From Logo to Object Segmentation

Fanman Meng, Hongliang Li, *Senior Member, IEEE*, Guanghui Liu, and King Ngi Ngan, *Fellow, IEEE*

***Abstract*—This paper proposes a method to segment object from the web images using logo detection. The method consists of three steps. In the first step, the logos are located from the original images by SIFT matching. Based on the logo location and the object shape model, the second step extracts the object boundary from the image. In the third step, we use the object boundary to model the object appearance, which is then used in the MRF based segmentation method to finally achieve the object segmentation. The key of our method is the object boundary extraction, which is achieved by searching a variation of the shape model that best fits the local edge of the image. Affine transform is used to consider the variations among the objects. Meanwhile, the Nelder-Mead simplex method with a simple initial rough search is used to run the boundary search. To verify the proposed method, we collect a LogoSeg dataset from the web such as Flickr and Google. The MOMI dataset is also used for the verification. The experimental results demonstrate that the proposed logo detection based segmentation method can improve the performance of the object segmentation.**

***Index Terms*—Specific object segmentation, logo detection.**

I. Introduction

IMAGE segmentation is a fundamental task in many high-level computer vision tasks, such as scene understanding [1], image retrieval [2], and object recognition [3]. In the past few decades, many image segmentation methods have been proposed, which can be roughly classified into two categories: non-semantic object segmentation [4], [5] and semantic object segmentation. The first one intends to extract some uniform and homogeneous regions from the image with respect to texture or color properties, such as superpixels based segmentation method [4], [5]. The second one instead extracts the semantic object from the images, which can provide the semantic prior of the object and benefit the high-level computer vision tasks. Meanwhile, semantic object segmentation remains challenging due to the difficulty of generating the semantic concepts by low-level features.

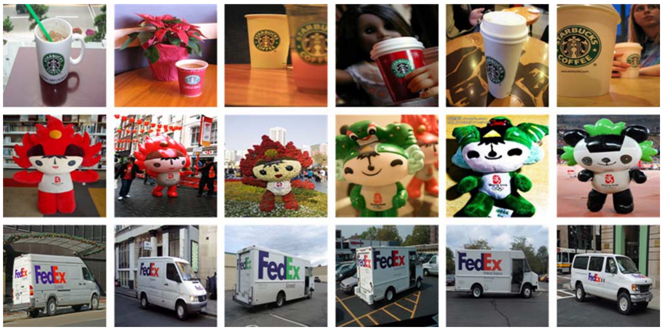


Fig. 1. Examples of the specific object (commodity) segmentation. *Starbucks*, *Fuwa* and *FedexTrucks* are shown in the first, second and third rows, respectively. The corresponding logos are shown in the last column.

Specific object segmentation is a semantic object segmentation method, which aims to extract a specific object from the images. For example, the commodities (the specific objects) may be required to be segmented from the retrieved images as shown in Fig. 1. The existing unsupervised or weakly supervised segmentation methods have been used to achieve the specific object extraction, such as co-segmentation [6]–[12] and weakly supervised object segmentation [13]–[15]. In these methods, the object prior is implicitly expressed by assuming that the object is contained in every image. In addition to these unsupervised or weakly supervised segmentation methods, the learning based method segments the specific object by first learning the object prior in a supervised manner and then applying the prior on the new image to conduct the specific object extraction [16]–[21]. Compared with the unsupervised manner, the learning based method can learn more accurate object prior and result in better segmentation results. This paper focuses on learning based specific object segmentation method.

Two problems are usually considered in the learning based specific object segmentation method. One is how to exactly construct the object prior from the training data. Considering there are many object variations among the objects, such as the changes of rotation, scale, translation, pose and shape, the generated object prior ought to be robust to these variations. The other problem is how to successfully detect and segment the object from the complex backgrounds. This is challenging due to the variations of the objects and the complexities of the backgrounds. Some examples can be found in Fig. 1, where there are many changes among the objects in *Fuwa*, such as the rotation, the scale and the poses. Meanwhile, the backgrounds may be complex, such as the street scene in *FedexTrucks*.

It is also observed from Fig. 1 that the logo usually stays fixed although the objects change significantly. The reason is that the logo is the identity of a brand and should be fixed to identify the referring commodity. Since the logos

Manuscript received September 17, 2012; revised February 24, 2013; accepted May 31, 2013. Date of publication September 05, 2013; date of current version November 13, 2013. This work was supported in part by the National Natural Science Foundation of China under Grants No. 61271289, The Ph.D. Programs Foundation of Ministry of Education of China under Grants No. 20110185110002, and the Fundamental Research Funds for the Central Universities under Grants E022050205. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wenwu Zhu.

F. Meng, H. Li, and G. Liu are with School of Electronic Engineering, University of Electronic Science and Technology of China, Cheng Du 610073, China (e-mail: fanmanmeng@gmail.com; hlli@uestc.edu.cn; guanghuiliu@uestc.edu.cn).

K. N. Ngan is with Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong, and also with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 610073 China (e-mail: knngan@ee.cuhk.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2013.2280893

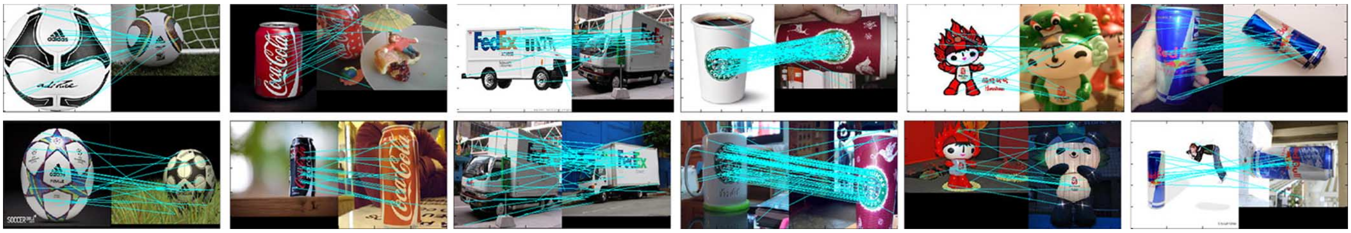


Fig. 2. The matching results by SIFT features. Six classes are shown. For each class, a pair of images randomly selected from the image group is shown.

usually share similar local texture, they can be easily matched by the existing local region matching method. Hence, logo extraction is easier than the object level based segmentation. Some of the matching examples are shown in Fig. 2, where the results of the six image pairs are displayed. SIFT features are used for the matching. We can see that the logos are successfully matched from the image pairs, while there are many matching mistakes on the object. Note that the matched logos can provide much information of the objects, such as the location and the local appearance of the objects, which can simplify the object extraction. Furthermore, the boundary of the objects can be located through learning the relationship between the logo and the object, which results in a new shape model and more accurate object segmentation.

By extending logo to be a more general concept such as the similar local region, the logo based segmentation method can be applied to many specific object extraction tasks, such as detecting and segmenting objects with specific characteristics from complex backgrounds or tracking a specific object from the video based on fixed local features. Besides apparent applications in image and video editing, the proposed method also can be used for several potential applications, such as content-based image retrieval [42].

In this paper, we propose a logo based object segmentation method, which consists of three steps (first reported in [22]). The first step is to detect the logos from the original images by logo matching. SIFT feature is used for the matching. In the second step, we extract the object boundary based on the logo and the object shape prior. Considering the changes between the object and the object shape prior, we take the shape variations into account. For each image, we vary the shape model by affine transform and search the transformed shape model that best fits the local edge of the image as the object boundary. In the third step, we use the regions inside and outside the boundary to form the unary term of the Markov Random Field (MRF) based energy function. The objects are finally segmented by the graph-cuts algorithm. We collect the LogoSeg datasets from the web to verify the proposed method. MOMI dataset is also used for the verification. The experimental results demonstrate the effectiveness of the proposed method.

The rest of the paper is organized as follows. The related work is introduced in Section II. In Section III, we present our proposed logo based object segmentation method. Experimental results are provided in Section IV to validate the efficiency of the proposed model. We discuss the proposed method in Section V. Finally, Section VI gives the conclusion of this paper.

II. Related Work

In the semantic object segmentation, an accurate object prior is required to distinguish the object from the backgrounds. In the existing methods, the object prior is usually generated by two methods: supervised method and unsupervised (weakly supervised) method. Supervised method learns the object prior from the training data. Compared with the second one, supervised method usually achieves better segmentation, since the object prior can be more accurately learned by the supervised manner. Many supervised methods have been proposed, such as the interaction based method [23]–[26] and training based method [16]–[21]. The interaction based method generates object prior through the location and appearance of the object labeled by the user, such as Grabcuts [25] and random walker based segmentation method [26]. However, the interaction based method is not suitable for the realistic applications with a large number of images. Compared with the interaction based method, the training based method can be used in more applications.

The training based method generates the appearance prior [27], [28] or the shape prior [16]–[20] of the object based on the training data. In this paper, we focus on the shape prior based model. Gabor filtering feedback was employed to describe local edge information by Wu *et al.* [16], which searched the structure of the Gabor filter feedback shared by the training images to form the common template. For a new image, the similar structure of the Gabor filter feedbacks was searched to obtain the common object. In the work of Ferrari *et al.* [17], the common template was obtained through finding, assembling and refining similar K adjacent segments (kAS) from original images. By using Hough-style voting and non-rigid point matching algorithm, similar objects can be detected from the new images. In the work of Jiang *et al.* [18], a common object detection method that was robust to object variations such as translation, rotation and scale was proposed. Thin plate spline (TPS) parameterizations were used to learn the mean shape from the annotated images, and the extended TPS-PRM algorithm was used for similar contours matching in new images. In the work of Ma *et al.* [19], contours were represented by fragments of edges. The common template was generated by affinity propagation clustering algorithm. Similar object was then detected from new images by maximal clique computation of the corresponding weighted graph. Region-based descriptor rather than edge-based descriptor was used to describe local area by Bagon *et al.* [20] which employed a self-similarity descriptor [21] to describe local region. The specific objects were extracted from the training images base on the descriptor matching. The common template was generated by averaging the obtained descriptors. For new images, objects were segmented through searching the region that was most similar to the object template.

There is also much research on imposing prior shape knowledge into image segmentation [29]–[35]. Cremers *et al.* in [29] added a variational integration of nonlinear shape statistics into a Mumford-Shah based segmentation process to guide curve evolution, which can segment objects with misleading information due to noise, clutter and occlusion. In [30], Cremers *et al.* measured the dissimilarity between two level sets based shape priors by computing the area of the set symmetric difference. Based on the invariant shape dissimilarity measurement, a statistical shape prior that can encode multiple fairly distinct shapes was introduced into level set based segmentation for accurate object segmentation. In the work of Schoenemann *et al.* [31], the shape prior was represented in the product space referring to the shape prior and images. The objects similar to the shape prior can be searched on the product space (represented by graph) by Minimum Ratio Cycle algorithm. Klodt *et al.* in [32] represented the shape prior in terms of moment constraints and added the shape prior into a convex framework for image segmentation. Several shape based moment constraints, such as area constraint and centroid constraint, were formulated as nested convex sets, which resulted in the simple minimization. Veksler *et al.* [33] proposed a generic shape prior named the star shape prior, which was not specific to any particular object, but rather applied to a wide class of objects, in particular to convex objects. The shape prior was then added into MRF based segmentation method for accurate segmentation. In [34], Lempitsky *et al.* added bounding box shape prior into interactive based graph cut segmentation to improve bounding boxes based segmentation. The shape prior was formulated as the tightness between the object boundary and bounding box, which can make the segment close to the bounding box. Das *et al.* [35] represented shape prior by compact shape, which used the directions of the connected boundary pixels to measure the semantic of current shape. The shape prior was added into MRF segmentation and can be minimized by graph cuts algorithm.

In addition to the supervised segmentation methods, unsupervised (weakly supervised) segmentation methods automatically generate object prior (or require less user interaction). Several unsupervised (weakly supervised) segmentation methods have been proposed, such as co-segmentation [6]–[12], [36]–[40] and weakly supervised semantic segmentation [13]–[15].

The co-segmentation extracts the specific object by segmenting common object from a group of images [41]. By introducing an additional foreground similarity constraint into single image segmentation, the object can be automatically segmented from the images. Several co-segmentation methods [6]–[12], [36]–[39], [42] have been proposed in the past few years.

MRF based co-segmentation method adds the constraint of foreground similarity into traditional MRF based segmentation model for common object segmentation [41]. The key is how to measure foreground similarity. Meanwhile, since the foreground similarity measurement makes the energy minimization difficult, another problem is how to minimize the energy. In the existing MRF based co-segmentation methods, the similarity measurements such as L1-norm [6], L2-norm [7], reward model [8], and Boykov-Jolly model [9], [43] have been used. The corresponding optimization methods, i.e., trust region graph cuts (TRGC) method [6], quadratic pseudo boolean optimization method [7], maximum flow procedure of graph [8], and dual decomposition [9] were also proposed for the minimization [40]. Instead of introducing foreground similarity into MRF model, Chu *et al.* in [42] proposed a MRF based co-segmentation method from the perspective of common pattern discovery. In the method, the foreground similarities were represented as feature matching and were discovered by common pattern discovery (using SIFT feature matching and density-based clustering). Then, based on the confidence maps obtained from the common pattern discovery, the unary term of MRF based segmentation was designed. Finally, the Graph-cuts algorithm was used to minimize the model for common object segmentation. Several logos have been tested to be successfully segmented by the method in [42]. Note that the logos in [42] are treated as common objects, and the purpose of [42] is to segment common logos from the image group. However, in our method, logos are treated as local information, which are used to locate the object and provide the appearances of the object. Furthermore, we also consider the shape variations among the objects for accurate object segmentation.

Apart from MRF based co-segmentation method, discriminative clustering and spectral clustering method were combined to segment common objects by Joulin *et al.* in [11]. The classifier by spectral clustering technique was treated as segmentation. The classifier that best discriminates the foregrounds and backgrounds was searched as the final co-segmentation classifier. An interactive co-segmentation method was proposed in the work of Batra *et al.* [43] which used an automatic recommendation system to achieve accurate common object segmentation. In the work of Vicente *et al.* [44], an object co-segmentation method was proposed. The authors first segmented the images into a set of overlapping local regions. Then, the co-segmentation was casted as MRF setting problem which was optimized by Loop Belief Propagation. Mukherjee *et al.* in [45] presented a scale invariant co-segmentation method which was robust to the scale variation. The fact that the matrix comprised of the common objects should have a rank of one was used to search the common objects. Chang *et al.* [46] proposed a novel global energy term that made the energy function submodular. Hence, Graphcut algorithm can be used to efficiently solve the optimization problem. Furthermore, the co-saliency map was employed to construct the unary term of the energy function. Kim *et al.* [12] segmented the common objects by diffusing the heat among the different images according to the region similarities.

To learn the object prior, several weakly supervised learning methods [13]–[15] focus on simultaneously learning the appearance model and labeling the pixels. For example, Vezhnevets *et al.* in [13] proposed a graphical model, i.e., multiple image model (MIM), to recover the pixel labels of the weakly labeled training images. The segmentation was achieved by applying the Alpha Expansion algorithm to find the approximate MAP state of MIM. In [14], a conditional random filed based weakly supervised learning method was proposed. In the model, the consistency between the segmentation and training samples

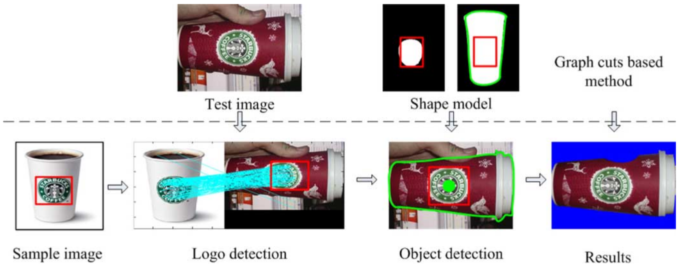
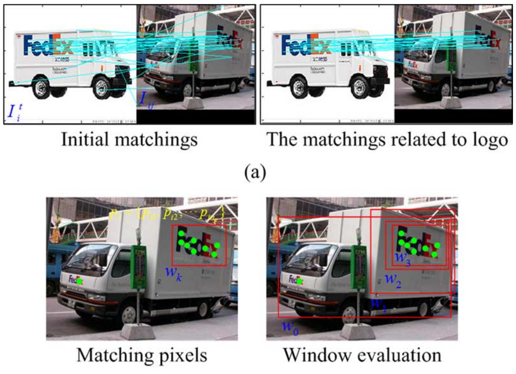


Fig. 3. The flowchart of the proposed method.



Initial matchings The matchings related to logo

(a)

Matching pixels Window evaluation

(b)

Fig. 4. An example to illustrate the logo detection.

TABLE I
VARIABLES DESCRIPTION

Symbols	Parameters
$l_i, i = 1, \cdots, m$	The i-th logo
I_i^t	The training image for l_i
m	The number of logos
I_i	Image set referring to l_i
I_{ij}	The j-th image in I_i
$b_i = \{b_{i1}, b_{i2}, \cdots, b_{in_b}\}$	The matching pixels of I_{ij}
$W = \{w_1, \cdots, w_{n_w}\}$	The initial windows for logo detection
r_m	The matching ratio
A_w	The area of the window
λ_1,σ	The scaling factor
$z_i \in \{0, 1\}$	The label of pixel p_i
$U(z_i)$	The unary term
$V(l_i, l_j)$	Pairwise term
$\overline{s}$, $\tilde{s}$	The original and changed shape model of the specific object.
$(\overline{x}, \overline{y})$	The location of Logo in $\overline{s}$
v	The shape variations operator
$\tau(v(\overline{s}, \alpha), I)$	Similarity measurement between $v(\overline{s}, \alpha)$ and original image I
B	The binary edge map
P_b	The set of edge pixels in B
P_s	The set of shape boundary pixels in B'
$d_s(p, B)$	The distance between pixel p and B
$f(d_s(p, B))$	The similarity between pixel p and B
(x, y)	The location of pixel p
T_r,T_s, T_s', T_q	The transform matrixes
α, θ, β	The parameters in shape transform

was considered. In [15], the object model was learned from the weakly labeled training images by active learning, which alternatively performed the weakly supervised learning and pixels' label for accurate learning. Because of the complexity of the background and the variations among the objects, there is still a substantial gap between the supervised methods and this type of weakly supervised methods.

III. The Proposed Method

The flowchart of the proposed method is shown in Fig. 3, where there are three steps, i.e., the logo detection, the object detection and object segmentation. In this section, we introduce these steps.

A. Logo Detection

To detect the logo, we give a training image I_i^t for each logo $l_i, i = 1, \cdots, m$, where m is the number of logos. To explicitly infer the variables used in this paper, we show these variables in Table I. For a new image I_{ij} where ij denotes the j-th image in the i-th image set I_i related to logo l_i, we detect the logo l_i by three steps. In the first step, image I_{ij} is matched with the training image I_i^t by SIFT matching. The matching pixels $b_i = \{b_{i1}, b_{i2}, \cdots, b_{in_b}\}$ in I_{ij} related to the logo pixels in I_i^t (as shown in Fig. 4(a)) are selected. In the second step, we consider the detection result as one of the initial windows $W = \{w_1, \cdots, w_{n_w}\}$ setting by sliding windows method. Each initial window is scored by window evaluation and the best window is selected as the final result. One example is shown in Fig. 4.

Two aspects are considered to score the window. One is the matching ratio (r_m) which is defined as the ratio of the number of matching pixels inside the windows n' to the total number of matching pixels n_b.

$$r_m = \frac{n'}{n_b} \tag{1}$$

The other is the area of the window A_w. Based on r_m and A_w, the score of window w_k is given by

$$score(w_k) = r_m - \lambda_1 \cdot A_w \tag{2}$$

where λ_1 is scaling factor. We can see that the windows covering most of the matching pixels will have large scores. Meanwhile, for a certain set of matching pixels (such as the matching pixels in w_1 in Fig. 4(b)), there are many windows (such as w_1, w_2 and w_3 in Fig. 4(b)) that cover these matching pixels and have the same value of r_m. Hence, the second term, i.e., area term, is introduced to order these windows. We believe that a small window is more compact and reasonable than a large window, such as the window w_3 compared with w_1 in Fig. 4(b). Furthermore, only using the matching ratio r_m will lead to the selection of the window covering all of the matching pixels, such as w_0. Because there may be several wrong matching pixels that are far away from the logo, covering these wrong pixels results in detection of a large unsuccessful window. Hence, we use the area term to avoid these wrong detections. It is seen that deleting these wrong matching pixels results in a small decrease of matching ratio but dramatically increases the area term, and finally results in a large score.

B. Object Segmentation

1) Segmentation Model: We segment the object by Markov Random Filed segmentation method, which searches label z of

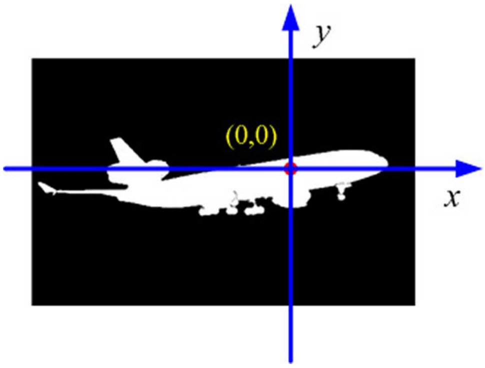


Fig. 5. The shape model used in the proposed method. The origin (0, 0) of the planar coordinate system xoy is the location of the logo.

image pixels that minimize the energy function represented as

$$E(z)=\sum_{z_i\in p}U(z_i)+\sum_{(i,j)\in N}V(z_i,z_j) \tag{3}$$

where $z_i\in\{0,1\}$ is the label of pixel p_i. p is the set of pixels. The foreground has label 1, and 0 for the background. $U(z_i)$ is the unary term which describes the probability of labeling z_i to p_i. The foreground and background models are represented as the Gaussian mixture model with parameters $N(\mu_1,\sigma_1)$ and $N(\mu_2,\sigma_2)$, respectively. To segment the specific objects of the proposed method, the foreground model and background model are required. Hence, we need to estimate the parameters μ_1,σ_1 and μ_2,σ_2 of the foreground model and background model.

In the second term of (3), $\mathcal{N}$ is the set of neighboring pixels. Tow pixels are neighbors if they satisfy 3×3 neighbor relationship. $V(l_i,l_j)$ denotes the pixels' consistency between two neighboring pixels, which is usually defined as the distance between the appearances of the pixel pair. The second term makes the labels of neighbor pixels consistency. Based on (3), the label with the minimum energy function is searched by the graph-cuts algorithm.

To segment the specific object, we require to obtain the appearance models of both foreground and background referring to μ_1,σ_1 and μ_2,σ_2. In our model, we estimate these parameters by searching the object boundary, where the regions inside and outside the boundary are used to generate the foreground appearance model and the background appearance model.

2) Boundary Extraction Based on the Logo: In our method, the detected logo simplifies the boundary extraction by two aspects. Firstly, the logo provides the roughly location of the objects. Secondly, the logo based object shape prior can easily handle the changes of the object and result in successful boundary extraction.

a) Logo Based Object Prior: The shape of the specific object $\overline{s}$ extracted from any image is used as the object prior. We also consider the location of the logo $(\overline{x},\overline{y})$ locating in the object shape. Hence, the object prior can be represented as $(\overline{x},\overline{y},\overline{s})$. In the object prior, shape is represented in a planar coordinate system xoy. The location of the logo $(\overline{x},\overline{y})$ is treated as the $(0,0)$ of xoy plane, and the shape boundary locates around the $(\overline{x},\overline{y})$, as shown in Fig. 5. The relationship between $(\overline{x},\overline{y})$ and the shape is represented by the location of the boundary pixels.

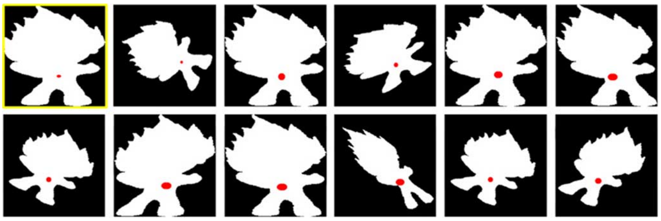


Fig. 6. The variations of the shape model.

b) The Model Prior Variations: Based on the logo based shape prior, the object is extracted by two steps. In the first step, we treat the image as a xoy planar coordinate system, and set the center of the detected window (x_0,y_0) as the origin $(0,0)$ in the planar coordinate system. The shape model $\overline{s}$ is then put into the image based xoy planar coordinate system by aligning $(\overline{x},\overline{y})$ with (x_0,y_0). Assuming there are no variations among the objects, the shape boundary is then considered as the object boundary.

Since there are variations among the objects, such as the changes in rotation, scale and pose, the initial object boundary usually does not locate along the object boundary. Hence, in the second step, we vary the $\overline{s}$ to cope with the shape changes, which can be represented as

$$\tilde{s}=v(\overline{s},\alpha) \tag{4}$$

where α is the parameter of the variations, v is the variation operator, $\tilde{s}$ is the new shape. By considering the shape changes, we obtain the object boundary by searching a variation of the shape $\tilde{s}$ that best fits the local edge of the image, which is represented as

$$\alpha=\arg\min_{\alpha}\tau(v(\overline{s},\alpha),I) \tag{5}$$

where α is considered here since α refers to $\tilde{s}$. $\tau(v(\overline{s},\alpha),I)$ is the measurement of the fitness between the varied shape $v(\overline{s},\alpha)$ and the local edge of the image I. By obtaining the transform parameter α, the final object boundary is obtained through varying $\overline{s}$ with the α by (4), and then aligning with the logo location (x_0,y_0).

c) The α and v for Shape Variations: We consider the shape changes by affine transform which contains the usual shape variations, such as the variations of rotation, scale and translation. For a pixel with location (x,y), the new location (x',y') after affine transform can be obtained by

$$(x',y',1)=(x,y,1)*\begin{pmatrix}\alpha_1&\alpha_2&0\\ \alpha_3&\alpha_4&0\\ \alpha_5&\alpha_6&1\end{pmatrix} \tag{6}$$

where $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_6)$ is a parameter. The original shape can be varied by adjusting different α. Fig. 6 shows several shape variations by selecting different α. Note that we also give the logo location $(\overline{x},\overline{y})$ (red node in each image) after the affine transform.

d) The measurement τ: To measure the fitness between shape $\tilde{s}$ derived from $\overline{s}$ and image I, we first extract the binary edge map B of the image I by edge detection method. In B, the edge pixels have value 1, and 0 for the background pixels. One

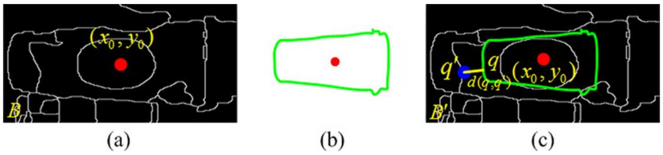


Fig. 7. (a): The binary edge map of the original image. (b): The shape model. (c): Example to calculate the fitness of the shape with an image edge map.

example is shown in 7(a). We denote the set of edge pixels as P_b. Then, we introduce the shape $\tilde{s}$ into B to obtain B' by aligning the logo $(\bar{x}, \bar{y})$ in $\tilde{s}$ with the logo (x_0, y_0) in B, as shown in Fig. 7(c). The set of pixels on the shape boundary is denoted as P_s. We can see that only the pixels on the image edge and the shape boundary are considered for measurement.

In the fitness evaluation, we only consider the edge pixel $p \in P_b$ around the shape boundary in $\tilde{s}$. For any pixel $p \in P_s$, we define the distance $d_s(p, B)$ between the pixel p and the image edge B as the smallest geometric distance between the $p \in p_s$ and the pixels $p' \in P_b$, i.e.,

$$d_s(p, B) = \arg \min_{p' \in P_b} d(p, p') \tag{7}$$

where $d(p, p')$ is defined as

$$d(p, p') = \sqrt{(x - x')^2 + (y - y')^2} \tag{8}$$

and (x, y) and (x', y') are locations of pixel p and p', respectively. One example is shown in Fig. 7(c), where the distance between q and image edge (white contour) is the geometric distance $d(q, q')$ between q and q'. The q' has the smallest distance to q among the pixels of the image edge.

Based on d_s in (7), we evaluate each pixel $p \in P_s$ by

$$f(p) = f(d_s(p, B)) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\frac{d_s(p, B)}{\sigma})^2\right) \tag{9}$$

where σ is variable referring to the width of the band around the shape contour. Based on (9), the fitness of shape $\tilde{s}$ to image I is defined as

$$\tau(\tilde{s}, I) = -\sum_{p \in P_s} f(p) = -\sum_{p \in P_s} f(d_s(p, B)) \tag{10}$$

We can see from (10) that the fitness is the sum of distances between all shape pixels to the image edge. When $\tilde{s}$ fits the local edge of the image, these shape pixels have small distances to the image edge and result in large $f(p)$. Thus, a small fitness value is obtained. Otherwise, a large value is given to the shape when the shape does not well fit the local edge of the image.

e) The Optimization: Based on (10), we can rewrite (5) by

$$\alpha^* = \arg \min_{\alpha^*} -\sum_{p \in P_s} f(d_s(p, v(\bar{s}, \alpha))) \tag{11}$$

The object boundary can be extracted by searching α that satisfies (11). To search α, we use gradient free based minimization method, such as Nelder-Mead simplex method. However, minimizing the cost function with respect to six parameters is a complicated computational task, since directly searching the parameters may lead to an undesirable local minimal value. In the proposed method, before applying the Nelder-Mead simplex method, we perform a rough search in the 6-dimensional parameter space, such as working on a coarse to fine set of grids. Since the dimension is still large for practice performance, we use a separation performance to divide the affine transform into 5 sub-transforms and lead to two dimension search space.

Four sub-operators are considered. They are rotation, scale, squeezing and translation, respectively. To simplify the solution, the scale is also divided into two sub-operators. One is the scale that has same scale factors on the width and the height. The other has different scale factors on the width and the height. Hence, there are five sub-transforms. The transform matrixes of these sub-transforms are expressed as

$$T_r = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}, T_s = \begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & \beta_1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$
$$T_s' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \beta_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, T_q = \begin{pmatrix} 1 & \beta_3 & 0 \\ \beta_4 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{12}$$

where the transform matrixes of rotation (T_r), scale (T_s and T_s' are the first scale and second scale transform, respectively), squeezing (T_q) are shown. θ is the angle in the rotation. $\beta_1, \beta_2, \beta_3, \beta_4$ are parameters related to the scale and squeezing transform. For the translation transform, we simply change the logo location $(\bar{x}, \bar{y})$ with distance β_5 and β_6 on the row and column, i.e., $(\bar{x} + \beta_5, \bar{y} + \beta_6)$. By considering T_t as the translation transform matrix, the original transform matrix T can be represented as:

$$T = T_r * T_s * T_s' * T_q * T_t \tag{13}$$

and the parameter α changes from $\alpha = \{\alpha_1, \cdots, \alpha_6\}$ to $\alpha = \{\theta, \beta_1, \cdots, \beta_5, \beta_6\}$.

Although the number of parameters increases, the advantage of the sub-transforms is that the objects usually have large changes on two transformations, i.e., the rotation T_r and scale (T_s). For the others, the objects have small changes. Hence, we can perform the rough search by only considering the transformations of rotation T_r and scale T_s. Note that it is a two dimensional search space that can be quickly achieved.

Based on the rough search, we then obtain the more accurate shape by minimizing (11) using Nelder-Mead simplex method. After obtaining α^*, the object boundary is obtained by transforming the original shape $\tilde{s}$ with respect to α^* by (4).

3) Object Segmentation: After extracting the object boundary, the images are classified into two regions, the region R_i inside the boundary and the region R_o outside the boundary. We consider R_i and R_o as the foreground region and background region respectively, which are used to estimate the parameters (μ_1, σ_1) and (μ_2, σ_2) for the foreground model and background model as introduced in Section III-B1. Then the foreground model and background model are introduced into the unary term of the segmentation energy function in (3). Finally, the energy function is minimized by the graph-cuts algorithm and the object is extracted through the obtained label.

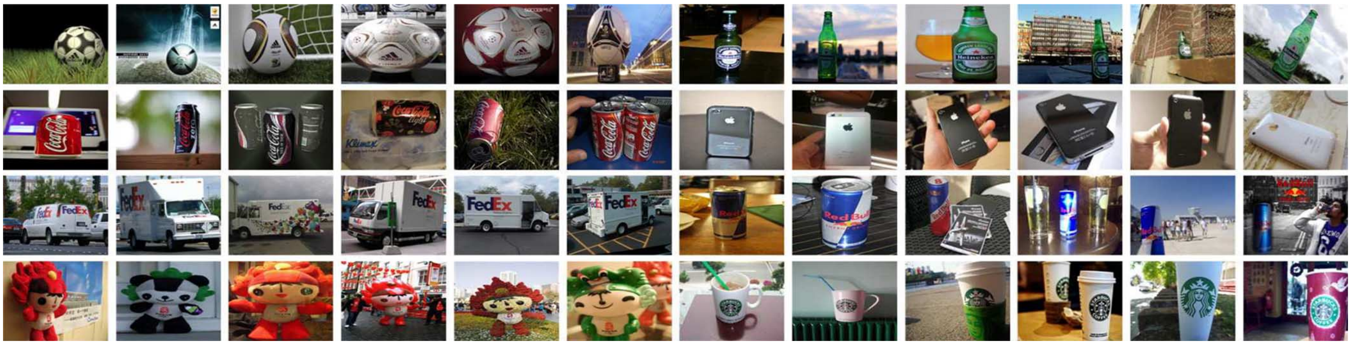


Fig. 8. Images in the logo dataset.

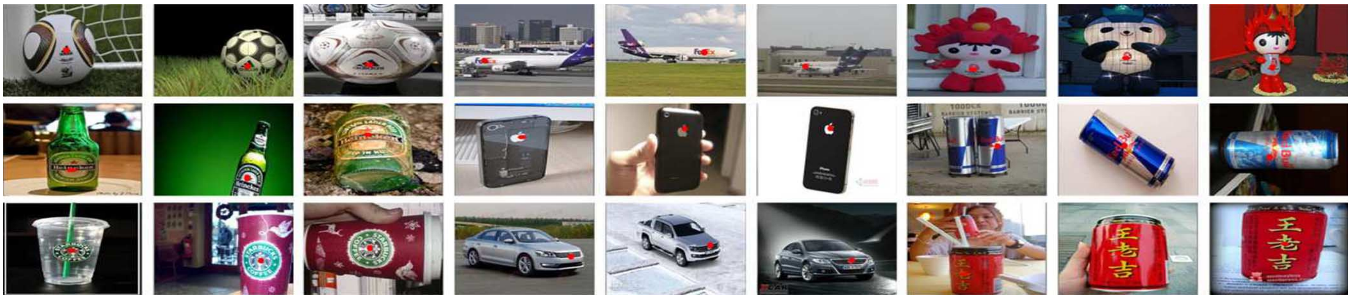


Fig. 9. The logo detection results. The logos are shown as the red points in the images.

IV. Experimental Results

In this section, we verify the performance of the proposed method. A Logo dataset is collected from web to verify the proposed method. We also use the MOMI dataset given by [42] for verification. Some subjective and objective assessments of segmentation results are reported.

A. Parameters Setting

We first introduce the parameter setting in our experiments. In the logo detection, the training image with simple background is selected as logo sample. The source code of the SIFT matching released by the author is used.[1] The initial window is set through the sliding windows method. In (2), $\lambda_1 = 0.1$. In the object segmentation, the number of Gaussian distribution in Gaussian mixture model is 5 for both foreground and background model. In the object detection, we use the method in [5] to obtain the edge map by setting $K = 50$. For the σ in (9), we set $\sigma = 5$ for the rough search and the minimization. In the rough search, the θ is searched in range of $[0, 2\pi]$ with step 0.2 and β_1 is searched in range of $[0.5, 4]$ with step 0.2.

B. Results of Our Proposed Method

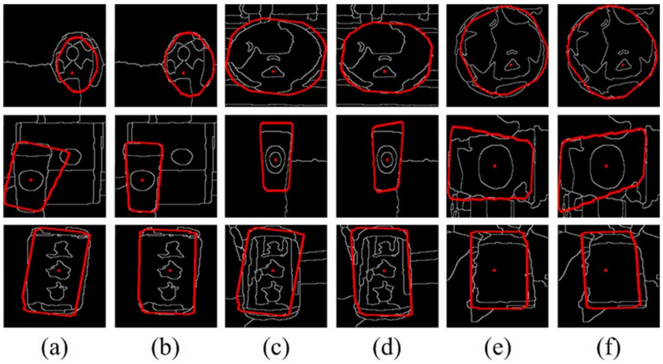


Fig. 10. The results of curve extraction by the proposed method. (a)(c)(e): The extracted boundaries by the rough search. (b)(d)(f): The final extracted object boundaries.

In order to completely verify the proposed method, we collect images from the web such as Flickr and Google to form LogoSeg dataset. In the LogoSeg dataset, there are 13 classes, such as *Adidas*, *Cocacola*, *Fedextrucks* and *Iphone*. Some of the images are shown in Fig. 8. It is seen that the objects have many variations among the objects, such as the color changes of the objects in *Cococola*. Meanwhile, many images have complex backgrounds, such as street scenes in *Fedextrucks*. The variations among the objects and the complexities of the background make the object extraction challenging.

We also verify the proposed method on MOMI dataset given in [42]. In the MOMI dataset, there are 12 classes, each of which contains three to eight images. In each class, the common pattern logos or regions are contained among the images. In our experiments, all the classes (12 classes) in MOMI dataset are used for complete verification. The common logos or regions are treated as the logos in our model. Because there are no training logos in MOMI dataset, we collect the similar logos and targets from the web to form training logos and object shape models. The ground-truth given by [42] are used for verification.

We first show the logo detection results by the proposed method. The results are shown in Fig. 9, where the logo is denoted as the red dot in each image. We can see from Fig. 9

[1] http://www.cs.ubc.ca/∼lowe/keypoints/

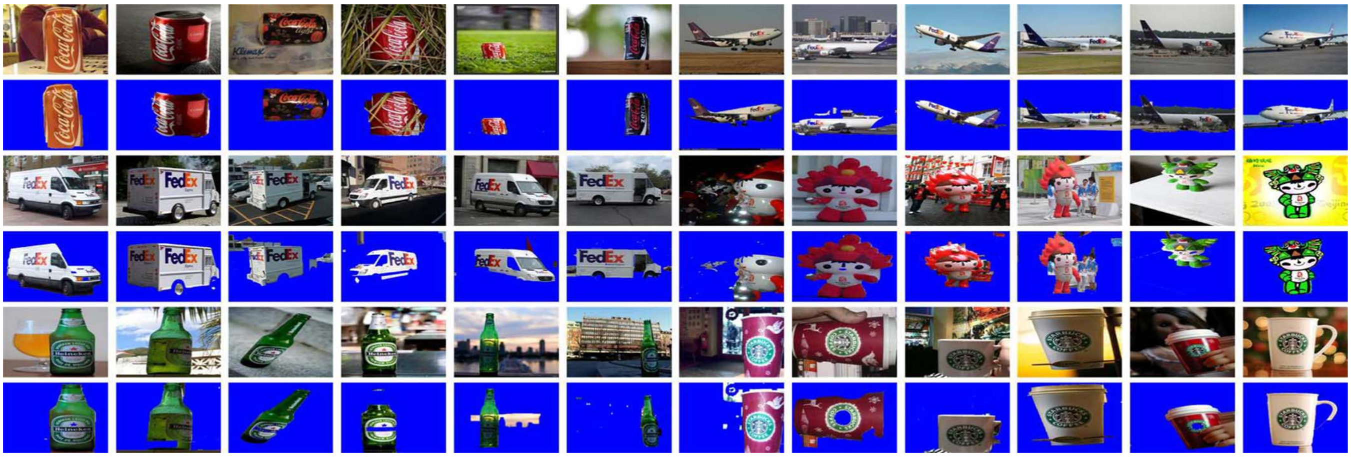


Fig. 11. The segmentation results of LogoSeg dataset by the proposed method. The original images are shown in the row 1, 3 and 5. The corresponding results are shown in the rows 2, 4 and 6, respectively.

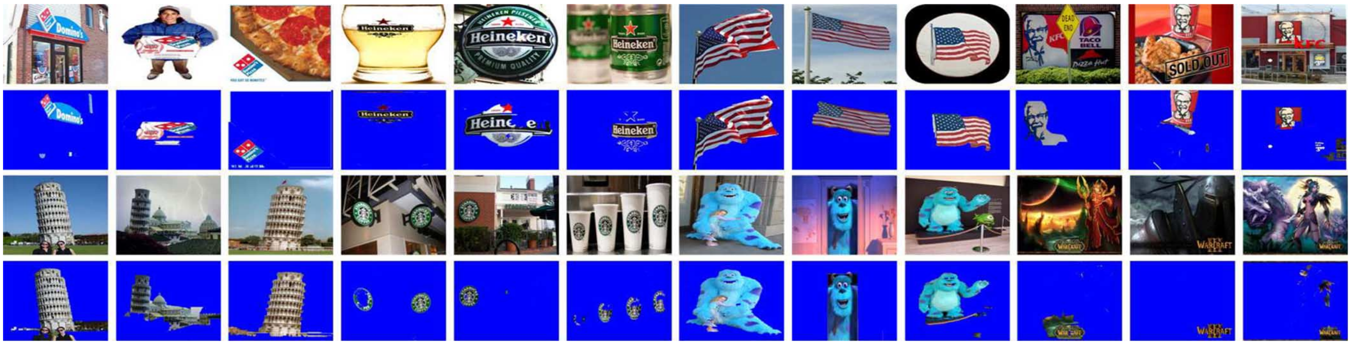


Fig. 12. The segmentation results of MOMI dataset by the proposed method. The original images are shown in the row 1, 3 and 5. The corresponding results are shown in the rows 2, 4 and 6, respectively.

that the logos are successfully detected from these images. The successful location of the logo is caused by the stable local textures shared by the logos. These stable local textures provide the stable logo location to benefit the object extraction of the following steps.

The extracted boundaries of the objects are shown in Fig. 10, where Fig. 10(a)(c)(e) show the extracted boundaries by the rough search. The final object boundaries are shown in Fig. 10(b)(d)(f). We can see that the rough search can provide an object boundary near the objects in these images. Meanwhile, the final searching can obtain more accurate object boundary from the background noises.

The segmentation results of the proposed method are shown in Figs. 11 and 12, where the segmentation results of the LogoSeg dataset and MOMI dataset are shown, respectively. In Fig. 11, the test images and the corresponding segmentation results of six classes in LogoSeg dataset are displayed. The original images are shown in the rows 1, 3 and 5. The rows 2, 4 and 6 display the corresponding segmentation results. It is seen from Fig. 11 that the proposed method successfully segments the objects from these images with complex backgrounds, such as the segments in *Fuwa*. The segmentation results of MOMI dataset are shown in Fig. 12, where the results of eight classes are represented. For each class, we display the results of three images. It is seen that the proposed method can also segment the objects from these images, such as "Pisa" in class *Pisa*.

C. *Objective Evaluation*

To evaluate the performance of the proposed method, we compare the proposed method with several existing related methods such as the methods in [16], [12], [11] and [42]. The method in [16] is a shape based object extraction method, which learns the shape model from the training images by discovering the similar local edge structures. Gabor filter is used to describe the local edge and the edge structure. In our experiments, the source code released by the authors is used.[2] The authors in [12] propose an anisotropic heat diffusion based co-segmentation, which first locates the similar regions among the images by clustering method. The location is then used as the seed to segment the object by anisotropic heat diffusion method. The code[3] released by the author is used in our experiment. To improve the performance of the method in [12], we also vary the parameters such as the segment number and Gaussian weight for better performance. The authors in [11] propose a co-segmentation method by combing discriminate clustering method and spectral clustering technique. The classifier that best discriminates the foregrounds and backgrounds was searched to achieve common object segmentation. We perform the method in [11] using the source code released by

[2] http://www.stat.ucla.edu/ywu/AB/ABbasicDec292010.zip

[3] http://www.cs.cmu.edu/∼gunhee

TABLE II
COMPARISON BETWEEN THE EXISTING METHODS WITH THE PROPOSED METHOD IN TERMS OF F-MEASURE

LogoSeg								
Method	Adidas	Coca	Fedextruck	Fedexplane	Fuwa	Heincken	Iphone	Nike
[11]	0.6475	0.5537	0.5163	0.3570	0.6734	0.5976	0.4792	0.4537
[16]	0.7025	0.5888	0.5146	0.5590	0.6178	0.5318	0.5727	0.4723
[12]	0.5555	0.4320	0.4461	0.3912	0.4930	0.3745	0.4286	0.4718
[42]	0.7418	**0.6911**	0.5252	0.2709	0.6210	0.5830	0.7672	0.7568
Ours	**0.7611**	0.6680	**0.6770**	**0.5590**	**0.7688**	**0.7554**	**0.7916**	**0.7751**
Method	Redbull	Starbucks	Thinkpad	Vw	Wang			
[11]	0.4882	0.5841	**0.8797**	0.7177	0.6885			
[16]	0.5351	**0.6588**	0.6098	0.5445	0.5737			
[12]	0.4120	0.4814	0.6740	0.4988	0.5522			
[42]	0.4854	0.5925	0.3336	0.7604	0.6550			
Ours	**0.6355**	0.6397	0.6262	**0.8851**	**0.7405**			
MOMI								
Method	Sulley	Starbucks	Magnet	USAFlag	Pisa	Superman	Heineken	Pringles
[11]	0.4387	0.1980	0.6652	0.5850	0.7904	0.1491	0.1516	0.2512
[16]	0.5115	0.1988	0.5231	0.2880	0.5310	0.2780	0.3929	**0.4265**
[12]	0.6065	0.1657	0.5744	0.3479	0.5800	0.1545	0.1056	0.1697
[42]	0.6122	0.4781	**0.8107**	0.8166	0.8207	0.2887	0.2244	0.2858
Ours	**0.9130**	**0.5777**	0.5688	**0.9232**	**0.8627**	**0.3495**	**0.4425**	0.2724
Method	Kfc	Warcraft	Domino	Lego				
[11]	0.1243	0.1332	0.1280	0.0320				
[16]	0.1466	0.0439	0.3770	**0.1930**				
[12]	0.0877	0.1209	0.1400	0.0433				
[42]	0.1939	0.3467	0.2546	0.0312				
Ours	**0.4391**	**0.5531**	**0.5551**	0.0792				

the authors.[4] In the experiments, Chi-square kernel is used. SIFT features are selected for local regions representation. In the method [42], common pattern discovery algorithm is first performed to obtain the confidence maps representing the potential of pixels belonging to common patterns. Then, the MRF based energy is calculated based on the confidence maps for common object segmentation. In our experiments, we implement the method in [42] by Matlab code. Level-set method is used for confidence map generation. The parameters of the density-based algorithm are adjusted for better results.

In our experiments, F-measure is used for objective evaluation. F-measure is defined as $\frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$, where $Precision$ is the ratio of the number of successful segmented foreground pixels to the number of segmented foreground pixels, and $Recall$ is the ratio of the number of successful segmented foreground pixels to the number of foreground pixels in Ground-truth. A large F-measure refers to an accurate segmentation. Note that the results of the method in [16] are contours rather than regions. We calculate the F-measure of the method in [16] by replacing the region and ground truth to the rectangles exactly covering the region and the ground truth respectively. Meanwhile, there are more than two images in each class. We use the mean F-measure over all images to evaluate the performance of each class. Table II concludes the F-measures of the methods in [16], [12], [11], [42] and the proposed method.

From Table II, we can see that the method in [16] achieves good performance on some image groups, such as *Adidas* and *Pringles*. Meanwhile, there are unsuccessful extractions, such as the results of *Heincken* and *Nike*. The reason for the unsuccessful extraction is caused by the shape variations among the objects. For the method in [12], the objects are segmented from the images such as *Wang* and *Sulley*. Furthermore, there are unsuccessful extractions, such *Iphone* and *Starbucks*. The unsuccessful segmentation is caused by the reason that the method in [12] focuses on the common objects segmentation containing similar colors, which is not suitable for the common objects segmentation containing similar contours. The method in [11] achieves successful objects segmentation in several classes, such as *Adidas* and *Pisa*. Unsuccessful segmentation is also obtained by the other classes, such as *Iphone*. The unsuccessful segmentation is caused by the pose variations among the objects. For the method in [42], the objects can be successful extracted on several datasets, such as *Adidas* and *USAFlag*. Unsuccessful segments are also obtained, such as *Fedextruck* and *Redbull*. The unsuccessful segmentation is caused by the fact that extracting common objects by common pattern discovery can be confused by the complex backgrounds. Compared with the existing methods, we can see that the proposed method achieves the largest F-measures in many classes. The improvements are mainly caused by two reasons. One is that logo detection simplifies the object detection. The other is that the shape transform is considered in the proposed method to overcome the shape changes among the objects.

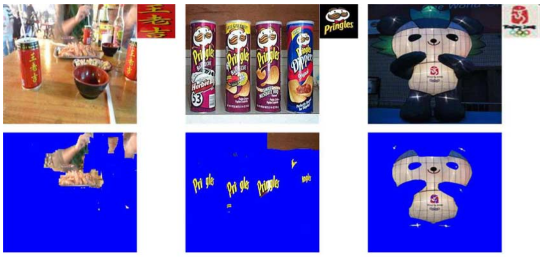
V. DISCUSSION

In this paper, we use classical SIFT feature for logo detection. Note that other matching methods can also be used for the logo matching. We also test the proposed method based on the other matching methods to verify the proposed method. The F-measures of the corresponding results are shown in Table III, where the results based on SIFT matching and SURF [47] matching are shown. We can see that the SURF based method achieves larger F-measures on several classes compared with SIFT matching, such as *Iphone* and *Starbucks*. But for some classes, the lower

[4] www.di.ens.fr/∼joulin

TABLE III
THE RESULTS BY DIFFERENT MATCHING METHODS. SIFT MATCHING AND SURF MATCHING ARE CONSIDERED

LogoSeg								
Method	Adidas	Coca	Fedextruck	Fedexplane	Fuwa	Heincken	Iphone	Nike
SIFT	0.7611	0.6680	0.6770	0.5590	0.7688	0.7554	0.7916	0.7751
SURF	0.7221	0.6848	0.6593	0.6884	0.7417	0.6997	0.8651	0.7230
Method	Redbull	Starbucks	Thinkpad	Vw	Wang			
SIFT	0.6355	0.6397	0.6262	0.8851	0.7405			
SURF	0.6439	0.6675	0.6430	0.8160	0.7258			
MOMI								
Method	Sulley	Starbucks	Magnet	USAFlag	Pisa	Superman	Heineken	Pringles
SIFT	0.9130	0.5777	0.5688	0.9232	0.8627	0.3495	0.4425	0.2724
SURF	0.8096	0.8125	0.5678	0.7857	0.8576	0.4133	0.4520	0.2853
Method	Kfc	Warcraft	Domino	Lego				
SIFT	0.4391	0.5531	0.5551	0.0792				
SURF	0.4512	0.4815	0.5868	0.0619				



(a) (b) (c)

Fig. 13. Failure cases by the proposed method. The proposed method fails for these classes due to cluttered background ((a) and (b)) and the similarity between the object and background ((c)).

F-measure values are obtained by SURF based method, such as *USAFlag* and *Sulley*. Considering all classes, the mean F-measure value of SURF based method (0.6338) is very close to the mean F-measure of SIFT based method (0.6328), which demonstrates that both SIFT and SURF can be used as logo detection in our method.

In addition, it should be noticed that false segmentation will be caused when the image contains significantly cluttered background or the object is depicted in very similar backgrounds. Examples can be found in Fig. 13(a) to (c), where some false segments can be observed. It is seen that the cluttered backgrounds can confuse the logo detection and then result in unsuccessful object segmentation, such as Fig. 13(a) and (b). Meanwhile, the similarity between the background and foreground increases the ambiguity of the segmentation task and leads to incomplete segmentation, such as Fig. 13(c).

In the proposed method, the gradient free based minimization method is used to minimize the model. In the minimization, a precise initial value is required for the accurate search of parameters. Because we can divide the affine transform used in the model into 5 sub-transforms and finally lead to two dimension search space, we use a rough search working on a coarse to fine set of grids to quickly obtain the initial value. When extending the method to the model with more complex parameter spaces, the large number of parameters will result in a large search space which leads to large computational cost of the grid search method. Note that for non-real-time applications, the grid search based method can also be used for initial value setting. For real-time applications, other approximate initial value setting methods such as random search [48] or manual search [49] can also be used for optimization.

VI. CONCLUSION

This paper proposes a specific object segmentation method based logo detection. In the method, the logo is firstly detected using SIFT matching. Then the objects are segmented based on the location of the logo and the shape model. To cope with the shape variations, affine transform of the shape model is considered. The best shape variation is searched by the Nelder-Mead simplex method with a simple initial rough search. We collect many images from the web to test the method. The experimental results demonstrate the effectiveness of the proposed method.

ACKNOWLEDGMENT

The authors would like to acknowledge Wen-Sheng Chu for his kind advice in the MOMI cosegmentation implementation.

REFERENCES

[1] M. R. Daliri and V. Torre, “Robust symbolic representation for shape recognition and retrieval,” *Pattern Recognit.*, vol. 41, no. 5, pp. 1799–1815, 2008.

[2] F. Jing, M. Li, H. J. Zhang, and B. Zhang, “Relevance feedback in region-based image retrieval,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 5, pp. 672–681, 2004.

[3] T. Nhon and K. Benjamin, “Skeleton search: Category-specific object recognition and segmentation using a skeletal shape model,” *Int. J. Comput. Vision*, vol. 94, no. 2, pp. 215–240, 2011.

[4] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[5] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, “From contours to regions: An empirical evaluation,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 2294–2301.

[6] C. Rother, V. Kolmogorov, T. Minka, and A. Blake, “Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, New York, NY, USA, Jun. 2006, pp. 993–1000.

[7] L. Mukherjee, V. Singh, and C. R. Dyer, “Half-integrality based algorithms for cosegmentation of images,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL, USA, Jun. 2009, pp. 2028–2035.

[8] D. S. Hochbaum and V. Singh, “An efficient algorithm for co-segmentation,” in *Proc. Int. Conf. Computer Vision*, Kyoto, Japan, Sep. 2009, pp. 269–276.

[9] S. Vicente, V. Kolmogorov, and C. Rother, “Cosegmentation revisited: models and optimization,” in *Proc. Eur. Conf. Computer Vision*, Crete, Greece, Sep. 2010, pp. 465–479.

[10] D. Batra, D. Parikh, A. Kowdle, T. Chen, and J. Luo, “Seed image selection in interactive cosegmentation,” in *Proc. IEEE Int. Conf. Image Processing*, Cairo, Egypt, Nov. 2009, pp. 2393–2396.

[11] A. Joulin, F. Bach, and J. Ponce, “Discriminative clustering for image co-segmentation,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, USA, Jun. 2010, pp. 1943–1950.

[12] G. Kim, E. P. Xing, L. Fei-Fei, and T. Kanade, “Distributed cosegmentation via submodular optimization on anisotropic diffusion,” in *Proc. Int. Conf. Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 169–176.

[13] A. Vezhnevets, V. Ferrari, and J. M. Buhmann, “Weakly supervised semantic segmentation with a multi-image model,” in *Proc. IEEE Int. Conf. Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 643–650.

[14] A. Vezhnevets, J. M. Buhmann, and V. Ferrari, “Active learning for semantic segmentation with expected change,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2012, pp. 3162–3169.

[15] A. Vezhnevets, V. Ferrari, and J. M. Buhmann, “Weakly supervised structured output learning for semantic segmentation,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2012, pp. 845–852.

[16] Y. N. Wu, Z. Si, H. Gong, and S.-C. Zhu, “Learning active basis model for object detection and recognition,” *Int. J. Comput. Vision*, vol. 90, no. 2, pp. 198–235, 2010.

[17] V. Ferrari, F. Jurie, and C. Schmid, “Accurate object detection with deformable shape models learnt from images,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[18] T. Jiang, F. Jurie, and C. Schmid, “Learning shape prior models for object matching,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Miami, FL, USA, Jun. 2009, pp. 848–855.

[19] T. Ma and L. J. Latecki, “From partial shape matching through local deformation to robust global shape similarity for object detection,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2011, pp. 1441–1448.

[20] S. Bagon, O. Brostovski, M. Galun, and M. Irani, “Detecting and sketching the common,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, USA, Jun. 2010, pp. 33–40.

[21] E. Shechtman and M. Irani, “Matching local self-similarities across images and videos,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[22] F. Meng, H. Li, and G. Liu, “Segmenting specific object based on logo detection,” in *Proc. IEEE Int. Symp. Circuits and Systems*, Beijing, China, May 2013, pp. 1–4.

[23] J. Zhang, J. Zheng, and J. Cai, “A diffusion approach to seeded image segmentation,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, USA, Jun. 2010, pp. 2125–2132.

[24] Y. Y. Boykov and M. P. Jolly, “Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images,” in *Proc. Int. Conf. Computer Vision*, Vancouver, BC, Canada, Jul. 2001, pp. 105–112.

[25] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” *ACM Trans. Graph.*, vol. 23, pp. 309–314, 2004.

[26] L. Grady, “Random walks for image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, 2006.

[27] O. M. Parkhi, A. Vedaldi, C. V. Jawahar, and A. Zisserman, “The truth about cats and dogs,” in *Proc. Int. Conf. Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 1427–1434.

[28] P. Arbeláez, B. Hariharan, C. Gu, S. Gupta, L. Bourdev, and J. Malik, “Semantic segmentation using regions and parts,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2012, pp. 3378–3385.

[29] D. Cremers, T. Kohlberger, and C. Schnorr, “Shape statistics in kernel space for variational image segmentation,” *Pattern Recognit.*, vol. 36, pp. 1929–1943, 2006.

[30] D. Cremers, S. J. Osher, and S. Soatto, “Kernel density estimation and intrinsic alignment for shape priors in level set segmentation,” *Int. J. Comput. Vision*, vol. 69, no. 3, pp. 335–351, 2006.

[31] T. Schoenemann and D. Cremers, “Globally optimal image segmentation with an elastic shape prior,” in *Proc. Int. Conf. Computer Vision*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–6.

[32] M. Klodt and D. Cremers, “A convex framework for image segmentation with moment constraints,” in *Proc. Int. Conf. Computer Vision*, Barcelona, Spain, Nov. 2011, pp. 2236–2243.

[33] O. Veksler, “Star shape prior for graph-cut image segmentation,” in *Proc. Eur. Conf. Computer Vision*, Marseille, France, Oct. 2008, pp. 454–467.

[34] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp, “Image segmentation with a bounding box prior,” in *Proc. Int. Conf. Computer Vision*, Kyoto, Japan, Sep. 2009, pp. 277–284.

[35] P. Das, O. Veksler, V. Zavadsky, and Y. Boykov, “Semiautomatic segmentation with compact shape prior,” *Image Vision Comput.*, vol. 27, no. 1–2, pp. 206–219, 2008.

[36] G. Kim and E. P. Xing, “On multiple foreground cosegmentation,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2012, pp. 837–844.

[37] A. Joulin, F. Bach, and J. Ponce, “Multi-class cosegmentation,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2012, pp. 542–549.

[38] M. Collins, J. Xu, L. Grady, and V. Singh, “Random walks for multi-image cosegmentation: Quasiconvexity results and gpu-based solutions,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2012, pp. 1656–1663.

[39] J. Rubio, J. Serrat, A. López, and N. Paragios, “Unsupervised co-segmentation through region matching,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2012, pp. 749–756.

[40] F. Meng, H. Li, G. Liu, and K. N. Ngan, “Image cosegmentation by incorporating color reward strategy and active contour model,” *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 725–737, Apr. 2013.

[41] F. Meng, H. Li, G. Liu, and K. N. Ngan, “Object co-segmentation based on shortest path algorithm and saliency model,” *IEEE Trans. Multimedia*, vol. 14, no. 5, pp. 1429–1441, Oct. 2012.

[42] W.-S. Chu, C.-P. Chen, and C.-S. Chen, “Momi-cosegmentation: Simultaneous segmentation of multiple objects among multiple images,” in *Proc. Asian Conf. Computer Vision*, Queenstown, New Zealand, Nov. 2010, vol. 6492, pp. 355–368.

[43] D. Batra, A. Kowdle, and D. Parikh, “icoseg: interactive co-segmentation with intelligent scribble guidance,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, USA, Jun. 2010, pp. 3169–3176.

[44] S. Vicente, C. Rother, and V. Kolmogorov, “Object cosegmentation,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2011, pp. 2217–2224.

[45] L. Mukherjee, V. Singh, and J. Peng, “Scale invariant cosegmentation for image groups,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2011, pp. 1881–1888.

[46] K. Chang, T. Liu, and S. Lai, “From co-saliency to co-segmentation: An efficient and fully unsupervised energy minimization model,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2011, pp. 2129–2136.

[47] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “Surf: Speeded up robust features,” *Comput. Vision Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[48] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012.

[49] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, “An empirical evaluation of deep architectures on problems with many factors of variation,” in *Proc. Int. Conf. Machine Learning*, New York, NY, USA, Jun. 2007, pp. 473–480.

Fanman Meng received the B.Sc. degree in computer science and technology and the M.Sc. degree in computer software and theory in 2006 and 2009 respectively. Since September 2009, he has been working toward the Ph.D. degree in the intelligent visual information processing and communication laboratory (IVIPC) at University of Electronic Science and Technology of China (UESTC). His research interests include image segmentation, object detection and visual attention.

Hongliang Li (SM'12) received his Ph.D. degree in electronics and information engineering from Xi'an Jiaotong University, China, in 2005. From 2005 to 2006, he joined the visual signal processing and communication laboratory (VSPC) of the Chinese University of Hong Kong (CUHK) as a Research Associate. From 2006 to 2008, he was a Postdoctoral Fellow at the same laboratory in CUHK. He is currently a Professor in the School of Electronic Engineering, University of Electronic Science and Technology of China. His research interests include image segmentation, object detection, image and video coding, visual attention, and multimedia communication system.

Dr. Li has authored or co-authored numerous technical articles in well-known international journals and conferences. He is a co-editor of a Springer book titled "Video segmentation and its applications". Dr. Li was involved in many professional activities. He is a member of the Editorial Board of the *Journal on Visual Communications and Image Representation*. He served as TPC members in a number of international conferences, e.g., ICME 2013, ICME 2012, ISCAS 2013, PCM 2007, PCM 2009, and VCIP 2010, and served as Technical Program co-chair in ISPACS 2009, and general co-chair of the 2010 International Symposium on Intelligent Signal Processing and Communication Systems. He will serve as a local chair of the 2014 IEEE International Conference on Multimedia and Expo (ICME). Dr. Li was selected as the New Century Excellent Talents in University, Chinese Ministry of Education, China, in 2008.

Guanghui Liu received the M.Sc. and Ph.D. degrees in electronic engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, Sichuan, China, in 2002 and 2005 respectively. In 2005, he joined Samsung Electronics, South Korea, as a senior engineer. Since 2009, he has been with the School of Electronics Engineering, UESTC, as an associate professor. His general research interests include digital signal processing and telecommunications, with emphasis on digital video transmission, and OFDM techniques. In these areas, he has published tens of papers in refereed journals or conferences, and received more than 10 patents (4 U.S. granted patents). Dr. Liu served as the publication chair of the 2010 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2010).

King Ngi Ngan (F'00) received the Ph.D. degree in electrical engineering from the Loughborough University in U.K. He is currently a chair professor at the Department of Electronic Engineering, Chinese University of Hong Kong. He was previously a full professor at the Nanyang Technological University, Singapore, and the University of Western Australia, Australia. He holds honorary and visiting professorships of numerous universities in China, Australia and South East Asia.

Prof. Ngan served as an associate editor of IEEE Transactions on Circuits and Systems for Video Technology, Journal on Visual Communications and Image Representation, EURASIP Journal of Signal Processing: Image Communication, and Journal of Applied Signal Processing. He chaired and co-chaired a number of prestigious international conferences on image and video processing including the 2010 IEEE International Conference on Image Processing, and served on the advisory and technical committees of numerous professional organizations. He has published extensively including 3 authored books, 6 edited volumes, over 300 refereed technical papers, and edited 9 special issues in journals. In addition, he holds 10 patents in the areas of image/video coding and communications.

Prof. Ngan is a Fellow of IET (U.K.) and IEAust (Australia), and an IEEE Distinguished Lecturer in 2006–2007.