Performance Bound of Position Estimation in Image Matching

Bao-shu XU\textsuperscript{1,2,3,4}, Ze-lin SHI\textsuperscript{1,3,4}

\textsuperscript{1}Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, PR China;
\textsuperscript{2}Graduate University of Chinese Academy of Sciences, Beijing, 100039, PR China;
\textsuperscript{3}Key Laboratory of Optical-Electronics Information Processing, Chinese Academy of Sciences, Shenyang 110016, PR China;
\textsuperscript{4}Key Laboratory of Image Understanding and Computer Vision, Liaoning Province, Shenyang 110016, PR China

Keyword: Performance prediction; Image matching; Cramer-Rao Lower Bound; Fisher information;

Abstract. Position estimation in image matching is a fundamental step in computer vision and image processing. To deal with the problem of performance prediction, we formulate it under the statistical parameter estimation aspect. The lower bound of position estimation variance is obtained based on Cramer-Rao lower bound (CRLB) theory. This paper analyses the impact of noise to 1-D signal matching, derives the lower bound of variance, and then extends it to 2-D image matching. Furthermore, we derive numerical expression that can be computed from observed data. Finally, we use Monte Carlo simulation method to verify the derived analytical expressions. Experimental results show that the derived CRLB is tight to simulation estimated variance. The CRLB can characterize the performance bound of position estimation in image matching.

Introduction

Position estimation in image matching is a fundamental step in the applications of computer vision and image processing, such as target recognition, target tracking, image mosaic, super-resolution image acquisition, image stabilization, etc. A lot of image matching algorithms have been reported in the last decades. The performance of image matching algorithms is usually evaluated through comparable experiments. But the experimental results may change as source images vary, and then the compared results are not steady. Unfortunately, few theoretical analysis studies have been reported in this issue. In target recognition field, some researchers point out that the ability to predict the fundamental performance of model-based object recognition is essential for transforming object recognition from an art to a science, and to speed up the design process for recognition systems\textsuperscript{[1]}. Stochastic factors in imaging process always lead to position estimation error. For example, the target to background apparent contrast is often low in remote imaging system application, so does signal to noise ratio (SNR). There is great difference between two noisy signals.

Being able to place a lower bound on the variance of any unbiased estimator proves to be extremely useful in practice. It provides a benchmark against which we can compare the performance of any unbiased estimator, and alerts us to the physical impossibility of finding an unbiased estimator whose variance is less than the bound\textsuperscript{[2]}. It can give algorithm developer a goal to shoot for. If performance is already near the bound, then there may be little point in spending more resources to further improving the algorithm\textsuperscript{[3]}. The position estimation problem can be formulated as a statistical parameter estimation problem and Cramer-Rao Lower bound can be introduced as performance measure. The CRLB is a lower bound on the variance (covariance) of any unbiased estimator. It can be asymptotically achieved by the maximum likelihood estimator (MLE), which can be defined as the inverse of the Fisher information matrix (FIM), such that\textsuperscript{[2][5]}

\[ \text{cov}(\hat{\theta}) = E(\theta - \hat{\theta})(\theta - \hat{\theta})^T \geq \text{FIM}^{-1} \]  

Where each element of the FIM is defined as\textsuperscript{[2][5]}

---

All rights reserved. No part of contents of this paper may be reproduced or transmitted in any form or by any means without the written permission of TTP, www.ttp.net.
In (2), $p(x;\theta)$ is the probability density function (PDF) of the observed data $x$ at the parameter vector $\theta$. All of previous work \[6\] \[7\] just gathered the PDF of reference data and current data, and didn’t obtain joint PDF of observed data authentically.

This paper is organized as follows. Section II gives the theoretical expression of 1-D signal matching under Additive Whiten Gaussian Noise (AWGN). Section III extends the expression of 1-D to 2-D position estimation in image matching, and discusses the covariance of the horizontal and vertical variables. Monte Carlo simulation method is used to verify the analytical expressions in section IV. Finally, section V gives the conclusion of this paper.

Preformance Bound for 1-d Signal

Performance prediction problem is nearly identical between 1-D and 2-D signals. For simplicity, we first discuss performance bound for 1-D signal matching. Noise, a significant source of matching error, is added in the model. This paper assume that the noise belongs to gauss distribution, and it is additive. So the reference data and current data is equal to original signal plus noise respectively

$$
\begin{align*}
\hat{s}_1(x) &= s(x) + n_1(x) \\
\hat{s}_2(x) &= s(x) + n_2(x)
\end{align*}
$$

(3)

In (3), $s$ is the original signal that is noiseless, $s_1$ is the reference data, $s_2$ is the current data, $x$ is the position and $t$ is the desired shift between $s_1$ and $s_2$. The joint probability density function of $s_1$ and $s_2$ at signal $s$ and parameter $t$ can be described as

$$
P(s_1, s_2 | s, t) = \prod_{x \in S} \frac{1}{\sqrt{2\pi(2\sigma)}} \exp \left\{ \frac{[s_1(x) - s_2(x) + s(x) + t - s(x)]^2}{2(2\sigma)^2} \right\}
$$

(4)

In (4), $S$ represents the sample space, its length is N, cardinality, the number of discrete elements in it. Therefore, the log-PDF function is that:

$$
\ln P(s_1, s_2 | s, t) = -\frac{1}{4\sigma^2}\sum_{x \in S} [s_1(x) - s_2(x) + s(x) + t - s(x)]^2 + c
$$

Where, the last symbol, c, is a constant that can be calculated from (4). The first derivative of log-PDF can be deduced as:

$$
\frac{\partial \ln P(s_1, s_2 | s, t)}{\partial t} = -\frac{1}{2\sigma} \sum_{x \in S} [s_1(x) - s_2(x) + s(x) + t - s(x)] \frac{\partial^2 s(x + t)}{\partial x}
$$

And, the second order derivative of log-PDF is that:

$$
\frac{\partial^2 \ln P(s_1, s_2 | s, t)}{\partial t^2} = -\frac{1}{2\sigma^2} \sum_{x \in S} \left\{ \frac{\partial^2 s(x + t)}{\partial x} + [s_1(x) - s_2(x) + s(x) + t - s(x)] \frac{\partial^2 s(x + t)}{\partial x^2} \right\}
$$

(5)

The expectation of the latter term on the right of (5) can be described as:

$$
E \left\{ \sum_{x \in S} [s_1(x) - s_2(x) + s(x) + t - s(x)] \frac{\partial^2 s(x + t)}{\partial x^2} \right\} = \sum_{x \in S} E \left\{ [s_1(x) - s_2(x) + s(x) + t - s(x)] \frac{\partial^2 s(x + t)}{\partial x^2} \right\}
$$

$$
= \sum_{x \in S} \frac{\partial^2 s(x + t)}{\partial x^2} E\left\{ [s_1(x) - s_2(x)] \right\}
$$

$$
= 0
$$

(6)
Because \( s \) is the original unknown signal, we use the observed data \( s_1 \) and \( s_2 \) to represent \( s \). From (3), the first order derivative of the equation, it holds that:

\[
E \left\{ \sum_{x} \left( \frac{\partial s(x)}{\partial x} \right)^2 \right\} = E \left\{ \sum_{x} \left( \frac{\partial s(x+1)}{\partial x} + \frac{\partial n(x)}{\partial x} \right)^2 \right\} = E \left\{ \sum_{x} \left[ \left( \frac{\partial s(x+1)}{\partial x} \right)^2 + \left( \frac{\partial n(x)}{\partial x} \right)^2 + 2 \frac{\partial s(x+1)}{\partial x} \frac{\partial n(x)}{\partial x} \right] \right\}
\]

We see that:

\[
E \left\{ \sum_{x} \left( \frac{\partial s(x+1)}{\partial x} \right)^2 \right\} = \sum_{x} \left( \frac{\partial s(x+1)}{\partial x} \right)^2 = 0 \quad E \left\{ \sum_{x} \left( \frac{\partial n(x)}{\partial x} \right)^2 \right\} = \sum_{x} \left( \frac{\partial n(x)}{\partial x} \right)^2 = N \left( x+1 - n(x) \right)^2 = 2N \sigma^2
\]

Therefore,

\[
E \left\{ \sum_{x} \left( \frac{\partial n(x)}{\partial x} \right)^2 \right\} + 2N \sigma^2 \quad (7)
\]

From (5), (6) and (7), we can obtain the desired fisher information matrix as that:

\[
FIM = -E \left[ \frac{\partial^2 \ln \mathbb{P}(s_1,s_2,s,t)}{\partial t^2} \right] = \frac{1}{2 \sigma^2} \sum_{x} \left( \frac{\partial s(x+1)}{\partial x} \right)^2 \quad (8)
\]

In (8), we found that the FIM is heavily dependent of \( s_1 \). The parameter estimation theory tells us that the performance prediction will be more accurate as more samples. If the observed data \( s_2 \) is also introduced, the fisher information matrix will be more accurate. Then, we modify the expression of FIM as

\[
FIM = \frac{1}{4 \sigma^2} \sum_{x} \left[ \left( \frac{\partial n(x)}{\partial x} \right)^2 + \left( \frac{\partial s(x)}{\partial x} \right)^2 \right] - N \quad (9)
\]

So far, we achieve a conclusion that, lower bound on the variance of any unbiased estimator is the inverse of the Fisher information matrix, see (1).

**Performance Bound for 2-d Image Matching**

In this section, we extend the results in 1-D position estimation to 2-D image matching. Image can be treated as two-dimensional signals. In imaging process, magnitude fluctuations cannot be avoided in many important steps, such as photo-detection, read-out circuits, etc. We assume that the noise is of independent identical distribution (IID) in any pixels, and it is AWGN. Likewise in section II, the reference image and current image is equal to original image plus noise respectively

\[
\begin{align*}
I_{1}(x,y) &= I(x+u,y+v) + n(x,y) \\
I_{2}(x,y) &= I(x,y) + n(x,y)
\end{align*}
\]

(10)

In (10), \( I \) is original and noiseless image, \( I_1 \) is reference image, and \( I_2 \) is current image, \( x \) and \( y \) is vertical and horizontal position respectively, \( u \) and \( v \) is shift between \( I_1 \) and \( I_2 \) in vertical and horizontal direction respectively. Then, the joint probability density function of \( I_1 \) and \( I_2 \) at image \( I \) and parameter \( [u \ v]^T \) can be described as

\[
P(I_1, I_2 | I, u, v) = \frac{1}{\sqrt{2\pi \sqrt{2\sigma}}} \exp \left[ -\frac{(I_{2}(x,y) - I_{1}(x,y) + I(x+u,y+v) - I(x,v))^2}{2(\sqrt{2\sigma})} \right] \quad (11)
\]

The joint PDF is different to previous work \([4][5][6]\), it is not the product of observed data and current data in this paper.
Next deriving process is similar with in 1-D except the covariance of horizontal and vertical parameters. Now we focus on the covariance of x and y

$$\sum \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} = \sum \left( I_x + \frac{\partial n}{\partial x} \right) \left( I_y + \frac{\partial n}{\partial y} \right) = \sum \left( I_x + \frac{\partial n}{\partial x} \right) \left( I_y + \frac{\partial n}{\partial y} \right)$$  \hspace{1cm} (12)

The second and third term in the right of (12) are both zeros

$$E \left( \sum \frac{\partial n}{\partial x} I_x \right) = E \left( \sum \frac{\partial n}{\partial y} I_y \right) = 0$$

The last term is dependent to the noise variance, since that:

$$E \left( \sum \frac{\partial n}{\partial x} \frac{\partial n}{\partial y} \right) = \sum E \left( \frac{\partial n}{\partial x} \frac{\partial n}{\partial y} \right)$$

$$= \sum \left( n_i (x+1,y) - n_i (x,y) \right) \left( n_i (x,y+1) - n_i (x,y) \right) = 0$$

We obtain the fisher information matrix for image matching through simplifying (12) and introducing it in fisher information matrix,

$$FIM = \frac{1}{2\sigma^2} \left[ \sum I_x^2 \sum I_y^2 \sum I_x I_y \right] = \frac{1}{2\sigma^2} \left[ \sum \left( \frac{\partial I}{\partial x} \right)^2 + \sum \left( \frac{\partial I}{\partial y} \right)^2 - N \left[ \begin{array}{cc} 2 & 1 \\ 1 & 2 \end{array} \right] \right]$$  \hspace{1cm} (13)

We also conclude, the lower bound on the covariance of any position estimator in image matching is the inverse of the fisher information matrix as same as in the case for 1-D signal matching.

**Experimental Results**

In this section, we present experimental results to verify our previous theoretical analysis. The experiments are carried out on matlab2007. There are three measures in each experiment, real CRLB, numerical CRLB and simulation results. Real CRLB is computed from the noiseless signal. Numerical CRLB is computed with (1) and (9) or (13). Simulation results are obtained from Monte Carlo simulation with 10000 repeats using MLE.

Two different type signals of one dimension have been created with length of 256. One is the rectangular signal with 32 maximum magnitude sample points in the middle. The other is sinusoidal signal with cycle period of $32\pi$. AWGN with different variance is added to the signal. We adopt the definition of signal to noise ratio (SNR) defined in [2] in our experiments.

$$\text{SNR} = 10 \log \left[ \frac{\sum \left( \frac{\partial I}{\partial x} \right)^2}{\sum \left( \frac{\partial n}{\partial x} \right)^2} \right]$$

$$-10 \log \left[ \sum \left( \frac{\partial n}{\partial x} \right)^2 / 2\sigma^2 \right]$$  \hspace{1cm} (14)

Fig. 1(a) and (b) show the distribution of matching position for rectangular signals and sinusoidal signals, respectively. It can be seen that the histogram is of nearly Gaussian distribution. Fig. 2 shows the logarithm of variance that decreases as SNR increases. The three measures are very close for the two type signals. Simulation results verified that the CRLB is the tight performance bound of position estimation in 1-D. Furthermore, the curves in Fig. 2 show that the position performance is SNR dependent rather than signal dependent. The same conclusion can be obtained from fisher information matrix in (8).

Fig. 3 shows a tank image and its performance bound. Adding noise with standard deviation 4 to tank image, we compute ellipses that represent the covariance. We found that the results obtained by theoretical and simulated methods are very similar, that verifies the correctness of the theoretical
expressions. It can be seen that the represented ellipses have similar rotation with the target tank in the source image. This observation also verified the effectiveness of the derived theoretical expressions.

Fig. 4 shows an airport image and its performance bound. We choose one patch cropped from the source image to be reference image, and the source image is current image. The MLE method is used to compute the simulated results. As in Fig. 4, the theoretical and simulated results show they are almost same, that means the precision of theoretical analysis is accurate. There is more information on the direction of long axis than the direction of short axis. From the source image, we can achieve the same conclusion. On the parallel direction of runway, the images are more similar to each other, so the distribution of matching position is more divergent than the vertical direction. The ellipses parallel to the runway approximately in Fig. 4, that means the probability of matching error matching is higher on the direction parallel to runway than other directions.

Position estimator variances of Fig. 3 and Fig. 4 are listed in Table I. The real CRLB is very close to the simulation result. The numerical CRLB is tempestuously fluctuant when SNR is low. Therefore, to obtain more accurate performance, more images are needed under low SNR condition.

Table 1 The variance of position estimation

<table>
<thead>
<tr>
<th>Image</th>
<th>Real CRLB</th>
<th>Numerical</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[2.69 -1.36; -1.36 2.13]×10^{-4}</td>
<td>[2.66 -1.30; -1.30 2.08]×10^{-4}</td>
<td>[2.72 -1.42; -1.42 2.17]×10^{-4}</td>
</tr>
<tr>
<td></td>
<td>[2.52 1.32; 1.32 3.91]×10^{-3}</td>
<td>[2.70 1.50; 1.50 3.84]×10^{-3}</td>
<td>[2.52 1.33; 1.33 3.93]×10^{-3}</td>
</tr>
</tbody>
</table>

Conclusion

We have presented a method to estimate the performance bound of position estimation in image matching under noise environment. We use CRLB to characterize performance bound. Theoretical expressions have been derived in 1-D and 2-D, respectively. They have been verified by Monte Carlo simulations. The CRLB can characterize the performance bound of position estimation on image matching.

It is known that there are many algorithms for other transformations such as skew, affine, projection, and so on. This paper just considers position estimation. In order to obtain performance bounds of complex image matching algorithms, our future work will take complete transformations into account.

References


Figure 1. Matching position distribution of two type signals: (a) rectangular (b) sinusoidal

Figure 2. The variance of position estimation varies with SNR in 1-d: (a) rectangular (b) sinusoidal

Figure 3. The performance of position estimation for tank image: (a) source image (b) distribution of matching points (c) view in the direction of long axis (d) view in the direction of short axis (f) represented ellipses
Figure 4. The performance of position estimation for airport image: (a) source image (b) distribution of matching points (c) view in the direction of long axis (d) view in the direction of short axis (f) represented ellipse