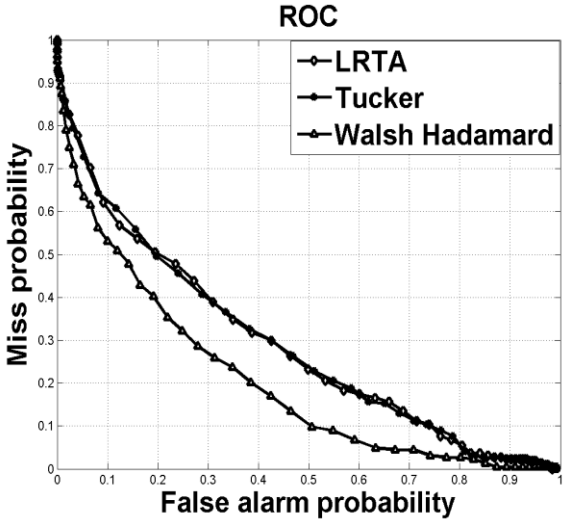

Fig -c: Contrast attack

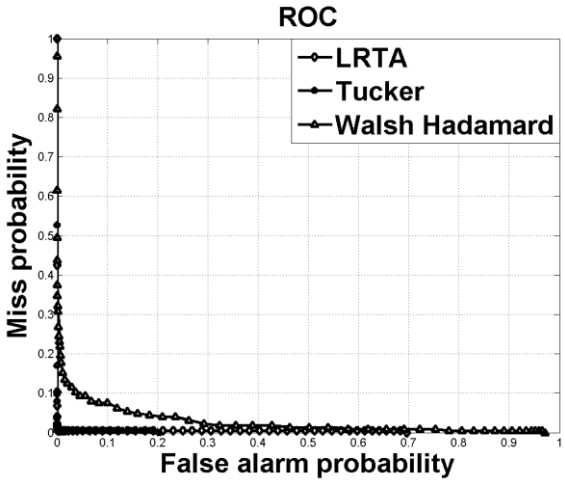

Fig -d: Malicious attack

Fig -5: ROC curves for single attacks

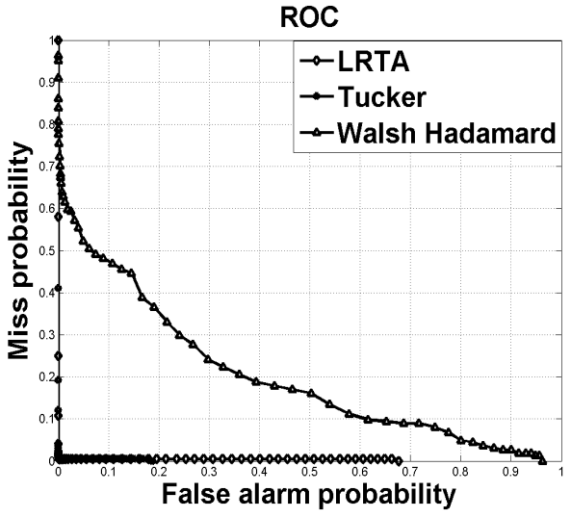

Fig -a: Bitrate and angle attack

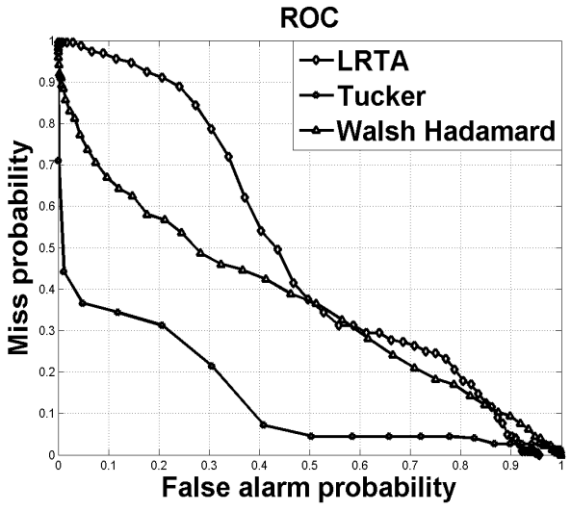

Fig -b: Noise and frame rate attack

Fig -6: ROC curves for multiple attacks

4. APPLICATION

The obtained hashes of two videos are compared to check whether the videos are same or not as shown in the below figure 7. Perceptual hash value of video one i.e. *V*1 and Perceptual hash value of second video i.e. *V*2 are compared by computing hamming distance between the hash values. If hamming distance measured amidst the two videos is less than 0.5 threshold value, then the videos are considered to be similar. If the hamming distances measured between two videos are more than 0.5 threshold then the videos are considered to be different.

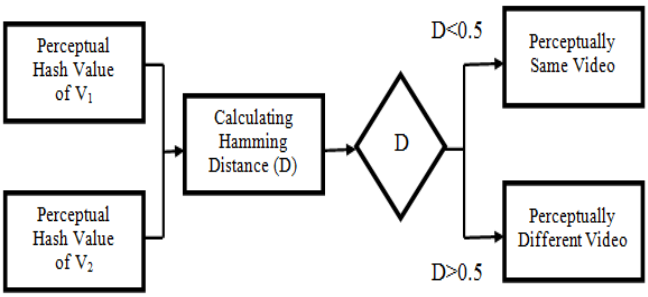

Fig -7: Block Diagram for Comparison of the Hash Values

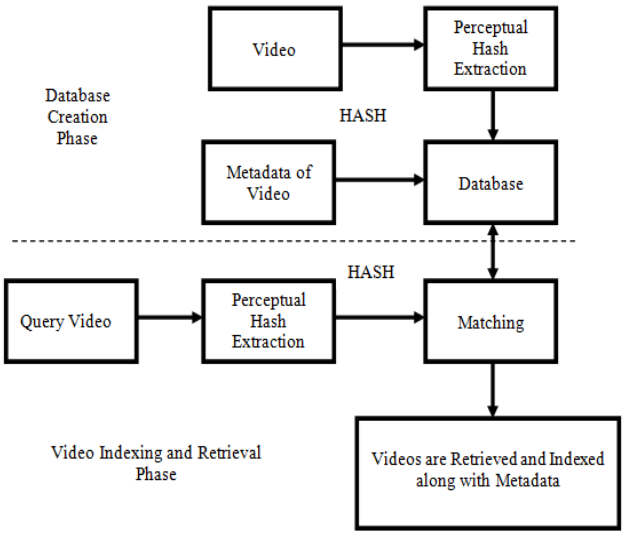

Fig -8: Block diagram of retrieval and indexing system for Video

The above figure shows the block diagram developed on the aspects of perceptual hashing for video indexing and retrieval application. A database of 1,792 videos is generated from 224 videos which consist of 224 original videos along with 8 variations of each original video in it. The 8 variations to the original video are,

1. Framerate change (15fps, 60fps)
2. Resolution change (big, small)
3. Bitrate change (64k, 100k)
4. Watermarking logo (16×16, 32×32)

For each video the hash is developed using the proposed method and the corresponding meta-data is stored in the database. The hash of the query video is generated and the computed hash is compared with the hashes in database. Based on hash similarity, near identical videos are retrieved and indexed. This is a query-by-example system which recovers all the near identical videos of the given query video. The relevance of recovered video indicates efficiency of the retrieval systems. The retrieval and indexing system's performance is validated on the aspect of average recall and precision rate curves [11]. Recall is the ratio of the number of pertinent videos recovered to the total number of pertinent videos. The recall value represents the capability to find each and every pertinent video in the data. Precision is the ratio of the number of pertinent videos recovered to the total number of videos recovered. The precision value represents the capability to recover highest rank videos that are predominantly pertinent. A graph of recall versus precision for the proposed retrieval and indexing system for a query video search i.e. *akiyocif.avi* video in the master database consisting of 1,792 videos including the given query video is shown in the below figure

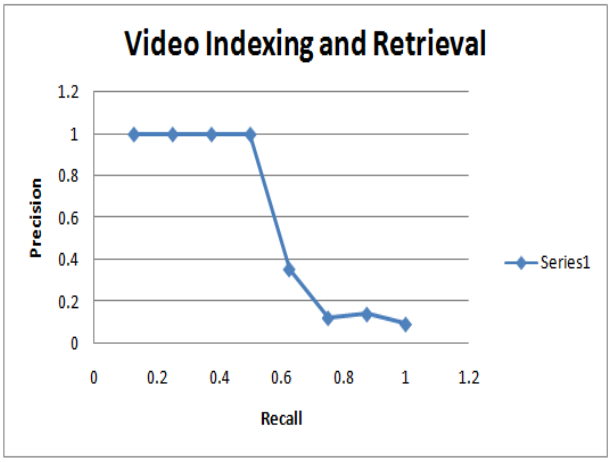

Fig -9: Recall versus precision graph for a query Video

The curve more towards top right corner indicates the system is more efficient and from the figure it can be seen that for the given query video, the video retrieval and indexing system is more efficient.

5. CONCLUSION

In the paper Walsh Hadamard Transform is employed to generate the perceptual hash of the videos. Proposed methods performance is compared with LRTA and Tucker decomposition algorithm based on the ROC curves. The proposed algorithm shows robustness under several single and multiple attacks such as brightness change, contrast modification, bitrate change, angle attack etc. The hashes generated using the proposed algorithm for two videos are compared based on the hamming distance, if the hamming distance measured is less than certain threshold; the videos are perceptually similar or different. Thus the proposed algorithm serves for video authentication. Video indexing and retrieval was developed using the proposed algorithm. The developed video indexing and retrieval system performance is estimated using the average recall and precision curves. By observing the precision and recall curve, the retrieval system is efficient for video retrieval and indexing based on the contents of the video.

REFERENCES

[1]. J. Fridrich and M. Goljan., "Robust hash functions for digital watermarking," in Proceedings of the IEEE Int. Conference on Information Technology: Coding and Computing, LasVegas, NV, USA, Mar 2000.

[2]. T. Kalker, J. A. Haitsma,and J. Oostveen, "Issues with digital watermarking and perceptual hashing," in Proc. SPIE 4518, Multimedia Systems and Applications IV, Nov. 2001.

[3]. R. Venkatesan, S. M. Koon, M. H. Jakubowski, and P. Moulin, "Robust image hashing," in Proc. ICIP, Sep. 2000.
[4]. M. Malekesmaeili, M. Fatourechi, and R. K. Ward, "Video copy detection using temporally informative representative images," in Proc. Int. Conf. Machine Learning and Applications, Dec. 2009, pp. 69-74.
[5]. Mu Li, Vishal Monga, "Twofold Video Hashing With Automatic Synchronization," IEEE Transactions on Information Forensics and Security 10(8): 1727-1738 (2015).
[6]. Xiushan Nie, Ju Liu, Jiande Sun, and Wei Liu, "Robust Video Hashing Based on Double-Layer Embedding," IEEE Signal Processing Letters, 2011, 18(5): 307-310.
[7]. Zhangyang Wang, Houqiang Li, Qing Ling, Weiping Li, "Robust temporal-spatial decomposition and its applications in video processing, " IEEE Transactions on Circuits and Systems for Video Technology, 23(3): 387-400, 2013.
[8]. Anil K. Jain, "Fundamentals of Digital Image Processing," Prentice Hall, 1989
[9]. Kolda TG, Bader BW (2009) Tensor decompositions and applications. SIAM Review 51:455-500
[10]. Fawcett T (2006) "An introduction to ROC analysis," Pattern Recognition Letters 27(8):861-874
[11]. Yu-Gang Jiang JWYudong Jiang (2014) Vcdb: A large-scale database for partial copy detection in videos. In: European Conference on Computer Vision (ECCV)
[12]. Li M, Monga V (2011) Desynchronization resilient video fingerprinting via randomized, low-rank tensor approximations. In: Multimedia Signal Processing (MMSP), 2011 IEEE 13th International Workshop on, pp1-6
[13]. Li M, Monga V (2012) Robust video hashing via multilinear subspace projections. Image Processing, IEEE Transactions on 21(10):4397 –4409
[14]. Sandeep R., Saksham Sharma, Mayank Thakur and P. K. Bora, "Perceptual video hashing based on Tucker decomposition with application to indexing and retrieval of near-identical videos," Journal: Multimedia Tools and Applications, 2015 DOI: 10.1007/s11042-015-2695-1.
[15]. G.G.Lakshmi Priya and S.Domnic, "Walsh Hadamard Transform Kernel based Feature Vector for Shot Boundary Detection," *IEEE Trans. on Image Processing*, 23(12), 5187-5196, 2014
[16]. Tucker L (1966) Some mathematical notes on three mode factor analysis. Psychometrika 31(3):279–311
[17]. Tucker LR (1963) Implications of factor analysis of three-way matrices for measurement of change. In: Harris CW (ed) Problems in measuring change., University of Wisconsin Press, MadisonWI, pp 122–137
[18]. (2012) http://media.xiph.org/video/derf/,test video sequences
[19]. (2012) http://www.reefvid.org/,test video sequences

BIOGRAPHIES

Asha Kiran M U received B.E. degree from Dr. T. Thimmaiah Institute of Technology, Kolar, India, in the year 2014. She as obtained M. Tech in Signal Processing (2016) from Cambridge Institute of Technology, Bengaluru, India. Both the institutes are affiliated to Visvesvaraya Technological University, Belagavi, India. She has academic and her areas of interests include perceptual image hashing, perceptual video hashing and video authentication.

Sandeep R. received the B.A. degree in Hindi from Mysore Hindi Prachar Parishad, the B.E. degree in Electronics and Communication Engineering and the M.Tech. degree in Electronics Engineering from Visvesvaraya Technological University in the year 2001, 2006 and 2009 respectively. Currently, he is pursuing Ph.D. in the department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, India. He has both academic and industry experience. His current research interests include perceptual image hashing, perceptual video hashing, biometric hashing and biomedical image processing.