OPTIMAL TEXTURE FILTERING

Trygve Randen and John Håkon Husøy

Høgskolen i Stavanger, Stavanger, Norway
t.randen@ieee.org, http://www.hsr.no/~tranden

ABSTRACT

This paper deals with the design of a finite impulse response filter for optimal feature extraction in texture segmentation/classification. Motivated by the Fisher criterion, we develop techniques for optimal filter design. Posing some restrictions on the data, we are able to find a closed form solution to the problem, and we develop a unifying framework for this and several other approaches. We also show how the optimal filters may be designed, using less restrictive assumptions and a gradient search.

1. INTRODUCTION

The design of *optimal* filters for texture discrimination is gaining increased interest, and several recent articles propose methods for the design of such filters. Dunn et al. [1] and Teuner et al. [2] present approaches to the design of optimal Gabor filters and Husøy et al. [3] present an approach to the determination of the optimality of a bank of critically sampled perfect reconstruction IIR filters. A common problem with all these approaches is that the number of free parameters is low, and there is no guarantee that a true global optimum will be reached.

Opening for a much less restricted set of possible solutions, Jain and Karu [4] proposes an approach to optimal filter design using a multi layer perceptron neural network for the filtering and local energy operations, and the back propagation training rule for the determination of the optimal parameters for multiple textures. Mahalanobis and Singh [5] uses yet another approach to the design of the filters, determining the optimal filter by stochastic texture analysis and maximizing the ratio of the energies of the filter responses of two textures. A related approach was presented by Unser [6], for design of an optimal *transform* for the two-texture classification problem.

In this paper we propose a new technique for the design of filters for two-class feature extraction which

Thanks to Geco-Prakla, a Schlumberger company, for financial support.

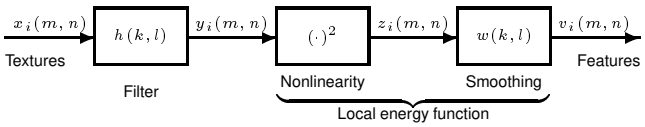


Figure 1: The experimental setup.

are optimal with respect to the Fisher criterion [7], i.e. the mean feature values for the two textures are maximally separated, while the feature variances are low. By assuming that the filter outputs are Gaussian random variables, we are able to solve the optimization problem.

If we make the assumption that the filter outputs are separable AR(1) processes (a rather restrictive assumption), we are able to find a closed form solution to the optimization problem. Otherwise, the optimization problem is solved by a gradient search. We finally develop a unifying framework, for both our closed form solution and several other approaches.

2. OPTIMAL FILTER DESIGN

The generic experimental setup for texture feature extraction with filters [8], Figure 1, is used as our model. We assume that we have an input image consisting of two textures, $x_1(m,n)$ and $x_2(m,n)$, and the output for texture $x_i(m,n)$ is the feature image $v_i(m,n)$. In the sequel, we denote the feature mean for texture i by $\mu_{v\,i}$ and the feature variance by $\sigma_{v_i}^2$. Expressions for the mean and variance will be given later.

The filtered image, $y_i(m,n)$ is given by $y_i(m,n) = h(m,n) * x_i(m,n)$. We define $\mathbf{h}$ as the vector formed by lexicographically ordering the rows of $h(m,n)$, and $\mathbf{x}_i(m,n) = [x_i(m,n) \cdots x_i(m,n-N+1) \cdots x_i(m-M+1, n-N+1)]^T$. Then, the output at pixel (m,n) for texture i may be written

$$y_i(m,n) = \mathbf{h}^T \mathbf{x}_i(m,n). \tag{1}$$

Similarly, the smoothing operation may be defined as

$$v_i(m,n) = \mathbf{w}^T \mathbf{z}_i(m,n).$$

We assume that the smoothing filter is an $N_w \times N_w$ rectangular filter.

The obvious aim of an optimal filter for two-texture discrimination is to make the local energies of the filter outputs maximally separable. If the feature separation is good we will achieve good segmentation – and the segmentation may be performed by a simple linear two-class discriminant classifier [7].

In order to achieve good separation, Mahalanobis and Singh [5] proposed the object function

$$J_1(\mathbf{h}) = \frac{\mu_{v1}}{\mu_{v2}} \tag{2}$$

and Unser [6] proposed to use

$$J_2(\mathbf{h}) = \frac{(\mu_{v1} - \mu_{v2})^2}{\mu_{v1}\mu_{v2}} \tag{3}$$

for optimal filter and transform design, respectively. (The Unser criterion [6] can easily be adapted to filter design.) The major drawback with both these criterions, is that they do not incorporate the feature variances. Hence we may achieve large separation of the feature *means*, but so large feature *variances* that the feature distributions overlap considerably. One object function which incorporates the variance is

$$J(\mathbf{h}) = \frac{(\mu_{v1} - \mu_{v2})^2}{{\sigma_{v1}}^2 + {\sigma_{v2}}^2}, \tag{4}$$

known as the Fisher criterion [7, p. 135]. It is previously not used in optimal texture feature extraction. In the rest of this paper, we will develop the necessay theory for optimal filter design with respect to the Fisher criterion.

In order to find the optimal filter, we have to find the partial derivative of $J(\mathbf{h})$ with respect to $\mathbf{h}$,

$$\begin{aligned}\frac{\partial J(\mathbf{h})}{\partial \mathbf{h}} &= \frac{2(\mu_{v1} - \mu_{v2})\left(\frac{\partial \mu_{v1}}{\partial \mathbf{h}} - \frac{\partial \mu_{v2}}{\partial \mathbf{h}}\right)({\sigma_{v1}}^2 + {\sigma_{v2}}^2)}{({\sigma_{v1}}^2 + {\sigma_{v2}}^2)^2} \\ &- \frac{(\mu_{v1} - \mu_{v2})^2\left(\frac{\partial {\sigma_{v1}}^2}{\partial \mathbf{h}} + \frac{\partial {\sigma_{v2}}^2}{\partial \mathbf{h}}\right)}{({\sigma_{v1}}^2 + {\sigma_{v2}}^2)^2}.\end{aligned} \tag{5}$$

Before we can calculate this expression, we have to develop expressions for the feature mean and variance.

2.1. Feature mean

Assuming wide sense stationarity, the expected energy at the output of the filter is defined by

$$\begin{aligned}\mu_{v\,i} &= E\{v_i(m,n)\} = E\{w(k,l) * z_i(m,n)\} \\ &= E\{z_i(m,n)\} = E\left\{y_i^2(m,n)\right\} \\ &= E\left\{(\mathbf{h}^T \mathbf{x}_i(m,n))^2\right\} = \mathbf{h}^T \mathbf{R}_{xxi} \mathbf{h},\end{aligned} \tag{6}$$

where

$$\mathbf{R}_{xxi} = E\left\{\mathbf{x}_i(m,n)\mathbf{x}_i^T(m,n)\right\}.$$

The partial derivative of $\mu_{v\,i}$ with respect to the filter vector $\mathbf{h}$ is needed in the subsequent sections, and since $\mathbf{R}_{xxi}$ is symmetric it is [9]

$$\frac{\partial \mu_{vi}}{\partial \mathbf{h}} = 2\mathbf{R}_{xxi}\mathbf{h}. \tag{7}$$

2.2. Feature variance

The variance of the feature image, $v_i(m,n)$, is

$$\begin{aligned}{\sigma_{v_i}}^2 &= E\left\{(v_i(m,n) - \mu_{v\,i})^2\right\} \\ &= \mathbf{w}^T E\left\{\mathbf{z}_i(m,n)\mathbf{z}_i^T(m,n)\right\}\mathbf{w} - 2\mu_{v\,i}\mu_{vi} + {\mu_{v_i}}^2 \\ &= \frac{1}{N_w^2}\sum_{k=0}^{N_w-1}\sum_{l=0}^{N_w-1} r_{zzi}(k,l) - {\mu_{v_i}}^2,\end{aligned} \tag{8}$$

where

$$\begin{aligned}r_{zzi}(k,l) &= E\{z_i(m,n)z_i(m+k,n+l)\} \\ &= E\{y_i(m,n)y_i(m+k,n+l) \\ &\qquad y_i(m,n)y_i(m+k,n+l)\}.\end{aligned} \tag{9}$$

For a one-dimensional Gaussian signal $y(k)$, we have [10] $E\{y(k)y(l)y(m)y(n)\} = r_{yy}(k-m)r_{yy}(l-n) + r_{yy}(l-m)r_{yy}(k-n) + r_{yy}(k-l)r_{yy}(m-n)$. If we assume that the filter output, $y(m,n)$, has a Gaussian distribution the fourth order moment in Equation (9) is easily found, and Equation (9) becomes

$$r_{zzi}(k,l) = 2r_{yyi}^2(k,l) + r_{yyi}^2(0,0).$$

Then, from Equation (8),

$$\begin{aligned}{\sigma_{v_i}}^2 &= \frac{1}{N_w^2}\sum_{k=0}^{N_w-1}\sum_{l=0}^{N_w-1}\left(2r_{yyi}^2(k,l) + r_{yyi}^2(0,0)\right) - {\mu_{v_i}}^2 \\ &= \frac{2}{N_w^2}\sum_{k=0}^{N_w-1}\sum_{l=0}^{N_w-1} r_{yyi}^2(k,l) + {\mu_{v_i}}^2 - {\mu_{v_i}}^2 \\ &= \frac{2}{N_w^2}\sum_{k=0}^{N_w-1}\sum_{l=0}^{N_w-1} r_{yyi}^2(k,l).\end{aligned} \tag{10}$$

Since

$$\begin{aligned}r_{yyi}(k,l) &= E\{y_i(m,n)y_i(m+k,n+l)\} \\ &= \mathbf{h}^T \mathbf{R}_{xxi}(k,l)\mathbf{h},\end{aligned}$$

where

$$\mathbf{R}_{xxi}(k,l) = E\left\{\mathbf{x}_i(m,n)\mathbf{x}_i^T(m+k,n+l)\right\},$$

we have from Equation (10) that

$$\sigma_{v_i}^2 = \frac{2}{N_w^2} \sum_{k=0}^{N_w-1} \sum_{l=0}^{N_w-1} (\mathbf{h}^T \mathbf{R}_{xxi}(k,l)\mathbf{h})^2. \quad (11)$$

It can easily be shown that

$$\frac{\partial \mathbf{h}^T \mathbf{A} \mathbf{h}}{\partial \mathbf{h}} = \mathbf{A}\mathbf{h} + \mathbf{A}^T\mathbf{h}$$

for arbitrary (i.e. also non-symmetric) $\mathbf{A}$. Then

$$\frac{\partial \sigma_{v_i}^2}{\partial \mathbf{h}} = \frac{4}{N_w^2} \sum_{k=0}^{N_w-1} \sum_{l=0}^{N_w-1} (\mathbf{h}^T \mathbf{R}_{xxi}(k,l)\mathbf{h}) \left(\mathbf{R}_{xxi}(k,l)\mathbf{h} + \mathbf{R}_{xxi}(k,l)^T\mathbf{h}\right). \quad (12)$$

2.3. Gradient search optimization

For any given filter, $\mathbf{h}$, we may now compute the object function $J(\mathbf{h})$ and its partial derivative. Hence, we may perform gradient search with some gradient search algorithm, to find the filter yielding the optimum filter $\mathbf{h}$.

2.4. Closed form optimization

With a gradient search optimization of the filter, we have no guarantee of convergence to a global optimum. In addition, depending on the surface of the search space, the convergence may be slow (remember that $\mathbf{h}$ typically has large dimensionality). Hence a closed form solution to the optimization problem would be desirable. We were, unfortunately, not able to find any closed form solution, i.e. any solution to the equation $\partial J(\mathbf{h})/\partial \mathbf{h} = 0$ from the results above.

If we assume that $y_i(m,n)$ is a separable AR(1) random process, we have $r_{yyi}(k,l) = r_{yyi1}(k) r_{yyi2}(l)$ where $r_{yyij}(k) = \rho_{ij}^k r_{yyij}(0)$ for $k > 0$, and hence

$$r_{yyi}(k,l) = \rho_{i1}^k \rho_{i2}^l r_{yyi}(0,0). \quad (13)$$

Thus, from Equation (10) we have

$$\begin{aligned}
\sigma_{v_i}^2 &= \frac{2}{N_w^2} \sum_{k=0}^{N_w-1} \sum_{l=0}^{N_w-1} \rho_{i1}^{2k} \rho_{i2}^{2l} r_{yyi}^2(0,0) \\
&= \frac{2}{N_w^2} \frac{1-\rho_{i1}^{2N_w}}{1-\rho_{i1}^2} \frac{1-\rho_{i2}^{2N_w}}{1-\rho_{i2}^2} r_{yyi}^2(0,0) \\
&= \frac{2}{N_w^2} \frac{1-\rho_{i1}^{2N_w}}{1-\rho_{i1}^2} \frac{1-\rho_{i2}^{2N_w}}{1-\rho_{i2}^2} (\mathbf{h}^T \mathbf{R}_{xxi}\mathbf{h})^2, \quad (14)
\end{aligned}$$

which, since $\mathbf{R}_{xxi}$ is symmetric, has the derivative [9]

$$\begin{aligned}
\frac{\partial \sigma_{v_i}^2}{\partial \mathbf{h}} &= \frac{8}{N_w^2} \frac{1-\rho_{i1}^{2N_w}}{1-\rho_{i1}^2} \frac{1-\rho_{i2}^{2N_w}}{1-\rho_{i2}^2} (\mathbf{h}^T \mathbf{R}_{xxi}\mathbf{h}) \mathbf{R}_{xxi}\mathbf{h} \\
&= \frac{4\sigma_{v_i}^2}{\mu_{v_i}} \mathbf{R}_{xxi}\mathbf{h}. \quad (15)
\end{aligned}$$

Inserting the expressions from Equations (7) and (15) into the expression for $\partial J(\mathbf{h})/\partial \mathbf{h}$, Equation (5), and equating to zero yields

$$(\mathbf{R}_{xx2}^{-1}\mathbf{R}_{xx1})\mathbf{h} = \frac{1 + \frac{(\mu_{v1}-\mu_{v2})\sigma_{v2}^2}{(\sigma_{v1}^2+\sigma_{v2}^2)\mu_{v2}}}{1 - \frac{(\mu_{v1}-\mu_{v2})\sigma_{v1}^2}{(\sigma_{v1}^2+\sigma_{v2}^2)\mu_{v1}}} \mathbf{h},$$

which we may rewrite as

$$(\mathbf{R}_{xx2}^{-1}\mathbf{R}_{xx1})\mathbf{h} = \lambda\mathbf{h} \quad (16)$$

with

$$\lambda = \frac{1 + \frac{(\mu_{v1}-\mu_{v2})\sigma_{v2}^2}{(\sigma_{v1}^2+\sigma_{v2}^2)\mu_{v2}}}{1 - \frac{(\mu_{v1}-\mu_{v2})\sigma_{v1}^2}{(\sigma_{v1}^2+\sigma_{v2}^2)\mu_{v1}}}.$$

We identify Equation (16) as an eigenvalue equation, where the filter, $\mathbf{h}$, is the eigenvector and λ is the eigenvalue. The optimal filter is therefore the eigenvector, $\mathbf{h}$, yielding maximum $J(\mathbf{h})$.

Note that the we do not expect the real world case to meet the assumptions above exactly, and we suspect that the assumption that $y_i(m,n)$ is a separable AR(1) process is the least realistic. We don't expect the textures to meet these assumptions exactly in a real world application, but we expect that useful results may be obtained by making the above assumptions. Our experimental results indicate that such is the case.

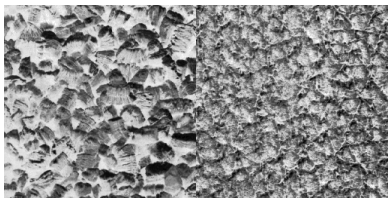
2.5. A unified framework

As we saw, the closed form solution above entailed the solution of the eigenproblem $(\mathbf{R}_{xx2}^{-1}\mathbf{R}_{xx1})\mathbf{h} = \lambda\mathbf{h}$. The same eigenproblem occurs as the closed form solution to the optimization with respect to the object function (2) [5] and the object function (3) applied to filter design [11].

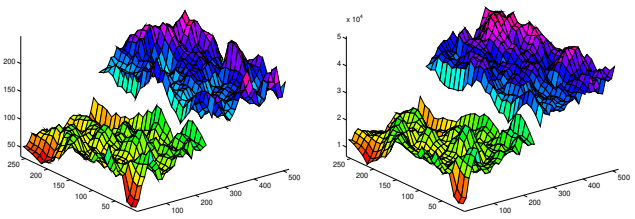
Since μ_{v_i} is a function of $(\mathbf{h}^T\mathbf{R}_{xxi}\mathbf{h})$ (Equation (6)), any object function, $J(\mathbf{h})$ expressing some relation between the feature means, like $J_1(\mathbf{h})$ and $J_2(\mathbf{h})$ in Equations (2) and (3), may be expressed as a function of $(\mathbf{h}^T\mathbf{R}_{xx1}\mathbf{h})$ and $(\mathbf{h}^T\mathbf{R}_{xx2}\mathbf{h})$. As shown in [11], the optimal solution then entails the solution of the eigenproblem $\mathbf{R}_{xx2}^{-1}\mathbf{R}_{xx1}\mathbf{h} = \lambda\mathbf{h}$.

Further, if we assume that the $y_i(m,n)$'s are separable AR(1) Gaussian processes, the feature variances, $\sigma_{v_i}^2$, are functions of $(\mathbf{h}^T\mathbf{R}_{xxi}\mathbf{h})$, and we similarly find that the optimization of object functions on the feature means and variances entails the solution of the same eigenproblem [11].

If we make the separable Gaussian AR(1) assumption, and assume that the textures have zero mean, it is possible to express the error probability of a two-class one-dimensional linear discriminant classifier [7] as a



(a)



(b) (c)

Figure 2: The texture pair D5-D92 (a), the feature surface from the closed form optimization (b), and from the gradient optimization (c). The discontinuity in the surface indicates the true texture boundary.

function of the μ_{v_i}'s and $\sigma_{v_i}^2$'s [11], thus the optimal solution is one of the eigenvectors of $\mathbf{R}_{xx2}^{-1}\mathbf{R}_{xx1}\mathbf{h} = \lambda\mathbf{h}$.

Hence, we see that the closed form solution of various optimization problems all enatils the same eigenproblem.

2.6. Experimental results

We do not propose any way for finding optimal filter sizes, so for demonstration purposes we optimize an 8×8 finite impulse response (FIR) filter with a 40×40 smoothing operator. As the initial value for the gradient search approach, we use the closed form solution.

Examples of the feature surfaces with the two techniques described here for the texture pair D5-D92 from the Brodatz [12] are given in Figure 2. We see clearly that it is feasible to use a simple linear discriminant classifier and get a satisfying segmentation in both cases. The features from the filter optimal with respect to the object function (2) had a considerable feature overlap due to higher variance for this texture pair.

We eveluated the proposed techniques on a large set of Brodatz [12] textures, and were not able to find any situation where the techniques failed. However, as expected the gradient optimized solution were somewhat better (it is based on fewer restrictions) and using estimated means and variances from the feature images, we found it to yield averagely around 50% higher $J(\mathbf{h})$, Equation (4).

3. CONCLUSION

We have developed two new approaches to optimal feature extraction for two textures with one filter. The techniques are based on modelling the textures with their autocorrelation functions and, with different assumptions, yield filters which are optimal with respect to the Fisher criterion. We experimentally found the techniques to be very robust.

The test images and programs (MATLAB) may be found on anonymous FTP at site `ftp.hsr.no` or on the World Wide Web at `http://www.hsr.no/~tranden`.

4. REFERENCES

[1] D. F. Dunn, W. E. Higgins, and J. Wakeley, "Texture segmentation using 2-D Gabor elementary functions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 130–149, Feb. 1994.

[2] A. Teuner, O. Pichler, and B. J. Hosticka, "Unsupervised texture segmentation of images using tuned matched Gabor filters," *IEEE Trans. Image Processing*, vol. 43, July 1995.

[3] J. H. Husøy, T. Randen, and T. O. Gulsrud, "Image texture classification with digital filter banks and transforms," in *Proc. SPIE International symposium on optical instrumentation and applied science: Applications of Digital Image Processing XVI*, (San Diego), pp. 260 – 271, July 1993.

[4] A. K. Jain and K. Karu, "Automatic filter design for texture discrimination," in *Proc. Int. Conf. on Patt. Rec.*, pp. 454–458, Int. Assoc. for Patt. Rec., Oct. 1994.

[5] A. Mahalanobis and H. Singh, "Application of correlation filters for texture recognition," *Applied Optics*, vol. 33, no. 11, pp. 2173–2179, 1994.

[6] M. Unser, "Local linear transforms for texture measurements," *Signal Processing*, vol. 11, no. 1, pp. 61 – 79, 1986.

[7] K. Fukunaga, *Statistical Pattern Recognition*. San Diego: Academic Press, 2 ed., 1990.

[8] T. Randen and J. H. Husøy, "Multichannel filtering for image texture segmentation," *Optical Engineering*, vol. 33, pp. 2617–2625, Aug. 1994.

[9] S. Haykin, *Adaptive filter theory*. Englewood Cliffs: Prentice Hall, 2 ed., 1991.

[10] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs: Prentice-Hall, 1975.

[11] T. Randen and J. H. Husøy, "Optimal filters for two-class texture feature extraction," *Pattern Recognition*, 1995. submitted.

[12] P. Brodatz, *Textures: A photographic album for artists and designers*. NY: Dover, 1966.