

Research Article

Nonuniform Video Size Reduction for Moving Objects

Anh Vu Le,¹ Seung-Won Jung,² and Chee Sun Won¹

¹ Department of Electrical and Electronic Engineering, Dongguk University-Seoul, Seoul 100-715, Republic of Korea

² Department of Multimedia Engineering, Dongguk University-Seoul, Seoul 100-715, Republic of Korea

Correspondence should be addressed to Chee Sun Won; cswon@dongguk.edu

Received 19 June 2014; Revised 4 August 2014; Accepted 15 August 2014; Published 31 August 2014

Academic Editor: Young-Sik Jeong

Copyright © 2014 Anh Vu Le et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Moving objects of interest (MOOIs) in surveillance videos are detected and encapsulated by bounding boxes. Since moving objects are defined by temporal activities through the consecutive video frames, it is necessary to examine a group of frames (GoF) to detect the moving objects. To do that, the traces of moving objects in the GoF are quantified by forming a spatiotemporal gradient map (STGM) through the GoF. Each pixel value in the STGM corresponds to the maximum temporal gradient of the spatial gradients at the same pixel location for all frames in the GoF. Therefore, the STGM highlights boundaries of the MOOI in the GoF and the optimal bounding box encapsulating the MOOI can be determined as the local areas with the peak average STGM energy. Once an MOOI and its bounding box are identified, the inside and outside of it can be treated differently for object-aware size reduction. Our optimal encapsulation method for the MOOI in the surveillance videos makes it possible to recognize the moving objects even after the low bitrate video compressions.

1. Introduction

Surveillance cameras are ubiquitous and play an important role in our daily life. The recorded video data from the surveillance cameras provide rich information to many applications ranging from human and machine interactions [1–3] to content indexing and retrieval [4, 5]. For such applications of digital video surveillance and digital video recording (DVR) systems [6, 7], it is often required to examine moving objects for a long period of frames in recorded videos. This naturally demands highly efficient compressions for a huge amount of video data. Here, the conflicting requirement is how to maintain high visual quality, especially for the important information in the video such as the moving objects, in low bit-rate compressions.

A wide range of advanced techniques has been proposed to improve the conventional video compression framework. For example, an efficient block mode determination algorithm [8] was applied for an efficient scalable video compression, where video data can change their resolution to use the limited bandwidth efficiently. The scalable compression scheme is particularly useful for surveillance videos.

Note that surveillance videos usually consist of alternating sequence of frames with static background and moving objects. Definitely, the moving objects are the important data to be preserved in the compressions. This requires the compression technique to distinguish the important moving objects of interest (MOOI) from the unimportant static background (non-MOOI) in the video and to treat them differently in the compression process. As a result, the natural user interface (NUI) via the face detection [9, 10] in the surveillance videos can be a feasible technique even for highly compressed videos. To differentiate the MOOI from the non-MOOI, the object segmentation and tracking processes can be applied [6, 7]. However, these methods need to identify accurate object boundaries, which often require expensive computations. Weng et al. [11] used Kalman filter for object detecting and tracking. This method can detect and track the object trajectory frame by frame accurately. However, the object boundary that differentiates the MOOI from the non-MOOI cannot be identified clearly. Goswami et al. [12] used a mesh-based technique to track moving objects in video sequence. Mesh-based motion estimation techniques

are more accurate than the block based method, but they are relatively slow due to the high computational complexity.

In this paper we differentiate the MOOI from the non-MOOI by detecting the bounding boxes surrounding the MOOIs for each group of frames (GoF). Then, the detected bounding boxes encapsulating the MOOIs are fixed throughout the GoF. To detect the bounding box we need to identify the pixels with spatiotemporal saliency. For this, we construct the spatiotemporal gradient map (STGM) of a GoF [13], where each pixel in the STGM represents the level of the temporal and spatial saliency. Then, the optimal size of the bounding box is determined to include the local pixels with the highest energy density of the STGM. Once the pixels including the MOOI are determined by the bounding boxes, we can apply linear transformations with different slopes to the inside and outside of the bounding boxes such that the MOOI is intact while those in the non-MOOI are the main target for size reduction. After this initial data reduction, the standard H.264/AVC compression is applied to the size-reduced frames for further compressions. At the receiver, the reverse processes including decompression by H.264/AVC and the size expansion by the inverse linear transformations are applied to restore the video data with the original size. The overall block diagram of our MOOI-based compression is shown in Figure 1.

As far as the image size reduction is concerned, various methods have been proposed for content-aware image and video retargeting context such as the seam carving methods [13–17]. These methods reduce the image size by removing the unimportant seam lines that have the low saliency. The output video has the reduced spatial resolution, where the rich texture areas are maintained but the homogenous areas are removed. These video retargeting methods are mainly for display purposes and it is not reversible to reconstruct the original image size from the retargeted videos unless the decoder knows exact locations of the discarded seam pixels. Therefore, the conventional video retargeting methods are not appropriate as the initial data reduction for video compression. Note that image pruning scheme with image downsampling as a preprocessing step of video compressions has been also used in Vo et al. [18], where one of the two consecutive image lines (i.e., even or odd lines) is to be discarded for image size reduction. Since the line dropping is limited for one of two consecutive lines and the criterion for line dropping is based on the least mean square errors (LMSE) of the interpolated image data, it is hard to differentiate the MOOI from the non-MOOI. A reversible nonuniform size reduction method was also proposed in Won and Shirani [19] without the bounding box. Finally, we note that this paper is the extended version of our previous single MOOI [20] to multiple MOOIs.

Our contributions of this paper can be summarized as follows: (i) we introduce a spatiotemporal gradient map (STGM) to trace the boundary of the MOOI within a GoF; (ii) based on the STGM, a cost function for determining the center and size of the bounding box encapsulating the MOOI is formulated; (iii) an optimization process for updating the center and size of the bounding box alternately is introduced; (iv) the subjective visual quality especially for the MOOI is

enhanced by nonuniformly reducing the size of the video frames as a preprocessing for the H.264 video compressions.

This paper is organized as follows. In Section 2, the algorithm for detecting multiple moving objects in video is presented. Then, different linear transforms are applied to MOOI and non-MOOI for size reduction in Section 3. Section 4 shows the experimental results of proposed method. Section 5 concludes this paper.

2. Detection of Multiple Moving Objects of Interest

Moving objects in video can be detected by motion estimation, which is a computationally expensive process. Instead, in this paper, all MOOIs are detected by using spatial and temporal gradients in a GoF of H.264/AVC structure. Specifically, a spatial gradient map $S_g^t(i, j)$ at pixel (i, j) of a frame I^t with a size of $N_r \times N_c$ can be defined as an average of the magnitude of spatial gradients within a window $(2\psi + 1) \times (2\psi + 1)$ as follows:

$$S_g^t(i, j) = \frac{1}{(2\psi + 1)^2} \sum_{u=-\psi}^{\psi} \sum_{v=-\psi}^{\psi} g^t(i + u, j + v), \quad (1)$$

where $g^t = |\partial I^t / \partial i| + |\partial I^t / \partial j|$ is the magnitude of the spatial gradient. Using S_g^t , we define the temporal saliency cost S_{tsc}^t by computing the temporal gradient of the spatial gradients between the two consecutive even numbered frames (even numbered frames give more temporal deviations with less computational load):

$$S_{\text{tsc}}^t = |S_g^t - S_g^{t+2}|. \quad (2)$$

As a result of the spatiotemporal gradients in (1) and (2), the boundary pixels of the moving objects are highlighted in the temporal saliency cost S_{tsc}^t . Then, we can construct a STGM R^k by mosaicking the maximum saliency cost S_{tsc}^t at each pixel location throughout the GoF with even numbered frames starting from the first even numbered frame m_0 to $m_0 + N_0 - 2$ for N_0 frames in the k th GoF of the H.264/AVC as follows:

$$R^k(i, j) = S_{\text{tsc}}^{t^*}(i, j), \quad (3)$$

where $t^* = \arg \max_{t \in \{m_0, m_0 + N_0 - 2\}} S_{\text{tsc}}^t(i, j)$. Note that, according to the definition of $R^k(i, j)$ in (3), the value of $R^k(i, j)$ corresponds to the spatial or temporal boundary of the MOOI in the GoF. Therefore, we can define a bounding box (or window) with $(2w_r + 1) \times (2w_c + 1)$ to encompass the trace of each MOOI, where the weighted sum of $R^k(i, j)$ within the optimal window at the center $C = (C_r, C_c)$ yields the peak value. Therefore, our problem boils down to the determination of the optimal window size (w_r, w_c) and the center of the bounding box $C = (C_r, C_c)$. In this paper, we propose a novel method to find bounding boxes and their centers for multiple MOOIs in a GoF. Our approach to determine (w_r, w_c) and $C = (C_r, C_c)$ is to take an alternate optimization process between (4) and (5). That is, starting

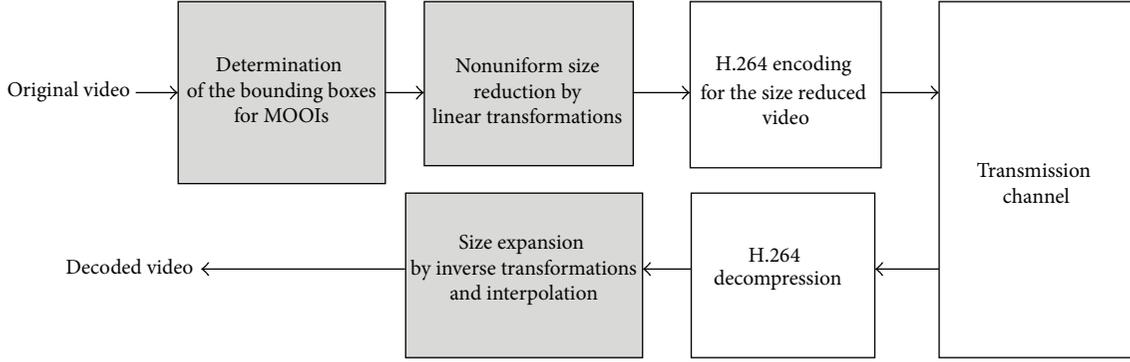


FIGURE 1: Flowchart of the proposed video encoding and decoding system.

from an initial value of $(w_r^{(0)}, w_c^{(0)})$ we apply (4) to have $(C_r^{(1)}, C_c^{(1)})$. Then $(C_r^{(1)}, C_c^{(1)})$ is used to update $(w_r^{(1)}, w_c^{(1)})$ by (5). This alternate process continues until there is no more change from $(w_r^{(n-1)}, w_c^{(n-1)})$ to $(w_r^{(n)}, w_c^{(n)})$. Consider

$$(C_r^{(n)}, C_c^{(n)}) = \arg \max_{1 \leq i \leq N_r, 1 \leq j \leq N_c} \left\{ \sum_{u=-w_r^{(n-1)}}^{w_r^{(n-1)}} \sum_{v=-w_c^{(n-1)}}^{w_c^{(n-1)}} R^k(i+u, j+v) \right\}, \quad (4)$$

$$(w_r^{(n)}, w_c^{(n)}) = \arg \max_{\substack{w_{\min} \leq w_r \leq w_{\max} \\ w_{\min} \leq w_c \leq w_{\max}}} \{ B_{w_r, w_c}^{\text{edge}} \chi(B_{w_r, w_c}^{\text{non-edge}} < T_l) \}, \quad (5)$$

where

$$B_{w_r, w_c}^{\text{edge}} = \sum_{u=-w_r}^{w_r} \sum_{v=-w_c}^{w_c} \chi(R^k(C_r^{(n)} + u, C_c^{(n)} + v) > T_g),$$

$$B_{w_r, w_c}^{\text{non-edge}} = \sum_{u=-w_r}^{w_r} \sum_{v=-w_c}^{w_c} \chi(R^k(C_r^{(n)} + u, C_c^{(n)} + v) < T_g) \quad (6)$$

and χ is an indicator function such that

$$\chi(\varphi) = \begin{cases} 1, & \text{if } \varphi \text{ is True} \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

T_g and T_l are predetermined thresholds. Note that, given the current window size $(w_r^{(n-1)}, w_c^{(n-1)})$, we find the center of the bounding box $(C_r^{(n)}, C_c^{(n)})$ for all pixels in the image by (4). Then, given the current center $(C_r^{(n)}, C_c^{(n)})$ of the bounding box, we examine all possible window sizes to find the maximum number of strong edges within the window (i.e., $B_{w_r, w_c}^{\text{edge}}$ in (5)) under the condition that the number of weak edges is less than a threshold T_l (i.e., $B_{w_r, w_c}^{\text{non-edge}}$ in (5)). Figure 2 shows the convergence of our alternate optimization process of (4) and (5) for $w_{\min} = 16$ and $w_{\max} = 128$, where our method converges after about the 5th iteration. Note that the center and the size of the bounding box improve after every iterative step. This tells that the center and size of the

bounding box are updated cooperatively, which eventually leads to the final convergence. Because our method is based on a binary decision by the threshold, the computation for each iteration is very simple and the convergence is very fast.

For the case of multiple MOOIs in the GoF, after the first MOOI (denoted as MOOI-1) and its bounding box are defined, the bounding box for the next MOOI (i.e., MOOI-2) can be found by repeating the alternating optimization process of (4) and (5). This time, in order not to detect the already-found bounding box again we set the pixel values of the STGM under the predetermined bounding boxes to zeroes before we search for the next MOOI (i.e., MOOI-2). This search process for the next MOOIs and their optimal bounding boxes continues until the sum of all pixel values of the STGM under the bounding box is less than a threshold T_s . After all moving objects and their bounding boxes in the GOF are found, we have P multiple MOOIs from MOOI-1 to MOOI- P .

3. Linear Transformations for Nonuniform Size Reduction

After all MOOIs and their bounding boxes in each GOF are determined, we can reduce the sizes for MOOIs and non-MOOIs nonuniformly. In this paper, to treat the inside and outside regions of the bounding boxes differently and to speed up the size reduction process, linear transformations with different slopes for MOOIs and non-MOOIs are applied. So, to squeeze the original frame of $N_r \times N_c$ to $M_r \times M_c$ ($M_r < N_r$ and $M_c < N_c$) we first apply 1D linear transformations to reduce the number of rows from N_r to M_r . Subsequently, the number of columns is reduced from N_c to M_c to have the squeezed frame of $M_r \times M_c$. Since these sequential 1D reductions for the rows and columns are similar, we describe only the row reduction in this section.

For each row n_r ($1 \leq n_r \leq N_r$) we need a linear mapping function $S_{\text{row}}[n_r]$ to convert the original row index n_r to $S_{\text{row}}[n_r]$ ($1 \leq S_{\text{row}}[n_r] \leq M_r$). Depending on the existence of the MOOI at n_r or not, the slope of the linear function $S_{\text{row}}[n_r]$ takes either α_r for the MOOI or β_r for the non-MOOI (see Figure 3). So, the slopes control the amount of the size reduction between the MOOI and

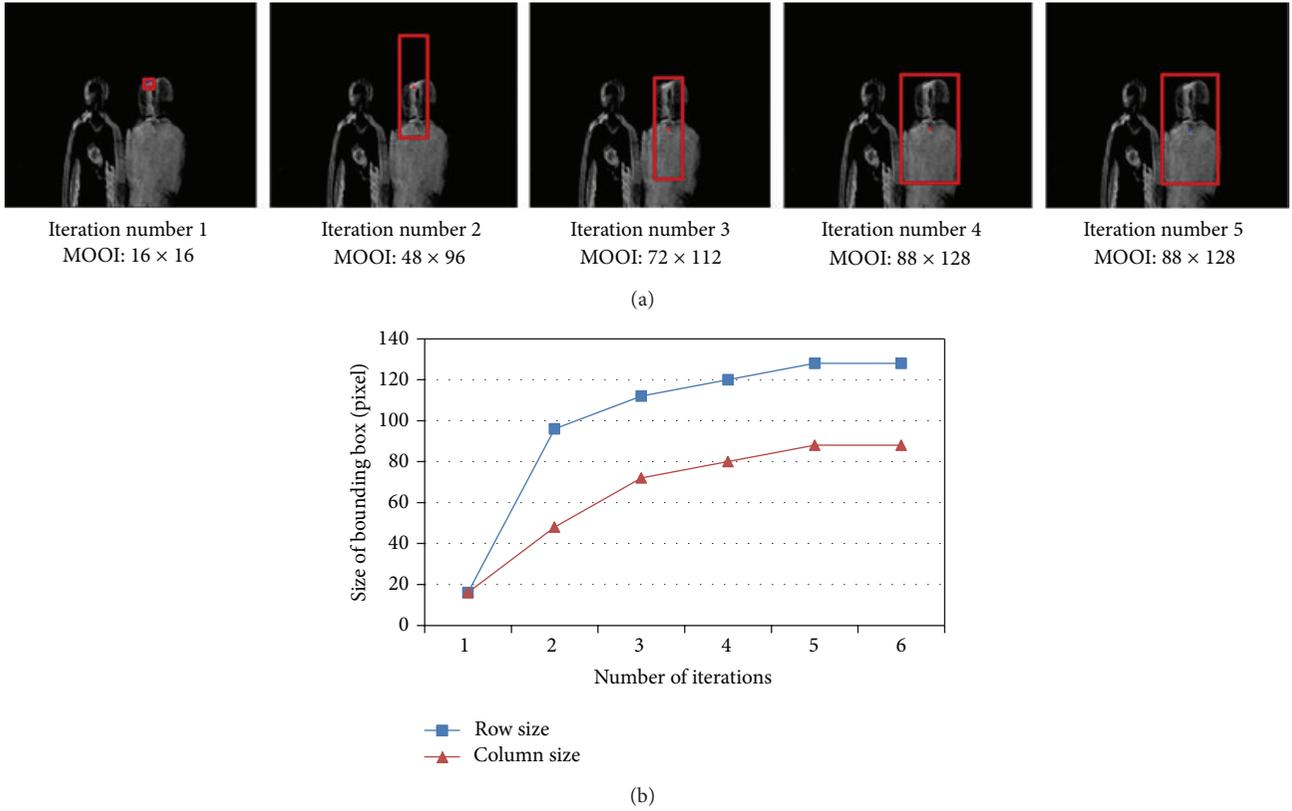


FIGURE 2: Convergence of the alternating optimization process: (a) detected bounding boxes for different iteration numbers and (b) convergence of the alternating optimization method with respect to the size and the number of iterations.

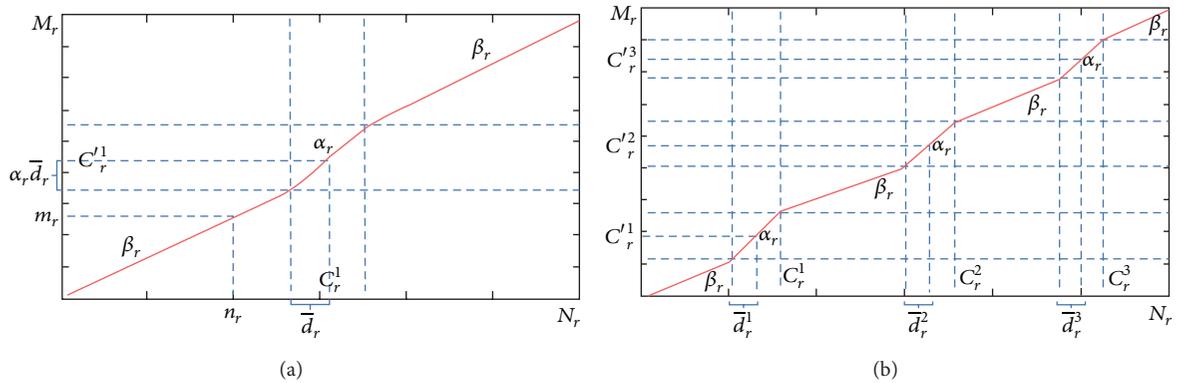


FIGURE 3: Linear transformation for (a) single MOOI and (b) three MOOIs.

the non-MOOI and we have $\beta_r < \alpha_r \leq 1$. Specifically, for the p th MOOI (i.e., MOOI- p), $1 \leq p \leq P$, we denote $d^p(n_r) = |C_r^p - n_r|$ as the absolute distance from the center of MOOI- p , C_r^p , to the row index n_r in the original frame and \bar{d}_r^p represents one-half of the vertical size of the MOOI- p . Also, C_r^p denotes the row index of the center of the MOOI- p at the reduced frame. Note that the index p of the MOOI- p is assigned sequentially from the left to the right of the image space and we start the linear mapping with MOOI-1 by the

following linear transformation $S_{\text{row}}[n_r]$ for each row n_r in $1 \leq n_r < C_r^1 + \bar{d}_r^1$:

$$S_{\text{row}}[n_r] = \begin{cases} C_r^1 - \alpha_r \bar{d}_r^1 - \beta_r (d^1(n_r) - \bar{d}_r^1) & \text{if } 0 < n_r < C_r^1 - \bar{d}_r^1, \\ C_r^1 - \alpha_r d^1(n_r) & \text{if } C_r^1 - \bar{d}_r^1 \leq n_r < C_r^1, \\ C_r^1 + \alpha_r d^1(n_r) & \text{if } C_r^1 \leq n_r < C_r^1 + \bar{d}_r^1. \end{cases} \quad (8)$$

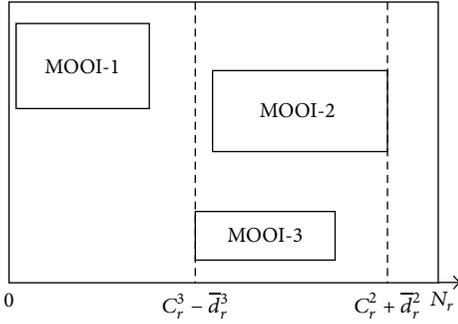


FIGURE 4: Illustration of the case when the two bounding boxes are overlapped horizontally.

Then, for the next rows in $C_r^p + \bar{d}_r^p \leq n_r \leq C_r^{p+1} - \bar{d}_r^{p+1}$ with $p \geq 1$ we have the following mapping function $S_{\text{row}}[n_r]$:

$$S_{\text{row}}[n_r] = \begin{cases} C_r^{lp} - \alpha_r \bar{d}_r^p + C_r^{lp-1} + \alpha_r \bar{d}_r^{p-1} + \beta_r (d^p(n_r) - \bar{d}_r^p) & \text{if } C_r^{p-1} + \bar{d}_r^{p-1} \leq n_r < C_r^p - \bar{d}_r^p, \\ C_r^{lp} - \alpha_r d^p(n_r) & \text{if } C_r^p - \bar{d}_r^p \leq n_r < C_r^p, \\ C_r^{lp} + \alpha_r d^p(n_r) & \text{if } C_r^p \leq n_r < C_r^p + \bar{d}_r^p, \\ C_r^{lp} + \alpha_r \bar{d}_r^p + \beta_r (d^p(n_r) - \bar{d}_r^p) & \text{if } C_r^p + \bar{d}_r^p \leq n_r < C_r^{p+1} - \bar{d}_r^{p+1}. \end{cases} \quad (9)$$

Finally, for the rows in the last MOOI- $PC_r^p - \bar{d}_r^p \leq n_r \leq N - 1$ we have the mapping function $S_{\text{row}}[n_r]$ as follows:

$$S_{\text{row}}[n_r] = \begin{cases} C_r^{lp} - \alpha_r d^p(n_r) & \text{if } C_r^p - \bar{d}_r^p \leq n_r < C_r^p, \\ C_r^{lp} + \alpha_r d^p(n_r) & \text{if } C_r^p \leq n_r < C_r^p + \bar{d}_r^p, \\ C_r^{lp} + \alpha_r \bar{d}_r^p + \beta_r (d^p(n_r) - \bar{d}_r^p) & \text{if } C_r^p + \bar{d}_r^p \leq n_r \leq N - 1. \end{cases} \quad (10)$$

Given the reduction rate α_r for the MOOIs, the reduction rate for the non-MOOI β_r should be determined by considering the overall reduction ratio from N_r to M_r , as well as α_r . So, for a single MOOI, β_r is given by

$$\beta_r = \frac{M_r - 1 - 2\bar{d}_r^1 \alpha_r}{N_r - 1 - 2\bar{d}_r^1}. \quad (11)$$

For multiple MOOIs β_r is calculated as follows:

$$\beta_r = \frac{M_r - 1 - 2\bar{d}_r^1 \alpha_r - \sum_{l=2}^p 2\bar{d}_r^l \alpha_r}{C_r^1 - \bar{d}_r^1 + N_r - 1 - C_r^p - \bar{d}_r^p + \sum_{l=2}^p C_r^l - \bar{d}_r^l - C_r^{l-1} - \bar{d}_r^{l-1}}. \quad (12)$$

The index of the center of the MOOI is also changed from C_r^p to C_r^{lp} after the reduction. Specifically, the indices for the first MOOI and the next MOOIs are given as the following equations, respectively:

$$C_r^{l1} = \beta_r (C_r^1 - \bar{d}_r^1) + \alpha_r \bar{d}_r^1, \quad (13)$$

$$C_r^{lp} = C_r^{lp-1} + \alpha_r \bar{d}_r^1 + \beta_r (C_r^p - \bar{d}_r^p - C_r^{p-1} - \bar{d}_r^{p-1}) + \alpha_r \bar{d}_r^p. \quad (14)$$

In practice, bounding boxes of MOOIs can be overlapped horizontally and/or vertically. Specifically, we define that bounding boxes are overlapped when their 1D projections are overlapped. This is because 1D linear transformations are applied to rows and columns separately. Figure 4 shows the example when the bounding boxes are overlapped horizontally. To deal with the overlapped bounding boxes, the boundaries of MOOI-2 and MOOI-3 are merged; that is, new left and right boundaries and center are determined as $C_r^3 - \bar{d}_r^3$, $C_r^2 + \bar{d}_r^2$, and $(C_r^2 + C_r^3 + \bar{d}_r^2 - \bar{d}_r^3)/2$, respectively. The row reduction is then performed using (8)–(14) for MOOI-1 and the merged MOOI. The column reduction is performed in a similar manner.

Our goal of the linear transformations is to keep the original image data in the MOOIs as much as possible after the size reduction, while achieving the major size reduction in the non-MOOI. Therefore, we first set $\alpha_r \approx 1$ and adjust β_r ($\beta_r < \alpha_r$) to meet the requirement of the size reduction. After the transformations of (8)–(10) the integer valued indices at the reduced rows are determined by the interpolation from the actual mapped indices. After the row reduction, the transformation-interpolation process is applied to the columns to complete the size reduction.

After the size reduction, the conventional H.264/AVC is used to further compress the size-reduced frames and the compressed bit stream is sent to the receiver. At the receiver, after decoding compressed bit-stream, the decompressed frames are expanded to the original size by the inverse transformation-interpolation for the columns and the rows sequentially. The sizes of bounding boxes, their centers, and the size reduction rate of MOOIs are sent to the decoder as side information for the size expansion. Note that we can use the same GoF boundaries as those from the H.264/AVC.

4. Experimental Results

Our experiments have been conducted to demonstrate two aspects: (i) accuracy of the proposed MOOI detection with



FIGURE 5: Comparison for the detection of boundary box: (a) Kalman filtering object detection and tracking [11] (b) the proposed method.

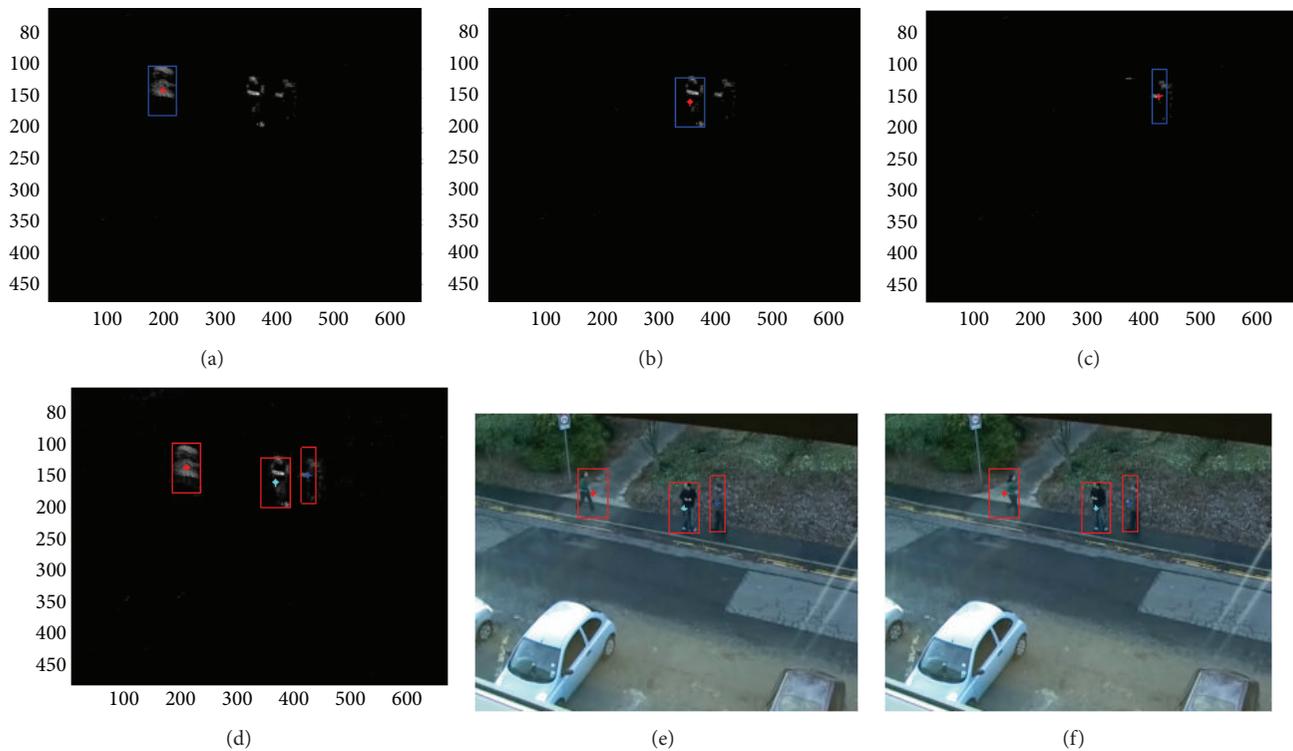


FIGURE 6: Proposed bounding box detection for multiple moving objects in a GOF: (a) first MOOI (MOOI-1), (b) second MOOI (MOOI-2), (c) third MOOI (MOOI-3), (d) bounding boxes for all MOOIs in the STGM, (e) bounding boxes for the first frame of the GoF, and (f) bounding boxes for the last frame of the GoF.

the bounding box and (ii) usefulness of the proposed MOOI detection. The accuracy of the proposed MOOI detection with the bounding box is judged by visual comparisons with the previous inter-frame based Kalman filtering approach [11]. The usefulness of the proposed MOOI detections is demonstrated by applying the detected MOOI to the content-aware image resizing with the comparison of the LMSE method [18] and to the image size reduction as a preprocessing for the H.264/AVC compressions.

The surveillance video sequences [21, 22] were used to evaluate the performance of the proposed method. In all our

experiments, the parameters are predetermined and fixed as follows: $\psi = 3$, $w_{\min} = 16$, $w_{\max} = 128$, $T_l = 0.8(2w_{\max} + 1)(2w_{\max} + 1)$, $T_g = 0.09$, $T_s = (2w_{\min} + 1)(2w_{\min} + 1)/2$. The threshold parameters affect the accuracy of the bounding box detection. The users can interact with the system by adjusting these parameter values. For comparisons, the proposed method and the LMSE method in Vo et al. [18] were applied to reduce the size of the video frames before we apply H.264/AVC. Then, the visual qualities after the H.264/AVC decompression and the size expansion are compared.



FIGURE 7: Size reduction by 30% with a single MOOI: (a) original, (b) LMSE [18], and (c) the proposed method.



FIGURE 8: Size reduction by 30% with three MOOIs: (a) original, (b) reduced frame by LMSE [18], and (c) the proposed method.

The proposed bounding box detection method is also compared to the moving object detecting and tracking method [11]. As shown in Figure 5, the Kalman filtering method [11] tends to detect only the moving part between the consecutive frames not the whole body of the moving object, which demonstrates the power of our STGM-based formulation of the cost function for a GoF. For the case of multiple MOOIs, Figure 6 shows the order of MOOI detection from the first MOOI (i.e., MOOI-1 in Figure 6(a)) at the leftmost side of the image to the last MOOI (i.e., MOOI-3 in Figure 6(c)) at the rightmost side of the image in a GoF. This demonstrates the extension of our previous work [20] to the problem of multiple MOOIs. Since our bounding box is determined on the basis of the GoF, the first bounding box of the walking person includes all pixels along the motion trajectories from the first frame (Figure 6(e)) to the last one (Figure 6(f)) of the GoF.

Once the MOOIs are detected with the bounding boxes, we can differentiate the image regions of the moving objects from the rest of the image regions of nonmoving objects. This allows us to treat MOOIs and non-MOOI separately for image size reduction. That is, we can nonuniformly reduce the size of the image frames in the video before compressions such that the non-MOOIs are the major target for the size reduction. Frame reductions by 30% and bitrates of 50–200 kbps were tested for the visual comparisons of the MOOIs after the decompressions and size expansions. Figure 7 and Figure 8 show the results of the size reduction by the LMSE in Vo et al. [18] and our method for a single MOOI and

three MOOIs, respectively. As shown in the figures, the moving objects are almost intact after the size reduction by the proposed method. Figure 9 for the Stair sequence in the database [21] demonstrates the differences more clearly. As one can see, the proposed method outperforms the LMSE inside regions of the MOOI in terms of PSNR and visual quality. Figure 10 compares the numerical results by the rate-distortion graphs. Although our method yields PSNR slightly lower than the LMSE for the whole image, inside the MOOI regions, it achieves 3~4 dB higher PSNRs than the LMSE and the H.264/AVC compressions without the size-reduction at bitrates lower than critical bitrate of 150 kbps.

5. Conclusions

Optimal bounding box detection method for the moving object of interest (MOOI) has been proposed. Multiple MOOIs as well as a single one can be automatically detected by the proposed method. Once the bounding boxes are identified, one can treat the MOOI and the non-MOOI differently to preserve the visual quality of the important MOOIs. Linear transformations with different slopes are used to nonuniformly reduce the sizes of the MOOIs and non-MOOI. Our size reduction method can be applied as an initial compression for the H.264/AVC video compression standard. Experimental results show that the decompressed videos of the H.264/AVC using the proposed method yield better PSNRs for the MOOI about 3 dB higher than the LMSE.



FIGURE 9: 300th frame of Stair sequence: size reduction by 30% and bitrate 100 kbps (MOOI inside the bounding box): (a) original, (b) H.264/AVC without the size reduction, (c) the LMSE [18], and (d) the proposed method.

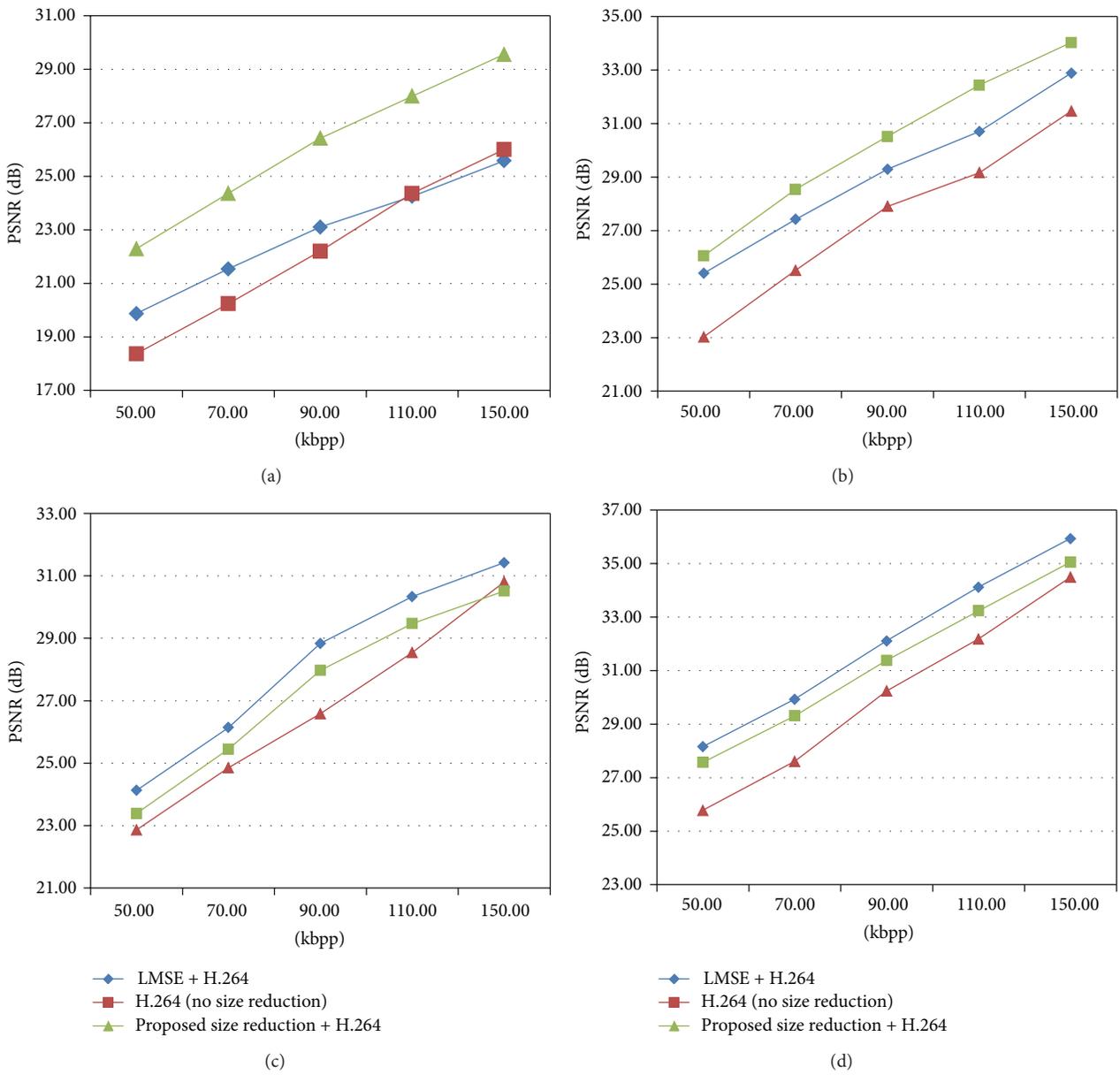


FIGURE 10: Rate and distortion graphs: (a) Stair sequence for the MOOI only, (b) Hallway sequence for the MOOI only, (c) Stair sequence for the whole image, and (d) Hallway for the whole image.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Republic of Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2014-H0301-14-4007) supervised by the NIPA (National IT Industry Promotion Agency).

References

- [1] S.-M. Chang, H. H. Chang, S. H. Yen, and T. K. Shih, "Panoramic human structure maintenance based on invariant features of video frames," *Human-Centric Computing and Information Sciences*, vol. 3, no. 1, pp. 1–18, 2013.
- [2] N. Howard and E. Cambria, "Intention awareness: improving upon situation awareness in human-centric environments," *Human-centric Computing and Information Sciences*, vol. 3, no. 1, pp. 1–17, 2013.
- [3] C. Shahabi, S. H. Kim, L. Nocera et al., "Janus—multi source event detection and collection system for effective surveillance of criminal activity," *Journal of Information Processing Systems*, vol. 10, no. 1, pp. 1–22, 2014.
- [4] S. K. Vipparthi and S. K. Nagar, "Color directional local quinary patterns for content based indexing and retrieval," *Human-Centric Computing and Information Sciences*, vol. 4, no. 1, article 6, 2014.
- [5] P. B. Patil and M. B. Kokare, "Interactive semantic image retrieval," *Journal of Information Processing Systems*, vol. 9, no. 3, pp. 349–364, 2013.
- [6] D. Venkatraman and A. Makur, "A compressive sensing approach to object-based surveillance video coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 3513–3516, April 2009.
- [7] S. Kim, B.-J. Lee, J.-W. Jeong, and M.-J. Lee, "Multi-object tracking coprocessor for multi-channel embedded DVR systems," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 4, pp. 1366–1374, 2012.
- [8] T. J. Kim, B. G. Kim, C. S. Park, and K. S. Jang, "Efficient block mode determination algorithm using adaptive search direction information for scalable video coding (SVC)," *Journal of Convergence*, vol. 5, no. 1, 2014.
- [9] X. Yang, G. Peng, Z. Cai, and K. Zeng, "Occluded and low resolution face detection with hierarchical deformable model," *Journal of Convergence*, vol. 4, no. 1, pp. 11–14, 2013.
- [10] D. Ghimire and J. Lee, "A robust face detection method based on skin color and edges," *Journal of Information Processing Systems*, vol. 9, no. 1, pp. 141–156, 2013.
- [11] S.-K. Weng, C.-M. Kuo, and S.-K. Tu, "Video object tracking using adaptive Kalman filter," *Journal of Visual Communication and Image Representation*, vol. 17, no. 6, pp. 1190–1208, 2006.
- [12] K. Goswami, G. S. Hong, and B. Kim, "A novel mesh-based moving object detection technique in video sequence," *Journal of Convergence*, vol. 4, no. 1, pp. 20–24, 2013.
- [13] H. T. Nguyen and C. S. Won, "Video retargeting based on group of frames," *Journal of Electronic Imaging*, vol. 22, no. 2, Article ID 023023, 2013.
- [14] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," in *Proceedings of the ACM SIGGRAPH*, 2007.
- [15] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," in *Proceedings of the 2008 ACM SIGGRAPH*, 2008.
- [16] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Discontinuous seam-carving for video retargeting," in *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 569–576, San Francisco, Calif, USA, June 2010.
- [17] J. Kiess, B. Guthier, S. Kopf, and W. Effelsberg, "SeamCrop: changing the size and aspect ratio of videos," in *Proceedings of the ACM 4th Workshop on Mobile Video (MoVid '12)*, pp. 13–18, February 2012.
- [18] D. T. Vo, J. Sole, P. Yin, C. Gomilaan, and T. Q. Nguyen, "Selective data pruning-based compression using high-order edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 399–409, 2010.
- [19] C. S. Won and S. Shirani, "Size-controllable region-of-interest in scalable image representation," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1273–1280, 2011.
- [20] A. V. Le and C. S. Won, "Bounding box and frame resizing for moving object of interest," in *Future Information Technology*, vol. 309 of *Lecture Notes in Electrical Engineering*, pp. 445–450, Springer, Berlin, Germany, 2014.
- [21] H. Sohn, W. de Neve, and Y. M. Ro, "Privacy protection in video surveillance systems: analysis of Subband-adaptive scrambling in JPEG XR," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 170–177, 2011.
- [22] S. J. Blunsden and R. B. Fisher, "The BEHAVE video dataset: ground truthed video for multi-person behavior classification," *Annals of the BMVA*, vol. 4, pp. 1–12, 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

