

An Effective Orientation-based Method and Parameter Space Discretization for Defined Object Segmentation

Huy Hoang Nguyen,¹ GueeSang Lee,² SooHyung Kim,² and HyungJeong Yang.²

¹ University of Science, Department of Computer Science, 227 Nguyen Van Cu Street, District 5, Ho Chi Minh City,
Vietnam
(84) 94-946-8684.

¹[e-mail: nguyenhuyhoang83@gmail.com]

² Department of Electronics and Computer Engineering, Chonnam National University
Gwangju 500 757, Republic of Korea.

²[e-mail: {gslee, shkim, hjyang}@jnu.ac.kr]

*Corresponding author: GueeSang Lee

Received July 19, 2013; revised October 20, 2013; accepted November 29, 2013; published December 27, 2013

Abstract

While non-predefined object segmentation (NDOS) distinguishes an arbitrary self-assumed object from its background, predefined object segmentation (DOS) pre-specifies the target object. In this paper, a new and novel method to segment predefined objects is presented, by globally optimizing an orientation-based objective function that measures the fitness of the object boundary, in a discretized parameter space. A specific object is explicitly described by normalized discrete sets of boundary points and corresponding normal vectors with respect to its plane shape. The orientation factor provides robust distinctness for target objects. By considering the order of transformation elements, and their dependency on the derived over-segmentation outcome, the domain of translations and scales is efficiently discretized. A branch and bound algorithm is used to determine the transformation parameters of a shape model corresponding to a target object in an image. The results tested on the PASCAL dataset show a considerable achievement in solving complex backgrounds and unclear boundary images.

Keywords: Solution space discretization, global optimization, super-pixel, object segmentation, branch-bound.

A preliminary version of this paper appeared in ACM ICUIMC 2013, Jan. 17-19, Kota Kinabalu, Malaysia. This version includes new figures and additional analysis compared to state-of-the-art methods as well as complete editing in English.

<http://dx.doi.org/10.3837/tiis.2013.12.013>

1. Introduction

First of all, it is necessary to differentiate image segmentation and object segmentation. Image segmentation is a classic problem in computer vision. It targets the partitioning of an image into many non-semantic regions, according to certain low-level features, such as color, intensity, gradient, and/or texture. The result of image segmentation does not indicate which regions belong to the foreground or background. Since it relies on low-level features, it consists of many assumptions, e.g. regions belonging to an object have homogenous color, and two adjacent regions are apparently different, in terms of color or intensity. On the other hand, object segmentation aims to find regions or boundaries of objects in an image. It explicitly distinguishes the foreground and background. Object segmentation is still a challenging problem, attracting the interest of many researchers.

Non-predefined object segmentation (NDOS) raised a lot of attention for decades and has been studied extensively. In active contour based methods [1], a contour is initialized around a target object. The process of minimizing the energy function that presents the curvature and the smoothness of the given contour evolutionally deforms it to the boundary of the target object. GrabCut [2] requires initializing a rectangle enclosing the target object. The object is segmented by minimizing an energy, relying on the knowledge of data inside and outside the rectangle. Another interactive object segmentation method is lazy snapping [3]. The user needs to provide two types of markers (i.e. foreground and background) appropriately near the boundary of the target object. The segmentation is achieved by global optimization using graph-cut [4]. Some non-initialization segmentation methods rely on the saliency map [5]. Instead of utilizing a certain prior, K. Fukuchi et al. [6] used high-saliency regions as an automatic prior, to minimize an energy function based on graph-cut. Ming-Ming Cheng et al. [7] presented an outstanding method to find a saliency map based on global contrast of both feature and spatial information. Furthermore, P. Mehrani et al. [8] combined saliency map, graph-cut, and learning in their work. All in all, NDOS methods aim to segment certain objects in an image, without knowing its actual shapes. They rely on basic characteristics, such as color continuity, and intensity discontinuity (high gradient magnitudes, or edges), to differentiate object and background.

In predefined object segmentation (DOS), the idea of utilizing prior shape for object detection was first raised by D. H. Ballard [9] in a paper about the generalized Hough transform. In the research, an object is described by a reference point, and a set of vectors presenting discrete points on the boundary of the object, with respect to the reference point.

Different from prior input in NDOS methods, prior input in DOS has a higher level representation, in terms of its structure. Such priors describe the whole shape of a target object, rather than its internal characteristics. V. Lempitsky et al. [10] proposed a segmentation framework in which priors are exemplars of many target objects. An object is defined by binary images of its plane shapes from various aspects. A. Toshev et al. [11] presented an object descriptor that is based on the holistic nature of an object. According to the descriptor, each boundary pixel is linked to all remaining ones, to form a 'chord' feature that records not only the length, but also the orientation relationships of pairs of boundary pixels. A chordiogram that is the histogram of all chords of an object is used as the descriptor of an object.

S. Abbasi [12] utilized curvature scale space (CSS) image to represent object boundary. In this approach, the original boundary (u -parameterized curve) of an object is smoothed by different σ -variance Gaussian kernels. The number of curvature zero-crossing points of a

curve is inversely proportional to σ . The image capturing the relationship between u and σ is the CSS image of the object boundary. In DOS, since a target object is explicitly indicated based on its own nature, the description is more objective, than low-level priors and interactive markers.

This paper presents a novel method to detect and segment objects given prior shapes, by globally optimizing an orientation-based objective function in a discretized parameter space. Each prior shape is described by a normalized set of bound vectors (i.e., the length of each bound vector is 1), whose initial points are discrete points on the boundary of an object model, and directions are identical to normal vectors of the corresponding initial points with respect to the object boundary. Initial points can be picked evenly or uniformly randomly on an object boundary. The original priors, however, are binary images of objects, so that they are much more intuitive and straightforward to produce.

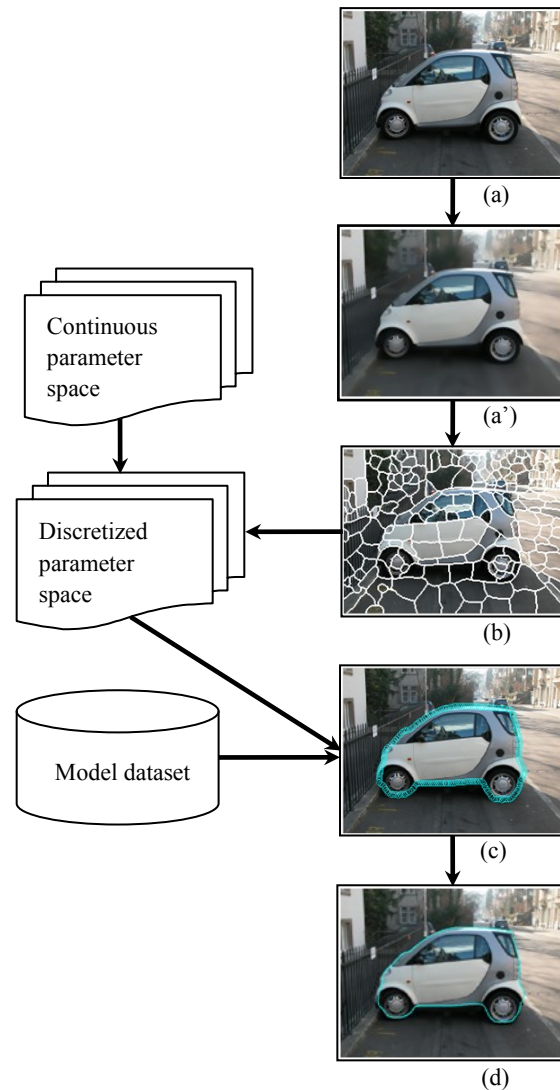


Fig. 1. The workflow of our method

Derived sets of bound vectors, then, are normalized into the unit square, and stored into a database as shape models. The greater the quantity of bound vectors that is chosen, the higher the accuracy of the target shape that can be described. Furthermore, besides not storing whole images of exemplars, our prior descriptor is easier to apply transformation compared to [10], which stores binary images of all feasible transformed exemplars. Indeed, we need to store only one normalized set of bound vectors, for each aspect of an object.

In a certain aspect, our object description method is similar to [12] which is primarily developed for the shape detection or identification of the object. While we use an ordered set of normal vectors of object boundary, [12] regarded curvature of the curve as the difference between adjacent tangent vectors. However, our approach relies on gradient orientation and magnitude while [12] worked on edge maps, where each curve in the derived edge map is considered as the shape of the object. It, then, is 'coded' in the form of CSS image. Consequently, CSS models of exemplars and CSS images are matched to find the closest pair. The problem of such methods is that edge map plays too large role, therefore it relies on the availability of exact edges in the image. However, in complex images, disrupted and merged edges are unavoidable. Also note that it targets on the identification of the shape, not the segmentation.

Global optimization is utilized to search through a parameter space for the most suitable configuration that makes a certain discrete model approximately fit the target object in the image. The dependency of transforming elements (i.e. translation, rotation, and scaling) on super-pixels generated by [13]-[14] is considered to expeditiously discretize the parameter space. Moreover, we propose an orientation-based objective function to measure the fitness of the transformed discrete model, and target object in the image. The objective function is designed to be compatible with the discretization phase. We utilized the well-known branch-bound global optimization algorithm to search for a solution. The consequent transformed model is linked to form a closed curve, which may need refining, by applying the optimization [15] in a small number of iterations.

The results tested on the PASCAL dataset [16] indicate that our method is robust at segmenting an object on a complex background, or with unclear boundary images. In the preliminary version of this paper [17], we proposed a simple, effective orientation-based object description, which is able to express arbitrary objects, together with an orientation-based objective function that measures the fitness of transformed object models and target objects in images. In addition, a robust solution space discretization was proposed.

In this paper, a preprocessing step using a bilateral filter is utilized to remove noise, but still preserve edges in images. Moreover, we demonstrate the effectiveness of orientation, by comparing segmented results generated with, and without, an orientation factor. The results show that when orientation is not utilized, not only is the dominance of the maximum objective value compared to others less, but also an inaccurate transformed model is chosen.

The next sections are organized as follows: our method is expressed in section 2; experiments are shown in section 3; and we make some conclusions in section 4.

2. Proposed Method

2.1 Overview

The overall workflow of the proposed method is described briefly in Fig. 1. Records of object models that describe shapes of objects are stored in a model dataset. The super-pixels in Fig. 1

(b) are generated by over-segmenting the input image, using [13]-[14]. The quantity of super-pixels should be small enough to discretize parameter space efficiently; it should also be large enough to include as much boundary of the target object as possible. Those super-pixels form an uneven grid, on which the discretization phase depends. Originally, the parameter space $(t_x, t_y, \lambda, \theta)$, which indicates x-, y-translation, scaling and rotation, is continuous. However, it includes a lot of infeasible configurations, because of the digitization of the image and characteristics of the object shape. Based on boundaries of super-pixels and the constraint of an aspect ratio (in 2.5), the parameter space is discretized efficiently, by merely considering feasible configurations. A branch-bound algorithm is applied to find the most suitable transformation configuration (TC). The resultant transformed model is shown in Fig. 1 (c). The tiny cyan circles represent the approximation (i.e. the parameter h) in the objective function (in section 2.4). Finally, the segmentation result is achieved after discrete points of the consequent model are chained and refined by the level set method [15], in a small number of iterations (Fig. 1 (d)).

The following subsections present in detail the preprocessing (2.2), the shape descriptor (2.3), orientation-based objective function (2.4), the discretization of parameter space (2.5), and a global optimization method (2.6).

2.2 Preprocessing by Bilateral Filtering

The object segmentation in section 2.5 relies on the oversegmentation of an input image. There are some factors that have negative influences on oversegmentation, which are noise and uncertain edges (i.e. where color or intensity changes). Moreover, the proposed object function, in section 2.4, utilizes gradient magnitude. Hence, removing uncertain edges enhances the fitness of the correct transformed model and target object in the image, compared to others.

The bilateral filter is a local, non-iterative, and simple method, which combines both feature and spatial information in the form of weight [12]. When applying a shift-invariant low-pass domain filter to an image, we have

$$h(x) = k_d^{-1} \sum_{\xi \in N} f(\xi) c(\xi - x) \quad (1)$$

where, $k_d = \sum_{\xi \in N} c(\xi - x)$, N is the neighboring region of x . Range filtering is defined as

$$h(x) = k_r^{-1} \sum_{\xi \in N} f(\xi) s[f(\xi) - f(x)] \quad (2)$$

where, $k_r = \sum_{\xi \in N} s[f(\xi) - f(x)]$. Combining geometric and photometric factors, we have the bilateral filter

$$h(x) = k^{-1} \sum_{\xi \in N} f(\xi) c(\xi - x) s[f(\xi) - f(x)] \quad (3)$$

where, $k = \sum_{\xi \in N} c(\xi - x) s[f(\xi) - f(x)]$.

2.3 Orientation-based Object Descriptor

The term ‘object’ is very general. It can be an arbitrary sort of visible thing, and can be found almost everywhere, in both the foreground and background of natural scene images. Despite

considering only the foreground of an image, we can realize a lot of objects existing individually or conjointly.

The scene of a little boy wearing a yellow t-shirt and handling a balloon aside, for instance, has at least three objects: the whole body of the boy, the yellow t-shirt, and the balloon. While the balloon stays separate, the body and t-shirt overlap each other. The question is, which thing is the target one? In some sense of subjective human mind, the body may be the most attractive one to detect or segment. However, is detecting or segmenting the t-shirt or the balloon wrong, when clues of the goal object do not exist? NDOS together with the implications of the human mind, in fact, is able to cause the ambiguity.

Object segmentation can be seen as either binary image labeling, or boundary finding. Those two views have a mutual relationship, and result in distinguishing object and background by one or many closed contours defining the plane shape (from a certain aspect) of the object. Shape, therefore, is an important factor to detect and segment an object. Use of the distinctness of colors, or the discontinuity of intensity, is eventually able to identify object shape. Besides, given this sort of high-level prior, we explicitly define the target object of segmentation. The ambiguity mentioned above can be prevented.

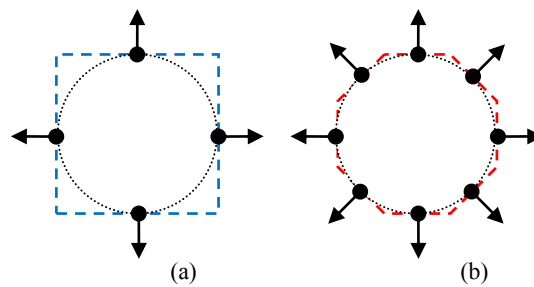


Fig. 2. Estimation of a circle by a square (a), and by an octagon (b)

Binary images are used for inputs to generate the database of normalized models (sets of bound vectors) are binary images, in which white pixels belong to the object area, and black pixels are in the background. Then, the boundary of the object is produced from its connected components, using [18]. Not all points on an object boundary are utilized. A set of discrete points are established by sampling points on the boundary evenly, in terms of the quantity of intermediate points. In the simpler case, they can be picked randomly, with uniform distribution. A record corresponding to each boundary point is a bound vector, whose initial point is the boundary point, and its normal vector, with respect to the boundary of the object. Hence, it is a 4-element vector, with the first two elements for the initial point, and the last two ones for the related normal vector. The quantity of records is chosen identically for all exemplars, so that the evaluation taking place later is effective. The number of bound vectors is proportional to the captured fineness of object shape. All those tasks are done in the pre-processing phase, which is separate from the segmentation process.

The use of normal vectors with discrete points provides an efficient and effective shape model for the object segmentation. It is simple in terms of computation and it can express significant structure of the object, using a set of bound vectors. Each bound vector consists of a boundary point, and its normal vector. The example in Fig. 2 expresses how bound vectors imply the structure of an object. In order to estimate the original circle, we can use 4 bound vectors, which imply a square, or 8 bound vectors, which imply an octagon. The more bound

vectors that are selected, the more accurate the estimation is. In other words, the number of bound vectors is proportional to the matching constraints.

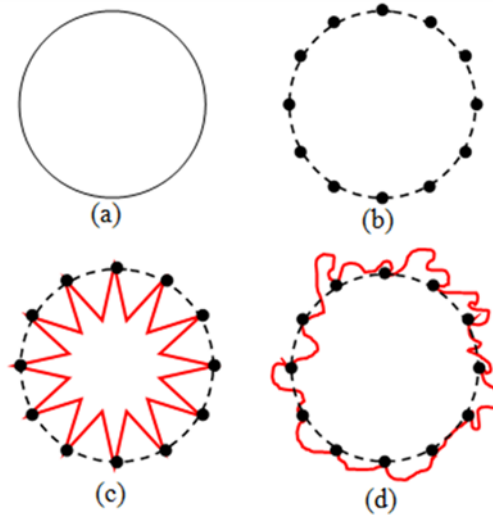


Fig. 3. The description (b) without an orientation feature can mismatch with red shapes (c) and (d).

The orientation factor fortifies the accuracy of object matching. In other words, random points have a low probability of mismatch to the object model. For instance, let us describe a circle without using orientation, as in **Fig. 3 (b)**. Not utilizing an orientation factor results in the mismatching of the model, as in **Fig. 3 (c)**, and **Fig. 3 (d)**. One of the factors reflecting a good object descriptor is the discrimination, i.e. it has to enhance the distinctness when a suitable transformed model fits and does not fit the target object in an image under a certain evaluation. In fact, such a factor is illustrated in **Fig. 5**, under our proposed objective function.

Exemplars are normalized into the unit square, and lists of derived bound vectors are stored as a prior dataset. Each exemplar corresponds to an object shape from a particular view point. Such relationship between transformation of view point, and set of bound vectors, is expressed by the function P

$$P: S \subset \mathbb{R}^4 \rightarrow D \subset \mathbb{R}^{2 \times N} \times \mathbb{R}^{2 \times N}$$

$$(t_x, t_y, \lambda, \theta) \mapsto v = \{(x_i, y_i), (u_i, v_i)\}_{i=1..N}$$

where, t_x, t_y are x- and y-translation, λ is scale, θ is rotating angle, (x_i, y_i) is initial point, and (u_i, v_i) is normal vector.

The exemplars can be extended to 3D object model with the proposed object descriptor. A number of sampled view points are chosen from the 3D model to create plane shapes of that object, by projecting it into the planes orthogonal to the view directions. This descriptor is followed by a suitable evaluation function, and a simple and effective global optimization framework.

2.4 Orientation-based Objective Function

To evaluate the fitness of each transformed model and target object in an image, an objective function is needed. The objective function maps sets of records, each of which consists of an initial point p and its normal vector d_p , into the set of real numbers

$$f : D \rightarrow \mathbb{R}$$

$$v = \{p_i, d_{p_i}\}_{i=1..N} \mapsto f(v)$$

where, N is the number of bound vectors. To build the objective function, we rely on potential boundaries created by [13]. The over-segmentation process generating super-pixels not only establishes regions of homogenous-colored pixels, but also forms potential boundaries of objects (i.e. pixels are potentially the boundary of the target object). The objective f is to find which model and its appropriate TC most fit a certain combination of potential boundaries, in terms of position and orientation. f is expressed as

$$f(v) = \sum_{i=1}^N \frac{1}{Q_{p_i+1}} \sum_{q \in PB} [\tau(q) K\left(\frac{p_i - q}{h}\right) |\cos(d_{p_i}, d_q)|] \quad (4)$$

where,

$$Q_{p_i} = \sum_{q \in PB} \tau(q) K\left(\frac{p_i - q}{h}\right), \quad (5)$$

$$\tau(q) = \begin{cases} 1 & \text{if } G_q \geq \epsilon_G \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$K\left(\frac{p_i - q}{h}\right) = \begin{cases} 1 & \text{if } \left\| \frac{p_i - q}{h} \right\| \leq 1, \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

PB is the set of pixels belonging to potential boundaries, h is the radius of circles whose centers are p_i 's, N is the quantity of bound vectors, G_q is the gradient magnitude of pixel q , and ϵ_G is a threshold of gradient magnitude.

The objective function f measures the summation of the averages of cosine similarity between each bound vector (p_i, d_{p_i}) of the transformed model and bound vectors, whose initial points are in its neighboring circle. In other words, f evaluates the correspondences, in terms of the orientation between the bound vectors, and the orientation trend of pixels in their neighborhoods. Our objective is to find an exemplar, and its corresponding TC, such that the function f is maximized

$$[n^*, k^*] = \arg_{n,k} \max_{n,k} f(v_k^n) \quad (8)$$

where, n is the index of the exemplar, and k is the index of the TC.

The objective function f aims to not only match point-to-point, which was done in [12], and [20], but it also seeks for a suitable discrete structure, based on orientation. Therefore, random noise points are restricted, to affect the measuring. In Fig. 5, we list all feasible transforming

configurations (t_x, t_y, λ) (rotating angle θ is fixed to 0) of the proper model of the red car, and their value computed by f . In order to enable the showing of the correlation of those variables, they are split into three 2-combinations. The maximum value of f in each combination is outstanding, in comparison with others. In addition, we demonstrate the need of orientation in f , by comparing how dominant the maximum value of f is, in the case of f with orientation, and without orientation. The comparison is exhibited in Fig. 6 and Fig. 7.

2.5 Parameter Space Discretization

The objective function can be given as follows by combining with the map P :

$$f(v_s) = f(\{(p_i, d_{p_i})\}_{i=1..N}) = f(P(s)) \tag{9}$$

where, $s \in S$. The actual variable of the objective function is the TC $(t_x, t_y, \lambda, \theta)$. Because models capture object shapes in many aspects, it is not necessary to have two separate scaling parameters for x- and y-coordinates. Therefore, we use only one variable λ , to indicate the scaling of the model.

While the transforming elements are treated individually in [21], the optimal result can be achieved by globally optimizing over the entire parameter space. However, there is a tradeoff between them. A potential set of TCs is very huge, and it is infeasible to consider all of them. The purpose of this phase is to obtain a feasible set of TCs, which is much smaller than the potential set, based on the invariance of the aspect ratio given a certain rotating angle.

Proposition 1 proves the invariance of aspect ratio of model when it is rotated by a fixed angle θ . Due to result of oversegmentation, a model transformed by an appropriate TC can fit into the super-pixel set which forms target object in image. Therefore, we only consider maximum and minimum of (x, y) coordinates of super-pixels. Let T, R, L and B be the set of top points, right points, left points, and bottom points respectively. One 4-point record is feasible if it contains 4 points $t \in T, r \in R, l \in L$, and $b \in B$ such that $\frac{|t_y - b_y|}{|r_x - l_x|}$ is equal to a specified aspect ratio C ; and t, r, l , and b belongs to corresponding sides of derived rectangle (Fig. 4.b and Fig. 4.c). Feasible set consists of all feasible 4-point records. In order to find 4-point records, proposition 2 firstly collects all feasible left-right pairs. Then, proposition 3 finds corresponding top or left points. In practical, we merely use 3 points to determine a rectangle (i.e., left-right-top or left-right-bot).

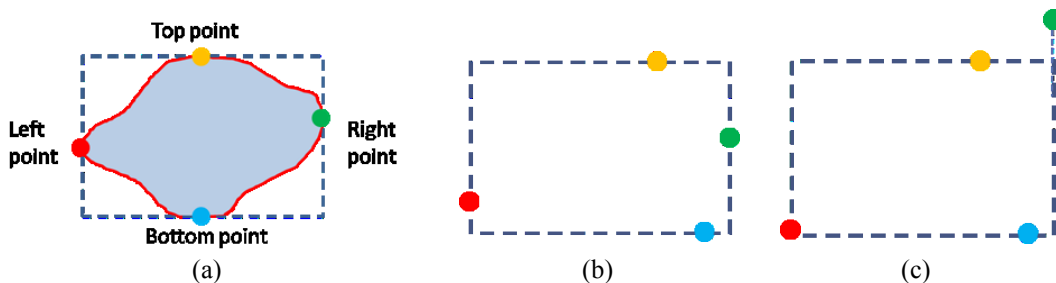


Fig. 4. (a) 4 optimal points of a super-pixel; (b) a feasible 4-point record; (c) an infeasible 4-point record.

Proposition 1

Given a rotating angle θ , the aspect ratio of the model that is transformed by arbitrary translation and scaling configuration (t_x, t_y, λ) is constant.

For each rotating angle $\theta \in [0; 2\pi)$, translation and scaling are discretized into high potential sets of values. We suggest utilizing proposition 1, and the outcome of over-segmentation, to discretize the parameter space. Super-pixels (patches) generated by [13] form an uneven grid, which plays the role as the basis of discretization. We consider the uneven grid, to find potential configurations of translation and scaling.

Definition

Given a set of points rotated by a certain angle θ . Let C be the constant aspect ratio of those points. A rectangle in the image is called a potential bounding box (with respect to θ), if its aspect ratio is equal to C .

A list of all potential bounding boxes implies a list of translations, and a list of scaling values (with respect to θ). After being over-segmented by [13], the input image is subdivided into many super-pixels (patches), based on color distinctness and size constraint. The minimum and the maximum x- and the y-coordinates of super-pixels generate an uneven grid. By choosing a significant large number of super-pixels, the boundaries of all super-pixels can cover approximately the boundary of the target object. We let potential bounding boxes snap into the uneven grid, such that its minimum x-coordinate x_{l_i} corresponds to the minimum x-coordinate of super-pixel i , its maximum x-coordinate x_{r_j} corresponds to the maximum x-coordinate of super-pixel j , and its minimum y-coordinate y_{t_k} corresponds to the minimum y-coordinate of super-pixel k . In the following propositions, we temporarily ignore the constraint of image size (i.e. the circumstance in which potential bounding boxes exceed the range of the input image is temporarily accepted). Violating ones will be eliminated afterward.

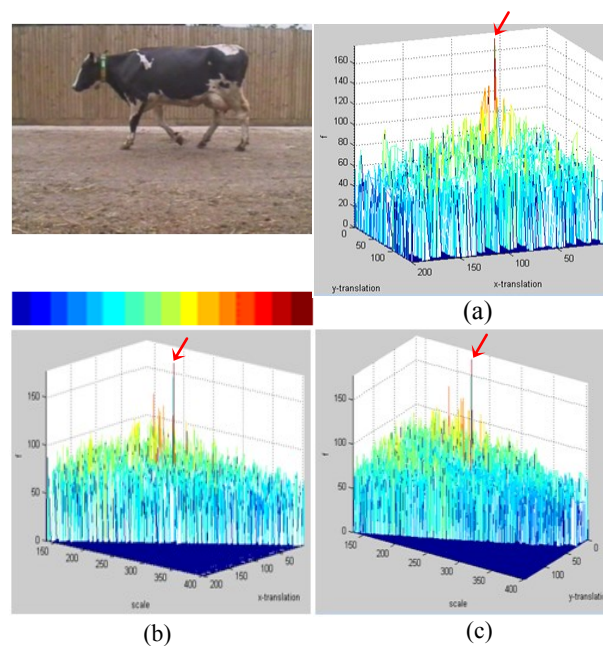


Fig. 5: (a) x- and y-translation; (b) x-translation and scale; (c) y-translation and scale.

Proposition 2

Ignore the constraint of image size. If there are two points $\forall(x_1, y_1), (x_2, y_2): |y_2 - y_1| \leq C|x_2 - x_1|$, then x_1 and x_2 are the bounds of the x-coordinates of a certain potential bounding box. (C is the aspect ratio of the normalized model (with respect to angle θ)).

Another proposition following proposition 2, is to completely show how to search for potential bound boxes; as a result, configurations of translation and scaling corresponding to a rotating angle θ are listed discretely.

Proposition 3

Ignore the constraint of image size.

If $\forall(x_1, y_1), (x_2, y_2): |y_2 - y_1| \leq C|x_2 - x_1|$, then $\forall y \in [-C\Delta x + y_M, C\Delta x + y_m], \exists x_0: (x_0, y)$ belongs to the top or bottom of a certain potential bounding box (where $\Delta x = |x_2 - x_1|$, $y_m = \min(y_1, y_2)$, and $y_M = \max(y_1, y_2)$).

According to the constant aspect ratio (in proposition 1) when given a rotating angle, potential bounding boxes of transformed models in the image are listed discretely, due to proposition 2 and proposition 3. Since the measuring of orientation in an objective function regards an h-radius circle, the set of y-coordinate values in proposition 3 can be enumerated as

$$y_k = y_m + k \frac{y_M - y_m}{h + 1}, k = \overline{0, h + 1}. \tag{10}$$

Also, it does not require having all partial edges of the target object. The necessary portions are the left, the right, and the top of the target object. For that reason, we choose over-segmentation [13]-[14], rather than Canny [18].

We utilize the branch-bound algorithm to find the global optimum in the discretized parameter space. In Fig. 5, based on the color map, in which values corresponding to the colors are increasing from left to right, blue areas indicate ignored sets of transforming configurations that are implicitly removed by the propositions, and the constraint of image size. It shows that our propositions are effective in discretizing the parameter space, by disregarding non-potential transforming configurations. In the combination of x-translation and scale (Fig. 5 (b)), there exist huge sets of ignored configurations. It is much more than the ones in Fig. 5 (a) and Fig. 5 (c), because the aspect ratio of the target object is small (its width is larger than its height).

2.6 Branch-Bound for Global Optimization

A global optimization is taken to find the optimal solution for the objective function in the discretized parameter space. Among several techniques for the global optimization, such as graph-cut [4], branch-bound [22], and space mapping [22], branch-bound is chosen because of its efficiency and capability. Branch-bound is a specialized discrete programming technique that provides a strategy to avoid full search through the entire solution space [22]. In the proposed object segmentation scheme, the solution space is a 4-D space of transformations (i.e. rotation, scale, x-, and y-translation). Different from [10], the objective of our method is to build based on discrete sets of bound vectors, and to evaluate the fitness based on the matching of boundary structure (position and orientation), rather than pixel labels.

Branch-bound is a well-known algorithm in discrete programming. It has the ability to get rid of exhaustively searching the whole solution space, by fathoming sub-spaces that are not feasible to contain the optimal solution. The two main factors of the branch-bound algorithm are a branching (subdivision) strategy, and a bounding function. The discretized parameter space is recursively divided into sub-spaces, each of which has a bounding value, reflecting its priority for being chosen. In our maximization problem, a bounding function g is designed to be an upper bound of the function f , such that

- $g(v) \geq f(v), \forall v$
- $\hat{g}(\mathbb{V}_1) \geq \hat{g}(\mathbb{V}_2)$ if $\mathbb{V}_1 \supseteq \mathbb{V}_2$
- $\exists v \in \mathbb{V}: |\hat{g}(\mathbb{V}) - f(v)| < (1 - \epsilon)\hat{g}(\mathbb{V})$, if \mathbb{V} is a ‘leaf’ of a searching tree.

where, $\hat{g}(\mathbb{V})$ is an upper bound of g in sub-space \mathbb{V} : $\exists v' \in \mathbb{V}: g(v') \equiv \hat{g}(\mathbb{V}) \geq g(v), \forall v \in \mathbb{V}$. Thus, g tends to approach f from above, when we branch the discretized parameter space progressively. Sub-spaces that have the largest bounds in comparison with others in the same level are eliminated, together with their children. The branching process can be terminated, if g is close enough to f . That is how the branch-bound algorithm can ease the size of the solution space.

The function g is supposed to be more straightforward to compute, compared to f . Consider the objective function

$$\begin{aligned} f(v) &= \sum_{i=1}^N \frac{1}{Q_{p_i+1}} \sum_q \tau(q) K\left(\frac{p_i-q}{h}\right) |\cos(d_{p_i}, d_q)| \\ &\leq \sum_{i=1}^N \frac{1}{Q_{p_i+1}} \sum_q \tau(p_i) \cdot K\left(\frac{p_i-q}{h}\right) \\ &\leq \sum_{i=1}^N \max_{q \in PB} \left[\tau(p_i) \cdot K\left(\frac{p_i-q}{h}\right) \right] = g(v), \forall v. \end{aligned} \quad (11)$$

The first inequality is for the property of the cosine function; the second is proven in the following proposition.

Proposition 4

$$\sum_{i=1}^N \frac{1}{Q_{p_i+1}} \sum_q \tau(p_i) \cdot K\left(\frac{p_i-q}{h}\right) \leq g(v)$$

Hence, g is the upper bound of f for all TCs. It counts the quantity of initial points (of the transformed model) that are close to an arbitrary potential boundary. The orientation factor is relaxed in g . At the same level, if $\hat{g}(\mathbb{V}_1) \geq \hat{g}(\mathbb{V}_2)$, \mathbb{V}_1 has more chance to be fathomed, and \mathbb{V}_2 has more chance to be picked soon.

The function g can also be viewed as

$$g(v, h) = \sum_{i=1}^N \max_{q \in PB'} [|p_i - q| \leq h] \quad (12)$$

where, $PB' = \{q \in PB | G_q > \epsilon_G\}$, and $[\cdot]$ is the predicate function that is 1, if the condition is true, and 0, if otherwise. $\hat{g}(\mathbb{V})$ can be calculated by choosing v_0 , and loosening the radius of the neighboring area of boundary points, such that $\hat{g}(\mathbb{V}) \equiv g(v_0, h + \delta) \geq g(v, h), \forall v$ (see [20]).

Even though the orientation-based objective function f is quite straightforward, it takes

more time to compute, compared to g . From the view of (9), g simply counts the quantity of the intersect sets of a small ball $B(p_i, h)$ and PB' . Thus, evaluating g is simpler and faster. In practice, it is about 10 times faster than evaluating f .

The branch-bound algorithm is applied for each exemplar, whose details can be found in [17].

2.7 Refinement












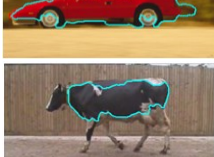
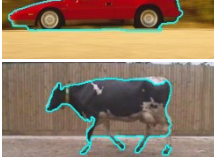
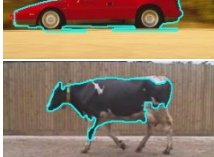
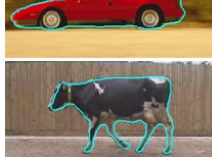





After applying the branch-bound algorithm to the discretized parameter space, we obtain the model and its corresponding TC that generates the fittest discrete ‘skeleton’ to the boundary of the target object. A continuous contour is created, by orderly chaining the resultant set of points. Because of the approximation (i.e. measuring the fitness of orientation in an h -radius circle) of the objective function, the contour is not completely accurate. In order to refine it, we utilize [15] to steadily deform the contour to a high gradient magnitude curve in a small number of iterations. When working on a complex background and/or low contrast images, the number of iterations should be very few, so that the structure of the resultant transformed model is preserved.

3. Experimental Results

The proposed method was implemented in Matlab version 7.11.0 (2010b). The configuration of the testing system is an Intel® Core™ i5-2500 CPU @ 3.30GHz 3.30GHz, and 4GB RAM. For evaluation, PASCAL and ETHZ datasets [16,24] are used, which provide both sample images and ground truths. Models (exemplars) are generated by extracting edges in ground truth images, then picking a specified number of boundary points, and collecting their normal vectors. The size of a model is chosen to be one third of the longest boundary. They are then normalized into the unit cube, and stored.

Super-pixels are generated using the method in [13], and the number of super-pixels is set to 200, which is large enough to cover most of the target object boundary. Then, adjacent

Table 1. Some examples to compare our method to the level set based method [15], GrabCut [2], and saliency based method [7].

Original image	LS [16]	GrabCut [1]	GC [6]	Our method
				
				
				
				

super-pixels that are similar in color (i.e. the difference of means of two adjacent ones is significantly small, compared to the sum of their standard deviations) are merged together.

The branch-bound algorithm is implemented, based on the best-first search. In contrast, the one having the smallest upper bound has the highest chance to be subsequently considered. The radius $h = 2$. The effectiveness of the refinement step is dependent on how different the boundary between the target object and background is. We only run from 3 to 8 iterations of [15].

Our method is robust at detecting and segmenting objects, especially rigid, because it relies on whole instances of object shape, and solves the problem globally. However, the payoff is that it needs many models (exemplars) of many types of objects and aspects. The average run-time for each exemplar is about 11 seconds. The relaxation variable h of the objective function decides how much variation of object, in terms of shape in image compared to models, the method can handle. Hence, rigid objects require less exemplars, than non-rigid objects do.

Firstly, we demonstrate results tested on PASCAL dataset [16]. **Table 1** shows some improvements of our method, compared to the level set based method (LS) [15], GrabCut [2] (not including interactive foreground and background editing), and the global contrast based method (GC) [7], due to its shape-based property. Initials of the LS are rectangles inside those target objects. In those examples, the boundary of object and background is not clear; as a result, they run 1410 iterations, without converging. The GC is strong at segmenting images in which the contrast between object and background is clear (e.g. the red car, and the yellow car in **Table 1**).

To evaluate the essence of orientation in the proposed objective function, we regard how dominant the maximum of f , which corresponds to the resultant configuration, is in comparison with others, in two opposite cases. In the first case, we use the proposed function (i.e. $f_1 = f$), and the second case is the same, except for getting rid of the orientation.

$$f_2(v) = \sum_{i=1}^N \frac{1}{Q_{p_i+1}} \sum_{q \in PB} [\tau(q) K \left(\frac{p_i - q}{h} \right)]. \quad (13)$$

Only correct models and appropriate rotating angles of target objects are considered; therefore, translation and scaling vary. Values of f_1 and f_2 are computed for a discretized set T of translations and scales. The dominance is measured by



Fig. 6. Above – Segmentation using f_2 ; below – Segmentation using f_1 .

$$D_k = \frac{1}{|T|-1} \sum_{t \in T} [\max_t f_k - f_k(t)] \tag{14}$$

where, $k \in \{1,2\}$.

According to Fig. 7 and the tested images, the percentage differences \bar{D} between D_1 and D_2 of those cases are 21.36%, 30.15%, 30.34%, and 47.67%. The difference \bar{D} increases, when the background is more complicated. The reason is that there are many noise points in complex background images (e.g., the silver and the yellow cars in Table 1 that have much influence on f_2 (not including orientation). Thus, f_1 with orientation measuring can enhance the solution stronger than f_2 does.

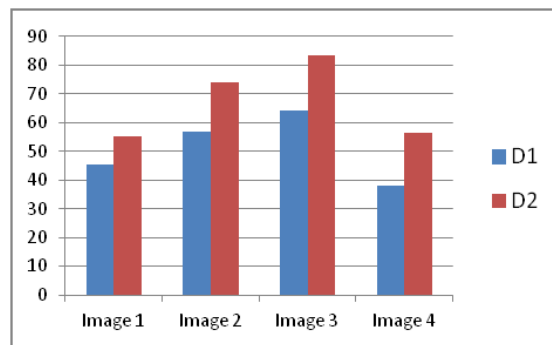


Fig. 7. Values of D_1 , and D_2 of images in Fig. 6.

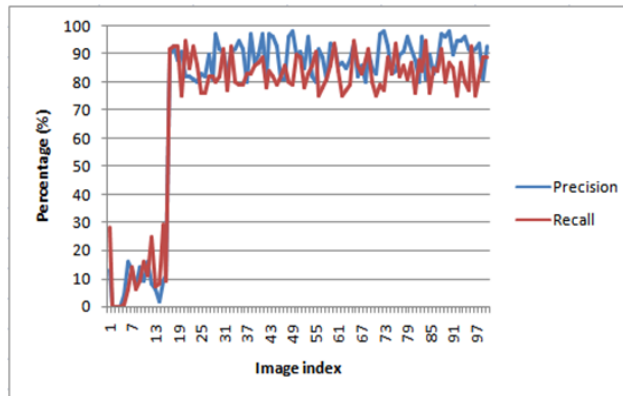


Fig. 8. Precision and recall of car images in PASCAL 2005 dataset.

Due to the fact that our approach finds the best TC for a certain model to fit into target object, improper TC may result in very low precision and recall. Therefore, there is a big gap between two major regions in Fig. 8. In 100 car images of PASCAL 2005, 82% of images exceeds 75% in terms of precision and recall. The average precision is 75.5%, and the average recall is 73.6%. Precision and recall are computed by following formula.

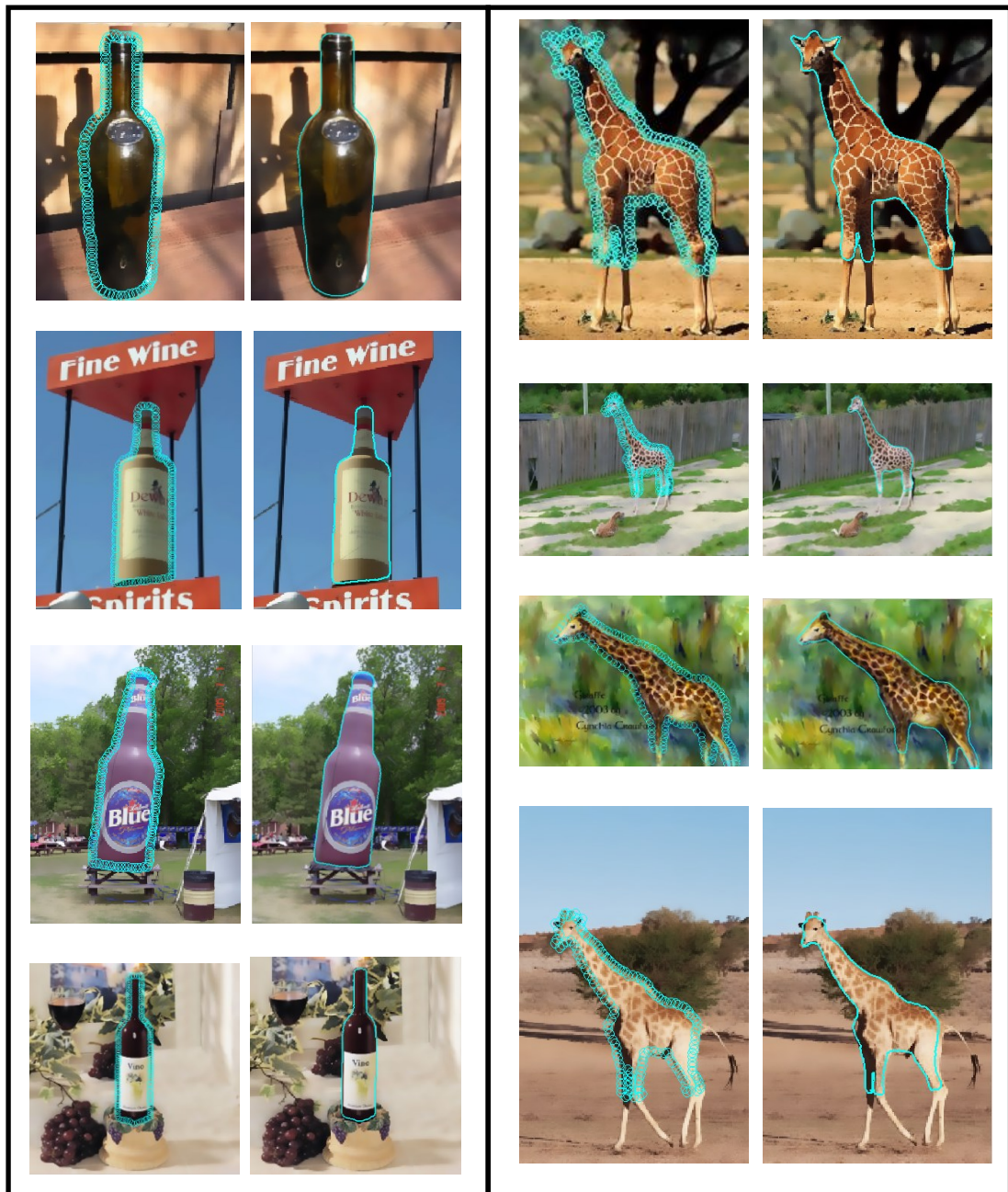


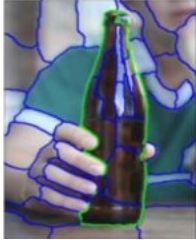
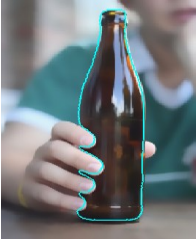


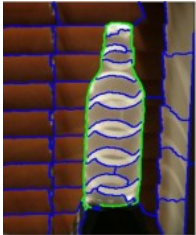







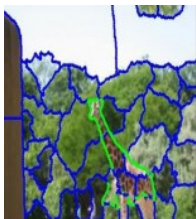

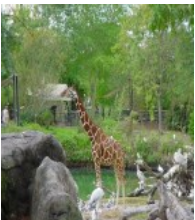

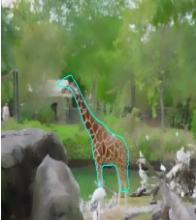


Fig. 9. Bottles and giraffes in ETHZ dataset segmented by our method. Images on the left of each column are transformed models with discrete points and their h-radius circles. Images on the right of each column are final results.

Table 2. Some examples to compare our method to branch-mincut [10], and Chordigram [11].

Original image	Branch-mincut [10]	Chordigram [11]	Our method
			
			
			
			
	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Fail</div>		

The source code of [10] is from the author's homepage. Its runtime varies from 29 to 102 seconds depending on image characteristic. The outputs of Chordigram are from [11].

Table 3. Run time of branch-mincut [10] with images in the ETHZ dataset.

Images	Size	Runtime (s)
tobias	327x495	77.9
spiral	375x500	62.75
mino	500x353	29.94
brookfield	500x375	55.01
blue3	375x500	102.44

$$Precision = \frac{TF}{TF+FP} \quad (15)$$

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

where TP is the overlapping area of transformed model and target object, FN is the area belonging to target object but not transformed model, and FP is the area belonging to transformed model but target object.

Besides, our method achieves considerable result when tested on ETHZ dataset (**Fig. 9**). Bottle class, rigid object, and giraffe class, non-rigid object, are chosen for evaluation. [10] inspired us in utilizing global optimization, while [11] raised the good idea of relying on superpixels. However, different from [11], which regards superpixels in the aspect of regions, our method concerns about seeking for partial superpixel boundaries which constitute a whole target object. We compared results of these 3 methods in **Table 2**.

4. Conclusions

A novel method for the predefined object segmentation was presented, by using global optimization in a discretization parameter space. By demonstrating the order of transformation elements with the derived constraints and super-pixels, the parameter space is discretized. An orientation-based objective function, which is capable of measuring the fitness of transformed models and objects, was proposed. The robustness of the orientation-based objective function turned out to overcome situations in which some partial edges are lost. The reason is that it recognizes an object as a whole instance. Our method is able to detect and segment objects, especially rigid and multi-color objects, because it relies on the whole instance of an object shape, and solves the problem globally.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST)(2012-047759 and 2013-006535). Also this research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA(National IT Industry Promotion Agency).

References

- [1] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active contour models," in *Proc. of ICCV'87*, 1987, p. 259-267. [Article \(CrossRefLink\)](#).
- [2] C. Rother, V. Kolmogorov and A. Blake, "GrabCut: interactive foreground extraction using iterated graph cuts," *ACM SIGGRAPH Journal*, vol. 23, no. 3, 2004, pp. 309-314. [Article \(CrossRefLink\)](#).
- [3] Y. Li, J. Sun, C.K. Tang and H.Y. Shum, "Lazy snapping," *ACM SIGGRAPH Journal*, vol. 23, no. 3, 2004, pp. 303-308. [Article \(CrossRefLink\)](#).
- [4] Y. Boykov, O. Veksler and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *IEEE PAMI Transaction*, vol. 23, no. 11, Nov. 2001, pp. 1222-1239. [Article \(CrossRefLink\)](#).
- [5] L. Itti, C. Koch and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE PAMI Transaction*, vol. 20, no. 11, Nov. 1998, pp. 1254-1259. [Article \(CrossRefLink\)](#).
- [6] K. Fukuchi, K. Miyazato, A. Kimura, S. Takagi and J. Yamato, "Saliency-based video segmentation with graph cuts and sequentially updated priors," in *Proc. of ICME'09*, 2009, p. 638-641. [Article \(CrossRefLink\)](#).
- [7] M.M. Cheng, G.X. Zhang, N.J. Mitra, X. Huang and S.M. Hu, "Global contrast based salient region detection," in *Proc. of CVPR '11*, 2011, p. 409-416. [Article \(CrossRefLink\)](#).
- [8] P. Mehrani and O. Veksler, "Saliency Segmentation based on Learning and Graph Cut Refinement," in *Proc. of BMVC'10*, 2010, p. 1-12. [Article \(CrossRefLink\)](#).
- [9] D.H. Ballard, "Generalizing the hough transform to detect arbitrary shapes," *PR Journal*, vol. 13, no. 2, Jan. 1981, pp. 714-725. [Article \(CrossRefLink\)](#).
- [10] V. Lempitsky, A. Blake and C. Rother, "Branch-and-Mincut: Global Optimization for Image Segmentation with High-Level Priors," *MIV Journal*, vol. 44, no. 3, Nov. 2012, pp. 315-329. [Article \(CrossRefLink\)](#).
- [11] A. Toshev, B. Taskar and K. Daniilidis, "Shape-Based Object Detection via Boundary Structure Segmentation," *CV Journal*, vol. 99, no. 2, Sep. 2012, pp. 123-146. [Article \(CrossRefLink\)](#).
- [12] S. Abbasi, F. Mokhtarian and J. Kittler. "Curvature scale space image in shape similarity retrieval," *Multimedia Syst.* 7, 6, Nov. 1999, pp. 467-476. [Article \(CrossRefLink\)](#).
- [13] J. Shi, and J. Malik, "Normalized Cuts and Image Segmentation," *PAMI Journal*, vol. 22, no. 8, Aug. 2000, pp. 888-905. [Article \(CrossRefLink\)](#).
- [14] J. Malik, S. Belongie, T. Leung and J. Shi, "Contour and Texture Analysis for Image Segmentation," *CV Journal*, vol. 43, no. 1, Jun. 2001, pp. 7-27. [Article \(CrossRefLink\)](#).
- [15] C. Li, C. Xu, C. Gui and M.D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IP Transaction*, vol. 19, no. 12, Dec. 2010, pp. 3243-3254. [Article \(CrossRefLink\)](#).
- [16] The homepage of PASCAL visual object classes: <http://pascalvin.ecs.soton.ac.uk/challenges/VOC/>
- [17] N. N. Nguyen, H. R. Kim, J. S. Cha, T. K. V. Le and G.S. Lee, "Global Optimization in Discretized Parameter Space for Predefined Object Segmentation," in *Proc. of ICUIMC '13*, 2013. [Article \(CrossRefLink\)](#).
- [18] J.F. Canny, "A Computational Approach to Edge Detection," *PAMI Journal*, vol. 8, no. 6, Nov. 1986, pp. 679-698. [Article \(CrossRefLink\)](#).
- [19] T.M. Breuel, "Implementation techniques for geometric branch-and-bound matching methods," *CVIU Journal*, vol. 90, no. 3, Jun. 2003, pp. 258-294. [Article \(CrossRefLink\)](#).
- [20] S. Belongie, J. Malik and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *PAMI Transaction*, vol. 24, no. 4, Apr. 2002, pp. 509-522. [Article \(CrossRefLink\)](#).
- [21] T.M. Breuel, "Fast recognition using adaptive subdivisions of transformation space," in *Proc. of CVPR '92*, 1992, p. 445-451. [Article \(CrossRefLink\)](#).
- [22] A. H. Land and A. G. Doig, "An Automatic Method of Solving Discrete Programming Problems," *Econometrica Journal*, vol. 28, no. 3, 1960, pp. 497-520. [Article \(CrossRefLink\)](#).

- [23] M. H. Bakr, J.W. Bandler, K. Madsen and J. Søndergaard, "Review of the space mapping approach to engineering optimization and modeling," *OE Journal*, vol. 1, no. 3, pp. 241-276. [Article \(CrossRefLink\)](#).
- [24] The homepage of ETHZ dataset: <http://www.vision.ee.ethz.ch/datasets/>



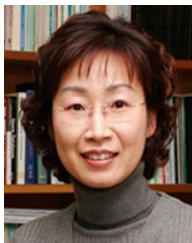
Huy Hoang Nguyen received the B.S. degree in Mathematics and Computer Science from the University of Science Ho Chi Minh City, Vietnam in 2009, and M.S. degree in Electronics and Computer Engineering from the Chonnam National University, Korea in 2013. His interesting studies are in multimedia and image segmentation, document processing, and pattern recognition.



GueeSang Lee received the B.S. degree in Electrical Engineering and the M.S. degree in Computer Engineering from Seoul National University, Korea in 1980 and 1982, respectively. He received the Ph.D. degree in Computer Science from Pennsylvania State University in 1991. He is currently a professor of the Department of Electronics and Computer Engineering in Chonnam National University, Korea. His research interests are mainly in the field of image processing, computer vision and video technology.



Soo-Hyung Kim received his B.S. degree in Computer Engineering from Seoul National University in 1986, and his M.S. and Ph.D degrees in Computer Science from Korea Advanced Institute of Science and Technology in 1988 and 1993, respectively. From 1990 to 1996, he was a senior member of research staff in Multimedia Research Center of Samsung Electronics Co., Korea. Since 1997, he has been a professor in the Department of Computer Science, Chonnam National University, Korea. His research interests are pattern recognition, document image processing, medical image processing, and ubiquitous computing.



Hyung Jeong Yang received her B.S., M.S. and Ph. D from Chonbuk National University, Korea. She is currently an associate professor at Dept. of Electronics and Computer Engineering, Chonnam National University, Gwangju, Korea. Her main research interests include multimedia datamining, pattern recognition, artificial intelligence, e-Learning, and e-Design.