Efficient image segmentation method based on resolution and region information fusion

Yanwei Ju (句彦伟)[1], Zheng Tian (田 铮)[1,2], and Yan Zhang (张 燕)[1]

[1]*Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an 710072*

[2]*National Key Laboratory of Pattern Recognition & Institute of Automation, Chinese Academy of Sciences, Bejing 100080*

A generalized multiresolution likelihood ratio (GMLR), which can increase the distinction between different signals by fusing their more features, is defined. Multiresolution representation of image characterizes inherent structure of image well, and the GMLR combines each resolution image features with corresponding region features. A spatially variant mixture multiscale autoregressive prediction (SVMMARP) model is proposed to estimate the parameters of GMLR based on maximum likelihood estimation via expectation maximization (EM) algorithm. In the parameter estimation, bootstrap sampling technique is employed. Experimental results demonstrate that the algorithm performs fairly well.

  OCIS codes: 100.0100, 150.0150, 070.4560.

Image segmentation plays a central role in low-level computer vision. It is a pre-requisite for solving many other computer vision problems, such as image classification, content-based image retrieval, and object recognition. Unsupervised image segmentation may be defined as the task of dividing an image into several homogeneous regions automatically based on some similarity measures. An efficient approach to measuring the similarity of regions is using probabilistic models. In recent years, most of probabilistic models are based on mixture model and Gibbs distribution[1−5]. For example, the multiresolution Gaussian autoregressive (MGA) model[5] takes into account the correlation between adjacent levels of resolutions. It was assumed that the parameters of the Gibbs distribution of the region process are known. Segmentation is achieved there as the maximum posterior marginals (MPM) estimate[6] rather than the maximum a posteriori (MAP) estimate, since MPM estimate minimizes the probability of classification error. However, it may only be approximately computed in a computationally expensive procedure similar to simulated annealing, and the MPM criterion does not consider the spatial placement of errors when distinguishing among the quality of segmentations. In MGA model, spatial interaction parameters are selected experimentally or prespecified. And likelihood ratio test is also often used[7,8].

In above almost all statistical approaches, the image model is frequently based on the assumption of statistically independent image sample with different marginal probability distributions in each region. After the distributions are recovered from the mixed empirical signal distribution over the image (see Refs. [9—13] to cite a few), initial segmentation is performed by low-level pixel-wise classification. Approximation and classification of scalar data using a mixture of probability distributions are also widely used for data clustering in pattern recognition[12,14,15]. These statistical approaches are to maximize likelihood function based on statistically independent image sample. In fact, these image samples are statistically dependent.

In this paper, a generalized multiresolution likelihood ratio (GMLR) is defined and GMLR test is given, which not only fuses several predictable features (predictable images) of image to increase distinction, but also it decision (pixel classification) is based upon regions. The statistically independence assumption of maximum likelihood estimation is satisfied better, and the algorithm saves computation greatly by using bootstrap sampling technique. The parameters of GMLR is estimated by a spatially variant mixture multiscale autoregressive prediction (SVMMARP) model[16]. Each image pixel is classified through GMLR test based on test window.

Considering different features of all kinds of signals (including images) and fusing these features better, we define GMLR. For simplicity and convenience of explanation, we give GMLR for binary hypotheses as

$$\lambda(x_1, \cdots, x_K) = \frac{\prod\limits_{k=1}^{K} p_k(x_k \,|H_1\,)}{\prod\limits_{k=1}^{K} p_k(x_k \,|H_0\,)}, \qquad (1)$$

where binary hypothesis H_1 (H_0) denotes two kinds of signals; $x_k = (x_{k1}, \cdots, x_{kp})$ $(k = 1, \cdots, K)$ are p dimension vectors; $p_k(x_k \,|H_i\,)$, $(k = 1, \cdots, K; \; i = 0, 1)$ are probability density functions (PDFs) and each of them may be different types for describing different features of signals, for example, Gaussian, Rayleigh, and so on. We also can get GMLR test

$$\frac{\prod\limits_{k=1}^{K} p_{r|H_0}(x_k \,|H_0\,)}{\prod\limits_{k=1}^{K} p_{r|H_1}(x_k \,|H_1\,)} \begin{array}{c} H_0 \\ > \\ < \\ H_1 \end{array} \frac{\prod\limits_{k=1}^{K} (C_{00}P_0 c_k + C_{01}P_1\beta_k)}{\prod\limits_{k=1}^{K} (C_{10}P_0\alpha_k + C_{11}P_1 r_k)}, \quad (2)$$

where C_{00}, C_{10}, C_{01}, C_{11}, P_0, c_k, P_1, β_k, α_k, r_k are some constants to get. It is easy to notice that Eq. (1) degenerates to classical likelihood ratio when $K = 1$.

The above GMLR can fuse different and more features of signals to consider together, and describe them with appropriate PDF, so the GMLR can make more precise

http://www.col.org.cn

decision for classification and pattern recognition.

For clarity of outline and explanation, all of the theory is binary hypotheses and all the G mentioned in the following is equal to 2. The detailed M-ary hypotheses and test of GMLR also can be gotten, similar to Ref. [17].

The starting point for our method development is a multiresolution sequence $X_K, X_{K-1}, \cdots, X_0$ of images, where X_K and X_0 correspond to the coarsest and finest resolution images, respectively. The detail can be found in Ref. [18]. As an example, Fig. 1 illustrates a multiresolution sequence of three images, together with the quadtree mapping. We use the notation $X(s)$ to indicate the pixel mapped to node s. The scale of node s is denoted by $k(s)$.

For image segmentation application, the GMLR of image is defined as

$$\lambda(x_{p1}, \cdots, x_{pK}) = \frac{\prod\limits_{k=0}^{K-3} p_k(x_{pk}^* | H_1)}{\prod\limits_{k=0}^{K-3} p_k(x_{pk}^* | H_0)}, \quad (3)$$

where x_{pk}^* ($k = 0, \cdots, K-3$) are prediction of bootstrap sample x_k^*. The $K-2$ predictable images capture the structure inherent in imagery well and accurately characterize the evolution in scale of homogeneous regions. By GMLR, all the structure inherent in imagery and the evolution in scale are fused.

When all the parameters in Eq. (3) are estimated, we classify each individual pixel based on a test window of pixels surrounding it. The classification of the image pixel $Y_0(s)$, denoted as $C(s)$, can be obtained via the rule

$$C(s) = \begin{cases} H_1 & \text{if} \quad \lambda(Y_0(s)) \geq \eta \\ H_0 & \text{if} \quad \lambda(Y_0(s)) < \eta \end{cases}, \quad (4)$$

where the threshold value η can be determined by detecting GMLR histogram.

Since Efron[19] introduced a very general re-sampling procedure, called the bootstrap model, to estimate the
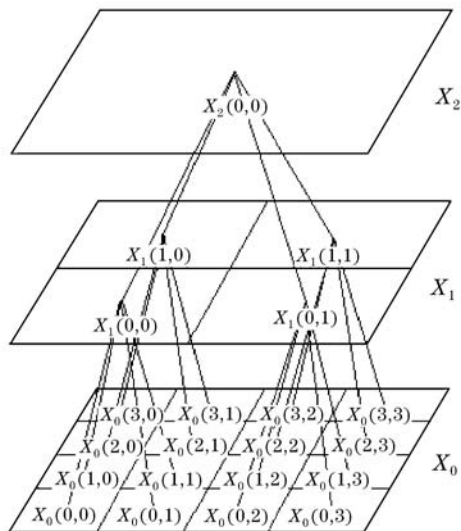

Fig. 1. Sequence of three multiresolution images mapped onto quadtree.

distributions of statistics based on independent observations, many researches applied a bootstrap method in pattern classification[20−22] and showed that it was a powerful non-parametric technique to evaluate classifier's performance. Here we employ bootstrap sampling technique to get bootstrap sample and sample size.

The criterion is chosen so that the sample would be representative only if each image grey level w_j occurs more than once into the bootstrap sample. Since the sample size of each calss $N_j's$ can be considered as independent Poisson random variables of parameter $n\pi_j$ (π_j is prior probability of each class), it must be subject to the condition $\sum_{j=1}^{G} n_j = n$, where n_j is a realization of N_j. Let us evaluate the probability P_n defined as

$$P_n = P(N_1 > 0, \cdots, N_j > 0, \cdots, N_L > 0)$$

$$= \prod_{j=1}^{G} (1 - e^{-n\pi_j}). \quad (5)$$

Then, the representative criterion requires P_n tending to 1. The problem to solve is to determine the optimal value n_0 of the nsS from which our bootstrap sample would be representative. Then, consider the function

$$p(x) = \prod_{j=1}^{G} (1 - e^{-\pi_j x}), \quad x \in R, \quad 0 < \pi_j < 1, \quad (6)$$

taking the derivation of log likelihood Eq. (6), and replacing x by n, we get

$$B(n) = \sum_{j=1}^{G} \frac{\pi_j e^{-\pi_j n}}{1 - e^{-\pi_j n}}, \quad (7)$$

where $B(n)$ is minimal when $P(x)$ is maximal. Hence, $B(n)$ could be regarded as the sampling characteristic function of the observed image, in such a way that it takes into account both the image pixel distribution and the bootstrap sample size. We can obtain bootstrap sample as the description in Ref. [23].

The use of the bootstrap sampling technique in image segmentation presents two advantages: 1) The choice of independent pixel sample which would allow an estimation of the statistical parameters of the image in the best conditions of independence; 2) the reduction of redundancy of information connected to the choice of a small representative sample, allows a gain in a factor N/n in times of calculation.

In order to estimate density function $p_k(\cdot | H_i)$ ($k = 1, \cdots, K; i = 0, 1$) in Eq. (3) for image unsupervised segmentation, we propose a SVMMARP model as[16]

$$\Phi(x_{pk}^*) | \theta^1 \cdots \theta^G) = \sum_{g=1}^{G} p_g^s \phi_g(x_{pk}^* | \theta^g), \quad (8)$$

where s is node at scale k; p_g^s denote the probability of the sth pixel belonging to the gth class at resolution k, $0 \leq p_g^s \leq 1$ and $\sum_{g=1}^{G} p_g^s = 1 \forall s$; $\phi_g(x_{pk}^* | \theta^g)$ is probability density function, and $\{\phi_g(\cdot | \theta^g)\}$ is a set of G density functions, each having its own vector of parameters θ^g. Here we relate Eq. (3) and model Eq. (8) so: as a

example, when classification number $G = 2$ and the resolution is k, $\phi_1(x^*_{pk}|\theta^1)$ and $\phi_2(x^*_{pk}|\theta^2)$ in Eq. (8) are same as $p_k(x^*_{pk}|H_0)$ and $p_k(x^*_{pk}|H_1)$ in Eq. (3); the parameters at each resolution in Eq. (3) can be estimated with a SVMMARP model. Here we consider $\phi_g(\cdot|\theta^g)$ is the PDF of a Gaussian distribution. We choose x^*_{pk} as a prediction of bootstrap sample x^*_k: firstly, we obtain multiresolution sequence $X_K, X_{K-1}, \cdots, X_0$ of image; secondly, we get bootstrap multiresolution sequence of image $x^*_0, \cdots, x^*_{K-1}$ from $X_K, X_{K-1}, \cdots, X_0$ as above; then we get x^*_{pk} as

$$x^*_{pk} = a_{k,1} x^*_{k+1} + a_{k,2} x^*_{k+2} + \cdots + a_{k,R} x^*_{k+p_k} + b_k. \quad (9)$$

One purpose of choosing x^*_{pk} as an autoregressive (AR) model is to filter and reduce the possible effect generated by the presence of noise of images. The order of regression associated with modeling x^*_{pk} from its ancestors will vary with the level $k \in \{0, 1, \cdots, K\}$ as defined by

$$p_k = \begin{cases} K - k & \text{if} \quad K - R + 1 \le k < K \\ R & \text{if} \quad 0 \le k < K - R + 1 \end{cases}. \quad (10)$$

Although regression order in Eq. (9) can be chosen by AR model order's criterion, here a maximum regression order $R = 3$ was selected as it was found to achieve a good trade-off between modeling accuracy and computational efficiency. The regression coefficients for resolution k are obtained by least squares minimization,

$$a_k = \arg \Big\{ \min_{a_l} \Big\{ \sum_s [x^*_k(s) - a_{k,1} x^*_{k+1}(s)$$

$$- \cdots - a_{k,R} x^*_{k+p_k}(s) - b_k]^2 \Big\} \Big\}, \quad (11)$$

where $a_k = [a_{k,1}, a_{k,2}, \cdots, a_{k,p_k}, b_k]'$ is regression coefficients, and $x^*_{k+i}(s)$ indicates pixel of bootstrap multiresolution sequence image $x^*_{k+i}(i = 0, \cdots, p_k)$.

The estimation of parameters is easily performed via expectation maximization (EM) algorithm[24], and the resulting iterations are given by

$$w_g^{s^{(i)}} = \frac{p_g^{s^{(i)}} \phi_g(x^*_{pk}(s))}{\sum\limits_{\{s|k(s)=k\}} p_g^{s^{(i)}} \phi_g(x^*_{pk}(s))}, \quad (12)$$

$$p_g^{s^{(i+1)}} = w_g^{s^{(i)}}, \quad (13)$$

$$u_g^{(i+1)} = \frac{\sum\limits_{\{s|k(s)=k\}} w_g^{s^{(i)}} x^*_{pk}(s)}{\sum\limits_{\{s|k(s)=k\}} w_g^{s^{(i)}}}, \quad (14)$$

$$\left[\sigma_g^2\right]^{(i+1)} = \frac{\sum\limits_{\{s|k(s)=k\}} w_g^{s^{(i)}} [x^*_{pk}(s) - u_g^{(i+1)}]^2}{\sum\limits_{\{s|k(s)=k\}} w_g^{s^{(i)}}}. \quad (15)$$

The estimates of the parameters are obtained by iterating above steps until convergence.

To verify the applicability of our approach to the segmentation of image, two simulated images with two different Gaussian distribution pixels and known class boundaries were created (Figs. 2(a,e)). Table 1 is the sampling characteristic function $B(n)$ from the two simulated images, and it is obvious that the bootstrap sample size $n_0 = 32 \times 32$ at scale 0 is enough to have the representation. We can estimate the parameters only using the bootstrap sample not whole sample, so our estimation is quite fast. We use the pixels in square windows of 3×3 pixels to classify the central pixel, and use a quadtree with $K = 4$ levels for the results presented here. The first simulated image has quite distinct mean and standard deviation, and the global histogram demonstrates a bimodal distribution (Fig. 2(b)). The segmentation with mixture model gives an accurate segmentation result, and our method gives a similar result (Table 2). The second simulated image has little distinct mean and standard deviation, and the global histogram demonstrates a unimodal distribution (Fig. 2(f)). The segmentation with mixture model gives a very poor segmentation result (65.76%), and our method gives an accurate result (98.02%) (Table 2), with all boundaries

Table 1. Sampling Characteristic Function $B(n)$ from the Two Simulated Images with Same Prior Probability

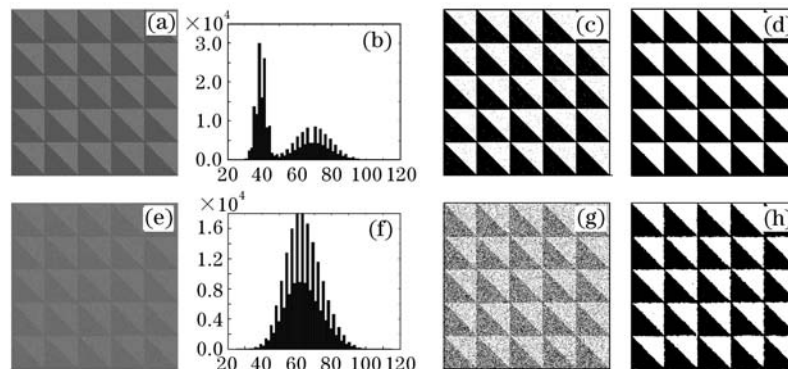| n	60	120	180	1024
$B(n)$	9.3576	8.7565	8.1940	4.3775
	$\times 10^{-14}$	$\times 10^{-27}$	$\times 10^{-40}$	$\times 10^{-223}$

Fig. 2. (a,e) Simulated images; (b,f) histograms of the simulated images; (c,g) segmentation based on classical mixture model; (d,h) segmentation based on our method.

Table 2. Comparison of Segmentation Accuracy

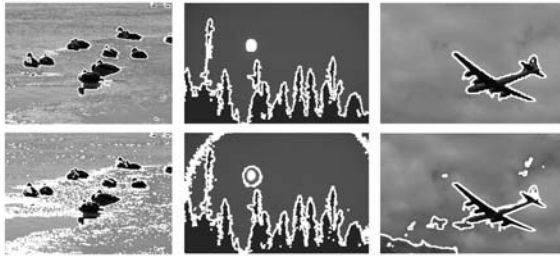| Method	First Simulated Image	Precision	Second Simulated Image	Precision
Mixture Model	Class Means $\vec{\mu} = (40, 70)$	99.15%	Class Means $\vec{\mu} = (60, 70)$	65.76%
Our Method	Standard Deviations $\vec{\sigma} = (3, 10)$	99.42%	Standard Deviations $\vec{\sigma} = (8, 10)$	98.02%

Fig. 3. Segmentation results based on our approach (the first row, and test window is 3×3) and mixture model (the second row). Edges are superimposed on original images.

and regions accurately identified whatever mean and standard deviation are. This demonstrates that our model produces improved segmentation performance (Figs. 2(c,d,g,h)).

In order to further test the method, we implemented it and presented our results on some true scene images (Fig. 3) as above. After computation and estimation, the bootstrap sample size $n_0 = 32 \times 32$ at scale 0 is enough to have the representation for each image. Segmentation results of these images using our method and mixture model are presented in Fig. 3. The segmentation using mixture model is very sensitive to noise and effects of lighting in images. It is obvious that a smoother and more precise segmentation results from our approach, and the method is robust to noise and effects of lighting in images. The analysis and experiments results are identical that the approach using bootstrap technique saves time considerably and the segmentation is precise. The reason for poor results of mixture model is that mixture model does not consider the dependence of neighboring pixels and spatial and each resolution relationships, and uses only intensity at finest multiresolution.

In conclusion, we define a GMLR for increasing the distinction between different signals by efficiently fusing more features. Based on GMLR test, we have proposed an efficient unsupervised image segmentation method, which combines bootstrap sampling technique with SVMMARP model to estimate all the parameters precisely and fast. The segmentation method takes advantage of resolution and region information of images, so it is not only precise and easy to perform but also robust to speckle. All the results support our analysis. The PDF in GMLR may be different types when image or image features are different. It is possible for wider application.

References

1. D.-B. Gu and J.-X. Sun, Vision, Image and Signal Processing, IEE Proceedings **152,** 184 (2005).
2. Z. H. Zhang, C. B. Chen, and K. L. Chan, Pattern Recognition **36,** 1973 (2003).
3. X. Yang and S. M. Krishnan, Image and Vision Computing **22,** 735 (2004).
4. P. Dickinson and A. Hunter, in *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance* 64 (2005).
5. M. L. Comer and E. J. Delp, IEEE Trans. Image Processing **8,** 408 (1999).
6. J. Aarroquin, S. K. Mitter, and T. Poggio, J. Amer. Stat. Assoc. **82,** 76 (1987).
7. H. Skriver, J. Schou, A. A. Nielsen, and K. Conradsen, in *Proceeding of IGARSS* 1011 (2002).
8. R. Cook, I. McConnell, C. Oliver, and E. Welbourne, Proc. SPIE **2316,** 92 (1994).
9. C. A. Glasbey, CVGIP: Graph. Models Image Processing **55,** 532 (1993).
10. J. Kittler and J. Illingworth, Pattern Recognition **19,** 41 (1986).
11. N. Otsu, IEEE Trans. Syst. Man. Cybern. **9,** 62 (1979).
12. N. R. Pal and S. K. Pal, Pattern Recognition **26,** 1277 (1993).
13. O. D. Trier and A. Jain, IEEE Trans. Pattern Anal. Mach. Intell. **17,** 1191 (1995).
14. R. C. Dubes and A. K. Jain, J. Appl. Stat. **16,** 131 (1989).
15. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification* (Wiley, New York, 2001).
16. Y. W. Ju and Z. Tian, Chin. J. Electronics **15,** 359 (2006).
17. H. L. Van Trees, *Detection, Estimation, and Modulation Theory* (Woly, New York, 2001) pp.46—96.
18. C. H. Fosgate, H. Krim, W. W. Irving, W. C. Karl, and A. S. Willsky, IEEE Trans. Image Processing **6,** 7 (1997).
19. B. Efron, Ann. Stat. **7,** 1 (1979).
20. I. Koch and G. Marshall, in *Proceedings of 13th International Conference on Pattern Recognition* 447 (1996).
21. S. V. Vijaya and M. N. Murty, Pattern Recognition **34,** 1047 (2001).
22. A. K. Jain, R. C. Dubes, and C. C. Chen, IEEE Trans. Pattern Anal. Mach. Intell. **9,** 628 (1987).
23. M. Zribi, Image and Vision Computing **22,** 1 (2004).
24. S. S. Gopal and T. J. Herbert, IEEE Trans. Image Processing **17,** 1014 (1998).