Fast Extraction of Objects of Interest from Images with Low Depth of Field

Changick Kim, Jungwoo Park, Jaeho Lee, and Jenq-Neng Hwang

In this paper, we propose a novel unsupervised video object extraction algorithm for individual images or image sequences with low depth of field (DOF). Low DOF is a popular photographic technique which enables the representation of the photographer's intention by giving a clear focus only on an object of interest (OOI). We first describe a fast and efficient scheme for extracting OOIs from individual low-DOF images and then extend it to deal with image sequences with low DOF in the next part. The basic algorithm unfolds into three modules. In the first module, a higher-order statistics map, which represents the spatial distribution of the high-frequency components, is obtained from an input low-DOF image. The second module locates the block-based OOI for further processing. Using the block-based OOI, the final OOI is obtained with pixel-level accuracy. We also present an algorithm to extend the extraction scheme to image sequences with low DOF. The proposed system does not require any user assistance to determine the initial OOI. This is possible due to the use of low-DOF images. The experimental results indicate that the proposed algorithm can serve as an effective tool for applications, such as 2D to 3D and photo-realistic video scene generation.

Keywords: Object of interest, low depth of field, image segmentation, video object extraction, immersive video.

Manuscript received June 26, 2006; revised Feb. 28, 2007.

This research was supported by the MIC, Korea, under the ITRC support program supervised by the IITA (IITA-2006-C1090-0603-0017)

Changick Kim (phone: + 82 42 866 6168, email: ckim@icu.ac.kr) and Jaeho Lee (email: jaeho@icu.ac.kr) are with Visual Information Processing Lab, Information and Communications University, Daejeon, Korea.

Jungwoo Park (email: jwpark@icu.ac.kr) is with Samsung Electronics, Suwon, Korea.

Jenq-Neng Hwang (email: hwang@u.washington.edu) with Department of Electrical Engineering, University of Washington, Seattle, USA.

I. Introduction

Image segmentation is one of the most challenging problems in computer vision. The objective of image segmentation is to partition an image into homogeneous regions, where pixels share the same attributes, such as texture, intensity, color, or even focus cue. In our preceding work [1], we proposed a novel unsupervised segmentation algorithm for images with low depth of field (DOF) (see Fig. 1). Low DOF is an important photographic technique commonly used to assist viewers in understanding the depth information within a two-dimensional photograph [2]. Unlike typical image segmentation methods [3]-[5], in which regions are discovered using properties of the intensity or texture, focus cue may play the most important role for the unsupervised extraction of the focused object-of-interest (OOI). The fact that we can effectively extract the OOI automatically from low-DOF images suggests a variety of applications, such as image indexing for content-based retrieval, object-based image compression, 3D microscopic image analysis, image enhancement for digital cameras, arbitrarily focused image generation from two differently focused images [6], range segmentation for depth estimation, 2D-to-3D conversion for 3D TV [7], [8], improvement of coding efficiency in multi-view coding [9], and fusion of multiple images which are focused to different degrees [10], [11].

Let us define $f(x, y)$ and $b(x, y)$ as the focused foreground and background, respectively. Also, we assume that an OOI lies in the depth of focus of the camera so the entire OOI appears sharp when focused, even if it is not necessarily completely planar. We model the low-DOF images by a linear combination of textures with blur functions as

$$g(x, y) = f(x, y) + h(x, y) * b(x, y), \quad (1)$$

Fig. 1. Low-DOF images.

where * indicates a 2D convolution operation, and h is a space-invariant blur function, which is usually assumed to be the Gaussian function.

As shown in (1), since the defocused region is lowpass filtered, high frequency components in the region are removed or reduced. Thus, by assuming that only sharply focused regions contain adequate high frequency components, it should be possible to distinguish the focused foregrounds from the defocused backgrounds by comparing the amount of the high frequency content.

There have been several approaches to the extraction of OOIs from low-DOF images. The edge-based method proposed in [12] extracts the boundary of the object by measuring the amount of defocus at each edge pixel. The algorithm has demonstrated high accuracy for segmenting man-made objects and objects with clear boundary edges. However, this approach often fails to detect boundary edges of natural objects, yielding disconnected boundaries [2]. Region-based segmentation algorithms [2], [13]-[15] rely on the detection of the high frequency areas in an image. A reasonable starting point is to measure the degree of focus for each pixel by computing the high frequency components. To this end, several methods have been used, such as spatial summation of the squared anti-Gaussian (SSAG) function [13], variance of wavelet coefficients in the high frequency bands [2], a multi-scale statistical description of high frequency wavelet coefficients [14], local variance [15], and so on. Note that exploiting high frequency components alone often results in errors in both focused and defocused regions. In defocused regions, despite blurring due to defocusing, there could be busy texture regions in which high frequency components are still strong enough. These regions are prone to be misclassified as focused regions. Conversely, we may have focused regions with nearly constant gray levels, which may be misclassified as defocused regions. As pointed out in [2], relying only on the sharp detail of the OOI can be a limitation for this region-based DOF image segmentation approach. Furthermore, the multi-scale approaches employed in both [2] and [14] tend to generate jerky boundaries, even though refinement algorithms for high resolution classification can be incorporated. To reduce the above-mentioned deficiencies, Won and others [15] proposed a block-based maximum *a posteriori* (MAP) segmentation algorithm. While it generates smooth boundaries of the segmented object, it tends to incorporate adjacent defocused regions into focused regions. In our preceding work [1], the initial detection of the high frequency areas is conducted by computing higher order statistics (HOS). In order to minimize the above mentioned misclassification due to solely depending on the sharp details, the obtained HOS map is simplified by morphological filtering which is followed by region merging. It is shown that the algorithm outperforms the existing ones when the shapes of the extracted focused regions are compared with those of ground truths in pixel accuracy. However, the use of time-consuming morphological filtering may become an obstacle to the extension of the algorithm to image sequence cases.

In this paper, we propose an efficient and fast extraction scheme. The algorithm consists of two parts: A fast and efficient block-based scheme is proposed in the next section, which is followed by an extension algorithm to deal with low-DOF image sequences as discussed in section III. Experimental results and conclusions follow in section IV and V, respectively.

II. Block-Based OOI Extraction from Low-DOF Images

Since focused regions are assumed to contain high-frequency areas, detecting such areas is the most critical step in ensuring successful extraction. In other words, without a good scheme to detect such areas, the extracted OOIs may not provide accurate boundary information. Finding high frequency areas may need to be supported by other cues to yield better performance. For instance, the cues may include some semantic assumptions such as "focused objects are usually located in the center of images" or "focused objects tend to have brighter colors compared to the defocused background," and so on.

In this section, we consider an unsupervised OOI segmentation algorithm, in which a focused OOI is automatically extracted, based on the HOS characteristics, from a single image with low DOF. The proposed focused OOI extraction algorithm consists of three modules as shown in Fig. 2 as a block diagram. First, the color-based HOS is calculated for every pixel from an input low-DOF image as conducted in prior works [16], [17]. Second, the HOS map is converted into the $m \times n$ blocked HOS. Then a block-based OOI extraction is

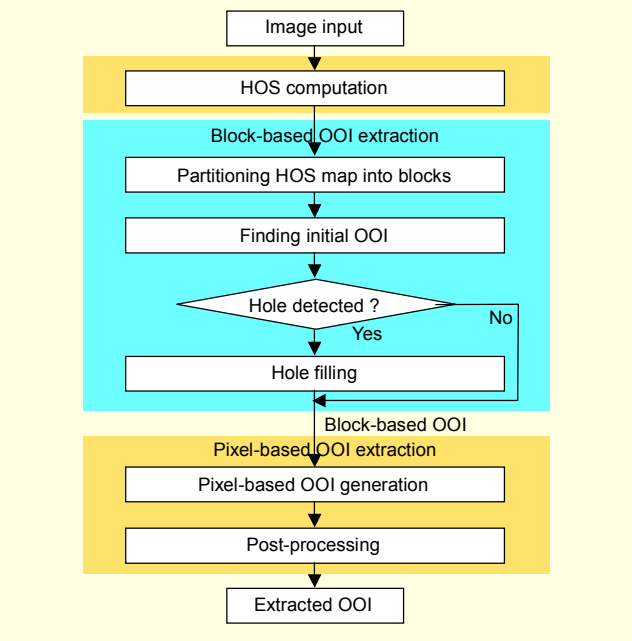

Fig. 2. System block diagram for the first frame.

conducted, which is followed by hole detection and filling techniques. The last stage is a pixel-based OOI extraction method based on applying a filling-in technique to the located block-based OOI.

1. Color-Based HOS Map Construction

The objective of low-DOF image segmentation is to extract the focused OOI from an image. In the first step, we transform the low-DOF image into an appropriate feature space, in which the spatial distribution of the high frequency components is represented. This is conducted by computing the HOS for all pixels in the low-DOF image. The HOS mapping method is known to be well suited to solving detection and classification problems because it can suppress Gaussian noise and preserve some of the non-Gaussian information [1], [16], [18]. The fourth-order moments are calculated for all pixels in the red, green, and blue channels for the $M \times N$ input image, respectively. For instance, the fourth-order moment at (x, y) in a red channel is defined as

$$\hat{m}_R^{(4)}(x,y) = \frac{1}{N_\eta} \sum_{(s,t)\in\eta(x,y)} (I_{\text{red}}(s,t) - \hat{m}_R(x,y))^4,$$

$$\hat{m}_R(x,y) = \frac{1}{N_\eta} \sum_{(s,t)\in\eta(x,y)} I_{\text{red}}(s,t),$$

$$(0 \le x < M, \quad 0 \le y < N), \qquad (2)$$

where $\eta(x, y)$ is a set of neighboring pixels centered at (x, y), $\hat{m}_R(x,y)$ is the sample mean of red channel $I_{red}(x, y)$ of $I(x, y)$, and N_η is the size of η. Only the maximum moment value among all three channels, $\hat{m}_R^{(4)}(x,y), \hat{m}_G^{(4)}(x,y)$, and $\hat{m}_B^{(4)}(x,y)$, at each pixel is used:

$$HOS_{\text{original}}(x,y) = \max(\hat{m}_R^{(4)}(x,y), \hat{m}_G^{(4)}(x,y), \hat{m}_B^{(4)}(x,y)). \quad (3)$$

Since the dynamic range of the HOS values is extremely large, the value for each pixel is down-scaled such that each pixel takes a value from [0, 255]. The outcome is a color-based HOS map. The HOS map value for each pixel (x,y) is thus defined as

$$HOS(x,y) = \min(255, HOS_{\text{original}}(x,y)/DSF), \qquad (4)$$

where DSF denotes the down-scaling factor to reduce the dynamic range, which may range from 0 to about 2.1 billion. For natural images, however, the maximum HOS takes values between 100 and 0.2 billion, and it is observed that using different DSF according to the maximum value yields better performance in the subsequent stages. Thus, the DSF is determined as follows:

$$DSF = \begin{cases} 300 & \text{for } \max(HOS_{\text{original}}(x,y)) < 10^7, \\ 700 & \text{for } \max(HOS_{\text{original}}(x,y)) < 10^8, \\ 1000 & \text{otherwise}. \end{cases}$$

Color-based HOS maps can be effectively used in the event that the foreground/background boundary is obscure due to similar gray levels. Figure 3 illustrates the color-based HOS map generation process. Figure 3(a) shows the input low-DOF image and Figs. 3(b)-(d) show HOS values for each channel. Figure 3(e) shows the HOS map derived from (4), considering all three channels.

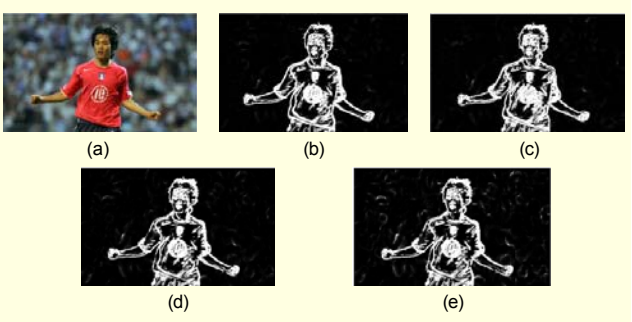

Fig. 3. Computing an HOS map: (a) low-DOF image, (b) HOS map in red channel, (c) HOS map in green channel, (d) HOS map in blue channel, and (e) color-based HOS map.

2. Block-Based OOI extraction

In our prior work [1], [16], the obtained feature space, which is

Table 1. Average processing time for each step in our prior work [1].

384 × 256 image	Processing time (ms)	(%)
HOS map	47	1.0
HOS simplification by morphological filtering by reconstruction	3,537	71.2
Region merging	1,378	27.7
Thresholding	7	0.1
Total	4,969	100.0

called a color-based HOS map, is refined by removing small dark holes and bright patches using a morphological closing-opening by reconstruction [19], that is, morphological closing by reconstruction followed by morphological opening by reconstruction.

However, the use of the filters is responsible for the degradation of system performance due to the required high computational complexity which takes up about 71% of the whole processing time as shown in Table 1. Thus the process needs to be replaced by an efficient and fast scheme for fast OOI segmentation applications dealing with still images or even image sequences in real-time. In this section, a fast block-based OOI extraction scheme is proposed, which is refined by the pixel-based OOI extraction scheme addressed in the following subsection.

A. HOS Characteristics of an Image from the Low DOF

Basically, the HOS calculation in (4) creates much higher values in focused regions than in defocused areas [1]. An original HOS computation from a low-DOF image (see Fig. 4(a)) has an extremely large dynamic range, in which a small number of strong peaks are present as shown in Fig. 4(b).

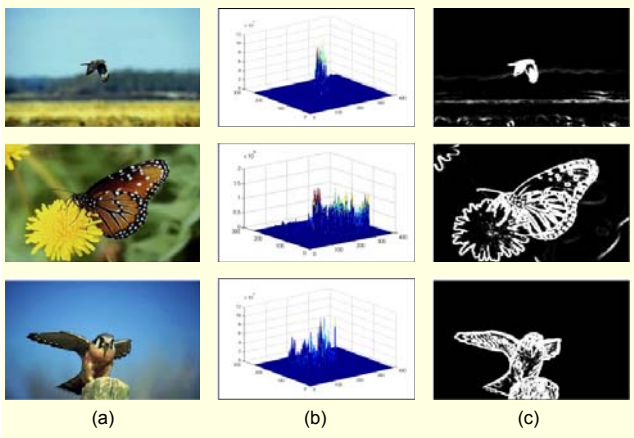

Fig. 4. HOS characteristics: (a) low-DOF image, (b) $HOS_{original}$ computation result, and (c) HOS map.

The HOS map (see Fig. 4(c)) is then used to distinguish focused regions from blurred background [1], [16]. However, as shown in Fig. 4(c), there may be unwanted bright patches in the background, which are removed by morphological filtering [1].

In this paper, we pay additional attention to the maximum (peak) values of $HOS_{original}$. We note that the locations of the maximum value can be exploited as an indication of the location of the OOI since it is very likely to exist in the focused OOI. Thus, it is reasonable to use the maximum HOS location as the starting point of the block-based reconstruction of the HOS map.

B. Block-Based Reconstruction of the HOS Map

In this subsection, we describe how to construct the block-based HOS map by reconstruction without the repetitive use of morphological filters which was used in prior work [1], [16]. First of all, the HOS map is partitioned into blocks with $m \times n$ pixels. The maximum value of each block is determined as

$$HOS_{\text{block}}(u,v) = \max_{\substack{mu \le x < m(u+1) \\ nv \le y < n(v+1)}} [HOS(x,y)], \qquad (5)$$

where the (u,v) range of the HOS_{block} is $0 \le u < M/m, 0 \le v < N/n$.

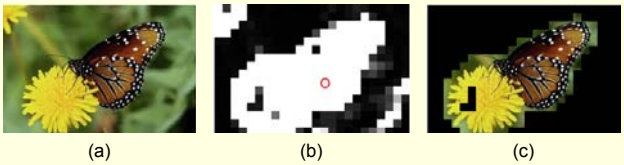

Fig. 5. Block-based HOS map: (a) low-DOF image, (b) blocked HOS in which the red circle denotes the location of the seed point, and (c) region-of-interest (ROI) extraction result from the HOS_{block}.

The pictorial example used to obtain the HOS_{block} is shown in Fig. 5. Focused objects in low-DOF images tend to contain the highest HOS. The coordinate corresponding to the maximum HOS_{block} value is used as the seed-point, as indicated by the red circle in Fig. 5(b), to start the OOI search.

$$(\hat{x}, \hat{y}) = \arg\max_{(x,y)} [HOS_{\text{original}}(x,y)], \qquad (6)$$

$$(seed\_u, seed\_v) = \left((\text{int})\frac{\hat{x}}{m}, (\text{int})\frac{\hat{y}}{n} \right). \qquad (7)$$

The reason we pay attention to the seed-point is that there may be two or more connected clusters of blocks in $HOS_{block}(u,v)$ as shown in Fig. 5(b). In this case, the cluster of

blocks containing the seed-point is used as the initial point for searching the block-based OOI map. To extract blocks connected to this point, a depth-first-based search technique is exploited. Starting at the seed-point, the four neighboring blocks are checked to see if their values are 255. If so, the block is connected and the procedure is continued until no block having the value of 255 is found.

Figure 5(c) shows the extracted connected blocks, which represent the block-based focused region. Since this initial block-based HOS map may contain small holes as shown in Fig. 5(c), a hole-filling procedure should be applied before proceeding to pixel-based OOI extraction.

The hole-filling module checks whether any holes are present in the initial block-based HOS map, and fills any holes that are found. To this end, we adopt the region merging algorithm [1], [16], [17]. More specifically, the hole is filled if it is surrounded by neighboring blocks (marked in red in Fig. 6(a)) belonging to the extracted initial OOI.

Since the task is block-based, the hole-filling process is fairly fast and effective even for larger holes. As shown in Fig. 6(b), the OOI with holes filled is still in a block-based shape, which needs to be refined in the pixel-based OOI extraction module.

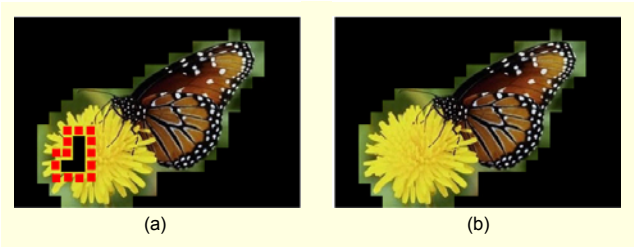

Fig. 6. ROI extraction: (a) initial OOI and (b) hole filled OOI.

3. Pixel-Based OOI Extraction

The block-based OOI is now able to serve as a mask, which confines the final locations of the OOI inside it. Figure 7(b) and (d) show refined HOS maps, with the removed pixels located outside the block-based OOI.

The next step is to fill in the OOI, which results from the smooth areas inside the focused objects. This is done by using the filling-in technique used in our prior work [20]. The results

Fig. 7. Refined HOS map: (a) original HOS maps, (b) refined HOS map with removed pixels located outside the block-based OOI, (c) original HOS maps, and (d) refined HOS map with removed pixels located outside the block-based OOI.

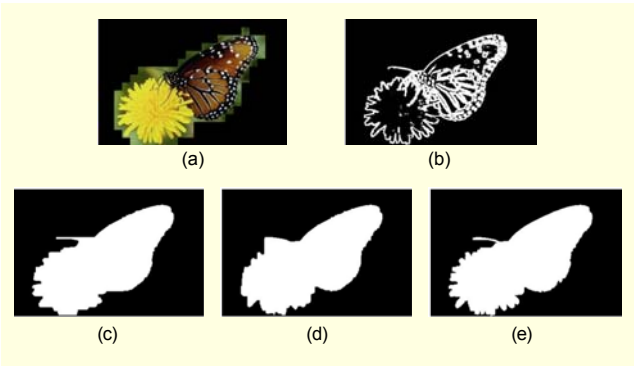

Fig. 8. The focused object-of-interest extraction process: (a) block-based ROI, (b) refined HOS map indicating OOI, (c) horizontally filled OOI, (d) vertically filled OOI, and (e) final OOI obtained by AND operation of (c) and (d).

are shown in Fig. 8(c) and (d).

The final OOI is extracted by a pixel-based AND operation as shown in Fig. 8(c) and (d), resulting in the final OOI, as shown in (e). There may still be some errors such as gulfs or peninsulas created at the boundaries of the object or isolated small false regions even though the segmented object produces the overall shape of the focused OOI. These errors may cause visual annoyance. The goal of the object post-processing step is to remove those errors and at the same time to smooth the object boundaries. Cascaded opening-closing morphological operations can thus be adopted to accomplish this task. The filtered result represents the final outcome of the proposed algorithm.

III. Extension to Low-DOF Image Sequences

In this section, we propose a scheme to extend our approach to deal with image sequences with low-DOF. To deal with low-DOF video sequences, the following issues should be taken into account:

- Input image sequence: The input image sequences should contain focused objects inside each image frame and the focused objects need to be well tracked throughout the sequence. The sequences can be captured using the low-DOF technique and can be used for various applications.
- Efficiency for image sequences: As previously mentioned, the possible applications can go beyond those in the traditional computer vision areas, such as range segmentation for depth estimation [13] and target recognition [21]. For instance, the low-DOF technique assisted video object segmentation could expedite the practical use of content-based interactivity for a variety of multimedia applications. By using this method, for example, a bird flying in the sky can be extracted from a video sequence, which is never feasible using chroma-key (or

blue screen) imaging. In such cases, one of the key factors for success is the reduction of computational complexity. The methods which use the MRF model [13], [15] are not suitable to handle image sequences due to their iterative relaxation process. An efficient deterministic method is critically required.

In order to achieve fast and effective processing of low-DOF image sequences, temporal redundancy needs to be considered. While motion information has been adopted to remove the temporal redundancy between consecutive image frames, we propose using a more efficient scheme to prompt the processing instead of conducting tedious motion estimation.

The initial block-based OOI is obtained by expanding that obtained in the previous frame using.

$$\begin{cases} D_n(x+i, y+j)=1 \text{ for } i, j \in \{-1,0,1\} \text{ if } D_{n-1}(x,y)=1, \\ D_n(x,y)=0 \qquad\qquad\qquad\qquad \text{otherwise,} \end{cases} \tag{8}$$

where $D_n(x,y)$ denotes the value at block position (x, y) and n denotes the current frame. Of course, the extent of expansion may be significant if the movement of the focused object between consecutive frames is somewhat large. The expanded region based on the block-based OOI from the previous frame is shown in Fig. 9(b). Then HOS is then computed on the pixels in the expanded regions and the resulting OOI can thus be derived on that basis for an individual image frame, as discussed in section II. By confining the pixels for HOS computation, the computational complexity and processing time can be greatly reduced.

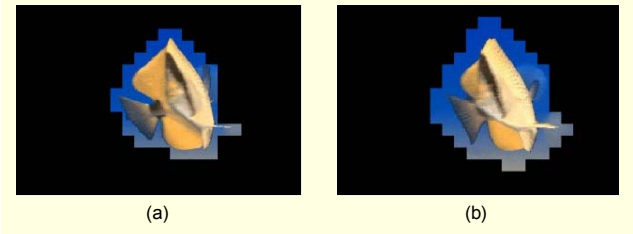

Fig. 9. Obtaining initial OOI: (a) block-based OOI at frame n-1 and (b) OOI expanded at frame n from (a).

IV. Experimental Results

1. Experiments on Low-DOF Images

The proposed algorithm was implemented and tested on low-DOF images. We used a 3×3 neighborhood for η in (2). Figure 10 shows some experimental results of the proposed algorithm. It also gives visual comparisons of outcomes from manual extraction with the algorithm from [1] on several test images. The elapsed processing time of each module executed for a 384×256 image is shown in Table 2. Compared with the speed of the previous algorithm [1] (see Table 1), the proposed scheme is about 19 times faster. This is because the time consuming morphological filtering with a 31×31 structuring element is eliminated with this proposed algorithm.

The performance of the proposed algorithm is also evaluated using an objective criterion. In [2], performance is evaluated in

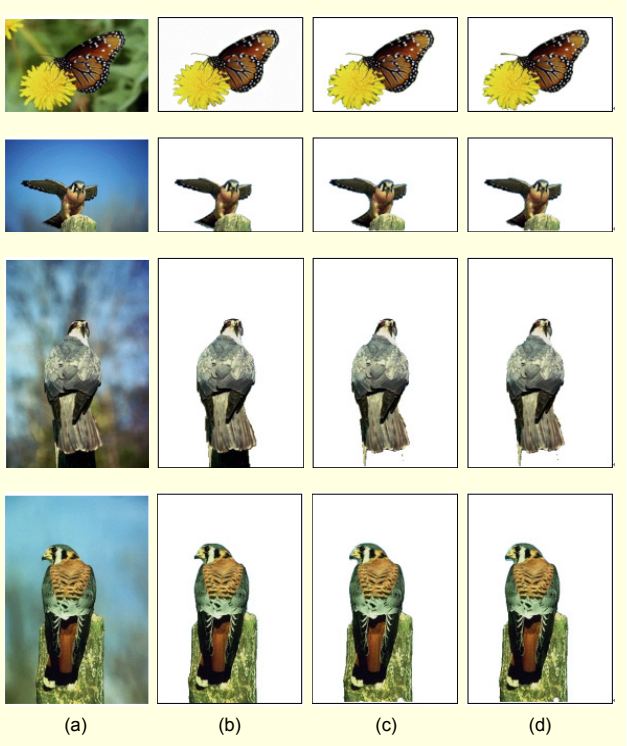

Fig. 10. Extracted OOI comparisons of the results in the prior work [1] and our proposed method: (a) low-DOF image, (b) ground truth from human manual extraction, (c) results from [1], and (d) results from the proposed algorithm.

Table 2. Elapsed processing time of each module experimented for 384×256 image.

Proposed method for 384×256 image	Processing time (ms)	(%)
Color-based HOS map	132	50.5
Block-based OOI segmentation & hole detection	3	0.9
Pixel-based OOI extraction	127	48.6
Total	262	100.0

terms of sensitivity, specificity, and error rate. However, since they are defined as the ratios of the areas, the different shapes of extracted areas can sometimes show high performance as long as the size of the extracted OOI (or background) is close to that of the reference binary mask.

We adopt a pixel-based quality measure [22], which is used to evaluate the performances of video object segmentation algorithms. The spatial distortion (error) of the estimated OOI from the reference OOI is defined as

$$d(O^{\text{est}}, O^{\text{ref}}) = \frac{\sum_{(x,y)} O^{\text{est}}(x,y) \otimes O^{\text{ref}}(x,y)}{\sum_{(x,y)} O^{\text{ref}}(x,y)}, \qquad (10)$$

where O^{est} and O^{ref} are the estimated and reference binary masks, respectively, and $\otimes$ is the binary XOR operation.

Table 3 shows the spatial distortion measures of the results from [1] and the proposed algorithm. Reference OOIs are extracted by manual segmentation, as shown in Fig. 10(b). For the binary XOR operation, the pixels of the OOI are set to one; otherwise, zero. As shown in Table 3, our algorithm yields quite low distortion values which are close to those from [1] and have much lower computational complexity.

Table 3. Performance evaluation via mis-segmentation rate by objective criterion. Images in Fig. 10 are numbered from top to bottom.

Image	Ref. [1]	Proposed
(a)	2.78%	2.79%
(b)	2.04%	1.97%
(c)	3.40%	3.43%
(d)	1.42%	1.35%

2. Experiments on Low-DOF Image Sequences

The proposed fully automatic algorithm has been implemented and tested on the several low-DOF image sequences as well. The MPEG-4 test sequence "Bream" was used with the background blurred in our experiment. The Bream sequence is in CIF format and contains moving objects and background. In our simulations, the OOI is a bream fish swimming around. Its shape changes quite a lot throughout the sequence due to its non-rigid motion.

Unlike the moving object extraction schemes based on object tracking [23]-[25], which require a manual selection of the OOI at the first frame or find the color feature by using a Bayes classifier [26], the proposed system does not require any user assistance to determine the OOI and can operate at fast speed. This is possible due to the use of low-DOF images. Throughout the image sequence, we can obtain the focused OOI without human intervention. The OOI extraction results from the sequence "Bream" with the background blurred are shown in Fig. 11.

Other experimental results are shown in Figs. 12 and 13, where a flying bird and blooming flowers are extracted from natural sequences. Note that the background of the input video is not necessarily Gaussian blurred. As we can clearly see, especially in Fig. 12, sequence tracking of an OOI can be very effective as long as the background appears motion blurred. In the image sequence shown in Fig. 12, only the flying bird is sharply focused. This is a real benefit of extending the proposed algorithm for low-DOF images to image sequences. To deal with sequences with shot changes as shown in Fig. 13, the system needs to be able to detect shot boundaries. We used the algorithm proposed in [27], where pixels are labeled with respect to the evolution of their intensities on several successive frames.

Table 4 shows the segmentation accuracy and spatial distortion rates of the proposed algorithm. For the binary XOR operation, pixels of the OOI are set to one; otherwise, zero. As shown in Table 4, our algorithm is robust regardless of the dynamics of motion in the scene.

The experiments were conducted on an Intel Pentium-IV 3.4 GHz PC. The average processing time for 352×288 sequences was about 0.13 seconds per frame. Table 5 shows the processing time of each step for both intra-frame (initial frame) and inter-frame, that is, the rest of the frames in the sequence. By confining the possible candidate area for inter-frames, even faster processing was attained, without degrading segmentation quality.

V. Conclusion

We developed an efficient and effective algorithm to automatically extract OOIs from individual images or image sequences captured with a low-DOF technique. Extending our prior work [1], a block-based scheme was adopted for faster processing, which is suitable for focused object extraction from image sequences. To track the OOI throughout a sequence, the previous block-based OOI is applied to exploit temporal redundancy between consecutive frames. The extraction is very fast while maintaining accurate object boundaries, even though the degree of speed improvement depends on the size of the focused object in the image.

In our future work we will apply the proposed system to various multimedia applications, such as photo-realistic scene generation, virtual reality, immersive video systems, and so on. Figure 14 shows an example of such applications. Our on-going endeavor is to reach real-time processing on low-end PCs

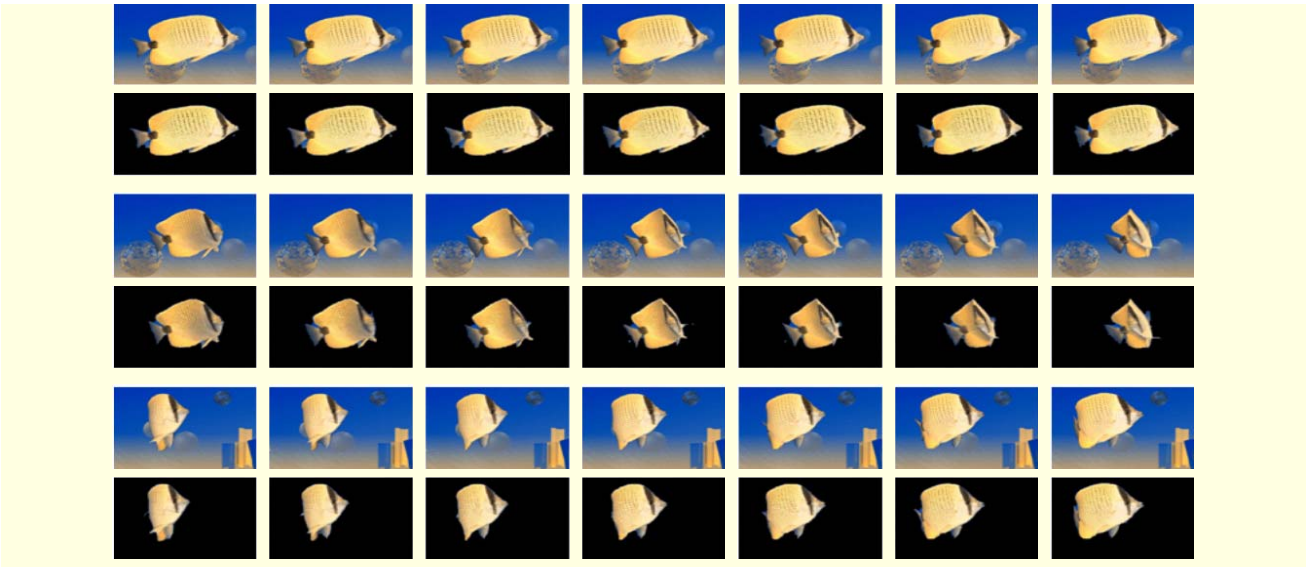

Fig. 11. Video object planes from sequence “Bream” (total 300 frames. 1st row: 52 to 58; 3rd row: 111 to 117; 5th row: 228 to 234).

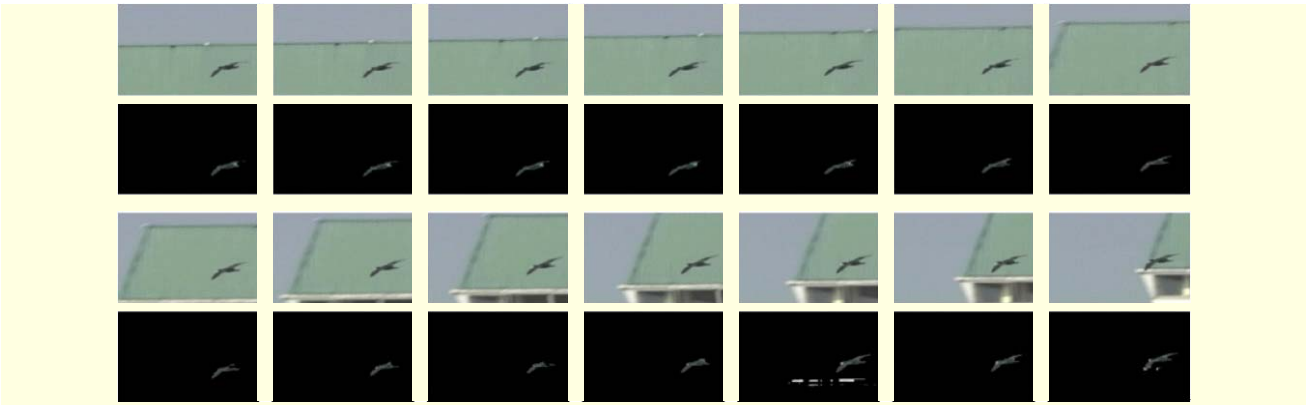

Fig. 12. Extracted OOI from a natural sequence 1 (total 390 frames).

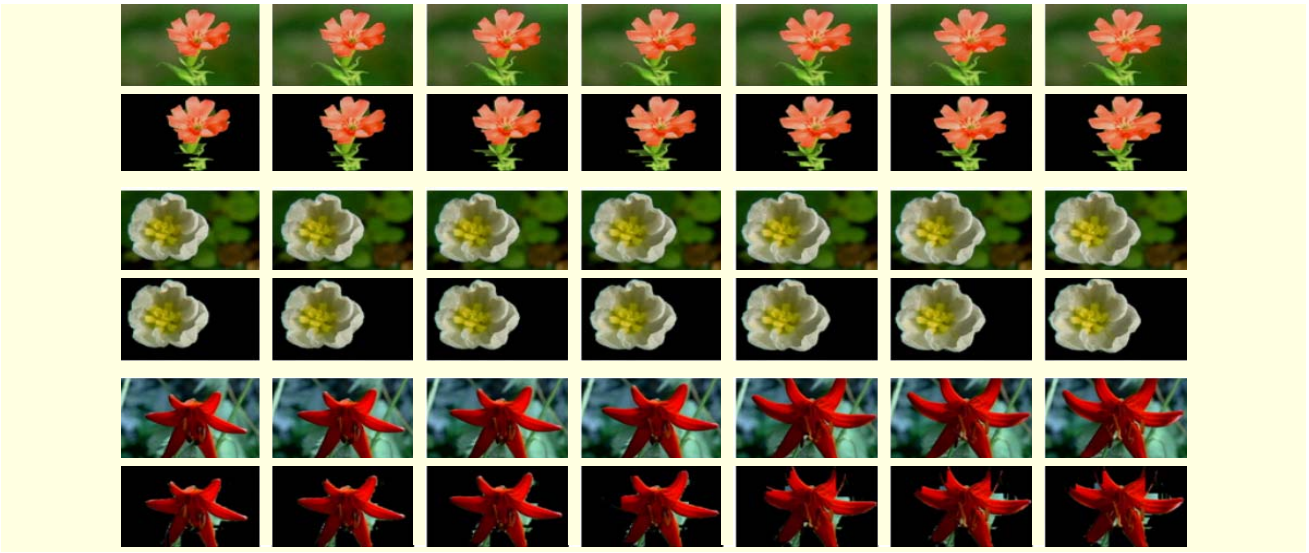

Fig. 13. Extracted OOIs from natural sequence 2 (total 180 frames).

Table 4. Extraction accuracy in "Bream"

Bream frames	Accuracy rate	Error rate (spatial distortion measures)
Steady motion	94.3%	5.7%
Rapid motion	92.2%	7.8%

Table 5. CIF format (Bream) processing time for each step.

Intra-frame	Processing time (ms)	Inter-frame	Processing time (ms)
Color-based HOS map	143	Color-based HOS map	41
Block-based OOI extraction	2	Block-based OOI extraction	1
Pixel-based OOI extraction	83	Pixel-based OOI extraction	89
Total	228	Total	131

Fig. 14. Example of photo-realistic scene generation using low-DOF sequence: (a) OOI in the low-DOF image, (b) natural background, and (c) OOI combined with (b).

through the reduction of processing time by code/algorithm optimization.

References

[1] C. Kim, "Segmenting a Low Depth-of-Field Image Using Morphological Filters and Region Merging," *IEEE Tr. on Image Processing*, vol. 14, no. 10, Oct. 2005, pp. 1503-1511.

[2] J.Z. Wang, J. Li, R.M. Gray, and G. Wiederhold, "Unsupervised Multi-Resolution Segmentation for Images with Low Depth of Field," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no.1, Jan. 2001, pp. 85-90.

[3] P. J. Besl and R.C. Jain, "Segmentation Through Variable –Order Surface Fitting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, Mar. 1988, pp. 167-192.

[4] L. Lucchese and S.K. Mitra, "Color Image Segmentation: A State-of-the-Art Survey: Image Processing, Vision, and Pattern Recognition," *Proc. of the Indian National Science Academy (INSA-A)*, vol. 67A, no. 2, Mar. 2001, pp. 207-221.

[5] D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR'97), San Juan, Puerto Rico, 1997, pp. 750-755.

[6] A. Kubota and K. Aizawa, "Reconstructing Arbitrarily Focused Images from Two Differently Focused Images Using Linear Filters," *IEEE Trans. on Image Processing*, vol. 14, no. 11, Nov. 2005, pp. 1848-1859.

[7] W.J. Tam and L. Zhang, "3D-TV Content Generation: 2D-to-3D Conversion," *IEEE Int'l Conf. Multimedia and Expo* (ICME'2006), Toronto, Canada, July 2006, pp. 1869-1872.

[8] S. Battiato, S. Curti, E. Scordato, M. Tortora, and M. La Cascia, "Depth Map Generation by Image Classification," *SPIE Electronic Imaging*, San Jose, CA, USA, Apr. 2004, pp. 95-104.

[9] X.Wei, M.-Y.Chu, and I. Ahmad, "Lowering the Complexity of Multi-View Video Encoding through Dynamic Segmentation and Registration of Video Object," *Proc. of IEEE Int'l Conf. on Image Processing*, Oct. 2006, pp. 549-552.

[10] K. Aizawa, A. Kubota, and K. Kodama, "Implicit 3D Approach to Image Generation: Object-Based Visual Effects by Linear Processing of Multiple Differently Focused Images," *Proc. 10th Int'l Workshop on Theoretical Foundations of Computer Vision*, Germany, Mar. 2000, pp. 226-237.

[11] H. Li, B.S. Manjunath, and S.K. Mitra, "Multi-Sensor Image Fusion Using the Wavelet Transform," *Proc. Int'l. Conf. Computer Vision*, 1993, pp. 173-182.

[12] D.-M. Tsai and H.-J. Wang, "Segmenting Focused Objects in Complex Visual Images," *Pattern Recognition Letters*, vol. 19, 1998, pp. 929-949.

[13] C. Yim and A.C. Bovik, "Multi-Resolution 3-D Range Segmentation Using Focused Cues," *IEEE Trans. Image Processing*, vol. 7, no. 9, Sep. 1998, pp. 1283-1299.

[14] Z. Ye and C.C. Lu, "Unsupervised Multiscale Focused Objects Detection Using Hidden Markov Tree," *Proc. Int'l Conf. Computer Vision: Pattern Recognition & Image Processing, 2002* (*CVPRIP '02*), Durham, North Carolina, USA, Mar. 2002, pp. 812-815.

[15] C.S. Won, K. Pyun, and R.M. Gray, "Automatic Object Segmentation in Images with Low Depth of Field," *Proc. Int'l. Conf. Image Processing*, vol. 3, Rochester, USA, Sep. 2002, pp. 805-808.

[16] J. Park and C. Kim, "Extracting Focused Object from Low Depth-of-Field Image Sequences," *Proc. SPIE Visual Communications and Image Processing*, vol. 6077, San Jose, Jan. 2006, pp. 607710-1–607710-8.

[17] J. Park and C. Kim, "Performance Improvement of Object-of-Interest Extraction from the Low Depth-of-Field Image Using Color-Based HOS (Higher-Order Statistics)," *Korean Signal Processing Conf. (KSPC'05)*, vol. 18, no. 1, 2005, p. 109.

[18] G. Gelle, M. Colas, and G. Delaunay, "Higher Order Statistics for Detection and Classification of Faulty Fanbelts Using Acoustical Analysis," *Proc. IEEE Signal Processing Workshop on Higher-*

Order Statistics (SPW-HOS '97), Banff, Canada, July 1997, pp. 43-46.

[19] J. Serra, *Image Analysis and Mathematical Morphology*, New York, Academic, 1982.

[20] C. Kim and J. -N. Hwang, "Video Object Extraction for Object-Oriented Applications," *J. VLSI Signal Processing Systems for Signal, and Video Technology, Special Issue on Multimedia Signal Processing*, vol. 29, no. 1/2, Aug. 2001, pp. 7-21.

[21] Y. Boykov and D. Huttenlocher, "A New Baysian Framework for Object Recognition," *Proc. of IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, vol. 2, June 1999, pp. 517-523.

[22] M. Wollborn and R. Mech, "Refined Procedure for Objective Evaluation of Video Generation Algorithms," *Doc. ISO/IEC JTC1/SC29/WG11 M3448*, Mar. 1998.

[23] J. Pan, S. Li, and Y. Zhang, "Automatic Extraction of Moving Object Using Multiple Features and Multiple Frames," *Proc. of IEEE Int'l Symp. on Circuits and Systems*, vol. 1, May 2000, pp. 36-39.

[24] C. Gu and M.C. Lee, "Semiautomatic Segmentation and Tracking of Semantic Video Objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, Sep. 1998, pp. 572-584.

[25] R. Castagno, T. Ebrahimi, and M. Kunt, "Video Segmentation Based on Multiple Features for Interactive Multimedia Applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, Sep. 1998, pp. 562-571.

[26] V. Mezaris, I. Kompatsiaris, and M.G. Strintzis, "Video Object Segmentation Using Bayes-Based Temporal Tracking and Trajectory-Based Region Merging," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 6, June 2004, pp. 782-795.

[27] Y. Taniguchi, A. Akutsu, and Y. Tonomura, "Panorama Excerpts: Extracting and Packing Panoramas for Video Browsing," *ACM Int'l Conf. Multimedia*, Seattle, WA, Nov. 1997, pp. 427-436.

Changick Kim (M'01–SM'89) received the BS degree in electrical engineering from Yonsei University, Seoul, the MS degree in electronics and electrical engineering from Pohang University of Science and Technology (POSTECH), Pohang, Korea, and the PhD degree in electrical engineering from the University of Washington, Seattle, in 1989, 1991, and 2000, respectively. From 2000 to 2005, he was a senior member of Technical Staff at Epson Research and Development, Inc., Palo Alto, CA. Since February 2005, he has been with the School of Engineering, Information and Communications University (ICU), Daejeon, Korea, where he is currently an assistant professor. His research interests include multimedia communications, 3-D video processing, image/video understanding, intelligent media processing, and advanced video coding for IPTV.

Jungwoo Park received the BS degree in information and communication engineering from Sungkyunkwan University, Seoul, in 2003, and the MS degree in the School of Engineering from Information and Communications University (ICU), Daejeon, Korea, in 2006. He currently works for the Corporate Technology Operations in Samsung Electronics, as an assistant engineer. His research interests include image/video understanding, computer graphics, and 3-D simulation for CAE.

Jaeho Lee received the BS degree in electronic, electrical and communication engineering from Pusan National University, Korea, in 2006. He is currently associated with the Visual Information Processing Laboratory at the Information and Communications University (ICU), where he is pursuing the MS degree. His research interests include object segmentation, virtual view generation, and 3D multimedia processing.

Jenq-Neng Hwang received his BS and MS degrees from National Taiwan University, and his PhD from the University of Southern California. In 1989, he joined the EE Department of the University of Washington, where he is currently a professor. He served as the Associate Chair of the EE Department from 2003 to 2005. He has published more than 200 journal papers, conference papers, and book chapters in the areas of multimedia signal processing (MMSP) and networking, neural networks signal processing (NNSP), and machine learning. He received the 1995 Annual Best Paper Award from the IEEE Signal Processing Society (SPS). Dr. Hwang is a fellow of IEEE. He is a founding member of MMSP Technical Committee of IEEE SPS. He served as the Chairman of the NNSP Technical Committee in IEEE SPS, and was the Society's representative to IEEE NNC. He served as an associate editor for IEEE T-SP and T-NN, and is now an associate editor for IEEE T-CSVT and an editor for the Journal of Information Science and Engineering. He was the conference Program Chair of the 1994 IEEE Workshop on NNSP, the General Co-Chair of the 1995 International Symposium on Artificial Neural Networks, the Tutorial-Chair of 1996 IEEE International Conference on Neural Networks (ICNN), the Program Co-Chair of I998 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), and the Chair of the 2006 IASTED conferences on Signal and Image Processing (SIP), as well as Internet Multimedia Systems and Applications (IMSA).