

# Two New Scale-Adapted Texture Descriptors for Image Segmentation

Miguel Angel Lozano and Francisco Escolano

Robot Vision Group  
Departamento de Ciencia de la Computación e Inteligencia Artificial  
Universidad de Alicante, Spain  
{malozano, sco}@dccia.ua.es  
<http://www.rvg.ua.es>

**Abstract.** In texture segmentation it is key to develop descriptors which provide acceptable results without a significant increment of their temporal complexity. In this contribution, we propose two probabilistic texture descriptors: polarity and texture contrast. These descriptors are related to the entropy of both the local distributions of gradient orientation and magnitude. As such descriptors are scale-dependent, we propose a simple method for selecting the optimal scale. Using the features at their optimal scale, we test the performance of these measures with an adaptive version of the ACM clustering method, in which adaptation relies on the Kolmogorov-Smirnov test. Our results with only these two descriptors are very promising.

## 1 Introduction

In the past, there have been many approaches to texture description: Gabor filters [6], quadrature filters [7], co-occurrence matrices [8], wavelets [9], second-order eigenstructure [10], and so on. As texture is not a pointwise feature but relies on a local neighborhood, there are two key problems to consider: (i) Find a good descriptor, like the ones listed above, and (ii) determine the optimal size of the neighborhood where such a descriptor is computed. In this paper, we address these two questions starting by revising two measures, polarity and texture contrast, which rely on the second-order eigenstructure. Later, we redefine such measures in terms of entropy and propose a way of automatically selecting the optimal scale of the measures. Finally, we test these measures in segmentation.

The polarity  $P_\sigma$  at a given pixel is defined in [1] as a measure of the extent to which the gradient vectors  $\nabla I$  in a certain neighborhood defined by the scale  $\sigma$  all point in the same direction:

$$P_\sigma = \frac{|E_+ - E_-|}{E_+ + E_-}, \quad (1)$$

where  $E_+$  and  $E_-$  are defined as follows

$$E_+ = \sum_{x,y} G_\sigma(x,y)[\nabla I \cdot \hat{n}]_+ \quad (2)$$

and

$$E_- = \sum_{x,y} G_\sigma(x,y) [\nabla I \cdot \hat{n}]_- , \quad (3)$$

where  $G_\sigma(\cdot)$  is a Gaussian smoothing kernel with variance  $\sigma^2$ ,  $[\cdot]_+$  and  $[\cdot]_-$  are the rectified positive and negative parts of their arguments, and  $\hat{n}$  is a unit vector perpendicular to  $\phi$ , the dominant direction in the neighborhood, which in turn is the argument of the principal eigenvector of the second-moment matrix

$$M_\sigma = \sum_{x,y} G_\sigma(x,y) (\nabla I) (\nabla I)^T . \quad (4)$$

Consequently,  $E_+$  and  $E_-$  measure, respectively, how many gradient vectors in the window defined by  $G_\sigma(\cdot)$  are in the positive side and negative side of  $\phi$ , and  $P_\sigma \in [0, 1]$ , will be close to zero when  $E_+ \approx E_-$ , that is, when we have a flow pattern; and it will be close to the unit for instance when  $E_- \approx 0$  and  $E_+ \neq 0$ , that is, when we have an edge.

## 2 Entropy-Related Measures

### 2.1 Probabilistic Polarity

As the underlying idea of polarity is to vanish as many different orientations appear in the neighborhood, we propose an alternative definition of polarity which does not rely on the eigenstructure of the local gradient, but on the structure of the distribution of local gradient orientations. Thus, the probability  $p_\sigma(z)$  of a given orientation  $z \in [0, 2\pi)$  at scale  $\sigma$  is defined by

$$p_\sigma(z) = \frac{\sum_{x,y} \delta(z - \theta(x,y)) W(x,y)}{\sum_w \{ \sum_{x,y} \delta(w - \theta(x,y)) W(x,y) \}} , \quad (5)$$

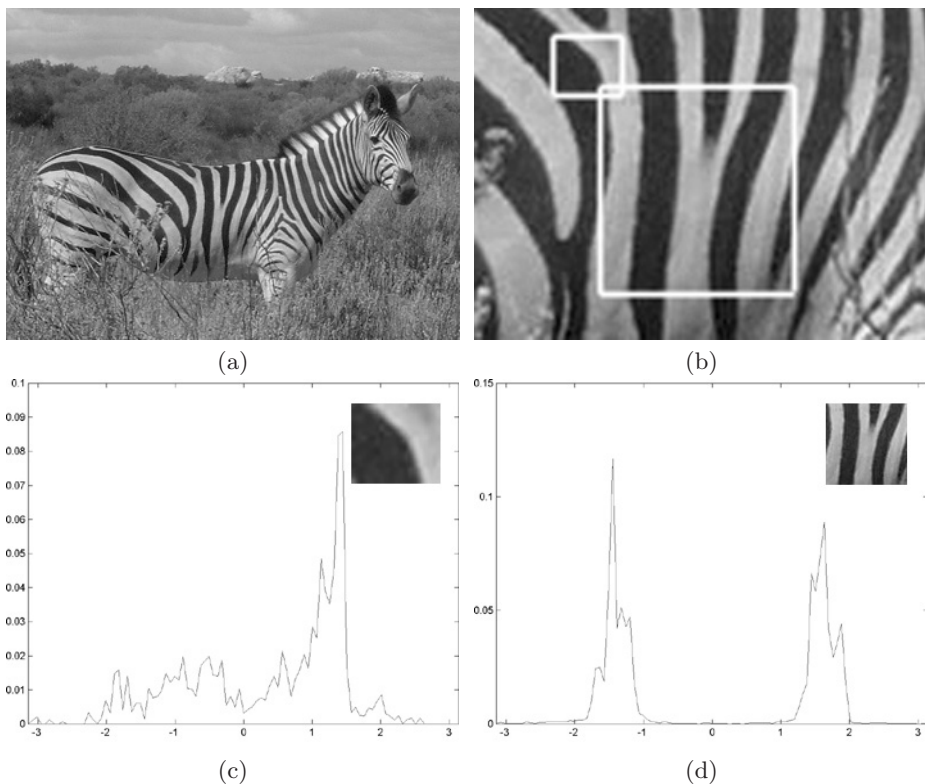
where

$$W(x,y) = G_\sigma(x,y) \|\nabla I\| , \theta(x,y) = \arctan I_y / I_x ,$$

that is, the weight of a given pixel in the neighborhood and the local orientation of its gradient, respectively. After quantizing the interval  $[0, 2\pi)$  into  $N$  bins of size  $\Delta = 2\pi/N$ , we define the empirical probability  $h_\sigma(k)$ , with  $k = 0, 1, \dots, N-1$ , which accumulates all probabilities  $p_\sigma(z) \in [k\Delta, (k+1)\Delta)$ . Using the latter  $N$ -component histogram, the entropy of the distribution is approximated by

$$H_\sigma = \sum_{k=0}^{N-1} h_\sigma(k) \log h_\sigma(k) . \quad (6)$$

In principle, the inverse entropy  $1 - H_\sigma$  is a good measure of polarity because it tends to the unit when all gradient vectors in the neighborhood have a similar orientation (minimal entropy, corresponding to a peaked distribution) and it vanishes when many different orientations appear (maximal entropy, corresponding to a uniform distribution). However, the inverse entropy decays too



**Fig. 1.** (a) Input image (b) Zoom showing both a polarized and a de-polarized zone (c) Histogram of the polarized zone:  $1 - H_\sigma = 0.2568$ ,  $P_\sigma = 0.2568$  (d) Histogram of the de-polarized zone:  $1 - H_\sigma = 0.3722$ ,  $P_\sigma = 0.0079$

slowly as the neighborhood is de-polarized. Particularly, a distribution with two closer peaks (or one wider peak) and one with the same peaks but distant, have similar entropies. Consequently, the latter measure captures the number of peaks but not their separation, and such a separation, in addition to the appearance of new peaks, occurs when we progressively de-polarize a texture edge while incrementing the size of its neighborhood. For instance, Fig. 1 polarity vanishes when two significant peaks appear in a de-polarized zone, whereas inverse entropy even gets incremented in the same zone.

In order to capture peak separation we re-define polarity in terms of the expression

$$P_\sigma = 1 - \sum_{k=0}^{N-1} h_\sigma(k) \sum_{r=-\lfloor N/2 \rfloor}^{\lfloor (N-1)/2 \rfloor} g(r) h_\sigma([k+r]_N), \tag{7}$$

where each component  $h_\sigma(k)$  is no longer weighted by its logarithm but by the result of convolving it with a kernel  $g(\cdot)$  defined in such a way that we ensure that  $P_\sigma \in [0, 1]$ . For a linear choice we have that

$$g(r) = ar = \frac{n}{\sum_{i=-\lfloor N/2 \rfloor}^{\lfloor (N-1)/2 \rfloor} |i|} r . \tag{8}$$

We also assume a cyclic histogram, because the orientation domain is also cyclic, where  $[k + r]_N \in \{0, 1, \dots, N - 1\}$  refers to  $(k + r) \bmod N$ .

### 2.2 Probabilistic Texture Contrast

Another texture feature is texture contrast. In [1] it is defined by  $2\sqrt{\lambda_1 + \lambda_2}$ , where  $\lambda_1$  and  $\lambda_2$  are the two eigenvalues of the second-moment matrix  $M_\sigma$ . Following the probabilistic rationale above, and in order to define texture contrast we consider the local intensity probabilities

$$q_\sigma(z) = \frac{\sum_{x,y} \delta(z - I(x,y)) G_\sigma(x,y)}{\sum_w \{ \sum_{x,y} \delta(w - I(x,y)) G_\sigma(x,y) \}} , \tag{9}$$

and proceed to quantize the normalized range of intensities  $[0, 1]$  yielding the  $M$ -component histogram  $c_\sigma(i)$ ,  $i = 0, 1, \dots, M - 1$ , which accumulates the probabilities  $q_\sigma(z)$ . Texture contrast must be close to the unit when we have two peaks at maximal distance, must vanish when the intensity distribution is peaked. Again, we find that the entropy is not a proper choice and we replace it by

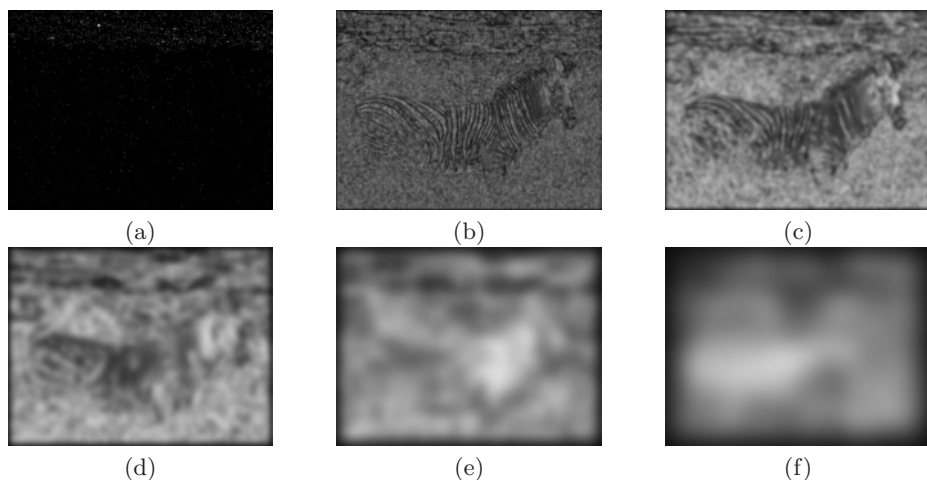
$$C_\sigma = \sum_{i=0}^{M-1} c_\sigma(i) \sum_{j=0}^{M-1} d(|i - j|) c_\sigma(j) , \tag{10}$$

where  $d(\cdot)$  is defined in such a way that  $C_\sigma \in [0, 1]$ . For the simple linear case, we have that  $d(r) = 2r/(N - 1)$ . With the latter definition we consider peak separation through a weighted correlation.

### 3 Scale Selection

As the probabilistic measures defined above depend on the scale, we are interested in a method for selecting them optimal scale for both of them. Few previous work has been done in this area [2]. However we follow the approach described in [1] and scale selection relies on polarity analysis. We will consider a sequence of scales  $\{\sigma_k\}$ , with  $k = 0, 1, 2, \dots, S$  and we will start by computing the polarity at the lower scale  $P_{\sigma_0}$  and assuming that the tentative optimal scale is  $\sigma_0$ . Thereafter, we will test whether an increment of scale is acceptable.

An increment of scale will always contribute either to de-polarize the pixel, or to leave its polarity unchanged. Consequently, given  $P_{\sigma_k}$ , and assuming that the temporary optimal scale is  $\sigma_k$ , we compute  $P_{\sigma_{k+1}}$ , and test whether the decrement  $P_{\sigma_k} - P_{\sigma_{k-1}} = \nabla P_{\sigma_k} \leq 0$  is *low enough*. If  $\nabla P_{\sigma_{k+1}} < \nu \nabla P_{\sigma_k}$ , with  $\nu \in [0, 1]$  then we will accept  $\sigma_{k+1}$  as a new temporary optimal scale because such scale de-polarizes the pixel significantly. Otherwise, we will assume that the optimal scale  $\sigma^* = \sigma_k$ . The coefficient  $\nu$  modulates the decrement needed to



**Fig. 2.** Polarity at different scales. (a)  $\sigma = 0.25$  (b)  $\sigma = 0.5$  (c)  $\sigma = 1.0$  (d)  $\sigma = 2.0$  (e)  $\sigma = 4.0$  (f)  $\sigma = 8.0$

increment the scale: When  $\nu \rightarrow 0$  we will change easily of scale, whereas with  $\nu \rightarrow 1$  we will be more restrictive.

In our experiments we have the set of scales  $\{0.25, 0.5, 1.0, 2.0, 4.0, 8.0\}$  and we have set  $\nu = 0.5$ , that is, we set  $S = 6$ . In Fig. 2, we show the polarity at those scales, and in Fig. 3 we show some results of optimal scale selection: Optimal-scale image, with dark greys corresponding to low scales and light greys corresponding to high scales, polarity image at the optimal scale (each pixel with its optimal polarity  $P_{\sigma^*}$ ), and texture-contrast image at the optimal scale (each pixel with its optimal texture contrast  $P_{\sigma^*}$ ). Low polarity appears in light grey and high polarity appears in dark grey. On the other hand, low texture contrast appears in dark grey and high contrast appears in light grey.

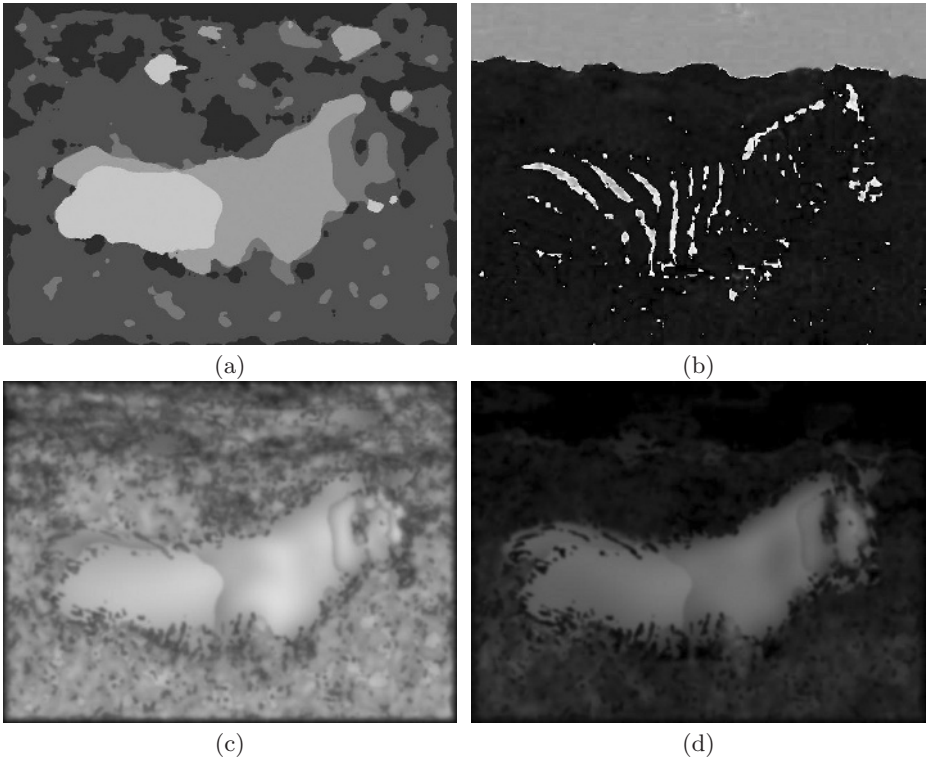
## 4 Adaptive Segmentation

### 4.1 EM Algorithm for Asymmetric Clustering

Given  $N$  image blocks  $x_1, \dots, x_N$ , each one having associated  $M$  possible features  $y_1, \dots, y_M$ , the Asymmetric Clustering Model (ACM) maximizes the log-likelihood

$$L(I, q) = - \sum_{i=1}^N \sum_{\alpha=1}^K I_{i\alpha} KL(p_{j|i}, q_{j|\alpha}), \quad (11)$$

where:  $p_{j|i}$  encodes the individual histogram, that is, the empirical probability of observing each feature  $y_j$  given  $x_i$ ;  $q_{j|\alpha}$  is the prototypical histogram associated to one of the  $K$  classes  $c_\alpha$ ;  $KL(\cdot, \cdot)$  is the symmetric Kullback-Leibler divergence; and  $I_{i\alpha} \in \{0, 1\}$  are class-membership variables.



**Fig. 3.** Texture features. (a) Optimal-scale image (b) Hue component (c)  $P_{\sigma^*}$  image (d)  $C_{\sigma^*}$  image

The following EM algorithm was proposed in [3][5]. The E-step consists of estimating the expected membership variables  $\hat{I}_{i\alpha} \in [0, 1]$  given the current estimation of the prototypical histogram  $q_{j|\alpha}$ :

$$\hat{I}_{i\alpha}^{t+1} = \frac{\hat{\rho}_{\alpha}^t \exp\{-KL(p_{j|i}, \hat{q}_{i|\alpha})/T\}}{\sum_{\beta=1}^K \hat{\rho}_{\beta}^t \exp\{-KL(p_{j|i}, \hat{q}_{i|\beta})/T\}}, \quad (12)$$

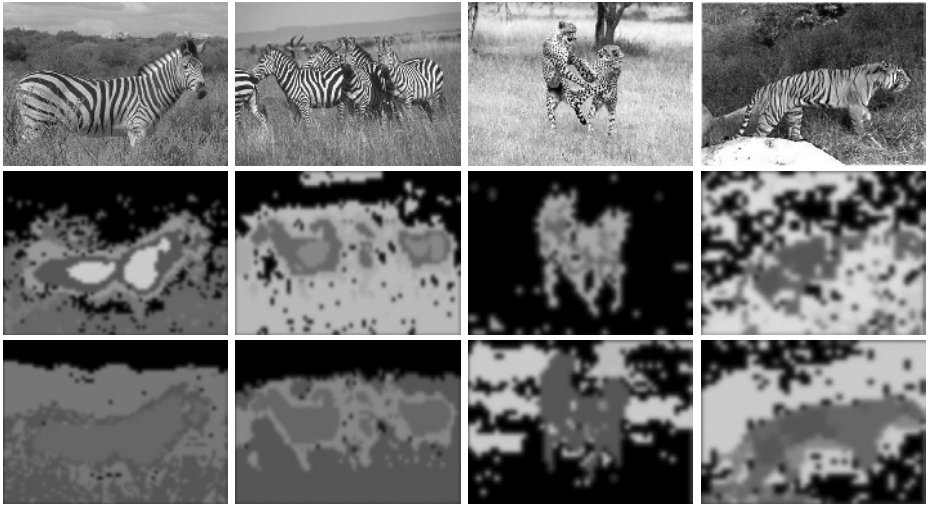
where

$$\hat{\rho}_{\alpha}^t = \frac{1}{N} \sum_{i=1}^N \hat{I}_{i\alpha}^t,$$

that is, the probability of assigning any block  $x_i$  to class  $c_{\alpha}$  at iteration  $t$ , and  $T$  the temperature, a control parameter which is reduced at each iteration.

In the M-step, given the expected membership variables  $\hat{I}_{i\alpha}^{t+1}$ , the prototypical histograms are re-estimated as follows:

$$\hat{q}_{j|\alpha}^{t+1} = \sum_{i=1}^N \pi_{i\alpha} p_{j|i}, \text{ where } \pi_{i\alpha} = \frac{\hat{I}_{i\alpha}^t}{\sum_{k=1}^N \hat{I}_{k\alpha}^t}, \quad (13)$$



**Fig. 4.** Segmentation results. Top: Input images; Middle: Only texture features; Bottom: Including color features.

that is, the prototype consists of the linear combination of all individuals  $p_{j|i}$  weighted by the  $\pi_{i\alpha}$ .

In a previous work we have introduced an adaptation mechanism for ACM [4] where we start with a high number of classes and class-fusion relies on considering whether the dispersion of the resulting class is lower than the sum of the dispersion of the two fusing classes. Herein we propose a method relying on the Kolmogorov-Smirnov test. As in our early work, we assume that the iterative process is divided in epochs, and our adaptation mechanism consists of starting by a high number of classes  $K_{max}$  and then reducing such a number, if proceeds, at the end of each epoch. At that moment we consider all the  $K(K-1)/2$  pairs of prototypes, where  $K$  is the current number of classes. For all these pairs we compute the Kolmogorov-Smirnov statistic with  $\alpha = 0.05$ , resulting from comparing their histograms, and then we select the pair  $\hat{q}_{j|\alpha}$  and  $\hat{q}_{j|\beta}$  with the lower statistic. If with such a statistic, the test does not succeed (both histograms are not different enough) we decide to fuse their classes.

Then, we compute the fused prototype  $\hat{q}_{j|\gamma}$  by applying Equation 13 and considering that  $I_{i\gamma} = I_{i\alpha} + I_{i\beta}$ , that is

$$\hat{q}_{j|\gamma} = \sum_{i=1}^N \pi_{i\gamma} p_{j|i}. \quad (14)$$

And then a new epoch starts, and proceeds until convergence is declared.

## 4.2 Segmentation Results

Dividing our input color images in blocks of  $8 \times 8$  pixels, we consider a histogram of 16 components both for polarity, texture contrast and hue component, and thus number of features is 32 when only texture is considered, and 48 when color is included. In Fig. 4 we compare the segmentation results obtained with and without color information, and assuming  $K_{max} = 10$ . In many cases texture features are enough for yielding acceptable segmentations, although color features usually improve the quality of the results.

## 5 Conclusion

In this paper we have proposed two entropy-related texture features, obtained through automatic scale selection. In order to demonstrate their utility in segmentation we have used them in an adaptive version of the ACM clustering model, and the obtained results were promising.

## References

1. Carson, C., Belongie, S., Greenspan, H., Malik, J. : Blobworld: Image segmentation using Expectation-Maximization and its application to image querying. *IEEE Trans. on Pattern Analysis and Machine Intelligence* (2002)
2. Gårding, J., Lindeberg, T.: Direct Computation of Shape Cues Using Scale-Adapted Spatial Derivative Operators. *International Journal of Computer Vision*, Vol 17, N 2 (1996) 163–191.
3. Hofmann, T., Puzicha, J.: Statistical Models for Co-occurrence Data. MIT AI-Memo 1625 Cambridge, MA (1998)
4. Lozano, M.A., Escolano, F.: Recognizing Indoor Images with Unsupervised Segmentation and Graph Matching. In: Garijo, F.J., Riquelme, J.C., Toro, M.(eds.): *Advances in Artificial Intelligence – Iberamia 2002. Lecture Notes on Artificial Intelligence*, Vol. 2527. Springer-Verlag, Berlin Heidelberg New York (2002) 933–942.
5. Puzicha, J.: Histogram Clustering for Unsupervised Segmentation and Image Retrieval. *Pattern Recognition Letters*, 20, (1999) 899–909.
6. A. Jain and F. Farrokhina. Unsupervised texture segmentation using gabor filters. *Pattern Recognition* 23, pp. 1167–1186, 1991.
7. H. Knutsson and G. Granlund. Texture analysis using two-dimensional quadrature filters. In *IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management*, pages 206–213, 1983.
8. Gotlieb C.C., Kreysszig. Texture descriptors based on cooccurrence matrices. *Comp. Vision, Graph. and Image Proc.* 51: 70–86, 1990.
9. Unser M., Texture Classification and Segmentation Using Wavelet Frames. *IEEE Trans. Image Processing*, 4, 11, 1549–1560, 1995.
10. N. Sochen, R. Kimmel and R. Malladi. A General Framework for Low Level Vision. *IEEE Trans on Image Processing*, Vol. 7, No. 3, pp. 310–318, 1998.