Video Perceptual Hashing Using Interframe Similarity

Xuebing Zhou
xuebing.zhou@igd.fraunhofer.de
Martin Schmucker
martin.schmucker@igd.fraunhofer.de
Christopher L. Brown
brown@spg.tu-darmstadt.de

Automatic content identification and verification are required in numerous application scenarios. This can be achieved by perceptual hash functions. This article describes a general structure of video perceptual hashing. An algorithm based on interframe similarity is investigated. We analyze its performance and compare it with an algorithm that uses mean luminance. The proposed algorithm shows increased identification reliability with respect to robustness and receiver operating characteristics.

1 Introduction

The perceptual hashing techniques, also called fingerprinting technologies, extract characteristic features from multimedia content to calculate compact digests, which are related to human perception and qualified for content identification. In comparison with watermark techniques and cryptographic hashing techniques, they do not modify the original data and have robustness against content manipulation. So perceptual hashing can be considered as robust hashing.

This paper is concerned with video perceptual hashing. In section 2, a general structure for video perceptual hashing is reported. Section 3 introduces a video perceptual hashing based on the interframe similarity of video signal. In section 4 we show that this algorithm extracts features with increased reliability in comparison with Philips' algorithm. A conclusion is given in section 5.

2 General structure of perceptual video hashing

![Diagram of video fingerprinting algorithm]

Figure 1: Structure of video fingerprinting algorithm

Figure 1 shows a general structure for video fingerprinting algorithms. The incoming video
content is firstly decoded into individual frames. In the block processing, individual frames are divided into small blocks. The statistical characteristics such as mean, variance, colour descriptor etc. are extracted. The resulting values should be processed and compressed in order to get robust and compact video fingerprints. The features of video signal can be extracted purely from spatial information like the algorithms in [DRVLM04], [CB04] and [MLM04], or from spatio-temporal information like [YTC04] and [OKH01]. The algorithms that extract features from individual frames offer good robustness to manipulations based on common image processing. However, they need a process to vanquish the highly temporal correlation of the video signal. The video perceptual hashes extracted from the spatio-temporal changing have less redundancy and good robustness, in particular for rapidly varying video. In the next section a perceptual hashing for videos that is based on the interframe similarity is described.

3 Video perceptual hashes using differential block similarity

Philips’ group uses spatio-temporal differentiation of mean luminance to get high frequency components of the video clip [OKH01]. This method performs well both in discernibility and in robustness to most video manipulations. However, its effectiveness is significantly impacted by noise including compression noise or slowly varying videos. Accordingly, we keep the structure of Philip’s algorithm and extract the interframe similarity to enhance the robustness of the perceptual hashes.

In the proposed algorithm, each frame is divided into $M + 1$ blocks. The luminance of pixel $j$ in the block $i$ of frame $n$ is denoted as $x(j, i, n)$ with $i \in [1, \ldots, M + 1]$. The similarity of temporal consecutive blocks is $S_{i, n}$ with:

$$S_{i, n} = \sum_{j} x(j, i, n) \cdot x(j, i, n - 1)$$  \hspace{1cm} (1)

![Figure 2: Block diagram of the algorithm](image)
The resulting hashes $H_{i,n}$ are the sign of the spatio-temporal difference of $S_{i,n}$. A block diagram, describing the algorithm, is shown in figure 2.

$$H(i,n) = \begin{cases} 1 & \text{if } (S_{i,n} - S_{i+1,n}) - a \cdot (S_{i,n-1} - S_{i+1,n-1}) \geq 0 \\ 0 & \text{if } (S_{i,n} - S_{i+1,n}) - a \cdot (S_{i,n-1} - S_{i+1,n-1}) < 0 \end{cases}$$ (2)

4 Analysis and evaluation

We implemented Philips perceptual hashing algorithm [OKH01] in Matlab as a reference implementation and compare it with the proposed algorithm in section 3. In figure 3 and 4 the boxplots of BER for different videos for noise contamination and MPEG2 compression are represented (BER is denoted as rate of mismatched bits between hashes of the original videos and those of their manipulated videos). Comparing the two algorithms, the median as well as the range of BER are strongly suppressed in the case of noise. For MPEG2 compression the BER decreases notably. The algorithm using similarity has better performance in robustness to video processing (the detailed results are shown in [ZSB05]).

The empirical tests show that the algorithm using similarity improves the identification performance. Figure 5 and 6 show ROC curves of Philips algorithm and our algorithm in the case of Gaussian noise. Generally, the dashed lines are under the solid lines. Therefore the algorithm using similarity has better identification performance. For $a = 1$ the enhancement is very significant (figure 5).

![Figure 3: Boxplot for different types of noise ($a = 0.95$)](image)

![Figure 4: Boxplot for MPEG2 compression in different bit rate ($a = 0.95$)](image)

5 Conclusion

The proposed perceptual hashing algorithm based on interframe similarity shows increasing robustness, especially to noise. Moreover it provides better identification performance than the Philips algorithm. Its complexity is a little higher in comparison with the original
method, however, it is more reliable.

6 Acknowledgment

The work presented in this paper has been supported in part by the European Commission through the IST program under Contract IST-2002-507932 (ECRYPT, www.ecrypt.eu.org).

References


