LARS Information Note 090274

Image Registration Error Variance As a Measure of Overlay Quality

C. D. McGillem
M. Svedlow

The Laboratory for Applications of Remote Sensing

Purdue University, West Lafayette, Indiana

1974

LARS Information Note 090274

IMAGE REGISTRATION ERROR VARIANCE AS A MEASURE OF OVERLAY QUALITY

C.D. McGillem

M. Svedlow

Laboratory for Application
of Remote Sensing
Purdue University
W. Lafayette, Indiana

This work was sponsored by the National Aeronautics and Space Administration under Grant Number NGL-15-005-112.

TABLE OF CONTENTS

IMAGE REGISTRATION ERROR VARIANCE AS A MEASURE OF OVERLAY QUALITY

Introduction

A model for the variance of the error in the registration of two different images of the same scene will be developed. The difference in the images is assumed to be due to additive noise. The method of solution employed is an adaptation of that used for the determination of the error in the measured delay time in a radar system. Several analyses of the radar problem have been carried out based upon different premises.[2,3,4] These approaches may be categorized as those which use the probability density function of the noise directly and those which do not. The first case utilizes maximum a posteriori, maximum likelihood, or minimum mean square error estimates. All three estimators are based upon knowledge of the noise probability density function. The second case is based only upon the output of a filter which gives a maximum output at the correct time delay when the input is noise free.

An analysis of this sort should prove useful in several respects. The results should give an indication of the best possible registration of two images given the models of the data and noise. Once the models of the parameters involved have been found or assumed, an optimum receiver to implement the overlaying procedure may be developed. Comparison of existing registration systems with the results obtained herein may also be performed. However,

---

The research described in this report was sponsored by the National Aeronautics and Space Administration (NASA) under Grant Number NGL-15-005-112.

one must keep in mind the assumptions the entire analysis will be based on, for different assumptions may yield different results.

It is assumed in the following investigation that the useful signal is present, reducing the problem to one of estimation only rather than detection as well as estimation. It is further assumed that the signal shape is known and nonrandom, although the parameter that is to be measured is a random variable. Since the original signal is known, it does not have a probability density function. However, the received signal does depend upon the noise and its distribution. The problem will be approached with this in mind.

The solution to the problem depends upon a cost function which is assigned to the error and the a posteriori distribution of the signal as a function of a parameter, $m(\tau)$, given the received signal, $p_f(m(\tau))$. A minimum mean square error estimate is the mean of $p_f(m(\tau))$; an absolute value cost function gives the median of the probability function; the maximum a posteriori estimate yields the maximum of $p_f(m(\tau))$.[4] The maximum likelihood estimate may be viewed as the same as the maximum a posteriori estimate when there is no prior knowledge of the density function of the parameter, $p(m(\tau))$, or $p(m(\tau))$ is assumed uniform over a given range. All four of the above cost functions will yield the same solution for $p(m(\tau))$ uniform, and a symmetric, unimodel density function, $p_f(m(\tau))$.[4] A Gaussian distribution, which $p_f(m(\tau))$ has been assumed in several analyses, is a member of this latter class. The reason for the use of the Gaussian distribution is the availability of a closed form analytical solution.

Method 1

The following derivation of the variance of the registration error is a direct adaptation of Zubakov's solution,[5] which assumes Gaussian statistics. The principal modification is adding a second dimension to the signal, for a two dimensional image. For this discussion the time domain and spatial domain will be synonymous.

The likelihood function, $\Lambda(\tau)$, is,

$$(1) \quad \Lambda(\tau) = p_m(\tau) \; \frac{p_{m(\tau)}(f)}{p_o(f)}$$

where,

$p_m(\tau)$ = conditional density of τ given $m(t,\tau)$

$p_{m(\tau)}(f)$ = conditional density of $f(t)$ given $m(t,\tau)$

$p_o(f)$ = conditional density of $f(t)$ given that $m(t,\tau)$ is absent

$m(t,\tau)$ = known signal as a function of time and the unknown parameter, τ

$f(t)$ = $m(t,\tau) + n(t)$ = received signal

$n(t)$ = additive noise

The case of interest here is one of signal delay in two spatial directions. The likelihood function may be extended to a dependence upon two parameters,

$$(2) \quad \Lambda(\tau_x,\tau_y) = p_m(\tau_x,\tau_y) \; \frac{p_{m(\tau_x,\tau_y)}(f)}{p_o(f)}$$

where these quantities are defined similarly to those above. However, $m(t,\tau)$ will become $m(x,y,\tau_x,\tau_y)$, a two dimensional signal with two parameters, the cartesian coordinate system being used. In working with discrete data the spatial dimensions can be indexed by a single subscript, thus allowing a further direct extension of Zubakov's derivation. A two dimensional array m_{ij}, $i = 1,\ldots, p_j$; $j = 1,\ldots, q$, is converted to a one dimensional data set $m_h, h = 1,\ldots,$ pq. This conversion is made for convenience in notation and loses nothing from the standpoint of the results to be derived. Similarly, the received signal and noise also will be indexed with only one subscript.

In the discrete case a continuous function has been sampled and may be denoted,

$$m_h = m(t_h)$$

$$n_h = n(t_h)$$

$$f_h = f(t_h) = m_h + n_h$$

$$h = 1,\ldots, H$$

$$H = pq = \text{total number of samples}$$

To arrive at an analytical result, the probability density function of the noise must be known. So for this purpose assume a Gaussian distribution function with zero mean,

$$(3) \qquad p_{\underline{n}}(\underline{n}) = \frac{1}{(2\pi)^{H/2}|\underline{R}|^{1/2}} \exp\left[-\frac{1}{2}\underline{n}^T \underline{R}^{-1} \underline{n}\right]$$

where $\underline{R}$ is the covariance matrix of the noise, $R_{gh} = E[n_g n_h]$.

The density functions in the likelihood equation then become,

$$p_{m(\tau_x,\tau_y)}(f) = p_{\underline{n}}(\underline{f} - \underline{m}(\tau_x,\tau_y))$$

$$p(f) = p_{\underline{n}}(\underline{f})$$

$$\underline{f}^T = (f_1,\ldots,f_H)$$

$$\underline{m}^T = (m_1,\ldots,m_H)$$

The likelihood function can be reduced to

$$(4) \quad \Lambda(\tau_x,\tau_y) = p_m(\tau_x,\tau_y) \exp\{\sum_g^H \sum_h^H Q_{gh} f_g m_h(\tau_x,\tau_y) - \frac{1}{2}\sum_g^H \sum_h^H Q_{gh} m_g(\tau_x,\tau_y) m_h(\tau_x,\tau_y)\};\ Q_{gh} = gh^{th} \text{ element of } \underline{R}^{-1}$$

Since it is only the maximum of $\Lambda(\tau_x,\tau_y)$ which is desired, the problem can be reduced even further. Let $p_m(\tau_x,\tau_y)$ be a uniform distribution over a given area. This is a reasonable assumption since there is no a priori knowledge about the actual distribution. The question in point here is concerned only with a spatial delay, so that the summation term,

$$(5) \quad \mu = \sum_g^H \sum_h^H Q_{gh} m_g(\tau_x,\tau_y) m_h(\tau_x,\tau_y)$$

will also be a constant function of τ_x and τ_y. The only factor which is not a constant with respect to τ_x and τ_y is,

$$(6) \quad \phi = \sum_g^H \sum_h^H Q_{gh} f_g m_h(\tau_x,\tau_y)$$

Therefore the maximum of $\Lambda(\tau_x,\tau_y)$ is determined solely by the maximum of ϕ. The optimum receiver is then the one which finds the maximum of ϕ. This type of receiver may be viewed as a correlator which is weighted according to the inverse noise covariance function, Q_{gh}. For the case in which the noise is white with spectrum $N_o/2$, the covariation matrix becomes $\frac{2}{N_o}\underline{I}$ ($\underline{I}$ = identity matrix), and the optimum receiver is simply a correlator.

$$(7) \qquad \phi = \frac{2}{N_o} \sum_h^H f_h m_h(\tau_x,\tau_y)$$

Given that the maximum point (denoted by $\hat{\tau}_x,\hat{\tau}_y$) of the likelihood function has been found, a measure of the accuracy of the estimates is necessary so that the results of the estimator may be evaluated. One such measure is the variance of the estimate about the maximum point of $\Lambda(\tau_x,\tau_y)$. For this analysis use the $\ln[\Lambda(\tau_x,\tau_y)]$ since it is a monotonic function of $\Lambda(\tau_x,\tau_y)$. In the once dimensional case the derivation continues by taking the Taylor series expansion of $\ln[\Lambda(\tau)]$ around $\hat{\tau}$. The two dimensional case necessitates a further evaluation of exactly what is wanted. A two dimensional Taylor series of a function $f(x,y)$ about $(\hat{x},\hat{y})$ which is truncated at the second order is,

$$(8) \qquad f(x,y) \simeq f(\hat{x},\hat{y}) + \frac{\partial f(\hat{x},\hat{y})}{\partial x}(x-\hat{x}) + \frac{\partial f(\hat{x},\hat{y})}{\partial y}(y-\hat{y})$$
$$+ \frac{\partial^2 f(\hat{x},\hat{y})}{\partial x \partial y}(x-\hat{x})(y-\hat{y}) + \frac{1}{2}\frac{\partial^2 f(\hat{x},\hat{y})}{\partial x^2}(x-\hat{x})^2$$
$$+ \frac{1}{2}\frac{\partial^2 f(\hat{x},\hat{y})}{\partial y^2}(y-\hat{y})^2 + \frac{1}{2}\frac{\partial^3 f(\hat{x},\hat{y})}{\partial^2 x \partial y}(x-\hat{x})^2(y-\hat{y})$$
$$+ \frac{1}{2}\frac{\partial^3 f(\hat{x},\hat{y})}{\partial x \partial y^2}(x-\hat{x})(y-\hat{y})^2 + \frac{1}{4}\frac{\partial^4 f(\hat{x},\hat{y})}{\partial x^2 \partial y^2}(x-\hat{x})^2(y-\hat{y})^2$$

Utilization of this formulation will present an intractable problem. An interpretation of the problem which will simplify the solution is needed. Once such viewpoint is to consider the variance only along the x-axis direction and y-axis directions, instead of a total variance in all directions. Expand $\ln[\Lambda(\tau_x,\tau_y)]$ in a Taylor series about $(\hat{\tau}_x,\hat{\tau}_y)$ first in the x-axis direction and then in the y-axis direction. Assume that the $\ln[\Lambda(\tau_x,\tau_y)]$ can be approximated by a second order polynomial in both directions.

x-axis direction,

$$(9) \quad \ln\Lambda(\tau_x,\hat{\tau}_y) \approx \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y) + \frac{\partial \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y)}{\partial \tau_x}(\tau_x-\hat{\tau}_x) + \frac{1}{2}\frac{\partial^2 \ln\Lambda(\tau_x,\hat{\tau}_y)}{\partial \tau_x^2}(\tau_x-\hat{\tau}_y)^2$$

y-axis direction,

$$(10) \quad \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y) \approx \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y) + \frac{\partial \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y)}{\partial \tau_y}(\tau_y-\hat{\tau}_y) + \frac{1}{2}\frac{\partial^2 \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y)}{\partial \tau_y^2}(\tau_y-\hat{\tau}_y)^2$$

A necessary condition for the maximum point of $\ln\Lambda(\tau_x,\tau_y)$ is that,

$$(11) \quad \frac{\partial \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y)}{\partial \tau_x} = 0 = \frac{\partial \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y)}{\partial \tau_y}$$

The Taylor series expansion may then be reduced to,

x-direction,

$$(12) \quad \ln\Lambda(\tau_x,\hat{\tau}_y) \approx \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y) + \frac{1}{2}\frac{\partial^2 \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y)}{\partial \tau_x^2}(\tau_x-\hat{\tau}_x)^2$$

y-direction,

$$(13) \quad \ln\Lambda(\hat{\tau}_x,\tau_y) \approx \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y) + \frac{1}{2}\frac{\partial^2 \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y)}{\partial \tau_y^2}(\tau_y-\hat{\tau}_y)^2$$

Rearranging the equations,

x-direction,

$$(14)\qquad \Lambda(\tau_x,\hat{\tau}_y) = \Lambda(\hat{\tau}_x,\hat{\tau}_y)\exp\left\{-\frac{1}{2}\frac{(\tau_x-\hat{\tau}_x)^2}{\Delta_x^{\,2}}\right\}$$

y-direction,

$$(15)\qquad \Lambda(\hat{\tau}_x,\tau_y) = \Lambda(\hat{\tau}_x,\hat{\tau}_y)\exp\left\{-\frac{1}{2}\frac{(\tau_y-\hat{\tau}_y)^2}{\Delta_y^{\,2}}\right\}$$

where,

$$(16)\qquad \Delta_x^{\,2} = -\left[\frac{\partial^2 \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_x^{\,2}}\right]^{-1} = \text{variance in the x-direction}$$

$$(17)\qquad \Delta_y^{\,2} = -\left[\frac{\partial^2 \ln\Lambda(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_y^{\,2}}\right]^{-1} = \text{variance in the y-direction}$$

Assuming $p_m(\tau_x,\tau_y)$ to be uniformly distributed,

$$(18)\qquad \frac{1}{\Delta_x^{\,2}} = \sum_g^H\sum_h^H Q_{gh}\left[m_g(\hat{\tau}_x,\hat{\tau}_y) - f_g\right]\frac{\partial^2 m_h(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_x^{\,2}} + \sum_g^H\sum_h^H Q_{gh}\frac{\partial m_g(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_x}\frac{\partial m_h(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_x}$$

$$(19)\qquad \frac{1}{\Delta_y^{\,2}} = \sum_g^H\sum_h^H Q_{gh}\left[m_g(\hat{\tau}_x,\hat{\tau}_y) - f_g\right]\frac{\partial^2 m_h(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_y^{\,2}} + \sum_g^H\sum_h^H Q_{gh}\frac{\partial m_g(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_y}\frac{\partial m_h(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_y}$$

If one further assumes a large signal-to-noise ratio, then

$$(20)\qquad \frac{1}{\Delta_x^{\,2}} = \sum_g^H\sum_h^H Q_{gh}\frac{\partial m_g(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_x}\frac{\partial m_h(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_x}$$

$$(21)\qquad \frac{1}{\Delta_y^{\,2}} = \sum_g^H\sum_h^H Q_{gh}\frac{\partial m_g(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_y}\frac{\partial m_h(\hat{\tau}_x,\hat{\tau}_y)}{\partial\tau_y}$$

since $[m_g(\hat{\tau}_x,\hat{\tau}_y) - f_g]$ is dependent only upon the noise and is small compared to $m_g(\hat{\tau}_x,\hat{\tau}_y)$. An alternative way of expressing the variance, which will be derived, is,

$$(22) \quad \Delta_x^{\,2} = \frac{1}{\Delta W_x^{\,2}\mu}$$

$$(23) \quad \Delta_y^{\,2} = \frac{1}{\Delta W_y^{\,2}\mu}$$

where,

$$(24) \quad \mu = \sum_g^H \sum_h^H Q_{gh} m_g(\hat{\tau}_x,\hat{\tau}_y)\, m_h(\hat{\tau}_x,\hat{\tau}_y), \text{ signal-to-noise ratio}$$

or equivalently,

$$(25) \quad \mu = \sum_u^p \sum_v^q \frac{|M(u,v)|^2}{S_R(u,v)}$$

$$(26) \quad \Delta W_x^{\,2} = \frac{4\pi \sum_u^p \sum_v^q \frac{u^2|M(u,v)|}{S_R(u,v)}}{\sum_u^p \sum_v^q \frac{|M(u,v)|^2}{S_R(u,v)}} = \text{effective bandwidth in the x-axis direction}$$

$$(27) \quad \Delta W_y^{\,2} = \frac{4\pi^2 \sum_u^p \sum_v^q \frac{v^2|M(u,v)|^2}{S_R(u,v)}}{\sum_u^p \sum_v^q \frac{|M(u,v)|^2}{S_R(u,v)}} = \text{effective bandwidth in the y-axis direction}$$

$M(u,v)$ = Fourier transform of the known signal

$S_R(u,v)$ = noise spectrum

The variance can be reduced to a function of the effective bandwidth and signal-to-noise ratio. A result such as this may yield easily obtainable first approximations to the error that can be expected.

Conversion of $\frac{1}{\Delta^2}$ to the frequency domain.

For the purpose of this proof subscript notation will be changed. The signals will be represented as a two dimensional array instead of the one dimension as used above.

$$\underline{M} = \begin{bmatrix} m_{11} & m_{1q} \\ & \\ m_{p1} & m_{pq} \end{bmatrix}$$

Where pq = H, the total number of samples. The elements of the inverse covariance matrix, $\underline{Q}$, will be denoted by Q_{cdab}. Elements of the covariance matrix, $\underline{R}$, will be R_{cdab}.

$$(29) \quad R_{cdab} = \overline{n_{ab} n_{cd}} \quad ; \quad \begin{array}{l} a, c = 1,\ldots,p \\ b, d = 1,\ldots,q \end{array}$$

Define an array $\underline{K}$ such that,

$$(30) \quad K_{cd} = \sum_{a}^{p} \sum_{b}^{q} Q_{cdab} \, m_{ab}$$

This may be equivalently expressed as,

$$(31) \quad M_{ab} = \sum_{c}^{p} \sum_{d}^{q} R_{cdab} \, K_{cd}$$

or, assuming stationarity in the x-axis and y-axis directions,

$$(32) \quad M_{ab} = \sum_{c}^{p} \sum_{d}^{q} R_{a-c,\, b-d} \, K_{cd}$$

With this notation,

$$(33) \quad \mu = \sum_{c}^{p} \sum_{d}^{q} \sum_{a}^{p} \sum_{b}^{q} Q_{cdab} \, M_{cd} \, M_{ab}, \text{ from equation (24)}$$

$$(34) \quad \mu = \sum_{c}^{p} \sum_{d}^{q} K_{cd}, \, M_{cd}, \text{ substitution from equation (30)}$$

$$(35) \quad \mu = \sum_{c}^{p} \sum_{d}^{q} \sum_{a}^{p} \sum_{b}^{q} K_{cd} \, R_{a-c,\, b-d} \, K_{ab}, \text{ from equation (31)}$$

$$\mu = \sum_{u}^{p} \sum_{v}^{q} \sum_{c}^{p} \sum_{d}^{q} \sum_{a}^{p} \sum_{b}^{q} S_R(u,v)\ K_{cd}\ K_{ab}\ \exp\ [j2\pi(a-c)u+(b-d)v]$$

(36) $$\mu = \sum_{u}^{p} \sum_{v}^{q} S_R(u,v)\,|K(u,v)|^2$$

Find K(u,v) in terms of M(u,v).

(37) $$M(u,v) = \sum_{c}^{p} \sum_{d}^{q} \sum_{a}^{p} \sum_{b}^{q} R_{a-c,\ b-d}\ K_{cd}\ \exp\ [-j2\pi(au+bv)]$$

(38) $$M(u,v) = \sum_{\lambda}^{p} \sum_{\gamma}^{q} \sum_{a}^{p} \sum_{b}^{a} R_{\lambda\gamma}\ K_{cd}\ \exp\ \{-j2\pi[(\lambda+c)u + (\gamma+d)v]\}$$

The conversion from equation (37) to (38) is true because the discrete Fourier transform assumes periodicity of the time function. Therefore,

$$M(u,v) = S_R(u,v)\ K(u,v)\ ;\ \text{from equation (38)}$$

or,

(39) $$K(u,v) = \frac{M(u,v)}{S_R(u,v)}$$

Substituting equation (39) into (36),

(40) $$\mu = \sum_{u}^{p} \sum_{v}^{q} \frac{|M(u,v)|^2}{S_R(u,v)}$$

Similarly, it can be shown that,

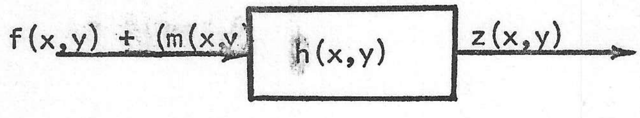
(41) $$\sum_{c}^{p} \sum_{d}^{q} \sum_{a}^{p} \sum_{b}^{q} Q_{cdab}\ \frac{\partial m_{cd}}{\partial \tau_x}\ \frac{\partial m_{ab}}{\partial \tau_x} = 4\pi^2 \sum_{u}^{p} \sum_{v}^{q} \frac{u^2|M(u,v)|^2}{S_R(u,v)}$$

(42) $$\sum_{c}^{p} \sum_{d}^{q} \sum_{a}^{p} \sum_{b}^{q} Q_{cdab}\ \frac{\partial m_{cd}}{\partial \tau_y}\ \frac{\partial m_{ab}}{\partial \tau_y} = 4\pi^2 \sum_{u}^{p} \sum_{v}^{q} \frac{v^2|M(u,v)|^2}{S_R(u,v)}$$

Method 2

A second method for estimating the variance of the registration error will now be developed. For this case, the only assumption about the processor is that in the absence of noise, the output of the filter will be a maximum at the correct time delay.[3] The only a priori knowledge of the noise that is necessary here is its correlation function. This is shown in the derivation. No assumptions about the probability distribution of the noise are needed. This derivation is carried out for the continuous signal case, but it may be extended directly to the discrete situation in which a continuous function has been sampled.

The received signal is made up of two components, f(x,y), the useful signal, and m(x,y) the additive noise. This signal is passed through a filter with impulse response h(x,y), yielding an output, g(x,y) + n(x,y) [g(x,y) = h(x,y) * f(x,y); n(x,y) = h(x,y) * m(x,y)]. The position at which the maximum of the output signal occurs is taken to be the correct registration position. However, the filter is designed to give the maximum of g(x,y) only at the correct delay. The difference between these two positions is the error. A measure of the variance of this error is derived below.



(43) $z(x,y) = g(x,y) + n(x,y)$

(44) $g(x,y) = \iint h(x-\alpha,y-\beta)\, f(\alpha,\beta)\, d\alpha d\beta$

(45) $n(x,y) = \iint h(x-\alpha,y-\beta)\, m(\alpha,\beta)\, d\alpha d\beta$

Let the maximum point of $g(x,y)$ occur at $(\tilde{x},\tilde{y})$. Expand $g(x,y)$ about $(\tilde{x},\tilde{y})$ using a two dimensional Taylor series.

(46) $$g(x,y) \approx g(\tilde{x},\tilde{y}) + g_x(\tilde{x},\tilde{y})(x-\tilde{x}) + g_y(\tilde{x},\tilde{y})(y-\tilde{y}) + g_{xy}(\tilde{x},\tilde{y})(x-\tilde{x})(y-\tilde{y}) + \frac{1}{2} g_{xx}(\tilde{x},\tilde{y})(x-\tilde{x})^2 + \frac{1}{2} g_{yy}(\tilde{x},\tilde{y})(y-\tilde{y})^2 + \ldots$$

where,

$$g_x(\tilde{x},\tilde{y}) = \left.\frac{\partial g(x,y)}{\partial x}\right|_{x=\tilde{x},\ y=\tilde{y}}$$; this subscript notation is used for the remainder of this section

Assume that $(x-\tilde{x})$ and $(y-\tilde{y})$ are small enough so that all higher order terms may be neglected.

A necessary condition for a maximum at $(\tilde{x},\tilde{y})$ is that $g_x(\tilde{x},\tilde{y}) = 0 = g_y(\tilde{x},\tilde{y})$. The Taylor series expansion may then be simplified.

(47) $$g(x,y) \approx g(\tilde{x},\tilde{y}) + g_{xy}(\tilde{x},\tilde{y})(x-\tilde{x})(y-\tilde{y}) + \frac{1}{2} g_{xx}(\tilde{x},\tilde{y})(x-\tilde{x})^2 + \frac{1}{2} g_{yy}(\tilde{x},\tilde{y})(y-\tilde{y})^2$$

Substitute equation (47) back into equation (43).

(48) $$z(x,y) = g(\tilde{x},\tilde{y}) + g_{xy}(\tilde{x},\tilde{y})(x-\tilde{x})(y-\tilde{y}) + \frac{1}{2} g_{xx}(\tilde{x},\tilde{y})(x-\tilde{x})^2 + \frac{1}{2} g_{yy}(\tilde{x},\tilde{y})(y-\tilde{y})^2 + n(x,y)$$

The maximum output of the filter may occur at a different position than $(\tilde{x},\tilde{y})$. Denote this new position as $(\bar{x},\bar{y})$. A necessary condition at $(x,y) = (\bar{x},\bar{y})$ is that $z_x(\bar{x},\bar{y}) = 0 = z_y(\bar{x},\bar{y})$. Use these conditions on equation (48).

$$(49)\quad z_x(\bar{x},\bar{y}) = 0 = g_{xy}(\tilde{x},\tilde{y})(\bar{y}-\tilde{y}) + g_{xx}(\tilde{x},\tilde{y})(\bar{x}-\tilde{x}) + n_x(\bar{x},\bar{y}).$$

$$(50)\quad z_y(\bar{x},\bar{y}) = 0 = g_{xy}(\tilde{x},\tilde{y})(\bar{x}-\tilde{x}) + g_{yy}(\tilde{x}-\tilde{y})(\bar{y}-\tilde{y}) + n_y(\bar{x},\bar{y})$$

Equations (49) and (50) may be set up in matrix form,

$$(51)\quad \begin{bmatrix} g_{xx} & g_{xy} \\ g_{xy} & g_{yy} \end{bmatrix} \begin{bmatrix} \bar{x} - \tilde{x} \\ \bar{y} - \tilde{y} \end{bmatrix} = \begin{bmatrix} -n_x(\bar{x},\bar{y}) \\ -n_y(\bar{x},\bar{y}) \end{bmatrix}$$

where the arguments, $(\tilde{x},\tilde{y})$, have been left out for ease in notation.

Solve for $(\bar{x}-\tilde{x})$ and $(\bar{y}-\tilde{y})$.

$$(52)\quad \begin{bmatrix} \bar{x} - \tilde{x} \\ \bar{y} - \tilde{y} \end{bmatrix} = \frac{1}{g_{xx}g_{yy} - g_{xy}^2} \begin{bmatrix} g_{yy} & -g_{xy} \\ -g_{xy} & g_{xx} \end{bmatrix} \begin{bmatrix} -n_x(\bar{x},\bar{y}) \\ -n_y(\bar{x},\bar{y}) \end{bmatrix}$$

Therefore,

$$(53)\quad (\bar{x}-\tilde{x}) = \frac{g_{xy}n_y - g_{yy}n_x}{g_{xx}g_{yy} - g_{xy}^2}$$

$$(54)\quad (\bar{y}-\tilde{y}) = \frac{g_{xy}n_x - g_{xx}n_y}{g_{xx}g_{yy} - g_{xy}^2}$$

where the arguments, $(\bar{x},\bar{y})$, have also been left out for notational convenience.

Next find the variance of $(\bar{x}-\tilde{x})$ and $(\bar{y}-\tilde{y})$, assuming the mean $= 0$, i.e., $E[\bar{x}-\tilde{x}] = E[\bar{y}-\tilde{y}] = 0$

Then,

$$(55) \quad Var[\bar{x}-\tilde{x}] = E[(\bar{x}-\tilde{x})^2] = \overline{(\bar{x}-\tilde{x})^2}$$

$$(56) \quad Var[\bar{y}-\tilde{y}] = E[(\bar{y}-\tilde{y})^2] = \overline{(\bar{y}-\tilde{y})^2}$$

$$(57) \quad \overline{(\bar{x}-\tilde{x})^2} = \frac{g_{xy}^2\overline{n_y^2} - 2g_{xy}g_{yy}\overline{n_y n_x} + g_{yy}^2\overline{n_x^2}}{[g_{xx}g_{yy} - g_{xy}^2]^2}$$

$$(58) \quad \overline{(\bar{y}-\tilde{y})^2} = \frac{g_{xy}^2\overline{n_x^2} - 2g_{xy}g_{xx}\overline{n_y n_x} + g_{xx}^2\overline{n_y^2}}{[g_{xx}g_{yy} - g_{xy}^2]^2}$$

where,

$$(59) \quad \overline{n_y^2(\bar{x},\bar{y})} = \iiiint h_y(\bar{x}-\alpha,\bar{y}-\beta)h_y(\bar{x}-\gamma,\bar{y}-\lambda)\overline{m(\alpha,\beta)m(\gamma,\lambda)}\,d\alpha\, d\beta\, d\gamma\, d\lambda$$

Letting,

$$(60) \quad R_m(\alpha-\gamma,\beta-\lambda) = \overline{m(\alpha,\beta)m(\gamma,\lambda)}$$

by assuming stationarity,

$$(61) \quad \overline{n_y^2(\bar{x},\bar{y})} = \iiiint h_y(\bar{x}-\alpha,\bar{y}-\beta)h_y(\bar{x}-\gamma,\bar{y}-\lambda)R_m(\alpha-\gamma,\beta-\lambda)\,d\alpha\, d\beta\, d\gamma\, d\lambda$$

Similarly,

$$(62) \quad \overline{n_y(\bar{x},\bar{y})m_x(\bar{x},\bar{y})} = \iiiint h_y(\bar{x}-\alpha,\bar{y}-\beta)h_x(\bar{x}-\gamma,\bar{y}-\lambda)R_m(\alpha-\gamma,\beta-\lambda)\,d\alpha\, d\beta\, d\gamma\, d\lambda$$

$$(63) \quad \overline{n_x^2(\bar{x},\bar{y})} = \iiiint h_x(\bar{x}-\alpha,\bar{y}-\beta)h_x(\bar{x}-\gamma,\bar{y}-\lambda)R_m(\alpha-\gamma,\beta-\lambda)\,d\alpha\, d\beta\, d\gamma\, d\lambda$$

$$(64) \quad g_{xx}(\tilde{x},\tilde{y}) = \iint h_{xx}(\tilde{x}-\alpha,\tilde{y}-\beta)\; f(\alpha,\beta)\,d\alpha\, d\beta$$

$$(65) \quad g_{yy}(\tilde{x},\tilde{y}) = \iint h_{yy}(\tilde{x}-\alpha,\tilde{y}-\beta)\; f(\alpha,\beta)\,d\alpha\, d\beta$$

$$(66) \quad g_{xy}(\tilde{x},\tilde{y}) = \iint h_{xy}(\tilde{x}-\alpha,\tilde{y}-\beta)\; f(\alpha,\beta)\,d\alpha\, d\beta$$

Convert equations (61) through (66) into their equivalent form in the frequency domain.

(67) $$\overline{n_y^{\,2}(\bar{x}-\bar{y})} = 4\pi^2 \iint v^2|H(u,v)|^2 S_m(u,v)\,du\,dv$$

(68) $$\overline{n_x(\bar{x},\bar{y})n_y(\bar{x},\bar{y})} = 4\pi^2 \iint uv|H(u,v)|^2 S_m(u,v)\,du\,dv$$

(69) $$\overline{n_x^{\,2}(\bar{x},\bar{y})} = 4\pi^2 \iint u^2|H(u,v)|^2 S_m(u,v)\,du\,dv$$

(70) $$g_{xx}(\tilde{x},\tilde{y}) = -4\pi^2 \iint u^2 H(u,v)\, F(u,v)\, e^{j2\pi(\tilde{x}\,u + \tilde{y}\,v)}\,du\,dv$$

(71) $$g_{yy}(\tilde{x},\tilde{y}) = -4\pi^2 \iint v^2 H(u,v)\, F(u,v)\, e^{j2\pi(\tilde{x}\,u + \tilde{y}\,v)}\,du\,dv$$

(72) $$g_{xy}(\tilde{x},\tilde{y}) = -4\pi^2 \iint uv\, H(u,v)\, F(u,v)\, e^{j2\pi(\tilde{x}\,u + \tilde{y}\,v)}\,du\,dv$$

The filter impulse response, $h(x,y)$, is assumed to be a real function. In order to arrive at a closed form solution as a function of the effective bandwidth and signal-to-noise ratio, a specific form for the filter will be chosen; viz., the so called "matched filter" which maximizes the output signal-to-noise ratio.

Let,,

(73) $$H(u,v) = \frac{F^*(u,v)\, e^{-j2\pi(\tilde{x}\,u + \tilde{y}\,v)}}{S_m(u,v)}$$

where,

$S_m(u,v) = \{R_m(x,y)\}$

$F(v,u) = \{f(x,y)\}$

$H(u,v) = \{h(x,y)\}$

Equations (67) through (72) then become,

(74) $$\overline{n_y^2(\bar{x},\bar{y})} = 4\pi^2 \iint v^2 \frac{|F(u,v)|^2}{S_m(u,v)}\, du\, dv$$

(75) $$\overline{n_x(\bar{x},\bar{y})n_y(\bar{x},\bar{y})} = 4\pi^2 \iint uv \frac{|F(u,v)|^2}{S_m(u,v)}\, du\, dv$$

(76) $$\overline{n_x^2(\bar{x},\bar{y})} = 4\pi^2 \iint u^2 \frac{|F(u,v)|^2}{S_m(u,v)}\, du\, dv$$

(77) $$g_{xx}(\tilde{x},\tilde{y}) = -\overline{n_x^2(\bar{x},\bar{y})}$$

(78) $$g_{yy}(\tilde{x},\tilde{y}) = -\overline{n_y^2(\bar{x},\bar{y})}$$

(79) $$g_{xy}(\tilde{x},\tilde{y}) = -\overline{n_x(\bar{x},\bar{y})n_y(\bar{x},\bar{y})}$$

Equations (77) through (79) give an important relationship. The equations for $\overline{(\bar{x}-\tilde{x})^2}$ and $\overline{(\bar{y}-\tilde{y})^2}$ may now be greatly simplified. The arguments $(\tilde{x},\tilde{y})$ and $(\bar{x},\bar{y})$ will be dropped for notational convenience in the following derivation.

$$\overline{(\bar{x}-\tilde{x})^2} = \frac{g_{xy}^2\,\overline{n_y^2} - 2\, g_{xy}g_{yy}\overline{n_y n_x} + g_{yy}^2\,\overline{n_x^2}}{[g_{xx}g_{yy} - g_{xy}^2]^2}, \text{ from equation (57)}$$

Substituting in equations (74), (75) and (76) yields,

(80) $$\overline{(\bar{x}-\tilde{x})^2} = \left[\frac{g_{xy}^2}{g_{yy}} - g_{xx}\right]^{-1}$$

Similarly,

(81) $$\overline{(\bar{y}-\tilde{y})^2} = \left[\frac{g_{xy}^2}{g_{xx}} - g_{yy}\right]^{-1}$$

These are the two basic equations for computing the variance of the estimate in the x-axis and y-axis directions. At first glance, equations (80) and (81) do not seem to resemble the analogous discrete signal bandwidth equations as derived in the previous section, equations (22) and (23). However, one must examine the terms more closely. The variables, $-g_{xx}$ and $-g_{yy}$, as in equations (74) and (78), contain the signal bandwidths.

$$-g_{xx} = 4\pi^2 \iint u^2 \frac{|F(u,v)|^2}{S_m(u,v)} \, du \, dv$$

This may be written as:

$$(82) \quad -g_{xx} = B_x^2 \, SNR$$

where,

$$B_x = \left[\frac{4\pi^2 \iint u^2 \frac{|F(u,v)|^2}{S_m(u,v)} \, du \, dv}{\iint \frac{|F(u,v)|^2}{S_m(u,v)} \, du \, dv} \right]^{1/2}$$ = effective bandwidth of input signal in the x-axis direction

and,

$$SNR = \iint \frac{|F(u,v)|^2}{S_m(u,v)} \, du \, dv$$ = output signal-to-noise ratio

Similarly,

$$-g_{yy} = 4\pi^2 \iint v^2 \frac{|F(u,v)|^2}{S_m(u,v)} \, du \, dv$$

$$(83) \quad -g_{yy} = B_y^2 \, SNR$$

where,

$$B_y = \left[\frac{4\pi^2 \iint v^2 \frac{|F(u,v)|^2}{S_m(u,v)} \, du \, dv}{\iint \frac{|F(u,v)|^2}{S_m(u,v)} \, du \, dv} \right]^{1/2}$$ = effective bandwidth of input signal in the y-axis direction

Substituting equations (82) and (83) into (80) and (81), one obtains,

$$(84)\qquad \overline{(\bar{x}-\tilde{x})^2} = \left(-\frac{g_{xy}^2}{B_y^2 SNR} + B_x^2 SNR \right)^{-1}$$

$$(85)\qquad \overline{(\bar{y}-\tilde{y})^2} = \left(-\frac{g_{xy}^2}{B_x^2 SNR} + B_y^2 SNR \right)^{-1}$$

The variances are again seen to be functions of the signal bandwidths, however, the relationships are not as simple as in the case considered earlier.

Now consider the the specific case where the ratio in the integrand is a constant and the signal spectrum is bandlimited.

$$(86)\qquad \frac{|F(u,v)|^2}{S_m(u,v)} = c, \text{ a constant}$$

This situation would be encountered if the noise spectrum has a similar shape to the signal spectrum within the bandlimits. In such a case it might be advantageous to model the two spectra as differing only by a constant factor for simplicity in estimating the variance to be expected. First look at g_{xy}.

$$g_{xy} = -4\pi^2 \int_{-a}^{a} \int_{-b}^{b} uv \frac{|F(u,v)|^2}{S_m(u,v)} \, du \, dv$$

$$= -4\pi^2 c \int_{-a}^{a} \int_{-b}^{b} uv \, du \, dv$$

$$(87)\qquad g_{xy} = 0$$

In this case, equations (84) and (85) reduce to,

$$(88)\qquad \overline{(\bar{x}-\tilde{x})^2} = \frac{1}{B_x^2 SNR}$$

$$(89)\qquad \overline{(\bar{y}-\tilde{y})^2} = \frac{1}{B_y^2 SNR}$$

which are analogous to the results in the first section.

The variance of the error again has been found to be a function of the effective bandwidth and signal-to-noise ratio. This time, however, there is no consideration of the probability density function of the noise. The correlation function of the noise is the only function that need be known.

Both the methods yield similar results, and therefore indicate that a measure of the error to be expected is available. A measure of this sort should be useful in evaluating the registration of one image upon another. The filters chosen in both cases are similar. The second case also provides for a measure of the error variance when the filter chosen in the derivation is not used: refer to equations (57) and (58). However, in this case, a simple analogy to the effective bandwidth and signal-to-noise ratio is not obtained. In the frequency domain, each of the chosen filters is the complex conjugate of the useful signal spectrum divided by the noise spectral density, times a phase shift factor, i.e. a matched filter. In the white noise case, the processors are correlators. This point should be of significance since one method of registration is to use a correlator, which is not an optimum processor in non-white noise. This is one aspect to be considered when designing the processor. A second aspect that must be dealt with, is how to obtain the noise covariance function, and if it is to be estimated, what is a good measure of the error in this estimate.

Sample Calculation of the Registration Error for Discrete Data

An expression for the registration error between two images was derived previously. The error was found to be a function of the signal-to-noise ratio and equivalent bandwidths in the x-axis and y-axis directions. For the analysis which follows, assume that the signal spectrum, $|F(u,v)|^2$, and noise spectrum, $S_n(u,v)$, are separable in u and v. This simplifies the expression for error to the one dimensional case,

$$(90) \qquad \frac{1}{E_x^2} = 4\pi^2 SNR_y \int_{-W}^{W} u^2 \frac{|F(u)|^2}{S_n(u)}\, du$$

In this example $\frac{|F(u,v)|^2}{S_m(u,v)}$ is chosen a constant. Therefore $g_{xy} = 0$, as was shown earlier.

E_x = error in the x-axis direction

SNR_y = signal-to-noise ratio in the y-axis direction

$$SNR_y = \int_{-W_y}^{W_y} \frac{|F(u)|^2}{S_n(u)}\, dv$$

W = bandwidth in the x-axis direction

W_y = bandwidth in the y-axis direction

This problem is explained as an adaptation of the continuous case to the discrete case, where equation (90) may be approximated by,

$$(91) \qquad \frac{1}{E_x^2} = 2\left\{4\pi^2 SNR_y \left[\frac{2W}{N} \sum_{k=0}^{N/2} \left(\frac{2kW}{N}\right)^2 \frac{|F(\frac{2kW}{N})|^2}{S_n(\frac{2kW}{N})}\right]\right\}$$

$$(92) \qquad SNR_y = 2\left[\frac{2W_y}{N}\sum_{k=0}^{N/2}\frac{\left|F\left(\frac{2kW_y}{N}\right)\right|^2}{S_n\left(\frac{2kW_y}{N}\right)}\right] - \frac{2W_y}{N}\,\frac{|F(0)|^2}{S_n(0)}$$

Refer to the figure below,

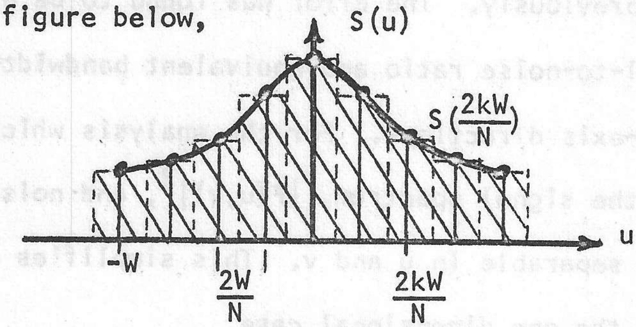


Figure 10

N = total number of samples

k = $0, 1, \ldots, \frac{N}{2}$

Δx = sampling interval

W = $\frac{1}{2\Delta x}$, cutoff (maximum) signal frequency

Δf = $\frac{2W}{N}$, frequency interval

Equation (91) yields a weighted sum of the shaded area in the figure and is a good approximation to the error expression for N large. In using equation (91) it is necessary that the frequency scales of the signal and noise spectrums be the same. This point is extremely important since equation (91) is meaningless otherwise.

Example

Let: $\frac{\left|F\left(\frac{2kW}{N}\right)\right|^2}{S_n\left(\frac{2kW}{N}\right)} = C$, a constant for all k.

SNR_x = signal-to-noise ratio in x-axis direction

SNR_y = signal-to-noise ratio in y-axis direction

Δx = 80 meters (approximately the sampling interval for ERTS data)

Find the expected error, E_x, in the x-axis direction.

Solution:

$$SNR_x = 2 \frac{2W}{N} \sum_{k=0}^{N/2} \frac{|F(\frac{2kW}{N})|^2}{S_n(\frac{2kW}{N})} - \frac{2W}{N} \frac{|F(0)|^2}{S_n(0)}$$

$$SNR_x = \frac{4W}{N} \sum_{k=0}^{N/2} C - \frac{2W}{N} C$$

$$SNR_x = 2WC \left[1 + \frac{2}{N} - \frac{1}{N}\right]$$

$$SNR_x = 2WC, \text{ for N large}$$

Therefore,

$$C = \frac{SNR_x}{2W}$$

From equation (91).

$$\frac{1}{E_x^2} = 2\ 4\pi^2 SNR_y \frac{2W}{N} \sum_{k=0}^{N/2} \left(\frac{2kW}{N}\right)^2 \frac{SNR_x}{2W}$$

$$\frac{1}{E_x^2} = \frac{4\pi^2}{N^3}\ 8\ W^2 (SNR_x)(SNR_y) \sum_{k=0}^{N/2} k^2$$

$$\frac{1}{E_x^2} = \frac{4\pi^2}{N^3}\ 8\ W^2 (SNR_x)(SNR_y) \frac{N(N+2)(N+1)}{24}$$

$$\frac{1}{E_x^2} \approx \frac{4\pi^2}{3} W^2 (SNR_x)(SNR_y), \text{ for large N}$$

$$E_x = \frac{1}{2\pi W} \sqrt{\frac{3}{(SNR_x)(SNR_y)}} \text{ unit length}$$

$$E_x = \Delta x \sqrt{\frac{3}{(SNR_x)(SNR_y)}} \text{ unit length}$$

Substituting in the value for Δx.

$$E_x = 41[(SNR_x)(SNR_y)]^{-1/2} \text{ meters}$$

For,

$$SNR_x = 1$$

$$SNR_y = 1$$

the error is,

$$E_x = 41 \text{ meters}$$

Error in the Estimation of Discrete Spectral Densities for the One-Dimensional Case

The variance of the registration error about the true point of overlay has been determined as a function of the Fourier transform of the signal and spectral density of the noise in section one. To proceed further to the analysis of a real problem it is necessary to obtain an estimate of the noise spectral density. Two approaches have been used for the estimation of this function in the one-dimensional case (this analysis is taken directly from Bendat[1]). The first is the calculation of the spectrum by finding the autocorrelation function and then taking its Fourier transform. The correlation function may be approximated by,

$$R_r = \frac{1}{n-r} \sum_{n=1}^{N-r} x_n x_{n+r} \, , \; r = 0, 1, \ldots, m$$

where,

x_n = $x(nh)$, the value of the function at the time, $t=nh$

h = time sampling interval

N = total number of data values

m = maximum lag in r

$E[X]$ = 0, by assumption

R_{-r} = R_r

The spectral density is then found by computing the Fourier transform of R_r from its samples as determined above. Using the discrete Fourier transform, the one-sided spectral density is given by,

$$\hat{G}_k = 2h\left[R_o + 2\sum_{r=1}^{m-1} R_r \cos\left(\frac{\pi r k}{m}\right) + (-1)^k R_m\right], \quad k=0, 1, \ldots, m-1$$

where,

$$\hat{G}_k = \hat{G}(f), \text{ the estimate of the spectrum } G(f) \text{ at } f = \frac{k}{2mh}$$

The two sided spectral density, S_k, is

$$S_k = S_{-k} = \frac{1}{2} G_k$$

The error for the spectral density estimation as given in Bendat[1] is,

$$E_r = \sqrt{\frac{m}{N}} = \sqrt{\frac{1}{B_e T}}$$

where,

$$B_e = \frac{1}{mh}, \text{ resolution bandwidth of spectrum}$$

$$T = Nh, \text{ total record length in time}$$

Where E_r is defined as,

$$E_r = \frac{\{Var[\hat{G}(f)]\}^{1/2}}{G(f)}$$

which is the ratio of the standard deviation of the spectral estimate at a frequency, f, to the true spectral density at that same frequency. It is seen that the error decreases as the sample size, N, grows large with respect to the maximum lag, m.

The second method for estimating the spectrum is to use the modulus squared of the Fourier transform of the data directly.

$$\hat{G}_k = \frac{2h}{N} |X_k|^2$$

$$X_k = \sum_{n=0}^{N-1} x_n \exp\left[-\frac{j2\pi kn}{N}\right]$$

This method yields an error,

$$E_r = 1, \text{ since } B_e = \frac{1}{T}$$

In this case, the variance of the estimate is equal to the actual spectral density at a given frequency. One way to reduce the error is to average the spectrum over a number of adjacent frequencies[2]. For "ℓ" equal to the number of points to be averaged,

$$E_r = \sqrt{\frac{1}{\ell}}$$

This averaging is in a sense a reduction of the bandwidth resolution.

This gives an indication of the quality of the spectral density estimate. For the purposes of determining the registration error, it may be desired to further approximate the spectrum by a general functional form. The justification for this step would lie in the ability to find a general functional model for the noise spectrum over different types of imagery.

Bibliography

1. Bendat, J.S., Piersol, A.G., Random Data: Analysis and Measurement Procedures, Wiley - Interscience, New York, 1971, pp. 125-133, 181-184.
2. DiFranco, J.V., Rubin, W.L., Radar Detection, Prentice-Hall, Inc., New Jersey, 1968.
3. McGillem, C.D., Thurman, L.A., Radar Resolution Enhancement and Target Surface Reconstruction, TR-EE 73-14, April 1973, Purdue University.
4. Van Trees, H.L., Detection, Estimation, and Modulation Theory, Part 1, John Wiley & Sons, Inc., New York, 1968.
5. Zubakov, V.D., Wainstein, L.A., Extraction of Signals From Noise, Prentice-Hall, Inc., New Jersey, 1962.