

A Garment Image Segmentation Method Based on Salient Region and JSEG

Liu Zhengdong*

Computer Information Center, Beijing Institute of Fashion Technology, Beijing 100029, China.

* Corresponding author. Email: jsjlzd@bift.edu.cn

Manuscript submitted July 22, 2015; accepted September 17, 2015.

doi: 10.17706/jsw.10.11.1274-1282

Abstract: Many model pictures will be adopted in the network marketing of garment products because this method is the most intuitive. Human eyes can easily identify the garment styles, while computer programs struggle to automatically do so. This paper proposes a method for garment image segmentation in which the JSEG algorithm is applied to the image salient region segmentation, and the background region and dressing region are identified with the help of face detection. This method was proven effective based on actual image samples tested at an online shopping mall, which lays a foundation for extracting the garment structure and characteristics as well as visualized retrieval.

Key words: Garment, image segmentation, JSEG, salient region detection, human face detection.

1. Introduction

The efficient and reliable retrieval of the dressing region from an image and the identification of the garment structure are vital to content-based garment image processing. The image segmentation method divides the image into several meaningful regions. The result can be used for post-processing, such as the recognition of garment styles. A reliable and efficient image segmentation algorithm will greatly simplify the post-processing.

Edge detection is commonly used in image segmentation processing. For example, in Canny operator detection [1], the brightness gradient can be used to detect edges, which can identify the entire border of the target object. However, edge detection generally will not create a closed contour. Isolated edge lines complicate post-processing steps, such as target identification. For the curve evolution method, the initial curve continuously evolves until the border surrounding the object is formed in the image. The curve evolution method [2] is applicable to dividing the entire image into a plurality of separate regions, but it will create small meaningless areas for image segmentation with a cluttered background. Furthermore, the entire image may not need to be segmented. To quickly segment the background and foreground, the graph cuts method and GrabCut method, which has been improved by Rother *et al.* [3], segment the image via minimal cut/maximum flow, but they require a small amount of human-computer interaction to separate the background and interest points.

The thought of priority in image segmentation [4], [5] is based on the detection of interest points. First, the region, like a face, that the observer is interested in will be detected and briefly positioned. For example: for a model themed photo, the observer will soon focus more on the model dress, ignoring other irrelevant parts. Hare [6] took the area beneath the face as the basis of garment region segmentation. Itti [7] defined

the salient concept of an image and proposed to acquire the salient detection by using the center-surrounding difference of a multi-scale image feature. Zhang and Ma [8] proposed an alternative approach of local contrast and analysis to produce salient images and extend them via a fuzzy growth model. Harel *et al.* [9] highlighted the salient parts by normalizing the characteristic pattern of Itti *et al.* and combined them with other salient images. Via the linear combination of the Gaussian pyramid image, Liu *et al.* [10] advanced the calculation method of multi-scale contrast. In recent years, some researchers [11] combined image visual saliency computation and the notion of image segmentation and proposed the object segmentation algorithm, which yielded good results. Huang [12] combined salient computing and Graph Cut and attempted garment region segmentation.

This paper proposed an automatic segmentation algorithm of the garment region to retrieve its structure by combining the JSEG image segmentation and visual saliency calculation. Unlike the existing method, this method can segment a garment region without training data and manual intervention and initially identify information about garment structure and shape.

2. JSEG Algorithm for Image Segmentation

The image segmentation method was used first in order to improve the general shape of a garment from the image and minimize the effects of texture, folds, shadows and other internal patterns. Vincent Martin [13] tested several classical segmentation algorithms, including Meanshift, JSEG, EGBIS, CSC, SRM and region growth, and validated the algorithm on the BSD (Berkeley Segmentation Dataset) database [14]. According to the results, the JSEG algorithm is more efficient for garment images. Li [15] studied image segmentation on textile printing images and validated the efficiency of the JSEG algorithm for garment image segmentation. The JSEG algorithm [16], [17] is an unsupervised image segmentation method based on the color and texture features of the region. The basic process consists of two processing stages: color quantization and spatial segmentation, both of which are independent of each other.

2.1. Color Quantization

Color quantization is the classification of all of the colors in the image. The distribution of colors does not need to be considered in the whole image. First, the image color space is converted to the LUV color space. This expression agrees with the perceptual features of the human visual sense, and the similarity between different colors can be measured via the Euclidean distance, which can help to describe colors. The PGF (Peer Group Filtering) [16] is then used to smoothly filter the image, which preserves the edge information while removing the particle noise influence.

Define a point in the image, $x_0(n)$, as the center of the window with size $W \times W$ in the neighborhood of $x_0(n)$ as $p(n) = \{x_i(n), i=0,1,\dots, m(n)-1\}$, where $x_i(n)$ is the image point in the window and $m(n)$ is the total number of image points. According to the Fisher criterion, the corresponding $m(n)$ is acquired. The $x_{new}(n)$ can be calculated with the weight ω of each point to replace the original center point.

$$x_{new}(n) = \frac{\sum_{i=0}^{m(n)-1} \omega_i p_i(n)}{\sum_{i=0}^{m(n)-1} \omega_i} p_i(n) \in P(n) \quad (1)$$

$$\omega_i = e^{\frac{(p_i(n) - \mu_i)^2}{-2\sigma_i^2}} \quad (2)$$

where μ_i and σ_i are the mean and variance of $p(n)$, respectively.

The number and categories of color are then determined. Assume $T(n)$ is the maximum distance of the corresponding neighborhood of $x_0(n)$. In $T(n) = d_{m(n)-1}(n)$, $d_i(n)$ represents the ascending distance between the center point $x_0(n)$, and all of the points in the window, $W \times W$, and the mean value T_{av} of $T(n)$ indicates the whole image smoothness. Thus, the number of color categories can be expressed as follows:

$$N = \beta T_{av} \quad (3)$$

where β is a constant (here, $\beta = 2$).

Finally, the GLA algorithm [16] is used for quantized processing. Color quantization will divide colors into several representative categories that divided the image into different regions and are marked with the appropriate label for each category. All pixels in the image will be replaced with the corresponding color category numbers, called a color category diagram.

2.2. Space Partitioning

Space partitioning is carried out based on the class diagram, which is formed during the color quantization stage. The class diagram generates the J image, and the region growth method then is applied to segment the J image.

The J image is calculated at the beginning of each iteration. If the pixel value and the size in a region are too small, it cannot become the new partition. The pixel values that are smaller than the average value of J will be integrated into the adjacent regions. This step is repeated until the J image of the small window is minimized. The final phase is the consolidation phase, which reduces excessive segmentation by calculating the histogram distance.

Assume Z as the set of N data points in the class diagram, $z = (x, y)$, $z \in Z$, where m is the mean value:

$$m = \frac{1}{N} \sum_{z \in Z} z \quad (4)$$

Divide Z into C classes Z_i ($i = 1, 2 \dots C$), where M_i is the mean value of N_i data in Z_i class. Thus,

$$m_i = \frac{1}{N_i} \sum_{z \in Z_i} z \quad (5)$$

S_T is defined as the total variance of the data points, and S_W is defined as the total variance of the data points that belong to the same class:

$$S_T = \sum_{z \in Z} \|z - m\|^2 \quad (6)$$

$$S_W = \sum_{i=1}^C S_i = \sum_{i=1}^C \sum_{z \in Z_i} \|z - m_i\|^2 \quad (7)$$

The calculated value of J is

$$J = \frac{(S_T - S_W)}{S_W} \quad (8)$$

Regional growth can be carried out after obtaining the J image. The threshold is defined as $T_j = \mu_j + \alpha\sigma_j$, where μ_j is the mean value of J in the non-seed region and σ_j is the variance of J . α is gradually selected from $[-0.6, -0.4, -0.2, 0, 0.2, 0.4]$. T_j is then calculated. The 4-way connector allows the value of J to be smaller than the candidate seed points formed by T_j .

The candidate region of the seed point should be larger than a certain area. Candidate regions that are too small can be considered non-seed areas. The average value of J in the unincorporated area is calculated, and the pixels smaller than the average value are connected to the growth region. After the region growth, the initial segmentation results of the images can be acquired. Generally, over-segmentation will be evident. These divided areas can be merged according to the color similarity. The Euclidean distance between the color histograms of two adjacent regions is defined as follows:

$$D_{CH}(i, j) = \|P_i - P_j\| \quad (9)$$

where P is a color histogram. The distance between the color histograms of two adjacent regions is calculated and stored in the distance table. The minimum distance between two regions in the table will be merged. This process continues until the distance between two adjacent regions reaches a preset threshold.



Fig. 1. Renderings after garment image segmentation by JSEG segmentation algorithm.

In Fig. 1 image segmentation groups simplify the inner pattern, effectively removing the textures, folds and shadows. White lines indicate the outline of each divided region. Each divided region is closed. The garment has advantages over other algorithms in terms of over-segmentation, and it also has some problems:

- 1) The JSEG algorithm does not distinguish the background from the foreground (dressing and body), it segments all images.
- 2) The existing shadow, words, jewelry and other objects in the images cannot be excluded via a priori knowledge, such as the body's projection on the background, brand and signs in the picture and others in Fig. 1.

To solve the above problems, improvements based on the segmenting JSEG are necessary.

3. Region Determination by Combining Saliency Calculation and Face Detection

Customers are only interested in the model's dressing region for a variety of apparel products listed in online shopping malls. Therefore, these regions need to be directly detected in order to obtain efficient processing results. People can easily determine the salient regions of the image; thus, researching the problem of image salient region detection in the field of artificial intelligence is important. Cheng [18] proposed the Histogram Contrast (HC) based the salient value detection algorithm according to the statistical characteristics of the input image colors.

The salient value of I_k in the image I is defined as follows:

$$S(I_k) = \sum_{\forall I_i \in I} D(I_k, I_i) \quad (10)$$

Formula (10) ignores the spatial relations of the colors in the images. Thus, pixels with the same color value will have the same salient calculation value. The pixels with the same color value, C_j are considered a collection; thus, the salient value of each color can be calculated as follows:

$$S(I_k) = S(C_l) = \sum_{j=1}^n f_j D(c_l, c_j) \quad (11)$$

where, C_l is the color value of the pixel I_k , n is the total number of colors contained in the images and f_j is the probability of C_l appearing in the images.

Next, the saliency value of each color can be improved via a smoothing operation. The saliency value of each color is replaced by the weighted average value of the similar color as follows:

$$S'(c) = \frac{1}{(m-1)T} \sum_{i=1}^m (T - D(c, c_i)) S(c_i) \quad (12)$$

where $T = \sum_{i=1}^m D(c, c_i)$ shows the relation of color c and its m distances nearest to c_i .



Fig. 2. Shows the calculation effects of the FT, HC, LC, RC and SR algorithm according to [18].

Cheng [18] presented the RC algorithm with better results, but the calculation efficiency is relatively low. Furthermore, Fig. 2 shows that the RC algorithm also expanded the area of the image salient region, which is not ideal for a garment image. The HC algorithm is fast with accurate results. According to the actual effect and calculation efficiency, we chose the HC algorithm. With help of the CHR method [19] to detect human faces, we can relatively easily identify the approximate range of the clothing area. The method is implemented in the saliency region as the mask of the segmentation image, and the regions with a high saliency value can become candidate regions. The identification results are connected to the detected human face region.

4. Experimental Results and Analysis

Because references for the garment image database are not available, we collected more than 20 types of model dress pictures by hand. These pictures mainly originated from several garment e-commerce sites. The pictures are all processed to a width of 220 and height of 220 pixels, converted into RGB color images and manually marked with the image structure. Using the collected images, we tested the performance and effectiveness of the algorithm. The experimental software environment was the development environment in Visual Studio 2013 for Desktop, and the implemented programming language was C#. The computer hardware environment was a CPU Intel Core Duo E8300 (2.83GHz) with a memory capacity of 4GB.

Fig. 3 shows the segmentation process and the final results of four samples. The processing of each sample includes four renderings, which are the original image, JSEG segmentation image, saliency detection figure and segmentation results figure from left to right. The white rectangle represents the detected face region.

To illustrate the effectiveness of the proposed algorithm, it was experimentally compared with other methods. Fig. 4 shows the effect of the contrast between the literature [6] algorithm (abbreviated as A) and GrabCut algorithm (abbreviated as B). Algorithm A obtains the garment region by detecting the region beneath the face. Because a source code is not available, a self-made program is used. B conducts Otsu threshold segmentation based on the GrabCut algorithm after the human-computer interaction. Algorithm A can identify the body's contours. However, the human body shade cannot be reasonably divided. Algorithm B is better than A in terms of retrieving the body shape, but it does not consider the segmentation between the garment and body region. This algorithm is suitable for extracting different human postures and can distinguish nearly every part of the garment.

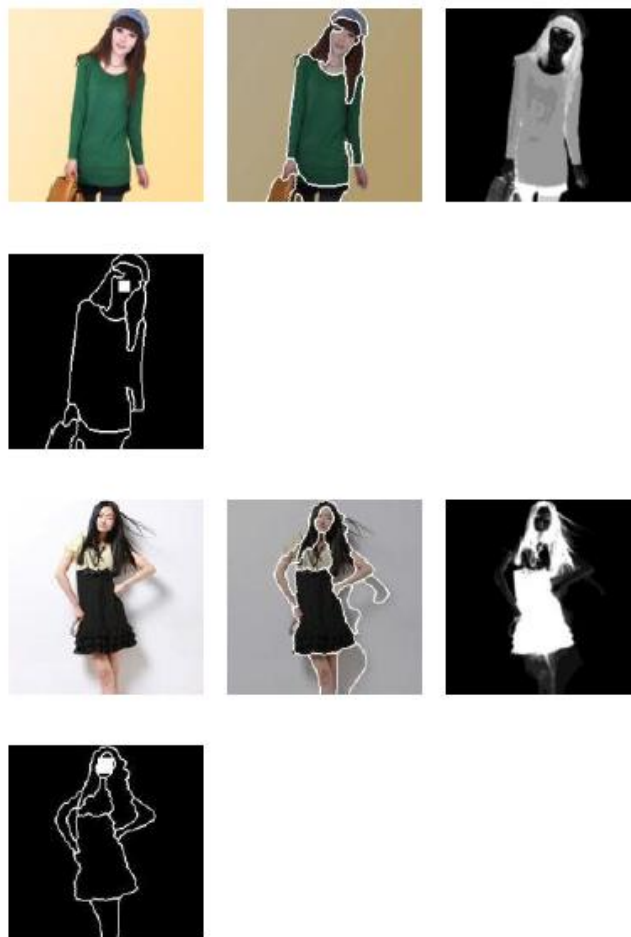


Fig. 3. Shows part of the sample image segmentation results of the garment (images from left to right are the original image, JSEG segmentation image, saliency detection figure and segmentation results figure).

To test the algorithm more quantitatively, the recall and precision rates are used as metrics to analyze the segmentation results, where the recall rate is $\text{Recall} = S'/M$, which represents the ratio of the segmentation results and manual marking. The precision is $\text{Precision} = S'/S$, which represents the proportion of the accurate part in the current segmentation results, where M is the number of divided regions in the manually marked sample images, S is the number of images after the division and S' is the number of accurate

segmentation results. Table 1 shows the index value, including the recall and precision rates, as well as the average running time.

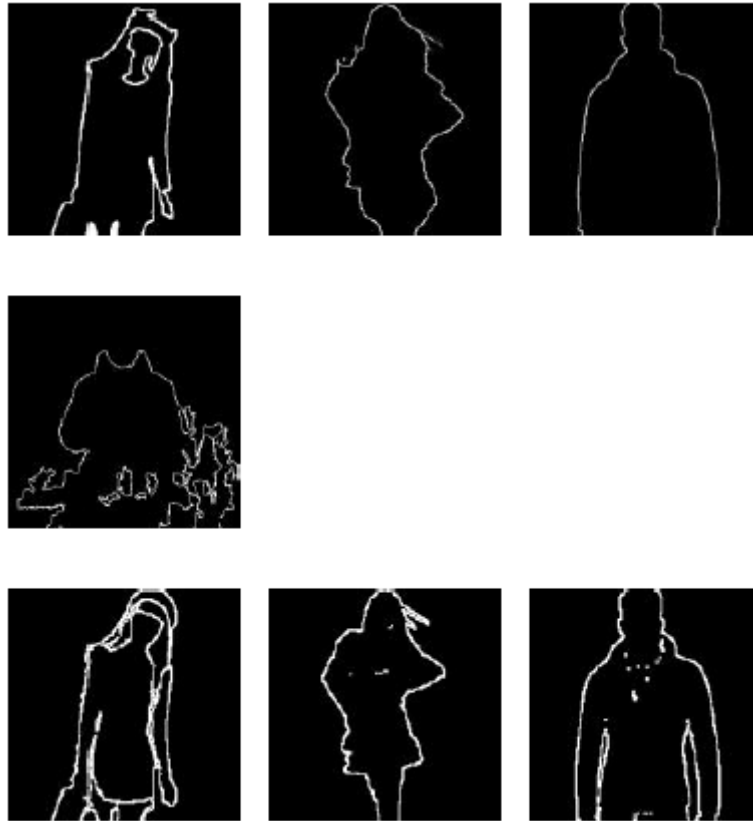


Fig. 4. Algorithm comparison experiments (the picture above shows the results of algorithm A, the middle picture shows the results of algorithm B, the picture below is the algorithm results of this paper, which are the same as those in Fig. 3 and without the face detection rectangle).

The statistics in Table 1 show that algorithm A is sensitive to the changes in body posture and that the recall and precision rates are relatively low. The operating efficiency of algorithm B is high, and the recall and precision rates have been greatly improved. The algorithm in this paper improved the precision rate while further shortening the running time, although the recall rate is slightly lower than that of the GrabCut algorithm.

Table 1. Comparison of Algorithm Indicators

Index	Method A	Method B	Our Method
Average time	1.5 sec	1.2 sec	0.9 sec
Recall	0.58	0.87	0.86
Precision	0.49	0.76	0.90

5. Conclusions

This paper presented an initial discussion of the garment region retrieval of human dressing images and the image segmentation of the structure; the negative impact of the flexible nature of the garment, which has folds, textures, lighting, draping and other non-rigid properties, are all solved; the segmentation of the

target garment in the image via the JSEG algorithm and saliency identification is advanced, avoiding the intervention step of human-computer interaction and improving the efficiency of batch processing. The experimental results show that the algorithm works well in terms of efficiency and accuracy. Taking into account its application in engineering, the garment region segmentation conditions on a complex background can be further studied to improve the timeliness and robustness of the system. The experiment also showed that the segmentation results of the algorithm can miss a garment's specific region. Future studies will examine methods to avoid this problem and conduct a more accurate saliency analysis of the image.

Acknowledgment

This work was financially supported by the National Science and Technology Support Program (2013BAH41F00) and the Open Project of the Beijing Key Laboratory of digital and interactive media (2014ZK-06).

References

- [1] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 679-698.
- [2] Cremers, D., & Funka-Lea, G. (2005). Dynamical statistical shape priors for level set based sequence segmentation. *Variational, Geometric, and Level Set Methods in Computer Vision*, Springer Berlin Heidelberg.
- [3] Rother, C., Kolmogorov, V., & Blake, A. (2004). Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3), 309-314.
- [4] Agarwal, S., Awan, A., & Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11), 1475-1490.
- [5] Tseng, C. H., Hung, S. S., & Tsay, J. J. et al. (2009) An efficient garment visual search based on shape context. *WSEAS Transactions on Computers*, 8(7), 1195-1204.
- [6] Hare, O. N., & Smeaton, A. F. (2009). Context-aware person identification in personal photo collections. *IEEE Transactions on Multimedia*, 11(2), 220-228.
- [7] Itti, L. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transaction on Pattern Analysis & Machine Intelligence*, 20(11), 1254-1259.
- [8] Ma, Y. F., & Zhang, H. J. (2003). Contrast-based image attention analysis by using fuzzy growing. *Proceedings of the Eleventh ACM International Conference on Multimedia* (pp. 374-381).
- [9] Harel, J., Koch, C., & Perona, P. (2006) Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 545-552.
- [10] Liu, T., Yuan, Z., Sun, J. et al. (2011). Learning to detect a salient object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2), 353-367.
- [11] Liu, Y., Huang, B., & Sun, H. (2013). Image segmentation based on visual saliency and graph cuts. *Journal of Computer-Aided Design and Computer Graphics*, 25(3), 402-409.
- [12] Huang, L., Tian, X., & Zhang, Y. D. (2012). Pose-invariant clothing segmentation based on saliency detection and graph cuts. *Journal of Computer-Aided Design & Computer Graphics*, 24(5), 620-627.
- [13] Image segmentation experimentations. Retrieved, from <http://www-sop.inria.fr/pulsar/personnel/Vincent.Martin/segmentation/>.
- [14] Berkeley segmentation dataset. Retrieved, from <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>
- [15] Pengfei, L. I., Wang, G., Jing, J., & Jiao, K. (2010). Segmenting color region of textile printing pattern

- image based on algorithm of JSEG. *Journal of Textile Research*, 31(5), 137-140.
- [16] Deng, Y., & Manjunath, B. S. (2001). Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 23(8), 800-810.
- [17] Victor, M. A. (2010). Preferential image segmentation using j segmentation based on color, Shape and texture. *International Journal of Engineering and Technology*, 2(2), 131-135.
- [18] Cheng, M. M., Zhang, G. X., Mitra, N. J., Huang, X., & Hu, S. M. (2011). Global contrast based salient region detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 409-416).
- [19] Yang, S., & Thomas, L. (2006). *Context-aided Human Recognition - Clustering*. Computer Vision - ECCV 2006, PT 3, Proceedings, 382-395.



Liu Zhengdong received the B.S. degree in electronic engineering from North China Electric Power Institute in 1993, the M.S. degree in computer science from Shenyang Institute of Technology in 2002, and the Ph.D. degree in the department of computer science of Nanjing University of Science and Technology in 2005. His research interests include image process and 3D modeling for garment industry.

He is currently an associate professor in the Computer Information Center of Beijing Institute of Fashion Technology (BIFT).