INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student	:	SANDEEP R
Roll Number	:	10610228
Programme of Study	:	Ph.D.

Thesis Title: Projection-based Perceptual Video Hashing

Name of Thesis Supervisor(s)	:	Prof. P. K. Bora
Thesis Submitted to the Department/ Center	:	EEE
Date of completion of Thesis Viva-Voce Exam	:	03-12-2019
Key words for description of Thesis Work	:	Perceptual video hashing; Content authentication; Near-duplicate video retrieval

SHORT ABSTRACT

A perceptual hash function for a video generates a fixed-length binary string called the perceptual hash on the basis of the perceptual contents of the video. This hash must be robust to the manipulations that preserve the perceptual contents of the video and fragile to the modifications that vary the perceptual contents of the video. Developing a perceptual video hashing method satisfying the conflicting properties is a challenge. This thesis generates the perceptual hash from the video based on the projection onto a sub-space by utilizing both the spatial and temporal properties of the video.

This thesis first generates a perceptual hash from the video using the 3D-radial projection of the pixels and assesses the differentiating capabilities and the perceptual robustness of the hash generated. This method is the 3D extension of the well established 2D radial projection based image hashing. The pixel luminance values of randomly located sub-cubes are radially projected to calculate the variance of the pixels along the radial lines. The hash is obtained in two different ways. In Scheme-I, the variances along the radial lines are averaged, and then projected onto the discrete cosine transform (DCT) basis to form a compact hash vector. This hash vector is binarized using the median-based quantization. In Scheme-II, the variances along the lines of each sub-cube are projected on the 2D DCT basis, and then averaged to form a compact hash vector. This hash vector is also binarized using the median-based quantization. The performance of this algorithm is assessed using the receiver operating characteristic (ROC) curves. Simulation results indicate that the method performs well against most of the typical content-preserving distortions but poorly against the addition of noise. It also performs well against the malicious attacks.

The thesis proposes to use random projection for the generation of perceptual hash from the video. This method is the 3D extension of the well established 2D random projection-based image hashing. The video is projected onto Achlioptas's random basis. This basis functions are written in the form of a matrix called as Achlioptas's random matrix (ARM), with each element taking +1 and -1 with an equal probability. Hence, the projection involves only addition and subtraction operation. Simulation results show that this perceptual hash is robust to common signal and image processing attacks, but has the scope for improvement. To improve the

performance of the perceptual hashing function, the temporal discrete wavelet transform (DWT), referred to as the temporal wavelet transform (TWT), is used for generating the temporally informative representative frames (TIRFs) by extracting the features in the temporal direction. The low-frequency components of the resultant transform domain data are then projected onto the Achlioptas's random basis to generate the hash. The TWT-ARM based perceptual video hashing function not only reduces the dimensions but also retains the essential features. Simulation results show that the TWT-ARM based video hashing algorithm performs better against the content-preserving attacks when compared to that of the existing video hashing algorithms. The drawback of this algorithm is that the computational complexity increases as the number of video frame increases.

Finally, the thesis exploits the multi-linear subspace projection for the extraction of the perceptual hash from the video. In the literature, the reduced rank PARAllel FACtor (PARAFAC) decomposition has been successfully applied to extract the perceptual hash for the videos. We propose a robust perceptual video hash algorithm based on a superior tensor decomposition known as the Tucker decomposition. We also propose a method to find the number of components in the factor matrices of the Tucker decomposition. The Tucker decomposition based video hashing algorithm shows superior performance against the most of the image processing attacks with the added advantage of computational efficiency. An application for indexing and retrieval of near-duplicate videos (NDVs) is developed using the proposed video hashing methods. The performance is evaluated using average precision recall curves. The experimental results on a moderate-size video database shows that the Tucker decomposition and the TWT-ARM based video hashing algorithms perform better in retrieving the NDVs.