χ^2 Test for Feature Detection

E-ren Chuang and David Sher
Department of Computer Science
State University of New York at Buffalo
224 Bell Hall
Buffalo, New York 14260

Abstract

We discuss χ^2 tests for detecting local visual features. We use simulated annealing to estimate the parameters of the blurring function in a given image automatically, and then include the blurring in the χ^2 test. In this paper, each feature with noise is modeled by a random vector $\mathbf{Y}$ with a multivariate normal distribution, denoted by $\mathbf{Y} \sim N(\mu_{\mathbf{y}}, \mathbf{\Sigma_y})$, and the mean vector μ_y and the variance-covariace matrix Σ_y include the intrinsic structure of the feature. The variance in the gaussian blurring is obtained by simulated annealing, and then a linear matrix of blurring $\mathbf{B}$ is generated to blur each feature $\mathbf{Y}$. Let $\mathbf{Z} = \mathbf{BY} + \mathbf{N_1}$, where $\mathbf{N_1} \sim N(\mathbf{0}, \mathbf{\Sigma_n})$, then $\mathbf{Z} \sim N(\mu_{\mathbf{z}}, \mathbf{\Sigma_z}) = \mathrm{N}(\mathbf{B}\mu_{\mathbf{y}}, \mathbf{B^t\Sigma_yB} + \mathbf{\Sigma_n})$. After the transformation $\mathbf{f(z)}$: $(\mathbf{Z} - \mu_{\mathbf{z}})^{\mathbf{t}}\mathbf{\Sigma_z^{-1}}(\mathbf{Z} - \mathbf{B}\mu_{\mathbf{z}})$, the random vector $\mathbf{Z}$ becomes a random variable with χ^2 distribution, then χ^2 test is applied to measure the similarity between data and the expectation vector of each model.

Key words: Simulated annealing χ^2 test Gaussian Blur Feature Detection

1 INTRODUCTION

Features are importance in object representation and pattern recognition. Early research ($\sim$ 1980) focused on extracting the features of an object directly from an image. Rapid intensity changes in an image imply the existence of the boundary between two regions. These changes, caused by edges, or corners, are significant because they convey important information, e.g. a few strokes in cartoon can describe the essence of an object. However, the process is difficult because the intensity of an object in the image depends on the following unknown factors: the reflectance of the object surface, the orientation of the surface, and the angle between the directions toward viewpoint and light source(1). In addition, noise and blurring make it harder to recover these properties from the image.

Template matching is a typical method to find a feature, $\mathrm{f}(\mathbf{x})$, in an image $\mathrm{g}(\mathbf{x})$. Euclidean distance can be used to measure the degree of match between f and g, i.e.

$$d(\mathbf{y}) = \sum_{\mathbf{x}} (g(\mathbf{x}) - f(\mathbf{x}-\mathbf{y}))^2$$

If we expand the square expression in summation, and neglect the constant term $\sum_{\mathbf{x}} f(\mathbf{x}-\mathbf{y})^2$, and $\sum_{\mathbf{x}} g(\mathbf{x})^2$, the term left $\sum_{\mathbf{x}} g(\mathbf{x})f(\mathbf{x}-\mathbf{y})$ is called the cross correlation between g and f. If the cross correlation is "high" at the position $\mathbf{y}$, then the feature $\mathbf{f}$ has a high probability. However. the result of template matching depends on the thresholding strategy, and the performance is sensitive to the variance of luminance and noise.

By modification on the values in a template, one can perform differentiation[2,3,4,5], surface fitting[6,7,8,9] or other operation[10], to detect features, like edges or corners. Various thresholding techniques have been devised, like zero-crossing, edge space project, or $F_{2,20}$ hypothesis test. By zero-crossing[2,6,11,12], edge detectors locate edges at places where the outcome of the second derivative of Gaussion on the image change in sign from positive to negative (zero-crossing), or vice versa. Another thresholding strategy [5] projects a local area in an image to an orthogonal space, and measure how close the area is to the edge basis. Nalwa[8] applies a sequence of least-squares tests, and $F_{2,20}$ test to estimate the direction of the edge, and the intensities between the edge, then interprets that edge by hyperbolic tangent **tanh** with these estimated parameters.

Most of those methods above do not provide flexible estimation of blurring in a real image, they are hard to generalize to detect different kinds of features, and the results of their operations have not been able to be combined statistically. This paper defines a feature such as an edge with multiple models. Each model is a random vector with multivariate normal distribution. The structure, and properties of each feature are included in the mean vector, and in the variance-covariance (or dispersion) matrix of the random vector. Noise and blurring which is modeled by Gaussian function is included in the the models. Those parameters in the blurring model is estimated by simulated annealing, then the Gaussian blur is approximated by a linear system. After establishing models, how hypothesis tests are applied to measure the similarity between data and each model is discussed. In the third section, practical implementation, and experimental results are presented. The conclusion and the applicability of this feature detector will be discussed at the last.

2 MODELING & DETECTION

Unlike probability which establishes mathematical models to predict the outcomes of an experiment, statistics[13] provides tools to estimate how well the given experiment outcomes fit into the predefined models, therefore, we apply the statistical approach to detect feature, includeing noise, and blurring.

2.1 Feature Models

Multiple models are proposed for ideal features on boundaries such as edges and corners. A ideal feature ω is modeled by a n dimensional multivariate normal random vector $\mathbf{X}_\omega = (X_1, \ldots, X_n)'$, denoted by $\mathbf{X}_\omega \sim N_n(\mu_{x_\omega}, \Sigma_{x_\omega})$, where $\mu_{\mathbf{x}_\omega}$, and $\mathbf{\Sigma}_{\mathbf{x}_\omega}$ are mean vector and variance-covariance (or dispersion) matrix. The joint distribution function of $\mathbf{X}_\omega$ is as follows:

$$f(\mathbf{x}_\omega) = \frac{1}{(2\pi)^{n/2}(\det\mathbf{\Sigma}_{\mathbf{X}_\omega})^{1/2}} exp(-\frac{(\mathbf{x}_\omega - \mu_{\mathbf{X}_\omega})^{\mathbf{t}}\mathbf{\Sigma}^{-1}_{\mathbf{X}_\omega}(\mathbf{x}_\omega - \mu_{\mathbf{X}_\omega})}{2}) \tag{1}$$

The expectation vector of $\mathbf{X}_\omega$ represents the brightness, and the structure of the feature ω, and the dispersion matrix includes the interrelationship among these random variables. Without losing generality, we assume the first m random variables in the model fall into a class, and the next $n-m$ random variables fall into the other class, and these random variables in a class have the following properties:

1. The relationship between any two r.v.'s in a class can be described by a linear equation, i.e. the correlation coefficient of them is one. This is justified because the reflectance of these two points in the same area is almost equal.

2. Any two r.v.'s between classes are uncorrelated because their intensities are not mutually dependent, and their the correlation coefficient is zero.

Hence, the dispersion matrix of $\mathbf{X}_\omega$ is $\begin{pmatrix} \mathbf{\Sigma_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_2} \end{pmatrix}$, where $\mathbf{\Sigma_1}, \mathbf{\Sigma_2}$ are $m \times m$ and $(n-m) \times (n-m)$ matrixs, and every element in them is σ_x^2, supposing each random variable in $\mathbf{X}_\omega$ has the same variance. This matrix $\mathbf{\Sigma_x}$ is singular, Intuitively, this singular normal distribution is concentrated in an two-dimensional sub-space of $\mathbf{R}^n$. However, in[14], let $\mathbf{\Sigma}_k = \mathbf{\Sigma}_x + \frac{1}{k}\mathbf{I}$, where $k \in 1, 2, \cdots$, $\mathbf{I}$ is an $n \times n$ identity matrix, and let $f_k(\mathbf{t}) = exp(i\mu\mathbf{t}' - \frac{1}{2}\mathbf{t}\mathbf{\Sigma}_k\mathbf{t}')$. It is the characteristic function of a multivariate normal vector. By the continuity theorem, the limiting function

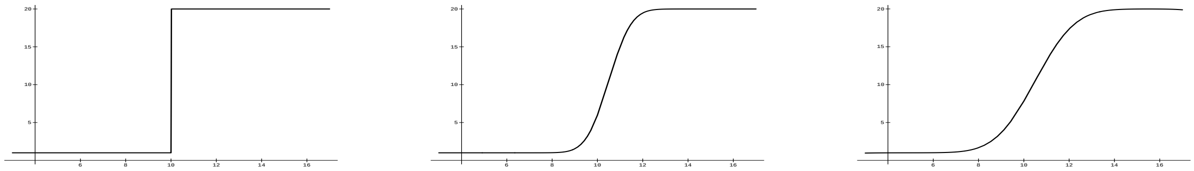
$$f(\mathbf{t}) = \lim_{k\to\infty} f_k(\mathbf{t}) = exp(i\mu\mathbf{t}' - \frac{1}{2}\mathbf{t}\mathbf{\Sigma}_{\mathbf{X}_\omega}\mathbf{t}')$$

is also a characteristic function. Hence, the corresponding random vector has singular normal distribution.

2.2 Noise Models

There are various kinds of noise in images. The roughness of a surface causes difference in intensity. During sampling, inaccurate measurement or quantization cause errors. Thermal noise or channel noise are also introduced during image formation or transmit. Those kinds of noise is frequently modeled by additive gaussion noise. It can be defined as a $n \times 1$ multivariate normal vector $\mathbf{N}$. Random vector $\mathbf{N}$ has normal distribution with zero mean and dispersion matrix $\mathbf{\Sigma_n}$. Elements in $\mathbf{N}$ are independent, and, $\mathbf{X}$ and $\mathbf{N}$ are also mutually independent.

Let $\mathbf{N}_b$, and $\mathbf{N}_a$ denote models for noise before, and after image formation, respectively. then the random vector $\mathbf{Y} = \mathbf{X} + \mathbf{N}_b$ still has multivariate normal distribution



(a) (b) (c)

Figure 1: This figure demonstrates the effect of the variance σ^2 in one dimentional gaussian function on a step edge. In (a) is an ideal step edge, and in (b) and (c) are the blurred ones of (a) by σ^2 0.7 and 2.0, respectively.

$N_n(\mu_{\mathbf{y}}, \mathbf{\Sigma_y}) = N_n(\mu_{\mathbf{x}}, \mathbf{\Sigma_x} + \mathbf{\Sigma n})^{(15)}$. Let each random variable in $\mathbf{X}$ or $\mathbf{N}_b$ has variance σ_x^2 or σ_n^2 respectively, then the dispersion matrix of $\mathbf{Y}$ is $\begin{pmatrix} \mathbf{\Sigma_1'} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_2'} \end{pmatrix}$, where $\mathbf{\Sigma_1'}$, and $\mathbf{\Sigma_2'}$ $m \times m$ and $(n-m) \times (n-m)$ matrix. Let a_{ij} denotes the element in the ith row, and the jth column of $\mathbf{\Sigma_1'}$, or $\mathbf{\Sigma_2'}$, then

$$a_{ij} = \begin{cases} \sigma_x^2 + \sigma_n^2 & \text{if } i = j \\ \sigma_x^2 & \text{otherwise} \end{cases}$$

In appendix A, we prove that the inverse of $\mathbf{\Sigma_y}$ is positive definite, *i.e.* its eigen values are all positive, so it satisfies that condition of dispersion matrix. The effect of noise is included in each model of features.

2.3 Blurring Models

Defocusing causes blurring in an image, so does spread out of scanning beam in vidicon, and imperfect point spread function which is supposed to be an exact unit impulse at every sampling location during image formation, therefore the intensity of a point spreads over to its neighborhood. Blurring in an image is defined as a convolution of an ideal image with a blurring function g. This function is usually modeled by a pillbox or a Gaussian function(1,16), and the dispersion matrix of the two dimensional Gaussian function (or the diameter of the pillbox) determines the spread out of blurring, shown in Fig. 1 This area between two planar area is called transition area, and a transition area is consistent if the width and the boundary directions of the area are equal.

Two dimensional (2D) Gaussian function is adopted as the model for blurring. There are three unknown parameters in the dispersion matrix, the variance σ_x^2 , and σ_y^2 in X, and

Y directions, and the correlation coefficient ρ_{xy}, therefore, the configuration domain of the dispersion matrix is a subset B of R^3. The optimal configuration, which blurs the model $\mathbf{Y}$, is estimated by simulated annealing which is one of the most popular methods to solve optimization problems. Simulated annealing is formalized as a pair (D, C), where D is parameter space (or configuration space), and C is a cost function $C : D \rightarrow R$, which maps a configuration to a real number. The inverse problem of blurring becomes choosing a configuration i in D which has a minimal result from a cost function C under equilibrium state of small control parameter in simulated annealing.

In this paper, the cost function in the simulated annealing algorithm is *Mahalanobis* distance [17]. Given a randomly chosen configuration i in *Monte Carlo algorithm*, the 2D Gaussian blurring function can be approximated by a $n \times n$ linear system $\mathbf{B}_i$. Given a feature model $\mathbf{Y}$, and noise $\mathbf{N}_a$, let $\mathbf{Z} = \mathbf{B}_i\mathbf{Y} + \mathbf{N}_a$, then $\mathbf{Z} \sim N_n(\mu_\mathbf{Z}, \mathbf{\Sigma_Z}) = N_n(\mathbf{B}\mu_\mathbf{Y}, \mathbf{B\Sigma_Y B} + \mathbf{\Sigma}_{n_a})$ [18], then the *Mahalanobis* distance between the data and the blurred ideal model is

$$(\mathbf{Z} - \mu_\mathbf{Z})^\mathbf{t}\mathbf{\Sigma}_\mathbf{Z}^{-1}(\mathbf{Z} - \mu_\mathbf{Z})$$

The algorithm repeat the above step until equilibrium state is sufficiently close, then decrease the value of the control parameter in the algorithm, and repeat all over again until converge to a globally minimal configuration $\mathbf{B}_{opt}$ [19]. In our implementation we find values for σ_x^2 and σ_y^2 initially using a one dimensional search. Given the values for σ_x^2 and σ_y^2, we search for σ_{xy} with a one dimensional search. This procedure is much more efficient and accurate than annealing in three dimensions directly.

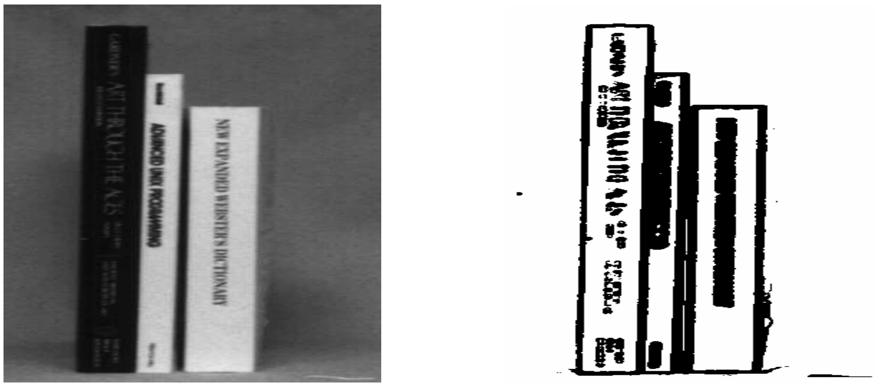
2.4 Detection

In previous three subsections, we have described how to model a feature by random vector $\mathbf{Y}$ with multivariate normal distribution, how to estimate a optimal blurring function $\mathbf{B}_{opt}$ by simulated annealing, and how to include the effect of blurring and independent noise $\mathbf{N}_a$ into models. Given $\mathbf{Y}, \mathbf{B}_{opt}, \mathbf{N}_a$, χ^2 hypothesis test will be applied to feature detection. In Statistics, hypothesis is an assurtion about a unknown parameter θ of one or more random variables A test of hypothesis is a rule (or statistic) which leads to the decision of the acceptance or rejection of the null hypothesis $H_0 : \theta \in \Theta$, or alternative hypothesis $H_a : \theta \in \Theta^c$. Let $\mathbf{Z} = \mathbf{B}_{opt}\mathbf{Y} + \mathbf{N}_a$, the hypothesis test of the feature is as follows:

$H_0 : \mu = \mu_\mathbf{Z}$
$H_1 : \mu \neq \mu_\mathbf{Z}$
The statistic: $S(\mathbf{Z}) = (\mathbf{Z} - \mu_\mathbf{Z})^\mathbf{t}\mathbf{\Sigma}^{-1}(\mathbf{Z} - \mu_\mathbf{Z})$
The critical region $C(\alpha) = \{y : S(\mathbf{z}) > c_\alpha\}$

Since $\mathbf{Z} \sim N_n(\mu_\mathbf{Z}, \mathbf{\Sigma_Z})$, let $Z = (\mathbf{Z} - \mu_\mathbf{Z})^\mathbf{t}\mathbf{\Sigma}^{-1}(\mathbf{Z} - \mu_\mathbf{Z})$, then Z is a random variable with χ_n^2 distribution of n degree of freedom [15].

In general, each hypothesis test has a corresponding critical region C. Given a sample point $\mathbf{z}$, if $\mathbf{z} \in C$, the H_0 is rejected. If $\mathbf{z} \notin C$, H_0 will not be rejected. The size of the critical region, α, (or the significance level of the test) is defined as the error probability that $\mathbf{x}$ falls



(a) (b)

Figure 2: In (b) is shown the transition regions of (a) which contains three books with noise and shading.

in C when H_0 is true. In the above hypothesis, the set $C(\alpha)$ denotes the upper α percent points of the χ^2_n distribution. Given the data $\mathbf{z}$ of a local area, if $(\mathbf{z}-\mu_{\mathbf{Z}})^{\mathbf{t}}\mathbf{\Sigma}_{\mathbf{Z}}^{-\mathbf{1}}(\mathbf{z}-\mu_{\mathbf{Z}}) > c_\alpha$, the null hypothesis H_0 is rejected, i.e. the local area does not contain this feature. When a significance level is chosen, c_α is determined accordingly by looking up χ^2 table.

It is very possible that the degree of freedom of a χ^2 distribution is very large so that the c_α is beyond the scope of tabulated χ^2 distribution. This problem can be solved by the following theorem [18].

Theorem 1. If $Z \sim \chi^2_n$, then

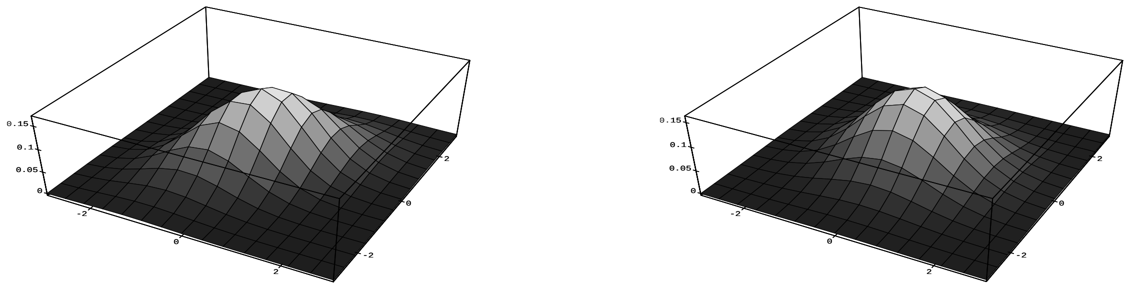
$$\lim_{n\to\infty} P(\sqrt{2Z}-\sqrt{2n-1}) \leq a = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Let $Z \sim \chi^2_n$ and $W = \sqrt{2Z}-\sqrt{2n-1}$. According to theorem 1, the distribution of W approaches to normal distribution while the degree of freedom, n, of X, is large. Hence, after this transformation, the threshold value is determined by looking up the table of standard normal distribution.

3 Implementation

In this section, we discuss the pratical implementation of the estimation of blurring function, and this feature detector. The experimental results of the blurring fuction are shown in Fig. 3 and 4, and comparisons among the results of the proposed feature detector, Nalwa's, and Marr's are shown in fig. 5, 6, 7. There are three steps in our implementation, described as follows:

1. Eliminating the horizontal regions from an input image. The output of the step, denoted by $\mathbf{O}_1$, is a map containing shading or transition areas in the input image.



(a) (b)

Figure 3: The estimated blurring function is shown in (a) which is not as circularly symmetric as a bivariate normal distribution $N_2(\mathbf{0}, \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix})$ in (b).

2. Estimating the blurring function by simulated annealing on consistent areas in $\mathbf{O}_1$.

3. Applying χ^2 tests on $\mathbf{O}_1$ to detect features.

The dispersion matrixes in the 5×5 models of horizontal plain and edges in different directions are automatically generated. In order to avoid enumerating all of the possible mean vectors for each feature, we directly estimate the vector of each feature by maximun likelihood estimation on the local area where the model is applying. This estimation is based on the first property in section 2.1 that pixels in a class have the same reflectance, *i.e.* if the amounts of light falling on pixels are equal, the brightness of these pixels should be equal. Given 0.05 probability that a true local horizontal region will not be detected, χ^2 tests apply to to eliminate those regions in the image first with threshold 37.652 from χ^2 table. Afterward, estimation of blurring fuction, and then χ^2 tests of the other features will only apply to areas of shading or transition area, shown in Fig. 2

We use Mitsubishi SC-7000 scanner to scan pictures taken by Nicon N4004 camara, shown in Fig. 5(a), and Fig. 6(a). By simulated annealing, the parameter σ_x, σ_y, and ρ_{xy} in the gaussian blur are estimated. However, we apply the following technique in our program to reduce the computation time of this estimation. First of all, filter consistent areas from $\mathbf{O}_1$, and choose four sets of points, $S_i, i = 1, \ldots, 4$ from different consistent areas, but points in S_1 belong to a consistent area which direction is equal (or close) to X axis, so do points in S_2 to Y axis. Secondly, assuming σ_y, σ_x are equal and independent, σ_x is estimated first by simulated annealing on S_1 and then take the average of these points which would converge

172	157	127	101	94
173	156	126	102	94
174	156	124	103	95
173	156	125	103	96
172	156	126	103	97

(a)

179	179	179	94	94
179	179	179	94	94
179	179	179	94	94
179	179	179	94	94
179	179	179	94	94

(b)

176	156	124	103	97
176	156	124	103	97
176	156	124	103	97
176	156	124	103	97
176	156	124	103	97

(c)

Figure 4: This figure show that the estimated 2D gaussian blur function can transform an ideal feature (b), which is a vertical edge in this case, to (c) which is close to a local area (a) in a real image. The *Mahalanobis* distance between (a) and (c) is 4.04. It is small in comparison to the threshold 34.38 which is determined by looking up χ^2 talbe of 25 degree of freedom.

as result. Similarly, σ_y is estimated on S_2. With those estimated σ_x, and σ_y, ρ_{xy} could also be obtained on S_3. Finally, these three parameters are fine-tuned by simulated annealing on the last set of points S_4. The experimental result shows that dispersion matrix in the gaussian blur function is as follows:

$$\begin{pmatrix} 1.05 & 0.5 \\ 0.5 & 1.2 \end{pmatrix},$$

The unknown parameter $\sigma_x^2 = 1.05, \sigma_y^2 = 1.2$, and $\rho_{xy}\sigma_x\sigma_y = 0.5$. These parameters suggest that the blurring function is not circularly symmetric, *i.e*, it has different effects in different directions, and the correlation between the effects in X, and Y direction is not independent. This blurring function is depicted in Fig. 3. The gaussian blur function can transform an ideal feature close to the one in the real image, shown in Fig. 4.

Comparisons are made among our operator, Nalwa's[1] [8], and Marr's[2] [2]. Fig. 5(a) has a stopsign in cloudy sky, Fig. 6(a) has three books, and Fig. 7(a) is a scene of an office. This second images is difficult because it is dark, noisy, defocusing, and has shading between books and at the lower left side of those books. Fig. 7(a) is also difficult because of different illumination, unclean blackboard, differect blurring from area to area. Experimental results show that our feature detector has less false positive on Fig. 5(a), and Fig. 6(a) and reasonable result on Fig. 7(a)

4 DISCUSSION & CONCLUSION

There are several merits in the feature detector in this paper. First, this feature detector can not only apply to edge detection, but also other features, such as corners. An ideal feature ω is modeled by random vectors $\mathbf{X}_\omega$ with multivariate normal distributions, and

[1]With courtesy of Vishvjit S. Nalwa, we apply his 5×5 edgel operator with $\sigma_{blur} = 0.6$ on images.

[2]With convenience, we apply Marr's Difference of Gaussians (D.O.G.) operator, in the HIPL software package. The Gaussians we applied are with variance 1.6, and 1.0 on 11×11 template. The zero-crossing were thresholded on their slope.

the intrinsic structure of this feature is included in the dispersion matrix of $\mathbf{X}_\omega$. Besides, the effect of the noise before and after sampling is embeded, and then generate a new model $\mathbf{Y}_\omega$ which is also a random vector with multivariate normal distribution.

Secondly, blurring which is caused by imperfect point spread function during sampling is modeled by 2D Gaussian function $\mathbf{B}$. These parameters σ_x, σ_y, and ρ_{xy} in this model are not fixed, but estimated directly from the image by simulated annealing. The cost function in the simulated annealing algorithm is *Mahalanobis* distance $(\mathbf{z}-\mu_\mathbf{z})^\mathbf{t}\boldsymbol{\Sigma}_\mathbf{z}^{-1}(\mathbf{z}-\mu_\mathbf{z})$ between data $\mathbf{z}$ and the blurred model $\mathbf{Z} = \mathbf{B}\ \mathbf{Y}$. Experimental result shows that $\mathbf{B}$ is not only circularly symmetric, but also not independent in X and Y direction.

Thirdly, we adopt statistical approach to detect feature. This blurred model $\mathbf{Z}$ can be transformed to a random variable Z with χ^2 distribution. Therefore, a χ^2 test determines whether the feature ω exists in the local area where the test applies to. Given a significance level, a threshold value is determined by looking up χ^2 table, instead of by heuristic, and the probability of falso negative is also determined accordingly. If the degree of freedom of Z, n, is too large so that the threshold value is beyond the scope of tabulated χ^2 table, we can choose one from the table of standard normal distribution after the transformation $\sqrt{2Z}-\sqrt{2n-1}$.

Finally, multiple statistics from different χ^2 tests are evidence of features. Those evidence represent the accuracy of those measurements. Those statistics from the tests of local features, such as edges and corners, are combined by summing rule[(20)] into statistics for larger features such as segments and polygons, and the geometric meaning of these addends are reserved in the sum. While applied to Hough transform, these statistics and summing rule have better performance in Hough transform than merely edge points as evidence[(21)].

A PROPERTIES OF THE DISPERSION MATRIX

In this appendix, it is proved that the inverse of the dispersion matrix $\boldsymbol{\Sigma}_{\mathbf{X}_\omega}$ which is made in **Section 3** is positive definite so that it satisfies the conditions of dispersion matrix.

Lemma 1. Let $a > b$, and $\boldsymbol{\Sigma}_{\mathbf{n},\mathbf{0}} = (\sigma_{i,j})$ be a $n \times n$ matrix, and $\sigma_{i,j}{=}a$ if $i = j$, $\sigma_{i,j}{=}b$, otherwise, then the determinant of $\boldsymbol{\Sigma}_{\mathbf{n},\mathbf{0}} = (a+(n-1)b)(a-b)^{n-1}$.

Proof:

After adding row 2, ..., and n to row 1, subtracting column 1 from column 2, ..., n, and then subtracting row $1 \times b/(a+(n-1)b)$ from row 2, ..., n, $\boldsymbol{\Sigma}_{\mathbf{n},\mathbf{0}}$ becomes a diagonal matrix $(\sigma_{i,j})$, the $\mathrm{Det}(\boldsymbol{\Sigma}_{\mathbf{n},\mathbf{0}}) = \prod_{i=1}^n \sigma_{i,i} = (a+(n-1)b)(a-b)^{n-1}$.

Lemma 2. Let $\boldsymbol{\Sigma}_{\mathbf{n},\mathbf{1}} = (\sigma_{i,j})$ be a $n \times n$ matrix, $\sigma_{1,j} = b$, $\sigma_{i,j} = a$ if $i = j$, and equals b, otherwise, then the determinant of $\boldsymbol{\Sigma}_{\mathbf{n},\mathbf{1}} = b(a-b)^{n-1}$.

Proof: Similar to Lemma 1.

Lemma 3. Let $\mathbf{A_{m,m}}, \mathbf{B_{n,n}}$ be $m \times m, n \times n$ matrices, respectively, then the determinent of $\begin{bmatrix} \mathbf{A_{m,m}} & \mathbf{0} \\ \mathbf{0} & \mathbf{B_{n,n}} \end{bmatrix}$= Det($\mathbf{A}$) $\cdot$ Det($\mathbf{B}$).

Proof: Can be easily proved by **Gauss-Jordan** elimination.

Theorem 2. Let $\mathbf{C_{m+n,m+n}} = (c_{i,j}) = \begin{bmatrix} \mathbf{\Sigma_{m,0}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_{n,0}} \end{bmatrix}$, then the eigen vectors of $\mathbf{C^{-1}_{m+n,m+n}}$ are $\begin{bmatrix} \mathbf{\Pi_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi_n} \end{bmatrix}$, where $\mathbf{\Pi_k} = \{\underbrace{(\overbrace{1,\cdots,1}^{k}), (\overbrace{1,-1,0,\cdots,0}^{k}), (\overbrace{1,0,-1,0,\cdots,0}^{k}), \cdots, (\overbrace{1,0,\cdots,0,-1}^{k})}_{k \quad vectors}\}$, and the eigen values of $\mathbf{C^{-1}_{m+n,m+n}}$ are all positive.

Proof: Let $\mathbf{D_{i,j}}$ be a $(m+n-1)\times(m+n-1)$ matrix, derieved from $\mathbf{C_{m+n,m+n}}$ by deleting jth row, and ith column, and $d_{i,j}$=Det($\mathbf{D_{i,j}}$). By $(i+j-3)$ row or column interchanges, then

$$\mathbf{D_{i,j}} = \begin{cases} \begin{bmatrix} \Sigma_{m-1,1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{n,0} \end{bmatrix}, & \text{if} i,j \le m, \\ \\ \begin{bmatrix} \Sigma_{m,0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{n-1,1} \end{bmatrix}, & \text{if} i,j > m \end{cases}$$

then $\mathbf{C^{-1}_{m+n,m+n}} = \frac{(c'_{i,j})}{\text{Det}(\mathbf{C_{m+n,m+n}})}$, where

$$c'_{ij} = \begin{cases} 0, & \text{if } c_{i,j} = 0 \text{ becuase } \mathbf{D_{i,j}} \text{ is linear dependent} \\ (-1)^{2(i+j)-3} \cdot d_{i,j}, & \text{otherwise} \end{cases}$$

Let $\alpha = (a+(m-1)b)(a+(n-1)b)(a-b)^{m+n-2}, \beta = (a+(m-2)b)(a+(n-1)b)(a-b)^{m+n-3}, \gamma = -b(a+(n-1)b)(a-b)^{m+n-3}$, then the $\mathbf{C^{-1}_{m+n,m+n}}$'s eigen vector $(\underbrace{1,\cdots,1}_{m},\underbrace{0,\cdots,0}_{n})\cdot\mathbf{C^{-1}_{m+n,m+n}} = \frac{\beta+(m-1)\gamma}{\alpha}(\underbrace{1,\cdots,1}_{m},\underbrace{0,\cdots,0}_{n}) = \frac{1}{a+(m-1)b}(\underbrace{1,\cdots,1}_{m},\underbrace{0,\cdots,0}_{n})$, and $(1,-1,\underbrace{0,\cdots,0}_{m+n-2}) \cdot \mathbf{C^{-1}_{m+n,m+n}} = \frac{\beta+\gamma}{\alpha}(1,-1,\underbrace{0,\cdots,0}_{m+n-2}) = \frac{1}{a-b}(1,-1,\underbrace{0,\cdots,0}_{m+n-2})$ By the same verification to the rest eigen vectors, we can obtain these eigen values, $\{ \frac{1}{a+(m-1)b}, \frac{1}{a+(n-1)b}, \underbrace{\frac{1}{a-b}, \cdots, \frac{1}{a-b}}_{m+n-2} \}$. Because $a > b$, these eigen values are all positive, then $\mathbf{C^{-1}_{m+n,m+n}}$ is positive definite. #

References

[1] Berthold Klaus Paul Horn. *Robot Vision.* The MIT Press and McGraw-Hill Book Company, 1986.

[2] D. Marr and E. Hildreth. Theory of edge detection. *Proc. R. Soc. Lond. B*, 207, 1980.

[3] J. F. Canny. A computational approach to edge-detection. *IEEE Trans. Pattern. Anal. Machine Intell.*, PAMI-8, 1986.

[4] R. A. Kirsch. Computer dermination of the constituent structure of biological images. *Computers and Biomidical Research*, 4, June 1971.

[5] R. C. Frei and C. C. Chen. Fast boundary detection: a generalization and a new algorithm. *IEEE Transaction on Computer*, 26, October 1977.

[6] Robert M. Haralick. Digital step edges from zero crossing of second derectional derivatives. *IEEE Trans. Pattern. Anal. Machine Intell.*, PAME-6, January 1984.

[7] A. Hueckel. An operator which locates edges in degital pictures. *Journal of the ACM*, 20, 1971.

[8] V. Nalwa. On detecting edges. *IEEE Trans. Pattern. Anal. Machine Intell.*, PAMI-8, 1986.

[9] David Sher. Tunable facet model likelihood generators for boundary pixel detection. In *IEEE ComputerSociety Workshop on Computer Vision*, November 1987.

[10] Krishnan Rangarajan, Mubarak Shah, and David Van Brackle. Optimal corner detector. *Computer Vision, Graphics, and Image Processing*, 48, 1989.

[11] A.L. Yuille and T.A. Poggio. Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:15–25, 1986.

[12] A.L. Yuille. Zero crossings on lines of curvature. *Computer Vision, Graphics, and Image Processing*, 45:68–87, 1989.

[13] Robert V. Hogg and Allen T. Craig. *Introduction to Mathematical Statistics*. The Macmillan Company, London, fourth edition edition, 1979.

[14] E. Lukacs and R.G. Laha. *Applications Of Characteristic Functions*. Hafner Publishing Company, New York, 1964.

[15] Robb J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley and Sons, Inc., 1982.

[16] Robert A. Hummel, B. Kimia, and Steven W. Zucker. Deblurring gaussian blur. *CVGIP*, 38, 1987.

[17] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, 1973.

[18] V. K. Rohatgi. *An introduction to Probability theory and mathematical statistics*. John Wiley and Sons, Inc., 1976.

[19] P.J.M. van Laarhoven and E.H.L. Aarts. *Simulated Annealing: Theory and Applications*. The Kluwer Academic Publishers, 1987.

[20] William James. *Psychology.* Holt, New York, 1890.

[21] E-Ren Chuang and David B. Sher. Evidence combination & verification in low-level vision. *Pattern Recognition*, (under review).

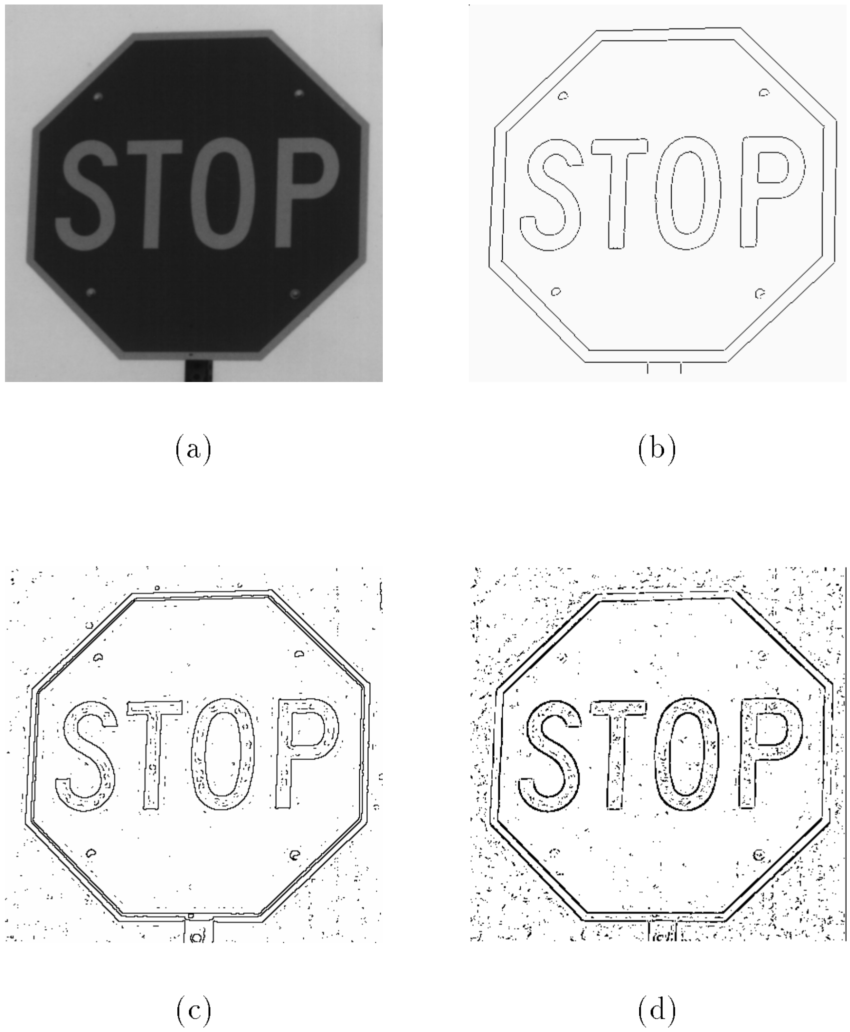
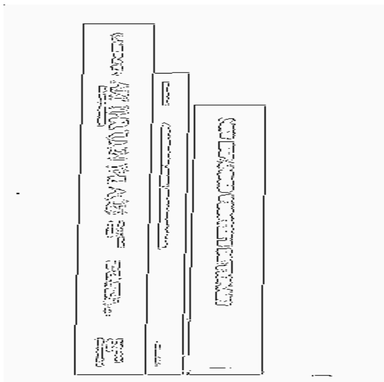
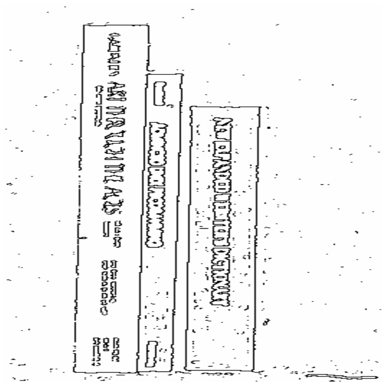


Figure 5: In (a) is the original image of a stop sign with background in cloudy sky. The results of our χ^2 feature detector, Nalwa's **tanh** edgel finder, and Marr's D.O.G (Difference of Gaussians) operator are shown in (b), (c), and (d), respectively.

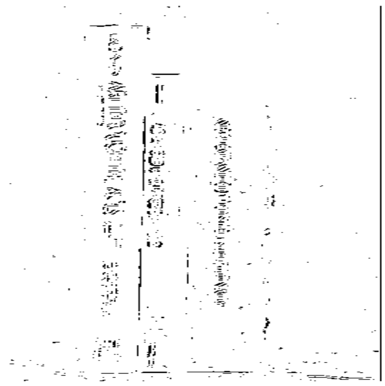
(a)



(b)



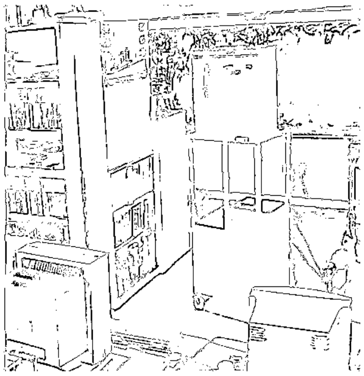
(c)



(d)

Figure 6: In (a) is the original image of books in dark and noisy background with shading at the lower left side, and between books. The results of our χ^2 feature detector, Nalwa's **tanh** edgel finder, and Marr's D.O.G (Difference of Gaussians) operator are shown in (b), (c), and (d), respectively.

(a)



(b)

(c)

(d)

Figure 7: In (a) is the original image of an office. This image has different illumination at different areas, and is ver noisy on that unclean blackboard. The results of our χ^2 feature detector, Nalwa's **tanh** edgel finder, and Marr's D.O.G (Difference of Gaussians) operator are shown in (b), (c), and (d), respectively.