2006－CVIM－155 (4)
2006／9／8

半教師付き物体抽出による自然画像マット合成

杜 偉薇† 浦浜 喜一†

† 九州大学 大学院芸術工学研究院 視覚情報部門
福岡市南区塩原 4-9-1

E-mail: †duweiwei@gsd.design.kyushu-u.ac.jp, ††urahama@design.kyushu-u.ac.jp

あらまし 自然画像から物体を切り出して他の画像に貼り付ける画像マット合成では，物体抽出を補助する情報として前景 (物体) と背景の 1 部を指示入力することがユーザに求められる．本論文では，ユーザの入力作業を軽減するために，類似度データからファジィクラスタを抽出する半教師付き手法を応用して，線形判別分析で変換した画素間カラー類似度に従ってメンバシップをユーザ入力領域から残りの領域へ伝搬して画像から物体を切り出す方法を提案する．本手法では，ユーザは物体領域か背景領域のどちらか一方だけに少しの指示入力を粗く塗るだけでよい．本方法は，従来のクラスタ分離タイプではなく，クラスタ抽出タイプであり、メンバシップ伝搬のウインドウも広いので物体や背景中の穴も飛び越えられ，少ない補助入力ですむ． 得られたメンバシップに基づいて各画素での前景の色を求めて物体を切り出し，他の画像を新たな背景としてマット合成する．

キーワード 自然画像マット合成，半教師付き物体抽出，メンバシップ伝搬，線形判別分析

Natural Image Matting with Semi-Supervised Object Extraction

Weiwei DU† and Kiichi URAHAMA†

† Faculty of Design, Kyushu University
Fukuoka-shi, 815-8540 Japan

E-mail: †duweiwei@gsd.design.kyushu-u.ac.jp, ††urahama@design.kyushu-u.ac.jp

Abstract In natural image matting where an object is extracted from a photograph with natural background and is composited with another image, a user is required to input strokes specifying both of foreground and background regions for supporting the extraction of the target object. In order to reduce user's inputs, we apply, in this paper, a semi-supervised algorithm for extracting a fuzzy cluster from similarity data to this task, and present a method for extracting an object from a natural image by propagating memberships from seeded pixels to the remaining area according to the similarity between pixel colors transformed with the Fisher's linear discriminant analysis. In our method, rough and coarse strokes are sufficient for users to draw them only either in foreground regions or in background areas. The color of the extracted object is estimated at each pixel based on the membership obtained with this algorithm and the extracted object is composited with another image as a new background.

Key words natural image matting, semi-supervised object extraction, membership propagation, linear discriminant analysis

1. Introduction

Natural image matting, instead of the blue screen matting, refers to a process of extracting an object from a natural image for compositing the extracted object with another image [1]. We can view this task at the standpoint of semi-supervised clustering which is one of transductive semi-supervised classification problems. The number of clusters, i.e. classes, is simply two, one is an object, i.e. foreground, and the other is background. The pixels in an input image are classified into these two classes on the basis of their colors. The clusters are generally fuzzy due to the partial transparency of the objects, e.g. hairs, fogs, spider webs. Since this problem is under-constrained, users are required to input supporting information by drawing strokes specifying these regions on the image by hand. Many matting methods [1]～

[3] require such user interaction in the form of trimap where definitely foreground, definitely background and the remaining unknown regions are painted in graded colors, e.g. white, black and gray as is shown in Fig.1. This trimap is difficult to draw manually in images including objects with many holes or gaps such as spider webs. Advanced methods accepting fewer user inputs still require strokes in both of foreground and background regions [4]~[6]. We present, in this paper, a semi-supervised object extraction method in which few and rough drawings of supporting strokes in only one of two regions are sufficient for extracting such a complex and fuzzy object. Our technique is derived from a graph spectral method for unsupervised extraction of clusters from similarity data [7].

Clustering algorithms are classified into a partition type and an extraction type. A representative method of partition type is the normalized cut algorithm [8] which use a continuous relaxation of [-1,1] integer programs. This algorithm gives good results of hard clustering but cannot be used for fuzzy clustering and cannot extract clusters buried in unstructured noise data [9]. On the other hand, the affinity factorization method [9], which is the same as the graph spectral methods [10], [11], is a continuous relaxation of [0,1] integer programs and is an extraction type algorithm. This method is robust to noise data and gives fuzzy clustering naturally. This algorithm gives, however, only clusters of spherical shapes. We have modified it for extracting fuzzy clusters of arbitrary shapes [7] and then extended it to a semi-supervised scheme [12]. We apply, in this paper, it to natural image matting tasks.

Almost all algorithms of natural image matting are classified into a cluster partition type. The methods using random walks [4] or graph cuts [13] deal with two classes symmetrically similarly to the graph spectral semi-supervised classification algorithms [14], [15], and hence inherently requires user's input of seed points in both of foreground and background regions. An excellent method by Levin and Weiss [16] also requires user inputs in both regions. The isoperimetric algorithm [17] is exceptionally asymmetric and can accept only one seed point but is different to our algorithm and its gap surmounting property and robustness to noises are unknown.

図 1 Example of trimap (left: input image, right: trimap).

We present, in this paper, a novel semi-supervised object extraction method derived from an unsupervised algorithm of extraction type for fuzzy clusters of arbitrary shapes. We use spatially wide window for propagating memberships and transform color coordinates with the linear discriminant analysis for enhancing color discriminability of an object. These devices enable our method to extract objects with holes or gaps and reduce user inputs of supporting strokes.

In section 2, we review an unsupervised object extraction method from which we derive a semi-supervised algorithm in section 3. We next utilize the Fisher's linear discriminant analysis for enhancing class separability in section 4 and devise some numerical techniques for saving the computational time in section 5. Finally in section 6, we exemplify the novel ability of our method for extracting an object holding many holes with user inputs of strokes in only one of foreground and background areas.

The color version of this paper is http://www.design.kyushu-u.ac.jp/~urahama/matting.pdf.

2. Graph Spectral Method with Regularized Normalization

Before presenting our matting method, we review an unsupervised object extraction method by using the graph spectral method with regularized normalization [7] from which our semi-supervised algorithm is derived. Let the color of pixel (i,j) be $c_{ij} = [R_{ij}, G_{ij}, B_{ij}]^T$ and the similarity between pixels (i,j) and $(i+k, j+l)$ be $s_{kl} = e^{-\alpha(k^2+l^2)-\beta\|c_{ij}-c_{i+k,j+l}\|^2}$ where s_{kl} is the abbreviation of $s_{ij;i+k,j+l}$. We use here the basic RGB color coordinate while we will introduce its transformation by dimensionality reduction in section 4.

The graph spectral method with regularized normalization [7] computes the proportion x_{ij} of the membership of pixel (i,j) in the cluster, i.e. object, by the formulation:

$$
\begin{aligned}
&\max_{x} && \sum_{i,j}\sum_{(k,l)\in W} x_{ij} s_{kl} x_{i+k,j+l} \\
&\text{subj.to} && \sum_{i,j} d_{ij} x_{ij}^2 = 1
\end{aligned}
\qquad (1)
$$

where $\sum_{i,j}$ denotes the sum over all pixels and W is a spatial window. We set here $d_{ij} = \max\{\sum_{(k,l)\in W} s_{kl}, \epsilon\}$ which is a form improved from $d_{ij} = \sum_{(k,l)\in W} s_{kl} + \epsilon$ in the original method [7] where the small regularization constant ϵ is added in all pixels. Alternatively to this uniform regularization, ϵ is added in eq.(1) only at pixels where $1/\sum_{(k,l)\in W} s_{kl}$ is singular. This selective regularization is effective for extracting elongated clusters buried in noisy data. The normalization, which denotes the multiplication of d_{ij} to x_{ij}^2 in the constraint condition in eq.(1), enables this method to extract arbitrarily-shaped clusters and the regularization, which means the thresholding of d_{ij} at ϵ, makes the method robust to noise data.

The solution of eq.(1) is the stationary point of its Lagrange function:

$$\max_x \min_\lambda \sum_{i,j} \sum_{(k,l)\in W} x_{ij} s_{kl} x_{i+k,j+l} - \lambda(\sum_{i,j} d_{ij} x_{ij}^2 - 1) \quad (2)$$

from which we know $x = [x_{ij}]$ is the principal eigenvector of the normalized similarity matrix $D^{-1}S$. The membership of pixel (i,j) is given by the normalization of x_{ij} as $u_{ij} = x_{ij}/\max_{i,j}\{x_{ij}\}$ [7]. This solution x can be calculated by using an iterative algorithm similar to the power method for computing the principal eigenvector of a matrix.

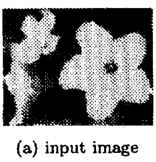
The membership u_{ij} obtained for the image of flowers in Fig.2(a) is shown in Fig.2(b). The parameters are set as $\alpha = 0, \beta = 0.1, \epsilon = 250$ and the window W is the entire image. Since the objects (flowers) are easy to extract in this example, we can extract them with no user interaction, however all of three flowers, not an isolated one, are extracted.

3. Semi-Supervised Object Extraction

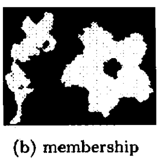
The above unsupervised method is hard to extract a specified object alone and also objects with complex textures including multiple colors. User interaction is prerequisite for specifying a target object and extracting it. We assume here that a user draws some strokes for specifying either a portion of regions included in a target object or those in the remaining background area. If a user draws strokes in an object, then x_{ij} is the membership of pixel (i,j) in the object and we fix $x_{ij} = 1$ at the pixels specified by a user in an object. Conversely if a portion of background regions is specified, we set x_{ij} as a membership of pixel (i,j) in the background, i.e. the foreground and the background are interchanged. If a user specifies both regions, x_{ij} represents the normal membership in the foreground.

If x_{ij} is given at some pixels $(i,j) \in T$ in this way, x_{ij} in itself becomes the membership and the constraint condition in eq.(1) becomes unnecessary and the Lagrange multiplier λ can be fixed to an arbitrary value. We fix here it to $\lambda = 1$, then eq.(2) becomes

$$\max_{(i,j)\notin T} \sum_{i,j} \sum_{(k,l)\in W} x_{ij} s_{kl} x_{i+k,j+l} - \sum_{i,j} d_{ij} x_{ij}^2 \quad (3)$$



(a) input image



(b) membership

図 2 Example of unsupervised object extraction.

We call the region T, where x_{ij} is given, the seed region. In cluster partition types such as the graph cut [13], users are required to input seed regions in both of foreground ($x_{ij} = 1$) and background ($x_{ij} = 0$) areas. In our method formulated by eq.(3) which is derived from eq.(1) of a cluster extraction type, it is sufficient for users to input seed regions in only one of foreground and background areas. This is a merit of the method of cluster extraction types against the cluster partition types. It is a matter of course that our method accepts strokes in both regions which raise its extraction performance.

The solution of eq.(3) is given by solving the following system of linear equations:

$$x_{ij} = \sum_{(k,l)\in W} s_{kl} x_{i+k,j+l}/d_{ij} \qquad (i,j) \notin T \qquad (4)$$

We solve it here with the Gauss-Seidel iterative scheme. During the iteration, $\{x_{ij} | (i,j) \in T\}$ is fixed to the value specified by a user. As the iteration progresses, x_{ij} propagates from pixels $(i,j) \in T$, where x_{ij} is known, to pixels $(i,j) \notin T$ where x_{ij} is unknown. We set the square window $W = \{k,l\} | (-p \leq k \leq p, -p \leq l \leq p)$ sufficiently broad for x_{ij} to propagate with surmounting gaps even if an object holds holes or gaps. In previous methods, the propagation is limited to only 4 or 8 neighborhoods, i.e. $p = 1$, hence it cannot jump over holes or gaps. Thus previous methods require many seed regions, contrastively a few and coarse seed regions are sufficient in our method. Some previous methods [6], [16] employ a hierarchical approach with pyramidal compression of images for remote propagation and acceleration of computation, however the object boundary is blurred in spatially compressed images and may cause errors in the object extraction. Our method does not use the image pyramid and deals with an input image in its intact size with the original resolution. Alternatively we reduce the computational time with effective setting of initial values of x_{ij} and successive reduction of processed pixels.

4. Color Coordinate Transformation by Linear Discriminant Analysis

Color coordinates are crucial for natural image matting

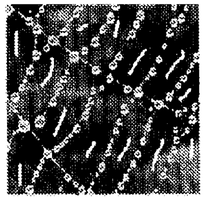


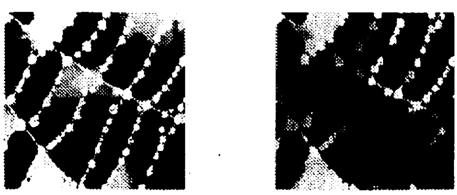
図 3 Image of spider web

where objects are extracted according to the similarity between pixel colors. For instance, Fig.4(a) shows the best one among the results obtained with the RGB color coordinate for the image of spider web in Fig.3 where white line scribbles are seed regions drawn only in the background area. Thus the RGB color coordinate gives unsatisfactory extraction of objects. We then examine the CIELAB color coordinate popularly used in various image processing but we get also an unsatisfactory result shown in Fig.4(b) at the best. Though their extractability varies from image to image, both of RGB and L*a*b* fail to extract objects in many images.

Grady [4] adopted the locality preserving projection (LPP) for the color coordinate transformation in his random walk extractor. The LPP is, however, an unsupervised algorithm neglecting user input information. We exploit user inputs and adopt the Fisher's linear discriminant analysis (LDA) for transforming the color coordinate. Since the matting is a classification problem of two classes, we project the colors into a 1-dimensional space.

We firstly enlarge the seed regions for improving the accuracy of the LDA projection. This enlargement of seed regions also serves for speeding up the iterative solution of eq.(4). The seed regions are dilated successively along the distance of pixels from the seed regions. We compute the 4-neighborhood distance transform of the seed regions with a fast 2-scan algorithm (we examined also the 8-neighborhood distance transform whose result was almost the same as that with the 4-neighborhood distance). We dilate the seed regions in the order of pixels with small distance to larger one from them if the color difference measured in the L_1 norm of RGB between the current pixel and its neighbor pixel is smaller than a small threshold. For instance, Fig.5(a) illustrates the dilated seed regions in Fig.3 with the threshold of color difference being 10.

We next perform the LDA with the dilated seed regions as one class and the remaining set of pixels as the other class. We compute their average colors m_1, m_2, and covariance matrices A_1, A_2 and their mean $A = (n_1A_1 + n_2A_2)/n$ where n_1 is the number of pixels in one class and n_2 is that in the other class, $n = n_1 + n_2$ is the size of the input image. The projection vector is then given by $q = A^{-1}(m_1 - m_2)$.



(a) RGB (b) L*a*b*

図 4 Memberships obtained with RGB (left) and with L*a*b* (right) for Fig.3.

The color c_{ij} of every pixel is then projected to scalar value $f_{ij} = q^T c_{ij}$. By this projection, the input color image is reduced to a monochromatic image in which the gray scale of each pixel is $g_{ij} = 255(f_{ij} - f_{min})/(f_{max} - f_{min})$ where f_{max} is the maximum value of f_{ij} and f_{min} is its minimum. The monochromatic image of Fig.3 is shown in Fig.5(b).

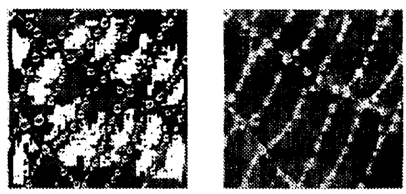
If strokes are drawn in both of foreground and background areas, we dilate both seed regions and execute the LDA with those dilated region as two classes.

5. Membership Propagation

We apply the method in section 3 to the monochromatic image obtained with the LDA. The similarity between pixels is now given by $s_{ij} = e^{-\alpha(k^2+l^2)-\beta(g_{ij}-g_{i+k,j+l})^2}$. We solve eq.(4) with the Gauss-Seidel iteration during which x_{ij} is fixed at the pixels in the dilated seed regions. The initial value for the iteration is set to $x_{ij} = 1/(1+e^{-\gamma(g_{ij}-\delta)})$. The above dilation of seed regions and this initial value setting reduce the convergence time of the iteration. In order to further speed up the convergence, we quantize x_{ij} to 1 if it becomes larger than 0.99 and similarly we set $x_{ij} = 0$ if it decreases below 0.01. The computation of x_{ij} is skipped at these pixels where x_{ij} reaches 0 or 1.

6. Experiments of Object Extraction

Figure 6 is an example of objects including many holes, i.e. green leaves are seen in many vacancies in petals of the red flower of a spider lily. The white scribbles are seed regions drawn in the background area, i.e. on the green leaves.



(a) dilated seed regions (b) monochoromatic image

図 5 Dilated seed regions in Fig.3 (left) and monochromatic image obtained from it with LDA(right)

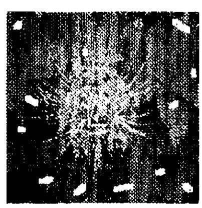


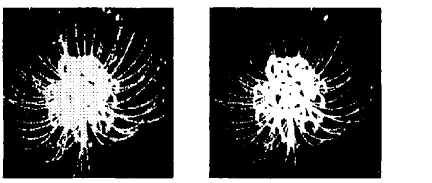
図 6 Image of spider lily

One example of result of our method is shown in Fig.7(a) where parameters were set as $\alpha = 0.01, \beta = 0.1, \gamma = 0.1, \delta = 100, \epsilon = 1, p = 1$ and the threshold for color distance in the dilation is 10. In Fig.7(a), many green spots in the red flower are erroneously mixed in the extracted object (some spots are classified correctly as background but this is solely owing to our effective setting of initial values of x_{ij}). This erroneous mixing of background spots into the object is due to the small window size of $p = 1$, i.e. 8-neighbors, which is too narrow for x_{ij} to propagate from surrounding background to inner spots by surmounting flower petals. Previous methods [4], [5] propagate alpha mattes through only 4 or 8 neighborhood, hence its propagation stops at these gaps. In order to get satisfactory object extraction in such methods, either many strokes must be drawn in every spot carefully or a hierarchical approach of pyramidal compression of the image size should be exploited [6]. However the sub-sampling of pixels in the pyramid blurs the image and may cause errors in object extraction. A result of our method with sufficiently broad window of $p = 20$ is shown in Fig.7(b) where every spot in petals is correctly classified to background.

The spider web in Fig.3 is another example of complex objects on which drawing strokes is hard. Contrastively drawing strokes in background areas is easy as is shown in Fig.3. The membership obtained with our method is shown in Fig.8.

Our method can extract an object even if the object and the background include similar colors as in Fig.9(a) where black colors are included in the eyes and in the background. Extracted memberships are shown in Fig.9(b).

In many images such as those shown in Fig.3, Fig.6 and Fig.9(a), our method can extract objects with user drawing of supporting strokes only either in foreground (object) or in background area in contrast to many previous methods which requires strokes in both regions (so, direct comparison of our results with them is impossible with the same input condition). It is however better for our method that strokes are drawn in both of object and background regions for preventing omission of details from extraction . For instance as is shown in an image of peacock in Fig.10 where strokes are drawn only in the background area, the body and legs of the peacock are missed to be extracted as in Fig.11(a). In such cases, if strokes are added on them, then the whole figure of peacock is extracted as in Fig.11(b). Since our algorithm is iterative, the membership is easily and quickly updated by restarting the iteration with appended seed regions. Our iterative algorithm is suitable for such interactive processes where a user repeats the cycle of verification and update of extracted memberships. Note the difference between the image in Fig.9(a) and that in Fig.10. Both images include similar colors in the forground and in the background. In Fig.9(a), they (black colors) are apart, while they (greenish colors) are contiguous in Fig.10.



(a) $p = 1$ (b) $p = 20$

図 7 Obtained memberships with $p = 1$ (left) and with $p = 20$ (right) for Fig.6.

We implemented the program with the C language on a computer with P4 2.79GHz CPU, 1GB RAM and the OS is Windows XP. The computational time of our method is 57, 29, 27 and 32 seconds for Fig.7(b) (200×200, $p = 20$), Fig.8 (281×281, $p = 5$), Fig.9(b) (278×333, $p = 5$) and Fig.11(b) (373×290, $p = 5$).

7. Object Composition with Another Image

Composition of an extracted object with another image needs the color of the object at each pixel. The color c_{ij}

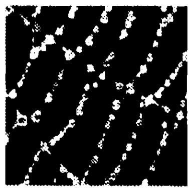
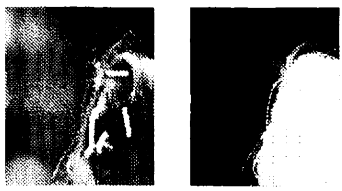


図 8 Obtained memberships for Fig.3.



(a) input image (b) memberships

図 9 Image of face with bright hairs

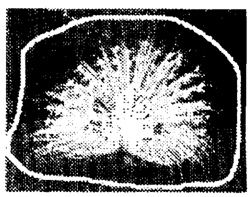
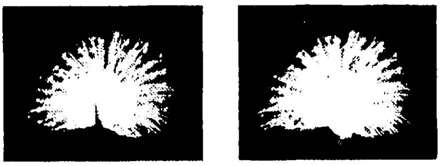


図 10 Image of peacock



(a) with strokes in Fig.10 (b) with additonal strokes

図 11 Obtained memberships with strokes in Fig.10 (left) and with additional strokes (right) for Fig.10.

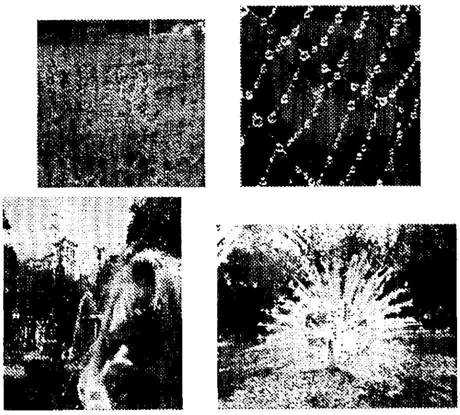


図 12 Composite images.

of the input image is composed with a foreground color c_{fij} and a background color c_{bij} with the mixing proportion x_{ij} as $c_{ij} = x_{ij}c_{fij} + (1 - x_{ij})c_{bij}$. In previous matting methods [1]~[16], both of c_{fij} and c_{bij} are estimated based on this relation, however their simultaneous estimation is hard and is also wasteful because only c_{fij} is needed for the image composition. We estimate here them by the following optimization formulation:

$$\min_{c_f} \sum_{i,j} \sum_{(k,l)\in W} s_{kl} x_{i+k,j+l} \| c_{fij} - c_{i+k,j+l} \|^2 \qquad (5)$$

Where s_{kl} is the similarity between pixels used in the above algorithm for object extraction and x_{ij} is the membership which is also already obtained with the above algorithm. Eq.(5) is solved analytically as

$$c_{fij} = \sum_{(k,l)\in W} s_{kl} x_{i+k,j+l} c_{i+k,j+l} \Big/ \sum_{(k,l)\in W} s_{kl} x_{i+k,j+l} \qquad (6)$$

from which we can get a new composite color $\tilde{c}_{ij} = x_{ij}c_{fij} + (1 - x_{ij})b_{ij}$ where b_{ij} is the color of pixel (i, j) of another image used as a new background. Examples of composited images with the objects in Fig.6, Fig.3, Fig.9(a) and Fig.10 are shown in Fig.12 where partially transparent objects such as spider webs and hairs are well extracted and composited with new backgrounds.

8. Conclusion

We have presented a method for natural image matting based on the semi-supervised algorithm for extracting a fuzzy cluster from similarity data. The novelty of our method lies in 1) application of the algorithm of cluster extraction type, 2) wide window for propagating memberships surmounting holes and gaps in an object, and 3) color coordinate transformation with the LDA. Main merit in our method is its ability of object extraction from a light user interaction drawing rough and coarse strokes in only one of foreground and background areas. Its further speed-up and application to videos are under study.

文 献

[1] M. Ruzon and C. Tomasi, "Alpha estimation in natural images", Proc. CVPR, pp.18-25, 2000.
[2] Y. Y. Chuang, B. Curless, D. Salesin and R. Szeliski, "A Bayesian approach to digital matting", Proc. CVPR, pp.264-271, 2001
[3] S. Lin and J. Shi, "Fast natural image matting in perceptual color space", Computers & Graphics, 29, pp.403-414, 2005.
[4] L. Grady, T. Schiwietz and S. Aharon, "Random walks for interactive alpha-matting", Proc. VIIP, pp.423-429, 2005.
[5] V. Vezhnevets and V. Konouchine, "GrowCut: Interactive multi-label N-D image segmentation by cellular automata", Proc. Graphicon, 2005.
[6] J. Wang and M. C. Cohen, "An iterative optimization approach for unified image segmentation and matting", Proc. ICCV, pp.936-943, 2005.
[7] K. Inoue and K. Urahama, "Robust extraction of arbitrarily-shaped clusters using graph spectral method with regularized normalization", Information Science Technology Letters, LI-008, pp.159-160 (in Japanese), 2005.
[8] J. Shi and J. Malik, "Normalized cuts and image segmentation", IEEE Trans. PAMI, 22, 8, pp.888-905, 2000.
[9] P. Perona and W. T.Freeman, "A factorization approach to grouping", Proc. ECCV, pp.655-670, 1998.
[10] S. Sarkar and K. L. Boyer, "Quantitative measures of change based on feature organization: eigenvalues and eigenvectors", Comput. Vision Image Und., 71, 1, pp.340-349, 1998.
[11] K. Inoue and K. Urahama, "Sequential fuzzy cluster extraction by a graph spectral methof", Patt. Recog. Lett., 20, 7, pp.699-705, 1999.
[12] W. Du, K. Inoue and K. Urahama, "Unsupervised and semi-supervised extraction of fuzzy clusters in similarity data", IEICE Tech. Report, PRMU2005-177, pp.165-170, Jan. 2006.
[13] C. Rother, V. Kolmogorov and A. Blake, "GrabCut: interactive foreground extraction using iterated graph cuts", ACM Trans. Graphics, 23, 3, pp.309-314, 2004.
[14] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation", Tech. Report, CMU-CALD-02-107, 2002.
[15] D. Zhou, O. Bousquet, T. N. Lal, J. Weston and B. Scholkopf, "Learning with local and global consistency", Proc. NIPS, pp.321-328, 2003.
[16] A. Levin, D. Lischinski and Y. Weiss, "A closed form solution to natural image matting", Proc. CVPR, 2006.
[17] L Grady and E. L. Schwartz, "Isoperimetric graph partitioning for image segmentation, IEEE Trans. PAMI, 28, 3, pp.469-475, 2006.