IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Adaptive appearance separation for interactive image segmentation based on Dense CRF

ISSN 1751-9659
Received on 25th February 2018
Revised 29th September 2018
Accepted on 22nd October 2018
doi: 10.1049/iet-ipr.2018.5073
www.ietdl.org

Zili Peng[1], Qiaoliang Li[1] ✉

[1]Key Laboratory of High Performance Computing and Stochastic Information Processing (HPCSIP, Ministry of Education of China), College of Mathematics and Computer Science, Hunan Normal University, Changsha, Hunan 410081, People's Republic of China

✉ *E-mail: liqiaoliang@hunnu.edu.cn*

Abstract: Interactive segmentation has recently become a hot topic for its wide application. The authors propose an efficacious appearance separation model for interactive binary segmentation, which incorporates the difference of foreground and background colour models and the difference of corresponding geodesic models into the popular densely connected conditional random field (Dense CRF) framework. The proposed method can adaptively set relevant parameter values in this framework according to the characteristics of target images in a per-image manner, therefore, it gets rid of the dependence on specific datasets. After accomplishing a mean-field inference, the authors are able to get satisfactory results without the time-consuming parameter learning process and multiple iterative optimisations. Overall, the proposed approach is highly efficient and mitigates the contradiction between accuracy and segmentation efficiency. In addition, the proposed approach reduces the efforts of scribble-style interaction from users. The experimental results on three famous datasets show that the proposed method is superior to the other five new algorithms released in recent years regarding accuracy, and is faster than or close to them in runtime.

1 Introduction

Many research institutions and universities have been enthusiastic about researching image segmentation for several decades [1, 2], however, it is still a great challenge for further study on computer vision. Due to intricacies and multiplicities of images in the real world, full-automatic segmentation is intrinsically ambiguous and cannot obtain satisfactory results in practical applications. On the other hand, interactive image segmentation [1, 3–6] allows users' interventions to appoint some foreground (FG) and/or background (BG) pixels, therefore, high-level understanding of contents is incorporated into the process, and more accurate results can be obtained using interactive segmentation.

Recently, it has been shown that the mean-field inference in densely connected conditional random fields (Dense CRFs) performed by Krähenbühl and Koltun [7, 8] is highly efficient and gains remarkable accuracy in computer vision applications. The two main reasons are: the long-distance modelling of pixel correlations, and the fast inferencing in Dense CRFs with Gaussian edge potentials performed by the approximation of Gaussian convolutions [9]. Vineet *et al.* [10, 11] show that generalised Gaussian kernels and models with higher-order pairwise terms can also be applicable for the inference of Dense CRF framework. Desmaison *et al.* [12] demonstrate the benefits of continuous relaxations over the mean-field inference and provide an excellent alternative solution to the Dense CRF framework. Ajanthan *et al.* [13] propose an efficient linear programming algorithm for minimising the potential of Dense CRF. Their method can achieve a potential drop in a relatively short period of time, however, they also point out that a lower potential does not mean a better segmentation result.

The goal of binary image segmentation is to separate pixels in an image into FG and BG [14]. Despite the development of the efficient inference in Dense CRF algorithms, binary segmentation based on the Dense CRF framework is difficult for applications. The reasons are that this requires a complex process of parameter learning and a comprehensive database for learning, in addition, little is known about parameter estimations for Dense CRF [8]. To overcome these difficulties, we propose an adaptive appearance separation method (AASM) for binary interactive image segmentation based on the Dense CRF, which improves efficiency and accuracy substantially.

We make the following contributions in this paper. First, we propose a utility appearance separation model (ASM) for interactive binary segmentation. This model combines spatial location information with colour features, reduces or even eliminates the irrelevant interference from the BG while segmenting out the FG. Second, instead of setting parameters in fixed statistics feature values based on the datasets, our method can adaptively fine-tune relevant parameter values in Dense CRF framework according to the characteristics of the images to improve the segmentation results. Finally, the adaptive appearance separation method (AASM) we created reduces the workloads of users in seeded image segmentation, especially in binary instance segmentation, and achieves better segmentation performance in terms of accuracy and efficiency with a mean-field inference.

2 Related work

Tang *et al.* propose the OneCut algorithm [15] in order to improve the efficiency of GrabCut [16]. To obtain accurate segmentation results, GrabCut needs time-consuming iterative updates of the colour models and corresponding parameters, while OneCut obtains the result by using graph-cut [17] only once. Tang *et al.* suggest that the difference between FG and BG colour models represented by unnormalised histograms based on L1-norm obtains the highest accuracy, and the L1-norm they utilised is better than any other forms of approximate appearance overlap terms regarding computing time. Therefore, they utilise this L1-distance to approximate the appearance overlap [15], and gain exciting results based on the graph-cut framework [17] with high efficiency.

DenseCut [18] proposed by Cheng *et al.* is based on the Dense CRF framework [7, 8]. Iterative improvement of the FG and BG colour models [15] is a key step in GrabCut [16] and is replaced by a fully connected CRF, so DenseCut achieves great efficiency. Also, Cheng *et al.* demonstrated high accuracy of DenseCut in their experiment. The following three factors contribute to its efficient and accurate segmentation results: the use of Dense CRF framework to achieve precise and efficient inference, the clever use of efficient colour classifier to describe the colour model, and the

setting of seven reasonable parameter values in their algorithm, according to some relevant statistical works on the testing datasets.

In the last few years, many different interactive segmentation techniques have been proposed. Here are a few:

(i) Random walk (RW) based approaches: For example, Shen *et al.* put forward the lazy RW (LRW) algorithm [19], which is a generalisation of RW image segmentation [20] method proposed by Grady. They set the probability of a RW staying at the current node as $1-p$, and the probability that the RW goes out along the edges adjacent to the current node as p. As a result, the image is over-segmented into several regions. In short, LRW is a good way to generate superpixels. Other examples are sub-Markov RW (SubRW) algorithm [21] and power watershed (PW) algorithm [22] proposed by Dong *et al.* and Couprie *et al.* separately. The former is based on the traditional RW method, adding the label priori auxiliary node has a good effect on twigs segmentation. The latter gives a new explanation for a series of fundamental algorithms, such as RW, graph cut, and shortest path, and puts forward a unified energy minimisation scheme.

(ii) Methods based on normalised cuts (NCuts) [23]: For example, the biased NCuts algorithm [24] proposed by Maji *et al.* makes a significant breakthrough in maintaining the integrity of FG targets by adding prior constraints. Rezvanifar and Khosravifard [25] use the superpixel algorithm (such as mean shift [26, 27] or simple linear iterative clustering (SLIC) [28]) to segment the image into the superpixel map, and then join the size of the region constraints on the basis of NCuts, their method improves the segmentation efficiency, to some extent, the segmentation effect is increased as well; Chew and Cahill [29] improve NCuts algorithm, by adding soft constraints that must be linked and cannot be linked on the basis of NCuts. The Dinkelbach NCut algorithm [30] proposed by Ghanem *et al.* provides an efficient solution to the NCuts problem of adding prior and convexity constraints. Shen *et al.* unify NCuts and graph cut algorithms into an energy optimisation framework and improve the image segmentation results by adding smoothing terms to the Laplacian energy cost function in the literature [31]. They also provide an acceleration strategy for large image segmentation.

(iii) Improved segmentation methods based on geodesic distance: Such as the works in [32–38], these methods incorporate the geodesic distance information in various ways and achieve quite exciting results. Criminisi *et al.* [32, 33] design a geodesic symmetric filter (GSF) for image segmentation based on geodesic distance and mathematical morphology. Their method is very efficient because it performs only morphological operations like opening and closing. However, in addition to requiring the users to provide seed point interaction information, this method needs subjectively specify two geometric parameters to remove the noise area in the FG and the BG. This is extremely tedious for the segmentation of appearance-complex images with different sizes or for the segmentation of images with large differences among FG and BG sizes. Wang and Yagi [39] add shape prior to GSF and preserve the integrity of FG objects. However, they set the two geometrical parameters of the GSF only related to the image size. The segmentation effect will undoubtedly be affected by the segmentation image's FG and BG appearances, specifically, that is their sizes and level of messiness. Instead of human subjectively specifying the two crucial parameters for removing noisy regions, our approach provides soft-segmented results based on the estimated appearance overlap, and then automatically set up the relevant parameters in the Dense CRF framework, finally, per-pixel accurate segmentation results are obtained based on a mean-field inference. We believe that is the most significant difference between our approach and the GSF-based approach.

(iv) Many other prominent segmentation methods: Such as segmentation based on level set method [40–42], segmentation combined the saliency [43–45], and method based on region merging [6, 46], and so on. In addition, there are some co-segmentation techniques [47–49] that are used for segmenting common objects in multiple images. Therefore, in addition to the information between FG and BG in a single image, the task of co-segmentation can also use the related information between images to be segmented.

3 Adaptive appearance separation based on Dense CRF

Our approach is based on the efficient filter-based inference in Dense CRF framework [7, 8, 11]. The Gibbs energy [8] of the Dense CRF is defined as:

$$E(x\Big|I,\theta)=\sum_i\varphi_i(x_i\Big|I,\theta)+\sum_{i<j}\varphi_{ij}(x_i,x_j\Big|I,\theta) \tag{1}$$

where $x\in\{B,F\}$ is a label assignment related to the model θ, B is the BG, F is the FG, i and j are pixel indices, ranging from 1 to N, and N is the number of pixels in image I. φ_i is the unary term, which measures the cost assignment x_i of pixel i; φ_{ij} is the pairwise term, which introduces a penalty for nearby similar pixels that are assigned different labels [18]. For notational convenience, we will use $\varphi_{ij}(x_i,x_j)$ to denote $\varphi_{ij}(x_i,x_j\big|I,\theta)$ in the rest of the paper. The form of the unary term can be chosen arbitrarily, while the pairwise term must be in the form of a linear combination of the weighted Gaussian kernels [7, 11]. Once the CRF formulation meets this form, [8] enables us to utilise Gaussian filtering techniques [9] within a mean-field approximation framework [7] to obtain solutions from a rapid maximum posterior marginal inference [11].

3.1 Proposed method

Motivated by the work of the authors in [15, 18, 35], we propose an efficient AASM based on the Dense CRF framework [7, 8].

For the unary term, we define it as

$$\varphi_i(x_i)=-\log P(x_i) \tag{2}$$

where $P(x_i)$ is not as simple as the use of naive FG/BG colour model Colour(x_i) in [18]

$$\text{Colour}(x_i)=\frac{P(\Theta_{x_i},I_i)}{P(\Theta_F,I_i)+P(\Theta_B,I_i)} \tag{3}$$

but the combination of the colour models difference $C(x_i)$ and the geodesic distance difference $G(x_i)$, which is defined as the proposed ASM (the form of $P(x_i=F)$ is as follows, a similar case $P(x_i=B)$ can be easily obtained):

$$P(x_i=F)=\lambda\times(C(x_i=F)+G(x_i=F)\times 8)\times 0.12 \tag{4}$$

$$c(x_i=F)=\frac{P(\Theta_F,I_i)-P(\Theta_B,I_i)}{P(\Theta_F,I_i)+P(\Theta_B,I_i)} \tag{5}$$

$$C(x_i=F)=\begin{cases}c(x_i=F) & \text{if } c(x_i=F)>0\\ 0 & \text{otherwise}\end{cases} \tag{6}$$

$$\text{Geo}(x_i=F)=\frac{D_F(pix_i)-D_B(pix_i)}{D_F(pix_i)+D_B(pix_i)} \tag{7}$$

$$G(x_i=F)=\begin{cases}-\text{Geo}(x_i=F) & \text{if Geo}(x_i=F)<0\\ 0 & \text{otherwise}\end{cases} \tag{8}$$

where λ is used to distinguish the importance of the unary term in comparison with the pairwise term. The reason why $P(x_i)$ is defined like that form is to limit it to be an effective probability (between 0 and 1) of pixel i belonging to the appearance model. $P(\Theta_l,I_i)\in(0,1)$ represents the probability density value of a colour I_i belonging to the colour model Θ_l, where label $l\in\{F,B\}$. Thus, $c(x_i=F)$ means the colour difference between the FG colour model and the BG one. $D_l(pix_i)$ is the geodesic distance term [35] that is defined as the geodesic distance from pix_i to the closest

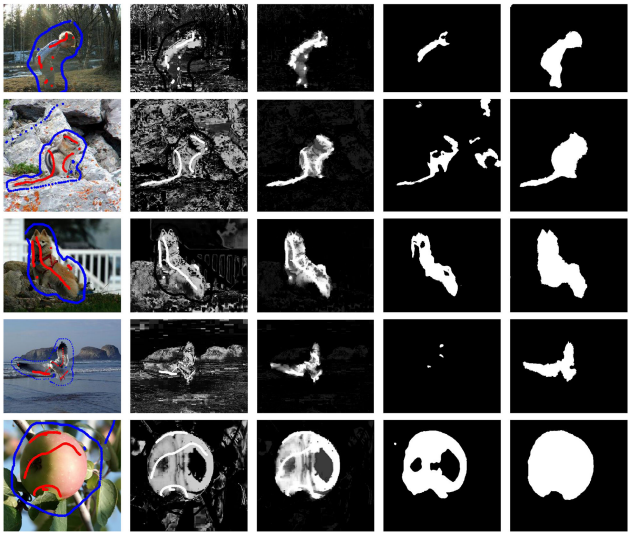

Fig. 1 *Examples of naive FG colour models and our FG ASMs that are derived from the given input images and FG/BG seeds. FG strokes are coloured red, and BG strokes are blue, and each row is an example*
(a) Images with FG/BG seeds, ***(b), (c)*** NCM maps and our ASM maps, respectively, ***(d), (e)*** Segmentation results from Dense CRF framework using the second and third columns of the models as unary term separately

seeded pixel F or B. It can be calculated using the classical Dijkstra's algorithm [38]. Thus, $\mathrm{Geo}(x_i = F)$ means the geodesic distance difference between the FG geodesic map and the BG one. Essentially, the geodesic distance is a shortest weighted path on the gradient map of the colour model [35, 50]. Rather than the straight line connecting the two pixels like in an Euclidean distance measure, geodesic distance is the weighted shortest distance from one point to another. Geodesic clue is better than the Euclidean measure on maintaining the internal geometry of the data globally. In addition, geodesic is apt to separating FG and BG parts that are nearby.

Similar to appearance overlap approximation [15] using the L1-norm between the colour histograms in OneCut, we use the difference between the estimated FG and BG colour model terms to approximate the objective appearance overlap mentioned in the formula of $C(x_i)$. After computing the values of $P(\Theta_l, I_i)$ and $D_l(pix)$, we can immediately get the exciting ASM, an excellent initial result of soft segmentation (see Fig. 1*c*).

Compared with the simple colour model, it can be found intuitively that the advantages of using our ASM as the unary term in Dense CRF framework include a large number of unrelated clutter BGs removed and the FG of some isolated holes filled (see Fig. 1*e*). There will be a detailed introduction of the advantages in Section 4.1.

For the pairwise term, we model it as the mixture of Gaussian functions related to colours and locations following [7, 18], and take the form of

$$\varphi_{ij}(x_i, x_j) = u(x_i, x_j)\left(w_1 K_1(f_i, f_j) + w_2 K_2(f_i, f_j)\right) \tag{9}$$

$$\begin{aligned} K_1(f_i, f_j) &= \exp\left(-\frac{\mathrm{dist}_{ij}^2}{\theta_d^2} - \frac{\Delta I_{ij}^2}{\theta_{I1}^2}\right) \\ K_2(f_i, f_j) &= \exp\left(-\frac{\Delta I_{ij}^2}{\theta_{I2}^2}\right) \end{aligned} \tag{10}$$

where $u(x_i, x_j)$ is a function used to reflect the label compatibility [7, 8, 11], we set it as the Potts model [7, 18]: $u(x_i, x_j) = 1_{[x_i \neq x_j]}$. K_1 and K_2 are Gaussian functions defined on feature f_i, f_j; w_1 and w_2 are the weighted factors of the kernel functions; dist_{ij} and ΔI_{ij} are the distance and colour differences between pixels located in i and j, respectively; θ_d, θ_{I1}, θ_{I2} are used to adjust the degree of smoothness, similarity, and tightness. The meaning of equality (9) and (10) is consistent with the expression of DenseCut [18]. Here, we discard the third kernel in the DenseCut algorithm because it still reflects the similarity of colours and is only used for fine-tuning. Our algorithm can completely replace that kernel by setting the corresponding coefficients reasonably. Details on how to set the values of the main parameters w_1, w_2, θ_d, θ_{I1}, θ_{I2} for our method will be discussed in the next subsection.

3.2 Setting of related parameters

It seems to be a considerable puzzle to set those parameters in our method at first glance. Through a large number of experiments, we find each parameter can be set adaptively according to the image data and get satisfactory segmentation results.

As naive colour model (NCM) $\mathrm{Colour}(x_i)$ (equality (3)) is an essential role of simple Bayesian classifier [35], we can estimate the error of classification

$$\varepsilon = \frac{1}{2}\left[\frac{\sum_{x_i \in F}(1 - \mathrm{Colour}(x_i))}{|\Omega_F|} + \frac{\sum_{x_i \in B}(1 - \mathrm{Colour}(x_i))}{|\Omega_B|}\right] \tag{11}$$

where $|\Omega_F|$ and $|\Omega_B|$ represent the number of seeds of the FG and BG, respectively. It is necessary to assign a large weight to the ASM when there is a minute error ($\varepsilon \simeq 0$); the classifier is unreliable when the error ε is too big (i.e. an error $\varepsilon \geq 0.5$ is caused by the fuzzy colour models), we need to give the ASM no weight or nearly zero weight [35]; the classified error ε will grow along with the indistinctness of the colour models, therefore, giving less weight to the ASM according to this error is a wise choice. Hence, we can introduce a confidence term κ for three weighted parameters

$$\kappa = \begin{cases} 0 & \text{if } \varepsilon \geq 0.5 \\ 1 - 2\varepsilon & \text{otherwise} \end{cases} \tag{12}$$

Based on the above analysis and experimental experience, we set the Dense CRF framework weight of unary terms as $\lambda = \kappa$; the weighted coefficient of the appearance kernel is set as $w_1 = 55 \times (1 - \kappa)$, that is used to weigh adjacent pixels with similar colours assigned the same label according to the classifying confidence; the weighted factor of the smooth kernel is set as $w_2 = 0.8 \times \kappa$, that is used to weigh the size of the outlier isolated points that need to be removed according to the classifying confidence. As a result, the relative weighting of the ASM, smoothness, and similarity will be adjusted on a per-image basis, which strengthen the adaptability and flexibility of our method.

Due to the sophistication of objects and scenes in the real world, setting θ_d, θ_{I1}, and θ_{I2} as fixed values cannot reflect the nature of the image itself and may excessively depend on particular image datasets. Therefore, we set them empirically according to each image data, rather than in a fixed way: $\theta_d = 0.093\sigma$, $\theta_{I1} = 1.8\sigma$, $\theta_{I2} = 1.5\sigma \times (1.1 - \kappa)$, where σ is the square root of the average ΔI_{ij}^2 over the image.

Just as the work in [18], our CRF formulation exactly meets the form of the Dense CRF framework. Hence, we can perform rapid message passing [7, 8] using high-efficiency Gaussian filtering [9, 11]. We only focus on the carefully designed AASM for interactive segmentation in this paper; for the theoretical and experimental analysis of this filter-based acceleration method, please refer to the works of Krähenbühl and Koltun [7, 8].

Our ASM provides promising initial values (as shown in Fig. 1*c*) for the methods based on Dense CRF framework, thus, we can refine the final segmentation results just by slightly adjusting the values of the parameters in the pairwise term. However, the ASM is closely related to the complexity of the image itself and the strokes from users, so the colour models difference $C(x_i)$ and the geodesic distance difference $G(x_i)$ will vary. To reduce interactive workload and to improve the result, it is necessary to adaptively adjust the values of λ, w_1, and w_2 according to the variations of κ, rather than only considering the ΔI_{ij}^2.

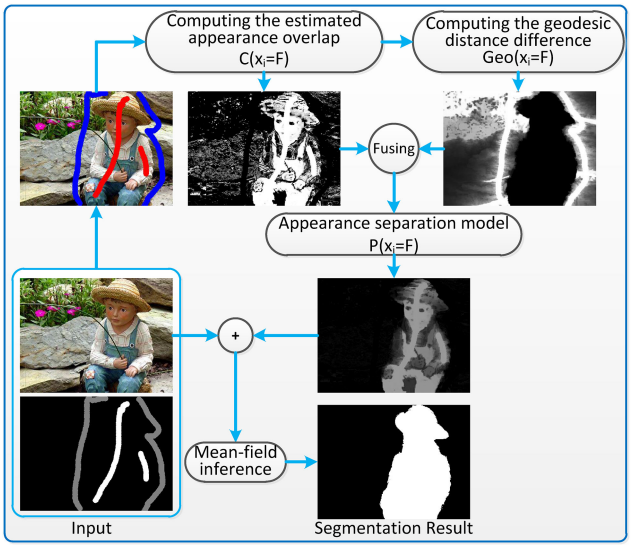

Fig. 2 *Pipeline of our AASM based on Dense CRF*

3.3 Pipeline and implementation

There are usually three user interaction forms in interactive image segmentation: (i) FG/BG clicks; (ii) FG/BG scribbles or strokes; (iii) bounding box around the desired FG [38]. However, we only focus on the second input type: scribbles. It is a more practicable and flexible way to get highly accurate segmentation results by using the form of scribbles for the images in which object edge is blurry, FG and/or BG colour model(s) is/are indistinct, or illumination varies greatly. Similar to most interactive image segmentation methods, we need to manually select the appropriate FG (the target of interest related to the specific application) and the BG seed points according to specific applications. Generally, there is no difference in seeds selection between our method and other interactive image segmentation methods, but for some images which are very difficult to segment, our approach allows for improved seed points selection. For example, for targets with small differences between FG and BG colours and with blurred boundaries, we usually place seed points on both sides of the boundary to prevent the target pixels 'overflowing' into the BG region, while suppressing BG pixels 'permeating' to the target region. As another example, for targets with obvious line barriers, multiple texture areas, or step type colour gradients in the region, to prevent the target area from being lost, we need to place seed points in the intermediate zones where these changes are dramatic.

The pipeline of our method is summarised in Fig. 2. Following [16, 18], we use the well-known Gaussian mixture models (GMMs) to represent the probability density value $P(\Theta_l, I_i)$, and each GMM consists of five Gaussian models. According to the seed point interaction information appointed by users, FG and BG classes can be initialised, and GMMs for this two classes can be created separately using the Orchard and Bouman colour clustering algorithm [51]. Then we modify the initial GMMs and add the interaction information as a rigid constraint to the estimated FG/BG appearance model. Different from the popular GrabCut [16], GMMs in our method are not required to be learned and updated iteratively, which significantly improve the efficiency of our algorithm. The simple difference in equality (5) and function (6) gives the estimated appearance overlap map: $C(x_i = F)$. The appearance overlap estimated here is very similar to that of the OneCut algorithm because the OneCut algorithm approximates the appearance overlap by utilising the L1-norm of the unnormalised histogram, whereas the appearance overlap of our algorithm uses the difference of the appearance models approximated by GMMs. With these estimated FG/BG appearance models, the geodesic distance can be quickly calculated using the Dijkstra's algorithm, followed by the differential of equality (7) and function (8), the geodesic distance differential map $\mathrm{Geo}(x_i = F)$ can be obtained. The ASM $P(x_i = F)$ presented in this work can be obtained by the fusion of the estimated appearance overlap map and the geodesic distance differential map according to equality (4). Finally, according to the parameter setting scheme in Section 3.2 and taking full advantage of the existing Dense CRF framework, we are able to obtain a satisfactory segmentation result by making a mean-field inference.

The most direct reasons for choosing the Orchard and Bouman colour clustering method [51] are as follows: one is because it is fast, which is twice the speed of the *K*-means method; another one is that it can get compact and well-separated clusters compared to other clustering methods, which have a huge contribution on the separation between the FG and the BG. In order to separate the FG from the BG as accurate as possible, we realise that clusters with a low variance can achieve better separation from other clusters, and the Orchard and Bouman colour clustering methods operate exactly like that. By the way, as long as the colour clustering method can efficiently obtain tight and well-separated clusters, then it is not a bad idea to switch to other clustering algorithms.

4 Experiments

We evaluated our adaptive appearance separating interactive segmentation algorithm on three famous datasets: MSRA Database (Image Set B) [52], GrabCut Dataset [16], and BSD500 Dataset [53]. We select 640, 30, and 300 pictures separately from these three databases as a test set. These pictures are either with the edge of the target blurry, or with messy FG/BG, or with non-homogeneous lighting, which bring great challenges to the segmentation; moreover, even in the interactive form of bounding box, segmentation will fail to obtain accurate results. After comparing our ASM with the NCM Colour(x_i), we contradistinguish our results from the state-of-the-art competitors (OneCut [15], GeoGC [35], SubRW [21], TRC [14], and DenseCut [18]) qualitatively and quantitatively. All the methods use default values for their parameters in this comparison. Since the official DenseCut [18] uses the bounding box to interactively segment the test image set we use, and the results are not in line with the results we expect, we modify the code to the way of strokes based on the algorithmic ideas described in the literature. The codes of the rest of other algorithms are officially released.

4.1 Appearance separation model versus NCM

The NCMs Colour(x_i) that are derived from the probability density values $P(\Theta_l, I_i)$ are cluttered usually. In contrast to the NCMs, our ASM $P(x_i)$ combines spatial location information with colour feature among pixels, therefore, it indirectly plays a smooth role in removing isolated points, reducing the effect from BG in the middle part of the FG and reducing or even eliminating the irrelevant interference from the BG, as shown in Fig. 1*c*. This property fits well with the request of interactive segmentation. The benefits of using our newly designed ASM in segmentation can be found naturally by comparing Figs. 1*d* and *e*. According to [11], the mean-field inference methods in the Dense CRF framework are vulnerable to initialisation that is not so superior, it may converge to local minima, nevertheless, our ASM provides superior fundamental values for the methods based on Dense CRF framework. In other words, our ASM is a better initial result of soft segmentation for Dense CRF framework. Bringing the colour and spatial information (geodesic distance) into the computation of the unary term can significantly benefit the segmentation. Examples of naive FG colour models Colour(x_F), ASMs $P(x_F)$, and corresponding segmentation results based on Dense CRF framework can be found in Fig. 1.

4.2 Reducing the user scribble-interaction workload

If the colour features of the FG and the BG are close to each other or the appearance overlap is severe, the segmentation result may easily lose some of the FG or be doped with irrelevant BG. To obtain accurate segmentation results, users often need to continue to join the seeded information based on the last segmentation result. At this point, if we merely consider the colour information, it may result in the phenomenon of 'attend to one thing and lose

Fig. 3 *Examples of binary instance segmentation*
(a) Input images that are marked with FG and BG scribbles, ***(b)–(g)*** Segmentation results from OneCut, GeoGC, TRC, DenseCut, SubRW, and ours, respectively, ***(h)*** Segmentation result of our algorithm overlaps with the original image to be segmented, the red indicates the instance mask, and the blue indicates the BG mask

Table 1 Binary classification confusion matrix

		Segmentation result	
		Foreground	Background
ground truth	foreground	TP	FN
	background	FP	TN

sight of another': retrieving the missing FG is likely to misjudge irrelevant BG as targets, or removing the redundant BG is likely to exclude the target part at the same time, which undoubtedly increases the user's workload of scribbling interaction. Our method integrates the spatial location information and takes full advantage of the image colour features and the positional relationship between pixels and scribble-seeded points, consequently, the workload of user interaction is reduced. Here, the instance segmentation [54] is provided as an example. Instance segmentation is not only required to classify a certain type of objects but also to distinguish the different individuals of such objects. For the sake of simplicity, we use binary image segmentation in scribble-interactive mode. The classifications are directly judged by users, and only one of the multiple instances in the image is segmented. Due to the common attributes of the same category among different instances, such as similar shapes, colours, textures, and so on, the conventional algorithms often require a large amount of user scribbles interaction to separate an instance from other instances and BGs completely. The more the similarity among adjacent instances, the greater the amount of interaction-workload required. However, our method requires only a small amount of scribbles to mark the FG specified by users and to remove the BG if necessary, then the specific instance can be cut out. In the same condition of scribble position and number of seeds, the segmentation results of other methods are inaccurate comparing to our method: either redundant or deficient, as shown in Fig. 3. They need to add further seeds to improve the segmentation results, but it is still possible facing the above-mentioned 'attend to one thing and lose sight of another' phenomenon. The cases of image segmentation with severe appearance overlap or with clutter appearances are similar to the case of instance segmentation, hence we will not repeat them.

4.3 Qualitative and quantitative comparison

To objectively quantify the accuracy of a segmentation result given a ground truth (GT) and strokes input, we follow the common practice by using six criteria: precision (Pr), true positive rate (TPR), *F*-measure [18], Jaccard coefficients (JaccardC) [38], rand index (RI), and false positive rate (FPR). The more the similarities between segmentation result and GT, the larger the score of Pr, TPR, *F*-measure, JaccardC, and RI will be, and the lower the value of FPR will be. However, high Pr, high TPR or low FPR does not guarantee better segmentation, we can find this rule according to the definition of these metrics. In our selected image test set, 30 images of the Grabcut dataset [16] and 300 images of the BSD500 dataset [53] have corresponding ground truths, and matching ground truths of the MSRA dataset [55] is only 270. For the sake of fairness, we give the same strokes input to all methods.

Before comparing, we give a binary classification confusion matrix in Table 1 to help define the evaluation criteria. As stated in the table, TP means the segmented FG is also the FG target in the GT; the meaning of the rest of notations (FN, FP, and TN) in this table can be easily derived in a similar way. JaccardC is defined as the ratio of segment intersection's size to segment union's size [38], formally: $\text{JaccardC} = \text{TP}/(\text{TP} + \text{FN} + \text{FP})$. $\text{TPR} = \text{TP}/(\text{TP} + \text{FN})$, which represents the rate at which all FG parts are correctly segmented. $\text{FPR} = \text{FP}/(\text{FP} + \text{TN})$, which represents the ratio of all BGs being incorrectly segmented into the FG. In the binary segmentation task, $\text{RI} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{FN} + \text{TN})$, it describes the percentage of the image's FG and BG being correctly segmented. *F*-measure measures the weighted harmonic mean value of precision ratio and recall ratio [15, 18]:

$$F_\beta = ((1 + \beta^2)\ \text{Pr}\ \times \text{Re})/(\beta^2 \times\ \text{Pr}\ + \text{Re}),$$

where $\text{Pr} = \text{TP}/(\text{TP} + \text{FP})$ is the rate of correctly segmented pixels [11]; Re is also called sensitivity of the classifier, which is equal to TPR; β is used for weighing the relative importance between Pr and Re. Note that TP, FN, FP, and TN in these evaluation metrics take the number of pixels as their unit. In order to facilitate the display of statistical results with other measures, the FPR is displayed with the corresponding criteria of the specificity ($\text{Specificity} = 1 - \text{FPR}$). Fig. 4 shows the average values of Pr, TPR, F-measure (the relative importance $\beta = 0.3$, as used in [43]), JaccardC, RI, and Specificity (or 1−FPR). Combined with these criteria's definitions, it can be seen from Fig. 4 that our segmentation results are similar to the GT, and are better than others on the three datasets on average, followed by GeoGC, DenseCut, SubRW, TRC, and OneCut.

In addition, in the binary image segmentation task, the misclassification rate (MR):

$$\text{MR} = (\text{FP} + \text{FN})/(\text{TP} + \text{FP} + \text{FN} + \text{TN}) = 1 - \text{RI}. \quad (13)$$

Thus, we show the average results of MR on each dataset in Fig. 4 (shown by the corresponding RI). Assuming MR>0.25 in this experiment indicates the segmentation fails, then after summarising the results on three datasets, the misclassification frequency of the six algorithms (Onecut, GeoGC, subRW, TRC, GCMF, and ours) are: 4.00%, 0.50%, 3.50%, 0.83%, 1.00%, and 0.33%, respectively.

Some qualitative results are shown in Figs. 5–7. In general, given the same seed point conditions (number of seed points, locations, FG and BG indications are the same), the overall effect of the existing method is not bad, but our approach is much better. The following are the three reasons: first, our process makes full use of the interactive information provided by users, second, we try to fully and accurately represent the inherent features of the image itself, finally, our method establishes the relationship between the interactive information and the characteristics of the image itself (The first two reasons are the basis for efficient and accurate interactive image segmentation. The last one is the key to improve

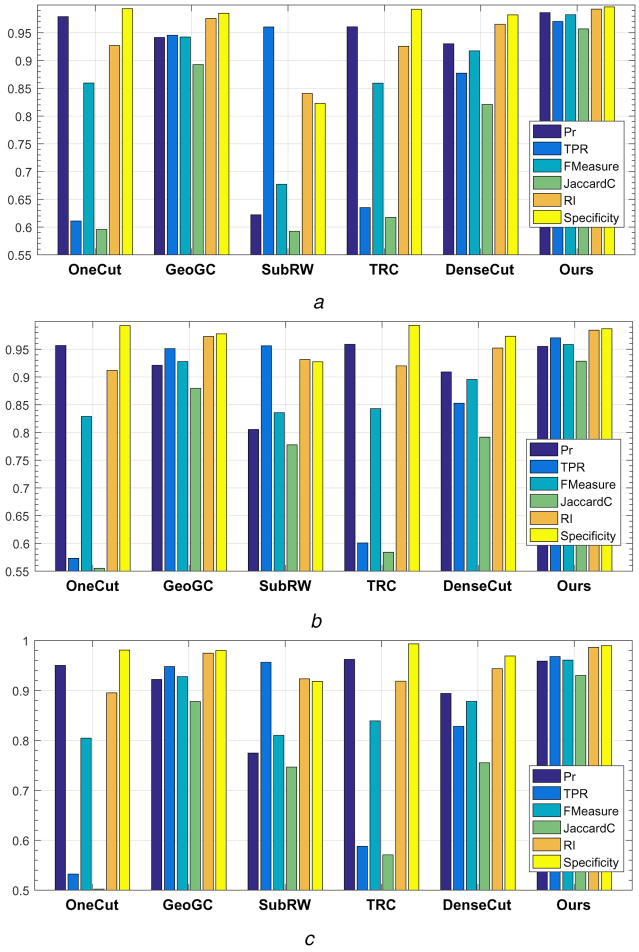

Fig. 4 *Quantitative comparisons of results from six interactive segmentation methods are OneCut, GeoGC, SubRW, TRC, DenseCut, and Ours, with the comparison measures of Pr, TPR, F-measure, JaccardC, RI and Specificity*

(a)–(c) Comparison on GrabCut, MSRA, and BSD500 dataset ground truth, respectively

segmentation and user experience.). Analysis of the reasons for the six algorithms segmentation results are as follows:

(i) As the spatial information among pixels and seeds, and object boundary information are not specially considered in OneCut, the segmentation may fail. Note that many of the images in the test set are either similar in FG/BG colour, or light range wide, as a result, OneCut is prone to isolated points, blocks, and fake FG results in the condition of different scribbles. The segmentation result of OneCut in Figs. 5*b*, 6*b*, and 7*b* is derived from officially released code and the default parameter values. Though the result is not as good as the official one and may conflict with Fig. 6 in [15], our experiment is rigorous and convincing. We can see from the fourth image in Fig. 5*b*, the first image in Fig. 6*b*, and the first image in Fig. 7*b* that OneCut only segments part of the objects. Note that the respective objects of the three images to be segmented are relatively close to the BG in colour feature, under this condition, OneCut's ability to describe the image attributes in such a way that the L1-norm approximate appearance overlap between FG and BG colour models is limited, moreover, the association between the interactive information and the characteristics of the image itself is not specifically set in OneCut, which leads to unsatisfied segmentation results. Users may need to provide a large number of seed points to improve segmentation, which reduces the user experience.

(ii) GeoGC alleviates the short-cutting problem [35] of ordinary graph cut algorithm. However, it lacks enough considerations of colour similarity, GeoGC can only get inaccurate segmentation results given weak interactions on some challenging segment tasks. From the second image in Fig. 5*c*, the first image in Fig. 6*c*, and the fourth image in Fig. 7*c*, it can be found that GeoGC has made some mistakes in the segmentation of the region near some seed points. The main reason is that GeoGC has limited ability to express image characteristics (colour, texture, boundary strength). Although GeoGC uses the geodesic distance to establish the spatial positional relationship between pixels and seed points, it does not make full use of the intrinsic relationship between the interactive information and the characteristics of the image itself, so, under the condition of limited seed points, the segmentation accuracy needs to be further improved. However on the whole, its segmentation result is good and is catching up with ours.

(iii) The TRC algorithm is actually an improvement on the OneCut algorithm, by adding convexity prior information, designing a dynamic programming algorithm and solving through an iterative approximation optimisation technique. The segmentation result from TRC is basically using convex polygon(s) to approximate the object. However, the result of this method is not particularly accurate because the objects in the real world are not always convex due to occlusion, perspective, and so on. From the second image in Fig. 5*d*, the second image in Fig. 6*d*, and the third image in Fig. 7*d*, it can be found that the segmentation result of TRC is almost a convex polygonal approximation to the object, and once positions of seed points destroy the convex structures of objects, it may result in abysmal results like the fifth image in Fig. 7*d*. This is mainly because the method is to solve the problem that OneCut is prone to isolated points, and it is an optimised solution scheme specially designed for the object of convex structure, which has strong application scenario restrictions.

(iv) As the mean-field inference methods are vulnerable to inferior initialisation, it may converge to local minima [11]. It is critical to estimate well fundamental values for the methods based on Dense CRF framework. Due to the intricacy of the selected test image set, the NCM is often cluttered. When this model is directly used as the unary term for DenseCut, we may get undesirable segmentation results. From the second image in Fig. 5*f* and the fifth image in Fig. 6*f*, it can be found that when there are distinct lines in the middle of the target to separate the different colour regions, it is difficult to achieve precise segmentation without adding more seed points. For the first image of Fig. 7*f*, because the target and BG textures are not distinguishable enough, even if enough seed points are placed, Densecut fails again. Overall, DenseCut does not adequately and reasonably utilise the user interaction information, and pays too much attention to the inherent attributes of the image. Besides, the relevant parameters are not adjusted according to the specific picture and the segmentation target, and the input to the mean-field estimation method is a poor initialisation result. All of this led to its unsatisfactory final segmentation results.

(v) According to the optimisation explanation of the SubRW's objective function, the algorithm considers three terms: unary term, smoothing term, and label priori term [21]. Essentially, the label priori term consists of the exact priori from user scribbles and the 'fuzzy' priori represented by probability distributions constructed by user scribbles using GMMs. SubRW is originally designed to solve the problem of twigs segmentation. To cut out of twigs that are similar to the main branch part, SubRW makes full use of the prior from the user scribbles in the trunk part of the object. However, the test images of this experiment are affected by illumination, colour, texture, blurry edges, and many other factors, GMMs constructed using only colour information tend to have many tedious components and may be highly ambiguous, so it is far from enough to simply add 'fuzzy' priori to the expanded graph for SubRW. Although SubRW considers the impact of noise resulting from added priori labels, it does not take consideration of the distance and positional relationship between pixels and seeds, and cannot eliminate the irrelevant BG. Therefore, the segmentation effect of SubRW on the three datasets in this experiment is not particularly good. From the last image in Fig. 5*e*, the first image in Fig. 6*e*, and the fifth image in Fig. 7*e*, it can be found that SubRW made some mistakes in the unknown area away from seed points. This is because the method has a limited degree of association between the interactive information and the characteristics of the image in the 'fuzzy' priori. Although the GMM used has a strong feature description capability, SubRW

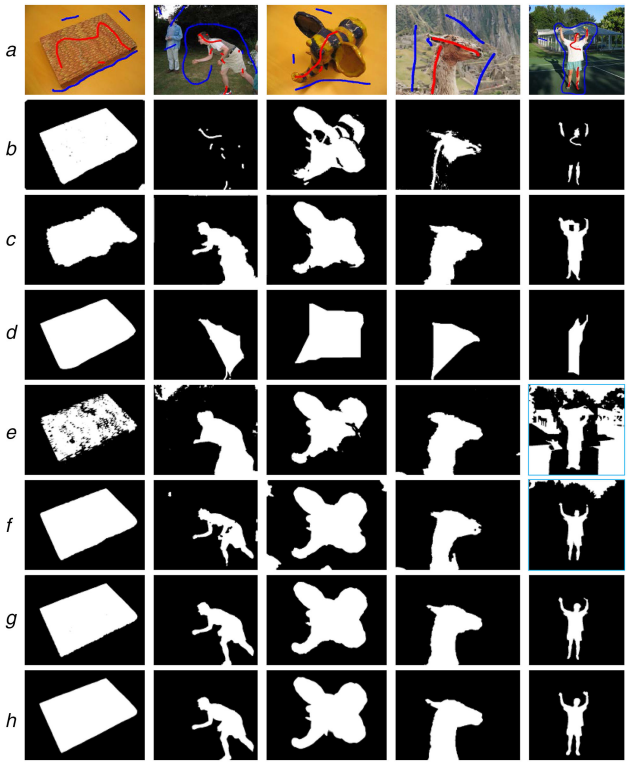

Fig. 5 *Examples of qualitative results on GrabCut datasets with the same user strokes. FG strokes are coloured red and BG strokes are blue, and each column is an example*

(a)–(g) Input image with FG/BG strokes, the output from OneCut, GeoGC, TRC, SubRW, DenseCut, and Ours, respectively, ***(h)*** Ground truth (GT)

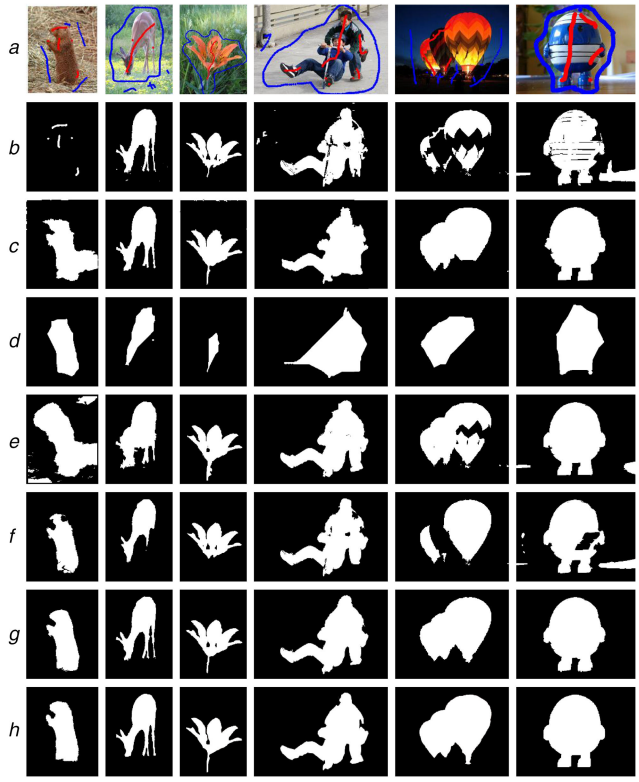

Fig. 6 *Examples of qualitative results on MSRA datasets with the same user strokes. FG strokes are coloured red and BG strokes are blue, and each column is an example*

(a)–(g) Input image with FG/BG strokes, the output from OneCut, GeoGC, TRC, SubRW, DenseCut, and ours, respectively, ***(h)*** Ground truth (GT)

does not involve the positional relationship between pixels and seed points, which makes impossible to focus on the target to be segmented, and the segmentation process is susceptible to interference with other contents which are similar to the target attribute.

(vi) In contrast, our ASM weakens the disordered effect of irrelevant BGs, so it can be used in providing an excellent starting point for Dense CRF. Besides, since the parameters in our method are adaptively set according to the nature of each image, our approach is more responsive to the characteristics of images that results in better segmentation results. For some other complex images that do not have ground truth, our algorithm can also acquire promising segmentation results, while the other four algorithms failed in various degrees.

As you may discover, the qualitative effects of DenseCut on different datasets are not the same. This is because some parameters in DenseCut are set to fixed values according to the global statistical characteristics (including colour mean, variance etc.) of the dataset. For different datasets, DenseCut needs to manually adjust its parameter values to achieve an average better segmentation effect on the images of the entire dataset. Although the images taken from different datasets are similar, and the given FG and BG information are almost the same, the overall statistics of the three datasets are different. In this experiment, we used the default fixed parameter values adopted by the official source code of DenseCut (set the parameter value for MSRA datasets), so it led to such a difference. The global statistics of the first two datasets are relatively close, while the global information of the BSD500 dataset is slightly different from the previous two datasets. In other words, DenseCut contains constraints of global information dataset. If only the global statistics of a particular dataset are used as the criteria in setting parameter values, and the relevant parameter values are not adjusted according to different datasets, the fixed parameter values set under such constraints may cause the segmentation accuracy of some images in different datasets to be significantly reduced, which can be clearly seen from the comparison of the three figures (Figs. 5*f*, 6*f*, and 7*f*). If you count the global information of the BSD500 dataset before performing DenseCut, you may be able to improve the existing segmentation result. However, this kind of statistical work has the drawback of no circumvention: if the dataset is too small, it is difficult to accurately reflect the global characteristics of the dataset; if the dataset is too large, the statistical calculation is very time consuming and resource consuming. The deficiencies of DenseCut that depend on a particular dataset have been overcome in our approach. Since our method adaptively sets and fine-tunes the relevant parameters in the manner of image-by-image in the Dense CRF framework, we are able to make the best segmentation for every individual image (not dependent on global statistics features of a particular dataset). It is more versatile than DenseCut.

It can be seen from Figs. 5–7 that the segmentation result of ours is obviously better than the other five latest algorithms. However, in the experiment, there are some cases of segmentation failure of the algorithm, such as the image segmentation shown in Fig. 8. The FG target of the segmentation is a bee. The following factors make segmentation difficult: (i) the wings of the bee are translucent; (ii) the target edge is not obvious; (iii) bee's legs belong to the slender target; (iv) the colour of that bee's legs is very close to the BG. The ASM (Fig. 8*c*) is still not clear enough even if more seeds labelling information is added (Fig. 8*b*), which makes the segmentation result (Fig. 8*d*) still unable to meet our expectations. The segmentation results of the other five algorithms seem less accurate (Figs. 8*e–i*).

4.4 Time and speed analysis

Table 2 shows the details of each algorithm's runtime on the test dataset, which reflects the computational speed of these algorithms. More precisely, given the same segmentation task, less runtime represents higher speed, and vice versa. Our approach is slightly faster than OneCut (an acknowledged efficient interactive segmentation method) on the three datasets. Due to so many convex structures [14] in images and convexity energy optimisations by using iterative trust region algorithm [56, 57], TRC is the slowest among the six comparing methods. In order to

expand the graph, SubRW needs to build a priori model by using GMMs based on user scribbles. Unlike GMMs used in our method and DenseCut, GMMs in SubRW are required to utilise recursive expectation-maximisation algorithm to learn model parameters [58, 59], which reduces the efficiency of SubRW. In addition, to calculate the probability that a random walker reaches a staying node or a priori node, SubRW needs to solve the linear equations. Although the coefficient matrix of the linear equations is sparse, the solution is still not efficient enough. Therefore, the segmentation efficiency of SubRW has a relative low ranking in this experiment. GeoGC computes probability density effectively through fast Gaussian transformation [60], and calculates geodesic distance by using an efficient implementation of Dijkstra's algorithm [35], in addition, a very efficient graph-cut algorithm [17] is used only once, thus, GeoGC becomes the fastest segmentation method compared to the other five. Comparing to GeoGC, our algorithm is based on Dense CRF framework which is temporary slower than the graph-cut algorithm [17] in implementation.

DenseCut in strokes manner is slower than the one in bounding-box manner reported in [18], because of the two following facts: some of the FG to be segmented are only a small part of the whole image (so that only a small portion of the original image can be cut out and used as the image to be segmented, greatly reducing the number of pixels); some literature like GrabCut [16] indicates the final segmentation result can be refined if the marking area (used to indicate determinate BG) outside the bounding box contains only the pixels within the narrow band surrounding bounding box. However, it is also critical to point out that DenseCut in bounding-box-manner does not improve the segmentation efficiency intrinsically, it only reduces the number of vertices in the graph for reducing segmentation time.

Theoretically, our algorithm has precisely the same computational complexity as DenseCut in estimating colour model and in mean-field inference (linear in the number of pixels [7, 8]), and only adds the calculation of geodesic distance; however, the computation complexity of geodesic distance of our method is near linear [61, 62]. The GeoGC method uses fast kernel density estimation, which is very close to the computational complexity of the efficient GMM in DenseCut. Therefore, the computation speed of our algorithm and GeoGC algorithm should be theoretically similar. However, it is slower than DenseCut and GeoGC in practice, the reason is that there may be lack of acceleration in our code implementation.

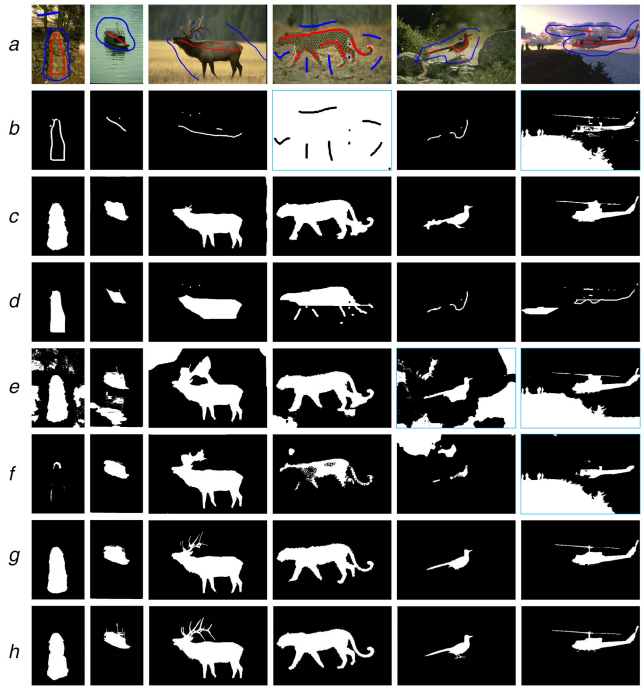

Fig. 7 *Examples of qualitative results on BSD500 datasets with the same user strokes. FG strokes are coloured red and BG strokes are blue, and each column is an example*

(a)–(g) Input image with FG/BG strokes, the output from OneCut, GeoGC, TRC, SubRW, DenseCut, and ours, respectively, ***(h)*** Ground truth (GT)

5 Conclusions and future works

We proposed a practical ASM for interactive segmentation, and got promising segmentation result after applying this model to the Dense CRF framework. Our approach gets rid of being kidnapped by the specific datasets because of owning good image-by-image adaptability that not only reduces the workload of user scribble interaction, but also eliminates the need for complex parametric learning processes. By using our AASM under the Dense CRF framework, precise segmentation results with mean filed inference in only once can be achieved, therefore, as with the OneCut algorithm, our approach is an efficient replacement of iterative optimisation techniques, such as GrabCut, TRC, and so on. In contrast to the colour separation used in OneCut, our adaptive appearance separation is less of a tendency to produce a large number of isolated points. The main limitation of our approach to get accurate segmentation results is that several interactions based on the previous results are required for some complex images. In the future, we plan to reduce the user input and extend the algorithm to multi-class segmentation. In addition, we intend to

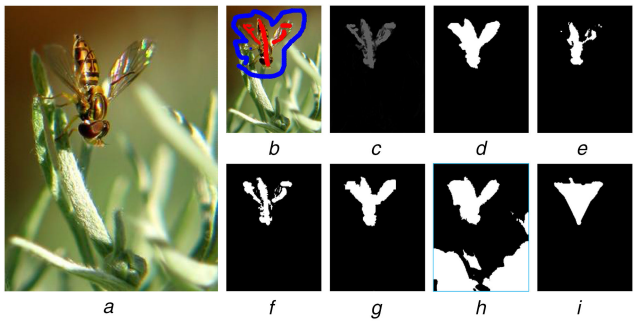

Fig. 8 *Failure segmentation example*

(a) Image to be segmented, ***(b)*** Image with FG and BG scribbles, ***(c)*** ASM map, ***(d)*** Segmentation result of our method, ***(e)–(i)*** Segmentation of DenseCut, OneCut, GeoGC, SubRW, and TRC algorithms, respectively, by using the same input (b) as ours

Table 2 Runtime for TRC, ours, DenseCut, OneCut, GeoGC, and SubRW

		TRC	Ours	DenseCut	OneCut	GeoGC	SubRW
MSRA	Avg.	10.243	0.525	0.345	0.719	0.257	2.145
	Max.	59.440	1.848	0.511	5.432	0.930	12.912
	Min.	0.541	0.120	0.122	0.228	0.104	0.385
GrabCut	Avg.	37.503	1.777	0.695	1.808	0.470	3.742
	Max.	182.695	4.670	0.859	8.050	0.880	5.851
	Min.	2.767	0.462	0.294	0.294	0.148	1.471
BSD500	Avg.	17.619	0.720	0.438	1.715	0.351	3.342
	Max.	288.037	0.908	0.542	5.065	0.502	9.448
	Min.	7.870	0.431	0.371	0.443	0.195	1.543

combine our method with deep learning methods to make better results.

6 Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC no. 11471002); Hunan Provincial Science and Technology Plan (no. 2013FJ4052); and the Construct Program of the Key Discipline in Hunan Province.

7 References

[1] Li, Y., Sun, J., Tang, C., *et al.*: 'Lazy snapping', *ACM Trans. Graph.*, 2004, **23**, (3), pp. 303–308

[2] Li, C., Xu, C., Gui, C., *et al.*: 'Level set evolution without re-initialization: A new variational formulation'. IEEE Conf. Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, pp. 430–436

[3] Boykov, Y., Jolly, M.: 'Interactive organ segmentation using graph cuts'. Medical Image Computing and Computer-Assisted Intervention, Third Int. Conf., Pittsburgh, Pennsylvania, USA, 2000 (LNCS, **1935**), pp. 276–286

[4] Boykov, Y., Jolly, M.: 'Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images'. IEEE Int. Conf. on Computer Vision, Vancouver, British Columbia, Canada, 2001, pp. 105–112

[5] Blake, A., Rother, C., Brown, M.A., *et al.*: 'Interactive image segmentation using an adaptive GMMRF model'. 8th European Conf. Computer Vision, Prague, Czech Republic, 2004, (LNCS, **3021**), pp. 428–441

[6] Ning, J., Zhang, L., Zhang, D., *et al.*: 'Interactive image segmentation by maximal similarity based region merging', *Pattern Recognit.*, 2010, **43**, (2), pp. 445–456

[7] Krähenbühl, P., Koltun, V.: 'Efficient inference in fully connected crfs with gaussian edge potentials'. Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems, Granada, Spain, 2011, pp. 109–117

[8] Krähenbühl, P., Koltun, V.: 'Parameter learning and convergent inference for dense random fields'. Proc. of the 30th Int. Conf. on Machine Learning, 2013. vol. 28 of JMLR Workshop and Conference Proceedings. JMLR.org, Atlanta, GA, USA, 2013, pp. 513–521

[9] Adams, A., Baek, J., Davis, M.A.: 'Fast high-dimensional filtering using the permutohedral lattice', *Comput. Graph. Forum*, 2010, **29**, (2), pp. 753–762

[10] Vineet, V., Warrell, J., Sturgess, P., *et al.*: 'Improved initialization and Gaussian mixture pairwise terms for dense random fields with mean-field inference'. British Machine Vision Conf., Surrey, UK, September, 2012, pp. 1–11

[11] Vineet, V., Warrell, J., Torr, P.H.S.: 'Filter-based mean-field inference for random fields with higher-order terms and product label-spaces', *Int. J. Comput. Vis.*, 2014, **110**, (3), pp. 290–307

[12] Desmaison, A., Bunel, R., Kohli, P., *et al.*: 'Efficient continuous relaxations for dense CRF'. European Conf. on Computer Vision, Amsterdam, The Netherlands, 2016, Proc., Part II (LNCS, **9906**), 2016, pp. 818–833

[13] Ajanthan, T., Desmaison, A., Bunel, R., *et al.*: 'Efficient linear programming for dense crfs'. IEEE Conf. on Computer Vision and Pattern Recognition, 2017. IEEE Computer Society, Honolulu, HI, USA, 2017, pp. 2934–2942

[14] Gorelick, L., Veksler, O., Boykov, Y., *et al.*: 'Convexity shape prior for binary segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, **39**, (2), pp. 258–271

[15] Tang, M., Gorelick, L., Veksler, O., *et al.*: 'Grabcut in one cut'. IEEE Int. Conf. Computer Vision, 2013. IEEE Computer Society, Sydney, Australia, 2013, pp. 1769–1776

[16] Rother, C., Kolmogorov, V., Blake, A.: ''Grabcut': interactive foreground extraction using iterated graph cuts', *ACM Trans. Graph.*, 2004, **23**, (3), pp. 309–314

[17] Boykov, Y., Kolmogorov, V.: 'An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision', *IEEE Trans. Pattern Anal Mach Intell.*, 2004, **26**, (9), pp. 1124–1137

[18] Cheng, M., Prisacariu, V.A., Zheng, S., *et al.*: 'Densecut: densely connected crfs for realtime grabcut', *Comput. Graph. Forum*, 2015, **34**, (7), pp. 193–201

[19] Shen, J., Du, Y., Wang, W., *et al.*: 'Lazy random walks for superpixel segmentation', *IEEE Trans. Image Process.*, 2014, **23**, (4), pp. 1451–1462

[20] Grady, L.: 'Random walks for image segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (11), pp. 1768–1783

[21] Dong, X., Shen, J., Shao, L., *et al.*: 'Sub-markov random walk for image segmentation', *IEEE Trans. Image Process.*, 2016, **25**, (2), pp. 516–527

[22] Couprie, C., Grady, L.J., Najman, L., *et al.*: 'Power watershed: a unifying graph-based optimization framework', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, (7), pp. 1384–1399

[23] Shi, J., Malik, J.: 'Normalized cuts and image segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (8), pp. 888–905

[24] Maji, S., Vishnoi, N.K., Malik, J.: 'Biased normalized cuts'. IEEE Conf. on Computer Vision and Pattern Recognition, 2011, IEEE Computer Society, Colorado Springs, CO, USA, 2011, pp. 2057–2064

[25] Rezvanifar, A., Khosravifard, M.: 'Including the size of regions in image segmentation by region-based graph', *IEEE Trans. Image Process.*, 2014, **23**, (2), pp. 635–644

[26] Cheng, Y.: 'Mean shift, mode seeking, and clustering', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1995, **17**, (8), pp. 790–799

[27] Comaniciu, D., Meer, P.: 'Mean shift: a robust approach toward feature space analysis', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (5), pp. 603–619

[28] Achanta, R., Shaji, A., Smith, K., *et al.*: 'SLIC superpixels compared to state-of-the-art superpixel methods', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (11), pp. 2274–2282

[29] Chew, S.E., Cahill, N.D.: 'Semi-supervised normalized cuts for image segmentation'. IEEE Int. Conf. on Computer Vision, 2015. IEEE Computer Society, Santiago, Chile, 2015, pp. 1716–1723

[30] Ghanem, B., Ahuja, N.: 'Dinkelbach NCUT: an efficient framework for solving normalized cuts problems with priors and convex constraints', *Int. J. Comput. Vis.*, 2010, **89**, (1), pp. 40–55

[31] Shen, J., Du, Y., Li, X.: 'Interactive segmentation using constrained laplacian optimization', *IEEE Trans. Circuits Syst. Video Techn.*, 2014, **24**, (7), pp. 1088–1100

[32] Criminisi, A., Sharp, T., Blake, A.: 'Geos: geodesic image segmentation'. European Conf. on Computer VisionProc., Part I., Marseille, France, 2008, (LNCS, **5302**), pp. 99–112

[33] Criminisi, A., Sharp, T., Rother, C., *et al.*: 'Geodesic image and video editing', *ACM Trans. Graph.*, 2010, **29**, (5), pp. 134:1–134:15

[34] Gulshan, V., Rother, C., Criminisi, A., *et al.*: 'Geodesic star convexity for interactive image segmentation'. IEEE Conf. Computer Vision and Pattern Recognition, 2010. IEEE Computer Society, San Francisco, CA, USA, 2010, pp. 3129–3136

[35] Price, B.L., Morse, B.S., Cohen, S.: 'Geodesic graph cut for interactive image segmentation'. IEEE Conf. on Computer Vision and Pattern Recognition, 2010. IEEE Computer Society, San Francisco, CA, USA, 2010, pp. 3161–3168

[36] Krähenbühl, P., Koltun, V.: 'Geodesic object proposals'. European Conf. on Computer Vision2014, Proc., Part V., Zurich, Switzerland (LNCS, **8693**), 2014, pp. 725–739

[37] Wang, W., Shen, J., Porikli, F.: 'Saliency-aware geodesic video object segmentation'. IEEE Conf. on Computer Vision and Pattern Recognition, 2015. IEEE Computer Society, Boston, MA, USA, 2015, pp. 3395–3402

[38] Feng, J., Price, B.L., Cohen, S., *et al.*: 'Interactive segmentation on RGBD images via cue selection'. IEEE Conf. on Computer Vision and Pattern Recognition, 2016. IEEE Computer Society, Las Vegas, NV, USA, 2016, pp. 156–164

[39] Wang, J., Yagi, Y.: 'Shape prior embedded geodesic distance transform for image segmentation'. Asian Conf. on Computer Vision, 2010, Part II., Queenstown, New Zealand, (LNCS, **6469**), 2010, pp. 72–81

[40] Li, C., Xu, C., Gui, C., *et al.*: 'Distance regularized level set evolution and its application to image segmentation', *IEEE Trans. Image Process.*, 2010, **19**, (12), pp. 3243–3254

[41] Yang, X., Gao, X., Tao, D., *et al.*: 'An efficient MRF embedded level set method for image segmentation', *IEEE Trans. Image Process.*, 2015, **24**, (1), pp. 9–21

[42] Khadidos, A., Sanchez, V., Li, C.: 'Weighted level set evolution based on local edge features for medical image segmentation', *IEEE Trans. Image Process.*, 2017, **26**, (4), pp. 1979–1991

[43] Cheng, M., Mitra, N.J., Huang, X., *et al.*: 'Global contrast based salient region detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, **37**, (3), pp. 569–582

[44] Ye, L., Liu, Z., Li, L., *et al.*: 'Salient object segmentation via effective integration of saliency and objectness', *IEEE Trans. Multimedia*, 2017, **19**, (8), pp. 1742–1756

[45] Wang, W., Shen, J., Yang, R., *et al.*: 'Saliency-aware video object segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2018, **40**, (1), pp. 20–33

[46] Syu, J., Wang, S., Wang, L.: 'Hierarchical image segmentation based on iterative contraction and merging', *IEEE Trans. Image Process.*, 2017, **26**, (5), pp. 2246–2260

[47] Dong, X., Shen, J., Shao, L., *et al.*: 'Interactive cosegmentation using global and local energy optimization', *IEEE Trans. Image Process.*, 2015, **24**, (11), pp. 3966–3977

[48] Han, J., Quan, R., Zhang, D., *et al.*: 'Robust object co-segmentation using background prior', *IEEE Trans. Image Process.*, 2018, **27**, (4), pp. 1639–1651

[49] Wang, W., Shen, J., Li, X., *et al.*: 'Robust video object cosegmentation', *IEEE Trans. Image Process.*, 2015, **24**, (10), pp. 3137–3148

[50] Bai, X., Sapiro, G.: 'A geodesic framework for fast interactive image and video segmentation and matting'. IEEE Int. Conf. on Computer Vision, 2007. IEEE Computer Society, Rio de Janeiro, Brazil, 2007, pp. 1–8

[51] Orchard, M.T., Bouman, C.A.: 'Color quantization of images', *IEEE Trans. Signal Process.*, 1991, **39**, (12), pp. 2677–2690

[52] Liu, T., Yuan, Z., Sun, J., *et al.*: 'Learning to detect a salient object', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, (2), pp. 353–367

[53] Arbelaez, P., Maire, M., Fowlkes, C.C., *et al.*: 'Contour detection and hierarchical image segmentation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, (5), pp. 898–916

[54] Dai, J., He, K., Sun, J.: 'Instance-aware semantic segmentation via multi-task network cascades'. IEEE Conf. on Computer Vision and Pattern Recognition, 2016. IEEE Computer Society, Las Vegas, NV, USA, 2016, pp. 3150–3158

[55] Achanta, R., Hemami, S.S., Estrada, F.J., *et al.*: 'Frequency-tuned salient region detection'. IEEE Conf. on Computer Vision and Pattern Recognition, IEEE Computer Society, Miami, Florida, USA, 2009, pp. 1597–1604

[56] Gorelick, L., Schmidt, F.R., Boykov, Y.: 'Fast trust region for segmentation'. IEEE Conf. on Computer Vision and Pattern Recognition, 2013. IEEE Computer Society, Portland, OR, USA, 2013, pp. 1714–1721

[57] Gorelick, L., Boykov, Y., Veksler, O., *et al.*: 'Submodularization for binary pairwise energies'. IEEE Conf. on Computer Vision and Pattern Recognition, 2014. IEEE Computer Society, Columbus, OH, USA, 2014, pp. 1154–1161

[58] Calinon, S.: '*Robot programming by demonstration - a probabilistic Approach*' (EPFL Press, 2009)

[59] Calinon, S., Guenter, F., Billard, A.: 'On learning, representing, and generalizing a task in a humanoid robot', *IEEE Trans. Syst., Man, Cybern., B*, 2007, **37**, (2), pp. 286–298
[60] Yang, C., Duraiswami, R., Gumerov, N.A., *et al.*: 'Improved fast gauss transform and efficient kernel density estimation'. IEEE Int. Conf. Computer Vision, 2003, IEEE Computer Society, Nice, France, 2003, pp. 464–471
[61] Toivanen, P.J.: 'Erratum to 'new geodesic distance transforms for gray-scale images', *Pattern Recognit. Lett.*, 1996, **17**, (13), pp. 437–450
[62] Yatziv, L., Bartesaghi, A., Sapiro, G.: 'O(N) implementation of the fast marching algorithm', *J. Comput. Phys.*, 2006, **212**, (2), pp. 393–399