Discriminative Image Thresholding

Jing-Hao Xue [a,*], D. Michael Titterington [a]

[a] *Department of Statistics, University of Glasgow, Glasgow G12 8QQ, UK*

Abstract

In this paper, we present discriminative approaches to histogram-based image thresholding, in which the optimal threshold is derived from the maximum likelihood based on the conditional distribution $p(y|x)$ of y, the class indicator of a grey level x, given x. The discriminative approaches can be regarded as discriminative extensions of the traditional generative approaches to thresholding, such as Otsu's method and Kittler and Illingworth's minimum error thresholding (MET). The generative approaches assume a model for the data-generating process for each class whereas the discriminative approaches do not. As illustrations, we develop discriminative versions of Otsu's method and MET by using discriminant functions corresponding to the original methods to represent $p(y|x)$. These two discriminative thresholding approaches are compared with their original counterparts on selecting thresholds for a variety of histograms of mixture distributions. Results show that the discriminative Otsu method consistently provides relatively good performance. Although being of higher computational complexity than the original methods in parameter estimation, robustness and model simplicity can justify the discriminative Otsu method for scenarios in which the risk of model mis-specification is high and the computation is not demanding.

Key words: Discriminative/Generative image thresholding; Logistic regression; Minimum error thresholding (MET); Mixture distributions; Otsu's method

1 Introduction

Image thresholding is a simple and widely-used technique for segmentation, partitioning a grey-level image into segments corresponding to different classes [1–3], given that the classes to some extent can be distinguished by their grey levels. Most thresholding approaches are proposed for two-class binarisation and are based on the grey-level histogram of an image [1, 3–5]. Two of the most popular approaches are Otsu's method [6] and Kittler and Illingworth's minimum error thresholding (MET) [7].

Given an image of N pixels, Otsu's method selects the optimal threshold t^* as

$$t^* = \underset{t\in[0,T-1]}{\operatorname{argmin}} \ \sigma_w^2(t) = \pi_0(t)\sigma_0^2(t) + \pi_1(t)\sigma_1^2(t) \ ,$$

where $[0, T]$ is the range of grey level, and $\pi_0(t)$ and $\sigma_0(t)$ are respectively the proportion of and standard deviation within class $\mathcal{C}_0(t)$, where $\mathcal{C}_0(t)$ includes all the pixels with grey levels x less than t, i.e., $\mathcal{C}_0(t) = \{i : 0 \leq x_i \leq t, 1 \leq i \leq N\}$; $\pi_1(t)$, $\sigma_1(t)$ and $\mathcal{C}_1(t)$ are defined similarly for the remaining pixels, and thus $\sigma_w^2(t)$ is called within-class variance. The MET method selects t^* as

$$t^* = \underset{t\in[0,T-1]}{\operatorname{argmin}} \ \pi_0(t)\log\frac{\sigma_0(t)}{\pi_0(t)} + \pi_1(t)\log\frac{\sigma_1(t)}{\pi_1(t)} \ ,$$

where $\pi_y \neq 0, y = 0, 1$, and in practice σ_y is nonzero. Research efforts have been made to unify these two approaches [8, 9].

* Corresponding author. Tel.: +44 141 330 2474; fax: +44 141 330 4814.
Email addresses: jinghao@stats.gla.ac.uk (Jing-Hao Xue),
mike@stats.gla.ac.uk (D. Michael Titterington).

Ref. [8] shows that Otsu's method is equivalent to maximisation of the log-likelihood based on the conditional distribution $p(x|y)$, where x is the grey level and $y \in \{0, 1\}$ is the class indicator corresponding to x, under the assumption that the grey level within each class (denoted by $x|y$) follows a normal distributions $\mathcal{N}(\mu_y, \sigma_y^2)$ and $\sigma_0^2 = \sigma_1^2$. Ref. [8] also shows that MET is equivalent to maximisation of the log-likelihood based on the joint distribution $p(x, y)$, under the assumption that $x|y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ and $\sigma_0^2 \neq \sigma_1^2$. Since $p(x, y) = \pi_y p(x|y)$, where $\pi_y = p(y)$, Otsu's method is also equivalent to maximisation of the log-likelihood based on $p(x, y)$ with $\pi_0 = \pi_1 = 0.5$. In this sense, both Otsu's method and MET assume a data-generating process (DGP) $p(x, y)$; therefore, we call such approaches generative thresholding approaches. As with Fisher's linear discriminant, the Otsu's original method does not assume normally distributed classes or that $\sigma_0^2 = \sigma_1^2$; therefore, hereafter we refer, as Otsu's method, to the generative method to which it is equivalent, shown in [8].

Since $p(x, y) = p(x)p(y|x) \propto p(y|x)$, the MET method is also equivalent to minimisation of the logistic loss, which is based on $-\log p(y|x)$. Meanwhile, under the assumption of normal distributions, both Otsu's method and MET are equivalent to minimisation of the expected misclassification error rate. In other words, both methods seek t^* such that $p(\mathcal{C}_1(t^*)|x = t^*) = p(\mathcal{C}_0(t^*)|x = t^*)$, leading to alternative iterative implementations by solving

$$\log\{p(\mathcal{C}_1(t)|x)/p(\mathcal{C}_0(t)|x)\} = 0$$

for x and then updating t, $p(\mathcal{C}_1(t)$ and $p(\mathcal{C}_0(t)$ in each iteration [7, 10].

For both Otsu's method and MET, the grey-level histogram is assumed to be an empirical realisation of a two-component normal mixture. However, such an assumption often cannot be guaranteed for real images, leading to a major potential risk of model mis-specification when generative thresholding is applied. In two-class discrimination, there are discriminative approaches which do not assume any DGP and which can be less sensitive to model mis-specification than are corresponding generative approaches [11,12]. Therefore, in this paper, we present discriminative approaches to histogram-based image thresholding. The optimal threshold is derived from the maximum log-likelihood based on the conditional distribution $p(y|x)$. The discriminative approaches can be regarded as discriminative extensions of the traditional generative approaches to thresholding, such as Otsu's method and MET.

2 Discriminative thresholding

For two-class discrimination, in terms of minimum misclassification error rate, an optimal discriminant criterion for classifying an observation x into class $\mathcal{C}_1$ with $y = 1$ (or $\mathcal{C}_0$ with $y = 0$) is a discriminant function $g(x, \alpha) = \log\{p(\mathcal{C}_1|x)/p(\mathcal{C}_0|x)\} > 0$ (or ≤ 0). For a pixel in grey-level images, x is in general its grey level as a scalar. The most widely used discriminant functions are a linear function $g(x, \alpha) = \beta_0 + \beta_1 x$, where $\alpha = (\beta_0, \beta_1)^T$, and a quadratic function $g(x, \alpha) = \beta_0 + \beta_1 x + \beta_2 x^2$, where $\alpha = (\beta_0, \beta_1, \beta_2)^T$.

The $g(x, \alpha)$ can be derived from a generative classifier, such as normal-based linear/quadratic discriminant analysis where $\mathcal{N}(\mu_y, \sigma_y^2)$ is assumed as the DGP for class y and where it is assumed that $\sigma_0^2 = \sigma_1^2$ for the linear case and $\sigma_0^2 \neq \sigma_1^2$ for the quadratic case. It can also be derived from a discriminative classifier, such as linear/quadratic logistic regression, in which no DGP is assumed.

Here we derive a discriminative thresholding approach from maximisation of the log-likelihood based on the conditional distribution $p(y|x)$, which can be represented as a function of $g(x, \alpha)$.

As $g(x, \alpha) = \log\{p(y = 1|x)/p(y = 0|x)\}$, after some algebra we obtain

$$p(y = 1|x) = e^{g(x,\alpha)} / \left(1 + e^{g(x,\alpha)}\right) \ , \ p(y = 0|x) = 1 / \left(1 + e^{g(x,\alpha)}\right) \ .$$

It follows that, for an image of N pixels $\{(x_i, y_i)\}_{i=1}^{N}$, where x_i and y_i are the grey level and class indicator of the i-th pixel, the log-likelihood $\ell(\alpha)$ based on $p(y_i|x_i)$ is

$$\ell(\alpha) = \sum_{i=1}^{N} g(x_i, \alpha) y_i - \sum_{i=1}^{N} \log\left(1 + e^{g(x_i,\alpha)}\right) \ .$$

Let $h(x)$, $x = 0, \cdots, T$, denote the grey-level histogram constructed from the N pixels. For histogram-based thresholding, a threshold t partitions $h(x)$ into two sets of grey levels and thus partitions the image into two classes of pixels, denoted by $\mathcal{C}_0(t)$ and $\mathcal{C}_1(t)$, such that $y_i = 0$ if $x_i \leq t$ and $y_i = 1$ otherwise. As y_i changes with t, and the parameter α of $g(x, \alpha)$ is estimated from $\{(x_i, y_i)\}_{i=1}^{N}$ by maximisation of $\ell(\alpha)$, we write $g(x, \alpha)$ as $g(x, \alpha(t))$ and $\ell(\alpha)$ can be rewritten as

$$\ell(\alpha(t)) = \sum_{x=t+1}^{T} h(x) g(x, \alpha(t)) - \sum_{x=1}^{T} h(x) \log\left(1 + e^{g(x,\alpha(t))}\right) \ .$$

In this context, the optimal threshold t^* can be determined discriminatively as

$$t^* = \underset{t}{\operatorname{argmax}} \ \ell(\hat{\alpha}(t)) \ ,$$

where $\hat{\alpha}(t)$, estimated from $\mathcal{C}_0(t)$ and $\mathcal{C}_1(t)$, is the maximum-likelihood estimator of α for a threshold t. Estimation of $\alpha(t)$ proceeds similarly to that for logistic regression models, using $\mathcal{C}_0(t)$ and $\mathcal{C}_1(t)$ as the training set. As there is no convenient analytical solution for α, discriminative thresholding is of higher computational complexity than generative thresholding.

The multi-threshold extensions of the discriminative thresholding approaches can be obtained by using the log-likelihood for a multinomial logit model, which is the multi-class generalisation of logistic regression.

When the DGP is known, a generative approach is to be preferred in general. However, for real-world application, the DGP is always unknown, in which case a generative approach has to assume a specific DGP. For different assumptions of the DGP, a generative approach can have different variants. For example, variants of MET include those for Poisson [13], Rayleigh [14], Nakagami-Gamma, Weibull, and log-normal distributions [15].

In contrast to generative thresholding, a discriminative approach to thresholding assumes the discriminant function $g(x, \alpha)$ rather than the DGP, and this may lead to more robust performance against the model mis-specification. As parameter estimation within discriminative approaches is in general harder than that in generative approaches [11], the computational complexity of discriminative thresholding is in general higher than that of generative thresholding, as in our implementation below.

For illustration, we present two discriminative thresholding approaches, which have the same formula but different α for $g(x, \alpha)$ as those for Otsu's method and MET, respectively.

As Otsu's method corresponds to a linear discriminant function and MET corresponds to a quadratic, we define the discriminative Otsu method as

$$t^* = \underset{t}{\mathrm{argmax}}\ \ell(\hat{\alpha}(t)) \text{ with } g(x,\alpha(t)) = \beta_0(t) + \beta_1(t)x\ ,$$

and the discriminative MET as

$$t^* = \underset{t}{\mathrm{argmax}}\ \ell(\hat{\alpha}(t)) \text{ with } g(x,\alpha(t)) = \beta_0(t) + \beta_1(t)x + \beta_2(t)x^2\ .$$

3 Experiments with discriminative thresholding

In this section, we compare the performance of generative and discriminative versions of Otsu's method and MET. Comparison of approaches to image thresholding requires appropriate evaluation method, and numerous methods have been developed based on various criteria [1, 3, 16, 17]. Roughly speaking, supervised evaluation is subjective, requiring a pre-segmented image as ground-truth; unsupervised evaluation is objective but prefers an approach appropriate for the underlying evaluation criteria.

As with Refs. [7] and [8], we compare the thresholding approaches by using histograms constructed from simulated data. The data for each class are simulated from normal, Poisson, log-normal and two-component normal mixture distributions. Normal distributions are, as used for Otsu's method and MET [8], the most-commonly used distributions in image processing; Poisson distributions are justified based on a theory of image formation [13]; log-normal distributions are used as heavy-tailed adaptions of Rayleigh distributions for the thresholding of synthetic aperture radar (SAR) amplitude images [15]; and, compared to normal, Poisson and log-normal distributions, a normal mixture can be a better approximation to the distribution of a class in the histogram.

Although, in our scenario, the underlying distributions for the simulated data are known, they are unknown for real images. Therefore, we do not compare discriminant thresholding approaches versus a generative thresholding approach developed for a specific distribution, such as MET for Poisson distributions in Ref. [13] or for log-normal distributions in Ref. [15].

For Otsu's method and MET, normally distributed classes can satisfy the underlying assumptions, while neither Poisson nor log-normally distributed data satisfy the assumptions.

For discriminant thresholding, as normal distributions are exponential families in canonical form, they satisfy the linear or quadratic formulation of $g(x, \alpha(t))$. Although Poisson distributions are also exponential families in canonical form, because of the equivalence of mean and variance, they only satisfy the linear formulation of $g(x, \alpha(t))$. Log-normal distributions are exponential families but not in canonical form; hence, they and normal mixture distributions satisfy neither the linear nor the quadratic formulation of $g(x, \alpha(t))$.

For Otsu's method and MET, the estimator of the parameter $\theta = (\pi_y, \mu_y, \sigma_y^2)^T$ is the maximum-likelihood estimator based on $p(x, y)$, which can be calculated directly from the histogram as in [6–8]. The thresholds obtained are denoted by t_O and t_M, respectively.

For discriminant thresholding, as for logistic regression, the estimator of the parameter α is implemented by an R function *glm* (from a standard package **stats**), which uses an iteratively re-weighted least squares algorithm to fit the model. The thresholds obtained are denoted by d_O and d_M, respectively.

We make following comments about our implementation. First, in order to avoid $\sigma_y = 0$, which may cause failure of MET, we only search for thresholds within the $[1, 99]$ percentile range of histograms. Secondly, since grey levels are in range of $[0, T]$, we left-truncate and right-censor the simulated data into that range.

We simulate six datasets, each with $10,000$ pixels, and set $T = 255$ as for 8-bit grey-level images. The datasets for normal distributions are unbalanced in terms of class proportions, while others are balanced. The setting of our simulated data is as follows.

The two datasets for normal distributions are the same as those used by [8]: one has $\pi_1 = 0.05$, $\mu_1 = 50$, $\mu_2 = 150$ and $\sigma_1 = \sigma_2 = 18$; the other has $\pi_1 = 0.25$, $\mu_1 = 38$, $\mu_2 = 121$, $\sigma_1 = 8$ and $\sigma_2 = 40$.

As a Poisson distribution can be well approximated by a normal distribution when its mean is larger, such as 10, as with [13], we simulate pixels with low grey levels. The dataset for Poisson distributions has $\mu_1 = 5$, $\mu_2 = 20$. As the mean is equal to the variance for Poisson distributions, the two classes have unequal variances.

The dataset for log-normal distributions has logarithms having $\mu_1 = 2$, $\mu_2 = 4$, $\sigma_1 = 1/2$ and $\sigma_2 = 1/4$.

One of the two datasets for normal mixture distributions has four components, two for each class with equal mixing weights. The two components $\mathcal{N}(\mu_{1,a}, \sigma_1^2)$ and $\mathcal{N}(\mu_{1,b}, \sigma_1^2)$ for the first mixture are specified with $\mu_{1,a} = 60$ and $\mu_{1,b} = 80$; and the two components $\mathcal{N}(\mu_{2,a}, \sigma_2^2)$ and $\mathcal{N}(\mu_{2,b}, \sigma_2^2)$ for the second mixture are specified with $\mu_{2,a} = 120$ and $\mu_{2,b} = 140$. In addition, $\sigma_1^2 = \sigma_2^2 = 10$, and hence the two classes have equal variances. The other dataset for normal mixture distributions is the same as the previous one but with $\sigma_1^2 = 5$ and $\sigma_2^2 = 15$, and hence the two classes have unequal variances.

The thresholding results for these six datasets are shown in Figure 1. We observe the following.

For the datasets from normal distributions, where the histograms are themselves normal mixtures, the discriminative Otsu method (d_O) gives almost the same results as MET (t_M), which is better than the Otsu's original method (t_O) [8] and the discriminative MET d_M. The same phenomenon appears for the Poisson dataset. For other three datasets, all the four methods of study show the similar thresholds and thus comparable performance.

Note that, for all six datasets, although the discriminative MET does not provide satisfactory results, the discriminative Otsu method consistently provides relatively good performance, compared to the original methods. In terms of the level of computational complexity, that of the discriminative Otsu method, which corresponds to a linear discriminant function, is lower than that of the discriminative MET, which corresponds to a quadratic, whereas those of both discriminative approaches are higher than those of the original approaches in parameter estimation.

4 Conclusions

The discriminative approach to histogram-based image thresholding proposed in this paper is based on maximum likelihood corresponding to the conditional distribution $p(y|x)$, rather than $p(x, y)$ as in the case of the traditional generative thresholding. For our simulated datasets, results show that the dis-

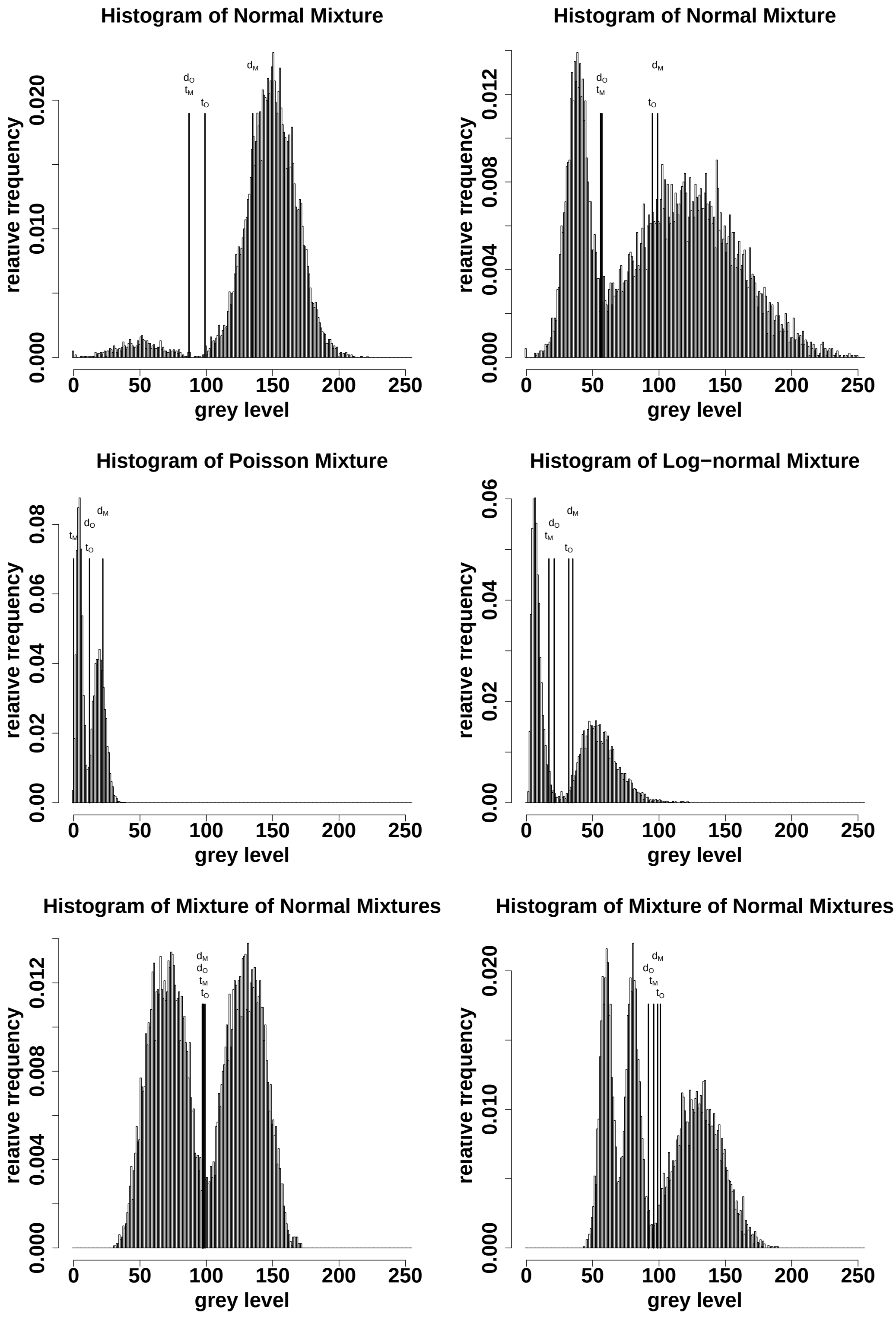

Figure 1. Thresholding results for 6 simulated datasets. Here t_O, t_M, d_O and d_M are thresholds from Otsu's method, MET and their discriminative counterparts, respectively.

criminative Otsu method consistently provides relatively good performance. Considering its robustness and model simplicity, we suggest the use of the discriminative Otsu method for scenarios in which the Otsu's original method and MET do not perform well due to model mis-specification and in which the computation is not demanding.

References

[1] P. K. Sahoo, S. Soltani, A. K. Wong, Y. C. Chen, A survey of thresholding techniques, Comput. Vision Graph. Image Process. 41 (2) (1988) 233–260.

[2] N. Pal, S. Pal, A review on image segmentation techniques, Pattern Recogn. 26 (9) 1277–1294.

[3] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, Journal of Electronic Imaging 13 (1) (2004) 146–165.

[4] C. A. Glasbey, An analysis of histogram-based thresholding algorithms, CVGIP: Graph. Models Image Process. 55 (6) (1993) 532–537.

[5] Ø. D. Trier, A. K. Jain, Goal-directed evaluation of binarization methods, IEEE Trans. Pattern Anal. Mach. Intell. 17 (12) (1995) 1191–1201.

[6] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Systems Man Cybernet. SMC-9 (1979) 62–66.

[7] J. Kittler, J. Illingworth, Minimum error thresholding, Pattern Recogn. 19 (1) (1986) 41–47.

[8] T. Kurita, N. Otsu, N. Abdelmalek, Maximum likelihood thresholding based on population of mixture models, Pattern Recogn. 25 (10) (1992) 1231–1240.

[9] H. Yan, Unified formulation of a class of image thresholding techniques, Pattern Recogn. 29 (12) (1996) 2025–2032.

[10] R. C. Gonzalez, R. E. Woods, Digital Image Processing, 2nd Edition, Prentice Hall, 2002.

[11] Y. D. Rubinstein, T. Hastie, Discriminative vs. informative learning, in: KDD, 1997, pp. 49–53.

[12] A. Y. Ng, M. I. Jordan, On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes, in: NIPS, 2001, pp. 841–848.

[13] N. R. Pal, D. Bhandari, Image thresholding: some new techniques, Signal Process. 33 (2) (1993) 139–158.

[14] J.-H. Xue, Y. J. Zhang, X. G. Lin, Rayleigh-distribution based minimum error thresholding for SAR images, Journal of Electronics (China) 16 (4) (1999) 336–342.

[15] G. Moser, S. Serpico, Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery, IEEE Transactions on Geoscience and Remote Sensing 44 (10) (2006) 2972–2982.

[16] Y. J. Zhang, A survey on evaluation methods for image segmentation, Pattern Recogn. 29 (8) (1996) 1335–1346.

[17] H. Zhang, J. E. Fritts, S. A. Goldman, Image segmentation evaluation: a survey of unsupervised methods, Comput. Vis. Image Understand. doi:10.1016/j.cviu.2007.08.003.