RESEARCH ARTICLE

Salient object segmentation based on active contouring
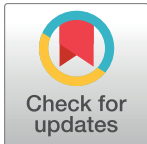
Xin Xia, Tao Lin*, Zhi Chen, Hongyan Xu

College of Computer Science, Sichuan University, Chengdu, Sichuan, China

* sculintao@gmail.com

Abstract

Traditional saliency detection algorithms lack object semantic character, and the segmentation algorithms cannot highlight the saliency of the segmentation regions. In order to compensate for the defects of these two algorithms, the salient object segmentation model, which is a novel combination of two algorithms, is established in this paper. With the help of a priori knowledge of image boundary background traits, the K-means++ algorithm is used to cluster the pixels for each region; in line with the sensitivity of the human eye to color and with its attention mechanism, the joint probability distribution of the regional contrast ratio and spatial saliency is established. The selection of the salient area is based on the probabilities, for which the region boundary is taken as the initial curve, and the level-set algorithm is used to perform the salient object segmentation of the image. The curve convergence condition is established according to the confidence level for the segmented region, thus avoiding over-convergence of the segmentation curve. With this method, the salient region boundary is adjacent to the object contour, so the curve evolution time is shorter, and compared with the traditional Li algorithm, the proposed algorithm has higher segmentation evaluation scores, with the additional benefit of emphasizing the importance of the object.

Introduction

Salient object segmentation refers to the segmentation of important and semantic objects from an image. The correctness of the segmentation determines the effectiveness of subsequent analysis and processing. It is a key step to the analysis and understanding of images. The technology can be used in the fields of object detection, image retrieval, and editing. The traditional image segmentation algorithm is based on image features (e.g., brightness, color, and edge), and the image will be divided into several *semantic concept* areas [1–3], but the contribution of each region to an understanding of the image is not indicated. The saliency detection uses regional contrast and spatial position to analyze the saliency of each region [4, 5], but each region only has one consistent color brightness and lacks semantic meaning.

The human visual system can identify salient objects in an image quickly and accurately [6]. From the perspective of low-level features of the image, an object is formed mainly by its own characteristics, such as edges and boundaries, and the saliency of the object depends primarily on relative differences in brightness and color along with its relative spatial location. Traditional image saliency detection algorithms generate a saliency chart of pixel importance

probability [7, 8]. Wang, Peng [9] proposed the binary salient segmentation estimation model [10] for images, using stochastic theory. In consonance with the color sensitivity of the human eye, a regional saliency analysis model [11] has been built that is based on the statistical characteristics of image color components. Using the different responses of the eye to signals of different frequencies, a saliency detection algorithm based on signal frequency has been designed [12]. At present, most saliency detection algorithms are based on the ability of the human eye to distinguish the boundary of blurred objects by its contrast sensitivity; through contrast analysis to estimate the image pixels and the contribution of regions to the visual composition, an image saliency analysis model can be designed. Saliency analysis is divided according to the region size into local and overall saliency models [13, 14]. Fu combined the relationship between regional contrast and relative position organically, proposing a clustering-based saliency detection algorithm [15]. Traditional saliency detection algorithms only highlight the importance of regions for image analysis; the regions lack object semantics.

In order to extract objects with semantic concepts from the image, an image segmentation algorithm is often used. A segmentation algorithm based on active contouring can effectively extract the complete objects from images [1]; a priori curve of this algorithm is combined with image areas or edge information to perform a priori object segmentation. For the sake of suppressing the influence of texture and noise on the segmentation, Mumford and Shah smoothed the images and evolved the curve segmentation model [16], but the Mumford–Shah (MS) functional is not convex, making it difficult to solve. Chan and Vess established the Chan–Vess (CV) segmentation model [17] by area means; this method is better for the segmentation of cartoon images and in addition is not sensitive to noise. In order to suppress texture, Tsai, Yezzi [18] proposed a piecewise approximation of the Piece-Smooth (PS) segmentation model, which can suppress texture to a certain extent, but the computation requirements are large. Li et al. proposed the active contour segmentation algorithm [19], which is based on edges; however, the algorithm uses the local information of the image edge, and the result is more sensitive to noise. To suppress noise, the segmentation image is often treated by Gaussian smoothing. However, Gaussian smoothing fuzzes object contours, and for weak edges the segmentation effect is not ideal. Traditional segmentation algorithms can effectively segment an image into meaningful regions, but they cannot evaluate the importance of the segmented regions to visual comprehension.

The attention mechanism of the human eye can quickly and accurately extract salient semantic objects. As the traditional saliency detection algorithm lacks object semantics and the segmentation algorithm cannot highlight segmentation region saliency, in order to address the individual deficiencies of the two algorithms, in this paper, two algorithms are organically combined to create a salient object segmentation model. Firstly, the model is based on the statistical finding that 85% of image boundaries belong to the background; this is used to establish features of the background image as prior knowledge, and through the K-means++ algorithm, the clustering pixels are divided into various regions. Secondly, regional contrast and spatial saliency are analyzed and estimated according to the human eye's sensitivity to color and its attention mechanism, the regional saliency joint probability distribution is established, and the region of saliency is selected by probability analysis. Finally, taking the saliency regional boundary as the initial curve, the level-set algorithm is used to perform the image segmentation, and the convergence condition of the curve is established according to the confidence level for the segmented region, to avoid over-convergence of the segmentation curve. In this model, each region has a clear semantic meaning, and the model highlights the importance of regional borders. After providing the details of the algorithms, this paper describes an experiment comparing the proposed method with the traditional Li algorithm in terms of the suitability of the salient region boundary, the system times required, and F-measure scores.

Detection of salient regions

Pixels in the same region of an image should have the same attributes in the feature space. This model uses the $L^*a^*b^*$ color model, which can describe any color in nature, as the feature space. In this space, the three components L^*, a^*, and b^* are perpendicular to each other, where a^*, and b^* indicate any color and hue characteristics, respectively.

$$\begin{cases} L^* = 116f(Y/Y) - 16 \\ a^* = 500[f(X/X_n) - f(Y/Y_n)] \\ b^* = 200[f(X/X_n) - f(Z/Z_n)] \end{cases} \tag{1}$$

Therein, X,Y,Z indicate the three components of the CIEXYZ color model and have the values 95.047, 100, and 108.083, respectively, and $f(t)$ is defined as

$$f(t) = \begin{cases} t^{\frac{1}{3}} & t > (\frac{6}{29})^3 \\ \frac{1}{3}\left(\frac{6}{29}\right)^3 + \frac{4}{29} & \text{otherwise} \end{cases} \tag{2}$$

In the $L^*a^*b^*$ feature space, the image pixels are clustered into a region $R^k(k = 1,2,\ldots,m)$, which enables each region to meet in the feature space: the similarity of pixels in the same region is higher, and the similarity of pixels in different regions is lower.

The K-means++ clustering algorithm is used, but the initial cluster center needs to be determined in advance; different initial cluster centers may lead to completely different clustering results. To determine the initial cluster center, images from the Internet were analyzed, and 85% of image edges were found to belong to the background region. In accordance with this property, the region which has 20 pixels to the image boundary is taken as the initial background region, and the regional mean value in the feature space is taken as the initial seed point. Then, the K-means++ algorithm is used to cluster the pixels in the image.

The sizes of different regions in the image are different, and the eye is generally more attentive to large regions, ignoring the smaller ones. In accord with this property, the model considers the contribution of the size of each region to the region's saliency, and sets a size weighting for R_k:

$$p_k = n_k/N \tag{3}$$

In the formula, n_k indicates the number of pixels in the region R_k, and N represents the number of pixels in the $W \times H$ image.

2.1 Analysis of region saliency

Image contrast is the relative change in the value of adjacent pixels: the greater the change, the higher the contrast. Contrast is a key factor affecting the image's visual effect: generally, an image with higher contrast is clear and bright; otherwise, the image is blurred. As the eye is more sensitive to areas of high contrast, a region's pixel saliency can be estimated through an analysis of the region's contrast. The region is represented by μ, which is the mean value of the region's pixels in the feature space, and the contrast of region R_k is defined by incorporating the eye's attention response to the area:

$$\omega_c(k) = \sum_{i=1,i\neq k}^{K} (p_k\|\mu_k - \mu_i\|_2) \tag{4}$$

In the proposed model, the contrast $\omega_c(k)$ of region R_k is weighted with respect to all regions of the image; it is an extension of the local contrast, which belongs to the overall contrast. The magnitude of the value shows the extent to which the region attracts more of the eye's attention than other regions; higher values indicate that the region is relatively more important for image analysis and understanding from a pixel perspective.

For the human visual system, the area near the center of an image attracts more attention, and the image boundary region is often neglected. As the distance between the object and the image center increases, the attention paid to the object decreases. This is called the *center deviation rule* of the image saliency detection algorithm. In our model, according to the above principle for the human eye, the spatial saliency for region R_k is represented by $\omega_s(k)$, which is defined as:

$$\omega_s(k) = \frac{1}{n_k} \sum_{z_i \in R_k, i=1}^{n_k} \exp\left(\frac{(z_i - o)^2}{2\sigma^2}\right) \tag{5}$$

Therein, o represents the center of the image, and the variance σ^2 is the normalized radius of the image. z_i indicates the ith pixel in the region R_k. According to the 3σ principle, σ is determined by the size of the image:

$$\sigma = \frac{1}{6} \max(W, H) \tag{6}$$

2.2 Mapping of salient regions

The image region contrast $\omega_c(k)$ and spatial saliency $\omega_s(k)$ respectively reflect the contribution of the region's pixels and that of the relative spatial position to the human eye. The eye often combines the above two features to determine the importance of a region. Common ways of combining contrast and spatial saliency are linear summation or point by point multiplication. Point by point multiplication is better than linear summation for suppressing the noise, but linear summation can obtain a higher recall rate. For the saliency test, the accuracy rate is more important than the recall rate. In our model, the probability $p(k)$ of joint saliency for region R_k is obtained by multiplying the integration contrast and the spatial saliency:

$$p(k) = \omega_c(k)\omega_s(k) \tag{7}$$

The regional contrast $\omega_c(k)$ of our model is based on the calculation of the mean feature space of the region, without considering the contrast for each pixel in the region. Whereas the regional spatial saliency $\omega_s(k)$ represents the weighted saliency of each pixel in the region relative to the image center, the joint salient probability $p(k)$ only indicates each region's saliency, without indicating the regional saliency of each pixel. In our model, it is assumed that the pixels in the region are subject to the Gaussian distribution, and the saliency of pixel z in region R_k is

$$p(z|k) = \exp\left(\frac{(z - \mu_k)^2}{2\sigma_k^2}\right), \quad z \in R_k \tag{8}$$

Therein, the variance σ_k^2 is the pixel variance of the region R_k. The pixel saliency probability $p(z)$ is obtained as the sum of all the regions jointly with saliency $p(z|k)p(k)$:

$$p(z) = \sum_{k=1}^{m} p(z|k)p(k) \tag{9}$$

The image saliency detection results are shown in Fig 1. Fig 1B shows that the saliency using regional pixels and relative position estimation indicates only the importance of regional

a) Original image   b) Saliency region
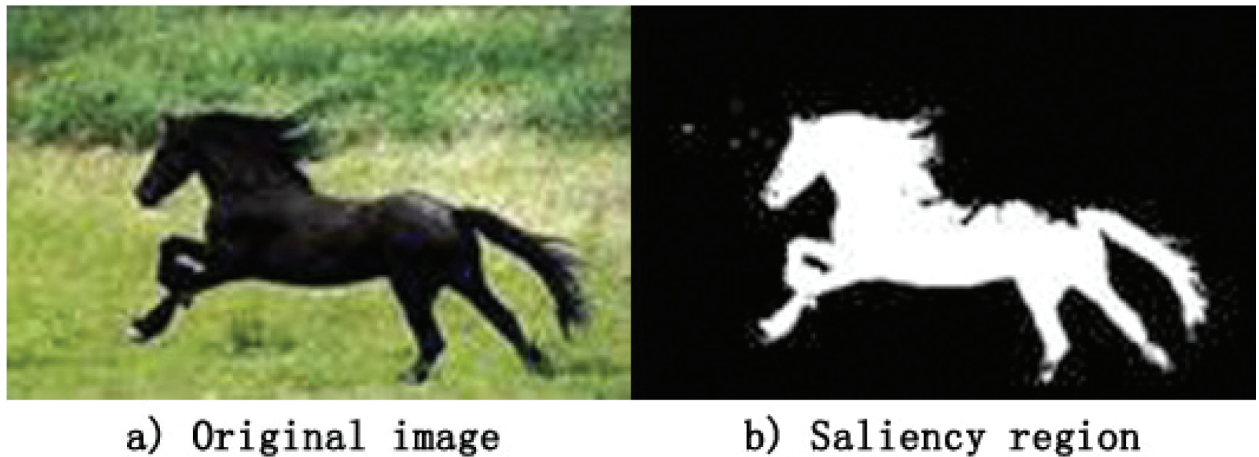
Fig 1. The salient region.

information and does not indicate whether the area has clear semantic meaning; however, the contour lines of salient regions are adjacent to the local edge lines of the object.

Segmentation of salient objects

In order to segment the semantic object from the image, this analysis assumes that there is no dividing line between objects. As for local neighborhood, the object boundary is represented as the margin of local neighborhood. Meanwhile, in the process of image segmentation, by the deformation of the topological structure of the segmentation curve based on the object contour edges, the salient objects of the image are segmented by the active contour segmentation algorithm. The object segmentation of this algorithm is implemented by curve evolution. In this model, $\varphi: \Omega \to R$ represents the Lipschitz function of three-dimensional space; for its level set representation curve C: $\varphi(x,y) < 0$ indicates an object within the region, and $\varphi(x,y) > 0$ indicates an external object. Its essence is to use function symbols to represent different regions. In order to facilitate the calculation, the Heaviside function $H(\varphi)$ is introduced to represent the inner and outer regions of the curve, and the curve C can be expressed as $\delta(\varphi)$, which is the measure of Dirac, and the Dirac is the derivative of $H(\varphi)$. $H(\varphi)$ and $\delta(\varphi)$ are respectively given as follows:

$$H(\varphi) = \begin{cases} 1 & \varphi \geq 0 \\ 0 & \varphi < 0 \end{cases} \quad , \quad \delta(\varphi) = \frac{dH(\varphi)}{d\varphi} \tag{10}$$

According to the local marginal of the object contour, the edge curves are seen as constraint conditions, and the segmented curve evolutionary energy function of image $\Omega \to [0,1]$ is

$$\varepsilon(\varphi) = \arg \min_{\varphi} \left\{ \lambda \int_\Omega g\delta(\varphi)|\nabla\varphi| dxdy + v \int_\Omega gH(-\varphi) dxdy + \frac{\mu}{2} \int_\Omega |\nabla\varphi - 1|^2 dxdy \right\} \tag{11}$$

The first term in the formula represents the curve length; the second term represents the object region area, and the third term is the curve regularization term. g is the image edge indication function:

$$g = (1 + |\nabla G_\sigma * \boldsymbol{u}|)^{-1} \tag{12}$$

If the curve is located in the smooth region, the gradient amplitude approaches zero, the edge indicator function tends to 1, and $\varepsilon(\varphi)$ is larger; if the curve is located at the edge, the gradient is larger, $\varepsilon(\varphi)$ is at a minimum, and the curve stops evolving.

In the traditional active contour algorithm, the initial curve is usually given first. If the initial curve is far from the object contour, the curve evolution time is longer. In our model, the region having a pixel saliency probability of more than 60% is seen as the initial region, and according to the region, the initial level-set function is defined as:

$$\varphi_0(z) = \begin{cases} -4 & p(z) \geq \max\{p(z)\} \\ 4 & \text{otherwise} \end{cases} \tag{13}$$

In this study, the fast descent algorithm was used for the discrete calculation (formula 12), and the negative gradient of φ was calculated by the forward difference method. The discrete calculation of $\partial\varphi/\partial t$ is

$$\frac{\varphi_{i,j}^{k+1} - \varphi_{i,j}^k}{t} = \mu \left[\Delta\varphi_{i,j}^k - div\left(\frac{\nabla\varphi_{i,j}^k}{|\nabla\varphi_{i,j}^k|} \right) \right] + \lambda\delta_a(\varphi_{i,j}^k) div\left(g_{i,j} \frac{\nabla\varphi_{i,j}^k}{|\nabla\varphi_{i,j}^k|} \right) + v g_{i,j}\delta_a(\varphi_{i,j}^k) \tag{14}$$

In order to avoid this situation, this paper defines a confidence level Pr for the segmented regions according to the different segmentation regions with different smoothing components:

$$Pr = \frac{card(A_k \cap A_{k-1})}{\max\{card(A_k), card(A_{k-1})\}} \tag{15}$$

Therein, A indicates the segmentation curve and its internal region $\{(x,y)|\varphi(x,y) \leq 0\}$. When the confidence for the segmented region meets the following condition, it indicates that the desired degree of similarity between the two iterative segmentation results is achieved, and thus the smoothing iteration will be stopped:

$$Pr \geq T \tag{16}$$

T is the segmentation region confidence threshold, and its value is close to 1. The segmentation results are different for different confidence thresholds.

Experiment, results, and analysis

The hardware environment of the experiment was an Intel Core™ i3-3220 processor with 4GB memory. The operating system was Microsoft Windows 7, and the experimental simulation environment was Microsoft Visual C++ 6.0. In this model, the image is divided into different regions via a clustering algorithm, and the number of cluster regions m needs to be given in advance, but initially the number of salient objects in different images is not known. m is selected by the method of this paper. For this study, 2000 natural images from the Internet were segmented using different m values, and Table 1 shows the F-measures obtained from different settings of m values. The F-measures are presented in the form of 'average±standard deviation', and the experimental results in Table 1 reveal that the best average F-measure is achieved under the setting of $m = 6$.

Table 1. Average F-measures obtained from different settings of *m*.

	m = 3	*m* = 4	*m* = 5	*m* = 6	*m* = 7	*m* = 8	*m* = 9
F-measure	0.9494±0.019	0.9464±0.018	0.9477±0.02	0.9650±0.022	0.9521±0.017	0.9496±0.017	0.9407±0.018
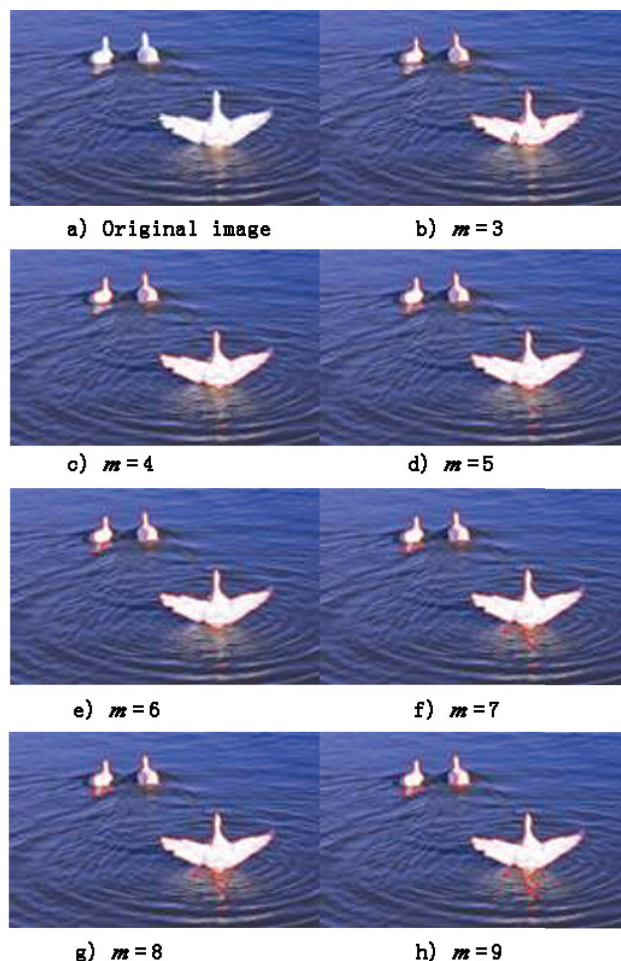
**Fig 2. Results of segmentation using different values of *m* (the number of cluster regions).**

To better illustrate the segmentation effects under different settings of *m*, Fig 2 shows the results of the segmentation of an image using different *m* values, and the segmentation effect is best when *m* = 6. The original image contains 3 visually salient objects, a semi-transparent region of 2 salient objects, and the background region containing the texture. If *m* = 3, the semi-transparent region is mis-segmented into background; if *m* = 9, the texture region is segmented as an object. (Fig 2E). If *m* = 6, the proposed algorithm will segment the semi-transparent regions of Fig 2A into salient objects.

In order to validate the effectiveness of the segmentation algorithm, the F-measures obtained by the proposed algorithm are compared with those of the method proposed in reference [19]. In order to check if the proposed algorithm is significantly better than the method given in reference [19], the Wilcoxon paired signed-rank test is conducted to compare the performance of two methods. The test results are reported in Table 2. The Wilcoxon test rejects

Table 2. Wilcoxon pairwise test result of our method vs. method in reference [19].

	Our method	Method in (19)	*p-value*
Avg F-measure	0.9650	0.9324	$p = 0.000 < 0.05$

a) Original image

b) Manual segmentation

c) Algorithm in this paper
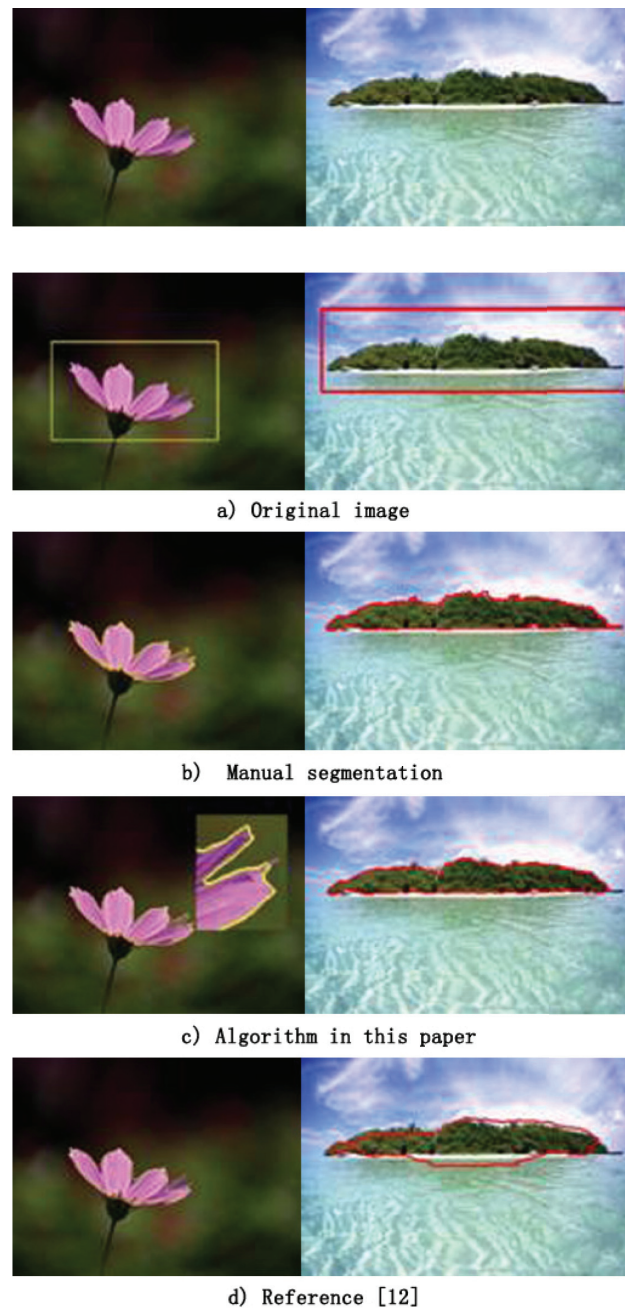
d) Reference [12]

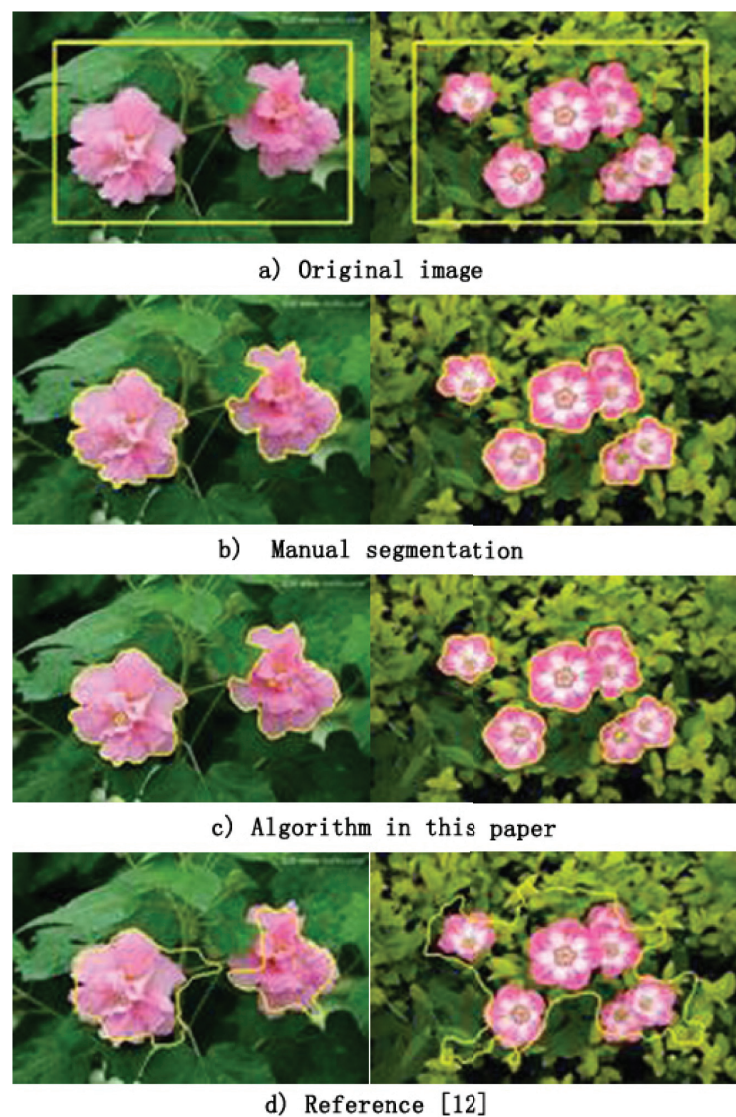**Fig 3. Results of segmentation on images with a single object.**

the hypothesis of equivalence with significance level at $p = 0.000$, indicating our proposed method is significantly better than the method given in reference [19].

To better illustrate the performance of our method, we show the segmentation results on images with different number of salient objects. We run our method and the method given in in reference [19] on an image with one salient object, and Fig 3 shows the results. In Fig 3, there is only one salient object. The brightness of the background region of the left image is nearly constant, and there is a change in the brightness of the background region of the right

image. The initial curve (adopting the algorithm of reference [19]) is in the original image (Fig 3A). Fig 3B shows the manual segmentation results. Fig 3C indicates the results using the algorithm proposed in this paper. Here, the left (800×600) image segmentation F-measure is 0.976, and the system time was 2.36 s. In addition to the cusp, the segmentation curve is accurate, and the curve of the cusp is over-convergent because the active contour segmentation algorithm requires that the curve be smooth. For the right (320×600) image, the segmentation F-measure is 0.987, and the system time was 1.89 s. Fig 3D shows the segmentation results using the method of reference [19]. Here, the F-measures of the left and right side image segmentations are 0.889 and 0.825, respectively, and the system operation times were 16.36 and 14.89 s, respectively. With the proposed method, the salient region boundary is adjacent to the contour of the target, and the curve evolution time was shorter.

In order to validate the effectiveness of the proposed algorithm for the segmentation of images with complex backgrounds, the images shown in Fig 4 were segmented, and the results

Fig 4. Results of segmentation on images with complex backgrounds.

are compared with those of the method given in reference [19]. In both images of Fig 4, there are at least 2 salient objects, the background is complex, and the initial curve (adopting the algorithm of reference [19]) is in the original image (Fig 4A). Fig 4B shows the manual segmentation results. Fig 4C shows the results using the proposed algorithm; the left (344×233) image segmentation F-measure is 0.986, and the system time was 1.36 s. The right (360×227) image segmentation F-measure is 0.992, and the system time was 1.89 s. Fig 3D shows the results of the segmentation method given in reference [19]; the left and right side image segmentation F-measures are 0.859 and 0.676, respectively, and the system times were 6.36 and 4.89 s, respectively. From the perspective of system operation time, the curve evolution time is shorter using the proposed method.

Conclusion

By an organic combination of region saliency analysis with the active contour segmentation algorithm, the salient object segmentation model is established in this paper. Using a priori background knowledge, the pixels are clustered for each region by the K-means++ algorithm; the attention mechanism of the human eye is used to analyze regional contrast and spatial saliency probabilities and to highlight the saliency of the segmentation regions, and the level-set algorithm is used to perform salient object segmentation of the image. With this method, each region segmented by our algorithm has clear semantic meaning, and its importance is emphasized; in addition, the salient region boundary is adjacent to the object contour, so the curve evolution time is shorter. However, when the image pixels are divided into different regions, the edge information of the image is not considered.

Author Contributions

Conceptualization: Xin Xia.

Data curation: Zhi Chen, Hongyan Xu.

Formal analysis: Xin Xia.

Funding acquisition: Tao Lin.

Resources: Zhi Chen.

Software: Zhi Chen, Hongyan Xu.

Writing – original draft: Xin Xia.

Writing – review & editing: Xin Xia.

References

1. Khan MW. A survey: Image segmentation techniques. International Journal of Future Computer and Communication. 2014; 3(2):89.

2. Kadir T, Brady M. Saliency, scale and image description. International Journal of Computer Vision. 2001; 45(2):83–105.

3. Sen Y, Qian Y, Avolio A, Morgan M. Image segmentation methods for intracranial aneurysm haemodynamic research. Journal of Biomechanics. 2014; 47(5):1014–9. https://doi.org/10.1016/j.jbiomech.2013.12.035 PMID: 24461575

4. Toet A. Computational versus Psychophysical Bottom-Up Image Saliency: A Comparative Evaluation Study. IEEE Transactions on Software Engineering. 2011; 33(11):2131–46.

5. Koene AR, Zhaoping L. Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. Journal of Vision. 2007; 7(7):6.1–14.

6. Judd T, Ehinger K, Durand F, Torralba A, editors. Learning to predict where humans look. IEEE International Conference on Computer Vision; 2009.

7.   Levin A, Weiss Y, editors. Learning to Combine Bottom-Up and Top-Down Segmentation. Computer Vision - ECCV 2006, European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings; 2006.

8.   Borenstein E, Sharon E, Ullman S, editors. Combining Top-Down and Bottom-Up Segmentation. Computer Vision and Pattern Recognition Workshop, 2004 Conference on; 2004.

9.   Wang X, Peng Z, Zhang P, Meng Y. Infrared small dim target detection based on local contrast combined with region saliency. High Power Laser & Particle Beams. 2015.

10. Liu T, Sun J, Zheng NN, Tang X, editors. Learning to Detect A Salient Object. Computer Vision and Pattern Recognition, 2007 CVPR '07 IEEE Conference on; 2007.

11. Cheng MM, Zhang GX, Mitra NJ, Huang X, Hu SM, editors. Global contrast based salient region detection. Computer Vision and Pattern Recognition; 2011.

12. Achanta R, Hemami S, Estrada F, Susstrunk S, editors. Frequency-tuned salient region detection. Computer Vision and Pattern Recognition, 2009 CVPR 2009 IEEE Conference on; 2009.

13. Fu H, Xu D, Zhang B, Lin S, Ward RK. Object-Based Multiple Foreground Video Co-Segmentation via Multi-State Selection Graph. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society. 2015; 24(11):3415–24. https://doi.org/10.1109/TIP.2015.2442915 PMID: 26068314

14. Han J, Zhang D, Wen S, Guo L, Liu T, Li X. Two-Stage Learning to Predict Human Eye Fixations via SDAEs. IEEE Transactions on Cybernetics. 2015; 46(2):487–98. https://doi.org/10.1109/TCYB.2015.2404432 PMID: 25730840

15. Fu H, Cao X, Tu Z. Cluster-based Co-saliency Detection. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society. 2013; 22(10):3766. https://doi.org/10.1109/TIP.2013.2260166 PMID: 23629857

16. Sedlák J, Morávek J. Optimal approximations by piecewise smooth functions and associated variational problems. Communications on Pure & Applied Mathematics. 1989; 42(5):577–685.

17. Chan TF, Vese LA. Active contours without edges. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society. 2001; 10(2):266. https://doi.org/10.1109/83.902291 PMID: 18249617

18. Tsai A, Yezzi A, Willsky AS. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. IEEE Transactions on Image Processing. 2001; 10(8):1169–86. https://doi.org/10.1109/83.935033 PMID: 18255534

19. Li C, Xu C, Gui C, Fox MD, editors. Level Set Evolution without Re-Initialization: A New Variational Formulation. Computer Vision and Pattern Recognition, 2005 CVPR 2005 IEEE Computer Society Conference on; 2005.