The Cramer-Rao Bound Applied to Image Fusion*

Rick S. Blum
ECE Department, Lehigh University
Bethlehem, PA 18015-3084
e-mail: rblum@eecs.lehigh.edu

Abstract – *Image fusion algorithms attempt to produce a single fused image that is more informative then any of the multiple source images used to produce the fused image. Analytical studies of image fusion performance have been lacking. Such studies can augment existing experimental studies by addressing some aspects that are difficult to study using experimental methods. Here, an estimation theory approach is employed using a mathematical model based on the observation that the each different sensor can provide a different quality when viewing a given object in the scene. One sensor may be better for viewing one object and a different sensor may be better for viewing a different object. The model also acknowledges that distortion and noise will enter into the sensor observations. This model allows us to employ known estimation theory techniques to find the best possible fusion performance, measured in terms the standard estimation theory measure of performance. This performance measure has not yet received attention in the image fusion community. Some interesting results include the demonstration that a particular weighted averaging approach is shown to yield optimum estimation performance for the model we focus on. It is also shown that it is important to employ a priori information that describes which sensor is able to provide a good view of the important objects in the scene. The essential aspects of some frequently employed fusion approaches are studied and the capabilities of these approaches are analyzed and compared to the best fusion algorithms. We hope this study will encourage further analytical studies of image fusion.*

*This material is based on work supported by the U. S. Army Research Office under grant number DAAD19-00-1-0431. The content of the information does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred.

1 Introduction

Multisensor data fusion is an area of great interest in the research community. Our focus here is on image-level fusion [1] (called image fusion here), which generates a fused image in which each pixel is determined from a set of pixels in each source image. The fused image should contain a better view of the scene than do any of the source images, thus improving computer or human interpretation. For a survey of the recent literature on image fusion we refer the interested reader to [2]. For the reader unfamiliar with image fusion, we give one example of using image fusion in the concealed weapon detection application discussed in [3, 4] and [5]. Figure 1 shows a pair of visual and 94 GHz millimeter-wave (MMW) images[1]. The visual image provides the outline and the appearance of the people while the MMW image shows the existence of a gun. For a law enforcement application (see [1]) one would like to produce a single image which would provide a good method for judging that there is considerable evidence to suspect that the person on the right has a concealed gun beneath his clothes. This fused image may be very helpful to a police officer, for example, who must respond promptly. Concealed weapon detection is an increasingly important topic in the general area of law enforcement and image fusion has been identified as a key technology to enable progress on this topic.

Measuring the performance of image fusion algorithms is an extremely important task which has received past study. The vast majority of investigations have been focused on developing approaches suitable for experimental evaluation of fused images. Thus, given a set of source images and a few fusion algorithms, the goal is to determine which fusion algorithm is more suitable for a given application. Such experimental procedures can be very useful. However, there

[1]The source images were obtained from Thermotex Corporation.

Figure 1: (a). Visual Image (b). Milimeter-wave Image (c). Fused Image (see [1])

are some inherent limitations to experimental tests. For example, experimental tests are only valid for the particular set of source images considered. Trying to argue that these results apply for a larger set of cases than the exact set considered (with the exact source images and any imperfections in them) is not possible in a rigorous way. Experimental studies can be considered only for a finite number of specific cases with a very specific set of conditions in each case. Thus trying to prove that a fusion approach is generally good is not possible using experimental tests. Further, whenever real data is taken there are always some unknown imperfections that enter into this data and these imperfections can influence the results. Of course one tries to minimize these imperfections, but typically they can not be completely eliminated. It is very easy to study a whole set of cases all at once using analysis and to vary any or all parameters in any way desired. Using mathematical analysis, one can even study the performance of the theoretically best fusion approach as we do in this paper. This can not be accomplished using the existing experimental approaches. Thus while experimental investigations can be useful, analytical investigations can be very helpful in filling in the gaps that such studies can never address. Thus, we want to urge the image fusion research community to begin employing analytical analysis in their research to augment experimental studies. Analytical investigations of image fusion have not received much attention to date. This appears to be one of the first attempts. We are hoping this paper will encourage further analytical image fusion research. It should be emphasized that we just take a first step in this paper with the goal of showing there is some value in analytical image fusion studies. Much work remains to be done on analysis with more comprehensive models.

To perform analytical studies, it is typically necessary to develop a model for the process under consideration. Here we propose a very simple model to show the utility of such studies. We hope that other researchers will build on our ideas and propose more comprehensive models and perform more comprehensive studies based on these models. The model, described in Section 2, relates the sensed images to the underling scene we wish to image. The model is simply a mathematical statement of two simple ideas which are true for the image fusion problem. Thus, analysis using this model should provide useful results. First, the model acknowledges that the different sensors have different capabilities to provide good views of different objects. Second, distortion and noise enter into the sensor images. We can explain the first idea using the example shown in Figure 1 where the underling scene includes the people and the gun. Here, only one sensor can "see" the gun while only the other sensor can "see" the people well.

It is also necessary to focus ones study on a particular aspect of performance. Here we focus on the estimation performance of the best possible image fusion schemes which we call the fundamental limits of image fusion algorithms. For the model we propose, these limits have not been studied. To study the fundamental limits on estimation performance of image fusion schemes, we find the smallest possible mean-squared error between the fused image and the underling scene, which we will compute using the Cramer-Rao [6] bound. The Cramer-Rao bound is well established in estimation theory but it has not yet been employed in the image fusion research community. Further we have not seen studies which calculate the Cramer-Rao bound for our model. We note that we have recently seen some work considering the Cramer-Rao bound for image registration, but image fusion is not discussed in this paper [7]. The Cramer-Rao bound tells us immediately the best achievable estimation performance. Any image fusion approach which achieves this performance is optimum in this regard. Further if simple approaches get us sufficiently close, these simple approaches can be justified.

2 Assumed Statistical Model

In multisensor image fusion it is true that each sensor can generally provide a good view of only a part of the scene in which we are interested, which we call the underling scene. In the example shown in Figure 1, the underlying scene would include the gun, the people and the other important objects. The visual sensor can provide a good view of the people, but not of the gun. The MMW sensor can provide a good view of the gun, but not of the people. We observe the same phenomenon in other multisensor image fusion applications (fusion of IR and low light visual images for night vision is one such application) and we feel this is an important aspect of the problem that we attempt to capture in our model. Of course, being a model, it is an abstraction of the physical world which captures some important aspects of the problem, but which ignores others for simplicity. Assume that we observe q registered sensor images. If an MSD method is to be employed, we assume the required transform has already been applied to these images[2]. We associate the corresponding coefficients in each of these q registered sensor images and then collect each set of q corresponding coefficients into a vector. Now consider one of these vectors. We can model this vector as (see [5])

$$\mathbf{z} = \beta s + \epsilon \tag{1}$$

where s (a scalar) is the value of the coefficient for the underling scene, ϵ (a length q vector) represents all unwanted degradation which includes distortion and noise (we often call ϵ noise for brevity), and $\beta = (\beta_1, \ldots, \beta_q)^T$ is the length q sensor selectivity vector which determines if the underling scene contributes to each component of $\mathbf{z}$. In the general case, we can let each component of β be a real valued quantity. For simplicity we will sometimes let each component of the vector β take on the values $-1, 0, +1$ to model polarity flips and that a given part of the scene is not represented in a given sensor image. The example given in Figure 1 would clearly motivate the choice of β taking on the values $0, +1$ with proper scaling (for example the gun is only visible in one of the two sensor images). When, for example, IR images are employed (see [5] for more details) the choice of -1 also becomes reasonable. In the model given in (1), s is the quantity we want to estimate using the image fusion approach.

Let us initially assume that ϵ is Gaussian with zero mean vector and covariance matrix $diag(\sigma_1^2, \ldots, \sigma_q^2)$. Initially assume s is an unknown deterministic scalar. Also assume β is deterministic. In the sequel we focus on the mean-squared error of these parameters when they are estimated from distorted and noisy observations as in (1). For example, the mean-squared error associated with the estimate $\hat{s}$ of the underling scene is $E\{(s-\hat{s})^2\}$. We will consider cases with Non-Gaussian ϵ and more complicated models later.

Let $f_{\mathbf{z}}(\mathbf{z})$ denote the probability density function (pdf) of the observations from (1). Collect the unknown (assumed deterministic) parameters into the vector $\xi = (\beta_1, \ldots, \beta_q, \sigma_1^2, \ldots, \sigma_q^2, s)^T$ and denote the estimated vector as $\hat{\xi}$. If we view the image fusion problem as an estimation problem then the Cramer-Rao bound [6] becomes applicable. In particular, the error correlation matrix satisfies

$$\begin{aligned} &\mathbf{MSE}(\xi, \hat{\xi}) = E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)^T\} \\ &\geq \frac{\partial E\{\hat{\xi}\}}{\partial \xi} \mathbf{J}^{-1} (\frac{\partial E\{\hat{\xi}\}}{\partial \xi})^T + (E\{\hat{\xi}\} - \xi)(E\{\hat{\xi}\} - \xi)^T \end{aligned} \tag{2}$$

where the elements of the Fisher information matrix (FIM) $\mathbf{J} = -E\{\mathbf{J}_E\}$ are

$$J_{ij} = -E\{\frac{\partial^2}{\partial \xi_i \partial \xi_j} \ln f_{\mathbf{z}}(\mathbf{z})\}. \tag{3}$$

The inequality $\mathbf{A} \geq \mathbf{B}$ between matrices in (2) is interpreted as implying that the matrix $\mathbf{A} - \mathbf{B}$ is positive semi-definite. Note that the diagonal elements of $\mathbf{MSE}(\xi, \hat{\xi})$ represent the mean-squared error in estimating each component of ξ by using $\hat{\xi}$. For an unbiased estimator, $E\{\hat{\xi}\} = \xi$, (2) becomes

$$\mathbf{MSE}(\xi, \hat{\xi}) \geq \mathbf{J}^{-1} \tag{4}$$

while the second term in (2), sometimes called the bias function, gives a partial measure of the impact of bias on the the estimation accuracy. It should be noted that the second term in (2) is a positive semi-definite matrix so it can only increase or maintain the mean-square errors, the diagonal elements of $\mathbf{MSE}(\xi, \hat{\xi})$.

For the zero-mean Gaussian noise model assumed, ξ_i has the pdf $\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp(-\frac{(\xi_i - \beta_i s)^2}{2\sigma_i^2})$. Since the distortion/noise samples at different sensors are assumed uncorrelated, we find

$$\begin{aligned} \ln f_{\mathbf{z}}(\mathbf{z}) &= \ln (\prod_{i=1}^{q} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp[-\frac{(z_i - \beta_i s)^2}{2\sigma_i^2}]) \\ &= \sum_{i=1}^{q} \ln (\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp[-\frac{(z_i - \beta_i s)^2}{2\sigma_i^2}]). \end{aligned} \tag{5}$$

Performing the calculations indicated in (3) and recalling $\mathbf{J} = -E\{\mathbf{J}_E\}$ yields (6) (on top of next page) where $\phi_i = \frac{-(z_i - \beta_i s)^2/\sigma_i^2 + 1/2}{(\sigma_i^2)^2}$ and $\kappa_i = \frac{-\beta_i s + (z_i - \beta_i s)}{\sigma_i^2}$ for $i = 1, \ldots, q$.

Using $\epsilon_i = z_i - \beta_i s$ and recalling that we assume zero

[2] Any multiscale transform or other linear preprocessing will be valid for our analysis, as would no preprocessing, in which case the coefficients are the original pixels.

$$
\mathbf{J}_E = \begin{pmatrix}
-\frac{s^2}{\sigma_1^2} & 0 & \dots & 0 & -\frac{s(z_1-\beta_1 s)}{(\sigma_1^2)^2} & 0 & \dots & 0 & \kappa_1 \\
0 & -\frac{s^2}{\sigma_2^2} & 0 & \dots & 0 & -\frac{s(z_2-\beta_2 s)}{(\sigma_2^2)^2} & 0 & \dots & \kappa_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \dots & 0 & -\frac{s^q}{\sigma_q^2} & 0 & \dots & 0 & -\frac{s(z_q-\beta_q s)}{(\sigma_q^2)^2} & \kappa_q \\
-\frac{s(z_1-\beta_1 s)}{(\sigma_1^2)^2} & 0 & \dots & 0 & \phi_1 & 0 & \dots & 0 & -\frac{\beta_1(z_1-\beta_1 s)}{(\sigma_1^2)^2} \\
0 & -\frac{s(z_2-\beta_2 s)}{(\sigma_2^2)^2} & 0 & \dots & 0 & \phi_2 & 0 & \dots & -\frac{\beta_2(z_2-\beta_2 s)}{(\sigma_2^2)^2} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \dots & \dots & \frac{s(z_q-\beta_q s)}{(\sigma_q^2)^2} & 0 & \dots & 0 & \phi_q & -\frac{\beta_q(z_q-\beta_q s)}{(\sigma_q^2)^2} \\
\kappa_1 & \kappa_2 & \dots & \kappa_q & -\frac{\beta_1(z_1-\beta_1 s)}{(\sigma_1^2)^2} & -\frac{\beta_2(z_2-\beta_2 s)}{(\sigma_2^2)^2} & \dots & -\frac{\beta_q(z_q-\beta_q s)}{(\sigma_q^2)^2} & -\sum_{i=1}^q \frac{\beta_i^2}{\sigma_i^2}
\end{pmatrix} \tag{6}
$$

mean noise yields $\mathbf{J} =$

$$
\begin{pmatrix}
\frac{s^2}{\sigma_1^2} & \dots & 0 & 0 & \dots & 0 & \frac{\beta_1 s}{\sigma_1^2} & \\
0 & 0 & \dots & 0 & 0 & \dots & \frac{\beta_2 s}{\sigma_2^2} & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \ddots & 0 & 0 & 0 & 0 & \frac{\beta_{q-1} s}{\sigma_{q-1}^2} & \\
0 & 0 & \frac{s^2}{\sigma_q^2} & 0 & 0 & 0 & \frac{\beta_q s}{\sigma_q^2} & \\
0 & 0 & 0 & \frac{1}{2(\sigma_1^2)^2} & \dots & 0 & 0 & \\
0 & 0 & 0 & 0 & 0 & \dots & 0 & \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\
0 & 0 & 0 & 0 & \dots & \frac{1}{(2\sigma_q^2)^2} & 0 & \\
\frac{\beta_1 s}{\sigma_1^2} & \dots & \frac{\beta_q s}{\sigma_q^2} & 0 & \dots & 0 & \sum_{i=1}^q \frac{\beta_i^2}{\sigma_i^2} &
\end{pmatrix} \tag{7}
$$

3 Bounds on Unbiased Fusion Performance

We now compute the Cramer-Rao bound for the model introduced in the previous section while varying our assumptions to accommodate a few cases of practical interest.

3.1 $\beta_1, \dots, \beta_q$ Known

First we investigate the impact of knowing $\beta_1, \dots, \beta_q$ on the image fusion. We note from the example shown in Figure 1 that the images, the scene, and our interpretation of it give some information on which sensor images will provide useful information in certain pixel locations. This is essentially equivalent to knowing something about $\beta_1, \dots, \beta_q$. Thus it is reasonable we can know something about $\beta_1, \dots, \beta_q$ and clearly in the best case we know their values exactly. What does this imply for the image fusion result? This is addressed in the following Theorem.

Theorem 1: Assume an unbiased estimator and the model in (1) for the case when $\beta_1, \dots, \beta_q$ are known. A bound on the smallest variance of an estimation of s is

$$
\hat{\sigma}_s^2 = E\{(s - \hat{s})^2\} \geq \frac{1}{\sum_{i=1}^q \beta_i^2/\sigma_i^2}. \tag{8}
$$

Outline of the proof:

If $\beta_1, \dots, \beta_q$ are known then we can take them out[3] of the vector ξ. Thus we should remove the corresponding columns and rows from $\mathbf{J}$. In this case we remove the first q rows and columns which results in $\mathbf{J}$ becoming

$$
\mathbf{J} = \begin{pmatrix}
\frac{1}{2(\sigma_1^2)^2} & 0 & \dots & 0 & 0 \\
0 & \frac{1}{2(\sigma_2^2)^2} & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \dots & \frac{1}{2(\sigma_q^2)^2} & 0 \\
0 & 0 & \dots & 0 & \sum_{i=1}^q \frac{\beta_i^2}{\sigma_i^2}
\end{pmatrix} = \mathbf{J}_{\overline{\beta}}. \tag{9}
$$

This gives $\mathbf{J}^{-1} = diag[2(\sigma_1^2)^2, 2(\sigma_2^2)^2, 2(\sigma_q^2)^2, 1/(\beta_1^2/\sigma_1^2 + \dots + \beta_q^2/\sigma_q^2)]$. Thus for an unbiased estimator ($E\{\hat{\xi}\} = \xi$) (4) is valid so that we find

$$
\hat{\sigma}_s^2 = E\{(s - \hat{s})^2\} \geq CRB_s = \frac{1}{\sum_{i=1}^q \beta_i^2/\sigma_i^2}. \tag{10}
$$

□

We can gain further insight into CRB_s by understanding how each sensor affects it. Note that the bound in (8) can be computed by calculating the effective variance of each sensor as $\hat{\sigma}_i^2 = \sigma_i^2/\beta_i^2$ and then combining them as one would combine resistors in parallel as in

$$
CRB_s = \frac{1}{\sum_{i=1}^q \frac{1}{\hat{\sigma}_i^2}}. \tag{11}
$$

[3]This can also be handed using the constrained Cramer-Rao bound, see [8].

In fact, just as for resistors in parallel, any sensor images with variances that are much larger than the others will have very limited impact on the bound in (8). This is summarized by the following Lemma.

Lemma 1: Assuming $\hat{\sigma}_q^2 >> \hat{\sigma}_1^2, \ldots, \hat{\sigma}_{q-1}^2$,

$$CRB_s = \frac{1}{\sum_{i=1}^{q} \frac{1}{\hat{\sigma}_i^2}} \approx \frac{1}{\sum_{i=1}^{q-1} \frac{1}{\hat{\sigma}_i^2}}. \quad (12)$$

Outline of the proof:
It is easy to see that $\hat{\sigma}_q^2 >> \hat{\sigma}_1^2, \ldots \hat{\sigma}_{q-1}^2$ implies $1/\hat{\sigma}_q^2 << 1/\hat{\sigma}_1^2, \ldots 1/\hat{\sigma}_{q-1}^2$ so ignoring the last term in the denominator of (8) is justified.

□

These same basic ideas lead to bounds based on each individual sensor image. In particular, we have the following Lemma.

Lemma 2: Assume that for all values of i we have $\hat{\sigma}_i^2 > 0$. Further define $\hat{\sigma}_{min}^2 = min[\hat{\sigma}_1^2, \ldots, \hat{\sigma}_q^2]$. For any i we have

$$CRB_s = \frac{1}{\sum_{i=1}^{q} \frac{1}{\hat{\sigma}_i^2}} < \hat{\sigma}_{min}^2 \leq \hat{\sigma}_i^2. \quad (13)$$

The Lemma follows easily from using the fact that each term in the denominator of (8) is positive. The asymptotic behavior of the bound in (13) is also of interest and this is addressed in the next Lemma.

Lemma 3: Assume an infinite number of sensors, all with the same noise variance $\sigma_i^2 = \sigma^2$. Further assume that $\beta_i^2 = 1$ for a percentage $P(V)$ ($\beta_i = 0$ for the others) of all $i = 1, \ldots, q$ as $q \to \infty$. Then

$$\lim_{q\to\infty} CRB_s = \lim_{q\to\infty} \frac{1}{\sum_{i=1}^{q} \beta_i^2/\sigma_i^2} = \lim_{q\to\infty} \frac{1}{\frac{P(V)q}{\sigma^2}} = 0 \quad (14)$$

The proof is summarized in (14). Lemma 3 implies that perfect performance is apparently possible with enough sensors. A comment on $P(V)$ is in order. Roughly speaking, with the proper ergodicity assumptions, $P(V)$ plays a role similar to the probability that a particular object is visible to a given sensor. Of course the result in Lemma 3 can be generalized for some cases with sensors with nonidentical sensor noise statistics as illustrated by the following Lemma

Lemma 4: Assume that $\beta_i^2 = 1$ for a percentage $P(V)$ of all $i = 1, \ldots, q$ as $q \to \infty$. Also assume that at least a certain percentage[4] $P(S)$ of the sensors have variance smaller than σ^2 and assume these sensors are distributed in a manner unrelated to those for which $\beta_i^2 = 1$. Then we can bound the impact of neglecting the positive terms in the denominator of (8) (similar as argued to obtain (13)) so that

$$\lim_{q\to\infty} CRB_s = \lim_{q\to\infty} \frac{1}{\sum_{i=1}^{q} \beta_i^2/\sigma_i^2} \leq \lim_{q\to\infty} \frac{1}{\frac{P(V)P(S)q}{\sigma^2}} = 0 \quad (15)$$

Lemma 4 (and similar constructions) indicate that in a very general sense, increasing the number of sensors to a sufficiently large number will drive the performance bound towards allowing perfect estimates. Further, due to the monotone decreasing nature of (15), we can choose the number of sensors to make performance as close to perfect as we desire. We note that it is possible to generalize these results for cases where β_i^2 can take values other than 0 or 1 and that this follows closely from the results given, but we do not discuss this further.

3.2 $\beta_1, \ldots, \beta_q$ Unknown

The following Theorem considers the case where we have no knowledge about $\beta_1, \ldots, \beta_q$.

Theorem 2: Assume an unbiased estimator and the model in (1) for the case when $\beta_1, \ldots, \beta_q$ are completely unknown (we have no knowledge about $\beta_1, \ldots, \beta_q$). In this case it is not possible to obtain an accurate estimate of the underling scene.

Outline of the proof:
For convenience, we break $\xi = (\beta_1, \ldots, \beta_q, \sigma_1^2, \ldots, \sigma_q^2, s)^T = (\alpha^T, s)^T$ into two parts. Now using these two parts, we let $\mathbf{J}$ be partitioned as ($J_{\alpha s} = -E\{\frac{\partial}{\partial \alpha}\frac{\partial}{\partial s} \ln f_{\mathbf{z}}(\mathbf{z})\}$)

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{\alpha\alpha} & \mathbf{J}_{\alpha s} \\ \mathbf{J}_{s\alpha} & \mathbf{J}_{ss} \end{pmatrix}. \quad (16)$$

Assuming an unbiased estimate and using standard results on the inverse of a partitioned matrix on (7) we have

$$\hat{\sigma}_s^2 = E\{(s-\hat{s})^2\} \geq \mathbf{J}^{-1} = \left(\mathbf{J}_{ss} - \mathbf{J}_{s\alpha}\mathbf{J}_{\alpha\alpha}^{-1}\mathbf{J}_{\alpha s}\right)^{-1} = \left(\sum_{i=1}^{q} \frac{\beta_i^2}{\sigma_i^2} - \sum_{i=1}^{q} \frac{\beta_i s}{\sigma_i^2}\frac{\sigma_i^2}{s^2}\frac{\beta_i s}{\sigma_i^2}\right)^{-1} = (0)^{-1} = \infty. \quad (17)$$

which implies that, given our model, we can't accuracy estimate the underling scene when we have no knowledge about the sensor selectivity factors.

□

[4] $P(S)$ is related to the probability of small variance in a similar manner as discussed for $P(V)$.

The conclusion of Theorem 2 is in stark contrast to the case where we know the sensor selectivity factors. In truth, the situation is between these two extremes. However, this result clearly emphasizes the difficult nature of the image fusion problem. It tells us that the image fusion problem can be very difficult if we do not have some a priori knowledge about the observed scene or if we are unable to extract information by interpreting the scene. This is not unreasonable since we recognize that humans tend to extract significant information by interpreting a scene, often without being completely aware that this scene interpretation process has taken place. The result in (16) seems to indicate that it is very important to try to extract such information.

The true situation is somewhere between the case of known $\beta_1, \ldots, \beta_q$ and unknown $\beta_1, \ldots, \beta_q$. Further, it is probably worth stating that one should be careful about overstating the implications of the result in (17). In particular, if we change the assumptions used to obtain (17) very slightly, we can change the conclusion that accurate estimation is not possible. As an example, consider the the problem formulation in [5] where the $\beta_1, \ldots, \beta_q$ take on constrained values. In this particular case each β_i can only take on the values $-1, 0, +1$. Now suppose that we can accurately determine which β_i are zero. This case is addressed in the following Lemma.

Lemma 5: Assume an unbiased estimator and the model in (1) for the case where $\beta_1, \ldots, \beta_q$ take on only the values $-1, 0, +1$. Assume we can accurately determine which β_i are zero. Then the constrained Cramer-Rao bound reduces to that in (8) which is for the case of known $\beta_1, \ldots, \beta_q$.

Outline of the proof:

If we can accurately determine which β_i are zero, we remove them from consideration as per (9) and definite a new matrix $\mathbf{J}$ with the rows and columns corresponding to these variables removed. The remaining β_i satisfy the constraint that $\beta_i^2 = 1$. The constrained Cramer-Rao bound [8, 9] is computed by first collecting these scalar constraint equations into a vector and next computing the gradient matrix with respect to the parameter vector ξ. Define an orthogonal matrix $\mathbf{U}$ whose columns span the null space of the gradient matrix. Then the constrained Cramer-Rao bound [8, 9] is $\mathbf{MSE}(\xi, \hat{\xi}) =$

$$E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)^T\} \geq \mathbf{U}\left(\mathbf{U}^T \mathbf{J} \mathbf{U}\right)^{-1} \mathbf{U}^T = \mathbf{J}^{-1} = \mathbf{J}_{\overline{\beta}}^{-1} \tag{18}$$

with $\mathbf{J}_{\overline{\beta}}$ from (9). Thus (18) once again leads to the result in (8).

□

3.3 $\beta_1, \ldots, \beta_q$ Known with a Window of Data

A useful generalization of the observation model in (1) jointly considers a set of coefficients which we henceforth call a window. In the general case we can index each coefficient in the window with the index $w \in \mathcal{W}$. If we assume that β and s (same definition as in (1)) are constant over the whole window $\mathcal{W}$ then (1) becomes

$$\mathbf{z}_w = (z_{1w}, \ldots, z_{qw})^T = \beta s + \epsilon_w \quad \forall w \in \mathcal{W} \tag{19}$$

where $\epsilon_w = (\epsilon_{1w}, \ldots, \epsilon_{qw})^T$ is a length q vector of independent Gaussian random variables, each with zero mean and variance σ_{iw}^2. We note that the log-likelihood function in (5) now becomes

$$\begin{aligned}\ln f_{\mathbf{z}}(\mathbf{z}) &= \ln\Big(\prod_{w\in\mathcal{W}} \prod_{i=1}^{q} \frac{1}{\sqrt{2\pi\sigma_{iw}^2}} \exp\Big[-\frac{(z_{iw} - \beta_i s)^2}{2\sigma_{iw}^2}\Big]\Big) \\ &= \sum_{w\in\mathcal{W}} \sum_{i=1}^{q} \ln\Big(\frac{1}{\sqrt{2\pi\sigma_{iw}^2}} \exp\Big[-\frac{(z_{iw} - \beta_i s)^2}{2\sigma_{iw}^2}\Big]\Big).\end{aligned} \tag{20}$$

For the case of $\beta_1, \ldots, \beta_q$ unknown, the Cramer-Rao bound blows up for the same reasons as in (17). On the other hand if $\beta_1, \ldots, \beta_q$ are known, then using calculations similar to (7) we obtain

$$\hat{\sigma}_s^2 = E\{(s - \hat{s})^2\} \geq CRB_s = \frac{1}{\sum_{w\in\mathcal{W}} \sum_{i=1}^{q} \beta_i^2 / \sigma_{iw}^2}. \tag{21}$$

It is immediately evident that the role of the window index w is similar to the role of the sensor index i. In particular, it is clear that we can drive the Cramer-Rao bound in (21) to zero if we can form a window of infinite size. Let $L_{\mathcal{W}}$ be the size of window, thus $L_{\mathcal{W}}$ is the number of elements in the set $\mathcal{W}$. Let $P(V)$ and $P(S)$ be defined similarly as in (15) where each now gives a percentage of all terms in the sequence $\sigma_{iw}^2, i = 1, \ldots, q, w \in \mathcal{W}$. Then from (21) we get

$$\begin{aligned}\lim_{L_{\mathcal{W}}\to\infty} CRB_s &= \lim_{L_{\mathcal{W}}\to\infty} \frac{1}{\sum_{w\in\mathcal{W}} \sum_{i=1}^{q} \beta_i^2 / \sigma_{iw}^2} \\ \leq \lim_{L_{\mathcal{W}}\to\infty} \frac{1}{\frac{P(V)P(S)qL_{\mathcal{W}}}{\sigma^2}} &= \lim_{L_{\mathcal{W}}\to\infty} \frac{\sigma^2}{P(V)P(S)qL_{\mathcal{W}}} = 0\end{aligned} \tag{22}$$

3.4 Non-Gaussian ϵ

Consider the model in (1) where ϵ is a random vector whose components are independently and identically distributed, each with the common univariate pdf f. Assume β known. For the case where f is a zero mean Gaussian pdf with variance σ^2, the result in (10) implies that the Cramer-Rao bound is

$$\hat{\sigma}_s^2 = E\{(s - \hat{s})^2\} \geq CRB_G = \frac{1}{\sum_{i=1}^{q} \beta_i^2 / \sigma^2}. \tag{23}$$

Using the chain rule [10, 11] and (4) one can show that the Cramer-Rao bound for the non-Gaussian f case, call it CR_{NG}, can be related to the Cramer-Rao bound in (23) as

$$CRB_{NG} = \frac{\sigma^2 CRB_G}{I_c} \quad (24)$$

where (ξ_i is the ith component of ξ and $f'(\xi_i) = \frac{d}{d\xi_i} f(\xi_i)$)

$$I_c = E\left\{ \left(\frac{f'(\xi_i)}{f(\xi_i)} \right)^2 \right\} = \int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)} \right)^2 f(x)dx. \quad (25)$$

We note that the value of I_c in (24) depends on the shape of f chosen. If f is a zero mean Gaussian pdf with variance σ^2 then $f'(x) = -\frac{x}{\sigma^2} f(x)$ and $I_c = \frac{1}{\sigma^2}$. See [10, 11] for a discussion of computing I_c for the Gaussian mixture model considered in [5], which is a fairly general nonGaussian model. We note [10, 11] that I_c in (25) is invariant to a scaling of the argument of f which can result in a change in variance.

Now we give the key result of this section. Using the Cauchy Schwartz inequality (under the assumption that $Lim_{x\to\infty} x f(x)$ is finite and f symmetric) we can find [10]

$$\begin{aligned} I_c \sigma^2 &= E\left\{ \left(\frac{f'(\xi_i)}{f(\xi_i)} \right)^2 \right\} E\left\{ \xi_i^2 \right\} \\ &\geq |E\left\{ \frac{f'(\xi_i)}{f(\xi_i)} \xi_i \right\}|^2 \\ &= \left| \int_{-\infty}^{\infty} \frac{f'(\xi_i)}{f(\xi_i)} \xi_i f(\xi) d\xi_i \right|^2 \\ &= |Lim_{\xi_i \to \infty} \xi_i f(\xi_i) - Lim_{\xi_i \to -\infty} \xi_i f(\xi_i) \\ &\quad - \int_{-\infty}^{\infty} f(\xi_i) d\xi_i|^2 = 1, \end{aligned} \quad (26)$$

where integration by parts is used in the last step. The result in (26) can be interpreted as implying that $I_c \geq \frac{1}{\sigma^2}$ which is the value of I_c for f Gaussian. Thus for any well behaved f with fixed variance σ^2, the Gaussian case gives the largest Cramer-Rao bound. For this reason, and the fact that the non-Gaussian results can be obtained by a multiplicative factor as shown in (24), we focus on the Gaussian case in the rest of this paper.

4 Performance of Some Fusion Algorithms

In the previous section we consider bounds on the performance of unbiased fusion algorithms. Here we consider some actual algorithms and the actual performance they obtain. Due to space limitations, we only give some of our investigations. A more complete presentation will be given in [12].

4.1 Optimum Estimator

First we consider an algorithm which achieves the optimum performance using the measures we have employed. Thus we consider an algorithm that is unbiased and achieves the Cramer-Rao bound. We consider the case where $\beta_1, \ldots, \beta_q$ are known to avoid unbounded variance. Further we consider the case where a window is employed to be general (see (19)). Thus using Equation (3.7) of Theorem 3.1 of [6] we can assert that the optimum estimator is

$$\hat{s} = \sum_{i=1}^{q} \sum_{w \in \mathcal{W}} \left(\frac{\frac{\beta_i}{\sigma_{iw}^2}}{\sum_{i=1}^{q} \sum_{w \in \mathcal{W}} \frac{\beta_i^2}{\sigma_{iw}^2}} \right) z_{iw} \quad (27)$$

which is unbiased and achieves the Cramer-Rao lower bound. It is interesting that (27) uses a weighted average, which is a linear function of the observed data. The calculation in (27) weights each observation by its reliability $\frac{\frac{\beta_i}{\sigma_{iw}^2}}{\sum_{i=1}^{q} \sum_{w \in \mathcal{W}} \frac{\beta_i^2}{\sigma_{iw}^2}}$ which intuitively should be proportional to the selectivity and inversely proportional to the noise power in the observation as (27) asserts. Since the observation z_{iw} is $\beta_i s$ plus zero-mean noise, the normalization in the denominator of (27) selects the proper contribution for each observation to give an unbiased estimator $E\{\hat{s}\} = s$. In case the noise averaging operations in (27) are not clear, we can consider the case where $\sigma_{iw}^2 = \sigma_i^2 \forall i$ for which (27) reduces to

$$\hat{s} = \sum_{i=1}^{q} \left(\frac{\frac{\beta_i}{\sigma_i^2}}{\sum_{i=1}^{q} \frac{\beta_i^2}{\sigma_i^2}} \right) \left(\frac{1}{L_{\mathcal{W}}} \sum_{w \in \mathcal{W}} z_{iw} \right). \quad (28)$$

In this case the reliabilities are constant across w. Intuitively (28) attempts to use all the observations to first average the affects of the noise by calculating $\frac{1}{L_{\mathcal{W}}} \sum_{w \in \mathcal{W}} z_{iw}$ which results in noise with a lower variance. Then the noise reduced observations from each sensor are weighted by their reliabilities. The case with no window $L_{\mathcal{W}} = 1$ (we observe $z_1, \ldots, z_q$) follows directly from (28) as

$$\hat{s} = \sum_{i=1}^{q} \left(\frac{\frac{\beta_i}{\sigma_i^2}}{\sum_{i=1}^{q} \frac{\beta_i^2}{\sigma_i^2}} \right) z_i \quad (29)$$

4.2 General Weighted Average

Consider a weighted average approach, with the weights $g_{iw}, i = 1, \ldots, q, w \in \mathcal{W}$, so that

$$\hat{s} = \sum_{i=1}^{q} \sum_{w \in \mathcal{W}} g_{iw} z_{iw}. \quad (30)$$

The bias when using such an estimator is

$$E\{\hat{s}-s\} = \left(\sum_{i=1}^{q}\sum_{w\in\mathcal{W}} g_{iw}\beta_i - 1\right) s. \quad (31)$$

Thus, in general, unless $\sum_{i=1}^{q}\sum_{w\in\mathcal{W}} g_{iw}\beta_i = 1$, which is true for the weights in (27), then the estimator is biased. Thus, while the optimum estimator uses a particular weighted average, not all weighted averaging estimators are optimum. The variance $Var(\hat{s}) = E\{(\hat{s}-E\{\hat{s}\})^2\}$ of a general weighted averaging estimator is

$$Var(\hat{s}) = E\{(\sum_{i=1}^{q}\sum_{w\in\mathcal{W}} g_{iw}(z_{iw}-E\{z_{iw}\}))^2\}$$

$$= \sum_{i=1}^{q}\sum_{w\in\mathcal{W}} g_{iw}^2 E\{(z_{iw}-E\{z_{iw}\})^2\} = \sum_{i=1}^{q}\sum_{w\in\mathcal{W}} g_{iw}^2\sigma_{iw}^2 \quad (32)$$

where the independence of ϵ_{iw} and $\epsilon_{jw'}$ (if either $i \neq j$ or $w \neq w'$) was employed. From (32) it is clear that even if an estimator is unbiased ($\sum_{i=1}^{q}\sum_{w\in\mathcal{W}} g_{iw}\beta_i = 1$), the variance may still not be optimum. However, inserting the weights from (27) gives a variance equal to the Cramer-Rao bound as suggested previously. If suboptimum weights are chosen, it is known [13] that the resulting performance can be quite bad in some cases.

5 Conclusions

We have described a simple but reasonable model for the image fusion problem and we have used this model to compute the Cramer-Rao bound and to find optimum fusion algorithms for this performance measure. We have given results for both Gaussian and non-Gaussian distortion-plus-noise and for some more complicated models. We also evaluate the estimation capabilities of the key parts of some common fusion approaches so we can compare these capabilities to the fundamental limits.

References

[1] Z. Zhang and R. S. Blum. A categorization and study of multiscale-decomposition-based image fusion schemes. Proceedings of the IEEE, Vol. 87, No. 8, pp. 1315-1326, Aug. 1999.

[2] R. S. Blum, Z. Xue, and Z. Zhang. An Overview of Image Fusion. Chapter in the book *Multi-Sensor Image Fusion and Its Applications, Edited by Rick Blum, Zheng Liu*, Signal and Image Processing Series of Marcel Dekker/CRC Press.

[3] P. K. Varshney, H. Chen, L. C. Ramac, et. al. Registration and fusion of infrared and millimeter wave images for concealed weapon detection. Proceedings of International Conference on Image Processing, Vol. 3, pp. 532-536, Japan, Oct. 1999.

[4] Z. Zhang and R. S. Blum. A region-based image fusion scheme for concealed weapon detection. Proceedings of 31st Annual Conference on Information Sciences and Systems, pp. 168-173, Baltimore, MD, Mar. 1997.

[5] J. Yang, and R. S. Blum. A statistical signal processing approach to image fusion for concealed weapon detecetion. IEEE International Conference on Image Processing, pp. 513-516, Rochester, NY, Sept. 2002.

[6] S. M. Kay. Fundamentals of Statistical Signal Processing: Estimation Theory. Upper Saddle River, NJ:Prentice Hall, 1993.

[7] D. Robinson, and P. Milanfar. Fundamental performance limits in image registration. Proceedings of the 2003 International Conference on Image Processing, Vol. 3, pp. 323-326, Sept. 2003.

[8] B. M. Sadler, R. J. Kozick, and T. Moore. Bounds on bearing and symbol estimation with side information. IEEE Transactions on Signal Processing, Vol. 49, No. 4, pp. 822-834, April 2001.

[9] P. Stoica and B. C. Ng. On the Cramer-Rao bound under parametric constraints. IEEE Signal Processing Letters, Vol. 5, No. 7, pp. 177-179, July 1998.

[10] S. Kay, and D. Sengupta. Recent advances in non-Gaussian autoregressive processes. In I. S. Haykin (Ed.), Advances in Spectrum Analysis and Array Processing, Englewood Cliffs, NJ: Prentice-Hall, 1991.

[11] R. J. Kozick and B. M. Sadler. Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures. IEEE Transactions on Signal Processing, Vol. 48, No. 12, pp. 3520-3535, Dec. 2000.

[12] R. S. Blum. On Multisensor Image Fusion Performance Limits from an Estimation Theory Perspective, to appear in *Information Fusion* 2004 (in press, available online).

[13] R. S. Blum. Robust Image Fusion using a Statistical Signal Processing Approach, to appear in *Information Fusion* 2004 (in press, available online).