Implementation of automatic map based on saliency detection

Liu Yuhong, Wang Xuan, Wang Xinyu and Cao Chunhong
School of Computer Science and Technology, Northeast University, Shenyang 110169, China
*Corresponding author’s e-mail: 1065678145@qq.com

Abstract. As one of the key technologies in film special effects production, virtual reality, augmented reality and digital image editing, image matting has been widely applied. In computer vision, art design and other aspects have played a very important role. In view of the limitations of traditional matting technology on background and the need for a large number of human interactions, this paper adopts an automatic matting algorithm based on saliency detection to solve this problem. Experimental results show that the improved RC algorithm can automatically mat out the images, and has a good effect and high efficiency in detail processing.

1. Introduction
The rapid development of the Internet makes images become the main way of information acquisition, transmission and processing with its intuitiveness. Compared with other media such as text and audio, people can quickly obtain information from images, which greatly improves the speed of information transmission. Therefore, image processing technology has gradually become one of the important research directions in the field of computer science.

Image matting technology was born in the film and television industry. In the process of film production, blue screen matting is used[5]. Actors shoot in a special studio and place it in another background through matting technology in the later stage. Although blue screen matting brings great convenience to people, it is difficult to be widely used in reality due to its high requirements for the background and the photographed object. Subsequently, the gradually developed natural matting technology overcomes the limitation of the processed image on the background, and provides prior knowledge by doodling or providing a trimap, so as to roughly distinguish the foreground and background of the image for further processing. However, this method also has some defects. The dependence on human interaction increases the workload and time of processing large-scale images, making it difficult to achieve high efficiency.

In the image editing software used in people's daily life, matting is a common basic function. For example, when using Photoshop to carry out image matting, users need to first carry out edge depiction on the desired matting target, or manually set some coefficients on the basis of experience. Obviously, these pre-treatment's have high experience and manual skills for matting. Most other algorithms of image matting also need artificial mark image or provide corresponding trimap to the next step which needs a lot of artificial processing and a lot of time. This kind of resource consumption is not optimistic, and the precision of image matting also influenced by the experience of artificial and so on, such as the low stability. If you can save the manual interaction process in the matting process and realize automatic image matting, it will save a lot of manpower and time, greatly improve the time efficiency of image matting and the stability of matting results. Therefore, how to
better, more conveniently and quickly mat and solve the dependence of automatic matting technology on human interaction is a problem with practical value and commercial benefits, which is worthy of our in-depth research and discussion.

2. Realisation Process
In order to realize automatic matting, this paper combines the matting algorithm with a reformed saliency detection algorithm. The specific process is illustrated in Fig1 as follows:

![Figure 1. specific process](image)

First, a saliency map is obtained from the original image via the modified RC algorithm, and then saliency map binarization is applied to it in preparation for further process. By the utilisation of morphological operation, the front and background of the trimaps are acquired, the difference between which is the desired unknown area. Hence we obtained a preliminary trimap. Next, the preliminary trimap is rectified through regional growth algorithm, resulting in a more precise trimap. Finally, shared matting algorithm is executed to matting from the refined trimap, thereby achieving automatic matting.

The followings are explanations of the algorithm involved in the process.

2.1 Region Contrast detection method based on global contrast (RC Algorithm)
According to human visual attention theory, the human visual system only fully processes part of the image, ignoring the other parts, when an image is presented. Likewise, in reality (complicated natural environment), not all information is perceived immediately. Instead, people would merely notice the area that stands out (significant area) and the information that is most striking[7]. The saliency detection of image, wherein the most remarkable elements in the image are extracted to form the saliency map via the saliency monitoring algorithm, is based upon this particular character of the visual system. This essay will explore the region Contrast detection method based on global contrast (RC algorithm) [2] by analysing and comparing available saliency detection method.

In RC algorithm, the image is divided according to graph-based superpixel segmentation algorithm[8]. Although this algorithm is able to produce superpixel swiftly, the amount and shapes of them are out of control. In order to address this issue, SLIC algorithm is adopted. It is discussed in detail as follows.
SLIC Algorithm is an algorithm with straightforward ideas that is easy to implement. It transforms the image into a 5-dimensional eigenvector, whose elements are Lab, colour, space, X- and Y-coordinates. Then measurements of distance are established, and image pixel are locally clustered[3]. The shape and size of the superpixel produced by SLIC are relatively “regular”, and the layout is intensive. Moreover, the outline of the object is well-preserved, and the algorithm is quite efficient as well. Thus, given those advantages as described above, it is used to further optimise saliency map. The basic steps of SLIC are as follows:

1) Initialise cluster centre. Construct cluster centre evenly throughout the image in compliance with the amount of superpixel. Assume there are N pixels with K superpixels of the same size. Then the size of every superpixel is N/K, and the distance between two adjacent cluster centres is about \( S = \sqrt{\frac{N}{k}} \).

2) Rearrange cluster centring its 3x3 (Normally) choose a new position in the neighbouring area. Compute gradient value of every pixel in that area, and move cluster centre to the place where gradient value is minimised.

3) Mark which cluster centre the pixel in the neighbourhood around each cluster canter belongs to.

4) Calculate the distance in colour and space between each pixel and cluster canter that it belongs. The computing method is displayed in (2.1):

\[
\begin{align*}
  d_c &= \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \\
  d_s &= \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \\
  D' &= \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2}
\end{align*}
\]

In (2.1), \( d_c \) is the distance in colour, \( d_s \) is the distance in space, and \( N_s \) is the maximum distance in genus. And among them \( N_s = S = \sqrt{\frac{N}{k}} \). \( N_c \) is the maximum distance in colour, which alters in accordance with the image and cluster change. Thus set the \( N_c \) value to a fixed constant \( m \) (in the range of \([1,40]\), is assigned to it. Under most circumstances, the value of \( m \) is taken as 10). Then the distance metric \( D' \) in equation (2.1) is transformed into the following equation (2.2):

\[
D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2}
\]

As in the process of searching pixel, a pixel will be linked to multiple cluster centres. The cluster centres with the least distance from the pixel is chosen to be the cluster canter.

5) Theoretically, after iteration of the steps above, the cluster centers of every pixel will not alter. As shown by experiments, the result is satisfactory when the number of iteration is 10, and that is the reason why the number of iterations takes value 10.

6) There may be issues related to iterations, including too small superpixel, multiple discrete superpixels when there should be a single one. Nevertheless, those problems can be resolved by enhancing connectivity. We can reallocate discontinuous superpixels, as well as too small superpixels, to their adjacent superpixels by using some pattern, and record the pixels that have been reallocated so that we will not allocate them again, until we finish reallocating all of the pixels.

For any given superpixels, the saliency value is defined in (2.3) as follows:

\[
S(r_k) = \sum_{r \in r_k} \omega(r_i)D_r(r_k, r_i)
\]

In (2.3), \( \omega(r_i) \) is the weight of superpixels, which is represented by the number of pixels in superpixel \( r_k \), so as to emphasis colour contrast of the larger region? \( D_r(\ldots) \) is the colour distance of any two given superpixel, and the definition of \( D_r(\ldots) \) is as follows:

\[
D_r(r_m, r_n) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f(c_m, i)f(c_n, j)D(c_m, i, c_n, j)
\]

In (2.4), \( f(c_m, i) \) is the probability that the ith colour \( c_m \) appears in the mth region \( r_n \) in all \( n_m \) colours. We use the color probability in the normalized color histogram as the weight of the color to represent the degree of color difference between the colors.
2.2 Saliency image binaries

Firstly, we binaries the saliency image from the improved RC algorithm. From further processing the saliency value, we briefly identified the foreground and background of the image.

This essay adopted the Otsu algorithm to binaries the image. Otsu is a dynamic threshold-segmentation calculating algorithm, which is simple and more efficient in calculation[6]. Besides, it is not affected by the brightness and contrast of the image. Next, we will introduce the principles of Otsu’s algorithm.

Define the variance in Otsu as shown in (2.5):

$$\sigma^2_{\omega}(t) = \omega_\omega(t)\sigma^2_\omega(t) + \omega_\nu(t)\sigma^2_\nu(t) \quad (2.5)$$

Through selecting the threshold $t$ when the variance within above is minimized, we could distinguish the foreground and background as much as possible. It means that the possibility of a pixel belongs to the foreground, is the possibility of a pixel belongs to the background, which is the variance between these two situations. If an image’s histogram has a number of $L$ levels (normally $L=256$), then under a given threshold $t$, and identifying respectively as shown in (2.6):

$$\omega_\omega(t) = \sum_{i=0}^{L-1} p(i)$$

$$\omega_\nu(t) = \sum_{i=L}^{\infty} p(i) \quad (2.6)$$

Since the variance in minimization equals to the variance in maximization, we get (2.7):

$$\sigma^2_{\omega}(t) = \sigma^2 - \sigma^2_{\omega}(t) = \omega_\omega(\mu_\omega - \mu_T)^2 + \omega_\nu(\mu_\nu - \mu_T)^2 \quad (2.7)$$

Whereas the sum of the average grey level of the foreground $\mu_\omega$ and background $\mu_\nu$ pixels, is the total average grey level of the image.

Based on $\omega_\omega \mu_\omega + \omega_\nu \mu_\nu = \mu_T$, $\omega_\omega + \omega_\nu = 1$, we get (2.8):

$$\omega_\omega(\mu_\omega - \mu_T)^2 + \omega_\nu(\mu_\nu - \mu_T)^2 = \omega_\omega(t)\omega_\nu(t)[\mu_\omega(t) - \mu_\nu(t)]^2 \quad (2.8)$$

By calculating the variance, the higher the variance is, the bigger the difference is. Which means when the image is binaries, the more distinct the segmentation is, the better the result is.

2.3 Morphology Calculation and Region Growth algorithm

Morphology is a branch of biology, involves animals’ and plants’ shapes and results. Moreover, mathematical morphology has been successfully applied to image processing and pattern fields. The basic concept behind using the mathematical morphology in image processing is: measuring the target image with structural elements of certain size and shape, to obtain relative information of the image’s form and structure. In result to achieve the goal of image analyzing and identification[4], we will expand and etch the binary image to obtain a preliminary trimap.

Through morphology calculation, the trimap is obtained, while the results might not be accurate. Its operation increases the chance that the background area in the image may be incorrectly included in the clear foreground area of the trimap. Therefore, in the next step, the obtained preliminary trimap is subjected to a region growing algorithm to be corrected, in order to get a more accurate final trimap.

The specific steps for applying regional growth algorithm on the preliminary trimap are as follows:

(1) Initialize the cluster pixel of the inner boundary of the unknown region in the trimap;
(2) Initialize the threshold $TH=0.13$;
(3) For all pixels $p \in U_5$;
(4) $Mask(p) \leftarrow \text{GenerateMask(image, }ind\text{, }TH\text{)}$
(5) $\text{FinalMask} \leftarrow \text{union of all } Mask(p)$;
(6) intersect the clear foreground area of the trimap with $\text{FinalMask}$;
(7) Obtain the final trimap.

Among which, the specific calculations in Generate Mask are shown as below:

(1) Initialize $Mask(p)$ to 0
(2) For all pixels $q \in image$;
(3) If $p - TH \leq Rq(image) \leq p + TH$ and $p - TH \leq Gq(image) \leq p + TH$ and $p - TH \leq Bq(image) \leq p + TH$
(4) $Maskq(p) = 1$;
(5) return Mask(p).

For the region growth algorithm, a cluster pixel and a threshold must be provided and grow with it. First, the inner boundary of the unknown region is selected as a set of cluster pixels, and the threshold depends on the contrast of the foreground and background regions. Experiments found that the threshold is 0.13 for all types of images.

For each cluster pixel p in the input image, whether the R, G, and B components of p are in this interval, if this applies to some pixels in the input image, these pixels are set to 1 in the Mask.

All Masks of the each cluster pixel jointly give the total growth area of all cluster pixels, i.e. the entire foreground area. The foreground area is then replaced by taking the intersection of the foreground areas. Since the foreground Mask grows only in the foreground rather than the background area, the intersection will correct the trimap by removing the unknown area that is misclassified as a clear foreground region.

2.4 shared matting algorithm

ESL Gastal et al. proposed a Shared Sampling for Real-Time Alpha Matting, referred to as shared matting algorithm[1]. The basic idea of the algorithm is that adjacent pixels usually have the same or similar foreground colour, background colour, and alpha value. So for unknown pixels, its foreground colour and background colour can be estimated by neighbouring known pixels, where a group of more related and appropriate pixels are selected from. The pixels selected are referred to as a pair of sample points of the foreground and the background. The foreground colour and the background colour of the unknown pixel point can be calculated, and the alpha value is obtained afterwards. In this way, the amount of calculation is reduced.

This algorithm has four steps in total: 1. Expansion. Slightly expand the known foreground and background of the input trimap. 2. Sample and Gather. Obtain samples of the known neighbouring pixels for the pixel in the unknown area, and select the best pair of foreground and background sample points. 3. Refinement. Combine the best foreground and background sample points for each pixel in the unknown area. 4. Local Smoothing. Locally smooth the calculated foreground, background, and the values to reduce noise.

3. Results

3.1 saliency detection algorithm Experimental Results

![Figure 2. comparison of the results of saliency detection](image)

As shown in Figure 2, (a) is the original image, (b) is ground truth, (c) is the original RC algorithm result, (d) is the improved RC algorithm result. From (c) and (d), it is observed that the pixel block in (d) is more even. The significant object’s edges are more accurate and smoother. For example, in the first row of pictures, the details of the faces, arms, etc. of the characters, and in the second row, the shapes of the ducks are more accurate and easier to detect from the improved RC algorithm result in
column (d), compared with the original RC algorithm results in column (c). Also, it can be observed that the results in (d) are more accurate than the ground truth in (b). Meanwhile, it can be seen from the first row of the pictures that there is a large amount of interference noise (flowers in the background) in the result of the original RC algorithm. And the improved RC algorithm results significantly weaken the noises. Since the purpose is to solve the trimap, and highlights the main area, the improved RC algorithm yields a more accurate detection result, which lays a good foundation for generating high-quality trimap results, which is beneficial for subsequent follow-up in the process afterwards.

3.2 Trimap Automatically Generates Experimental Results

In Figure 3, (a) in the original image, (b) is the preliminary trimap, (c) is the final trimap obtained by the region growing algorithm. It can be observed that the main area of the significant target in the image is marked as the foreground area, and the details such as the edge and the indistinguishable parts (the hair of the elephant, the hair of the character) are better marked than column (b), and a final trimap is obtained. Since the accuracy of the trimap will influence the subsequent result of matting, therefore, the region growth algorithm will be more helpful to the subsequent matting result.

3.3 Automatic Matting Experimental Result

(a) the original image (b) the transparency image (c) the foreground image

Figure 4. matting results
In figure 4, the first one is the original image, the second one is the transparency image obtained after the image is softly matting, the third one is the matting foreground image. It can be seen that after the matting, the foreground image is completely extracted. Especially for the foreground region image with complex edge details and translucent properties, the improved shared matting algorithm is well processed. Comparing with the original algorithm, the details such as edge, translucent, and hair is better processed, which cannot be achieved by image segmentation. Besides, it will take a huge amount of time to get the same result by artificially interacting matting.

4. Conclusion

We have based on the image saliency detection to design an algorithm that automatically generates a trimap and combines the matting algorithm to achieve an automatic matting algorithm without human interaction.

For the image saliency detecting algorithm, we have chosen the algorithm based on the overall contrast, which is referred to as RC algorithm. We have introduced the basic concept of RC algorithm and improved the algorithm. We have designed an algorithm that automatically generates a trimap. For the saliency graph obtained by the improved RC algorithm, the preliminary graph is obtained by the algorithm of threshold segmentation to get the preliminary trimap, an then the region growth algorithm is used to obtain the final trimap. We have chosen the shared matting algorithm for matting. Since there might be errors from the previous steps, that is, the unknown region is not accurate enough. In order to resolve this problem, the trimap obtained in the previous step is taken as an input, the algorithm is changed to the Lab color space to measure the similarity of pixel points in the image to further reduce the error, and an automatic matting algorithm is implemented.

Through the experimental research and analysis of the results, we can see that the proposed algorithm improves the original one, improves the efficiency of the algorithm, realizes the requirements of automatic matting, and also eliminates the manual interaction process. It saves time and effort to get the most accurate matting results.

Acknowledgments

Supported by “the Fundamental Research Funds for the Central Universities”(N161602001)
Supported by Natural Science Foundation of Liaoning Province of China(20170540312)
Supported by Natural Science Foundation of Liaoning Province of China(2018520001)
Project 201910145170 Supported by National Training Program of Innovation and Entrepreneurship for Undergraduates(201910145170).

References

[1] Gastal E S L, Oliveira M M. Shared sampling for realtime alpha matting [C]. Computer Graphics Forum, 2010, 29(2): 575-584.
[2] Cheng M M, Zhang G X, Mitra N J, et al. Global contrast based salient region detection[C]// Computer Vision and Pattern Recognition. IEEE, 2011:409-416.
[3] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(11):2274-2282.
[4] Esmael Hamuda, Brian McGinley, Martin Glavin, Edward Jones. Automatic crop detection under field conditions using the HSV color space and morphological operations[J].Computers and Electronics in Agriculture, 2017, 133: 97-107.
[5] Smith A R, Blinn J F. Blue screen matting[C]. In Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. ACM, 1996: 259-268.
[6] Otsu N. A Threshold Selection Method from Gray-Level Histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1):62-66.
[7] Dongdong Gao, Xinsheng Zhang. Saliency Image Detection Based On Spatial Convolutional Neural Network Model[J]. Computer Engineering, 2018, 44(5):240-245.
[8] Yongjie Zhang. Gesture Recognition Based On Kinect Depth Superpixel[D]. 2017