International Journal of Pattern Recognition and Artificial IntelligencePrint ISSN: 0218-0014
Online ISSN: 1793-638

International Journal of Pattern Recognition and Artificial Intelligence
Print ISSN: 0218-0014 Online ISSN: 1793-638

ACCEPTANCE LETTER

Ref.: Ms. No. IJPRAI-D-17-00379R1
Image Cosegmentation using Shape Similarity and Object Discovery Scheme
International Journal of Pattern Recognition and Artificial Intelligence

Dear Ph.D Haiping Xu,

I am pleased to inform you that your work has now been accepted for publication in International Journal of Pattern Recognition and Artificial Intelligence.

You will be contacted by the publisher for the source files shortly before the date of publication.

Thank you for submitting your work to this journal.

With kind regards

Managing Editor
International Journal of Pattern Recognition and Artificial Intelligence

Comments from the Editors and Reviewers:

Accepted, Congratulations !

Reviewer #2: The authors have made a commendable effort to address most of reviewers' concerns within the available space. I am satisfied with the revised version of the paper. I recommend to publish in this journal..

Editors-in-Chief
X Jiang
Department of Computer Science
University of Münster
Einsteinstrasse 62 D-48149 Münste
http://www.worldscientific.com/worldscinet/ijprai

Image Cosegmentation using Shape Similarity and Object Discovery Scheme

Haiping Xu
Center for Discrete Mathematics and Theoretical Computer Science, Fuzhou University, Fuzhou, Fujian,
xuhaiping_fzu@163.com
Meiqing Wang[1]

mqwang@fzu.edu.cn
Fei Chen
College of Mathematics and Computer Science, Fuzhou Univ
chenfei314@fzu.edu.cn
Choi-Hong Lai

c.h.lai@gre.ac.uk

：Image cosegmentation is a newly emerging research area in image processing. It refers to the problem of segmenting the common objects simultaneously in multiple images by utilizing the similarity of foreground regions among these images. In this paper, a new active contour model is proposed by using shape-similarity and foreground discovery scheme. The foreground discovery scheme is used to obtain the rough contours of the common objects which are used as initial evolution curves. The energy function includes two parts: an intra-image energy and an inter-image energy. The intra-image energy explores the differences between foreground regions and background regions in each image. And the inter-image energy is used to explore the similarities of the common objects among target images, which composes of a region color feature energy term and a shape constraint energy term. The region color feature term indicates the foreground consistency and the background consistency among the images; and the shape constraint energy term allows the global changes of shapes and truncates the local variation caused by misleading features. Experimental results show that the proposed model can improve the accuracy of the image cosegmentation significantly through regularizing the changes of shapes.

Keywords： Image cosegmentation; shape similarity; foreground discovery scheme.

1. Introduction

Image segmentation is a fundamental problem in image processing, which is beneficial to many applications, e.g., image editing, object recognition and object retrieval. During the past decades, segmentation methods have been developed in a variety of directions, among which the active contour model (ACM) [1] is a very influential one. Generally, ACMs transform image segmentation into an energy minimization problem where the energy function describes image features and the unknown variables represent the contours of different objects. The ideal result is that the contour curve right stops on the true boundary of the object.

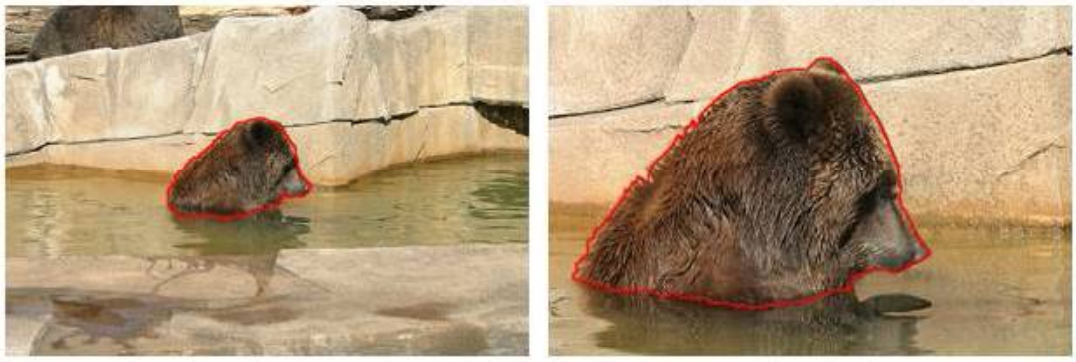
Although ACMs can achieve high precision in edge detection and keep the continuity of boundaries, they are prone to be corrupted by missing or misleading features. Furthermore, without

[1] Corresponding author.

clear definitions of subsequent applications, segmentation is not well defined since the definitions of complete objects vary according to their applications. For example, an image contains a bear and a rock, as illustrated in Fig. 1(a), which object do you want to segment? This problem is eased when a common object exists in multiple images. In this case the common object becomes prior information to guide segmentation.

Recently, extracting common objects (i.e. foreground objects) from multiple images has become an active research topic. A variety of methods have been proposed to solve multiple-image-based segmentation problems. Among them, a popular one is the image cosegmentation (IC) method. Given a group of images that contain a common object, the goal of IC is to simultaneously identify the common object in each image. This approach was firstly proposed by Rother et al. [2], in which a common object shared by two images is detected by measuring the similarity between their foreground histograms in L1-norm. Early IC methods [2-5] only use a pair of images as input under the assumption that they share a common foreground object. In recent years, numerous approaches [6-9] have been proposed to co-segment multiple images. All these IC methods have achieved more accurate result than traditional single-image-based segmentation methods. However, these IC methods do not perform well when foreground regions and background regions have overlapping color distribution. In this case, it is difficult to find the common object automatically.

Meng et al. [10] proposed an IC method in the ACM framework for images with similar background, as illustrated in Fig. 1. The method simultaneously considers the foreground consistency and the background consistency, and it can achieve right segmentation results for similar scenes cosegmentation. However, the performance of this method is not good when it is confused by misleading features. Meanwhile when multiple images come from complex scenes, the method is time-consuming.



(a) (b)

Fig. 1. An example of image cosegmentation. Bear (enclosed with red boundary) is the common foreground object between the two images.

We observe that images from similar scenes usually have similar topological structures. Based on this observation, this paper proposes a shape-similarity-based and foreground discovery scheme image cosegmentation model, which does not require knowing the shape models of the segmented objects in multiple images beforehand. Instead of using the time-consuming shape learning mechanism, a low-rank shape matrix strategy is used to regular changes of shapes. The foreground discovery scheme is used to find common objects and then rough shape curves of the objects are

extracted. Shape curves are stacked into a matrix and the shape similarity of the foreground objects is measured by the rank of the matrix.

As we all known, it is required to give an initial contour firstly to start the evolution of ACMs. Different initial contour usually leads to the different result. Manually initializing the contour for each image is time-consuming, tedious and impractical, especially in large scale image datasets. Inspired by the newly active research on saliency detection, we build a saliency-inspired way to discover the common foreground object to deal with this initialization problem, which can detect an initial curve around the true object boundary.

The main contributions of this paper are follows: (1) We propose a framework that introduces shape similarity and foreground discovery scheme for IC. (2) A saliency-inspired method is designed for discovering foreground objects. (3) The rank of the matrix consisting of multiple weighted shape curves is used to measure the shape similarity. Each curve is assigned a weight coefficient to balance the trade-off between the similarity and the correctness of the curve.

The remainder of this paper is organized as follows: Section 2 briefly reviews the related works. Section 3 is to detail our proposed model, and the minimization is described in Section 4. Section 5 conducts the extensive experiments and comparisons. Finally, some conclusions are presented in Section 6.

2. Related Works

The idea of IC was firstly proposed by Rother et al. [2] which exploited the histogram matching of the foreground in the Markov Random Field (MRF) framework. Due to its success in other computer vision applications, such as image classification and image retrieval, IC has been actively studied in recent years. The authors of [3-5] tried to use other measuring methods to compare the two foreground histograms for cosegmentation. Recent works [6-8, 11, 12] extended previous IC methods which use only a pair images by cosegmenting multiple images. The work of [9, 12, 13] also extended the foreground-background binary segmentation to multiple-regions segmentation being able to co-segment multiple images with multiple objects. The IC method for a large scale dataset was proposed in [14]. All the methods mentioned above are unsupervised cosegmentation methods. These methods may not perform well when there are overlapping color distribution between foreground regions and background regions. Interactive IC methods [15, 16] can avoid this problem by indicating the foreground objects with sparse scribbles. For example, Batra et al. [16] proposed an interactive cosegmentation method, which enable the user to adjust the inconsistent segmentation by adding scribbles. They proposed a recommendation method to help users choose the regions where the scribbles are necessary. All these methods use region features such as color histogram and SIFT to guide segmentation.

IC methods based on region color features have achieved great success in the past several years. However, the similarity of aforementioned region color features cannot provide powerful and precise auxiliary information. One potential problem of region color features is that it is easy to segment unnecessary objects when foreground and background regions have overlapping color distribution, so they may not provide strong prior information for segmentation. Therefore, it is necessary to add some other

constraints to assist the cosegmentation.

IC based on shape feature is less involved. The earliest research is the work of Kumar et al. [17], where the shape model should be learned beforehand. Cosketch [18] is an unsupervised learning framework. It performs well on images where similar shape patterns repeat themselves in the same image. However, sketch learning is a time-consuming process and the influence of similar background sketch cannot be ignored. Zhou et al. [19] used the similarity of object shapes to guide the cardiac segmentation for ultrasound images. They used the rank of the matrix formed by the shape curves to measure shape changes and avoid trapping into misleading features. Inspired by this idea, we introduce a weighted shape similarity into a cosegmentation model to improve the performance of segmentation. We will give more details of our model in the following section.

3. Proposed Framework

The keys of the proposed model are twofold: (1) Common foreground objects discovery; (2) Foreground similarity measurement. We use region saliency and region repeatability to discover the common foreground object. The foreground similarity measurement is composed of two parts: color similarity and shape similarity.

3.1 Problem Formulation

Given a set of N images $\{I_i\}_{i=1}^{N}$ for image cosegmentation, each image I_i is firstly segmented into a group of regions $R_i = \{r_{ij}\}_{j=1}^{m_i}$ by an over-segmentation method [20] for computational efficiency. A curve C_i divides the image I_i into the foreground region set ω_i^{in} and the background region set ω_i^{out}. In the context of the proposed ACM model, the problem of co-segmenting these N images is converted to find the optimal curves $\{C_i\}_{i=1}^{N}$ minimizing the following energy function:

$$E(\{C_i\}) = \sum_{i=1}^{N} E_{intra}(C_i) + \lambda E_{inter}(C_i). \quad (1)$$

Where E_{intra} is the conventional single image segmentation term used to describe the distinctions between the foreground set ω_i^{in} and the background set ω_i^{out} of each image, E_{inter} is the multiple image term used to measure the foreground consistency between the image I_i and the image I_p, and λ is the trade-off parameter. Fig. 2 gives a general schematic framework of the proposed model, which not only utilizes the global color information but also exploits the shape information.

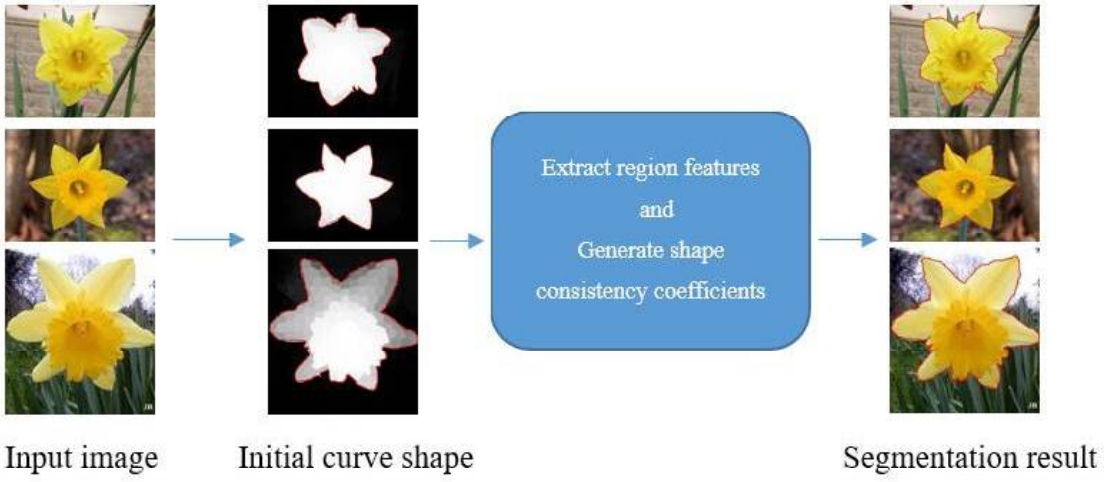


Fig. 2. Workflow of our proposed model.

In what follows, we shall firstly explain how the foreground object discovery is derived, and how to use the foreground information to guide the following segmentation process.

3.2 Foreground Object Discovery Scheme

The iteration method will be used to solve the minimization problem of Eq.1. To do this, initial positions of the curves $\{C_i\}_{i=1}^{N}$ will be set firstly. In this paper, the saliency detection is used to obtain the initial curves. Saliency detection is used to calculate a saliency map that highlights the salient foreground objects and suppresses the background. In this paper, we assume that the regions considered as salient areas have high probabilities of containing parts of the common objects. That is, the object regions are not only common but also salient ones among images. The saliency maps $\Phi_i = \{\phi_{ij}\}_{j=1}^{m_i}$ of the image I_i can be obtained by the cellular automate [21].

We use a high-dimensional feature vector to measure similarities between any two regions from different images. The color information of each region is represented by the HSV, and the texture of the region is represented by a bag of SIFT visual words. Let c_{ij} and t_{ij} be the HSV and SIFT feature vector of the region r_{ij}, respectively. For each region r_{ij} of image I_i, its most similar region in image I_p is decided by the distance

$$\mathrm{d}(r_{ij}, I_p) = \min_q(||c_{ij} - c_{pq}|| + \gamma||t_{ij} - t_{pq}||). \quad (2)$$

r_{ij} is now associated with $N-1$ distances $\{\mathrm{d}(r_{ij}, I_p)\}_{p\neq i}$. Then we choose the first half smallest distances to compute the average distance d_{ij}. The sigmoid function [11] is used to define the region repeatability

$$\varphi_{ij} = \frac{1}{1 + \exp(-\frac{\mu - d_{ij}}{\sigma})}. \quad (3)$$

Where μ and σ are the parameters adjusting the shape of the sigmoid function ($\mu = 0.1$ and $\sigma = 0.2$ in all our experiments based on the empirical study).

According to the above region saliency prior and region repeatability measurement, the likelihood of a region is defined by

$$\psi_{ij} = \phi_{ij} \times \varphi_{ij}. (4)$$

It is not easy to use thresholding methods, i.e., Otsu's method [22], to choose a proper threshold for all saliency images to get initial foreground objects from background regions. Additionally, it needs a way to obtain a connected contour curve. Considering that there is no a sharp contrast between the saliency values from foreground objects and background regions, we use the CV [23] model to get the binary segmentation. Thus, the resultant curve is our initial curve shape set $S_0 = \{C_i\}_{i=1}^{N}$.

3.3 Image Feature Extraction

In this paper, we use the HSV color histogram to describe color feature. The HSV has a uniform color distribution which not only reflects the human perception of color but also brings benefits to uniform sampling. Moreover, the HSV is appropriate when illumination condition changes significantly. For any region r_{ij} to be evaluated, we use Hellinger [25] distance to define the color similarity of the region r_{ij} and the foreground region set ω_i^{in} as follows:

$$f(r_{ij}, \omega_i^{in}) = \sqrt{1 - \sum_{k=1}^{c} \sqrt{h_{rk} \cdot h_{\omega k}}} \ (5).$$

Where c is the number of bins, h_{rk} and $h_{\omega k}$ are the normalized color histograms of r_{ij} and the foreground region set ω_i^{in}, respectively. A smaller value means a higher similarity between them.

The second feature used in our paper is related to shape. For each curve C_i, it could be represented by a series of landmark points (x_{ij}, y_{ij}) on its curve, that is, the curve can be represented as a vector: $C_i = [x_{i1}, \dots, x_{ik}, y_{i1}, \dots, y_{ik}]^T$. Assuming that C_i is generated from C_1 through an affine transformation, then C_1 and C_2 are two identical shapes and would satisfy the following condition:

$$C_2 = \Phi_1 T_1. (6)$$

Where

$$\Phi_1 = \begin{bmatrix} C_1^x & 0 & C_1^y & 0 & 1 & 0 \\ 0 & C_1^x & 0 & C_1^y & 0 & 1 \end{bmatrix}. (7)$$

And T_1 is an affine transformation matrix defined by $T_1 = [m_{11}, m_{21}, m_{12}, m_{22}, t_1, t_2]$, which describes the shape changes, such as rotation, translation and scaling. Therefore, a shape matrix which consists of N identical shapes also will be satisfied the following condition:

$$S = [C_1, \dots, C_N] = \Phi_1 [T_1, \dots, T_N]. (8)$$

Since $\text{rank}(\Phi_1) \leq 6$, the shape matrix also satisfies $\text{rank}(S) \leq 6$. The more difference between a curve shape and other curve shapes are, the harder for the curve shape is represented by others. Therefore, $\text{rank}(S)$ will increase when some shapes

change. The rank of shape matrix can be used to represent the difference between different shapes in multiple images.

However, when some inconsistent shapes (caused by the foreground object discovery or other possible conditions) are significantly different from the latent common shapes of the image set, the curves are prone to be affected to converge to false boundaries during the evolution. Although we can increase penalty to reduce the error, most specific details of shapes would be removed and inaccurate segmentation results would be obtained. Thus, it is unreasonable to directly constrain all curves by a low rank matrix. The consistency of one curve shape to the common curve shapes need to be considered. In this paper, we use the similarities between any two curves to measure the consistency. Hu moments [24] are invariant to image changes, such as rotation and scaling. Thus it is suitable for describing the shape contexts of foreground objects. The similarity between two curves is defined by

$$d_{ip}(C_i, C_p) = \max_{t=1,\dots,7} \frac{\left|sign(h_t^i) \cdot logh_t^i - sign(h_t^p) \cdot logh_t^p\right|}{\left|sign(h_t^i) \cdot logh_t^i\right|}. (9)$$

Where h_t^i and h_t^p are Hu moments of C_i and C_p, respectively. For each curve C_i, we use v_i to represent the average of the distances from C_i to other curves as follows:

$$v_i = \frac{1}{N}\sum_{p=1}^{N} d_{ip}(C_i, C_p). (10)$$

Since smaller d_{ip} indicates higher similarity between two curves, we choose the top K smallest values and then rewrite v_i as

$$v_i = \frac{1}{K}\sum_{p=1}^{K} d_{ip}(C_i, C_p). (11)$$

Where K is a constant, which is set to $K = \left\lceil \frac{N}{2} \right\rceil$.

We use logistic function to compute the consistency coefficient of each curve as follows

$$\omega_i = \frac{1-\tau}{1+e^{-\sigma(v_i-0.5)}} + \tau. (12)$$

Both τ and σ are constant.

For each curve obtained by foreground object discovery scheme, it would be represented by a weighted shape matrix as follow:

$$S = w \cdot C = [\omega_1 C_1, \dots, \omega_N C_N]. (13)$$

3.4 Energy Function Design

As mentioned in Eq.1, the energy function in the proposed model is composed of two parts: the intra-image energy term and the inter-image energy term.

3.4.1 Intra-image Energy

The intra-image energy is used to measure the consistency within a single image, which ensures the smoothness of the object boundaries and keeps the distinction between the foreground and the background in each image. The intra-image energy consists of three components: the regularization term, foreground consistency term and background consistency term, which is expressed as

$$E_{intra}(C_i) = \lambda_{sc}E_i^{sc}(C_i) + \lambda_{sf}E_i^{sf} + \lambda_{sb}E_i^{sb}. (14)$$

Regularizing term E_i^s is used to describe the attributes of the curve C_i:

$$E_i^{sc}(C_i) = \mu \cdot Area(\omega_i^{in}) + \nu \cdot Length(C_i). (15)$$

Where $Area(\omega_i^{in})$ refers to the area enclosed by the curve, $Length(C_i)$ refers to the curve length, μ and ν represent the weights.

The foreground consistency term E_i^{sc} is used to measure similarities of the foreground region of a single image, which is defined as the sum of the color similarities between the region r_{ij} in ω_i^{in} and the interior region set ω_i^{in} itself, which is expressed as

$$E_i^{sf} = \int_{r_{ij}\in\omega_i^{in}} f(r_{ij}, \omega_i^{in})dr. (16)$$

Small value means high consistency in the interior region set ω_i^{in}.

The background consistency term E_i^{sb} is used to measure similarities of the background region of a single image. Similar to the foreground consistency energy, the energy of background consistency is defined as the sum of the color similarities between the region r_{ij} in ω_i^{out} and the exterior region set ω_i^{out} itself, which is expressed as

$$E_i^{sb} = \int_{r_{ij}\in\omega_i^{out}} f(r_{ij}, \omega_i^{out})dr. (17)$$

Similar to the foreground consistency, small values means high consistency in the exterior region ω_i^{out}.

3.4.2 Inter-image energy

The inter-image energy term is designed to measure the foreground consistencies among multiple images, which is described by an inter-image color energy term and an inter-image shape energy term:

$$E_{inter}(C_i) = \lambda_s E_{inter-shape} + E_{inter-color}. (18)$$

The inter-image color energy term is designed to measure the foreground consistencies and background consistencies of multiple images. The foreground consistency of multiple images measures similarities of foreground regions among images. For a given image I_i, the energy E_i^f is defined as the sum of the color similarities between the interior region set ω_i^{in} of the $i\text{-}th$ image and the interior region set $\omega_p^{in}, p = 1,2,\dots,N, P \neq i$ of the rest of the images, which is expressed as

$$E_i^f = \sum_{p=1,p\neq i}^{N} \int_{r_{ij}\in\omega_i^{in}} f(r_{ij},\omega_p^{in})\mathrm{dr}. (19)$$

E_i^f will have a small value when it is consistent to most of the foreground regions of other images, otherwise, a large value will be obtained.

Similarly, the background consistency can be expressed as

$$E_i^b = \sum_{p=1,p\neq i}^{N} \int_{r_{ij}\in\omega_i^{out}} f(r_{ij},\omega_p^{out})\mathrm{dr}. (20)$$

Thus, the energy functional of inter-image color feature is

$$E_{inter-color}(C_i) = \lambda_f E_i^f + \lambda_b E_i^b. (21)$$

When the backgrounds of images are not similar, λ_b will be set as zero for computation efficiency.

In real applications, the performance of ACMs sometimes is corrupted due to missing or misleading features. Shape prior as a kind of auxiliary information has been introduced to improve the robustness of cosegmentation. The authors of [19] use rank of shape matrix to measure the correlation among the shapes, and describe the degree of freedom of the shape change. The low-rank constraint will allow the global change of curves such as rotation, scaling, translation and principal deformation to fit the image data and truncate the local variation caused by image misleading feature. Since rank is a discrete operator, nuclear norm replacing the rank operator is used to describe the shape change. Following the idea of [19], the inter-image shape energy term $E_{inter-shape}$ is defined as

$$E_{inter-shape} = \|w \cdot C\|_*. (22)$$

In the proposed model, E_{intra} and $E_{inter-color}$ are focused on measuring the region features, and $E_{inter-shape}$ constrains the similarity of shapes. The similarity shape means the outline of foreground always keep similar no matter how the poses and viewpoints change. Spired by [19], we firstly evolve the curve according to the region feature and then impose the shape similarity regularization via singular value thresholding to aid the segmentation.

4. Iterative optimization algorithm

The intra-image energy term and inter-image region energy term are differentiable with Lipschitza continuous gradient, thus we use the Proximal Gradient method [19] to minimize Eq.1. Our energy function belongs to the following optimization category

$$\min_C F(w \cdot C) + \alpha R(w \cdot C). (23)$$

Where $R(w \cdot C) = \|w \cdot C\|_*$. The differentiable function $F(\cdot)$ is given by

$$F(w \cdot C) = \sum_{i=1}^{N} \lambda_1 E_{intra}(\omega_i C_i) + \lambda_2 E_{inter-color}(\omega_i C_i)$$

$$= \sum_{i=1}^{N} E_i(\omega_i C_i) = \sum_{i=1}^{N} \omega_i E_i(C_i) = w \cdot F(C). (24)$$

PG method produces a sequence of following iteration $C^{k+1}, k = 0,1,2,...,$ that converges to the optimum C^*,

$$w \cdot C^{k+1} = arg\min_{C} \frac{1}{2} \left\| w \cdot C - \left[w \cdot C^k - \frac{1}{\beta} \nabla F(w \cdot C^k) \right] \right\|_F^2 + \frac{\alpha}{\beta} R(w \cdot C). (25)$$

Where $\|\cdot\|_F$ denotes the Frobenius norm, α and β are constants. $\nabla F(w \cdot C^k) = [\omega_1 \nabla E_1(C_1^k), ..., \omega_N \nabla E_N(C_N^k)]$. According to [26], the Eq.25 could be solved by using singular value thresholding algorithm. Hence, we obtain a new iterative procedure as follow:

$$C^{k+1} = w^{-1} \cdot D_{\frac{\alpha}{\beta}} \left(w \cdot C^k - \frac{1}{\beta} \nabla F(w \cdot C^k) \right). (26)$$

Where $D_{\frac{\alpha}{\beta}}$ is a singular value thresholding operator [26]. After several iterations, more accurate and stable results are obtained. The pseudo-code of the proposed algorithm is summarized in Algorithm1.

Algorithm1 Shape Similarity and foreground discovery scheme based image Cosegmentation

Input: Image set $\mathcal{I} = \{I_1, I_2, ..., I_N\}$.

Output: The cosegmentation results $\mathcal{C} = \{C_1, C_2, ..., C_N\}$.

1: For $\mathcal{I} = \{I_1, I_2, ..., I_N\}$, do foreground objects discovery by Eq. 4 to get initial curve shape set $S_0 = \{C_i^0\}_{i=1}^N$.

2: For S_0, calculate consistency coefficient by Eq. 12 to get weighted curve shape set $S = w \cdot C = [\omega_1 C_1, ..., \omega_N C_N]$.

3: For $\mathcal{I} = \{I_1, I_2, ..., I_N\}$, calculate region similarity measurement by Eq. 5.

4: Initialize $C^0 = S$.

5: for k = 0:n do

6: $C^{k+1} = w^{-1} \cdot D_{\frac{\alpha}{\beta}} \left(w \cdot C^k - \frac{1}{\beta} \nabla F(w \cdot C^k) \right)$.

7: if $\left\| C^{k+1} - C^k \right\|_F < tolerance\ then$

8: return

9: end if

10: end for

11: Obtain the cosegmentation results of $\mathcal{C} = \{C_1, C_2, ..., C_N\}$.

5. Experiments

In this section, some experiments are designed to test the performance of the proposed model on two commonly used datasets: iCoseg dataset [16] and MSRC dataset [27]. Both iCoseg dataset and MSRC dataset include the ground truth segmentation masks

that are used for quantitative evaluation. The parameters of our algorithm are set up by lots of experiments. Considering the relationship between each item in Eq. 1, we set $\lambda = 1, \lambda_{sc} = \lambda_{sf} = \lambda_{sb} = 1, \lambda_b = 3, \lambda_f \in [4,7], \alpha = \lambda_s = 45, \beta \in [3,5]$. For Eq. 2, we set $\gamma = 3$. In Eq. 11, we set $\sigma = 10$ and $\tau = 0.95$. According to [23], we set $\mu = 0.01, \upsilon = 0.001$ in Eq. 15. All the experiments were implemented in MATLAB R2012a on a Win XP system.

5.1 Analysis of inter-image energy

In this section, we demonstrate the effectiveness of the inter-image energy by running out the algorithm on the sheep image group (from MSRC) with and without the inter-image energy in our total energy function. There is a common sheep in the middle of a single image. The background region has some soils with the same color distribution as the sheep. When the algorithm without inter-image shape constraint term only includes the inter-image region color energy term, and it resembles the SBCS model [10], which does not consider single image foreground consistency. When the algorithm without inter-image region color energy term only includes the inter-image shape constraint and single image energy terms, and it is similar to GS model [19], which draws a circle at the center of the image as the initial curve. As shown in Fig. 3, the SBCS model and the GS model cannot perform well due to these soils.

The SBCS model can yield good segmentation results for similar scene cosegmentation, which combines the multiple image foreground consistency constraint with the background consistency constraint to form the energy function. However, when the foreground or background has misleading feature, SBCS model may still fail. For example, there are some soils in the background region (meadow). These soils influence the segmentation. In this case SBCS cannot get the right segmentation results. Region color feature is a global feature. Color histogram describes the global distribution of colors in the image. It is very easy to segment unnecessary objects when the original image contains multiple object classes and these object classes have overlapping color distribution. So region color feature is not enough to provide strong prior information for segmentation. GS uses the similarity of object shapes to assist the segmentation. From the experiment results one can find that the results obtained with the shape similarity constraint are more robust than the results of SBCS model. The shape constraint regularizes the segmentation so as to keep the shape similar in image set. GS only focuses on intra-image feature and inter-image shape consistency. Hence, it cannot perform well for complex scene. One can find that the curve cannot evolve to the boundary of object. Although the curve keep consistency in the image set, there is no enough external force to push them to the boundary. This is due to the lacking of the inter-image region color feature. As shown in Fig. 3, our proposed model can segment the common sheep and avoid the influence of soils.

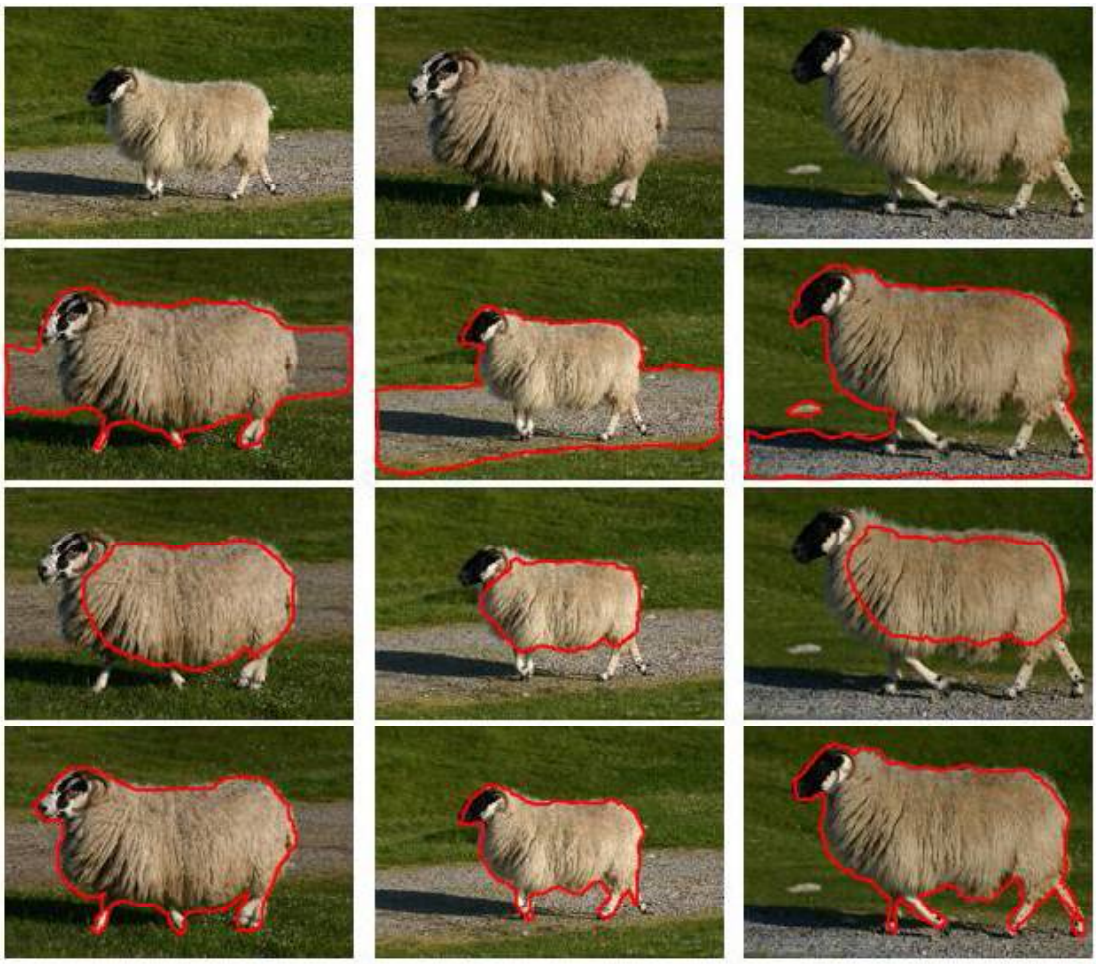


Fig. 3. A sheep example to illustrate the inter-image energy function. Row 1: the original images. Row 2: the segmentation results without the inter-image shape constrain term. Row 3: the segmentation results without the inter-image region color energy term. Row 4: the segmentation results by our model.

5.2 Cosegmentation on MSRC and iCoseg Dataset

We firstly show segmentation results on the iCoseg dataset and MSRC dataset. Example results include eight datasets where five datasets (goose, hot balloon, panda, bear and pyramid) are from iCoseg dataset shown in Fig. 4, and the other threes datasets (cat, cow and sheep) are from MSRC dataset shown in Fig. 5. We select four representative images from each class to illustrate the performance of our proposed algorithm. Some background regions have similar color distribution to the foreground object, such as the rock region in the fourth row of Fig. 4 and the ground region in the fifth row of Fig. 4. From the experiment results one can see that foregrounds (goose, hot balloon, panda, bear, pyramid, cat, cow and sheep) are correctly segmented from the backgrounds by the proposed model, and the shapes are globally consistent with each other throughout the dataset due to the shape constraint. Hence our model is more resistant to local misleading features.

Fig . The cosegmentation result by the proposed model on iCoseg dataset. Row 1: the results of goose dataset. Row 2: the segmentation results of hot balloon dataset. Row 3: the segmentation results of panda dataset. Row 4: the segmentation result of bear dataset. Row 5: the segmentation results of pyramid dataset.

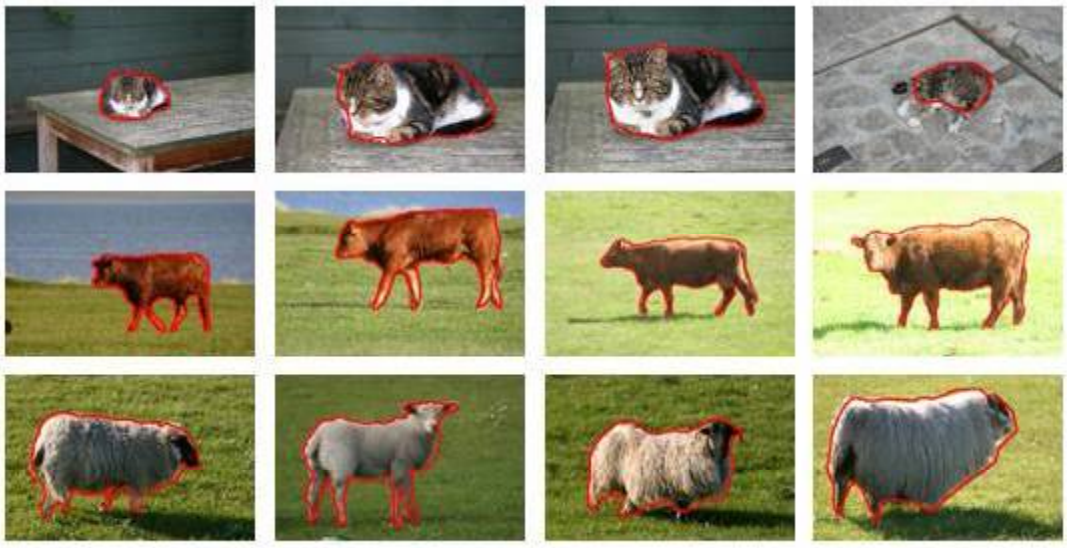
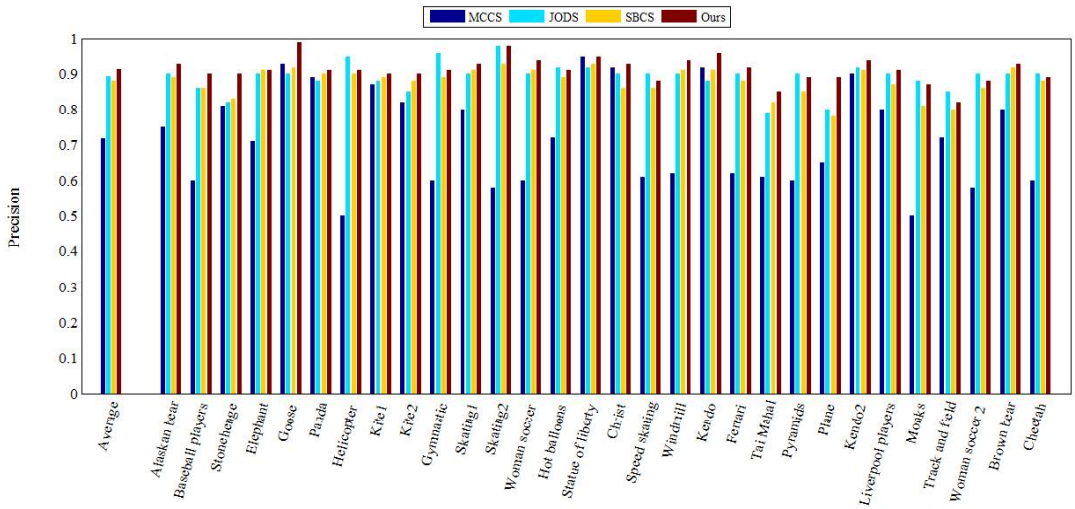


Fig. 5. The cosegmentation result by the proposed model on MSRC dataset. Row 1: the results of cat dataset. Row 2: the segmentation results of cow dataset. Row 3: the segmentation results of sheep dataset.

5.3 Comparison

Next, we quantitatively evaluate the cosegmentation results on both iCoseg and MSRC dataset to examine the performance of the proposed model. Two metrics (Precision P and Jaccard Similarity J) are used to do quantitative evaluation. P denotes the ration of correctly classified labeled pixels (both foreground and background) to the total number of pixels, and J represents the intersection over the segmentation results and ground truth masks.

Fig. 6 shows the per-class precision on iCoseg (top) and MSRC (bottom). Table 1 shows the average precision and Jacarrd similarity on iCoseg and MSRC. For the sake of fair comparison, we use all the images to avoid bias, and compare with multi-class cosegmentation [9], similar background cosegmentation (SBCS) [10] and joint object discovery and segmentation (JODS) [28]. Experimental results of [28] are produces by running the implementation codes from their websites. According to Fig. 6 and Table 1, our algorithm has better performance than other methods. Benefiting from successfully integrating the region color information and shape constraint, the average precision of our method is higher than the average precision of other methods.



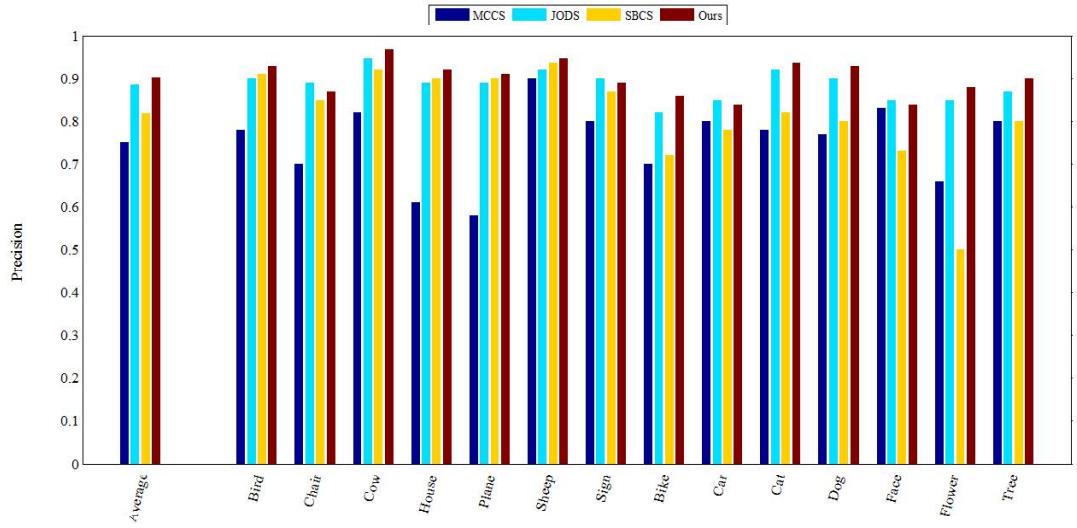


Fig. 6. Comparison results on iCoseg (top) and MSRC (bottom) between our method and other method: MCCS [9], JODS [28] and SBCS [10]. The ratios of correctly labeled pixels are measured, and each plot shows the average per-class precision.

Table 1 Average Precision ($\bar{P}$) and Jaccard Similarity ($\bar{J}$)

Method	iCoseg		MSRC	
	$\bar{P}$	$\bar{J}$	$\bar{P}$	$\bar{J}$
MCCS [9]	69.6	44.7	74.6	49.7
JODS [28]	88.7	69.8	88.1	71.6
SBCS [10]	86.5	67.9	87.8	79.9
Ours	90.8	90.3	90.4	88.5

5.4 Computational complexity

The experiment results listed in Table 2 demonstrates that the proposed model with the shape constraint term converged faster than that without the shape constraint term. Due to the shape constraint term, the proposed model has faster convergence and fewer iterations. In our model, for an image set with N images, it at most performs singular value decomposition for $2\mathrm{p} \times N$, where p is the number of landmarks. The SVD is not very time-consuming.

Table 2 Image groups including the number of images, the rank of output, the number of iterations for the algorithm to stop, and the computation time. The value in parentheses corresponds to the results without the shape constraint.

datasets	#image	rank	#iter	Time(min)
Goose	12	3	20(50)	3.1(7.0)
Bear	9	4	35(97)	6.0(10.5)

Hot Balloon	24	3	18(40)	7.0(15.0)
Panda	10	3	45(100)	6.1(9.7)
Pyramid	8	2	20(44)	6.0(9.0)
Cat	12	4	28(56)	5.2(11.0)
Cow	13	4	32(48)	6.2(11.2)
Sheep	16	5	25(59)	6.3(11.5)

6. Conclusion

In this paper, we propose a shape-similarity based and foreground object discovery scheme ACM for solving the cosegmentation problem. The energy function describes region color and shape features, which consists of an intra-image energy term, an inter-image region color term and an inter-shape constraint energy term. The shape constraint is used to regularize the segmentation. Object discovery scheme can locate initial curve to start the evolution. Experiment results demonstrate that the proposed model is more robust than the other models in terms of accuracy and computation efficiency, and demonstrates the significance of shape constraints. Although our method performs well on the benchmark datasets in general, there are some problems: (1) The proposed model is not able to address multi-class foregrounds. The rank of shapes matrix just simply describes the object shape change, such as scaling and rotation, and cannot achieve topological structure change, such as merging and splitting. (2) It is very tedious to adjust the value of parameters to get good segmentation result. Different image group has different parameters. In the future, we plan to explore more superior shape constraint to make the propose model flexible to image feature.

Acknowledgements

This work was supported partially by the National Natural Science Foundations of China (61401098, 61771141), and Natural Science Foundation of Fujian (China) under Grant No. 2015J01013.

References

[1] M. Kass, A. Witkin, D. Terzopoulos, Snake: active contour models, Int. J. Comput. Vis. 1, Jan. 1987, pp.321-331.

[2] C. Rother, T. Minka, A. Blake, and V. Kolmogorov, Cosegmentation of image pairs by histogram matching incorporating a global constraint into mrfs, In Proc. IEEE CVPR, Jun. 2006, pp.993-1000.

[3] H. Zhu, F. Meng, J. Cai, and S. Lu, Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation, Journal of Visual Communication and Image Representation, vol. 34, pp. 12-27, 2016.

[4] D. S. Hochbaum and V. Singh, An efficient algorithm for cosegmentation, In Proc. IEEE ICCV, Sep/Oct. 2009, pp.269-276.

[5] S. Vicente, V. Kolmogorov, and C. Rother, Cosegmentation revisited: Models and optimization, In Proc. ECCV, 2010, pp.465-479.

[6] A. Joulin, F. Bach, and J. Ponce, Discriminative clustering for image cosegmentation, In Proc. IEEE CVPR, Jun. 2010 pp. 1943-1950.

[7] S. Vicente, C. Rother, and V. Kolmogorov, Object cosegmentation, In Proc. IEEE CVPR, Jun.2011, pp. 2217-2224.

[8] L. Mukherjee, V. Singh, and J. Peng, Scale invariant cosegmentation for image groups, In Proc. IEEE CVPR, Jun.2011, pp.1881-1888.

[9] A. Joulin, F. Bach, and J. Ponce, Multi-class cosegmentation, In Proc IEEE CVPR, Jun. 2012, pp.542-549.

[10] F. Meng, H. Li, K. Ngan, B. Zeng, N. Rao, Cosegmentation from similar backgrounds, ISCAS, 2014, pp.353-356.

[11] K.-Y. Chang, T.-L. Liu, and S.-H. Lai, From co-saliency to cosegmentation: An efficient and fully unsupervised energy minimization model, In Proc. IEEE CVPR, 2011.

[12] J. C. Rubio, J. Serrat, A. L'opez, and N. Paragios, Unsupervised cosegmentation through region matching, In Proc. IEEE CVPR, 2012.

[13] L. Mukherjee, V. Singh, J. Xu, and M. D. Collins, Analyzing the subspace structure of related images: concurrent segmentation of image sets, In ECCV, 2012.

[14] E. Kim, H. Li, and X. Huang, A hierarchical image clustering cosegmentation framework, In Proc. IEEE CVPR, 2012.

[15] W. Wang and J. Shen, Higher-order image co-segmentation, IEEE Trans. Multimedia, vol. 18, no. 6, pp. 1011–1021, Jun. 2016.

[16] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, iCoseg: Interactive cosegmentation with intelligent scribble guidance, in Proc. IEEE CVPR, Jun. 2010, pp. 3169–3176.

[17] M. P. Kumar, P. H. S. Ton, and A. Zisserman, OBJ CUT, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 1. San Diego, CA, USA, Jun. 2005, pp. 18-25.

[18] J. Dai, Y. N. Wu, J. Zhou, and S.-C. Zhu, Cosegmentation and cosketch by unsupervised learning, in Proc. Int. Conf. Comput. Vis., 2013, pp. 1305-1312.

[19] X. Zhou, X. Huang, J. S. Duncan, and W. Yu, Active contours with group similarity, In Proc. IEEE CVPR, 2013, pp.2969-2976.

[20] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, Slic superpixels. Technical report, 2010.

[21] Y. Qin, H. Lu , Y. Xu, H. Wang, Saliency Detection via Cellular Automata, In Proc. IEEE CVPR, 2015, pp.110-119.

[22] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybem. 9(1), 1979, pp.62-66.

[23] T. Chan, L. Vese, Active contours without edges, IEEE Trans. Image Process. 10, Feb. 2001, pp. 266–277.

[24] R. Sivaramakrishna and N. Shashidhar, Hu's moment invariants: how invariant are they under skew and perspective transformations? in IEEE Pronc. 97: Commun, Power and Comput. IEEE, 1997, pp. 292-295.

[25] A. Bhattacharyya, On a measure of divergence between two multinomial populations, Sankhya: Ind. J. Stat. 1946, pp.401-406.

[26] J. Cai, E. Candes, and Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim, vol. 20, no.4, 2010, pp. 1956-1982.

[27] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Texton-boost: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In ECCV, 2006.

[28] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu, Unsupervised joint object discovery and

segmentation in internet images, in Proc. IEEE CVPR, (2013) 1939-1946.

Haiping Xu received her B.S. degree in Information and Computing Science from the College of Mathematics and Computer Science, Fuzhou University. She is currently a Ph.D. student at Center for Discrete Mathematics and Theoretical Computer Science, Fuzhou University. Her research interests include computer vision, image processing and partial differential equations.

Prof. Wang received her PhD in June, 2002 and have been appointed as Professor of Applied Math ematics in the College of Mathematics and Computer Science, Fuzhou University, China, since Se ptember, 2006. From 1997, her research interests have been focused on image and video compress ion using fractal theory, image processing methods based on PDE, close-
range photogrammetry, the parallel implementation of image processing algorithms in distributed or parallel computing environments, data analysis and processing, and computational finance.

Fei Chen received the Ph.D. degree in signal and information processing from Zhejiang University, Hangzhou, China, in 2013. He is currently an Associate Professor with the College of Mathematics

and Computer Science, Fuzhou University. His current research interests include machine learning, partial differential equations, and shape-driven techniques in image processing.

Choi-Hong Lai received the Ph.D. degree in Aerodynamics and Parallel Computing from Department of Aeronautical Engineering, Queen Mary, University of London, in 1985. He is the head of the Numerical and Applied Mathematics Research Unit and the programmer leader and year tutor for MSc Applicable Mathematics. Currently he is visiting professor at Jiangnan University and Fuzhou University, China, and at Buckingham University. His current research interests include numerical algorithms and applied mathematics.