Quality Guided Handbag Segmentation

Yan Wang, Sheng Li, and Alex C. Kot
Rapid-Rich Object Search (ROSE) Lab
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore 639798
Email: {wang0696,lisheng,eackot}@ntu.edu.sg

***Abstract*—In this paper, we address the problem of handbag segmentation, which is a challenging while important pre-processing for fashion related applications such as handbag tagging and search. Inaccurate segmentation will easily lead to other descriptions of color and shape of the handbag. We first design and extract a set of features for measuring the quality of the handbag segmentation based on some prior knowledge of handbag images. The quality of the handbag segmentation is then measured based on the weighted combination of these features. Guided by such quality measurement, we propose to segment the handbag image by a bottom-up super-pixel fusion. We conduct the experiment on a newly built handbag dataset as well as an existing branded handbag dataset. The results show that our segmentation algorithm performs favorably for handbags. The performance of handbag tagging and recognition is shown to be improved by incorporating such algorithm as pre-processing.**

***Index Terms*—Handbag segmentation, quality measurement, search, tagging**

I. Introduction

Over the years, the rapid growth of world wide online shopping attracts the attention of the public. Handbag, which has become an irreplaceable item in women's wardrobe ever since 1920s, is seldomly studied [1], [2]. Purchasing handbags online is common for many shoppers, but sometimes it is not convenient for them to find the handbags they want only based on key-word search. It is necessary to develop techniques for tagging or searching these handbags.

In order to have a good handbag tagging (e.g., color, shape) or search, proper segmentation is necessary before further processing because those attributes depends highly on the segmentation results rather than a rough localization. Several existing techniques can be adopted to identify the interest handbag region and separate it from the background. Image segmentation is to group and cluster pixels according to their similarity, proximity and good continuation [3], and each segmented region is corresponding to a meaningful part of the image. Normally, given a window that contains the object of interests, Grab Cut [4] settles the problem of separating objects from background. In order to solve the limitation that Grab Cut is likely to get stuck at weak local minima, a recent work in [5] propose a globally optimal solution in segmentation. Kim *et al.* [6] segmented objects by introducing shape prior. The figure-ground segmentation method proposed in [7] extracted overlapping windows from images and transferred segmentation masks from training images that are visually similar to windows in the testing images. In these method, class-specific knowledge is added either automatically or manually to the category-independent segmentation. Besides, saliency maps usually highlight entire salient object regions, which are widely used in object-of-interest image segmentation [8], [9], [10]. These existing works make use of energy terms which measure the distance between the object and background in graph cut, category-specific properties or saliency cues, however, their lack of handbag priors does not consider the handbag as a whole when handbag patterns appear distinctively, which would result in inferior performance when applied on the handbag images.

(a) (b) (c)

Fig. 1. (a) Examples of handbag images. (b) Foreground segmentation results by using existing algorithms (e.g., figure-ground segmentation [7]). (c) Annotated segmentation ground truth.

In this paper, we first propose a quality measurement for handbag specific segmentation problem, which is based on a weighted combination of a set of features. These features indicate the quality of the segmentation in different aspects, which are designed based on some prior knowledge of the handbag image. While the weights are learned through a binary SVM classifier. We next propose a super-pixel (following [11], we call the small region obtained from an over-segmentation as super-pixel) based handbag segmentation. This algorithm searches the foreground segment (in terms of super-pixels [12]) iteratively based on the quality of the segmentation, the one with the best quality will be chosen for the segmentation. The experiments show that the proposed quality measurement and segmentation algorithms are effective on handbags. We also demonstrate that our segmentation algorithm offers a great help for the applications such as handbag tagging or search.

978-1-4799-8058-1/15/$31.00

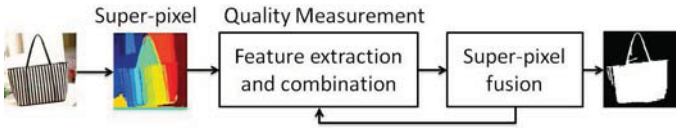

Fig. 2. Overview of proposed quality guided handbag segmentation.

Fig. 3. Shift each side of R into four new bounding boxes (R_l, R_r, R_u, R_b) of corresponding sub-segments.

II. Handbag segmentation

As examples shown in Fig. 1(a), handbags can appear in any colors and patterns. Some of the colors appear distinctively. Examples of handbag segmentations are shown in Fig. 1(b), where the white region indicates the foreground segment (terms as segment in the following discussions) and the black refers to the background. Inaccurate segmentation will lead to other colors or shapes of handbags, such that designing a proper segmentation strategy is challenging but useful. From Fig. 1 (a), we observe that handbags appearing on the handbag images of some e-commerce websites or someone's blogs share the following common properties: 1) they occupy a large proportion of the images; 2) they are roughly placed around the center; 3) there exists a certain level of color contrast between the handbags and the background and 4) theirs shapes are close to rectangles. A good handbag segmentation should be consistent with these properties. In this section, we propose a set of features to measure these properties, which are combined to measure the quality of the handbag segmentation. Then guided by the quality measurement, a super-pixel fusion technique is proposed to segment the handbag image, the flowchart of which is shown in Fig. 2.

A. Feature extraction

Given a handbag image I with a set of super-pixels S which is considered as one of its segmentation results, we design the following four features measuring the four aforementioned properties including area proportion, location, color contrast and shape.

Area proportion measures the occupation of the segment, which is computed as the ratio of areas of segment and the whole image.

Location: measures the location of the segment, which is computed as the average Euclidean-distance between each foreground pixel and the image center.

Color Contrast: measures the color contrast between the foreground and background, which is computed as the Chi-square distance of the normalized color naming features [13] between the foreground and background of the handbag image.

Shape: measures how close the segment is to a rectangular. The most intuitive way is to compute the proportion of the segment with its bounding box, as shown in Fig. 4(a). However, when the segment contains handbag strap, the direct computation of proportion does not have a good measure of the shape as shown in Fig. 4(b). Therefore, instead of calculating the proportion of the segment with its bounding box for the shape measurement, we measure the shape by computing the proportion of a part of the segment (termed as sub-segment) with its corresponding bounding box, as shown in Fig. 4(c).

Algorithm 1 Algorithm for finding the best sub-segment.

Require: Binary image mask I^B, tradeoff parameter λ, step-size $Step$

$R \Leftarrow boundingbox(I^B)$ ▷ Initialize R as the bounding box of the foreground segment in I^B

$[c_l, c_r, c_u, c_b] \Leftarrow coordinate(R)$ ▷ Coordinate of R's left, right, upper and bottom side

initialize $\boldsymbol{R} \Leftarrow R$ ▷ $\boldsymbol{R}$ is a sub-segment set

repeat

  $\{R_l, R_r, R_u, R_b\} \Leftarrow Shift(R)$ ▷ Shift each side of R based on Fig. 3

  $R \Leftarrow \arg\min_{R_i \in \{R_l, R_s, R_u, R_b\}} Cost(I^B, R_i)$ ▷ See Eq. 2

  $\boldsymbol{R} \Leftarrow \boldsymbol{R} \cup R$

until $|A_2| = 0$ ▷ $|A_2|$ is the number of pixels in the background region inside R

$R^* \Leftarrow \arg\min_{R \in \boldsymbol{R}} Cost(I^B, R)$

return R^*

Next, we introduce how to find the sub-segment which is the best for measuring the shape. Let I^B denote the binary segmentation mask (which is computed based on S) and R denote the bounding box of a sub-segment. We define the best R^* for measuring as

$$R^* = \arg\min_R Cost(I^B, R), \tag{1}$$

where $Cost(I^B, R)$ is a cost function defined by

$$Cost(I^B, R) = \frac{|A_2|}{|A_1|} + \lambda \frac{|A_3|}{|A_1|}, \tag{2}$$

where A_1 indicates the region of the sub-segment inside R, A_2 indicates the background region inside R, A_3 is the region of the segment outside R (see Fig. 4(c)), $|\cdot|$ counts the pixels inside different regions and λ is a tradeoff parameter. Finding the best R^* is an exhaustive search problem, which is very time-consuming. We propose an iterative procedure to estimate R^*. First of all, we initialize R as the bounding box of the segment. At each iteration, we shift the four sides of R into four new bounding boxes, which are termed as R_l, R_r, R_u and R_b respectively by a step-size of $Step$ (see Fig. 3). R is updated by $R = \arg\min_{R' \in \{R_l, R_r, R_u, R_b\}} Cost(I^B, R')$. This procedure continues until $|A_2| = 0$. Each iteration produces an updated bounding box, and from all such bounding boxes, we choose R^* which minimizes the cost function. The detailed process is shown in Algorithm 1.

Fig. 4. Bounding boxes of two segments (a)(b) and a sub-segment (c). The bounding boxes are shown in the red boxes.

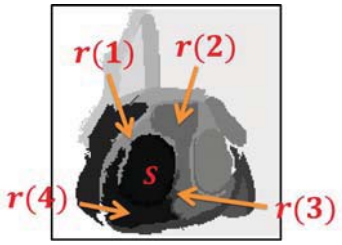

Fig. 5. S and its connected super-pixels at one iteration. A super-pixel is denoted as a region with the same gray scale level.

Once R^* is decided, the shape of the segment is measured as

$$area(I^B, R^*) = \frac{|A_1^*|}{|A_1^*| + |A_2^*|}. \quad (3)$$

where A_1^* and A_2^* indicate the regions of the sub-segment and background inside R^*, respectively.

B. Feature combination

For simplicity, we denote the four features (i.e., Area proportion, Location, Color Contrast and Shape) of image I with foreground segment S as $A(I,S)$, $L(I,S)$, $C(I,S)$, and $H(I,S)$, respectively. Intuitively, the larger the better for each of the proposed feature in terms of segmentation quality, however, the importance of each feature is different. We propose to measure the quality of S by combining these features as given below.

$$\begin{aligned} Q(I,S) =& w_1 A(I,S) + w_2 L(I,S) + \\ & w_3 C(I,S) + w_4 H(I,S), \end{aligned} \quad (4)$$

where $\{w_1, ..., w_4\}$ denote the weights for different features. The larger the value of Q, the better the segment S is for the handbag image I. The feature weights are obtained by the following off-line training process. Among a set of training images, we extract the proposed features from their ground truth segments as positive data and the features from other randomly sampled segments as negative data. We sample positive and negative data sets with equal size, and the weights $\{w_1, ..., w_4\}$ of the four features are learned by adopting a linear SVM.

It should be noted that our quality measurement does not limited to measure S which is a combination of super-pixels, it can be generalized to any segmentation results.

C. Quality based super-pixel fusion

The concept of our quality based super-pixel fusion is to find the best combination of super-pixels S^* such that the quality measurement of the segment is maximum

$$S^* = \arg\max_S Q(I,S) \quad (5)$$

We propose a bottom-up fusion technique by iteratively searching for the best super-pixels. We initialize S to be the super-pixel around the central of I. At each iteration, as shown in Fig. 5, we search among the super-pixels (termed as $r(1)$ - $r(4)$) which are connected with S, and merge the one which leads to the highest segmentation quality Q into S. The iterative procedure ends when all super-pixels are merged. Each iteration produces an updated merged super-pixel segment, and among all such merged super-pixel segments, we choose the one S^* that maximizes the segmentation quality.

III. EXPERIMENTS

We evaluate our algorithm on a newly built handbag dataset for segmentation and quality measurement of handbag segmentation, the images of which are collected from *Amazon.com*. There are 835 handbag images in total from the dataset, some examples are given in Fig. 1(a). Each image is annotated with a binary segmentation mask which serves as the ground truth as shown in Fig. 1(c). we adopted the work in [12] as the initial result for handbag segmentation. A set of 235 images are randomly selected from the dataset for the feature weight training, and the rest images are used for testing.

A. Evaluation on handbag segmentation

We compare the proposed segmentation algorithm with several prior works on the handbag images, including the saliency maps extracted from Low Rank matrix recovery (LR) [14] and integration of Global uniqueness and Color spatial distribution cues (GC) [9], figure-ground segmentation (ST) [7] and GrabCut in One Cut (One-Cut) [5]. The evaluation is based on the mean absolute error (MAE) and f-measure. For saliency map based segmentation methods LR and GC, we binarize the saliency maps by thresholding to obtain the segmentation mask. As suggested in [15], [16], we report the f-measure of LR and GC using the segmentation masks which achieve the maximum f-measure. Similarly, we report the MAE of LR and GC using the segment masks which achieve the minimum MAE. For ST, One-Cut and our proposed method, both directly output binary segmentation masks, from which the f-measure and MAE can be computed. The comparison results are shown in Table I. Noted that for both LR and GC, the thresholds are different in the test set that produces the maximum f-measure or the minimum MAE. It can be seen that our proposed handbag segmentation algorithm outperforms other methods in terms of both the f-measure and MAE.

TABLE I
COMPARISON RESULTS FOR HANDBAG SEGMENTATION AND AVERAGE HANDBAG SEGMENTATION QUALITY.

Measurement	LR	GC	ST	One-Cut	Ours
MAE(%)	32.22	6.25	8.50	6.25	**5.71**
f-measure(%)	70.83	92.75	89.32	91.88	**93.72**
$Q(I,S)$	2.09	3.55	3.32	3.52	**4.49**

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

Fig. 6. Visual comparison of segmented masks computed by different techniques. (a) original image, (b) LR, (c) GC, (d) ST and, (e) One-Cut, (f) ours.

We also show the visual comparison of the handbag segmentation masks computed using different techniques in Fig. 6(b)-(e). For LR or GC, we binarize the saliency map with a fixed thresholding such that the MAE is minimized. It can be observed that our proposed algorithm is more likely to segment the handbag regions rather than distinct patterns that appear on handbags.

B. Evaluation on quality measurement of handbag segmentation

In order to evaluate the effectiveness of the proposed quality measurement for handbag segmentation, we compute the average segmentation quality on the test set for different segmentation methods. We also show in Table I the average quality measurements for the handbag segmentation using different method. It can be seen that the measured quality Q is in consistent with the actual performance based on MAE and f-measure.

C. Applications of handbag segmentation

We also evaluate our segmentation algorithm for different applications and prove that segmentation can be treated as a useful pre-processing. First, we give some quantitative results on attribute identification. Attribute identification can be useful for handbag tagging or retrieval. We adopt color attribute here and manually label each image in our dataset with one of the colors defined by the 11-dimensional color naming feature [13] , including *black*, *blue*, *brown*, *cyan*, *gray*, *green*, *orange*, *pink*, *purple*, *red*, *silver*, *white*, *yellow* and *multi-color*. We adopt a multi-class linear SVM as the classifier and obtain 74.62% in color identification accuracy for foreground segment and 70.78% for the whole image.

Then, we incorporate our segmentation algorithm as pre-processing for handbag recognition. Concretely, a handbag image is pre-processed by adopting proposed segmentation algorithm, and the foreground image is extracted by placing a bounding box that covers the foreground segment. From the bounding box for each handbag image, we extract Caffe feature [17] which is an implementation of the 7 layer ConvNets. Here we used the Caffe feature of 6th layer. We follow the same way of splitting the dataset into training and testing as in [2] and adopt a linear SVM classifier afterwards. The recognition accuracy obtained by proposed segmentation achieves 82.40%, which is superior than the whole image which only obtains 76.80% in accuracy.

IV. CONCLUSIONS

In this paper, we present a handbag segmentation method to facilitate handbag tagging or search. In this method, a set of features are proposed based on handbag priors, and the features are integrated to measure the quality of handbag segmentation. Then guided by the features, we propose an iterative procedure to search for the best combination of super-pixels. We evaluate the segmentation performance and the quality measurement on a newly built handbag dataset. We also evaluate our segmentation algorithm for different applications such as automatic attribute tagging and handbag search. The results show that our method achieves encouraging results.

ACKNOWLEDGMENT

This research was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its IDM Futures Funding Initiative and administered by the Interactive and Digital Media Programme Office.

REFERENCES

[1] T. L. Berg, A. C. Berg, and J. Shih, "Automatic attribute discovery and characterization from noisy web data," in *Proceedings of the 11th European conference on Computer vision: Part I*, 2010, pp. 663–676.

[2] Y. Wang, S. Li, and A. C. Kot, "Category-separating strategy for branded handbag recognition," in *Communications, Control and Signal Processing (ISCCSP), 2014 6th International Symposium on*, May 2014, pp. 61–64.

[3] B. Peng, L. Zhang, and D. Zhang, "A survey of graph theoretical approaches to image segmentation," 2012.

[4] C. Rother, V. Kolmogorov, and A. Blake, ""grabcut": Interactive foreground extraction using iterated graph cuts," in *ACM SIGGRAPH 2004 Papers*, 2004, SIGGRAPH '04, pp. 309–314.

[5] M. Tang, L. Gorelick, O. Veksler, and Y. Boykov, "Grabcut in one cut," in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, 2013, ICCV '13, pp. 1769–1776.

[6] J. Kim and K. Grauman, "Shape sharing for object segmentation," in *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VII*, 2012, pp. 444–458.

[7] D. Kuettel and V. Ferrari, "Figure-ground segmentation by transferring window masks," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 0, pp. 558–565, 2012.

[8] J. Han, K.N. Ngan, M. Li, and H.-J. Zhang, "Unsupervised extraction of visual attention objects in color images," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, no. 1, pp. 141–145, Jan 2006.

[9] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *IEEE ICCV*, 2013, pp. 1529–1536.

[10] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE TPAMI*, 2014.
[11] B. Fulkerson, A Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *Computer Vision, 2009 IEEE 12th International Conference on*, Sept 2009, pp. 670–677.
[12] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vision*, vol. 59, no. 2, pp. 167–181, Sept. 2004.
[13] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *Trans. Img. Proc.*, vol. 18, no. 7, pp. 1512–1523, July 2009.
[14] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 0, pp. 853–860, 2012.
[15] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned Salient Region Detection," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, 2009, pp. 1597 – 1604.
[16] A. Hornung, Y. Pritch, P. Krahenbuhl, and F. Perazzi, "Saliency filters: Contrast based filtering for salient region detection," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 0, pp. 733–740, 2012.
[17] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.