Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Real-Time Imaging 11 (2005) 290–299
REAL-TIME IMAGING
www.elsevier.com/locate/rti

Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model

Wei Zeng[a,*], Jun Du[b], Wen Gao[c], Qingming Huang[b]

[a]Department of Computer Science and Engineering, Harbin Institute of Technology, No. 92 West Da-Zhi Street, Harbin 150001, People's Republic of China
[b]Graduate School of Chinese Academy of Sciences, No. 19 Yu-Quan Road, Beijing 100039, People's Republic of China
[c]Institute of Computing Technology, Chinese Academy of Sciences, No. 8 Ke Xue Yuan Nan Road, Beijing 100080, People's Republic of China

Available online 22 June 2005

Abstract

Moving object segmentation in compressed domain plays an important role in many real-time applications, e.g. video indexing, video transcoding, video surveillance, etc. Because H.264/AVC is the up-to-date video-coding standard, few literatures have been reported in the area of video analysis on H.264/AVC compressed video. Compared with the former MPEG standard, H.264/AVC employs several new coding tools and provides a different video format. As a consequence, moving object segmentation on H.264/AVC compressed video is a new task and challenging work. In this paper, a robust approach to extract moving objects on H.264/AVC compressed video is proposed. Our algorithm employs a block-based Markov Random Field (MRF) model to segment moving objects from the sparse motion vector field obtained directly from the bitstream. In the proposed method, object tracking is integrated in the uniform MRF model and exploits the object temporal consistency simultaneously. Experiments show that our approach provides the remarkable performance and can extract moving objects efficiently and robustly. The prominent applications of the proposed algorithm are object-based transcoding, fast moving object detection, video analysis on compressed video, etc.

1. Introduction

Moving object segmentation aims at partitioning an image sequence into meaningful regions along the time axis. In general, moving object segmentation techniques in pixel domain have to fully decode the compressed video first. Such algorithms are quite accurate but cannot fulfill the requirement of real-time applications. As a consequence, fast algorithms to segment moving objects performed directly on compressed video are desired. Such type of video processing in compressed domain plays an important role in many real-time applications, e.g. video indexing, video coding, video surveillance, video manipulation, etc. [1–4].

Moving object segmentation algorithms in compressed domain usually rely on two types of features in terms of macroblock (MB): motion vector (MV) and DCT coefficients. MVs are obtained in the motion compensation between the current frame and its reference frames block by block. MV presents the temporal correlation between two frames and provides the displacement of the block. All MVs in one frame can be treated as a sparse motion vector fields. On the other hand, the DCT coefficients of an MB carry the image information. For the inter-coded block, DCT coefficients contain the residues of the motion compensation. For the intra-coded block, DCT coefficients are transformed signal of the original image. Therefore, the block DCT coefficients can be used to reconstruct the DC image [5] or treated as the texture feature to measure the similarity of blocks [6].

However, as the up-to-date video coding standard, H.264/AVC employs several new coding tools and

*Corresponding author. *Post address*: JDL, P.O. BOX 2704, Beijing 100080, China. Tel.: +86 10 58858300(348); fax: +86 10 58858301.
E-mail addresses: wzeng@jdl.ac.cn, wzeng163@163.com (W. Zeng), jdu@jdl.ac.cn (J. Du), wgao@jdl.ac.cn (W. Gao), qmhuang@jdl.ac.cn (Q. Huang).

1077-2014/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.rti.2005.04.008

provides a different video format [7]. As a consequence, moving object segmentation on H.264/AVC compressed video is a challenging work. Very little work has been carried out in the area of video analysis on H.264/AVC compressed video. H.264/AVC has some new characters from the video analysis point of view. In H.264/AVC, the intra-coded block is spatial intra-predicted according to its neighbor pixels. So, the DCT coefficients provide the spatial prediction residues information for blocks now. On the other hand, H.264/AVC supports variable block-size motion compensation. An MB may be partitioned into several blocks and has several MVs. As a result, the MV field for H.264/AVC compressed video consists of MVs with variant block size. This is quite different from the former MPEG standard video with regular block size MVs. Therefore, there is a requirement of efficient object extraction technique for the H.264/AVC compressed video. In this paper, we propose a novel object segmentation algorithm using the block-based Markov Random Field (MRF) model to extract moving objects based on the MVs and block DC coefficients with decoding. The proposed approach treats the object segmentation as a Markovian block labeling process and integrates object tracking in the uniform MRF model simultaneously. The method can extract moving objects efficiently and robustly.

This paper is organized as follows. First, Section 2 briefly reviews some related work of moving object segmentation over MPEG compressed domain. Then, an overview of the proposed algorithm is presented in Section 3. Our algorithm consists of two stages: the MV classification and the MRF classification. They are detailed in Sections 4 and 5, respectively. The process of I-frame segmentation is discussed in Section 6. Experimental results are presented in Section 7, and finally concluding remarks are provided in Section 8.

2. Related work

Recently, some moving object segmentation algorithms over MPEG compressed domain have been reported. In MPEG compressed video, pictures are encoded in terms of I-frame, P-frame and B-frame. P-frames and B-frames store the motion information and residues of the motion compensation, I-frame has no motion information and stores the DCT transformed signals of the original image. Thus, I-frame can provide texture or color information without decoding. Most of the object segmentation algorithms in compressed domain employ MVs and DCT coefficients to extract moving objects. Based on the features and decision rules these algorithms used, we group the algorithms into three categories as the motion-based segmentation, change-based segmentation and spatio-temporal segmentation.

The motion-based segmentation techniques extract regions with homogeneous motion on the MV field. Zen et al. [8] have proposed a simple approach to detect moving objects from the MPEG stream without decoding process. In their algorithm, if the magnitude of an MV is above a predefined threshold, the block is classified as moving block. All moving blocks are spatially merged into moving objects according to the MV angle similarity. Zen's method can extract moving objects from the fixed camera video and is used for video surveillance applications. Babu et al. [9] have proposed automatic video object segmentation algorithm for the MPEG video. They first estimate the number of independently moving objects in the scene using a block-based affine clustering method. The object segmentation is then obtained by the exception maximization (EM) clustering algorithm. Their approach can handle multiple moving objects and extract video object precisely by refining the decoded edge blocks. Jamrozik and Hayes [10] have developed a compressed domain video object segmentation system. Their system merges the regions through the leveled watershed technique on the motion map. The motion map is the MV accumulation over a number of frames and represents the MV distribution. Their approach provides a foundation of spatial segmentation of compressed video and achieves very fast segmentation.

The change-based segmentation techniques extract moving objects relying on local motion-related measurements. The changed parts conducted by the objects motion in a frame are considered as moving regions. Therefore, change-based segmentation approach segments foreground objects through detecting changed regions in frames. Ji and Park [6] have proposed the dynamic region decision approach to detect and track moving objects in DCT-based compressed video. Yu et al. [11] have developed a robust object segmentation algorithm using MV and DCT coefficients jointly. The MBs that are different from the reconstructed background or have the true MVs are identified as the moving blocks in their algorithm. Zeng et al. [12] have proposed an approach to extract moving objects based on change detection of the DC image difference. Their method can efficiently segment moving objects even if the MV field has majority of false MVs. Benzougar et al. [13] have proposed an MRF-based moving object detection algorithm based on MVs and the displaced frame difference of DC images. Their approach combines the motion and image information in a Markovian labeling framework and achieves the robust moving object segmentation.

The spatio-temporal segmentation techniques integrate spatial information into the object segmentation process. The region information of image segmentation is used to guide the object pixel classification and guarantees that the extracted object boundaries coincide with the region boundaries. Eng and Ma [14] have proposed a spatio-temporal moving object segmentation algorithm based on the spatial segmentation of DC coefficients and the temporal segmentation of MB MVs. They adopt the maximum entropy fuzzy clustering algorithm in the spatial and temporal segmentation process. In their algorithm, the spatial segmentation is used to refine the object boundary of the temporal segmentation result. Sukmarg and Rao [15] have proposed a fast object and segmentation approach in MPEG compressed domain. They first cluster the blocks into homogeneous regions based on the color and AC-energy. The regions are then merged into large ones based on the proposed spatio-temporal similarity measurement. The large regions are classified as background and foreground according to their average change. High temporal changed regions are classified as objects finally. Wang et al. [16] have proposed a confidence measure based moving object extraction system. Their system combines the spatial, temporal and texture confidence to obtain the robust motion information first. Moving blocks are then extracted by the outlier detection of dominant motion. The object masks are obtained in the spatial clustering finally.

The proposed method is motion-based segmentation on H.264/AVC compressed video. In our algorithm, the moving object segmentation is treated as a Markovian labeling procedure. Similar MVs are merged into moving objects through the minimization of the MRF energy. To employ the temporal coherence of object labels, object tracking is integrated in the uniform MRF model. Because H.264/AVC supports variable block-size motion compensation, the blocks size information is also considered as a clue in the proposed model.

3. Overview of the proposed algorithm

Three types of temporally interleaved frames are supported in H.264/AVC bitstream. The first is I-frame that is intra-coded on 16×16 or 4×4 pixel blocks. The second is P-frame that is motion compensated in the forward direction from I-frame or other P-frame. The third is B-frame that is motion compensated in both directions. A group-of-pictures (GOP) refers to the frames between I-frames. Because consecutive P-frames can provide continues motion information through the whole video, only P-frames and I-frames are used in our implementation. The overall algorithm is illustrated in Fig. 1.

In our algorithm, MVs of P-frames are first buffered to conduct the MV fields. Moving objects are then extracted from the MV field through the MRF classification process. The algorithm consists of two stages. In the first stage, MVs are classified into several MV types. Different MV types will provide different contributions to the next step's segmentation. In the second stage, moving blocks are extracted by the MRF classification. For object segmentation of I-frames, we implement an object label projection scheme to track the segmentation results of the previous P-frame, and project the labels to the current I-frame by means of inverting the MVs in the previous P-frame.

4. Motion vector classification

Because only the MV corresponding to the true motion will provide reliable motion information, the MV related to the real object should be identified first. Since MVs are issued from a coding-oriented criterion only, the MV field is quantized and noisy. This leads to a

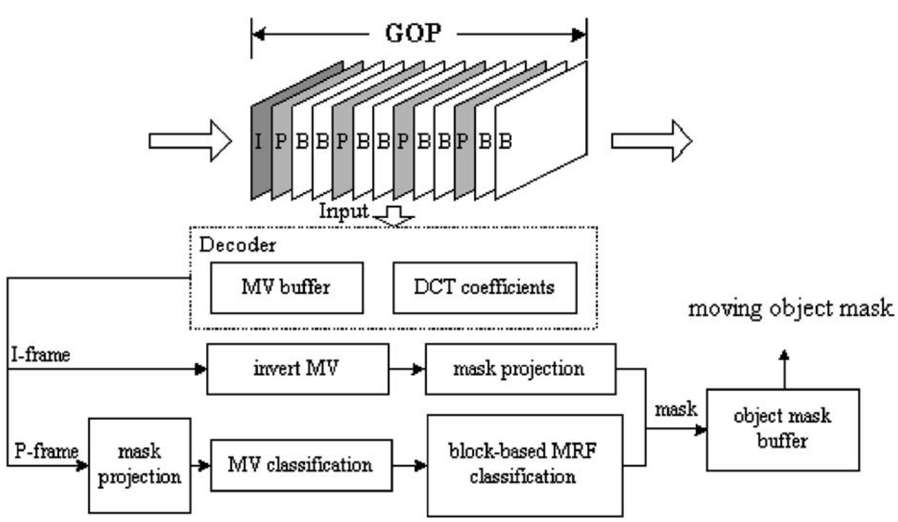

Fig. 1. Block diagram of the proposed algorithm.

fundamental drawback that constrains the video-processing algorithm on compressed video from achieving satisfactory results. In order to filter out noise and recover the true MV, we design a classification process to determine the regularity of MVs.

Four MV types are defined in our algorithm: (1) background MV (BMV), (2) edge MV (EMV), (3) foreground MV (FMV), and (4) noise MV (NMV). Different types of MVs will provide different contributions to the MRF classification stage. BMV is defined as a small MV whose magnitude is lower than a predefined threshold Th_b. The blocks with BMV are usually corresponding to background regions in the scene. EMV is defined as the salient MV. The blocks with EMV may imply the object boundaries. Starting from the upper-left corner of the MV field, an EMV is detected by comparing the ith MV (MV_i) with the average MV ($MV_{avg}(i)$) of its four neighbors by the following rule:

$$|MV_i - MV_{avg}(i)| > Th_e. \tag{1}$$

Because the blocks in H.264/AVC are variable size in motion compensation, the number of neighboring MVs is variant, and determined by the related MB partition type. FMV is defined as the MV whose magnitude is bigger than the threshold Th_f. Blocks with FMV will indicate they have true MV. NMV is defined as the noise MV whose magnitude is between Th_b and Th_f. It should be noted that both FMV and NMV are not EMV. The thresholds mentioned above are determined by the noise level of MV field and depend on the sequence. These thresholds can be set manually according to the user's experience or calculated through the statistic from the training data.

5. MRF classification

Moving object segmentation can be treated as a Markovian labeling procedure on the classified MV field. MRF model provides spatial continuity that is inherent in nature images, and is used to guide the block merging process under the maximum a posteriori (MAP) criterion. In the proposed model, three clues are taken into account and integrated in the uniform MRF model. The first clue is the MV similarity that is used to merge blocks into the motion homogeneous region. The second clue is the temporal consistency that is used to track object labels through several frames, and provides the temporal homogeneity in the segmentation process. The third clue is the block size that is used to remove noises in the final segmentation. Different from the rectangular lattice that composes the image, we define the MRF model over the set of blocks obtained from the compressed bitstream. Due to the fact that the output of the MRF labeling is the object identification labels, we also call the labeling process as the MRF classification.

Denote a random field $X = \{x_i | i \in \Lambda\}$, where Λ is a set of indices for blocks. Each element x_i corresponds to one of the labels related to the objects in the scene. Given the observation (the classified MV field, MVF), the goal of moving object segmentation is to estimate the configuration x of the random field X, in which each x_i has been assigned a most probable label. Two object labels are defined in our model. One is the foreground label (FL), and the other is the background label (BL). Using the MAP criterion, the segmentation can be solved by the following minimization problem:

$$\widehat{x} = \arg\min_x U(x|MVF), \tag{2}$$

where $U(x|MVF)$ is the posterior energy function and formulated as

$$U(x|MVF) = \sum_i V^M(x_i, MV_i) + \sum_i V^T(x_i, MV_i) + \sum_{(i,j)\in C_i} V^S(x_i, MV_i, x_j, MV_j). \tag{3}$$

The energy function in Eq. (3) consists of three terms. The first is a data-driven term, which describes the likelihood between the label and the MV type. The second is a tracking term, which describes the temporal continuity of the moving objects. The last is a spatial smoothness term, which enables the segmented object mask to be compact. $V^M(x_i, MV_i)$ can be written as

$$V^M(x_i, MV_i) = \begin{cases} -a & \text{if } x_i = FL\&FMV \text{ or } x_i = BL\&BMV, \\ 0 & \text{if } EMV \text{ or } NMV, \\ +a & \text{if } x_i = BL\&FMV \text{ or } x_i = FL\&BMV, \end{cases} \tag{4}$$

where EMV and NMV contribute zero energy. If the label x_i coincides with the type of MV_i, the term energy will decrease and vice versa. The tracking energy term $V^T(x_i, MV_i)$ is computed by projecting foreground labels from the previous segmentation

$$V^M(x_i, MV_i) = \begin{cases} -\beta P(MV_i) & \text{if } x_i = FL, \\ +\beta P(MV_i) & \text{if } x_i = BL. \end{cases} \tag{5}$$

$P(MV_i)$ is the projection function that calculates the proportion of projected foreground labels of current block i. The spatial smoothness term $V_i^s(x_i, MV_i, x_j, MV_j)$ is defined on the pair-site clique and formulated as

$$V^S(x_i, MV_i, x_j, MV_j) = \begin{cases} -\gamma \dfrac{n_{ij}}{n_i} S(MV_i, MV_j) - \eta W(i)_i & \text{if } x_i = x_j, \\ +\gamma \dfrac{n_{ij}}{n_i} S(MV_i, MV_j) + \eta W(i) & \text{if } x_i \neq x_j, \end{cases} \tag{6}$$

where n_{ij} and n_i are the boundary lengths. $S(MV_i, MV_j)$ is the MV similarity function that is deduced from the normalized Euclidean distance. $W(i)$ is the weight function for block i. We set $W(i)$ to 1.0 for the 4×4 blocks and 0.5 for the larger blocks. Such strategy can efficiently remove isolated blocks. α, β, γ and η are constants, and determined by the relative weights of each energy term in Eq. (3).

The minimization of energy function is performed by the iterative deterministic relaxation algorithm known as highest confidence (HCF) [17]. Inspired by the work of Tsaig and Averbuch [18], we only change the labels on the object boundaries in each step. The process is very effective on keeping compactness for the segmented object mask as well. The modified HCF algorithm can be described as follows.

(1) Assign each x_i, as
$$x_i = \begin{cases} \text{FL} & \text{if } MV_i = \text{FMV or EMV}, \\ \text{BL} & \text{if } MV_i = \text{BMV or NMV}. \end{cases}$$
(2) Create heap.
(3) Insert boundary blocks to the heap.
(4) While the stability for the top element k in the heap is lower than zero:
   (a) Change the label of block k.
   (b) Update the stability of block k.
   (c) Adjust the heap for the block k.
   (d) For each block j that belongs to the pair-site clique of block k:
      (i) Update the stability of block j.
      (ii) Adjust the heap for the block j.
(5) If label is changed, remove non-boundary blocks from the heap and insert new boundary blocks to the heap, go to step 4; else stop.

6. Object label projection for I-frame

In H.264/AVC compressed video, only P-frame and B-frame have the MVs. I-frame is intra-coded picture and has no MVs. Two block coding types are defined in H.264/AVC: Intra_4×4 and Intra_16×16. Both the types of blocks are coded by the spatial prediction from the adjacent pixels. The Intra_4×4 type is corresponding to the block with 4×4 size, while the Intra_16×16 type to the 16×16 block. Thus, the number of blocks in an I-frame vary and depend on the frame content.

To obtain the segmentation results for I-frames, we implement an object label projection approach to transfer the previous segmentation results of P-frame to the current I-frame. The idea underlying this method is that object motion can be regarded as the uniform speed motion approximately. Therefore, the MVs of current I-frame can be obtained through inverting the MVs of the previous P-frame. Fig. 2 is an illustration.

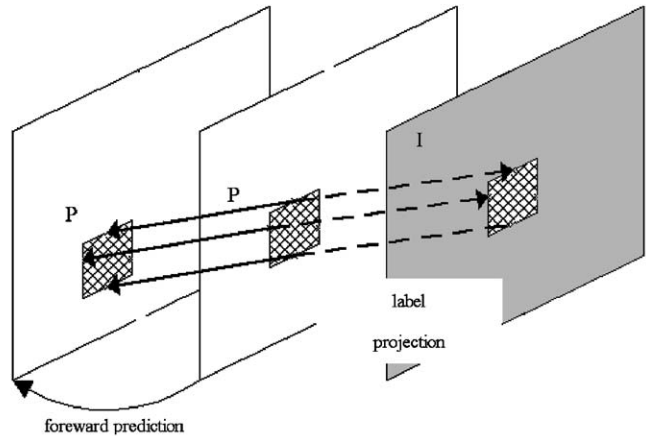

Fig. 2. Illustration of object label projection for I-frames.

The broken lines show the MV inversion. Once the segmentation result of the previous P-frame is obtained, the segmentation of I-frame can be computed by projecting the object labels to the current I-frame according to the inverted MVs. In I-frame, if a block has a majority of one kind of object labels, the block is assigned with this type of label at last.

7. Experimental results

The proposed algorithm is evaluated by two outdoor surveillance sequences. The first is *PIE* sequence with CIF (352×288) format. It is captured with a long focal length in a high speedway. The second is *ETRI_od_A* sequence with SIF (352×240) format. It is a middle focal length sequence and selected from the 30th CD in the MPEG-7 content set. Figs. 3(a) and 4(a) show two original frames of the sequences. In our experiments, both the test sequences are compressed using the H.264 encoder of version JM 6.1e.

Figs. 3 and 4 display the segmentation results for the PIE sequence and the *ETRI_od_A* sequence, respectively. The segmented foreground object blocks are marked out with rectangles. Fig. 3(b,c) shows the masks of the blocks with FMV, EMV or NMV. If the block with FMV, EMV or NMV is treat as moving block, this process is equal to a threshold-based motion detection approach. That is to say, moving blocks are detected according to their MV magnitudes. Therefore, the output of the MV classification provides the simply moving object detection. From Fig. 3(b,c), it can be seen that some background blocks are false classified and there exist many noise blocks. Figs. 3(d–f) and 4(b–d) show the segmentation results of the proposed algorithm. Most of the noise blocks are removed, and the obtained object masks are more compact. Some false classified background blocks have been corrected, even if these blocks have the spurious MVs. Although the

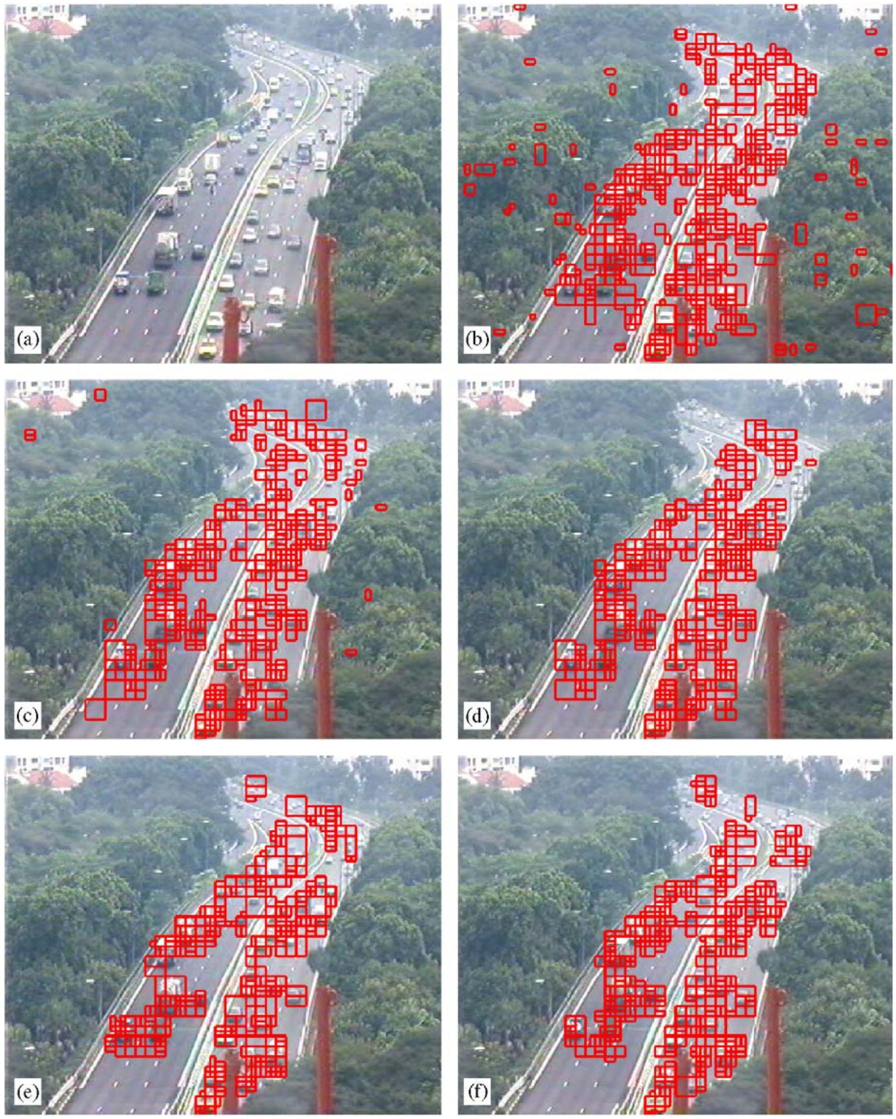

Fig. 3. Segmentation results for the *PIE* sequence: (a) the 21st frame, (b,c) FMV, EMV and NMV masks for the 18th and 21st frame, (d–f) segmentation results for the 21st, 24th, 27th frame.

human body exhibits a non-rigid motion in the *ETRI_od_A* sequence, our algorithm still achieves satisfactory results (see Fig. 4). Fig. 5 displays the results of object label projection for I-frames. Due to the MV noise, the projected object labels have some errors in the results.

To evaluate the performance of the proposed MRF model, we compare the segmentation results between the proposed algorithm and the MV classification. Two measurements are defined to describe the segmentation performance. The first measurement is recall that is defined as the percentage of the correctly detected pixels to the real moving objects. The second measurement is precision that is defined as the percentage of the detected pixels to all detected moving object pixels. High recall means less miss-segmentation, while high precision means less false segmentation. To obtain the ground truth, we manually labeled the moving objects of 105 frames in both of the sequences. Fig. 6 shows the example of the labeled real moving objects of the two sequences. It is worth mentioning that as the segmentation is based on blocks, the labeled object mask is also based on blocks. The block shape and number will depend on the configuration of the used H.264/AVC encoder and the content of sequence. Fig. 7 illustrates the number of blocks for the labeled real objects and the frames. The block numbers of the *PIE* sequence are high because there are too many small cars in the scene (see Fig. 7(b)). The precision and recall curves of the *ETRI_od_A* sequence and the *PIE* sequence are shown in Fig. 8(a,b) and (c,d), respectively. From Fig. 8, it can be clearly seen that the proposed algorithm improves the precision and maintains the recall at the same time. Table 1 shows the average precision and recall of the

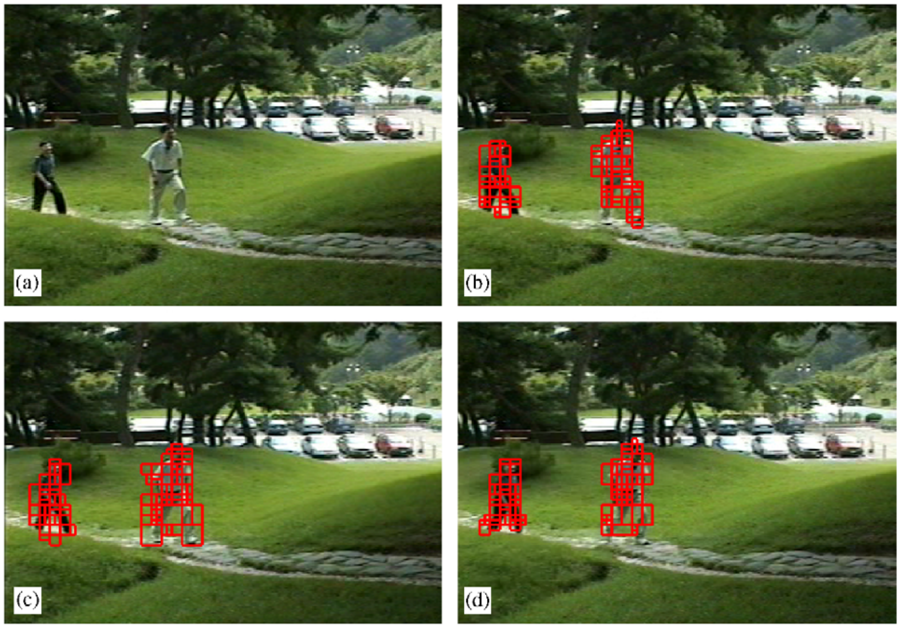

Fig. 4. Segmentation results for the *ETRI_od_A* sequence: (a) the 66th frame, (b–d) segmentation results for the 66th, 69th, 72nd frame.

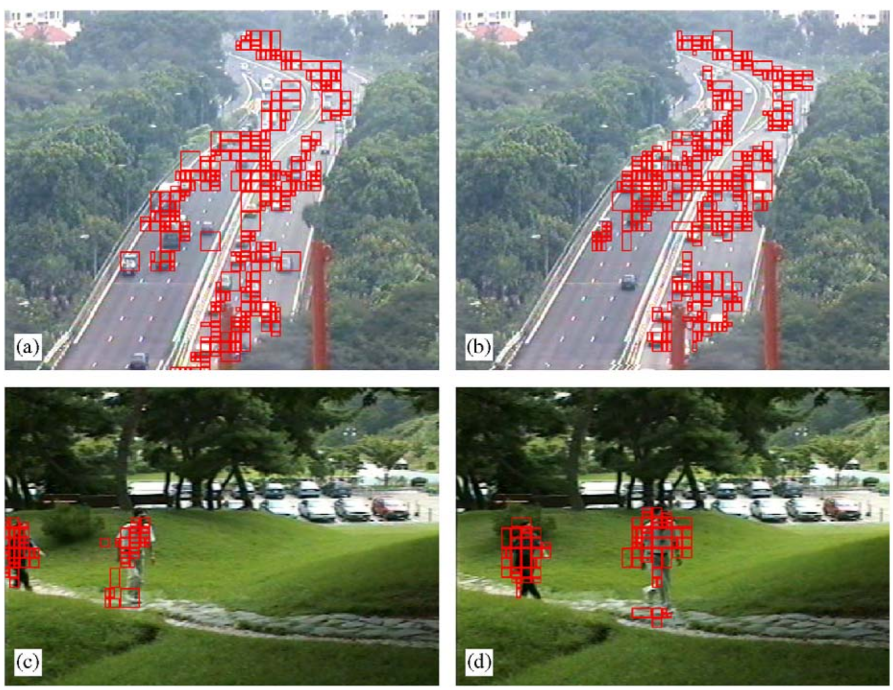

Fig. 5. Results of object label projection for I-frames: (a,b) the projection results for 45th, 90th frame of the *PIE* sequence, (c,d) the projection results for 45th, 90th frame of the *ETRI_od_A* sequence.

ETRI_od_A sequence and the *PIE* sequence. The precision of the *PIE* sequence is a little low in a sense. It is caused by the following three reasons. The first is that too many occluded blocks exit in the *PIE* sequence. Sometimes, occluded blocks may be classified as the moving blocks when they have spurious MVs. The second reason is that the moving cars are small and close to each other. The blocks among the closed moving blocks probably have little moving object pixels and will be imposed as spurious MVs. The third reason is that

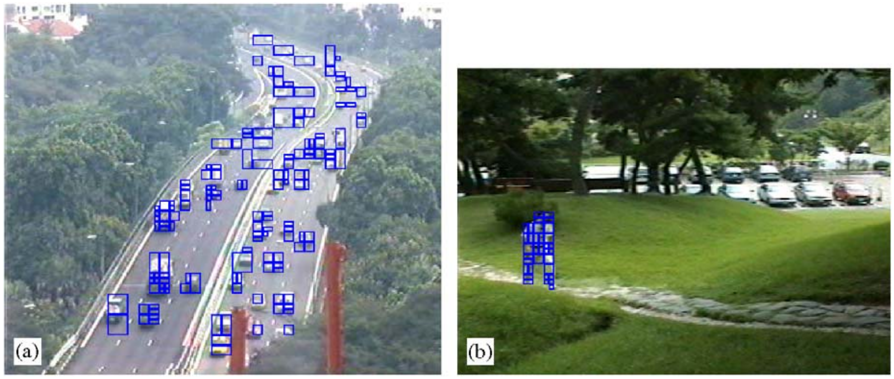

Fig. 6. Example of the manually labeled real moving objects: (a) the labeled objects for 15th frame of *PIE* sequence, (c) the labeled objects for 15th frame of the *ETRI_od_A* sequence.

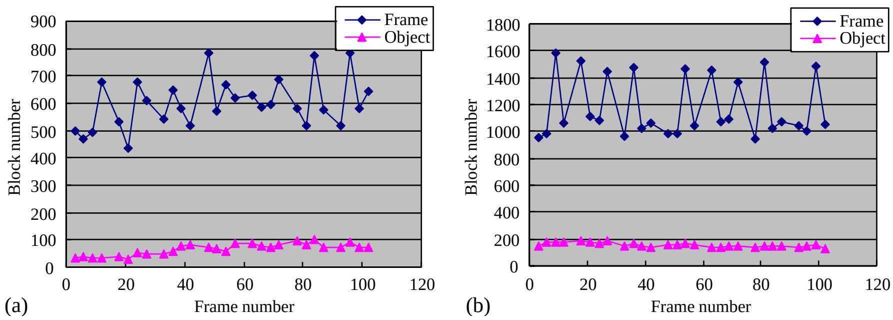

Fig. 7. Block numbers of object and frame for the two sequences: (a) the *ETRI_od_A* sequence, (b) the *PIE* sequence.

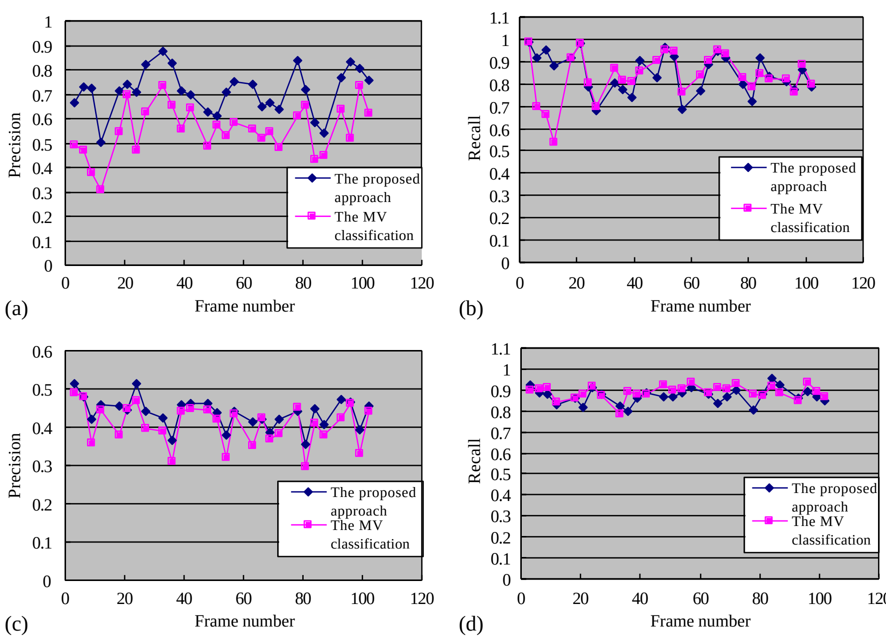

Fig. 8. Precision and recall of the two sequences: (a) the precision of the *ETRI_od_A* sequence, (b) the recall of the *ETRI_od_A* sequence, (c) the precision of the *PIE* sequence, (d) the recall of the *PIE* sequence.

Table 1
Average precision and recall of the two sequences

Method	The sequence *ETRI_od_A*		The sequence *PIE*	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
The proposed approach	71.3	87.2	43.7	83.5
The MV classification	55.4	90.0	40.7	84.8

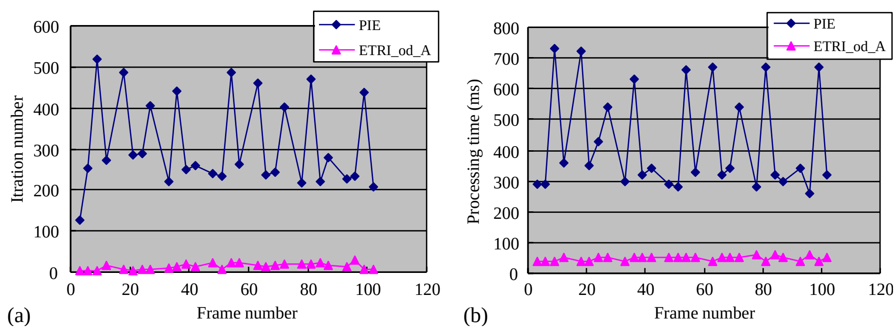

Fig. 9. Iteration number and processing time of the two sequence: (a) the number of iteration number, (b) the processing time.

the size of object is varying. The cars are changing their size along the time axis, and the cars are moving slowly in the far end on the way. To improve the precision for the *PIE* sequence, we may use the edge information to refine the segmentation further. Partially decoding the images may be a solution to improve the segmentation precision as proposed in [9].

Fig. 9 shows the processing speed of the proposed algorithm. The hardware platform for testing is a Pentium III 700M Hz portable PC with 256M ram. Fig. 9(a) illustrates the number of iteration in the minimization of the energy function in Eq. (3). Fig. 9(b) shows the processing time of the algorithm. It can be seen that the iteration number and CPU time for the PIE sequence is high. This is because the number of moving blocks is large in the *PIE* sequence as in Fig. 7. The processing time of MRF classification will be influenced by the size of foreground object and depends on the algorithm initialization. From Fig. 9, it can be seen that our algorithm has fast processing speed and can be used in the real-time applications.

8. Conclusions and future work

This paper presents a robust approach to segment moving objects on H.264/AVC compressed video. Using the block-based MRF model, the proposed algorithm efficiently segments moving objects from the noisy MV field. Our method can be used in object-based transcoding, fast moving object detection, video analysis on compressed video, etc. Although the proposed approach is designed for H.264/AVC, it can be easily extended to other video format, such as MPEG-1, MPEG-2, H.261, etc. The algorithm is suitable for videos captured from the fixed cameras currently, and it can be extended to the videos from the moving camera by adopting the global motion compensation technique [19].

Because H.264/AVC is the up-to-date video-coding standard, we just provide a fundamental investigation on video processing in H.264/AVC compressed domain in this paper. As the intra-mode provides a spatial correlation among the current block and its neighbors, future work will focus on employing the spatial information to improve the segmentation precision.

Acknowledgements

The work is supported by the NEC-JDL joint project and One Hundred Talents Plan of Chinese Academy of Science. The authors thank Dr. Siwei Ma for providing the H.264/AVC CODEC. The authors appreciate the help from the anonymous reviewers for their valuable and constructive comments.

References

[1] Mandal MK, Idris F, Panchanathan S. A critical evaluation of image and video indexing techniques in the compressed domain. Image and Vision Computing 1999;17(7):513–29.

[2] Zhang HJ, Low CY, Smoliar SW. Video parsing and browsing using compressed data. Multimedia Tools Application 1995;1(1): 89–111.

[3] Chang SF. Exploring functionalities in the image/video compressed domain. ACM Computing Surveys 1995;27(4):573–5.

[4] Lee SW, Kim YM, Choi SW. Fast scene change detection using direct feature extraction from MPEG compressed video. IEEE Transactions on Multimedia 2000;2(4):240–54.

[5] Yeo BL, Liu B. Rapid scene analysis on compressed video. IEEE Transaction on Circuits Systems for Video Technology 1995;5(6): 533–44.

[6] Ji S, Park HW. Moving object segmentation in DCT-based compressed video. Electronics Letters 2000;36(21):1769–70.

[7] Wiegand T, Sullivan GJ, Bjøntegaard G, Luthra A. Overview of the H.264/AVC video coding standard. IEEE Transaction on Circuits Systems for Video Technology 2003;13(7):560–76.

[8] Zen H, Hasegawa T, Ozawa S. Moving object detection from MPEG coded picture. Proceedings of the IEEE International Conference on Image Processing 1999;4:25–9.

[9] Babu RV, Ramakrishnan KR, Srinivasan SH. Video object segmentation: a compressed domain approach. IEEE Transaction on Circuits Systems for Video Technology 2004;14(4): 462–74.

[10] Jamrozik ML, Hayes MH. A compressed domain video object segmentation system. Proceedings of IEEE International Conference on Image Processing, 2002.

[11] Yu X, Duan LY, Tian Q. Robust moving video object segmentation in the MPEG compressed domain. Proceedings of the IEEE International Conference on Image Processing 2003;3: 933–6.

[12] Zeng W, Gao W, Zhao D. Automatic moving object extraction in MPEG video. Proceedings of the IEEE International Symposium on Circuits and Systems 2003;2:524–7.

[13] Benzougar A, Bouthemy P, Fablet R. MRF-based moving object detection from MPEG coded video. Proceedings of the IEEE International Conference on Image Processing 2001;3:402–5.

[14] Eng HL, Ma KK. Spatiotemporal segmentation of moving video objects over MEPG compressed domain. Proceedings of the IEEE International Conference on Multimedia and Expo 2000;3: 1531–4.

[15] Sukmarg O, Rao KR. Fast object detection and segmentation in MPEG compressed domain. Proceedings of the TENCON 2000 2000;2:364–8.

[16] Wang R, Zhang HJ, Zhang YQ. A confidence measure based moving object extraction system built for compressed domain. Proceedings of the IEEE International Symposium on Circuits and Systems 2000;5:21–4.

[17] Li SZ. Markov random field modeling in computer vision. Berlin: Springer; 1995 ISBN:4-431-70145-1.

[18] Tsaig Y, Averbuch A. Automatic segmentation of moving objects in video sequences: a region labeling approach. IEEE Transactions on Circuits Systems for Video Technology 2002;12(7): 597–612.

[19] Smolic A, Ohm JR. Robust global motion estimation using a simplified M-estimator approach. Proceedings of the IEEE International Conference on Image Processing 2000;1:868–71.