LABEL FIELD INITIALIZATION FOR MRF-BASED SONAR IMAGE SEGMENTATION BY SELECTIVE AUTOENCODING

Sanming Song, Bailu Si*, Xisheng Feng*

Sanming Song, *Bailu Si*[†], *Xisheng Feng*

Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China

ABSTRACT

The optimal solution of a Markov random field (MRF) can be solved by constructing a Markov chain that eventually goes to a balance state. However, in most situations, only an suboptimal solution can be obtained, because it is hard to choose the ideal initial state and the updating strategy. While the updating strategy has been extensively investigated, the initialization issue has been fully neglected. Though *k-means-clustering* has been used exclusively in initializing the label field, it suffers from the lack of account of the local constraints, which is the most essential part of the MRF model. A structural method based on selective autoencoding (SAE) is proposed for the label field initialization of MRF model in the task of sonar image segmentation. SAE is similar to the AutoEncoder, with the largest difference on the activation function, where a piece-wise sigmoid activation function with two different slop parameters is used to selectively encode image patches that resemble shadow areas or other areas. The synapse matrixes of SAE network act as information filters, preserve specific area adaptively and selectively, generating a label field that is much closer to the balance state. Experiments on sonar image segmentation demonstrate the efficiency of the SAE algorithm.

Index Terms— selective autoencoding, sonar image, label field, Markov random field

1. A GOOD BEGINNING IS HALF DONE

A statistical image segmentation task is to find the label field L^* that maximizes the posterior probability [1], $P_{L/X}\left(l/x\right) = \frac{P_L(l)P_{X/L}(x/l)}{P_X(x)}$, where X is an observed image and L is the label field. Now that $P_X\left(x\right)$ is fixed for a given image X, finding L^* is equivalent to finding the label field L which maximizes the following energy function (i.e. *MAP* estimation): $E = \ln P_L\left(l\right) + \ln P_{X/L}\left(x/l\right)$.

*Thanks to the support of China Postdoctoral Science Foundation (2014M561266), and Jiang Xinsong Innovation Fund (No.Y4FC012901).

†Acknowledges financial support from "Hundred Talents Program" of the Chinese Academy of Sciences (No.Y3F9060901) and the support of State Key Laboratory of Robotics, Shenyang Institute of Automation (No.Y5A1203901).

In sonar images, with the existence of strong speckle noises, the classification rules that solely based on global statistics, like maximizing the conditional probability $P_{X/L}\left(x/l\right)$, are insufficient to "clearly" segment foreground objects from the background. To introduce the local constraints, a Markov assumption is introduced as a kind of prior knowledge in MRF to model the local dependent relationships between the labels of neighboring pixels [1]. Intuitively, the prior probability $P_L\left(l\right)$ is used to "correct" the mistakes introduced by the conditional probability. MRF model is an ideal choice for texture image segmentation, like sonar images and geodesic images. The conditional distribution can be described by Weibull law, and the Markov assumption can be described by Gibbs distribution [2].

The MRF model aims at a balance between global optimization and local constraints. Such a delicate balance cannot be easily obtained by an explicit approach. Instead, an iterative optimization process, like iterated conditional method (ICM) [1], is necessary. Simply, ICM tries to construct a Markov chain that will eventually converge to a balance state. However, ICM tends to get into a local energy minimum quickly. Many other transition strategies, like Markov Chain Monte-Carlo (MCMC) [3] or simulated annealing (SA) [4], have been proposed to jump out of the local energy barricade.

However, the balance state of a dynamical system is not only determined by the dynamics equation, i.e. the transition strategy, but also largely depends on the initial state [5]. Unfortunately, the label field initialization issue in the MRF model has been fully neglected. So far, the *k-means* algorithm has been exclusively used to initialize the Markov chain. Though *k-means* algorithm is easy to be understood, it has the following drawbacks:

- *k-means* pursues the local optimization by greedy iteration. The initial centroids delineate the local searching boundaries. Therefore, *k-means* tends to find a local extremum.
- Essentially, the shape of the histogram determines the last segmentation result. *K-means* tries to choose several appropriate partition points, like the local minimum of the vallies, to divide the histogram. *K-means* algorithm will fail if the histogram has no obvious multiple peaks.

- *K-means* doesn't take the local configurations into account, leading to a hard segmentation.

The conditional probability parameters are directly estimated from the initial label field. An inappropriate initial label field contains a large number of misclassifications, which means that it will take a much longer time for the Markov chain to arrive the balance state.

In this paper, a selective autoencoding algorithm is proposed to binarize the sonar image. With the binary label field, we can then estimate the prior probability $P_L\left(l\right)$ and the conditional probability $P_{X/L}\left(x/l\right)$ and run the ICM iterations. SAE incorporates the local constraints in the initialization stage, exploiting the local spatial-structural dependent relationships by selective encoding. In this way, SAE generates a label field that is much closer to the balance state.

The SAE algorithm is described in Section 2 and the sonar image segmentations are shown in Section 3. We discuss related works in section 4 and conclude in Section 5.

2. SELECTIVE AUTOENCODING

The SAE network has the same structure as the AutoEncoder[6, 7]. An AutoEncoder is a feedforward, non-recurrent, three-layer neural network, trying to reconstruct the input (I) at the output layer (O) with the compressed representation in the hidden layer (H). In the training stage, each randomly sampled 8×8 image patch [8] is fed into the input layer. In the labelling stage, the output is binarized to generate a binary label field.

2.1. Training

Let x_i^I be the input, y_j^H and z_k^O be the output of the hidden layer and the output layer respectively, then

$$y_j^H = \mathcal{R}\left(v_j^H\right), z_k^O = \mathcal{R}\left(v_k^O\right), \tag{1}$$

where $v_j^H = \sum_i \omega_{ji}^{HI} x_i^I - b_j^H$, $v_k^O = \sum_j \omega_{kj}^{OH} y_j^H - b_k^O$. $\mathcal{R}\left(\cdot\right)$ is the activation function, ω and b stand for the weight and the bias.

The reconstruction error is $E = \frac{1}{2}\sum_k {e_k}^2 = \frac{1}{2}\sum_k \left(z_k^O - x_k^I\right)^2$.

The learning rules for ω_{kj}^{HO} and b_k^O are the same as the traditional BP (*back-propagation*) algorithm [9],

$$\Delta\omega_{kj}^{OH} = -\eta\frac{\partial E}{\partial\omega_{kj}^{OH}} = -\eta\delta_k y_j^H, \Delta b_k^O = -\eta\delta_k, \tag{2}$$

where $\delta_k = e_k\mathcal{R}'$. $\mathcal{R}'$ is the derivative, η is the learning rate.

Similarly, we derive the learning rules for ω_{ji}^{HI} and b_j^H,

$$\Delta\omega_{ji}^{HI} = -\eta\mathcal{R}'\delta_j x_i^I, \Delta b_j^H = -\eta\mathcal{R}'\delta_j, \tag{3}$$

where $\delta_j = \sum_k \omega_{kj}^{OH}\delta_k$.

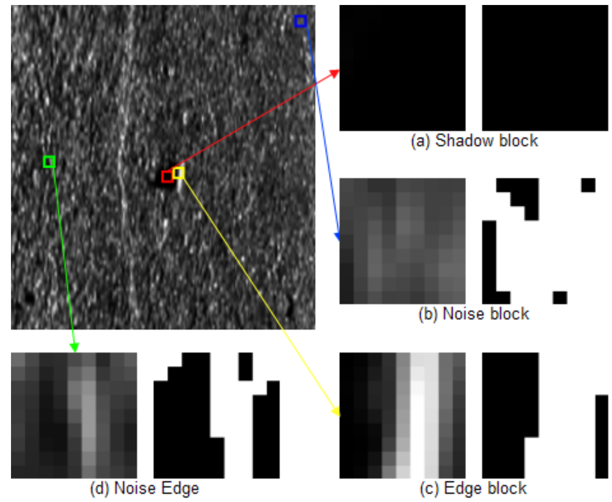

Fig. 1. Different kind of blocks. In panel (a)-(d), the left column is the megascopic image patch, the right column is the binarized result when the average of the image, $\bar{f}$, is removed. Note that, in the binarized patch, -1 is shown in black and 1 in white.

2.2. Activation function $\mathcal{R}$

A piecewise sigmoid function is designed when the following issues are considered.

Firstly, the shadow area (see Fig.1(a))should be preserved as much as possible, because it is the most salient feature that can be used to detect objects in the post-processing stages. The rule is applicable for the reverberation areas. Such requirements could be satisfied by any sigmoid function, because it saturates when the input is sufficiently large or small.

Secondly, *the negative branch of the activation function should saturate quickly*. Sonar image is full of speckle noises (see Fig.1(b)), we hope the weighted summation of the noisy shadow patch input, i.e. v_j^H and v_k^O, is close to the negative saturation areas. A larger β is helpful not only in getting rid of the noises in the region of interest (shadow area or meta-shadow area), but also in preserving more transition area between the shadow area and other areas.

Thirdly, *β in the negative axis should be larger than the positive axis*, because the image patch that contains more shadow area should be more precisely coded. For example, when an image patch is taken from the transition zone between the shadow area and the object area, the total energy of an image patch is very often to be a small positive value (see Fig.1(c)). Inversely, the total energy of a patch from the transition zone between the shadow and the reverberation area is more likely to be a small negative value (see Fig.1(d)).

Then, the activation function (see Fig.2) is

$$\mathcal{R}\left(\mathrm{x}\right) = \begin{cases} g\left(\mathrm{x}\right), & if\ \mathrm{x} > 0 \\ g\left(\beta\mathrm{x}\right), & if\ \mathrm{x} \leq 0 \end{cases} \tag{4}$$

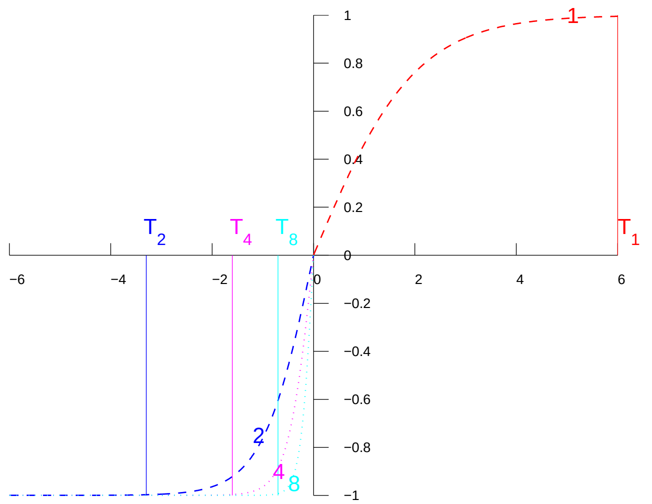

Fig. 2. The piecewise sigmoid function $\mathcal{R}\left(\mathrm{x}\right)$. $\beta = 1$ in the positive axis and $\beta > 1$ in the minus axis. T_β marks the saturation point.

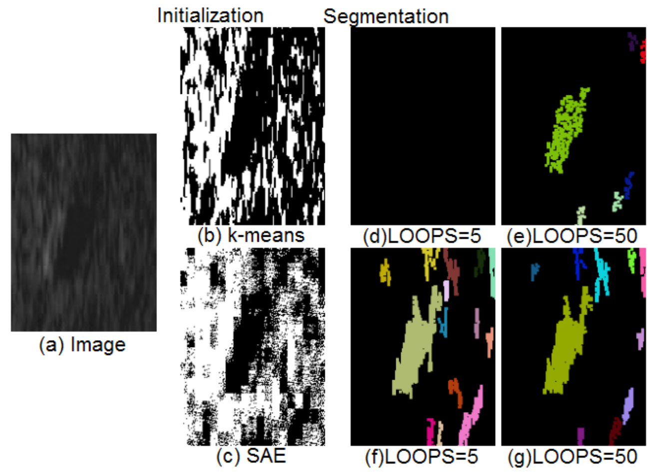

Fig. 3. Label field initialization and the MRF segmentation for an image taken by a forward-looking sonar.(a) is the original image, (b) and (c) is the initial label field obtained by *k-means* algorithm and the proposed SAE method respectively. The corresponding image segmentation results are shown in (d) and (f) for 5 loops, (e) and (g) for 50 loops."$LOOPS$" is the iterations in the ICM method.

where $(\beta > 1)$ is a selective factor, and

$$g\left(\mathrm{x}\right) = \frac{2}{1 + e^{-\mathrm{x}}} - 1. \tag{5}$$

The derivative is

$$\mathcal{R}'\left(\mathrm{x}\right) = \begin{cases} \frac{1}{2}\left(1 - g^2\left(\mathrm{x}\right)\right), & if\ \mathrm{x} > 0 \\ \frac{1}{2}\beta\left(1 - g^2\left(\beta\mathrm{x}\right)\right), & if\ \mathrm{x} \leq 0 \end{cases} \tag{6}$$

3. EXPERIMENTS

In this section, we compare the performance of the SAE method with the dominant *k-means-clustering* algorithm in the label field initialization. To further evaluate its benefits in MRF-based image segmentation, we run the ICM method on three different kinds of sonar images. Panels (a) of Fig.3-5 present three pictures taken by the forward-looking sonar, the side-scan sonar and the multi-beam high-resolution sonar respectively.

There are $N_I = 64$,$N_H = 8$,$N_O = 64$ neurons in layer I,H and O respectively. We experimentally set $\beta = 2$. 40 blocks are sampled and trained for *40* iterations.

3.1. Label field initialization

The label fields provided by the *k-means* algorithm and the SAE method are shown in panels (b) and (c) of Fig.3-5. Two observations can be drawn from the comparisons:

Firstly, the SAE method highlights the shadow area, which is the region of interest. For example, in Fig.3, the label field provided by the *k-means* algorithm, i.e. Fig.3(b), contains many "solid" noise blobs. But these blobs become much sparse when the SAE method is applied. The results show that the SAE method reduces the effects of speckle noises and preserves better structural information.

Secondly, the performance difference enlarges when the noise increases. For example, compared with the difference between 5(b) and (c), Fig.3(b) and (c) differs much, because the resolution of multi-beam sonar is far higher than the forward-looking sonar. It demonstrates that the SAE method is more reliable when strong noise exists.

3.2. Image Segmentation

The first task of the MRF model is to segment the sonar image into two areas, the shadow area and the seafloor reverberation area. Similar to [2], the conditional parameters of Weibull law are estimated by the *Maximum-Likelihood* method, and the prior parameters are estimated by the *Least-Square* method. The *Iterated Conditional* method is adopted to construct the Markov chain, i.e. update the label field state. In the post-processing stage, morphological operators, like image erosion and dilation are adopted sequentially to get rid of the pepper-and-salt noises. In the results, candidate areas larger than 100 pixels are displayed with different colors. Note that it can be extended to the Potts-MRF [10] by further dividing the positive branch .

As it can be seen from panels (d)-(g) of Fig.3-5, the SAE-initialized label field largely accelerates the segmentation process. For example, with the forward-looking sonar and the side-scan sonar, even after only $LOOPS = 5$ iterations, the SAE-initialized MRF returns acceptable segmentations (Fig.3(f) and 4(f)), while the *k-means*-initialized MRF re-

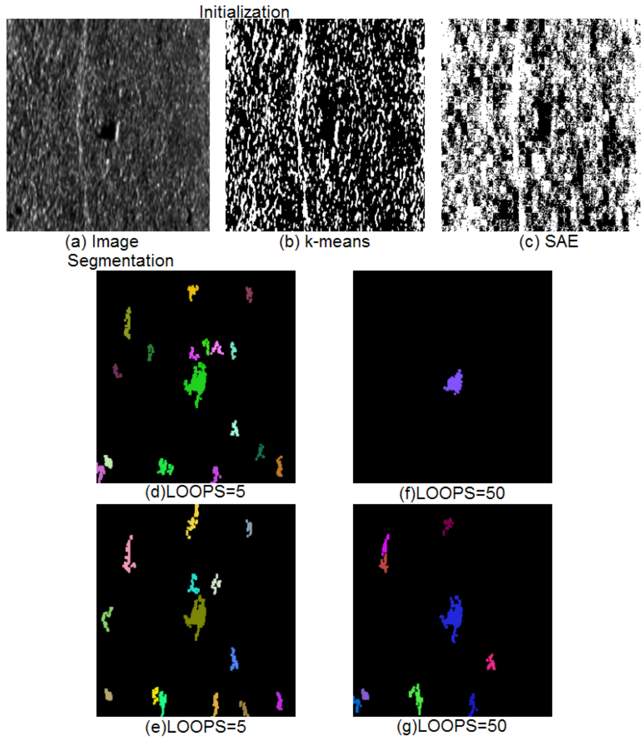

Fig. 4. Label field initialization and the MRF segmentation for an image taken by a side-scan sonar.

Table 1. The switching ratio of labels

	Fig. 3	Fig. 4	Fig. 5
k-means	51.39%	67.21%	45.77%
SAE	27.51%	42.11%	33.49%
$\downarrow$	***46.47%***	***37.35%***	***26.83%***

turns relatively worse results (subfig (d) of Fig.3 and 4). It demonstrates that the SAE method is able to provide a more appropriate initial state. After $LOOPS = 50$ iterations, the segmentation results returned by the SAE-initialized MRF contain less background noise blobs. Even with the same blob, the SAE-initialized MRF has smaller size. Therefore, the SAE-initialized Markov chain is able to arrive more optimized balance state.

The distance between the initial label field, L, and the final segmentation result, L_s, can be calculated by the percentage of switchings,

$$d = \frac{L \odot L_s}{N} \tag{7}$$

where $\odot$ is the *xnor* operator.

The switching ratio of Figs 3-5 with $k-means$ algorithm and the proposed SAE method are shown in Table.1. The mean relative reduction ratio is 36.88%, which demonstrates that SAE is able to generate a initial state that much closer to the balance state.

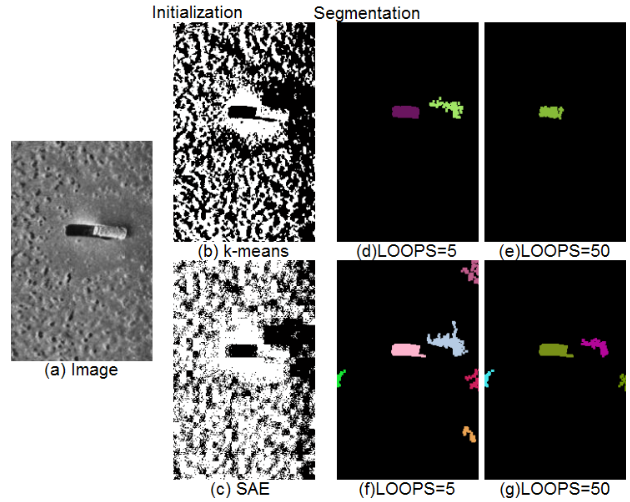

Fig. 5. Label field initialization and the MRF segmentation for an image taken by a multi-beam high-resolution sonar.

4. DISCUSSION

To our knowledge, this is the first work that devotes to the problem of label field initialization in the MRF model. Compared with the dominant $k-means$ algorithm [1, 2], the SAE algorithm takes structural information, i.e. local constraints, into consideration, generates an initial label field that is far closer to the balance state.

Though the SAE network is the same as an autoencoder[6], or a RBM [7], they differ much on the activation function and the binarization strategy. Firstly, a RBM learns statistical features from a large number of data samples. However, due to the limit of the sampling frequency of sonar equipments, SAE has to learn the structural information with very few sonar pictures. Secondly, in a RBM, the raw samples has to be pre-processed with normalization and a zero-phase crossing analysis (ZCA) filtering. Only the simple zero-mean processing is needed in the SAE method. Thirdly, the binary representation of a RBM is generated by alternating Gibbs sampling. However, in SAE, the binary output are obtained by a *sgn* function. Lastly, different activation strength is applied to different kinds of image patches selectively.

5. CONCLUSION

A label field initialization method for the MRF model based on selective autoencoding is proposed in the task of sonar image segmentation. SAE applies different activation functions to different kinds of areas. With an adaptive autoencoding mechanism, SAE takes local configuration constraints into the initial state, generating a label field that is closer to the balance state. Experiments on different kinds of sonar images demonstrate that SAE not only accelerates the segment process, but also converges to a more optimized balance state.

6. REFERENCES

[1] Haluk Derin and Howard Elliott, "Modeling and segmentation of noisy and textured images using gibbs random fields," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-9, no. 1, pp. 39–55, 1987.

[2] Max Mignotte, Christophe Collet, Patrick Perez, and Patrick Bouthemy, "Sonar image segmentation using an unsupervised hierarchical mrf model," *Image Processing, IEEE Transactions on*, vol. 9, no. 7, pp. 1216–1231, 2000.

[3] Marcelo Pereyra, Nicolas Dobigeon, Hadj Batatia, and Jean-Yves Tourneret, "Estimating the granularity coefficient of a potts-markov random field within a markov chain monte carlo algorithm," *Image Processing, IEEE Transactions on*, vol. 22, no. 6, pp. 2385–2397, 2013.

[4] Sridhar Lakshmanan and Haluk Derin, "Simultaneous parameter estimation and segmentation of gibbs random fields using simulated annealing," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 8, pp. 799–813, 1989.

[5] Steven H Strogatz, *Nonlinear dynamics and chaos: with applications to physics, biology and chemistry*, Perseus publishing, 2001.

[6] Geoffrey E Hinton and Richard S Zemel, "Autoencoders, minimum description length, and helmholtz free energy," *Advances in neural information processing systems*, pp. 3–3, 1994.

[7] Tijmen Tieleman, "Training restricted boltzmann machines using approximations to the likelihood gradient," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1064–1071.

[8] Garrison W Cottrell, Paul Munro, and David Zipser, "Learning internal representations from gray-scale images: An example of extensional programming," in *Ninth annual conference of the cognitive science society*, 1987, pp. 462–473.

[9] S Haykin, *Neural Networks: A Comprehensive Foundation*, Printice-Hall, Inc., New Jersey, 1999.

[10] Xavier Descombes, Robin D Morris, Josiane Zerubia, and Marc Berthod, "Estimation of markov random field prior parameters using markov chain monte carlo maximum likelihood," *Image Processing, IEEE Transactions on*, vol. 8, no. 7, pp. 954–963, 1999.