References

[1] O. H. Schade, "Optical and photoelectric analog of the eye," *J. Opt. Soc. Amer.*, vol. 46, pp. 721–739, 1956.
[2] F. Ratliff, *Mach bands: Quantitative studies on neural networks in the retina.* San Francisco: Holden Day, 1965.
[3] D. H. Kelly, "Spatial frequency selectivity in the retina," *Vision Res.*, vol. 15, pp. 665–672, 1975.
[4] J. P. Thomas, "Spatial resolution and spatial interactions," in *Handbook of Perception*, vol. V, "Seeing," E. C. Carterette and M. P. Friedman, Eds. New York: Academic 1975, pp. 233–261.
[5] D. H. Kelly, "Visual contrast sensitivity," *Optica Acta*, vol. 24, pp. 107–129, Feb. 1977.
[6] B. E. A. Saleh, *Photoelectron Statistics.* New York: Springer, 1978.
[7] H. K. Hartline and F. Ratliff, "Inhibitory interaction of receptor units in the eye of Limulus," *J. Gen. Physiol.*, vol. 40, pp. 357–376, 1957.
[8] F. Ratliff and H. K. Hartline, "The response of Limulus optic nerve fibers to patterns of illumination on the retinal mosaic," *J. Gen. Physiol.*, vol. 42, pp. 1241–1255, 1959.
[9] H. B. Barlow, R. Fitzhugh, and S. W. Kuffler, "Change of organization in the receptive field of the cat's retina during dark adaptation," *J. Physiol.*, vol. 137, pp. 338–354, 1957.
[10] H. B. Barlow, "Temporal and spatial summation in human vision at different background intensities," *J. Physiol.*, vol. 141, pp. 337–350, 1958.
[11] H. R. Blackwell, "Contrast thresholds of the human eye," *J. Opt. Soc. Amer.*, vol. 36, pp. 624–643, 1946.
[12] F. L. Van Nes and M. A. Bouman, "Spatial modulation transfer in the human eye," *J. Opt. Soc. Amer.*, vol. 57, pp. 401–406, 1967.
[13] A. Van Meeteren, "Spatial sinewave response of the visual system (a critical literature survey)," Institute for Perception, RVO-TNO, Soesterberg, 1968.
[14] G. H. Jacobs and R. L. Yolton, "Center-surround balance in receptive fields of cells in the lateral geniculate nucleus," *Vision Res.*, vol. 10, pp. 1127–1144, 1970.
[15] D. H. Kelley, "Adaptation effects on spatio-temporal sinewave thresholds," *Vision Res.*, vol. 12, pp. 89–101, 1972.
[16] R. L. DeValois, H. Morgan, and D. M. Snodderly, "Psychophysical studies of monkey vision—III. Spatial luminance contrast sensitivity tests of macaque and human observers," *Vision Res.*, vol. 14, pp. 75–81, 1974.
[17] M. H. Pirenne, *Vision and the Eye.* London: Science Paperbacks, 1967.
[18] M. H. Pirenne and E. J. Denton, "Accuracy and sensitivity of the human eye," *Nature*, vol. 170, pp. 1039–1042, 1952.
[19] G. V. BéKésy, "Neural inhibitory units of the eye and skin," *J. Opt. Soc. Amer.*, vol. 50, pp. 1060–1070, 1960.
[20] I. Morishita and A. Yajima, "Analysis and simulation of networks of mutually inhibiting neurons," *Kybernetik*, vol. 11, pp. 154–165, 1972.
[21] K. P. Hadeler, On the theory of lateral inhibition," *Kybernetik*, vol. 14, pp. 161–165, 1974.
[22] B. E. A. Saleh and T. E. Cohn, "Reduction of the human detective quantum efficiency due to lateral neural interactions," *Biological Cybern.*, vol. 28, pp. 137–141, 1978.
[23] A. Rose, "The sensitivity performance of the human eye on an absolute scale," *J. Opt. Soc. Amer.*, vol. 38, pp. 196–208, 1948.

Correspondence

Bayesian Classification in a Time-Varying Environment

PHILIP H. SWAIN, MEMBER, IEEE

Abstract—**The problem of classifying a pattern based on multiple observations made in a time-varying environment is dealt with. The identity of the pattern may itself change. A Bayesian solution is derived, after which the conditions of the physical situation are invoked to produce a "cascade" classifier model. Experimental results based on remote sensing data illustrate the effectiveness of the classifier.**

Introduction

We pose the following pattern classification problem: a series of observations is made on a pattern in a time-varying environment. The identity of the pattern itself may change. It is desired to classify the pattern after the current observation is made, drawing on information derived from earlier observations plus knowledge about the statistical behavior of the environment.

Manuscript received May 5, 1978; revised July 24, 1978. This work was supported in part by NASA Contract NAS9-14970.

The author is with the School of Electrical Engineering and the Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN 47907.

An example of such a situation arises in remote sensing applications in which the sensor system can make multiple passes over the same ground area [1]. The identity of the ground cover may change between passes. In general, it is desired to determine the current identity of the ground cover, but past observations can be helpful in accomplishing the identification.

Approach

The classification strategy we shall develop is a Bayes optimal (minimum risk) strategy [2]. In the ordinary single observation case, the approach is to select a decision rule so as to minimize the conditional average loss

$$L_X(\omega_i) = \sum_{j=1}^{m} \lambda_{ij} p(\omega_j \mid X), \tag{1}$$

where

X	n-variate observation (feature) vector,
$\{\omega_j, j = 1, 2, \cdots, m\}$	set of m classes,
λ_{ij}	cost resulting from classifying into class i a pattern actually from class j, and
$p(\omega_j \mid X)$	conditional probability that, given observation X, its class is ω_j.

That is, $L_X(\omega_i)$ is the expected loss incurred if an observation X is classified as ω_i. Commonly [2], λ_{ij} is taken to be the "0–1 loss function," i.e.,

$$\lambda_{ij} = \begin{cases} 0, & i = j \text{ (no cost for correct classification)} \\ 1, & i \neq j \text{ (unit cost for an error).} \end{cases}$$

Then (1) becomes

$$L_X(\omega_i) = 1 - p(\omega_i | X), \quad (2)$$

and an appropriate decision rule which will minimize $L_X(\omega_i)$ is

decide $X \in \omega_i$, if and only if

$$p(X | \omega_i)p(\omega_i) = \max_j p(X | \omega_j)p(\omega_j), \quad (3)$$

where $p(X | \omega_i)$ is the probability density function for the observations associated with class ω_i and $p(\omega_i)$ is the *a priori* probability of class ω_i. Thus the set of products $\{p(X | \omega_i)p(\omega_i), i = 1, 2, \cdots, m\}$ is a set of discriminant functions for the classification problem.

We now generalize this Bayes optimal approach to the case of a series of observations. It will be convenient to assume that observations are made at two times. Generalization to a larger number of observation times is straightforward.

Let $X_1 = X(t_1)$ and $X_2 = X(t_2)$ be n-variate random vectors, the pattern observations at times t_1 and t_2, respectively.

Let $\{v_i = v_i(t_1) | i = 1, 2, \cdots, m_1\}$ be the set of possible classes at time t_1, and let $\{\omega_i = \omega_i(t_2) | i = 1, 2, \cdots, m_2\}$ be the set of possible classes at time t_2.

We define a compound conditional average loss

$$L_{X_1X_2}(\omega_i) = \sum_{j=1}^{m_2} \lambda_{ij} p(\omega_j | X_1, X_2), \quad (4)$$

where λ_{ij} is the cost resulting from classifying into class i, at time t_2, a pattern actually from class j. In this case $p(\omega_j | X_1, X_2)$ is the *a posteriori* probability that, given the observations X_1 at time t_1 and X_2 at time t_2, the class of the pattern at time t_2 is ω_j.

Once again assuming a "0–1 loss function," (4) becomes

$$L_{X_1X_2}(\omega_i) = 1 - p(\omega_i | X_1, X_2), \quad (5)$$

which is minimized if we choose ω_i to maximize the *a posteriori* probability $p(\omega_i | X_1, X_2)$. Thus an appropriate set of discriminant functions for a Bayes optimal classification strategy is the set of *a posteriori* probabilities; i.e.,

$$\{p(\omega_i | X_1, X_2), i = 1, 2, \cdots, m_2\}.$$

As usual, however, we wish to derive a set of equivalent discriminant functions expressed in terms of class-conditional density functions and *a priori* probabilities as in (3). This may be accomplished proceeding as follows. First we write

$$p(\omega | X_1, X_2) = \frac{p(\omega, X_1, X_2)}{p(X_1, X_2)} \quad (6)$$

For fixed X_1 and X_2, the denominator in (6) is constant. Let $c = 1/p(X_1, X_2)$ and write (6) as

$$\begin{aligned} p(\omega | X_1, X_2) &= cp(\omega, X_1, X_2) \\ &= c \sum_v p(X_1, X_2, v, \omega) \\ &= c \sum_v p(X_1, X_2 | v, \omega)p(v, \omega) \\ &= c \sum_v p(X_1, X_2 | v, \omega)p(\omega | v)p(v). \end{aligned} \quad (7)$$

The summation is over the classes which can occur at time t_1. The factor $p(X_1, X_2 | v, \omega)$ is a joint class-conditional density; $p(\omega | v)$ may be interpreted as a transition probability (the probability that the class is ω at time t_2 given the class was v at time t_1); and $p(v)$ is an *a priori* probability.

Thus the multiobservational decision rule analogous to (3) is

decide $X_2 \in \omega_i$, if and only if

$$\sum_{k=1}^{m_1} p(X_1, X_2 | v_k, \omega_i)p(\omega_i | v_k)p(v_k) = \max_j \sum_{k=1}^{m_1} p(X_1, X_2 | v_k, \omega_j)p(\omega_j | v_k)p(v_k), \quad (8)$$

and the set of discriminant functions is the set of sums of products:

$$\left\{ \sum_{k=1}^{m_1} p(X_1, X_2 | v_k, \omega_i)p(\omega_i | v_k)p(v_k), \quad i = 1, 2, \cdots, m_2 \right\} \quad (9)$$

A "Cascade" Implementation

In practice, the terms in the discriminant functions must be estimated from "training samples." The most formidable job is estimating the $m_1 \cdot m_2$ joint class-conditional densities $p(X_1, X_2 | v_k, \omega_i)$, each of which is of dimension $2n$.[1] Clearly, a large number of training samples will be required. When certain approximations can be justified, the situation is eased considerably. We shall now show that these approximations lead to a rather attractive model for a multitemporal classifier.

We are accustomed to assuming class-conditional independence in the spatial domain; i.e., given the class at a particular point, the random variable which is the measurement vector at that point is independent of the class or measurement vector at any other point. Applying this same idea to multitemporal measurements at a given point, we say that given the classes v_k at t_1 and ω_i at t_2, the random variables X_1 and X_2 are independent. Then we can write

$$p(X_1, X_2 | v_k, \omega_i) = p(X_1 | v_k, \omega_i)p(X_2 | v_k, \omega_i), \quad (10)$$

and furthermore,

$$\begin{aligned} p(X_1 | v_k, \omega_i) &\cong p(X_1 | v_k) \\ p(X_2 | v_k, \omega_i) &\cong p(X_2 | \omega_i). \end{aligned} \quad (11)$$

Imposing these conditions, it follows that

$$p(X_1, X_2 | v_k, \omega_i) = p(X_1 | v_k)p(X_2 | \omega_i).$$

The discriminant functions in (9) then become

$$\left\{ \sum_{k=1}^{m_1} p(X_1 | v_k)p(X_2 | \omega_i)p(\omega_i | v_k)p(v_k), \quad i = 1, 2, \cdots, m_2 \right\}. \quad (12)$$

From (12) we can model the discriminant function calculations as indicated in Fig. 1, from which we derive the term "cascade classifier" to describe this multistage classifier.

Simulation and Experimental Results

The cascade classifier model was programmed and applied to the analysis of a set of Landsat multispectral data. The data, collected by the satellite on two successive passes, 18 days apart,

[1] The observation vectors need not be of the same dimensionality. If X_1 has n_1 components and X_2 has n_2 components, the $p(X_1, X_2 | v, \omega)$ is N-variate, where $N = n_1 + n_2$.

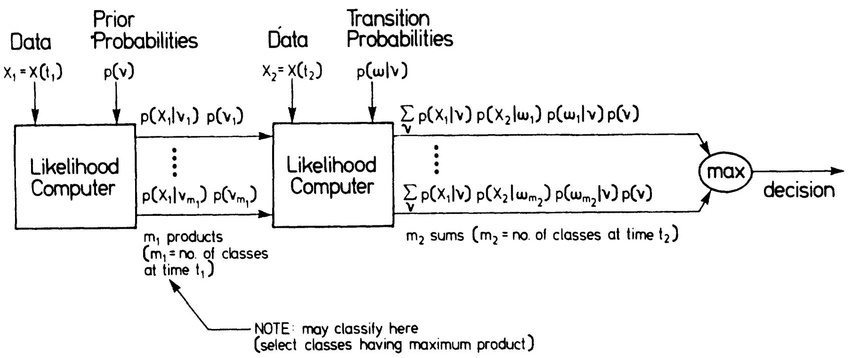

Fig. 1. Cascade classifier model.

TABLE I
TEST RESULTS FOR CLASSIFICATION OF THE FAYETTE COUNTY, IL, DATA

(a) June 29, 1973 data

Group	No. of Samples	Percent Correct	No. of Samples Classified into CORN	OTHERS	SOYBEAN	WOODS
CORN	186	65.1	121	36	24	5
OTHERS	100	40.0	33	40	22	5
SOYBEAN	227	82.4	10	30	187	0
WOODS	44	72.7	0	4	8	32
TOTAL	557		164	110	241	42

OVERALL PERFORMANCE = 68.2 percent correct

(b) July 17, 1973

Group	No. of Samples	Percent Correct	No. of Samples Classified Into CORN	OTHERS	SOYBEAN	WOODS
CORN	186	89.2	166	16	1	3
OTHERS	100	45.0	38	45	15	2
SOYBEAN	227	73.6	24	36	167	0
WOODS	44	56.8	4	9	6	25
TOTAL	557		232	106	189	30

OVERALL PERFORMANCE = 72.4 percent correct

(c) Multitemporal results (cascade classifier)

Group	No. of Samples	Percent Correct	No. of Samples Classified Into CORN	OTHER	SOYBEAN	WOODS
CORN	186	90.3	168	11	4	3
OTHERS	100	48.0	29	48	20	3
SOYBEAN	227	94.3	3	10	214	0
WOODS	44	84.1	0	5	2	37
TOTAL	557		200	74	240	43

OVERALL PERFORMANCE = 83.8 percent correct

over Fayette County, IL (see Table I), were geometrically registered at Purdue University's Laboratory for Applications of Remote Sensing. The objective of the analysis was to discriminate among the ground cover classes "corn," "soybeans," "woods," and "other," where the last category was simply a catchall consisting of water, pasture, fallow, and other relatively minor ground covers. Each class was actually decomposed in the analysis process into a union of subclasses, each having a data distribution describable as approximately multivariate normal.[2]

To provide a baseline for comparison, the data from each of the passes was first analyzed separately. The *a priori* probabilities of the classes were approximated as being equal, and 557 test samples, independent of the training samples, were used to evaluate the results. As shown in Table I(a) and (b), the performance of this conventional maximum likelihood classifier was 68 percent correct for the June 29, 1973 data, and 72 percent correct for the July 17, 1973 data.

[2] All probability densities were assumed to be multivariate normal (Gaussian), characterized by mean vector and covariance matrix.

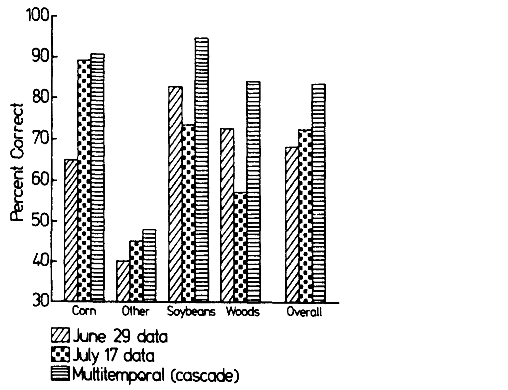

Fig. 2. Test results for Fayette County, IL, data.

To implement the cascade analysis, it was assumed unlikely that the ground cover would change identity over so short a time span. Accordingly, the transition probabilities were estimated as follows:

$$p(\omega_i \mid v_k) = 0.8, \qquad \text{for } \omega_i = v_k, \tag{13a}$$

and all other transition probabilities were set equal and such that

$$\sum_{\substack{i \\ \omega_i \neq v_k}} p(\omega_i \mid v_k) = 0.2. \tag{13b}$$

Again the *a priori* probabilities were assumed equal, and the same test samples were used to evaluate the results.

The results of this multitemporal classification, Table I(c), were substantially better than either of the unitemporal analyses. The overall results were 84 percent correct. In addition, the performance for each class was better than the best attained for the class in either of the unitemporal analyses. The unitemporal and multitemporal results are compared in Fig. 2.

The results can be sensitive, however, to the specification of the transition probabilities and *a priori* probabilities. This is demonstrated in the following experiment.

Landsat data from two passes over Grant County, KS, were analyzed in a manner similar to that used for the Fayette County data. In this case, the two passes were separated by more than two months, and a different set of classes was involved (Table II). The transition probabilities were specified as in (13a) and (13b); equal *a priori* probabilities were assumed.

TABLE II
TEST RESULTS FOR CLASSIFICATION OF THE GRANT COUNTY, KS, DATA

(a) May 9, 1974

Group	No. of Samples	Percent Correct	No. of Samples Classified Into ALFALFA	CORN	FALLOW	PASTURE	WHEAT
ALFALFA	58	84.5	49	0	0	0	9
CORN	428	57.0	0	244	183	1	0
FALLOW	526	54.4	0	196	286	36	8
PASTURE	1513	52.6	127	148	220	796	227
WHEAT	913	82.5	97	17	0	49	767
TOTAL	3455		273	605	689	882	1006

Overall Performance = 62.0 percent correct

(b) July 20, 1974

Group	No. of Samples	Percent Correct	No. of Samples Classified Into ALFALFA	CORN	FALLOW	PASTURE	WHEAT
ALFALFA	58	5.2	3	3	0	10	42
CORN	428	53.0	15	227	105	15	66
FALLOW	526	62.9	0	113	331	5	77
PASTURE	1513	42.4	64	329	213	641	266
WHEAT	913	76.2	22	108	33	58	709
TOTAL	3455		104	780	682	729	1160

Overall Performance = 55.3 percent correct

(c) Multitemporal results (cascade classifier)

Group	No. of Samples	Percent Correct	Number of samples classified Into ALFALFA	CORN	FALLOW	PASTURE	WHEAT
ALFALFA	58	41.4	24	0	0	2	32
CORN	428	59.6	5	255	165	1	2
FALLOW	526	76.4	0	107	402	2	15
PASTURE	1513	46.3	101	205	224	701	282
WHEAT	930	88.3	77	19	0	13	821
TOTAL	3455		207	586	791	719	1152

Overall Performance = 63.8 percent correct

As shown in Table II and Fig. 3, in this case the overall performance of the multitemporal cascade classifier was only marginally better than the best unitemporal result. A closer look at the class-by-class results is revealing. The largest detractors from the multitemporal results were the classes "alfalfa" and "pasture." In both of these cases, the unitemporal results for the second pass were substantially lower than those obtained in the first pass. (There are physical explanations for why this is reasonable, but this is not germane to our exploration of classifier behavior.)

Let us examine the impact that the relatively arbitrary assignment of transition probabilities has on the classification results. In case the actual transition probabilities are not known (which was true for the cited examples), the assignment can be made anywhere between two extremes. On the one hand, it could be assumed that

$$p(\omega_i \mid v_k) = \frac{1}{m_1}, \qquad k = 1, 2, \cdots, m_1,$$

i.e., equiprobable transitions. Then the discriminant functions have the form

$$\sum_{k=1}^{m_1} p(X_1 \mid v_k) p(X_2 \mid \omega_i) \frac{1}{m_1} p(v_k)$$

$$= \frac{1}{m_1} p(X_2 \mid \omega_i) \sum_{k=1}^{m_1} p(X_1 \mid v_k) p(v_k)$$

$$= \frac{1}{m_1} p(X_2 \mid \omega_i) p(X_1).$$

Since $1/m_1$ and $p(X_1)$ will be common to each of the discriminant functions, the decision will depend only on $p(X_2 \mid \omega_i)$ and will be independent of the first-stage results.

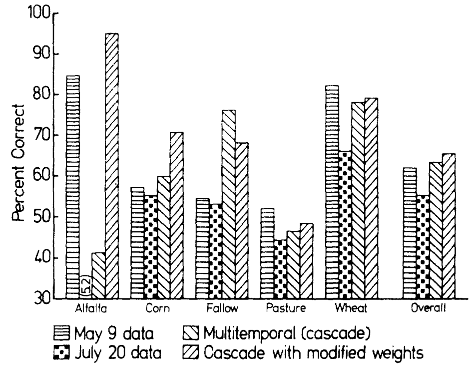

Fig. 3. Test results for Grant County, KS, data.

TABLE III
CASCADE CLASSIFIER RESULTS FOR ADJUSTED TRANSITION PROBABILITIES (GRANT COUNTY DATA)

Group	No. of Samples	Percent Correct	Number of samples classified Into ALFALFA	CORN	FALLOW	PASTURE	WHEAT
ALFALFA	58	94.8	55	0	0	0	3
CORN	428	70.3	5	301	122	0	0
FALLOW	526	68.1	0	139	358	7	22
PASTURE	1513	48.1	105	211	195	727	275
WHEAT	930	89.1	82	9	0	10	829
TOTAL	3455		247	660	675	744	1129

Overall Performance = 65.7 percent correct

On the other hand, we could make $p(\omega_i \mid v_i) = 1$ and $p(\omega_i \mid v_j) = 0, j \neq i$. Then the discriminant functions become

$$p(X_1 \mid v_i) p(X_2 \mid \omega_i) p(v_i).$$

Thus, in a sense, the contributions from the two stages are weighted equally. There is no way to make the first stage input dominate the second stage.

In view of these considerations, another classification of the Grant County data was performed. In this case, the transition probabilities $p(\omega_i \mid v_i)$ were set equal to unity for the "alfalfa" and "pasture" classes in order to give as much strength as possible to the first stage results. Table III and Fig. 3 show the outcome of this classification. The confusing influence resulting from the second stage data has been reduced.

It is interesting to compare the results obtained using the cascade classifier to results produced by a "conventional" maximum likelihood classifier using all of the multitemporal features simultaneously. To perform the latter classifications, equal *a priori* probabilities were assumed. The results were:

Fayette County	80.8 percent correct,
Grant County	64.1 percent correct.

It is curious that neither of these results is any better than the cascade classifier results achieved. It is possible that these slightly poorer results represent the price paid for having to estimate eight-dimensional statistics as opposed to four-dimensional statistics in the face of limited training data.

DISCUSSION AND CONCLUSIONS

The approach we have adopted for classifying data in a nonstationary environment was based on application of classical statistical decision theory in a straightforward manner. However, we used the conditions of the problem to approximate some of the statistical quantities involved. This step simplified the interdependencies of the data involved and led to a "cascade classifier" model. In the time-varying environment, this model is seen to

1) successfully incorporate the temporal information in the classification process, resulting in improved classification accuracy;

2) reduce the dimensionality of the probabilty functions used and thereby make less stringent demands with respect to the size of the training set required; and
3) facilitate distribution of the computational load over time.

Each time a set of observations becomes available, discriminant functions are calculated which can be used, if desired, to make a classification. However, the values of the discriminant functions are also passed along and contribute to a new set of discriminant functions calculated when the next set of observations is obtained. Although we have demonstrated the use of the cascade model only for the case of two stages, extension to an arbitrary number of stages presents no difficulty.

The prospective user of this approach should be aware that a casual implementation of the likelihood computers may result in computational difficulties of two sorts: loss of precision and very large computation times as compared with, let us say, a conventional Gaussian maximum likelihood classifier. Both of these difficulties can be overcome, or at least substantially reduced, by appropriate measures (scaling, ignoring zero terms, etc.) in carrying out the likelihood computations.

Finally, we have said nothing about how the transition probabilities might be obtained in practice. Certainly the transition probabilities will depend on the specific application under consideration, but conceptually they should be no harder to obtain than the prior probabilities used by "conventional" maximum likelihood classification (and by the first of the cascaded stages, for that matter). They may be estimated from historical observations, from ground sampling, or from observation of areas having similar ground cover and utilization.

ACKNOWLEDGMENT

The author wishes to thank Carlos A. Pomalaza for programming the cascade classifier model and testing it with the remote sensing data.

REFERENCES

[1] D. A. Landgrebe, "The quantitative approach: Concept and rationale," *Remote Sensing: The Quantitative Approach*. P. H. Swain and S. M. Davis, Eds. New York: McGraw-Hill, Ch. 1, 1979.
[2] N. J. Nilsson, *Learning Machines*. New York: McGraw-Hill, 1965.

Image Restoration Utilizing Spatial Masking of the Visual System

TOMA HENTEA AND BAHAA E. A. SALEH, MEMBER, IEEE

Abstract—**The fact that the visual system masks details near high-contrast edges is utilized in the design of filters for the restoration of blurred images. Psychophysically measured visibility functions are used to determine the size of oscillations that are allowed in an edge response. Filters are obtained that correspond to an edge response of the highest slope subject to this psychophysical constraint. Examples of the restoration of one- and two-dimensional images blurred by diffraction are discussed.**

Manuscript received June 20, 1978; revised July 25, 1978. This work was supported in part by the NSF under Grant ENG75-10063, and in part by a Fulbright Grant monitored by the Institute of International Education under program number 33737.
T. Hentea was with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA. He is now with the Facultatea de Electrotehnica, Timisoara, Romania.
B. E. A. Saleh was with the School of Optometry, University of California, Berkeley, CA. He is now with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706.

I. INTRODUCTION

The human visual system is usually the final stage of systems that involve coding, transmission, enhancement, or restoration of pictures. Many efforts have been dedicated to the incorporation of some aspects of the complex properties of the visual channel in the design of these systems. In this correspondence, we present some results on the restoration of blurred images using a criterion of fidelity that involves the spatial masking effect in the vicinity of high-contrast edges.

The spatial masking effect has been the subject of many experimental studies by visual scientists [1]–[7]. Its importance in the efficient coding of pictures [8] and in the elimination of noise in images [9] has been realized. We extend these studies to the problem of restoring images that are blurred by known amounts and that are to be viewed by human observers from fixed distances.

We limit ourselves to linear restorations, and we use the method of restoration by superposition of images [10]–[14]. In this method, the restored image at a certain position is taken as a weighted sum of the values of the measured image at a finite number of points surrounding that position. The weights, which determine the restoration filter, are chosen to satisfy some criterion of fidelity. Usually the point spread function (PSF) of the restoration system (i.e., the system that relates the restored object to the true object) is made as narrow as possible without containing large sidelobes [10]. If the image contains noise, the weights may be chosen to minimize the variance of the noise in the restored image [11]. In this correspondence, we are interested in pictures that contain high-contrast edges. The quality of a restoration scheme depends largely on how well the edges are restored. It is therefore more appropriate to consider the unit-step response of the restoration system. A unit-step function corresponds to a perfect edge. Success in restoring the edge is measured by the slope of the edge response at the position of the edge. But, as is known, an increase in this slope is always accompanied by ringing or oscillation effects on both sides of the edge. Those oscillations cause an unpleasant subjective appearance of the image. An optimum trade-off between the sharpness of the edge and the amplitude of these oscillations depends on subjective factors as well as on the nature of the picture to be restored.

At this point we utilize the masking property of the visual system. Details in the vicinity of the edge will not be perceived by the viewer. Although present, these ringing oscillations are masked or attenuated in the viewer's visual system. If such oscillations are not visible, why eliminate them at the expense of a reduction in the sharpness of the edge? The masking effect therefore sets our criterion for the amplitude of the permissible oscillations and consequently decides the optimum restoration weights.

In Section II, we discuss the masking effect. In Section III we formulate the restoration problem. Numerical examples of one- and two-dimensional systems blurred by diffraction are presented in Section IV.

II. THE EDGE MASKING EFFECT

In general, masking is defined as the action of one visual stimulus (primary stimulus) on the visibility of another (secondary stimulus) [1]–[7]. In our case, the primary stimulus is a high-contrast edge (Fig. 1(a)), and the secondary is some small detail located in the vicinity of the edge. It is known that the visibility of the secondary stimulus increases exponentially as a function of the distance from the edge [1]–[7]. The effect increases with the contrast and sharpness of the edge. In order to use the masking property in the design of filters for image restoration, knowledge of the actual dependence of the visibility on the distance is necessary. We have therefore performed some psychophysical experi-

0018-9472/78/1200-0883$00.75