Pattern Recognition Letters 34 (2013) 1531–1539

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Low complexity object detection and tracking with inter-layer graph mapping and intra-layer graph refinement in H.264/SVC bitstreams

Houari Sabirin, Munchurl Kim *

Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

ARTICLE INFO

Article history:
Received 5 October 2012
Available online 25 June 2013

Keywords:
Graph-based method
H.264/SVC
Object tracking
Spatio-temporal graph
Video surveillance

ABSTRACT

In this paper we present a novel method of detecting and tracking moving objects in H.264/SVC bitstreams for video surveillance applications. Efficient detection and reliable tracking of real moving objects are first performed in the spatial base layer of H.264/SVC based on a spatio-temporal graph which is constructed from the block partitions with non-zero motion vectors and/or non-zero residue information. The spatio-temporal graph is utilized in reliably maintaining the real moving objects of being detected and tracked by removing false detected objects via graph pruning and graph projection. Graph matching is then performed to precisely identify the real moving objects over time even under occlusion. For low-complex but accurate detection and reliable tracking of moving objects in spatial enhancement layer of H.264/SVC, inter-layer graph mapping and intra-layer graph refinement are used without performing graph pruning, graph projection and graph matching which are mostly performed in the spatial base layer. For this, the identified block groups of the real moving objects in the spatial base layer are then mapped to the spatial enhancement layer to provide accurate and efficient object detection and tracking in the bitstreams of higher spatial resolution. Experimental results show the proposed method can reliably detect small objects, object occlusions and object separation. It also produces efficient processing time down to 27% compared to fully performing graph processing in both spatial base and enhancement layers of H.264/SVC test bitstreams.

1. Introduction

With the growing necessity of video surveillance that can be utilized in web-based home monitoring applications or smart-phone-based applications, the surveillance video applications should be able to produce different video data in regard with the computational capabilities of the terminal devices. For this, scalable video coding can be a good answer. H.264/SVC (Scalable Video Coding) is the extension to H.264/AVC (Advanced Video Coding) video coding standard, and has a combined spatial, temporal and quality scalability features (Schwarz et al., 2007) so that the scalable bitstreams of H.264/SVC can easily be adaptable to the various terminal devices and networks (Schäfer et al., 2005). For surveillance video applications based on H.264/SVC, the object detection and tracking can efficiently be performed in spatial base layers for multi-channel surveillance input video on a server side and the detection and tracking results in spatial base layers can effectively be utilized in spatial enhancement layers. Therefore, the handheld/mobile devices can take an advantage of only receiving the H.264/SVC-encoded bitstreams of spatial base layers with the information of object detection and tracking results for the surveillance video. The terminal devices connected to electric power and broadband networks may receive the H.264/SVC-encoded bitstreams of spatial enhancement layers with object detection and tracking results.

Käs et al. proposed a moving object method that utilizes motion information and partial decoding in H.264/SVC bitstream domains for traffic surveillance video (Käs et al., 2009). Although the proposed method in Käs et al. (2009) is performed in H.264/SVC, the object detection and tracking results are not related between a lower layer and its just upper layer. Käs and Nicolas proposed an object detection and tracking method which is applicable in both H.264/AVC and H.264/SVC (Käs and Nicolas, 2009). However, the inter-layer encoded parameters of H.264/SVC are not utilized. While in Käs et al. (2009) they reported that they only process the spatial base layer, detection and tracking methods in Käs and Nicolas (2009) seems to be always performed in the highest scalability layer of H.264/SVC.

We also distinguish our work from some previous works performed in H.264/AV(Käs and Nicolas, 2009) bitstreams for segmentation of moving objects (Szczerba et al., 2009; Chen et al., 2011; Poppe et al., 2009; De Bruyne et al., 2009; Moura and Hemerly, 2010; Kapotas and Skodras, 2010). The goals of the previous methods are to provide accurate object segmentation from background where some limited information such as motion information

* Corresponding author. Tel.: +82 42 350 7419; fax: +82 42 350 7619.
E-mail addresses: houari@kaist.ac.kr (H. Sabirin), mkim@ee.kaist.ac.kr (M. Kim).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.06.011

(Szczerba et al., 2009; Chen et al., 2011; Poppe et al., 2009; De Bruyne et al., 2009; Moura and Hemerly, 2010) and block partition information (Kapotas and Skodras, 2010) is used.

Graph-based approaches have been proven as one of the effective solutions for object segmentation and tracking, especially in pixel domains (Lombaert et al., 2005; Mooser et al., 2007; Gomila and Meyer, 2003; Guanling et al., 2009; Pallavi et al., 2008; Cesar et al., 2005; Caetano and McAuley, 2009). The methods based on graph cut algorithms (Lombaert et al., 2005; Mooser et al., 2007) or graph-based object tracking (Gomila and Meyer, 2003; Guanling et al., 2009; Pallavi et al., 2008), and graph matching (Cesar et al., 2005; Caetano and McAuley, 2009) are shown to be effective to detect/segment moving objects from background, and to identify/track the detected moving objects.

In this paper, a novel spatio-temporal graph based method is proposed to detect and track moving objects in spatial scalability layers of H.264/SVC (Käs and Nicolas, 2009) bitstreams by utilizing both the inter-layer encoded parameters such as motion and residual information between two spatial layers and the intra-layer coded parameters in a same spatial layer. In the proposed method, for lower processing time, the detection/tracking in a higher spatial scalability layer is performed by utilizing the detection and tracking results in its lower spatial scalability layer instead of performing object detection and tracking in both base and enhancement layer.

The proposed method in this paper is the extension to our previous work in H.264/AVC (Sabirin and Kim, 2012). A simple spatio-temporal graph for effective detection, reliable tracking and precise identification for moving objects is built. Since the spatial base layer of H.264/SVC has the same structure as H.264/AVC, our proposed method in Sabirin and Kim (2012) can easily be applied in the spatial base layer H.264/SVC and can then be extended to the spatial enhancement layers. It is noted that the proposed method in this paper considers the detection and tracking problems only in the spatial scalability layers of H.264/SVC bitstreams, which is simple and can be applicable to scalable video surveillance. Fig. 1 shows a block diagram of our proposed detection and tracking method in spatial scalability layers of H.264/SVC bitstreams.

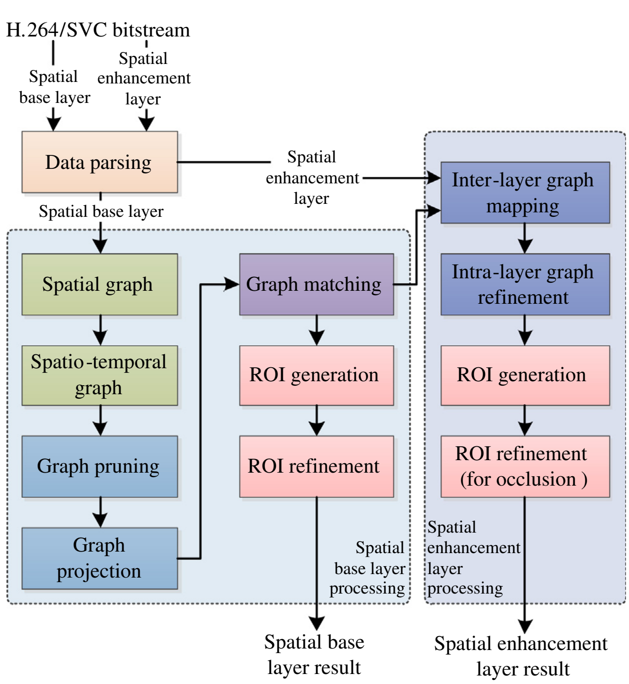

Fig. 1. A block diagram of the proposed method.

As shown in Fig. 1, the proposed method consists of two processing phases: (i) in the spatial base layer processing, spatial graph and spatio-temporal graph are constructed based on the block partitions with non-zero motion vectors and/or non-zero residue information. Then graph pruning and projection are performed to remove false-detected blocks and to recover missing blocks and their identities are defined using graph matching (Sabirin and Kim, 2012); (ii) from the spatial base layer processing, the spatial graphs that are survived from the graph pruning and that are correctly identified by the graph matching are then mapped to their corresponding spatial graphs which are constructed in the spatial enhancement layer. The vertex positions of the mapped spatial graphs are scaled up from the spatial base layer. When there is any overlap between the vertices of a spatial graph in the spatial enhancement layer and the vertices of a mapped spatial graph from the spatial base layer, that spatial graph is regarded as the corresponding real moving object to the mapped spatial graph. Next, the spatial graph is refined by performing temporal filtering on its vertices, thus resulting in an approximate region of the real object in the spatial enhancement layer. By mapping the result of graph mapping instead of the final ROI refinement result into the higher spatial layer, intra-layer graph refinement can be performed to ensure reliable object segmentation in the higher spatial layer.

It is noted that the proposed method in this paper only utilizes the spatial scalability features of H.264/SVC for the sake of efficient storage in lower spatial resolution and high-quality monitoring in higher spatial resolution. Also, the proposed method deals with H.264/SVC bitstreams recorded under fixed cameras.

This paper is organized as follow: In Section 2, we first describe the construction of the spatial and spatio-temporal graphs; In Section 3, graph pruning and graph projection are described to remove false-detected blocks, and to recover missing blocks, respectively; Section 4 discusses the tracking of moving objects using graph matching and the mapping of the identified graphs from the spatial base layer to the spatial enhancement layer; In Section 5, we present our experimental results; and we conclude the paper in Section 6.

2. Graph structure

Each frame of encoded bitstreams of H.264/SVC contains macroblock (MB) structure which is similar to that of H.264/AVC. Each MB in H.264/AVC is encoded in a block partition mode among 16×16–4×4 block partitions for Inter prediction coding or among 4×4, 8×8 and 16×16 block modes for Intra prediction coding. Therefore, each object region is represented in a group that consists of clustered 4×4 blocks with non-zero motion vectors and/or non-zero residue. Clustering of the 4×4 blocks is performed by progressively scanning the blocks and recursively detect any neighboring 4×4 blocks with non-zero motion vectors and/or non-zero residue in the 8-connection of the current scanned block.

The definitions of the spatial graph and the spatio-temporal graph are derived from the graph definitions proposed in H.264/AVC method (Sabirin and Kim, 2012). Let $G_f = \left\{ g_m^f : 0 \leqslant m \leqslant N_f - 1 \right\}$ be a spatial graph that consists of subgraphs g_m^f in a frame f where N_f is the total number of the detected block groups. Each subgraph is an undirected attributed graph, $g_m^f = (V_m^f, E_m^f, a_m^f)$, that represents a moving object candidate. A set of vertices $V_m^f = \left\{ v_{km}^f : 1 \leqslant k \leqslant \left| g_m^f \right| \right\}$ denotes the 4×4 blocks for which each 4×4 block corresponds to a vertex v_{km}^f in a block group that corresponds to the subgraph g_m^f. The order $|g_m^f|$ is the number of 4×4 blocks in the group. The edge $E_m^f \equiv E\left(v_{jm}^f, v_{km}^f \right) = 1$ if two vertices v_{jm}^f and v_{km}^f are in touch each other where $v_{jm}^f, v_{km}^f \in V_m^f$ and