Journal of Visual Communication and Image Representation 50 (2018) 263–269

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

journal homepage: www.elsevier.com/locate/jvci

Short Communication

Encoder settings impact on intra-prediction-based descriptors for video retrieval☆

Amir H. Rouhi*, James A. Thom

School of Science, RMIT University, Melbourne, Australia

ARTICLE INFO

Keywords:
H.264/AVC
Encoder impact
Profile impact
Feature sensitivity
Video copy detection
Intra-prediction modes
I-frames
Correlogram
Visual features
Content-preserving transformations

ABSTRACT

Intra-prediction modes (IPMs) of H.264/AVC, as visual feature components, are easily extracted from the compressed domain. However, in spite of their efficiency and effectiveness, they can be sensitive to encoder settings such as encoder-type and encoding-profile. In content-based video retrieval using different encoders between a video and its copy can affect performance. Similarly, applying different encoding-profiles can also have a negative impact. Experiments with different IPM-based descriptors, on a subset of the TRECVID/CCD 2011 collection with content-preserving visual distortions, compare the performance of four H.264/AVC encoding profiles (*High10*, *High*, *Baseline* and *Main*) and also investigate the impact of three encoders (*FFmpeg*, *MediaCoder* and *JM*). The research shows that ordinal pattern distribution of unified IPMs can improve their tolerance to changes in encoder settings. Furthermore, it is shown that a Correlogram rather than a Histogram of the IPMs is more robust with respect to the impact of encoding-profiles and encoder-types.

1. Introduction

Efficient multimedia searching to retrieve near-duplicate video segments is required in big multimedia archives and data banks. Content-based approaches for multimedia retrieval, extract audio-visual features directly from multimedia files, as compared to tag-based approaches that use metadata such as annotations. Content-based approaches are the method of choice for copy detection content-based video search, advertisement tracking, automated tagging, and copyright infringement detection; applications where annotations are either nonexistent or inapplicable. The prohibitive size of the available multimedia content in large collections also makes a manual annotation approach infeasible.

A major challenge of content-based retrieval is its computational cost. Utilising light-weight yet effective features is a solution. Compressed domain processing is a promising approach that saves a considerable amount of computational resources by avoiding fully decompressing the signal with the advantages of faster processing of less data, lower storage requirements and reduced bandwidth utilization. However compressed domain methods have been criticized for their inability to benefit from the spatial information available in the pixel domain and their sensitivity to the compression technique [1,6,19] and the encoder settings [7].

This sensitivity can affect the components used in the structure of features which are extracted from different encoders. We describe this phenomenon as **encoder-type impact**. Moreover, within each bitstream syntax, a limited number of subsets, known as *Profiles* [13], enable the encoder to adapt the encoded video in different signal transmitting channels to maintain the highest quality available. Profiles can affect the visual feature components in the compressed domain, for a given encoder. We describe this issue as **encoding-profile impact**.

Intra-prediction modes (IPMs) in H.264 video streams utilise a block-based approach for compression. Despite the sensitivity of block sizes and modes to the picture texture, it is shown in this paper that using an ordinal distribution of unified block sizes can produce robust features. We previously proposed using IPM Histograms (IPMHs) for near-duplicate video copy detection in the compressed domain [16,18]. IPMH features utilise the information stored in the IPMs of key frames in H.264/AVC videos. These descriptors take a *unification process* to unify similar modes of different block sizes which is needed when comparing different encoding-profiles. They also exploit an ordinal approach instead of a probability distribution of IPM blocks. Ordinal Pattern Distribution (OPD) is the key technique that enables IPMH descriptors to be computationally efficient, more effective in content-preserving visual distortions, and robust against encoder settings. This is similar to nonparametric methods in statistics that use order statistics and ranking of observations, which makes them independent of the assumptions about the underlying distribution of data.

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail addresses: amir.rouhi@rmit.edu.au (A.H. Rouhi), james.thom@rmit.edu.au (J.A. Thom).

https://doi.org/10.1016/j.jvcir.2017.12.009
Received 10 June 2017; Received in revised form 10 October 2017; Accepted 16 December 2017
Available online 19 December 2017

IPMC (IPM Correlogram), proposed in this paper, operating in the lower-dimensional sub-space of the compressed-domain is more efficient compared to the conventional methods which operate in the pixel domain. Compared with IPMH, IPMC is also more robust to the content-preserving video distortions and less sensitive to the impact of encoder-type and encoding-profile at the expense of marginally higher computational cost. Moreover, with respect to the unification process and OPD, IPMC incorporates spatial information stored in correlation among pairs of blocks, instead of single block distributions used by IPMH. The locality information provided by IPMC can compensate for the lack of local information in the compressed domain. Furthermore a sensitivity study is missing from the existing literature on IPM-based descriptors; to address this, a comprehensive sensitivity analysis is conducted on both the IPMH and IPMC descriptors using data from the TRECVID/CCD task. The sensitivity of IPM-based descriptors is studied for three encoder-types and four encoding-profiles on four content-preserving video distortions. A non-parametric Kolmogorov–Smirnov (K-S) statistical test [20] has been employed for testing the null hypothesis.

Section 2 provides some background information. An overview of the sensitivity of IPMs and the role of the OPD and unification of block sizes is provided in Section 3. Section 4 introduces IPMC and the other two IPM-based descriptors. Section 5 is dedicated to the video collection, performance measures and the methodology for evaluating the impact of encoder settings. The results of the performance evaluation are analysed in Section 6. Finally, the conclusions are summarised in Section 7.

2. Background and related work

2.1. Visual transformations

Detecting an undistorted copy of a video is a straightforward task and most visual features handle this easily. TRECVID/CCD in 2011 addressed a much more complicated scenario by providing a dataset containing videos that are affected by 8 types of visual and 7 types of audio distortions, known as *transformations* [2]. Fig. 1 shows the 8 visual transformations categorised as *content-altering (geometric)* (top row) versus *content-preserving (non-geometric)* (bottom row). The functional scope of global visual descriptors are content-preserving transformations. So, to evaluate the performance of global descriptors, Gupta et al. [8,9] recommend testing the effectiveness of the descriptors on the content-preserving visual transformations.

2.2. Research on IPM-based descriptors

IPMs are the essential component in I-frames in H.264/AVC and H.265/HEVC encoders. With respect to the increased computational complexity of HEVC compared to H.264/AVC, there are advanced parallel framework techniques to improve the encoder's efficiency [22,23]. Similarly there are numerous studies investigate the efficiency of mode-decision function techniques for H.26x encoders. Although the mode-decision function can directly affect the IPM components, there is no research on how the changes of IPM components can affect the performance of IPM-based descriptors.

The use of IPM components as a visual descriptor, was initially proposed by Zargari et al. [25]. They used the probability distribution of IPM as a texture descriptor on I-frame coded pictures. This idea was adopted for video copy detection [16,18]; where IPMH descriptors, that utilise the probability distribution of IPM modes, have been investigated.

Intra-prediction modes are also used in other applications. For example video watermarking, where Horng et al. [11] utilise both intra- and inter-prediction modes.

3. Unification and ordinal pattern distribution

Using the same standard of video compression does not guarantee a similar structure in intra-prediction modes. Some procedures used in compression algorithms (such as mode-decision, rate control, rate allocation, and motion estimation) are considered as non-normative components and are not standardised [13,15]. The standard provides a syntax to be followed and tools available to achieve a compression target. Deciding how to use these tools is optional, as long as the resulting bit-streams operate according to the syntax defined in the standard. So any distortion in the image can affect the distribution of intra-prediction modes making IPM-based descriptors for video retrieval algorithms fragile, even for simple tasks. Moreover, different H.264/AVC encoders and profiles can have a major impact on the IPM components. Fig. 2 shows this phenomenon given different encoder-types (FFmpeg, MediaCoder and JM) and encoding-profiles (High, High10, Baseline and Main). IPM components of a video frame, encoded by different encoders and profiles, can result in significantly different distributions of block sizes and modes. Consequently, a simple probability distribution (histogram) of IPM blocks is not a robust visual descriptor, because it is not capable of detecting similarity in the presence of visual distortions and encoder settings.

Two techniques are employed in the structure of proposed IPM-based descriptors: a *unification process* and *Ordinal Pattern Distribution* (OPD) of IPM components. The *unification process*, deals with the size of the blocks, and enables us to compare different encoder-types and encoding-profiles with different block sizes together. High profile utilises all the 3 types of the block sizes while the other profiles utilise 4×4 and 16×16 blocks only. The unification process, depicted in Fig. 3, transforms all the blocks to the smallest block size, 4×4, with consideration to the appropriate transforming weight (each 8×8 block has a weight of 4 and each 16×16 block has weight of 16).

The second technique, *Ordinal Pattern Distribution* (OPD), is applied

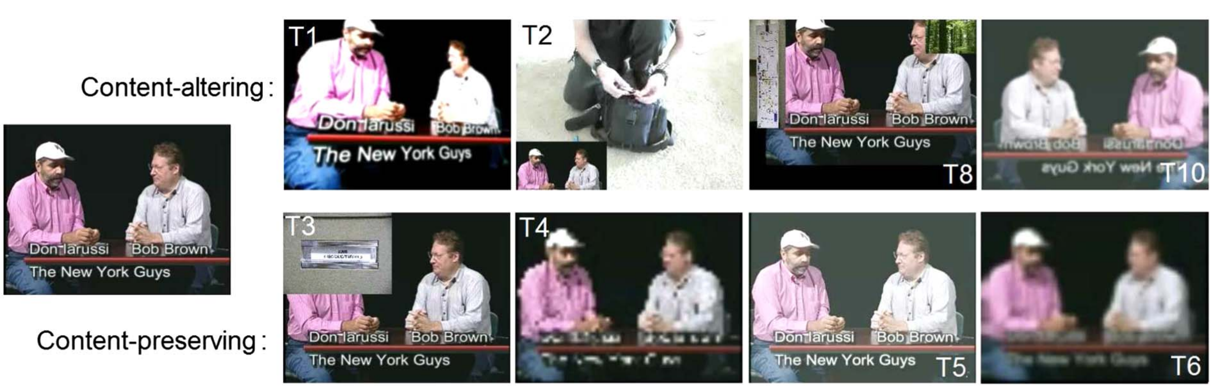

Fig. 1. TRECVID/CCD 2011 visual transformations. Content-preserving transformations are shown in bottom row (T3 insertions of pattern; T4 resolution change; T5 change of gamma; T6 decrease in quality). Transformations T6, T8 and T10 are combinations of different transformations.