Optimal Fisher discriminant ratio for an arbitrary spatial light modulator

Richard D. Juday
NASA Johnson Space Center
Houston TX 77058

Abstract

Optimizing the Fisher ratio is well established in statistical pattern recognition as a means of discriminating between classes. I show how to optimize that ratio for optical correlation intensity by choice of filter on an arbitrary spatial light modulator (SLM). I include the case of additive noise of known power spectral density.

1. Introduction

There is a long and venerable history[1] in optical correlation pattern recognition (OCPR) of building filters to permit the discrimination between two classes of objects on the basis of their correlation intensities. Usually we would like for objects in the "accept" class to have large intensity and conversely for the "reject" class. The classical Fisher linear discriminant (FLD) reduces a highly dimensioned (and possibly complex) vector to a single quantity with the intention that it can be thresholded as a discriminant between classes. The classical FLD operates as a linear transform of the input object, and its optimizing filter best separates the means of the transformed classes, as normalized to their widths. In general it is a good metric to optimize, and under some circumstances (e.g. identical normal distributions) the FLD is an exactly optimal (minimal error) classifier. I have not previously seen the Fisher ratio analytically optimized for OCPR. In OCPR we work with intensities, not just the complex field amplitude that is the last linear stage in the optical correlation process. Thus we can not tell the difference between complex correlations of the same amplitude even if their values separate well in the complex plane. We take the Fisher ratio for OCPR to be the squared difference between mean correlation intensities for two classes of objects, divided by the sum of the correlation variances for those two classes (see Eq. (1)). I show how to maximize the optical Fisher ratio by choice of a filter to be realized on an arbitrary SLM. I include additive noise in determining the normalizing variances.

2. Formulating the problem

We assume that there is an optimal filter; a necessary condition for its optimality is that the partial derivative of the filter with respect to allowed changes is zero. In practice this laboratory has not found any problem with local maxima, nor with the fact that the specification for the *worst* filter nominally looks the same. A frequency's filter value can not be optimally chosen without regard to all other frequencies' filter values, the signals to be discriminated, and the noise. An explicit feature here is that all such information is condensed to a comparatively very few parameters to search over (there is essentially one complex scalar per training image to search over – not bad, compared with the typical filter's tens of thousands of frequencies!). Then the optimal filter value at the frequency under consideration must be chosen from the set of realizable values.

We adopt the following nomenclature. C is a class (Λ for accept, Ψ for reject). $\langle \cdot \rangle_C$ is the expectation of a quantity over C. The m-th frequency's value of the

transformed signal is $S_m = A_m \exp(j\phi_m)$ and of the filter, $H_m = M_m \exp(j\theta_m)$. The central correlation field is $D = B\exp(j\beta) = \sum_k H_k S_k$ where the indicated sum is over all frequencies (we use the one-dimensional notation), and the correlation intensity is $I = B^2$. Total variance over class C is σ_{CT}^2, arising partly in different correlation intensities for the objects within training class C, $\sigma_C^2 = \left\langle \left(I - \langle I \rangle_C\right)^2 \right\rangle_C$, and partly in additive input noise. The power spectral density of the input noise is P_{nm} at the m-th frequency, and the input noise's contribution to variance is $\sigma_n^2 = \sum_k |H_k|^2 P_{nk} = \sum_k M_k^2 P_{nk}$.

We assume the effect of noise on intensity has zero mean (else its effect can be incorporated into the class means for intensity, and so we will take no further notice of it here). The total variance is modeled as $\sigma_{CT}^2 = \sigma_C^2 + \sigma_n^2$. With these definitions we can set up the optical Fisher ratio, J, and optimize it.

$$J = \frac{\left(\langle I \rangle_\Lambda - \langle I \rangle_\Psi\right)^2}{\sigma_{\Lambda T}^2 + \sigma_{\Psi T}^2} =: \frac{\text{Num}}{\text{Den}} \tag{1}$$

defining the numerator and denominator of J.

3. Optimizing the filter

The optimization strategy is based on that of Juday[2,3]. We take the radial component of the gradient of J in the complex plane of values for H_m, and from that infer the azimuthal component, and thus deduce the optimal realizable values for H_m.

Taking the radial derivative in the complex plane of H_m,

$$\text{Den}^2 \frac{\partial J}{\partial M_m} = \text{Den}\frac{\partial \text{Num}}{\partial M_m} - \text{Num}\frac{\text{Den}}{\partial M_m}$$

$$= 2\text{Den}\sqrt{\text{Num}}\frac{\partial}{\partial M_m}\left(\langle I \rangle_\Lambda - \langle I \rangle_\Psi\right) - \text{Num}\frac{\partial}{\partial M_m}\left(\sigma_{\Lambda T}^2 + \sigma_{\Psi T}^2\right) \tag{2}$$

and we see that we need several partial derivatives. As shown by Juday[2],

$$\frac{\partial I}{\partial M_m} = 2BA_m \cos\theta_m \cos\left(\beta - \phi_m\right) + 2BA_m \sin\theta_m \sin\left(\beta - \phi_m\right) \tag{3}$$

and restriction to a class and taking the expectation is straightforward. From the modeling of the variance,

$$\frac{\partial \sigma_{CT}^2}{\partial M_m} = \frac{\partial \sigma_C^2}{\partial M_m} + \frac{\partial \sigma_n^2}{\partial M_m} = \frac{\partial \sigma_C^2}{\partial M_m} + 2M_m P_{nm}. \tag{4}$$

In a little more detail than in the definition,

$$\sigma_{\mathrm{C}}^2 = \frac{1}{N_{\mathrm{C}}-1}\sum_{i\in \mathrm{C}}\left(I_i - \langle I\rangle_{\mathrm{C}}\right)^2 = \frac{1}{N_c-1}\sum_{i\in \mathrm{C}}\left(I_i - \frac{1}{N_{\mathrm{C}}}\sum_{\ell\in \mathrm{C}}I_\ell\right)^2 \tag{5}$$

from which

$$\frac{\partial \sigma_{\mathrm{C}}^2}{\partial M_m} = \frac{1}{N_{\mathrm{C}}-1}\sum_{i\in \mathrm{C}}2\left(I_i-\langle I\rangle_{\mathrm{C}}\right)\left(\frac{\partial I_i}{\partial M_m} - \left\langle\frac{\partial I}{\partial M_m}\right\rangle_{\mathrm{C}}\right)$$

$$= \frac{4}{N_{\mathrm{C}}-1}\sum_{i\in \mathrm{C}}\left(I_i-\langle I\rangle_{\mathrm{C}}\right)\Big[\cos\theta_m B_i A_{im}\cos\left(\beta_i-\phi_{im}\right)+\{\sin\text{ term}\} \tag{6}$$

$$-\frac{1}{N_{\mathrm{C}}}\sum_{\ell\in \mathrm{C}}\left(\cos\theta_m B_\ell A_{\ell m}\cos\left(\beta_\ell - \phi_{\ell m}\right)+\{\sin\text{ term}\}\right)\Big].$$

Reorganizing and inserting Eq. (6) into Eq. (4),

$$\frac{\partial \sigma_{\mathrm{C}T}^2}{\partial M_m} = \cos\theta_m\frac{4}{N_{\mathrm{C}}-1}\sum_{i\in \mathrm{C}}\left(I_i-\langle I\rangle_{\mathrm{C}}\right)\Big[B_i A_{im}\cos\left(\beta_i-\phi_{im}\right)-\frac{1}{N_{\mathrm{C}}}\sum_{\ell\in \mathrm{C}}B_\ell A_{\ell m}\cos\left(\beta_\ell-\phi_{\ell m}\right)\Big]$$

$$+\sin\theta_m\{\text{similar}\} \tag{7}$$

$$+2M_m P_{nm}.$$

Now consider an operator $\square$ such that

$$\exp(j\theta)\,\square\; R\exp(j\alpha) = R\cos\theta\cos\alpha + R\sin\theta\sin\alpha \tag{8}$$

gives the projection of $R\exp(j\alpha)$ onto the unit vector in the complex direction $\exp(j\theta)$. We regard an equation like the first part of Eq. (7) as expressing the gradient of a quantity as it interacts with the complex unit vector $\exp(j\theta)$ – and further, that the gradient is uniform [later we shall look at the last term in Eq.(7)]. From this perspective we build $\nabla_m J$, the gradient of J as a function of position in the complex plane of values for H_m. We assemble Eqs.(3) and (7) into Eq. (2), with the result

$$\frac{\mathrm{Den}^2}{4}\nabla_m J = \mathrm{Den}\sqrt{\mathrm{Num}}\left[\frac{1}{N_\Lambda}\sum_{i\in\Lambda}D_i S_{im}^* - \frac{1}{N_\Psi}\sum_{\ell\in\Psi}D_\ell S_{\ell m}^*\right]$$

$$-\mathrm{Num}\left[\frac{1}{N_\Lambda-1}\sum_{i\in\Lambda}\left(I_i-\langle I\rangle_\Lambda\right)D_i S_{im}^* - \frac{1}{N_\Lambda\left(N_\Lambda-1\right)}\sum_{i\in\Lambda}\left(I_i-\langle I\rangle_\Lambda\right)\sum_{\ell\in\Lambda}D_\ell S_{\ell m}^*\right.$$
$$\left.+\{\text{similar for }\Psi\}\right] \tag{9}$$

$$-\mathrm{Num}M_m P_{nm}.$$

In the term of Eq. (9) that is the product of sums, we interchange summation order and swap indices so that it becomes a sum on S_{im}^*. Then we can express $\nabla_m J$ as a sum over the training images' spectral conjugated transforms S_{im}^* (and the $M_m P_{nm}$ term).

$$\frac{\text{Den}^2}{4}\nabla_m J = \sum_{i\in\Lambda} S_{im}^* \left[+\frac{D_i\text{Den}\sqrt{\text{Num}}}{N_\Lambda} - \frac{D_i\text{Num}\left(I_i - \langle I\rangle_\Lambda\right)}{N_\Lambda-1} + \frac{D_i\text{Num}}{N_\Lambda\left(N_\Lambda-1\right)}\sum_{\ell\in\Lambda}\left(I_\ell - \langle I\rangle_\Lambda\right) \right]$$

$$+ \sum_{i\in\Psi} S_{im}^* \left[-\frac{D_i\text{Den}\sqrt{\text{Num}}}{N_\Psi} - \frac{D_i\text{Num}\left(I_i - \langle I\rangle_\Psi\right)}{N_\Psi-1} + \frac{D_i\text{Num}}{N_\Psi\left(N_\Psi-1\right)}\sum_{\ell\in\Psi}\left(I_\ell - \langle I\rangle_\Psi\right) \right]$$

$$-\text{Num}M_m P_{nm}$$

$$(10)$$

The terms in square braces are not functions of m, but there is one for each reference image. Therefore each can be replaced by a complex constant T_i, now giving

$$\nabla_m J = \sum_{i\in\Lambda\cup\Psi} T_i S_{im}^* - T_{noise}M_m P_{nm} \qquad (11)$$

which implicitly defines the set of complex coefficients $\{T_i; i\in\Lambda\cup\Psi\cup noise\}$, one per training object plus a real-valued one for the input noise (if that noise is present). The coefficients represent the necessary information as mentioned in the first paragraph of Section 2. We do not know the coefficients *a priori*. However, we know they exist, and we can search for their values and confirm them by comparing Eqs. (10) and (11) when we have maximized J in the search. Following the gradient-of-metric logic developed in section 13 of Juday[2], Eq. (11) is our guide to the selection of the optimal filter value for the m-th frequency. If P_{nm} is zero, we select for H_m the realizable value that has the largest projection in the direction of $\nabla_m J$. If P_{nm} is not zero, we compute an ideal value

$$H^I = \frac{\left|\sum_{i\in\Lambda\cup\Psi} T_i S_{im}^*\right|}{T_{noise}P_{nm}}\exp\left[j\text{Arg}\left(\sum_{i\in\Lambda\cup\Psi} T_i S_{im}^*\right)\right] \qquad (12)$$

and then select for H_m the realizable value that is closest by Euclidean measure. The effect of the additive noise is to change the maximum-projection filter toward matched-filter behavior.

Interestingly, there is a strong similarity with the initial formulation of the synthetic discriminant function[4] (SDF) for noiseless input. In that approach a linear sum of training images was sought that would cause exactly the desired central correlation intensities. There were two flaws in that approach; the computed filter was not realizable and the tools to handle the mapping onto a filter encoding domain were not at hand, and the method needlessly specified certain complex correlation values that were not founded in the observation of intensity.

4. Physical results

We have gotten confirming optical bench results, but there is not room to show them here. A subsequent paper will explore some practical issues including search strategies, convergence in the search, limitation on how many training images can be put into a filter, selecting among operating curves for various noise environments, etc.

5. Acknowledgement

[1] B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators", Applied Optics **31**, 4773-4801 (1992).

[2] Richard D. Juday, "Generalized Rayleigh quotient approach to filter optimization", JOSA-A **15**(4), 777-790 (April 1998).

[3] Richard D. Juday, "Optimal realizable filters and the minimum Euclidean distance principle", Applied Optics **32**, 5100-5111 (10 September 1993).

[4] C. F. Hester and David Casasent, "Multivariant technique for multiclass pattern recognition", Applied Optics **19**, 1758-1761 (1980).