Segmenting a Low-Depth-of-Field Image Using Morphological Filters and Region Merging

Changick Kim, *Senior Member, IEEE*

Abstract—We propose a novel algorithm to partition an image with low depth-of-field (DOF) into focused object-of-interest (OOI) and defocused background. The proposed algorithm unfolds into three steps. 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 higher order statistics (HOS) for all pixels in the low-DOF image. Next, the obtained feature space, which is called HOS map in this paper, is simplified by removing small dark holes and bright patches using a *morphological filter by reconstruction*. Finally, the OOI is extracted by applying region merging to the simplified image and by thresholding. Unlike the previous methods that rely on sharp details of OOI only, the proposed algorithm complements the limitation of them by using morphological filters, which also allows perfect preservation of the contour information. Compared with the previous methods, the proposed method yields more accurate segmentation results, supporting faster processing.

Index Terms—Image segmentation, low depth-of-field (DOF), morphological filter, object of interest (OOI).

I. INTRODUCTION

AUTOMATIC image segmentation [1] is one of the most challenging problems in computer vision. The objective of image segmentation is to partition an image into homogeneous regions. In this paper, we describe a novel 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 (2-D) 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 automatic extraction of the focused OOI. The fact that we can extract a semantically meaningful object automatically from low-DOF images suggests a variety of applications, such as image indexing for content-based retrieval, object-based image compression, video object extraction, three-dimensional (3-D) microscopic image analysis, image enhancement for digital camera, range segmentation for depth estimation, and fusion of multiple images, which are differently focused [6].

In order to model defocusing of a focused image, blurring effect by defocusing has been often described by a 2-D Gaussian function

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \qquad (1)$$

where σ is a filter scale or spread parameter which controls the amount of defocusing. Thus, a defocused image $I_d(x,y)$ at a pixel (x,y) can be modeled as the linear convolution of a focused image $I_f(x,y)$ and a Gaussian function $G_\sigma(x,y)$

$$I_d(x,y) = G_\sigma(x,y) * I_f(x,y). \qquad (2)$$

As shown in (2), since the defocused image is low-pass filtered, high-frequency components in the image are removed or reduced. Thus, by assuming only sharply focused regions contain adequate high-frequency components, it should be possible to distinguish the focused regions from the defocused regions by comparing the amount of the high-frequency contents.

There are two approaches to the segmentation of the low-DOF images: edge-based and region-based approaches. The edge-based method in [7] 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 the natural object, yielding disconnected boundaries [2]. The region-based segmentation algorithms in [2], [8]–[10] rely on the detection of the high frequency areas in the image. A reasonable starting point is to measure the degree of focus for each pixel by computing high-frequency components. To this end, several methods have been used, such as spatial summation of the squared anti-Gaussian (SSAG) function [8], variance of wavelet coefficients in the high-frequency bands [2], a multiscale statistical description of high-frequency wavelet coefficients [9], local variance [10], 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 also generate errors in these 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 multiscale approaches employed in both [2] and [9] tend to generate jerky boundaries, although refinement algorithms for high-resolution
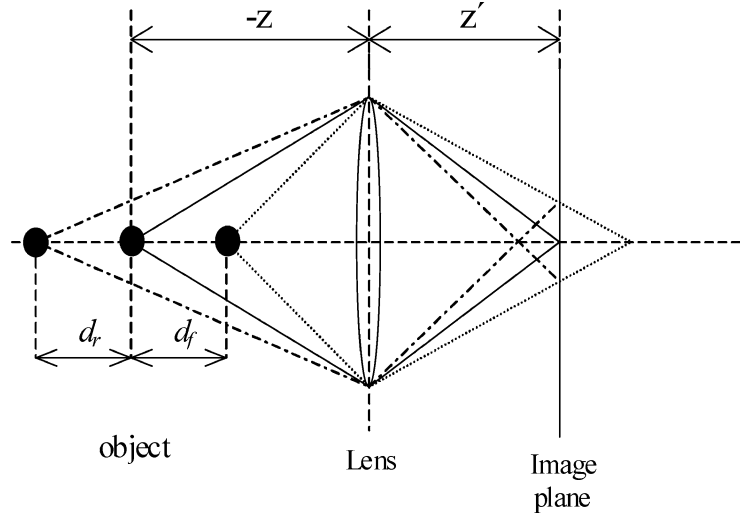
Fig. 1. Optical geometry of a typical camera.

classification are incorporated. To reduce the above-mentioned errors, Won *et al.* [10] proposed a block-wise 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.

Considering the limitations of the existing methods, we derive the following criteria for desirable low-DOF image segmentation.

- Dependency on high-frequency contents: Even if multiscale schemes are incorporated, relying on high-frequency components only does not yield accurate outputs when focused smooth regions exist in an image. Finding focused regions using high-frequency components should be supported by some auxiliary methods or other cues.

- Efficiency: These days, its possible applications go beyond the traditional computer vision problems such as range segmentation for depth estimation [8] and target recognition [11]. 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, birds flying in the sky can be extracted from a video sequence. In such case, one of the key factors is computational complexity. The methods using MRF model [8], [10] are not suitable to handle video sequence due to their iterative relaxation process. Some deterministic methods must be a better choice.

In this paper, we consider an efficient and fast segmentation algorithm to satisfy the above criteria. We compute higher order statistics (HOS) for each pixel [12], which effectively assesses the amount of high-frequency components in the focused regions, whereas less sensitive to noises in the defocused regions. Then, we employ a morphological approach so that even focused smooth areas can be merged into the surrounding areas with high-frequency components (i.e., edges). The final decision of the focused regions is conducted by region merging and thresholding.

The outline of this paper is as follows. In Section II, the concept of depth of field and low DOF is described. In Section III, the proposed algorithm is introduced. Experimental results and conclusion follow in Sections IV and V, respectively.

II. DEPTH OF FIELD AND LOW DOF

An ideal lens has the disadvantage that it only brings to focus light from points at a distance $-z$ given by the familiar lens equation [13]

$$\frac{1}{z'} + \frac{1}{-z} = \frac{1}{f} \tag{3}$$

where z' is the distance of the image plane from the lens and f is the focal length (see Fig. 1). Points at other distances are imaged as small blur circles. The size of the blur circle can be determined as follows. A point at distance $-\overline{z}$ is imaged at a point $\overline{z}'$ from the lens, where $1/\overline{z}' + 1/-\overline{z} = 1/f$, and so

$$(\overline{z}' - z') = \frac{f}{(\overline{z} + f)} \frac{f}{(z + f)} (\overline{z} - z). \tag{4}$$

If the image plane is situated to receive correctly focused images of object at distance $-z$, then points at distance $-\overline{z}$ will give rise to blur circles of diameter $(d/z') |\overline{z}' - z'|$, where d is the diameter of the lens. The *depth of field* is the range of distances over which objects are focused "sufficiently well," in the sense that the diameter of the blur circle is less than the resolution of the imaging device. The DOF depends, of course, on what sensor is used, but in any case it is clear that the larger the lens aperture, the less the DOF. Clearly, errors in focusing become more serious when a large aperture is employed. As shown in Fig. 1, d_f and d_r are the front and rear limits of the "*depth of field*." With low DOF, only the OOI is in sharp focus, whereas objects in background are blurred to out of focus. Photographers often use this photographic technique to point their interest in the image or to help viewers understand the depth information from the 2-D image. Examples are shown in Fig. 2.

Fig. 2. Low-DOF images.

III. Proposed Algorithm

Let R represent a set of pixels, $R = \{(x, y); 1 \leq x \leq X, 1 \leq y \leq Y\}$, where the image size is $X \times Y$. Our goal is to partition R into sharply focused objects-of-interest, denoted by OOI, and remaining regions, expressed by OOI^c.

Let $P = \{R_i, i \in \{1, \ldots, N\}\}$ denote a partition of R. The OOI of an image is defined as follows:

$$\text{OOI} = \bigcup_{i=1}^{N_{\text{OOI}}} R_i \tag{5}$$

where R_i is the ith connected region and N_{OOI} denotes the number of regions belonging to OOI. In other words, OOI represents the focused objects of interest, composed of N_{OOI} regions of P. Equation (5) naturally allows for the definition of multiple OOIs, i.e., OOI can be composed of separated sub-OOIs.

A. Feature Space Transformation Using HOS

The first step toward segmentation consists in transforming the input low-DOF image I into an appropriate feature space. The choice of the feature space depends on the applications that the algorithm is aimed at. For instance, the feature space may represent the set of wavelet coefficients [2], or local variance image field [10].

In our case, we compute HOS for feature space transformation. HOS are well suited to solving detection and classification problems because they can suppress Gaussian noise and preserve some of the non-Gaussian information [12], [14]. In this paper, the fourth-order moments are calculated for all pixels in the image. The fourth-order moment at (x, y) is defined as follows:

$$\hat{m}^{(4)}(x, y) = \frac{1}{N_\eta} \sum_{(s,t) \in \eta(x,y)} (I(s,t) - \hat{m}(x,y))^4 \tag{6}$$

where $\eta(x, y)$ is a set of neighboring pixels centering at (x, y), $\hat{m}(x, y)$ is the sample mean of $I(x, y)$ (i.e., $\hat{m}(x, y) = (1/N_\eta) \sum_{(s,t) \in \eta(x,y)} I(s,t)$), and N_η is a size of η. Since the dynamic range of the fourth-order moment values is extremely large, the value for each pixel is down scaled and limited by 255 such that each pixel takes a value from [0, 255]. The outcome image is called a HOS map. At a pixel $(x, y) \in R$, a component of the HOS map, $\text{HOS}(x, y)$, is defined as follows:

$$\text{HOS}(x, y) = \min\left(255, \frac{\hat{m}^{(4)}(x, y)}{\text{DSF}}\right) \tag{7}$$

where DSF denotes down scaling factor. For a variety of test images, it is observed that 100 is appropriate for DSF. Hence, DSF has been set to 100 throughout the paper. By applying (7) for all pixels, we have a HOS map, $O = \{\text{HOS}(x, y); (x, y) \in R\}$. Fig. 3(c) shows HOS map generated from an image with low DOF [Fig. 3(a)]. Comparing it to a local variance map shown in Fig. 3(b), we can observe that the HOS map yields denser and higher values in the focused areas, suppressing noise in the defocused regions.

B. HOS Map Simplification by Morphological Filtering by Reconstruction

In the previous subsection, feature space transformation was addressed such that more adequate feature space is exploited for segmentation. The HOS map transformed from the low-DOF image has gray levels ranging from 0 to 255, where high values indicate the existence of high-frequency components (i.e., possibly focused regions). However, as mentioned earlier, there could be some focused smooth regions, which may not be easily detected by HOS transformation. Similarly, defocused texture regions may generate noise [see Fig. 4(b)]. Therefore, a proper tool for HOS map simplification is needed to remove these errors, appearing in the form of small dark and bright patches in focused and defocused regions, respectively.

Morphological filtering is well known as a useful approach to smooth noisy gray-level images by a determined composition of opening and closing with a given structuring element. A large number of morphological tools rely on two basic sets of transformations known as *erosion* and *dilation*. Let B denote a window or flat structuring element and let $B_{x,y}$ be the translation of B so that its origin is located at $(x, y) \in R$. Then, the erosion $\varepsilon_B(O)$ of a HOS map O by the structuring element B is used in constructing a morphological filter for image simplification $\varepsilon_B(O)(x, y) = \min_{(k,l) \in B_{x,y}} \text{HOS}(k, l)$. Similarly, the dilation $\delta_B(O)(x, y) = \max_{(k,l) \in B_{x,y}} \text{HOS}(k, l)$. Elementary erosions and dilations allow the definition of morphological filters such as morphological opening and closing: Morphological opening, $\gamma_B(O)$ and closing, $\varphi_B(O)$ are given by $\gamma_B(O) = \delta_B(\varepsilon_B(O))$ and $\varphi_B(O) = \varepsilon_B(\delta_B(O))$, respectively.

The morphological opening operator $\gamma_B(O)$ applies an erosion $\varepsilon_B(\cdot)$ followed by a dilation $\delta_B(\cdot)$. Erosion leads to darker images, whereas dilation to brighter images. A morphological opening (resp. closing) simplifies the original signal by removing the bright (resp. dark) components that do not fit within the structuring element B. This morphological operators can also be directly applied to binary image without any modification.

The disadvantage of these operators is that they do not allow a perfect preservation of the contour information. To overcome this problem, so-called filters by reconstruction are preferred [15], [16]. Although similar in nature, they rely on different erosion and dilation operators, making their definitions slightly more complicated. The elementary geodesic erosion $\varepsilon^{(1)}(O, O_R)$ of size one of the original image O with respect to the reference image O_R is defined as

$$\varepsilon^{(1)}(O, O_R)(x, y) = \max\{\varepsilon_B(O)(x, y), O_R(x, y)\} \tag{8}$$
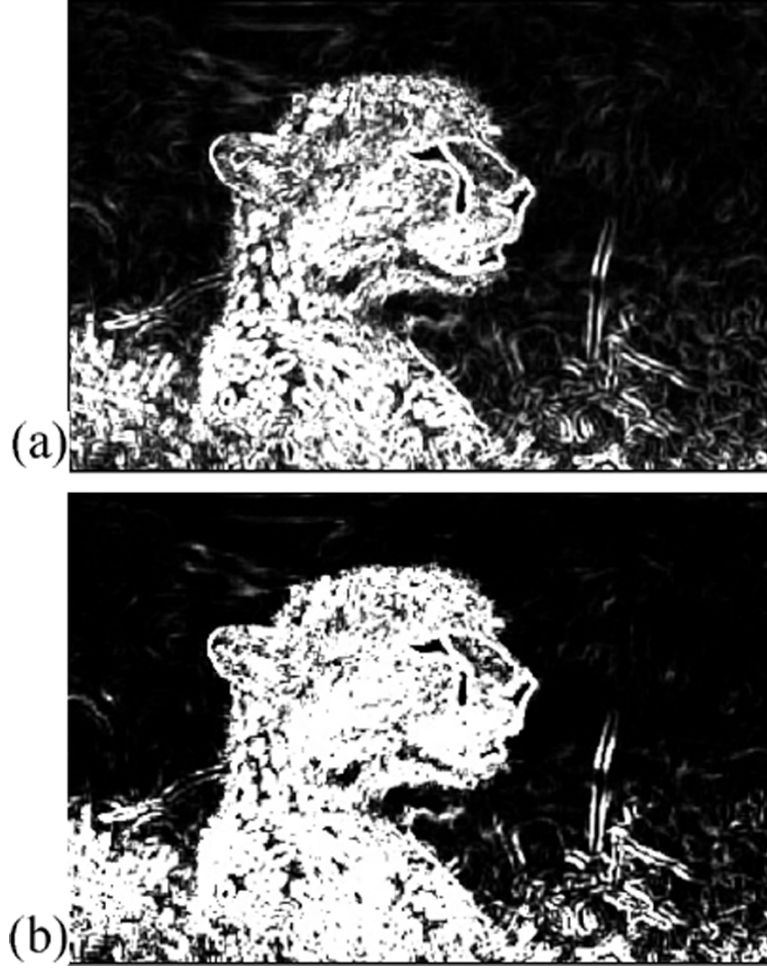
Fig. 3.   Feature space transformed by (a) local variance and by (b) higher order statistics.
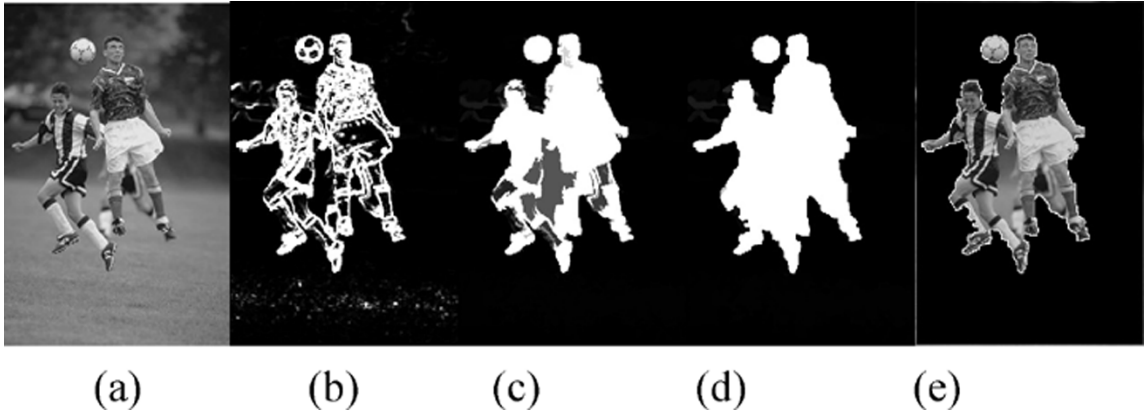

Fig. 4.   Pictorial illustration of the proposed algorithm. (a) Low-DOF image. (b) HOS map. (c) Simplified image. (d) Region merging. (e) Final decision by thresholding.

and the dual geodesic dilation $\delta^{(1)}(O, O_R)$ of O with respect to O_R is given by

$$\delta^{(1)}(O, O_R)(x, y) = \min\left\{\delta_B(O)(x, y), O_R(x, y)\right\}. \quad (9)$$

Thus, the geodesic dilation $\delta^{(1)}(O, O_R)$ dilates the image O using the classical dilation operator $\delta_B(O)$. As we know, dilated gray values are greater or equal to the original values in

O. However, geodesic dilation limits these to the corresponding gray values of O_R. The choice of the reference image O_R will be discussed shortly.

Geodesic erosions and dilations of arbitrary size are obtained by iterating the elementary versions $\varepsilon^{(1)}(O, O_R)$ and $\delta^{(1)}(O, O_R)$ accordingly. For example, the geodesic erosion (dilation) of infinite size, which is so-called reconstruction by erosion (by dilation) is given by the following.

• **Reconstruction by erosion:**

$$\varphi^{(\text{rec})}(O, O_R) = \varepsilon^{(\infty)}(O, O_R) = \varepsilon^{(1)} \circ \varepsilon^{(1)} \circ \cdots \circ \varepsilon^{(1)}(O, O_R). \quad (10)$$

• **Reconstruction by dilation:**

$$\gamma^{(\text{rec})}(O, O_R) = \delta^{(\infty)}(O, O_R) = \delta^{(1)} \circ \delta^{(1)} \circ \cdots \circ \delta^{(1)}(O, O_R). \quad (11)$$

Notice that $\varphi^{(\text{rec})}(O, O_R)$ and $\gamma^{(\text{rec})}(O, O_R)$ will reach stability after a certain number of iterations.

Finally, the two simplification filters, *morphological opening by reconstruction*, $\gamma^{(\text{rec})}(\varepsilon_B(O), O)$, and *morphological closing by reconstruction*, $\varphi^{(\text{rec})}(\delta_B(O), O)$, become special cases of $\gamma^{(\text{rec})}(O, O_R)$ and $\varphi^{(\text{rec})}(O, O_R)$.

Morphological opening by reconstruction first applies the basic erosion operator $\varepsilon_B(O)$ to eliminate bright components that do not fit within the structuring element B. However, unlike morphological opening which applies just a basic dilation afterwards, the contours of components that have not been completely removed are restored by the reconstruction by dilation operator $\gamma^{(\text{rec})}(\cdot, \cdot)$. The reconstruction is accomplished by choosing O as the reference image O_R, which guarantees that for each pixel the resulting gray-level will not be higher than that in the original image O.

In the proposed system, we employ *morphological closing-opening by reconstruction*, which is *morphological closing by reconstruction* followed by *morphological opening by reconstruction*. The strength of the *morphological closing-opening by reconstruction* filter is that it fills small dark holes and removes small bright isolated patches, whereas perfectly preserving other components and their contours. Obviously, the size of removed components depends on the size of the structuring element. As shown in Fig. 4(c), the focused smooth regions are well covered whereas the scattered small regions in the background are removed by the filter.

C. Region Merging and Adaptive Thresholding

In typical morphological segmentation techniques, the simplification by morphological filters is followed by marker extraction and watershed algorithm to partition an image or scene into homogeneous regions in terms of intensity. The marker extraction step selects initial regions, for instance, by identifying large regions of constant gray level obtained in the simplification step. After the marker extraction, the number and the interior of the regions to be segmented are known [17]. However, a large number of pixels are not yet assigned to any region. These pixels correspond to uncertainty areas mainly concentrated around the contours of the regions. Assigning these pixels to a given region can be viewed as a decision process that precisely defines the partition. The classical morphological decision tool is the *watershed*, which labels pixels in a similar fashion to region growing techniques.

Unlike the abovementioned conventional intensity-based segmentation schemes, the task of the low-DOF image segmentation is to extract focused region (i.e., OOI) from the image rather than partitioning the image. In this case, the reasonable way is to grow initially detected focused regions until they occupy all the focused regions. In the following, we propose a decision process, which is composed of two steps: region merging and final decision by thresholding.

1) Region Merging: Our region merging is started based on seed regions, which can be regarded as definitely focused regions of OOI. First, every flat zone is treated as a region regardless of its size, which means even one pixel zone can become a region. Then, we define regions having the highest value v_h in the simplified HOS map as seed regions and these seed regions become initial OOI [see white areas in Figs. 4(c) and 5(a)]. We also define regions having values less than or equal to a predefined value T_L ($T_L < v_h$) as definitely defocused regions. Those regions are labeled as initial OOI^c. Then, the remaining regions are labeled as uncertainty regions with pixel values (T_L, v_h). A pictorial example is shown in Fig. 5(a), where the initial OOI and OOI^c are denoted by white and black regions, respectively, whereas the dashed regions R_i, R_j, and R_k indicate uncertainty regions. Our goal in this region merging process is to assign uncertainty regions to either OOI or OOI^c. Such an assignment is iteratively conducted by considering bordering relationship between uncertainty region and current OOI and OOI_n (i.e., OOI at the nth iteration). Specifically, we develop an algorithm that assigns an ith uncertainty region $R_{n,i}$ in the nth iteration to either OOI_n or OOI_n^c by computing *normalized overlapped boundary* (nob).

Given a partition P_n, nob between ith uncertainty region $R_{n,i} \in P_n$ and OOI_n is given by

$$\text{nob}_{n,i} = \frac{\text{card}(\text{BR}_{n,i} \cap \text{OOI}_n)}{\text{card}(\text{BR}_{n,i})} \quad (12)$$

where $card(\text{A})$ denotes the cardinality of a set A, and the set of boundary pixels of $R_{n,i}$ is defined as $\text{BR}_{n,i} = \{ x \notin R_{n,i} \mid \min_{r \in R_{n,i}} \|r - x\| \leq T_b \}$. The threshold value T_b for defining boundary pixels of a region is simply set to be one in our system. Note that $\text{BR}_{n,i}$ can also be defined morphologically as the difference between $R_{n,i}$ and its dilation. In this case, the structuring element is of size 3×3 since T_b is set to be one.

$\text{nob}_{n,i}$ takes zero when the uncertainty region $R_{n,i}$ is not touching OOI_n and one when $R_{n,i}$ is perfectly enclosed by OOI_n, and will be used for labeling the uncertainty regions in P_n. Obviously, the uncertainty region $R_{n,i} \in P_n$ must belong to either OOI_n or OOI_n^c. It is expressed as follows in hypothesis-testing terms

$$H_1 : R_{n,i} \subseteq \text{OOI}_n \; ; \; H_0 : H_1^c. \quad (13)$$

The normalized overlapped boundary can be modeled as a continuous random variable **nob**, taking values of nob in [0, 1]. If $\text{nob}_{n,i}$ is larger than a threshold value, the region $R_{n,i}$ is merged to OOI_n. Then, the partition P_n is updated, yielding an increasing sequence of OOI_n which eventually converges to OOI. A reasonable starting point for finding the threshold value, which is denoted by T_{nob}, is to use the likelihood ratio test as follows (we drop the iteration index n in the followings for simplicity):

Assign R_i to OOI if $P(\text{OOI} \mid \text{nob}_i)$

$$> P(\text{OOI}^c \mid \text{nob}_i), \text{otherwise assign to} \text{OOI}^c$$
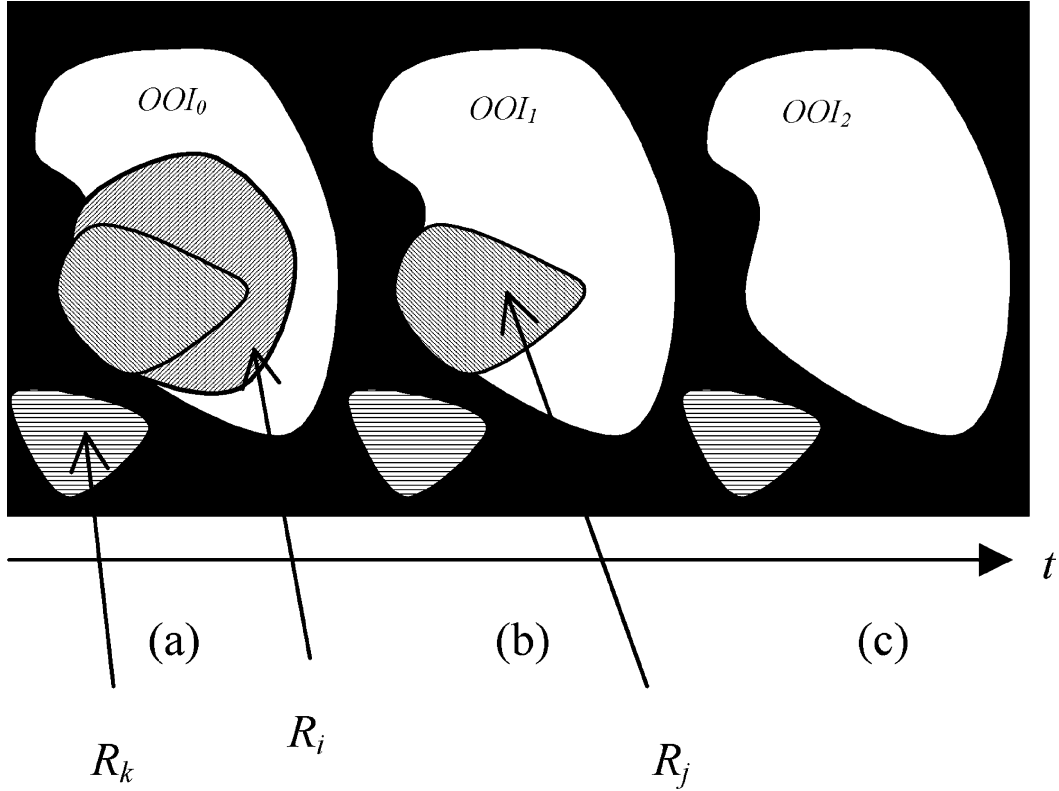
Fig. 5. Evolution of OOI by the proposed region merging. (a) Initial OOI and three uncertain regions with pixel values (T_L, v_h) in the simplified HOS map. (b) R_i is merged into OOI. (c) Final OOI, after R_j is merged into OOI. Note that R_k is not decided as OOI since it has a value less than that of OOI_2.

where OOI represents the class for the OOI with prior probability $P(\text{OOI})$ and OOI^c denotes the class for the non-OOI with prior probability $P(\text{OOI}^c) = 1 - P(\text{OOI})$. $P(\text{OOI} \mid \text{nob}_i)$ and $P(\text{OOI}^c \mid \text{nob}_i)$ represent the *a posteriori* conditional probabilities that correspond to H_1 and H_0, respectively. If we apply Bayes theorem on both sides of the expression and rearrange terms as follows:

$$\frac{p(\text{nob}_i \mid \text{OOI})}{p(\text{nob}_i \mid \text{OOI}^c)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(\text{OOI}^c)}{P(\text{OOI})} \qquad (14)$$

the left-hand ratio is known as the *likelihood ratio* and the entire equation is often referred to as the *likelihood ratio test*. Since the test is based on choosing the region class with maximum *a posteriori* probability, the decision criterion is called the *maximum a posteriori* (MAP) *criterion*. It is also called the *minimum error criterion*, since on the average, this criterion yields the minimum number of incorrect decisions. Since the OOI and background may have any size and shape, we assume equal priors $(P(\text{OOI}) = P(\text{OOI}^c))$, and, thus, the expression reduces to the *maximum likelihood* (ML) *criterion*

$$\frac{p(\text{nob}_i \mid \text{OOI})}{p(\text{nob}_i \mid \text{OOI}^c)} \underset{H_0}{\overset{H_1}{\gtrless}} 1. \qquad (15)$$

We propose to model the class-conditional probability density functions by exponential distributions

$$p(\text{nob}_i \mid \text{OOI}^c) = \lambda_1 e^{-\lambda_1 \text{nob}_i} u(\text{nob}_i)$$
$$p(\text{nob}_i \mid \text{OOI}) = \lambda_2 e^{-\lambda_2(1-\text{nob}_i)} u(1 - \text{nob}_i) \qquad (16)$$

where $u(x)$ denotes the step function. We believe that these distributions are suited to the real data because $p(\text{nob}_i \mid \text{OOI})$ would have high values around $\text{nob}_i = 1$ and rapidly decay as $\text{nob}_i \to 0$, whereas $p(\text{nob}_i \mid \text{OOI}^c)$ would have high values around $\text{nob}_i = 0$ and rapidly decay as $\text{nob}_i \to 1$. Finally, optimal threshold for nob_i can be obtained by rearranging (15) and (16)

$$\text{nob}_i \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\lambda_2}{\lambda_1 + \lambda_2} + \frac{\ln\left(\frac{\lambda_1}{\lambda_2}\right)}{\lambda_1 + \lambda_2} = T_{\text{nob}}. \qquad (17)$$

The parameters λ_1 and λ_2 can be estimated from the actual data. However, if we assume symmetry between the exponential distributions $(\lambda_1 = \lambda_2)$, the expression for the optimal threshold can be approximated and simplified as

$$T_{\text{nob}} = \frac{\lambda_2}{\lambda_1 + \lambda_2} + \frac{\ln\left(\frac{\lambda_1}{\lambda_2}\right)}{\lambda_1 + \lambda_2} \approx \frac{1}{2}. \qquad (18)$$

Hence, if nob_i is larger than T_{nob}, R_i is merged to OOI and OOI is updated. This process is iterated until no merging occurs. A pictorial illustration is provided in Fig. 5. In Fig. 5(a), nob_i is greater than T_{nob}, thus R_i merges into OOI, whereas R_k does not since nob_k is less than T_{nob}. In the next iteration, as shown in Fig. 5(b), R_j merges into OOI since $\text{nob}_j > T_{\text{nob}}$. In order to expedite the process, very small regions can be merged to the neighbor region with nearest value in advance.

2) Final Decision: In the preceding subsection, the focused regions (i.e., OOI) are updated by region merging. In other words, an uncertainty region whose nob is larger than T_{nob}, has
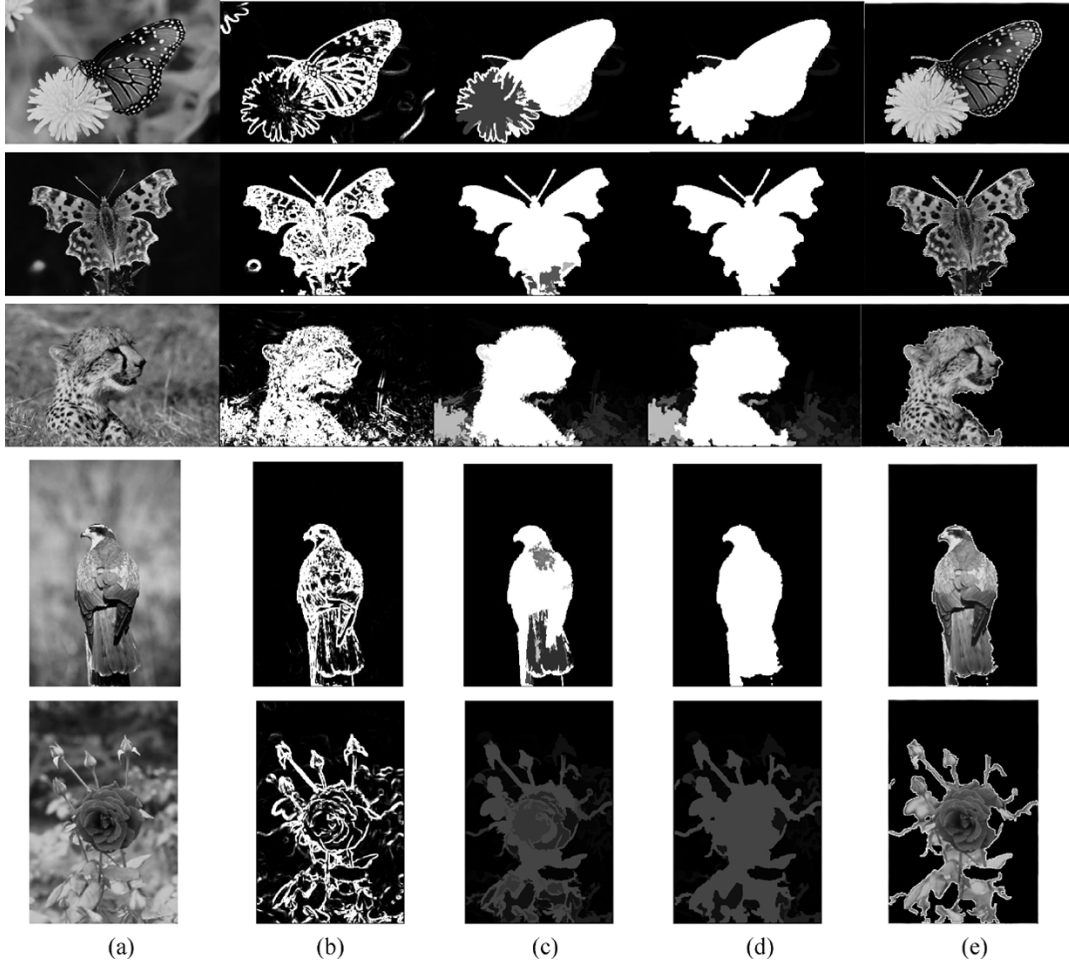
Fig. 6. Experimental results from each process. (a) Low-DOF image. (b) HOS map. (c) Simplified image. (d) Region merging. (e) Final decision by thresholding.

been incorporated into OOI. Now, the final decision becomes to extract OOI from the final partition P. It is easily done by extracting regions having the highest value. For instance, in Fig. 5(c), OOI_2 will be extracted whereas R_k will not be decided as OOI since it has a value than less than that of OOI_2.

IV. EXPERIMENTAL RESULTS

The proposed algorithm has been implemented and tested on low-DOF images selected from the JPEG compressed COREL CD-ROM image collection. Color images are first transformed into gray level images in our system. We used a neighborhood of size 3×3 for η in (6). The threshold value T_L to determine the initial OOI^c was set to be 20 in the tests. One of the most important parameters is the size of the structuring element (SE) of the morphological filter. We used rectangular SE and set the size to be 31×31 for all experiments except the image shown in Fig. 4(a). Since the size of the ball shown in the figure is too small, it is removed by the filter when 31×31 of SE is used. For a better subjective result, 21×21 of SE was employed on this image only.

Fig. 6 gives some experimental results of the proposed algorithm. It shows outcomes of each process for several test images. Fig. 7 shows the segmented results from [2], [9], [10] and the proposed algorithm for visual comparison. As one can see from the figures, the results obtained from [2] and

[9] are blocky due to block-wise initial classification, even if refinement algorithms for high-resolution classification were applied. The algorithm from [10] provides pretty reasonable segmentation results. However, due to its smoothness constraint adopted in the Markov random field (MRF) model, it tends to connect adjacent non-OOI regions. The proposed algorithm yields more accurate results over various images with low DOF.

The performance of the proposed algorithm is also evaluated by using objective criterion. In [2], the performance is evaluated by *sensitivity*, *specificity*, and *error rate*. However, since they are defined as the ratios of the areas, even different shape of areas can show high performance as far as the size of the extracted OOI (or background) is close to that of reference. We propose to use a pixel-based quality measure [18], which was used to evaluate the performances of video object segmentation algorithms [19]. The spatial distortion of the estimated OOI from the reference OOI is defined as

$$d(O^{\text{est}}, O^{\text{ref}}) = \frac{\sum\limits_{(x,y)} O^{\text{est}}(x,y) \otimes O^{\text{ref}}(x,y)}{\sum\limits_{(x,y)} O^{\text{ref}}(x,y)} \tag{19}$$

where O^{est} and O^{ref} are the estimated and reference binary masks, respectively, and $\otimes$ is the binary "XOR" operation.
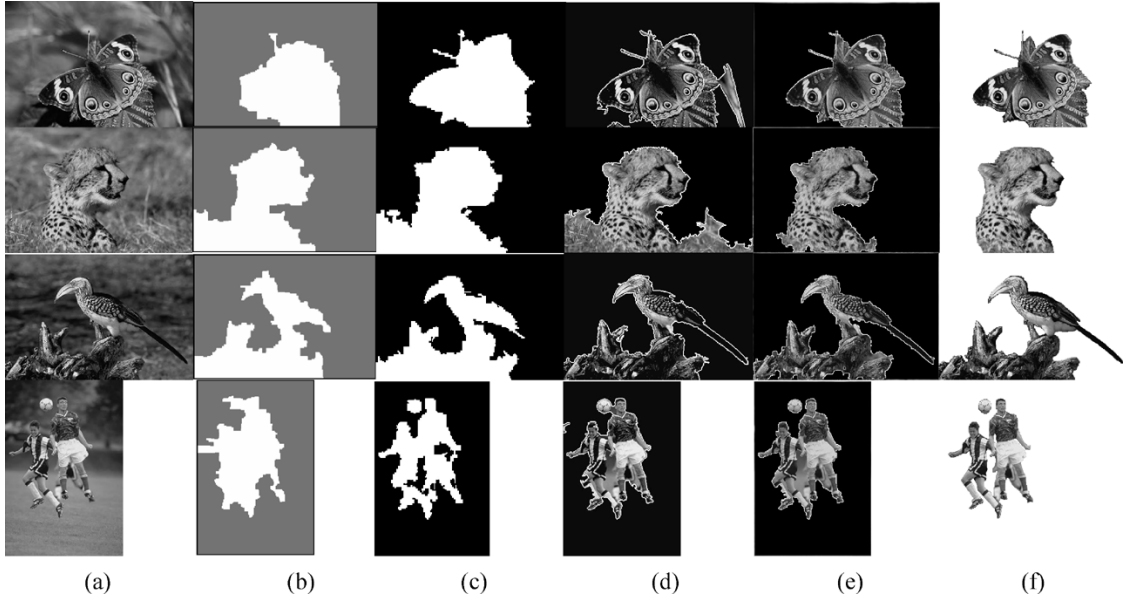
Fig. 7. Visual comparison of segmentation results. (a) Low-DOF images. (b) Results from [2]. (c) Results from [8]. (d) Results from [9]. (e) Results from the proposed algorithm. (f) References by human manual segmentation.

TABLE I
PERFORMANCE EVALUATION BY OBJECTIVE CRITERION. IMAGES IN FIG. 7(a) ARE NUMBERED FROM TOP TO BOTTOM. BOLD FIGURES INDICATE THE BEST OF THE FOUR METHODS

Image	Ref. [2]	Ref. [8]	Ref. [9]	proposed
(1)	0.13	0.05	0.16	**0.04**
(2)	0.29	0.26	0.44	**0.11**
(3)	0.21	0.17	**0.12**	0.16
(4)	0.31	0.22	0.23	**0.17**

Table I shows the spatial distortion measures of the results from [2], [9], [10], and the proposed algorithm. Reference maps are obtained by manual segmentation, as shown in Fig. 7(f). For binary "XOR" operation, pixels on OOI are set to be one, otherwise zero. As shown in Table I, our algorithm has lower distortion measures than those from other methods and these measure are well matched with subjective evaluation.

V. CONCLUSION

We developed an algorithm that separates the pixels in the low-DOF images into two regions based on their higher order statistics. To this end, the low-DOF image was transformed into an appropriate feature space, which was called HOS map in this paper. *Morphological filter by reconstruction* was applied to simplify the HOS map, followed by region-merging technique and thresholding for final decision.

By employing the powerful morphological tool for simplification, the proposed scheme performs well even for focused smooth regions as far as their boundaries contain high frequency components (i.e., edges). Also, it shows its robustness to scattered sharp areas in the background thanks to the powerful morphological simplification and the following region merging. Nonetheless, if the focused smooth region is too large, the proposed algorithm may need to incorporate some semantic or human knowledge.

Our future work is to extend the algorithm to video object segmentation in cooperation with the low-DOF photographic technique. Since extracting video objects from arbitrary video sequences is still difficult to solve, this low-DOF photographic technique-assisted segmentation could be a pertinent choice until a powerful scheme on arbitrary videos is attained. For the use in a variety of video applications, we are improving the processing time to reach real-time on a low-end PC. Currently, the processing time per QCIF format image is one second on an Intel Pentium IV 2.78 GHz PC.

REFERENCES

[1] R. G. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1992.
[2] J. Z. Wang, J. Li, R. M. Gray, and G. Wiederhold, "Unsupervised multiresolution segmentation for images with low depth of field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 1, pp. 85–90, Jan. 2001.
[3] P. J. Besl and R. C. Jain, "Segmentation through variable –order surface fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-10, no. 3, pp. 167–192, Mar. 1988.
[4] L. M. Lifshitz and S. M. Pizer, "A multiresolution hierarchical approach to image segmentation based on intensity extrema," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 6, pp. 529–540, Jun. 1990.
[5] D. Comaniciu and P. Meer, "Robust analysis of feature spaces: Color image segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1997, pp. 750–755.

[6] 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," in *Proc. 10th Int. Workshop on Theoretical Foundations of Computer Vision*, vol. 2032, Dagstuhl Castle, Germany, Mar. 2000, pp. 226–237.

[7] D.-M. Tsai and H.-J. Wang, "Segmenting focused objects in complex visual images," *Pattern Recognit. Lett.*, vol. 19, pp. 929–949, 1998.

[8] C. Yim and A. C. Bovik, "Multiresolution 3-D range segmentation using focused cues," *IEEE Trans. Image Process.*, vol. 7, no. 9, pp. 1283–1299, Sep. 1998.

[9] Z. Ye and C.-C. Lu, "Unsupervised multiscale focused objects detection using hidden Markov tree," in *Proc. Int. Conf. Computer Vision, Pattern Recognition, and Image Processing*, Durham, NC, Mar. 2002, pp. 812–815.

[10] C. S. Won, K. Pyun, and R. M. Gray, "Automatic object segmentation in images with low depth of field," in *Proc. Int. Conf. Image Processing*, vol. III, Rochester, NY, Sep. 2002, pp. 805–808.

[11] Y. Boykov and D. Huttenlocher, "A new Baysian framework for object recognition," presented at the Proc. DARPA Image Understanding Workshop, 1998.

[12] G. Gelle, M. Colas, and G. Delaunay, "Higher order statistics for detection and classification of faulty fanbelts using acoustical analysis," in *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics*, Jul. 1997, pp. 43–46.

[13] B. Horn, *Robot Vision*. Cambridge, MA: The MIT Press, 1986.

[14] M. K. Tsatsanis and G. B. Giannakis, "Object and texture classification using higher order statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 7, pp. 733–750, Jul. 1992.

[15] P. Salembier and M. Pardas, "Hierarchical morphological segmentation for image sequence coding," *IEEE Trans. Image Process.*, vol. 3, no. 9, pp. 639–651, Sep. 1994.

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

[17] P. Salembier, L. Torres, F. Meyer, and C. Gu, "Region-based video coding using mathematical morphology," *Proc. IEEE*, vol. 83, no. 6, pp. 843–856, Jun. 1995.

[18] M. Wollborn and R. Mech, "Refined procedure for objective evaluation of video generation algorithms," Doc. ISO/IEC JTC1/SC29/WG11 M3448, 1998.

[19] C. Kim and J.-N. Hwang, "An integrated scheme for object-based video abstraction," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2000, pp. 303–311.

Changick Kim (M'01–SM'89) was born in Seoul, Korea. He received the B.S. degree in electrical engineering from YonSei University, Seoul, the M.S. degree in electronics and electrical engineering from Pohang University of Science and Technology (POSTECH), Pohang, Korea, and the Ph.D. 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 communication, 3-D video processing, image/video understanding, intelligent media processing, and error resilient video coding.