https://doi.org/10.2352/ISSN.2470-1173.2017.9.IRIACV-267

Unsupervised Video Segmentation and Its Application to Region-based Local Contrast Enhancement

Sungbum Park [a], Woosung Shim [a], Yong Seok Heo [b]
[a] Samsung Electronics, Suwon, Korea
[b] Ajou University, Suwon, Korea

Abstract

In this paper, we propose a simple but efficient video segmentation scheme for real-time video applications. First, we temporally separate video frames into scenes, comparing the chi-square distance between consecutive frames. To partition each frame into disjoint regions, then, a pixel-wise color clustering scheme is employed, which is based on K-means clustering and EM algorithm. Finally, we regularize computational complexity to apply the proposed scheme into embedded video processing system. Due to pixel-wise video segmentation with very low complexity, the proposed scheme yields a realistic framework for real-time video applications.

1. Introduction

Recently, much interest has been made on the quality of pixels than only increasing image resolution in video system. Regarding the pixel quality, for example, the UHD alliance announced the high dynamic range (HDR) and the wide color gamut as examples of the next-generation visual experience [1].

For enhanced pixel quality, contrast enhancement (CE) could provide HDR experience in legacy contents [2, 3]. For example, scene contrast was globally enhanced by luminance histogram analysis in each frame [2]. Also, local contrast was adaptively improved by referring neighbor pixels from the target pixel [3]. When video is partitioned into several regions and contrast is adaptively tuned in each region, more dynamic enhancement could be expected.

In [4, 5], a Gaussian mixture model (GMM) fitting has presented unsupervised segmentation of image [4, 5]. Specifically, a combined approach between K-means initialization and expectation-maximization (EM) method accelerates the convergence of GMMs faster than EM only case [5].

In this work, therefore, we propose a GMM-based video segmentation scheme for real-time video analysis and applications. Inspired by the recent work on color clustering based image segmentation work in [5], we partition each video frame into several regions with similar color statistics, shown in Figure 1. Specifically, for temporal consistency of region boundary, an input video is separated into disjoint scenes by measuring chi-square distance between consecutive video frames [6]. Then, each frame in the same scene is segmented by the same GMMs like [4, 5]. As computation complexity in both k-means clustering and EM method is challenging due to iterations, we regularize the complexity by reducing the processing domain size in both k-means and EM iterations.

In experiment, we apply the proposed scheme to various video contents. Also, the proposed video segmentation scheme is employed to CE application, yielding visual better performance. Therefore, a complexity-regularized video segmentation scheme is proposed for real-time video applications in h/w-constrained system.

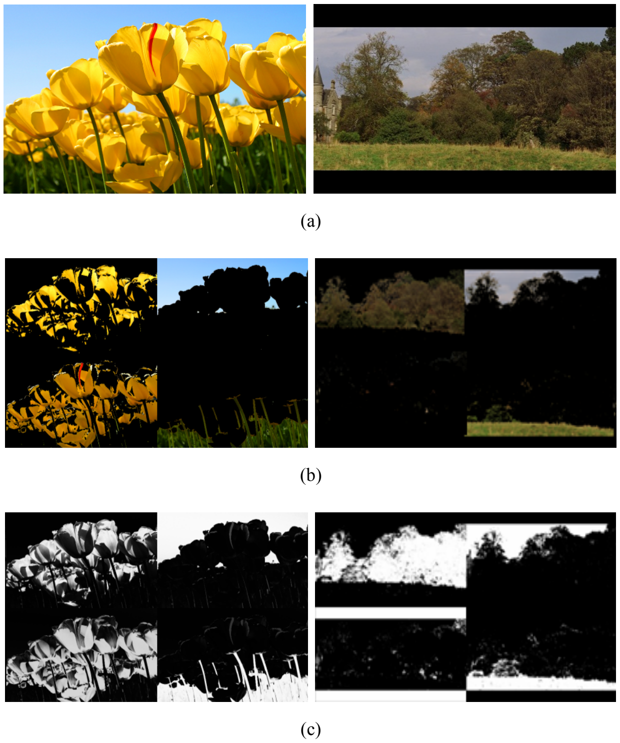


Figure 1. Video segmentation example using the proposed scheme: (a) Original image, (b) region map, (c) probability map. The input frames are partitioned into regions (region map) using the GMM fitting. Then, the probability map reveals the reliability of each pixel mapped onto the specific regions (bright value means high reliability).

2. Unsupervised Video Segmentation

Fig. 2 demonstrates the overall block diagram of the proposed video segmentation scheme. Like [5], a GMM fitting is employed combining K-means and EM method. The GMM fitting yields a region map and a probability map for each image. Then, a scene change detection scheme is presented for scene partitioning. In video frames with the same scene, therefore, the GMM is not updated, yielding temporally consistent segmentation result.

2.1 GMM-based unsupervised segmentation

A GMM fitting is widely used in unsupervised learning problem. Specifically, it has been widely used for color clustering in image segmentation [4, 5]. In this work, we extend a GMM scheme into the proposed video segmentation framework to partition frame into regions with similar color statistics.

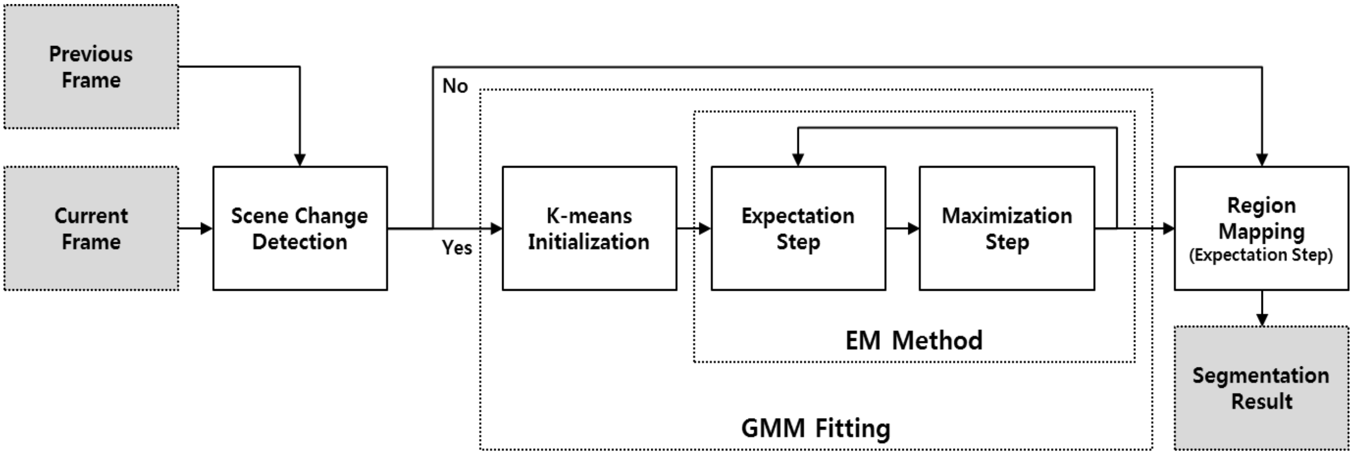


Figure 2. Block diagram of the proposed video segmentation scheme. In each image, a GMM fitting yields a region map and a probability map, respectively. For temporally consistent video segmentation, a scene change detection algorithm where the Gaussian models are identical in the same scene frames.

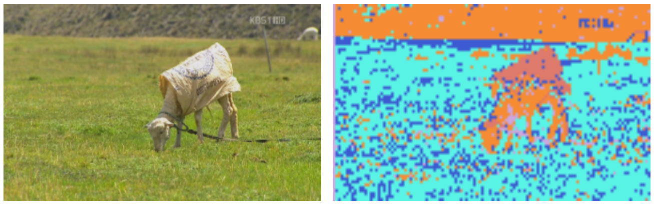

(a) (b)

Figure 3. K-means clustering example: (a) Input image and (b) color clustering result using K-means algorithm.

An expectation-maximization (EM) method is applied to fitting a mixture of Gaussians. As convergence speed of a EM method is slow, however, a K-means clustering is employed for the initial setup of the EM method in this work [5]. For example, the K-means clustering example is illustrated in Figure 3. Specifically, the initial value of mean color μ^i and the corresponding covariance matrix Σ^i for i-th GMM, $i = \{1, \ldots, K\}$, are evaluated using the K-means clustering.

In the EM algorithm, given the initial μ^i and Σ^i, the expectation step updates the probability $P^i(p)$ for each pixel p to the i-th segment, evaluated as

$$P^i(p) = \frac{w^i exp\left\{-0.5\left(I(p)-\mu^i\right)^T \left(\Sigma^i\right)^{-1} \left(I(p)-\mu^i\right)\right\}}{(2\pi)^{\frac{3}{2}} \left|\Sigma^i\right|^{\frac{1}{2}}}, \quad (1)$$

where $I(p)$ is a RGB value vector for p and w^i is the weight for i-th segment. Note that in the first iteration, all the weight values are initialized as $w^i = Z_i / Z_{tot}$, where Z_{tot} denotes the total number of pixels and Z_i is the number of pixels that are assigned to the i-th segment after the K-means clustering.

Then, the maximization step updates μ^i and Σ^i using $P^i(p)$, defined as

$$\mu^i = \frac{\Sigma P^i(p) I(p)}{\Sigma P^i(p)}, \quad (2)$$

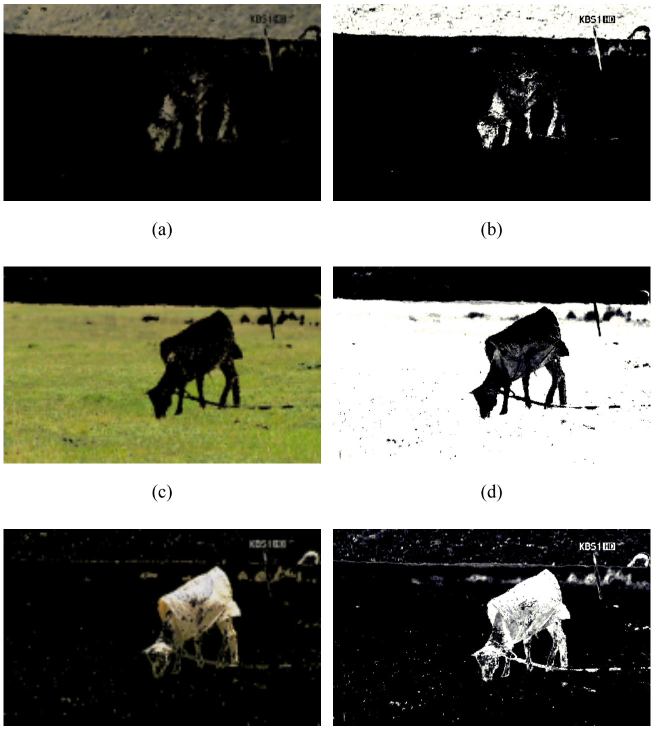

(a) (b) (c) (d) (e) (f)

Figure 4. Clustering using a EM method: (a) First region map, (b) probability map of (a), (c) second region map, (d) probability map of (c), (e) third region map, and (f) probability map of (e). In this example, input image is clustered into three regions.

$$\Sigma^i = \frac{\Sigma P^i(p)\left(I(p)-\mu^i\right)\left(I(p)-\mu^i\right)^T}{\Sigma P^i(p)}, \quad (3)$$

$$w^i = \frac{\Sigma P^i(p)}{Z_{tot}}, \quad (4)$$

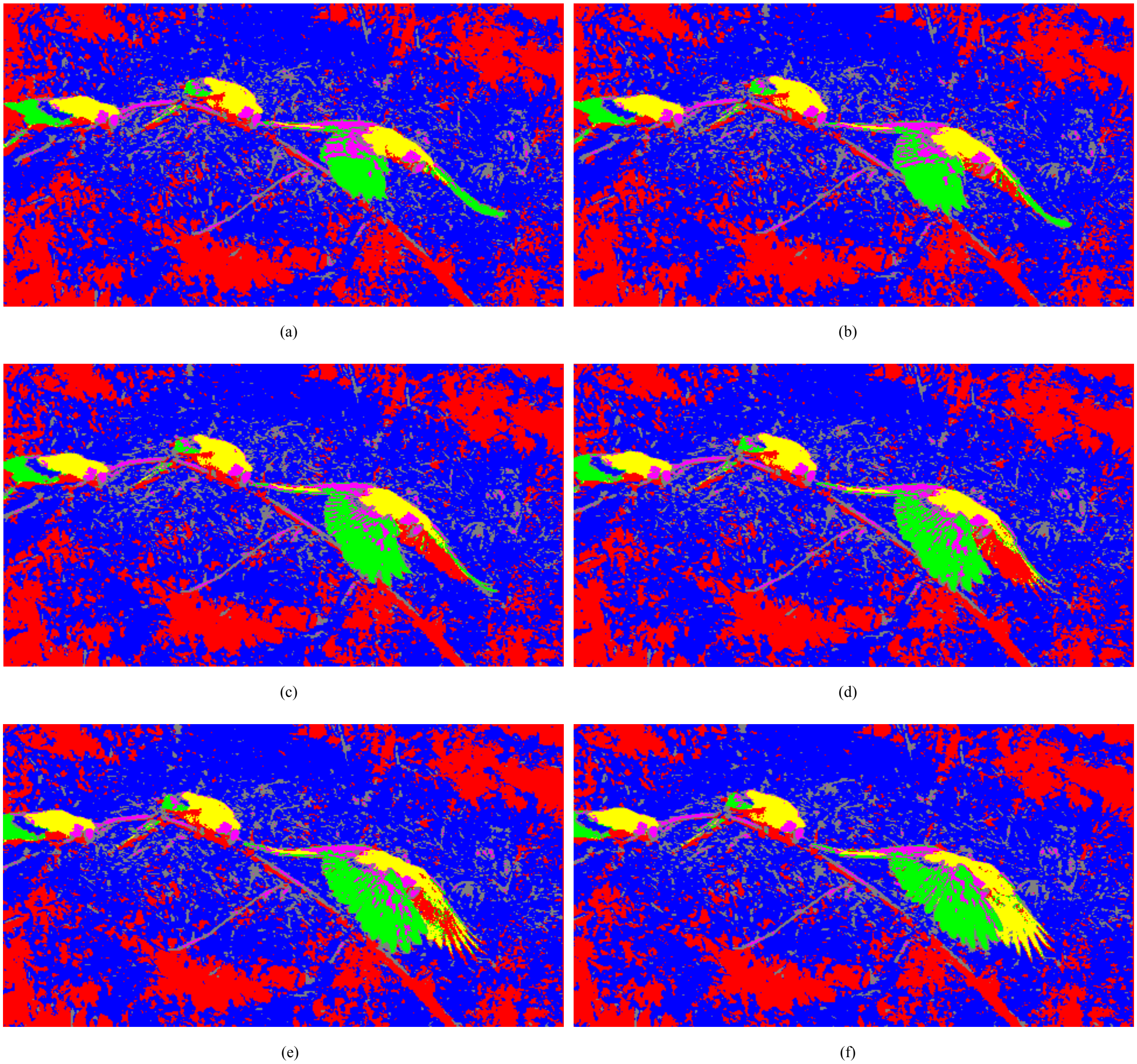


Figure 5. Segmentation result on consecutive frames in the same scene: (a) 1st frame, (b) 3rd frame, (c) 5th frame, (d) 7th frame, (e) 9th frame, and (f) 11th frame. Note that every odd frame is demonstrated to show motion of birds.

where Z_{tot} is the total number of pixels in the image.

Both steps iterate until convergence. At the final region mapping stage in Figure 2, the final expectation step reveals the region map and the corresponding probability map, shown in Figure 4. Also, we perform an $O(1)$ bilateral filtering on P^i to enforce spatial smoothness in the region map [7], shown in Figure 4-(a), (c), and (e).

2.2 Temporally consistent video segmentation

For temporal consistency of segmentation results on consecutive frames, GMM parameters are only evaluated for key frames in input video. Specifically, a frame is assigned as a key frame when the current frame is significantly different from the previous frame.

In this work, the inter-frame difference is measured using the histogram chi-square distribution approach [6]. The chi-square metric between the current frame I^t and the previous frame I^{t-1} is defined as

$$C(h^t, h^{t-1}) = \sum_{i=1}^{255} \frac{\left(h_i^{t-1} - h_i^t\right)^2}{h_i^{t-1} + h_i^t}, \qquad (5)$$

where h_i^{t-1} and h_i^t are the i-th bin value of the normalized luminance histogram for I^{t-1} and I^t, respectively. The scene change is detected if $C(h^t, h^{t-1})$ is greater than the threshold θ, set as $\theta = 0.005$ in this work.

In Figure 5, for example, region maps are shown for consecutive video frames in the same scene. Due to same GMMs in the scene, each region keeps temporal consistency. Specifically, there is no flickering on the background regions (blue and red). Also, in foreground objects, the boundary of the moving object is consistent through motion changes.

3. Complexity Regularization for Real-time Video Applications

In the proposed scheme, the initial K-means clustering and the EM method cause much complexity due to large number of iterations. To alleviate the computational loads, the initial K-means clustering and the EM iterations are processed on the reduced domain down-sampled from the original frame size.

Figure 6 summarizes the complexity regularization scheme in this work. Specifically, the scene change detection compares only luminance histogram differences between two consecutive frames. Then, both K-means initialization and EM method work on the reduced domain (120×67 resolution in this work) sampled from the original domain ($1{,}920\times1{,}080$ FHD resolution). In final expectation stage for the region mapping and the probability mapping, the original domain is utilized for fine resolution segmentation result.

4. Video Segmentation and Application to Contrast Enhancement

First, the proposed video segmentation scheme is evaluated on various video contents. As shown in Figure 7, each region is effectively separated from the original video frames (6 GMMs in this work). While each region contains pixels with similar colors, however, it does not extract explicit object boundaries, which is the limitation of unsupervised video segmentation.

Then, for real-time video application, the proposed video segmentation scheme is coupled to video contrast enhancement (CE). The locally enhanced image $\tilde{I}(p)$ is defined as

$$\tilde{I}(p) = \sum_{k=1}^{6} P^k(p) CE^k(I(p)), \qquad (6)$$

where $CE^k(\cdot)$ is a local CE function for k-th region and $P^k(p)$ is the probability value of the pixel p at the k-th region.

In Figure 8, the CE result is demonstrated. Compared with the global CE result in Figure 8-(c), the adaptive CE result in Figure 8-(d) shows more dynamic picture quality due to local processing using the proposed video segmentation scheme.

5. Conclusion

In this paper, we presented a simple but efficient video segmentation system for real-time video applications. The pixel-wise region segmentation partitioned pixels with similar colors using a GMM fitting. Then, we employed a chi-square test for scene change detection, and the GMMs are unchanged in the same video scene for temporal consistent segmentation result. Finally, the complexity regularization in the proposed GMM-based video segmentation scheme enabled a real-time implementation in h/w-constrained environment. Therefore, we believe that the proposed system is a promising framework for real-time video analysis applications in embedded video processing system.

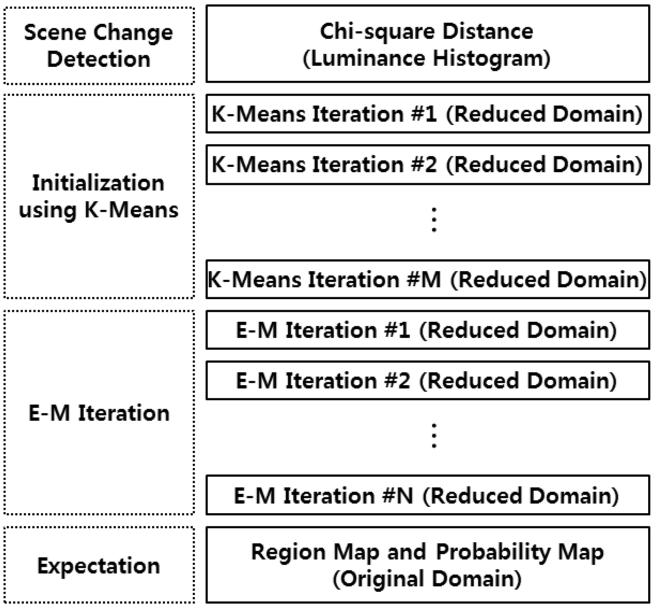

Scene Change Detection
Chi-square Distance (Luminance Histogram)
Initialization using K-Means
K-Means Iteration #1 (Reduced Domain)
K-Means Iteration #2 (Reduced Domain)
⋮
K-Means Iteration #M (Reduced Domain)
E-M Iteration
E-M Iteration #1 (Reduced Domain)
E-M Iteration #2 (Reduced Domain)
⋮
E-M Iteration #N (Reduced Domain)
Expectation
Region Map and Probability Map (Original Domain)

Figure 6. Block diagram of the proposed video segmentation scheme in implementation. Note that full processing time is under single frame duration for real-time implementation (33 ms in 60 Hz video input, for example).

References

[1] UHD Alliance, http://www.uhdalliance.org/.

[2] Wang Qing and R. K. Ward, "Fast image/video contrast enhancement based on weighted thresholded histogram equalization," IEEE Trans. Consumer Electronics, vol. 53, no. 2, pp. 757-764, May 2007.

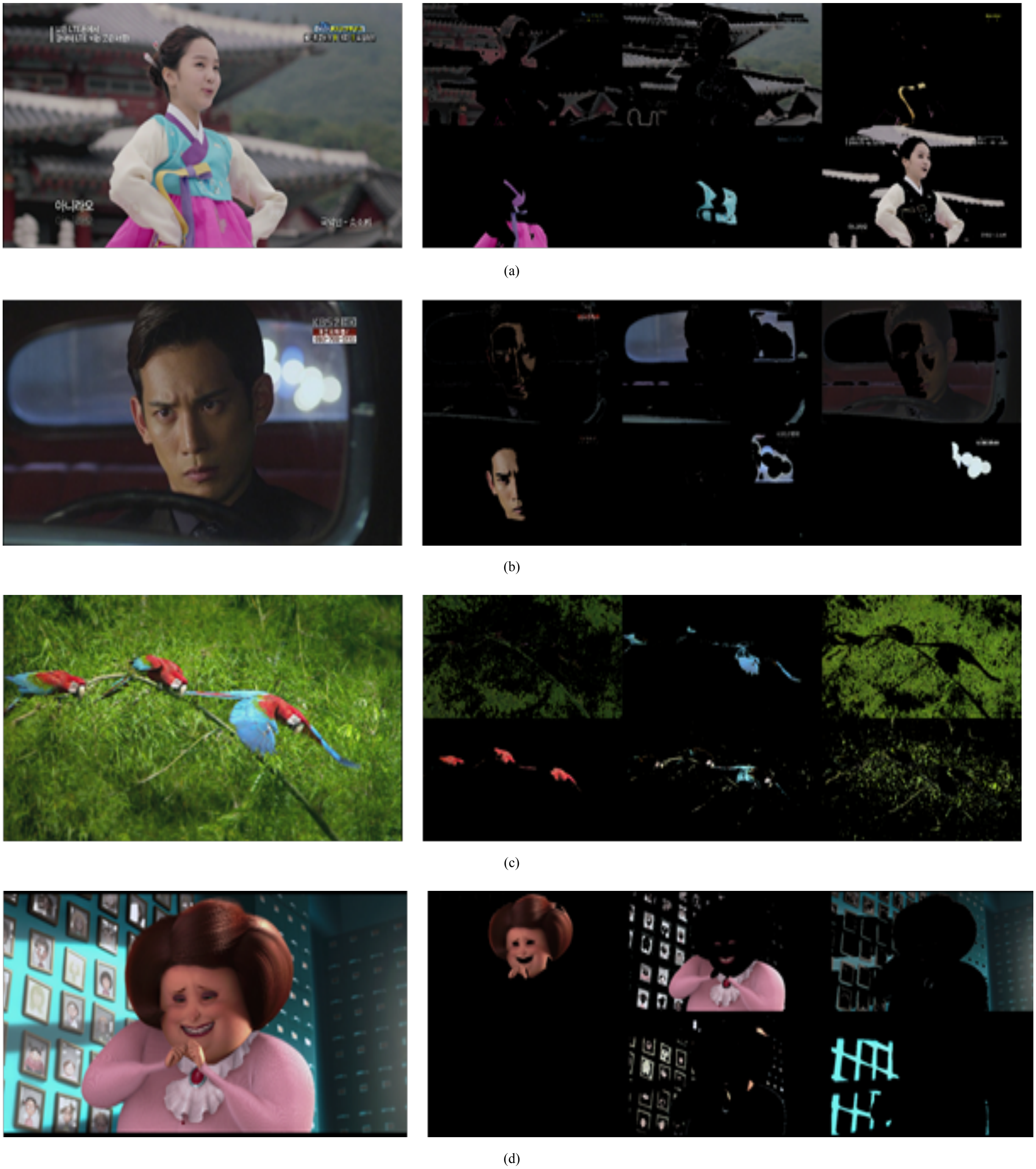
[3] J.-Y. Kim, L.-S. Kim, S.-H. Hwang, "An advanced contrast enhancement using partially overlapped sub-block histogram equalization," IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no. 4, pp. 475-484, Apr. 2001.

[4] S. Belongie, C. Carson, H. Greenspan, and J. Malik, "Color- and texture-based image segmentation using EM and its application to content-based image retrieval," in Proc. IEEE ICCV, Jan. 1998, pp. 675-682.

[5] Y.-W. Tai, J. Jia, and C.-K. Tang, "Soft color segmentation and its applications," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 9, pp. 1520-1537, Sept. 2007.

[6] U. Gargi, R. Kasturi, and S. H. Strayer, "Performance characterization of video-shot-change detection methods," IEEE Trans. on Circuits and Systems for Video Technology, vol. 10, no. 1, pp. 1-13, Feb. 2000.

[7] Q. Yang, K.H. Tan, and N. Ahura, “Real-time O(1) bilateral filtering,” In Proc. IEEE Conference on CVPR, 2009.



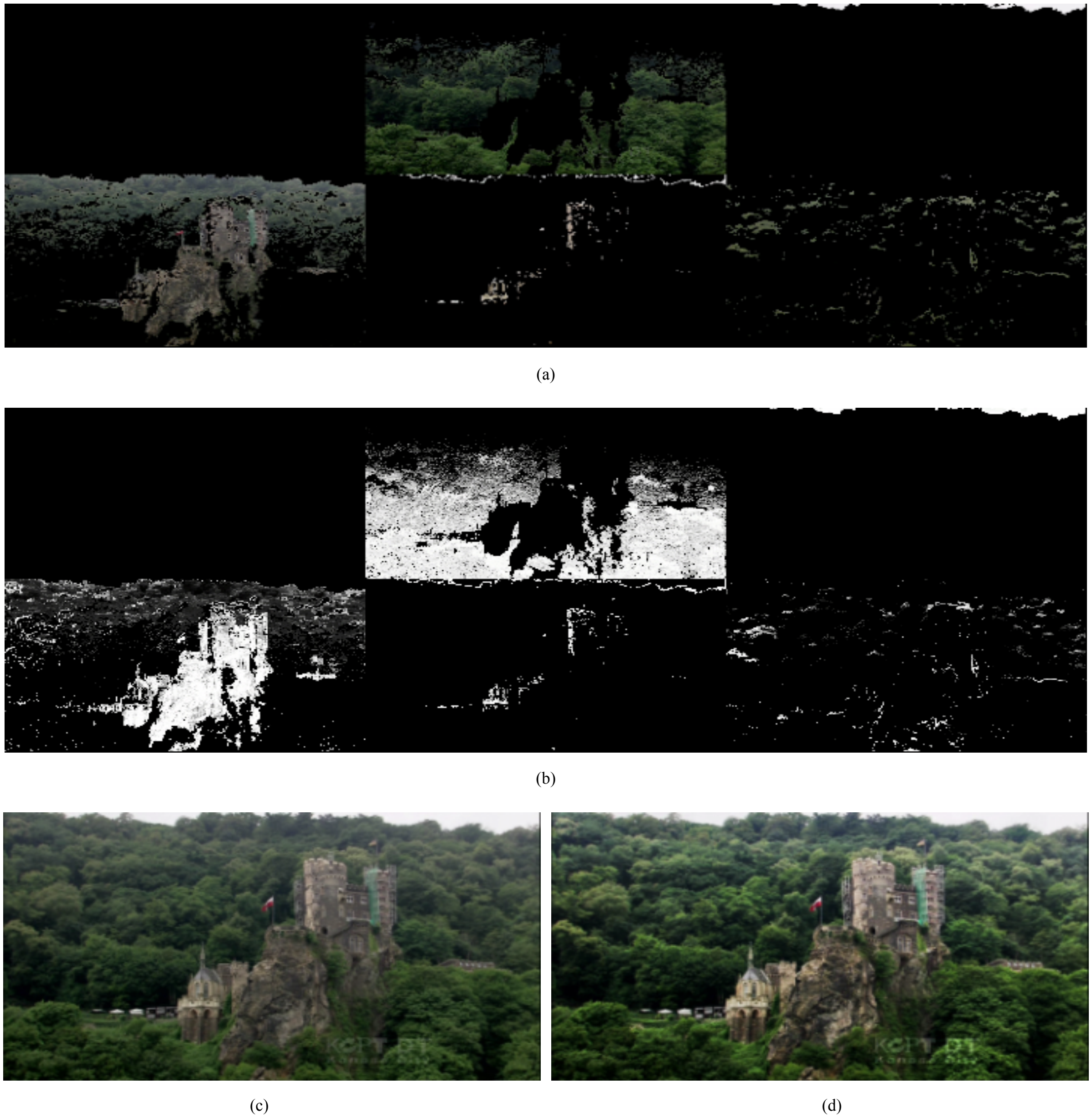
(a)

(b)

(c)

(d)

Figure 7. Unsupervised video segmentation results on various video contents: (a) Girl, (b) Man, (c) Bird, and (d) Animation.



(a)

(b)

(c)

(d)

Figure 8. Contrast enhancement using the proposed pixel-wise region segmentation scheme: (a) Region map, (b) corresponding probability map for (a), (c) global contrast enhancement, and (d) region-based local contrast enhancement.

Author Biography

Sungbum Park received the BS degree in Electrical Engineering in 2000, from Yonsei University, Korea, and the MS and the Ph.D. degrees in Electrical Engineering and Computer Science in 2002 and 2007, respectively, from Seoul National University, Korea. Since 2007, he has joined in Samsung Electronics. During 2014 and 2015, he was a visiting scholar for Berkeley Vision and Learning Center, UC Berkeley. Currently, in Samsung, he is leading a computer vision and machine learning group, covering unsupervised video segmentation, semantic video segmentation, human/facial object detection and segmentation, and efficient machine learning framework for visual scene understanding.

Woosung Shim received the M.S. degree and Ph.D. degree in electronic engineering from Wonkwang University, Korea, in 1996 and 2000, respectively. Since 2000, he has been working for Samsung electronics. His research interests include image processing and computer vision.

Yong Seok Heo received the BS degree in Electrical Engineering in 2005, and the MS and the Ph.D. degrees in Electrical Engineering and Computer Science in 2007 and 2012, respectively, from Seoul National University, Korea. During 2012–2014, he was with Samsung Electronics, in the Digital Media and Communications R&D Center. Currently, he is with the Department of Electrical and Computer Engineering at Ajou University as an assistant professor. His research interests include segmentation, stereo matching, 3D reconstruction, and computational photography.