

Image Segmentation Based on Visual Perception Model

Shuai Shao, Fuqing Duan

Image Processing and Pattern Recognition Laboratory
Beijing Normal University,
Beijing, China
oliveshao@126.com

Ping Guo

Image Processing and Pattern Recognition Laboratory
Beijing Normal University
Beijing, China
pguo@ieee.org

Abstract—Image segmentation is the basis of image processing and image analysis. However, there are no common method that can be used in natural images, and present methods fail to explain understandings of human’s visual system. In this paper, we propose to apply Karklin’s visual perception model to extract feature vectors of images, and the features are clustered with K-means method. The results obtained in feature space are projected back to the image space to finish segmentation. A comparison with the Normalized Cuts (Ncut) method is done, and it turns out that proposed method outperform Ncut in texture rich images.

Keywords—Visual perception model; Image segmentation; Normalized cuts; Feather extraction

I. INTRODUCTION

Image segmentation has been one of the basic problems of image processing. So far, there has been a lot of image segmentation algorithms, such as methods based on clustering [1], on active contour [2][3], on active appearance model (AAM) [4], on motion [5], and on graph theory [6], etc.

Recently, many researchers have focused on graph theory based segmentation methods, and one of the most notable works is the Normalized Cuts (Ncut) [6]. Ncut is better than other methods, as it provides a standard segmented criterion and will not tend to small areas in segmentation. However, when it comes to natural images, the results may mismatch human’s understanding. In fact, human visual structure outperforms the best machine vision system in any measurements [7], and building a system that emulates working mechanism in human cortex has always been attracting. Will it be better than the present methods if an image is segmented by visual perception model? In this paper, we propose to apply Karklin’s visual perception model [8][9] to image segmentation, and make a comparison with the Ncut method [6].

II. SEGMENTATION BASED ON KARKLIN’S MODEL

A. Karklin’s Model

It is known that perceptually distinct image regions produce response patterns that are highly overlapping and cannot be easily distinguished using low-level, linear representations [8]. To solve this problem, Karklin’s model intends to transform image pixels to a new feature space where image patches of different vision patterns can be easily discriminated. The model makes an inference of

covariance of image intensities, that is to say, the greater the difference in shapes of covariance of images, the easier the latent variables y_j can be distinguished.

Karklin’s model describes individual image patches x with multivariate Gaussian probability distributions [9]:

$$p(x|y) = N(0, C) \quad (1)$$

The logarithm of the covariance matrix C is given by the combination of a group of linear base function:

$$\log C = \sum y_j A_j \quad (2)$$

where A_j is a symmetric matrix whose size is equal to the covariance matrix C . A_j can be represented as:

$$A_j = \sum_k \omega_{jk} \mathbf{b}_{jk} \mathbf{b}_{jk}^T \quad (3)$$

Where \mathbf{b}_{jk} are the unit vectors in K -dimensional space, specifying orientation of image features. \mathbf{b}_{jk} are weighted by ω_{jk} so that the encoded distribution can be lengthened or shortened along them. y_j describes changes along \mathbf{b}_{jk} using weights ω_{jk} , and it can be seen as texture feature vectors of images. In the model, \mathbf{b}_{jk} and ω_{jk} are shared by all neurons in the model, while y_j differs from patches to patches.

The goal of Karklin’s algorithm is to estimate latent variables and model parameters with MAP(Maximum a posterior). In each iteration of training process, y_j is calculated using gradient ascent, and \mathbf{b}_{jk} and ω_{jk} are then estimated with the new y_j . This is repeated until the model reaches convergence.

B. Model Parameter Selection

With Karklin’s visual model, image features are extracted [10] for image segmentation. Model parameters are determined in the experimental methods. The training set contains 40 grayscale images with the size of 640×500 pixels from kyo_dat [11]. Most of the images have similar visual characteristic.

Before training, overlapping $n \times n$ -image patches are randomly extracted from the images. Mean luminance value was subtracted from each patch. Then the image patches are whitened to remove global correlations and to normalize the variance, so their covariance is of canonical distribution.

Sizes of image patches used to train model parameters have great influence on results of segmentation, so we will discuss performance of models trained with different sizes of image patches first. If sizes of images are too large, it will take excessive computation to train the model, and if sizes are too small, performance of the model will decline. As a

result, sizes of image patches in our experiments are 20×20 as mentioned in [8], as well as 10×10.

If 20×20 image patches are used for training, the number of covariance coefficient is set to 150 and the number of b_{jk} 1000 [9].

When it comes to the 10×10 image patches, the number of covariance coefficient is set to 60, and the number of b_{jk} 300 [9]. Besides, training images have to be downsampled to one-fourth the size of the original image. Because the larger the image is, the more texture information is contained in the randomly extracted patches. But the model can not learn too much texture information at one iteration, or it will degenerate.

After training, the model are applied to extract features from different kinds of images, the features are then clustered and projected onto the two-dimensional. The results are shown in Figure.1.

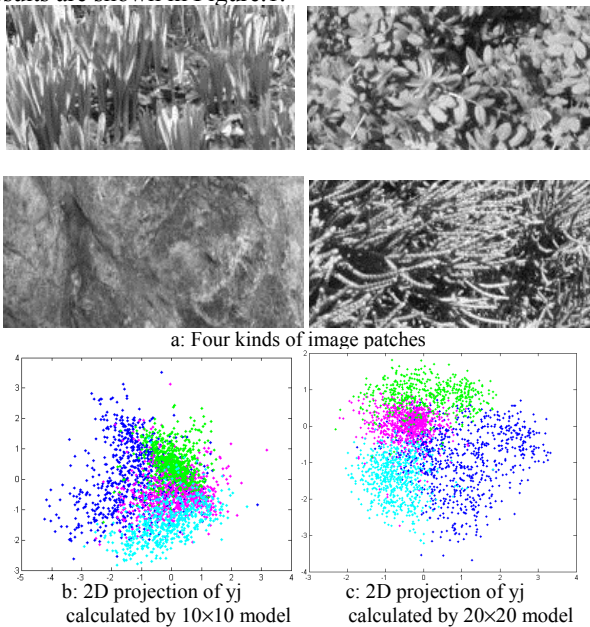


Figure 1. Patches and model

Points in Figure.1.b represent two-dimensional projections of the model's representation calculated for 10×10 patches extracted from images in Figure.1.a, and Figure.1.c for 20×20 patches. It can be seen that points in Figure.1.c are separated more distinctly, which means the model in Figure.1.c behaves better. In consequence, model trained with 20×20-image-patches is used in the following experiment to extract features [10].

C. Algorithm Implementation

The algorithm using Karklin's model in image segmentation is designed as follows,

- The input image is divided into $n \times n$ ($n=20$) patches;
- Latent variable y_j is calculated for each patches;
- All latent variables are clustered with K-means;
- Clusters of y_j 's are then projected back into the original image.

As K-means results are projected back, there might be certain singular value points existing in partitions. To solve this problem, all of the four-connected regions of the singular value point are traversed to find the nearest color of it, and then it is replaced by the closest color.

III. RESULTS

Images used for our experiments consisted of 500 photographs downloaded from the Web. The images varied widely in compression ratio and size, ranging from about 300000 to 500000 pixels; all of them are outdoor photographs.

In this section, Karklin's model and the Normalized Cuts are used in image partition respectively, and are compared with each other.

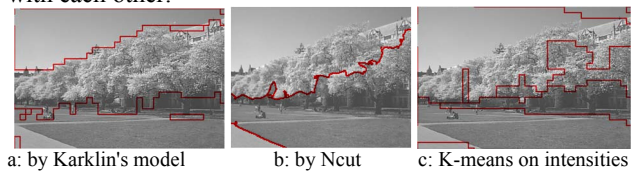


Figure 2. segmented

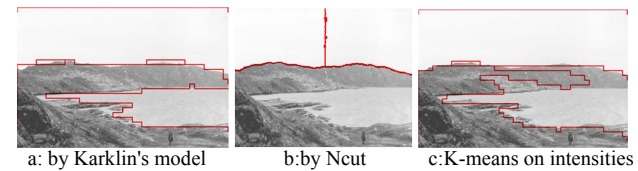


Figure 3. segmented

The Ncut algorithm supposes an undirected graph $G = (V, E)$ can be partitioned into two disjoint sets A and B with $A \cup B = V, A \cap B = \Phi$, let x be an $N = |V|$ dimensional indicator vector, $x_i = 1$ if node i is in A , and -1 otherwise.

The degree of node i is $d_i = \sum_j w(i, j)$, which can be calculated as the total connection from node i to all other nodes. Then non-correlation between the two sets of can be calculated using Ncut:

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)} \quad (4)$$

where $assoc(A, V) = \sum_{u \in A, t \in V} W(u, t)$.

In our experimental environment, sizes of images are restricted to 300×300 when Ncut is used, or the computer will go out of memory.

The results are shown in Figure.2 to Figure.6. Images segmented by K-means on intensities directly are presented on the right for comparison.

Pixel ratio of error segmentation is shown in Table 1.

TABLE I. MISCLASSIFIED PIXELS RATIO.

Objects	Karklin Cut	Ncut	Figure
Cherries	0.91%	40.6%	2
Mountain	3.33%	11.71%	3
Buildings	5.19%	49.47%	4

Boat	18.52%	9.78%	5
Pavilion	5.99%	49.87%	6

It can be seen from the experiments that,

- if images are composed of several regions whose texture information differs greatly, our methods tend to perform better than Ncut; for example, in Figure.2.a, Karklin's model partitions the three regions distinctly;
- segmented regions in the experiment will always separate perceptually identical categories into the same regions; However, as Ncut works on pixels, it will cut an object into different parts if an object contains edges itself, for example, in Figure.6.b, Ncut cuts the pavilion apart, for there are edges within it;
- segmentation with Karklin's model are inclined to cut the object out of its background, for example, in Figure.5.a, the boat is partitioned from the images;
- edges of partitions may not be as dedicate as Ncut, as the image is segmented with 20×20 patches.

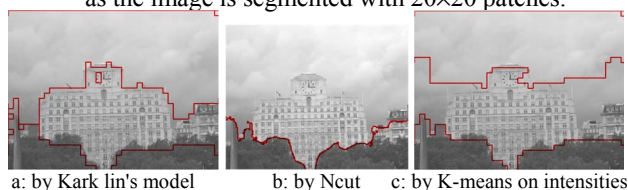


Figure 4. segmented

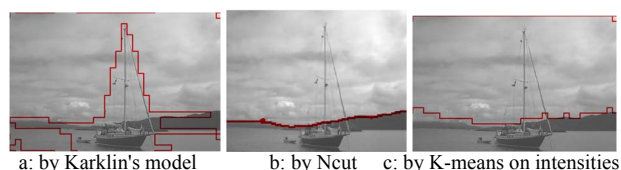


Figure 5. segmented

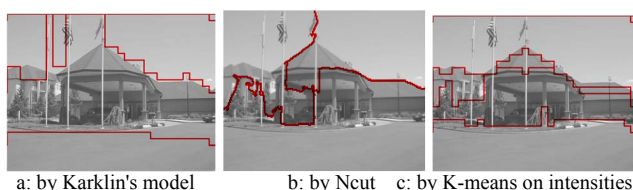


Figure 6. segmented

The reason is that, Karklin's model transforms image intensities to a new feature space where patches of different vision patterns can be easily discriminated. Moreover, as the patches are 20×20 , edges in local areas have less influence in partitions than Ncut. It is because Karklin's model is based on latent variables y_j rather than on luminance values of pixels that the results of our methods conform to human visual perception better than Ncut.

Besides, under our experimental condition, Karklin's model is not sensitive to the size of images, and it can work on images of large size; On the other side, Ncut has size restrictions.

IV. CONCLUSIONS

The Karklin's visual perception model is applied to image segmentation and a comparison with the Ncut algorithm is performed in this paper. Features are extracted from the input image with Karklin's model, clustered using K-means, and projected back to the original image to finish segmentation. It turns out that results with proposed method are more consistent with visual patterns and less affected by edges of objects in images than Ncut. In a word, proposed method outperforms Ncut on large, texture-rich natural images. However, the outlines of image partitions is not as smooth as those calculated by the Ncut algorithm. This is a limit of our methods.

In the future, we will seek more effective method to improve the result of model parameters trained with small patches to smoothen outlines of the partitions.

ACKNOWLEDGEMENTS

The research works describe in this paper was fully supported by the grants from the National Natural Science Foundation of China (Project No. 90820010, 60911130513). Prof. Ping Guo is the author to whom the correspondence should be addressed, his e-mail address is pingguo@ieee.org.

REFERENCES

- [1] XU Rui. Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*. 2005, 16(3), PP: 645 - 678.
- [2] KASS Michael, WITKIN Andrew, and TERZOPOULOS Demetri. Snake: active contour models. *Computer Vision*. 1987, 1(4), PP: 321-331.
- [3] CASELLES Vicent, KIMMEL Ron, and SAPIRO Guillermo. Geodesic active contours. *International Journal of Computer Vision*. 1997, 22(1), PP: 61- 79.
- [4] LELIEVELDT B.P.F., UZUMCU M., GEEST R.J. van der, REIBER J.H.C., SONKA M.. Multi-view active appearance models for consistent segmentation of multiple standard views: application to long - and short - axis cardiac MR images. *Proceedings of The 17th International Congress and Exhibition*. 2003, PP: 1141- 1146.
- [5] CREMERS D.. A variational framework for image segmentation combining motion estimation and shape regularization. *IEEE Conference on Computer Vision and Pattern Recognition*. C. Dyer and P. Perona (Eds.). 2003. 1, PP: 53-58.
- [6] SHI Jianbo and. MALIK Jitendra. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000, 22(8), PP: 888-905.
- [7] SERRE Thomas, WOLF Lior, BILESCHI Stanley. Maximilian Riesenhuber, and Tomaso Poggio, Robust Object Recognition with Cortex-Like Mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2007, 28(3), PP: 411-426.
- [8] KARKLIN Yan, LEWICKI Michael S.. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*. 2009, 457(7225), PP: 83-86.
- [9] KARKLIN Yan. Hierarchical statistical models of computation in the visual cortex. [D]. School of Computer Science Carnegie Mellon University Pittsburgh, PA 15213. 2007.
- [10] GUO Ping; WAN Tao; MA Jin. Experimental Studies of Visual Models in Automatic Image Annotation, *Lect. Notes in Comp. Sci.*, 2011, LNCS 6761, PP.562-570.
- [11] DOI Eizaburo, INUI Toshio, LEE Te-Won, WACHTLER Thomas, SEJNOWSKI Terrence J.. Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes. *Neural Computation*. 2003, 15(2), PP: 397-417.