Signal Processing: *Image Communication* 24 (2009) 484–498

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Compressed domain indexing of scalable H.264/SVC streams

Christian Käs *, Henri Nicolas

Laboratoire Bordelais de Recherche en Informatique (LaBRI), University of Bordeaux 1, 351, Cours de la Libération, 33405 Talence Cedex, France

A R T I C L E I N F O

Article history:
Received 16 February 2009
Accepted 18 February 2009

Keywords:
H.264/SVC
Trajectory estimation
Object distance estimation
Copy detection

A B S T R A C T

We present methods to efficiently analyze scalable, compressed H.264/scalable video coding (SVC) video streams. Relying solely on information present in the compressed stream, we estimate the global camera motion, perform motion segmentation and use a simple matching process to track moving objects over time. Object energy images are constructed in order to help resolve the problem of object correspondence during the occlusions of multiple objects. To save computing time, we analyze lower spatial layers of the stream and add higher layer information only if necessary. We draw 2-D object trajectories in the view plane of the camera and use the temporal evolution of the objects' properties to estimate the relative distance to the camera, resulting in a pseudo 3-D representation of the trajectories. Finally, the suitability of the motion parameters to perform video retrieval/copy detection tasks is demonstrated. We therefore form two simple descriptors that are invariant to a series of transformations.

1. Introduction

Scalable video streams gain increasingly more importance in the context of modern applications and distribution networks. Heterogeneous networks with end-user devices ranging from smart phones to high-definition (HD) TV sets raise the need for scalable solutions. H.264/scalable video coding (SVC) addresses these issues and provides support for spatial, temporal and quality scalability in a coding-efficient manner. SVC is the scalable extension of H.264/AVC, which is also known as MPEG-4/Part 10 and which gained popularity because of its solid performance. H.264/AVC is already in widespread use on mobile devices and is one of the three mandatory codecs to use on HD Blue-ray discs. Although other, wavelet-based scalable video codecs are available, we focus on SVC coded material in this work.

Compressed domain indexing is an interesting and challenging research topic, where the main benefit is to save computing time by reusing information already present in the stream. Possible applications of such algorithms include video surveillance networks, retrieval tasks and fast scene analysis. Though image and video processing tasks working in the pixel domain are usually more precise and reliable, they are often computationally too expensive for real-world applications. Decreased robustness of compressed domain approaches usually results from the unavailability of color information, block-based coding and from the sparse and noisy nature of the motion vectors (MVs), which are optimized in terms of coding efficiency and not with respect to the real optical flow. These MVs, which are associated with inter-predicted macro-blocks (MBs), are a valuable source of information in MPEG-based video streams. Motion estimation is a computationally expensive task that has already been performed by the encoder. This enables fast processing speed in the compressed domain. In MPEG-1/2 indexing algorithms, DC and AC coefficients of the underlying DCT are also widely used, because they are easily accessible and their extraction does not require stream decoding. H.264 introduces more advanced intra- and inter-prediction structures, so the re-construction of the transform coding coefficients is computationally more expensive because more decoding steps are required. Due to its superior design and performance, H.264 is

* Corresponding author. Tel.: +33 5 40 00 84 27.
E-mail addresses: kaes@labri.fr (C. Käs), nicolas@labri.fr (H. Nicolas).
URL: http://www.labri.fr (H. Nicolas).

0923-5965/$ - see front matter
doi:10.1016/j.image.2009.02.007

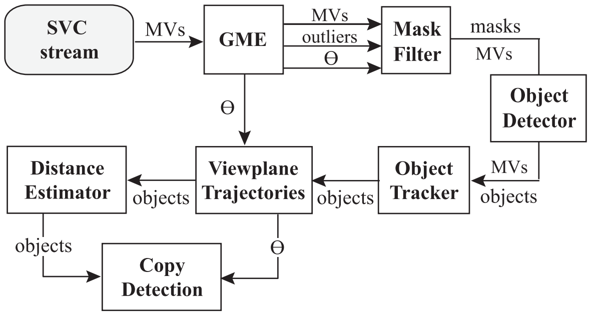

Fig. 1. System overview.

successively adopted in favor of MPEG-2 in a wide range of applications, ranging from the distribution of HD content and media storage to streaming on mobile devices.

This article presents methods and results for global motion estimation (GME), trajectory estimation of moving objects and the estimation of their relative distance to the camera, resulting in a pseudo 3-D representation. Furthermore, we demonstrate the suitability of the motion parameters to perform video copy detection in databases. Since robust algorithms for certain compressed domain indexing tasks already exist, we ported some of the techniques to the scalable SVC domain. We also present novel extensions and approaches to further exploit and process the information that can be easily extracted from H.264/SVC compressed streams.

The suite of this article is organized as follows: after a brief overview of existing methods in Section 2 we provide a short description of the used test sequences and some important encoding parameters in Section 3. In Section 4 we show how we ported an existing GME algorithm to the scalable SVC domain. The segmentation and extraction of moving objects is explained in Section 5, followed by the object matching/tracking algorithm in Section 6. The construction of motion-compensated object trajectories in the view-plane of the camera and the estimation of the relative object distance to the camera are provided in Section 7. Finally, we show in Section 8 how the estimated camera motion parameters can be used to form a robust descriptor to perform video copy detection in databases. We provide separate results at the end of each section. The schema of our approach and its building blocks is depicted in Fig. 1.

We assume that we have separate video scenes without any cuts or transitions. This can be achieved by first applying a compressed domain shot boundary detector, one of which was proposed by Bruyne [1] specifically for H.264 streams.

2. Related work

A large number of compressed domain indexing approaches appeared over the years, including object segmentation, tracking, mosaic construction, video summarization and video retrieval.

The basis for numerous indexing and retrieval tasks builds the GME. The observed global background motion in a scene is usually due to camera operation. Kobla [2] employed directional MV histograms to determine the dominant, translational camera movement and performed focus-based zoom detection. Wang et al. [3], Bouthemy et al. [4] and Durik et al. [5] used iterative motion estimation and outlier rejection algorithms with some form of weighted least squares estimation applied to the MV field extracted from the MPEG-2 stream. Hesseler et al. [6] applied two-dimensional MV histograms and used the AC coefficients for outlier detection. The existing algorithms do not incorporate scalable video streams with differently sized MBs and bi-predicted B-frame MVs.

Some publications concerning object segmentation in the MPEG domain include [7–12]. Babu et al. [7,8] proposed an accumulation of MVs over time, followed by a K-means clustering to determine the number of objects in the scene and the EM algorithm for object segmentation. Zeng et al. [9] employed a block-based Markov random field (MRF) model to segment moving objects from the sparse MV field, which is extracted from H.264 compressed streams. Their method is limited to static cameras. Sukmarg et al. [10] presented a MPEG-2-based approach that works solely with DC images and AC coefficients. They apply a region-merging algorithm on an initial set of regions obtained by adaptive K-means clustering. Mezaris et al. [11] used an iterative rejection scheme based on the bi-linear motion model for foreground/background segmentation. Yu [12] used a combination of MV clustering and background subtraction of DC images to segment moving objects.

The proposed tracking approaches in the compressed domain rely either on MVs, residual information, or both. Most of these works exploit the information found in MPEG-1/2 streams, where MVs and DCT coefficients are easily accessible. Hesseler et al. [6] performed the tracking initialization on decoded I-frames and used histograms of MPEG-2 MVs to track the objects. The method does not support rotatingobjects and changes in size. Lie et al. [13] proposed a system that tracks single MBs under consideration of residual information. Trajectories are afterwards merged to obtain a moving object segmentation. The interested reader can find some other MPEG-2-based methods in [14–21].

Though most of the MPEG-2-based work can in principle be ported to the H.264-AVC/SVC domain, some basic assumptions are no longer valid. The often used AC and DC coefficients (e.g., [14,18–20]) of intra-coded blocks in H.264/AVC are transformed from spatially intra-predicted values instead of the original pixel values, so full decoding is necessary. A few approaches specific to MPEG-4 and H.264-AVC/SVC have been proposed in the literature. Sutter et al. [22] presented a lightweight tracking algorithm for MPEG-4/FGS. No indication for the performance in the case of multiple occluding objects is given and the system has to be initialized by the user. You et al. [23] performed the tracking of feature points selected by the user. The matching of these points uses the dissimilarity energies related to texture, form, and motion. Therefore, they partially decode the stream around the region-of-interest (ROI) back to pixel level and fully decode I-frames.