Pattern segmentation method based on super-pixel multi-feature fusion

Xiaohong Gu, and Shuyuan Shang*
Beijing Institute of Fashion Technology, Beijing, 100029, China

Keywords: Clothing patterns, Super pixel, Image segmentation, Deep learning.

Abstract. The paper presents an optimized segmentation algorithm for clothing patterns. It uses the depth model to improve the image quality. Super-pixel pre-segmentation enhances the ability to capture the edge information of patterns. LBP texture information, HSV color information and image pixel space information are used to optimize the parameters of Gaussian mixture model. The energy function is optimized by maxflow/mincut to realize the fast segmentation of clothing patterns. The simulation results show that the segmentation accuracy and precision of the algorithm are improved.

1 Introduction

With the development of digital era, artificial intelligence, big data and real industry are deeply integrated. The transformation and upgrading of traditional industries realized the digital storage and display of intangible cultural heritage. At present, the means of extracting patterns mainly depends on manual copying. It is time-consuming and laborious. As one of the industries with the most commercial interests at present, the digital research of clothing patterns is of great value to the industry and fashion industry. Hou Xiaogang[1] made graying and binarization on the pre-processed true color image. Combining morphological connected domain labeling and CV model contour evolution method. Realize the automatic segmentation of ethnic costume pattern gene sub map. Wang Hui[2] used bilateral filtering for image enhancement of printed fabric patterns. GrabCut algorithm uses saliency map as prior information input to realize pattern element extraction and reduce manual interaction. Zhao Lei[3] used Gaussian mixture model (GMM) clustering to achieve unsupervised segmentation by preprocessing image super pixels, combining HSV color information, spatial information and LC saliency map. Furthermore, on the basis of multi-scale, the edge connectivity is fused to optimize the saliency map, and the pattern is extracted using the graph cut method. According to the current research, the paper proposes an image segmentation algorithm based on multi-feature fusion for clothing patterns.

* Corresponding author: shang@bift.edu.cn

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
2 Image enhancement

In the process of image collection, images usually carry a lot of unnecessary information due to unavoidable environmental or technical limitations. Image enhancement can selectively highlight ROI areas or suppress unwanted features, and improve image effect to a certain extent. Traditional algorithms can be divided into histogram equalization based methods and Retinex theory based methods. The HE[4] method only optimizes a single pixel to achieve overall brightness improvement, ignoring the correlation between regional pixels and destroying image details. Retinex[5] theory considers the correlation of image area features, but it will cause color distortion artifacts for uneven lighting. LLNet[6] pioneered the use of stacked denoising self encoder to achieve weak light image enhancement and denoising. Its source code is difficult to achieve and is aimed at single channel gray image. Other CNN based methods, such as LLCNN, based on VSDR and ResNet[7], propose a dual branch and residual module. Although it is convenient to implement, it has color distortion, and can not handle brightness/contrast and denoising at the same time. LightenNet[8] will result in underexposure[9], as shown in Fig. 1.

It is difficult to achieve high quality image enhancement with simple network. MBLLEN[10] (Multi-branch weak light network enhancement model) decomposes the image enhancement problem into sub-problems of denoising, texture preservation, color correction, etc., and features are fused and output by multi-branches. Feature extraction module (FEM), enhancement module (EM) and fusion module (FM) constitute the core components of the network. The network structure is shown in Fig 2. The feature extraction module is a single branch network composed of 10 convolution layers. No pooling layer is added to deal with the image surface loss. The final enhancement effect is obtained by concatenating the output of all EMs modules to the color channel and using the weighted sum of 1 convolution kernel. Compared with other methods, MBLLEN can achieve satisfactory visual enhancement effect.

The weight of the whole network model is initialized with the pretrained VGG19 benchmark network and the parameters of the training model with Poisson noise provided by the author to reduce the training time cost of the neural network model. For clothing
patterns of any scale $I \in RW \times H \times 3$, the contrast is improved by adjusting the parameter percentage to 5% and 95% linear amplification network output. Set the gamma transform value to 8 to increase saturation. Obtain multiple 32 channel feature map weighted fused enhancement images $\hat{R} = F(I; \theta)$, where $F$ represents trainable parameters $\theta$’s network. Image enhancement effect is shown in Fig. 3.

![Fig. 3. MBLLEN image enhancement.](image)

### 3 Prior knowledge

Since only rectangular boxes are determined, as long as a rough threshold value is taken, the requirements for rectangular boxes can be met. Therefore, binary connected domain analysis is carried out (Fig.4.). In terms of threshold processing, the threshold value is generally set as the empirical value of 127, but it will not always get better results in actual situations; OTSU is used to automatically generate a threshold to divide the image into target and background, which is suitable for the case of double peaks in the image. After threshold segmentation, the pixel value of the image background is 0, and the pixel value of the target is 1. The formula is:

$$g = w_0 \times (u_0 - u)^2 + w_1 \times (u_1 - u)^2$$

wherein, $u_0$ is the mean of foreground pixels, and $w_0$ is the probability of pixels whose gray level is less than $T$; $u_1$ is the mean of the background pixels and $w_1$ is the probability that the gray scale is greater than $T$. $u$ is the mean value of all pixels, and $T$ is the best threshold when $g$ is the maximum. The division is as follows:

$$dst(i, j) = \begin{cases} \max \text{val, img}(i, j) > \text{threshold} \\
0, \text{otherwise}
\end{cases}$$

![Fig. 4 Border coordinate acquisition.](image)

After threshold segmentation, a black and white image is obtained, that is, each pixel on the image has only two possible gray levels. For any pixel $(x, y)$ and its surrounding pixels to form a neighborhood, the Euclidean distance between each other is controlled within $\sqrt{2}$. Label distribution is carried out for different connected areas through connectivity analysis of binary images. The formula is as follows:

$$L_e = \begin{cases} 0, \cdots I_e = 0 \\
I_e(l \leftarrow l + 1), I_e \neq 0 \text{ and } I_1 = 0, \forall l \\
\min_{i | I_i = 1(L_1)} I_e \neq 0 \text{ and } I_1 \neq 0, \exists l
\end{cases}$$
where, \( I_e \) is the pixel value of pixel \( e \), \( L_e \) is the label to be assigned, and \( i \) is the pixel in the 8 neighborhood of pixel \( e \). In the array listed in the last figure in Fig. 4, the first row is the coordinate point of the entire image. Based on the image area information in the last column of the vectors, small areas of the connected area are filtered and the corresponding area-size coordinate vectors are extracted.

In order to improve the segmentation efficiency and reduce the complexity of the algorithm, the simple iterative clustering algorithm is used to segment the image, and the image processing is extended from the pixel level to the image block stage, and the image is characterized on the basis of superpixel area. At present, typical superpixel algorithms mainly include Superpixel Lattice, SLIC, Watersheds, TurboPixel, NormalizedCuts, etc. SLIC is ideal in terms of running speed, compactness of generating superpixels, edge dependence, etc[11].

SLIC greatly improves the execution speed of the algorithm by reducing the search space calculated by the distance between each cluster center. However, SLIC algorithm needs to manually configure the compactness parameter size for all hyperpixels during image pre segmentation, especially when the texture feature areas are inconsistent, there will be hyperpixel blocks of different sizes. Due to factors such as production technology and fabric attributes, clothing patterns are basically highly textured, so it is necessary to use SLIC Zero to adaptively select the correct compactness parameters for pattern images, as shown in Fig. 5.

![SLIC and SLIC0](https://example.com/slic.png)

**Fig. 5** Super pixel segmentation.

After simple iterative clustering, the super pixel edge can cover the image object edge (black solid line) in the given texture prominent area, and has good boundary adhesion. In terms of overall appearance, it is obviously different from the dense edges obtained by SLIC algorithm. At the same time, pixels with similar characteristics are aggregated to achieve image dimensionality reduction, which is helpful to eliminate abnormal pixels in the pattern background.

The essence of image texture features is to quantify the gray levels of different levels in the region composed of multiple pixels, which reflects the local texture information of the image. After constant repetition, the global characteristics of the object appear. Different objects show different regional characteristics. Through texture analysis, images can be better classified. LBP operator[12] constructs the relationship between the center pixel and its surrounding pixels, which is an operator to describe local texture features. The core idea is to assign labels according to the relationship between the pixel center threshold and its neighborhood gray value. LBP feature vector is represented by image block LBP histogram. Divide the image into \( N \times N \) regions. Each sub block area \( T=(f_c, f_0, ..., f_{p-1}) \) is composed of the central pixel point \( f_c \) and its neighboring pixel values \( f_0 \sim f_{p-1} \). In a defined \( 3 \times 3 \) window, the center pixel \( f_c \) of the window is taken as the threshold. When the gray value of eight adjacent pixels is greater than the center pixel value, the position is marked as 1, otherwise it is 0. The calculation formula is:

\[
LBP(g_c) = \sum_{p=0}^{p-1} s(f_p - f_c) * 2^p
\]  

(4)
In which, \( p \) represents the neighborhood pixels; \( g_c \) is the coordinate of the central pixel \( c \). Normalize and connect all image sub block histograms to obtain global texture features. \( s(i) \) is expressed as follows:

\[
s(i) = \begin{cases} 
0, & f_p < f_c \\
1, & f_p \geq f_c 
\end{cases}
\]  

(5)

HSV can perceive color more accurately than traditional RGB color space, and keep simple calculation. The super pixel basically retains the boundary information between the main regions and the basic topological structure. The K-means clustering algorithm is improved by combining the image position information after preprocessing. The dimensional feature is \( x_i = [Z_H, Z_S, Z_V, Z_x, Z_y, Z_{LBP}]^T \), where \( Z_H, Z_S, Z_V \) represent color information; \( Z_x, Z_y \) represent the spatial position of the image; \( Z_{LBP} \) represents the pixel texture eigenvalue.

### 4 Gaussian mixture model

SLIC algorithm represents the image as a set of super pixel images \( P = \{P_1, P_2, \ldots, P_n\} \). The Gaussian mixture model is used to cluster the foreground and background of the pre segmented image. The model is:

\[
p(x) = \sum_{k=1}^{K} \pi_k N(x | \mu_k, \Sigma_k)
\]  

(6)

means algorithm clusters foreground and background features to obtain the probability \( p(x) \) and mean value of super pixel value in samples \( \mu(\alpha, k) \), covariance matrix \( \Sigma(\alpha, k) \) and weight \( \pi(\alpha, k) \). The parameters of the model are obtained:

\[
\theta = \{\pi(\alpha, k), \mu(\alpha, k), \sum (\alpha, k), \alpha = 0,1, k = 1\ldots K \}
\]  

(7)

According to Gibbs energy term, S-T network diagram and the corresponding energy function are constructed. The formula is:

\[
E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z)
\]  

(8)

where: \( \alpha \) is the pixel marker, \( K \) is the Gaussian component, \( \theta \) is the foreground/background probability of the pixel, and \( Z \) is the image pixel. Due to the introduction of LBP texture features, a new region term is constructed according to the literature[13]:

\[
U'(\alpha, k, \theta, z, \gamma) = \rho U_{hsv}(\alpha, k, \theta, z) + (1-\rho)U_{lbp}(\gamma, z)
\]  

(9)

\( U_{hsv} \) and \( U_{lbp} \) are the penalty values of color and texture at GMM respectively, and \( \rho \) is the weight. The smoothing term corresponds to the weight of the graph model n-link, which is the weighted accumulation of the color distances of all adjacent pixels. The final energy function is expressed as:

\[
E(\alpha, k, \theta, z, \gamma) = U'(\alpha, k, \theta, z, \gamma) + V(\alpha, z)
\]  

(10)

Maxflow/mincut algorithm is used to minimize the energy function.
By introducing texture features and HSV color space into GrabCut algorithm, over segmentation results in true color images can be effectively improved. It improves the efficiency of the algorithm and reduces the number of iterations.

5 Experiment and analysis

5.1 Simulation experiment

The experimental environment is 64 bit Windows, CPU: Intel Core i5-6500, RAM: 8G. In order to verify the performance of the proposed algorithm, experiments were conducted on the collected clothing pattern samples (Fig.6.). The segmentation comparison test shall be conducted in Berkeley, DUTS and MSRA-B public data sets containing pixel level truth annotation. The segmentation performance of the algorithm is analyzed qualitatively and quantitatively.

Fig. 6. Pattern data set.

Fig.7.shows the segmentation effect of patterns under different algorithms. GrabCut(GC)[14] has a good segmentation effect when the texture highlights and the contrast is obvious. Saliency Cuts(SC)[15] can better process the background in different situations by using Saliency map to obtain good edges, but cannot retain the complete image when the background is similar. FH[16] algorithm has a serious phenomenon of over segmentation and under segmentation in low contrast images. This algorithm can not completely remove the background when the main body color is light or blended into the background, but the segmentation effect is basically the same as GrabCut and FH algorithm when the boundary is clear and the pattern is single. It is superior to other algorithms in the integrity of pattern segmentation, edge detail preservation and other aspects.
Select 6 images of different scenes in the public database for display (Fig. 8). The images listed cover a variety of situations, such as obvious objects touching the image boundary, simple background and the target is located in the image center, background and target are similar, multi target concentration, low light images, redundant objects, etc. The segmentation effect of this algorithm in different backgrounds is more accurate, closer to the true value image, and can effectively suppress the image background. GrabCut algorithm can not completely segment the target image when the image touches the boundary. SaliencyCuts algorithm has a poor segmentation result when the background is similar to the background before and after. In the end, the background is wrongly segmented when there is other interference. FH algorithm can not segment salient targets well in any of the scenes shown.
Fig. 8. Comparison of algorithm segmentation effects.

5.2 Analysis of results

The experimental results are analyzed separately, and the segmentation results are evaluated using evaluation indicators, namely Precision, Recall, F-measure, and aveMAE.

Table 1. Segmentation index of clothing pattern algorithm

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>aveMAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC</td>
<td>0.701</td>
<td>0.825</td>
<td>0.725</td>
<td>0.138</td>
</tr>
<tr>
<td>SC</td>
<td>0.903</td>
<td>0.587</td>
<td>0.800</td>
<td>0.123</td>
</tr>
<tr>
<td>PH</td>
<td>0.482</td>
<td>0.482</td>
<td>0.720</td>
<td>0.187</td>
</tr>
<tr>
<td>Ours</td>
<td>0.872</td>
<td>0.883</td>
<td>0.865</td>
<td>0.065</td>
</tr>
</tbody>
</table>

It can be seen from Table 1 that the average F-measure results of other algorithms are 0.725, 0.800 and 0.720, respectively. The average F-measure results of the proposed method are 0.865, which is less accurate than SaliencyCuts, mainly because the SC can better highlight the target area by introducing the saliency map. Compared with GrabCut, SaliencyCuts and FH, this algorithm has good positioning performance for patterns with the minimum average absolute error, and the model performance is better than other methods.

Table 2. Algorithm segmentation indicators

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC</td>
<td>0.847</td>
<td>0.657</td>
<td>0.778</td>
<td>0.875</td>
<td>0.627</td>
<td>0.788</td>
<td>0.845</td>
<td>0.237</td>
<td>0.526</td>
</tr>
<tr>
<td>SC</td>
<td>0.751</td>
<td>0.529</td>
<td>0.335</td>
<td>0.368</td>
<td>0.602</td>
<td>0.306</td>
<td>0.649</td>
<td>0.348</td>
<td>0.100</td>
</tr>
<tr>
<td>FH</td>
<td>0.417</td>
<td>0.658</td>
<td>0.455</td>
<td>0.351</td>
<td>0.733</td>
<td>0.398</td>
<td>0.603</td>
<td>0.748</td>
<td>0.629</td>
</tr>
<tr>
<td>Ours</td>
<td>0.932</td>
<td>0.878</td>
<td>0.914</td>
<td>0.946</td>
<td>0.834</td>
<td>0.909</td>
<td>0.933</td>
<td>0.931</td>
<td>0.930</td>
</tr>
</tbody>
</table>

Compared with GrabCut algorithm, the segmentation accuracy of this algorithm is effectively improved under complex or single background.

6 Conclusion

The research improves the image quality by image enhancement and pre segmented region adjacency graph, reduces the scale of image post-processing, and improves the problem that the graphics cutting algorithm is based on pixel computing and has low interaction efficiency. According to the color, texture and position information of the pattern, it can be effectively segmented. The experimental results show that the algorithm in this chapter has a great improvement in the segmentation effect, and can accurately obtain significant objects in the image compared with other algorithms.
Fund project: Research Innovation Project for Postgraduates of Beijing Institute of Fashion Technology (X2022-114), Beijing Philosophy and Social Science Planning Research Project (No. 13JDWYA005)

References


