

Performance of Optimal Registration Estimators

Tuan Q. Pham¹, Marijn Bezuijen¹, Lucas J. van Vliet¹, Klamer Schutte², Cris L. Luengo Hendriks³

¹Quantitative Imaging Group, Delft University of Technology, Lorentzweg 1, 2628 CJ, Delft, the Netherlands

²TNO Physics and Electronics Laboratory, P.O. Box 96864, 2509 JG, The Hague, the Netherlands

³Life Science Division, Lawrence Berkeley Lab, One Cyclotron Road, Berkeley, CA 94720, USA

ABSTRACT

This paper derives a theoretical limit for image registration and presents an iterative estimator that achieves the limit. The variance of any parametric registration is bounded by the Cramer-Rao bound (CRB). This bound is signal-dependent and is proportional to the variance of input noise. Since most available registration techniques are biased, they are not optimal. The bias, however, can be reduced to practically zero by an iterative gradient-based estimator. In the proximity of a solution, this estimator converges to the CRB with a quadratic rate. Images can be brought close to each other, thus speedup the registration process, by a coarse-to-fine multi-scale registration. The performance of iterative registration is finally shown to significantly increase image resolution from multiple low resolution images under translational motions.

Keywords: registration limit, optimal registration, Cramer-Rao bound, registration bias, super-resolution.

1. INTRODUCTION

Image registration is a fundamental task that plays a key role in many applications such as motion estimation,¹ panoramic image construction,¹⁰ super-resolution,⁴ ... It is generally accepted that registration performs better on some types of images than others and that noise degrades the registration accuracy in a linear fashion. However, not until Cramer-Rao analysis⁵ is applied to image registration that these observations are supported with a well-formulated theory. The Cramer-Rao bound (CRB) imposes a lower bound on the mean squared error for *any estimate* of a deterministic parameter. The CRB analysis in this paper is motivated from the analysis of shift estimation in⁹ and affine registration in.⁶ It is however extended to all parametric registration including but not limited to: translation and 2D projective registration. Common to all these parametric motion models, the CRB is signal-dependent and is proportional to the variance of input noise.

The paper also proposes optimal shift and projective estimators that achieve the Cramer-Rao bound. Since most available registration estimators are biased, they are not optimal. However, through an iterative process, the biases of the gradient-based shift estimator⁷ and 2D projective estimator can be removed. The iterative estimator is not only unbiased but is also optimal since their variances approach the CRB. Analysis of convergence rate shows that the error of the iterative shift estimator reduces to less than one-hundredth of a pixel after only two iterations. The Levenberg-Marquardt optimization¹⁰ for 2D projective registration also achieves this level of accuracy within 10 to 15 iterations under a coarse-to-fine multi-scale strategy.

The paper is organized as follows: section 2 derives the Cramer-Rao bound for all parametric registration with special cases for translational and 2D projective registration. Section 3 presents a gradient-based shift estimator as derived from a Taylor series expansion of the local signal. Biases of the gradient-based estimator due to Taylor series truncation and input noise are also characterized. Section 4 corrects the bias in an iterative manner and shows that the iterative gradient-based estimator is optimal. Section 5 finally presents a super-resolution application that uses the proposed registration estimators.

Corresponding author: Tuan Pham: E-mail: tuan@qi.tnw.tudelft.nl Phone: +31 15 278 6054 Fax: +31 15 278 6740

2. MINIMUM VARIANCE BOUND FOR IMAGE REGISTRATION

The mean squared error of any estimate of a deterministic parameter has a lower bound known as the Cramer-Rao Bound⁵ (CRB). Specifically, if a parameter vector $\mathbf{m} = [m_0 \ m_1 \ \dots \ m_n]^T$ is estimated from a given set of measurements, the CRB provides a bound on the error correlation matrix:

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] \geq \left(\mathbf{I} + \frac{\partial \mathbf{b}}{\partial \mathbf{m}} \right) \mathbf{F}^{-1}(\mathbf{m}) \left(\mathbf{I} + \frac{\partial \mathbf{b}}{\partial \mathbf{m}} \right)^T + \mathbf{b}\mathbf{b}^T \quad (1)$$

where $\boldsymbol{\varepsilon} = \hat{\mathbf{m}} - \mathbf{m}$ is the estimation error (the hat sign denotes an estimate of the underneath variable); $\mathbf{b} = E[\hat{\mathbf{m}}] - \mathbf{m}$ is the bias of the estimator ($E[\cdot]$ is the expectation of the enclosed expression); \mathbf{I} is the identity matrix; $\mathbf{F}(\mathbf{m})$ is the Fisher Information Matrix (FIM) that characterizes how well the unknown parameter vector \mathbf{m} can be estimated from the observed data. The \geq sign in (1) means that the difference between the left and the right matrices is non-negative definitive. As a result, the inequality holds for all diagonal terms.

If the estimator is unbiased (i.e. $\mathbf{b} = 0$), the expected variance of the parameters can be found directly from the main diagonal entries of the inverse matrix of the FIM:

$$E[(\hat{m}_i - m_i)^2] \geq \mathbf{F}_{ii}^{-1}(\mathbf{m}) \quad (2)$$

The Fisher information matrix is derived from the maximum likelihood principle. Let $\Pr(r|\mathbf{m})$ be the probability density function of an observed noisy data $r(\mathbf{m})$, the Fisher information matrix is a measure of the steepness of the likelihood function around its peak:

$$\mathbf{F}(\mathbf{m}) = E \left[\frac{\partial}{\partial \mathbf{m}} \log \Pr(r|\mathbf{m}) \right] \left[\frac{\partial}{\partial \mathbf{m}} \log \Pr(r|\mathbf{m}) \right]^T \quad (3)$$

Since the peak of a steep likelihood function is less sensitive to noise than that of a smooth one, the FIM characterizes how accurate \mathbf{m} can be estimated from the observed data.

2.1. Fisher information matrix for image registration

Image registration searches for a parametric transformation between two images' coordinates based on their intensity correlation. Assuming that both images I_1 and I_2 are corrupted from a noiseless scene I by zero mean Gaussian noise with variance σ^2 :

$$\begin{aligned} I_1(x, y) &= I(x, y) + n_1(x, y) \\ I_2(x, y) &= I(x', y') + n_2(x, y) \end{aligned} \quad (4)$$

where $x' = f(x, y, \mathbf{m})$ and $y' = g(x, y, \mathbf{m})$ are the coordinate transformation, and $\mathbf{m} = [m_1 \ m_2 \ \dots \ m_n]^T$ is the unknown registration parameter (for example, under translation $x' = x - v_x$, $y' = y - v_y$ and $\mathbf{m} = [v_x \ v_y]^T$). If the noises n_1 and n_2 are normally distributed over the registration region S , the total probability of the unknown scene I given an estimate of \mathbf{m} is:

$$\Pr(I|\hat{\mathbf{m}}) = \prod_S \frac{1}{\sigma\sqrt{2\pi}} \left(-\frac{(I_1 - I)^2}{2\sigma^2} \right) \prod_S \frac{1}{\sigma\sqrt{2\pi}} \left(-\frac{(I_2 - I')^2}{2\sigma^2} \right) \quad (5)$$

where the implicit coordinates for I_1 , I_2 and I is (x, y) except for $I' = I(x', y')$. The log-likelihood function therefore is:

$$\log \Pr(I|\hat{\mathbf{m}}) = -\frac{1}{2\sigma^2} \sum_S \left\{ (I_1 - I)^2 + (I_2 - I')^2 \right\} + const \quad (6)$$

From (3), the Fisher information matrix for a n -parameter vector \mathbf{m} is thus a $n \times n$ matrix \mathbf{F} with its entries computed as:

$$\begin{aligned} F_{ij} &= -E \left[\frac{\partial^2}{\partial m_i \partial m_j} \log \Pr(I|\hat{\mathbf{m}}) \right] = -E \left[\frac{\partial}{\partial m_i} \left(\frac{1}{\sigma^2} \sum_S (I_2 - I') \frac{\partial I'}{\partial m_j} \right) \right] \\ &= -E \left[\frac{1}{\sigma^2} \sum_S \left(n_2 \frac{\partial^2 I'}{\partial m_i \partial m_j} - \frac{\partial I'}{\partial m_i} \frac{\partial I'}{\partial m_j} \right) \right] = \frac{1}{\sigma^2} \sum_S \frac{\partial I'}{\partial m_i} \frac{\partial I'}{\partial m_j} \end{aligned} \quad (7)$$

where the derivative of the noiseless second image $I' = I(x', y')$ with respect to each unknown parameter m_i can be computed from the derivatives of the second image and the registration model:

$$\frac{\partial I'}{\partial m_i} = \frac{\partial I'}{\partial x'} \frac{\partial x'}{\partial m_i} + \frac{\partial I'}{\partial y'} \frac{\partial y'}{\partial m_i} \quad (8)$$

2.2. Cramer-Rao bound for 2D shift estimation

Using the general derivation of the Fisher information matrix in the previous subsection, the Cramer-Rao bound for any unbiased shift estimator can be derived. Two-dimensional shift estimation searches for a translational vector $\mathbf{v} = [v_x \ v_y]^T$ between two images: $x' = x - v_x$ and $y' = y - v_y$. The Fisher information matrix can be computed from (7) and (8)⁹:

$$\mathbf{F}(\mathbf{v}) = \frac{1}{\sigma^2} \begin{bmatrix} \sum_S I_x^2 & \sum_S I_x I_y \\ \sum_S I_x I_y & \sum_S I_y^2 \end{bmatrix} \quad \text{where } I_x = \frac{\partial I}{\partial x} \text{ and } I_y = \frac{\partial I}{\partial y} \quad (9)$$

According to (2), the Cramer-Rao bounds of the variances of the estimated shifts are:

$$\text{var}(v_x) \geq \mathbf{F}_{11}^{-1} = \sigma^2 \sum_S I_y^2 / \text{Det}(\mathbf{F}) \quad \text{var}(v_y) \geq \mathbf{F}_{22}^{-1} = \sigma^2 \sum_S I_x^2 / \text{Det}(\mathbf{F}) \quad (10)$$

where $\text{Det}(\mathbf{F}) = \sum_S I_x^2 \sum_S I_y^2 - \left(\sum_S I_x I_y \right)^2$ is the determinant of the FIM. The inequality in (10) is stricter than the one given in,⁹ which ignores the second term of $\text{Det}(\mathbf{F})$ to arrive at a simpler CRB:

$$\text{var}(v_x) \geq \sigma^2 / \sum_S I_x^2 \quad \text{var}(v_y) \geq \sigma^2 / \sum_S I_y^2 \quad (11)$$

The simplified CRB in (11) clearly shows that the shift variance is linearly proportional to the input noise variance σ^2 and inversely proportional to the total gradient energy in the shift direction. As a result, big image patches with high-gradient details and low noise are likely to result in accurate shift estimation. However, the equality of the loose bound in (11) is hardly achievable (since $(\sum I_x I_y)^2$ is set to 0 in⁹). As a result, the true CRB in (10) should be used instead.

Note that the CRB characterizes the shift variances based on an uncorrupted signal I , which is not available in practical situations. However, if the region of interest S is large, the total gradient energies of I can be computed from that of the corrupted image I_1 and an instance of noise with the same variance:

$$\sum_S I_x^2 \approx \sum_S \left(\frac{\partial I_1}{\partial x} \right)^2 - \sum_S \left(\frac{\partial n_1}{\partial x} \right)^2 \quad \sum_S I_y^2 \approx \sum_S \left(\frac{\partial I_1}{\partial y} \right)^2 - \sum_S \left(\frac{\partial n_1}{\partial y} \right)^2 \quad \sum_S I_x I_y \approx \sum_S \frac{\partial I_1}{\partial x} \frac{\partial I_1}{\partial y} \quad (12)$$

2.3. Cramer-Rao bound for 2D projective registration

Similar to shift estimation, the Cramer-Rao bound is applicable to other motion models such as 2D projective transformation (which includes translation, Euclidean, similarity, and affine transformations). The bound is also computed from the Fisher information matrix as done with shift estimation. Planar projective registration looks for an 8-parameter vector $\mathbf{m} = [m_1 \ m_2 \ \dots \ m_8]^T$ that transforms the two images' coordinates by:

$$x' = \frac{m_1 x + m_2 y + m_3}{m_7 x + m_8 y + 1} \quad y' = \frac{m_4 x + m_5 y + m_6}{m_7 x + m_8 y + 1} \quad (13)$$

Since $\begin{bmatrix} \frac{\partial x'}{\partial \mathbf{m}} & \frac{\partial y'}{\partial \mathbf{m}} \end{bmatrix}^T = \frac{1}{D} \begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -x'x & -x'y \\ 0 & 0 & 0 & x & y & 1 & -y'x & -y'y \end{bmatrix}$ where $D = m_7 x + m_8 y + 1$, (8) results in:

$$\begin{bmatrix} \frac{\partial I'}{\partial \mathbf{m}} \end{bmatrix}^T = \frac{1}{D} \begin{bmatrix} x \frac{\partial I'}{\partial x'} & y \frac{\partial I'}{\partial x'} & \frac{\partial I'}{\partial x'} & x \frac{\partial I'}{\partial y'} & y \frac{\partial I'}{\partial y'} & \frac{\partial I'}{\partial y'} & -x(x' \frac{\partial I'}{\partial x'} + y' \frac{\partial I'}{\partial y'}) & -y(x' \frac{\partial I'}{\partial x'} + y' \frac{\partial I'}{\partial y'}) \end{bmatrix} \quad (14)$$

The 8×8 Fisher information matrix is then derivable from (7) and (14). Due to a complex 8×8 matrix inversion, the exact formula for Cramer-Rao bound is not given here. However, similar to shift estimation, the variances of the projective registration parameters are proportional to the input noise variance and inversely proportional to the total gradient energy.

3. BIAS OF GRADIENT-BASED SHIFT ESTIMATOR

The Cramer-Rao bound presented in the previous section is applicable to unbiased shift estimators. Biased estimators therefore have an even higher variance bound. While most current shift estimators are biased, the biases are not easily characterizable.⁹ In this section, we present a bias analysis of the one-dimensional gradient-based shift estimator. Bias for 2D shift estimator is similar and is included in the Appendix. The original work in² is presented entirely in the spatial domain as opposed to a mix spatial-frequency analysis in.⁹

According to the Taylor series expansion, a continuous signal can be approximated from a close-by signal by a polynomial expansion. This approximation can be used to relate two slightly shifted signals:

$$s_2(x) = s_1(x + v_x) = s_1(x) + v_x \frac{\partial s_1}{\partial x} + \frac{1}{2!} v_x^2 \frac{\partial^2 s_1}{\partial x^2} + \frac{1}{3!} v_x^3 \frac{\partial^3 s_1}{\partial x^3} + \dots \quad (15)$$

where the 1D signal s_2 equals to a shifted version of s_1 by a small displacement v_x . Because v_x is small, high powers of v_x in (15) are negligible. s_2 can thus be approximated by truncating the Taylor series after the first order derivative. The displacement v_x can then be solved from a set of linear equations by minimizing the Mean Square Error (MSE) over N samples in a supported region S :

$$MSE = \frac{1}{N} \sum_S \left(s_2(x) - s_1(x) - v_x \frac{\partial s_1}{\partial x} \right)^2 \quad (16)$$

which yields a least squares solution⁷:

$$v_x^{LQ} = \frac{\sum_S (s_2(x) - s_1(x)) \frac{\partial s_1}{\partial x}}{\sum_S \left(\frac{\partial s_1}{\partial x} \right)^2} \quad (17)$$

The approximation is only accurate when the second and higher derivatives in (15) are small. As a result, only displacements smaller than the derivative scale ($\sigma_s \approx 1$ pixel for discrete signals) can be estimated reliably using the Taylor expansion method. Signals with displacements larger than one pixel should be aligned to integer accuracy first using other methods such as cross-correlation. A multi-resolution Taylor shift estimation that estimates and refines the displacements from coarse to fine scales could also be used in such cases.

One problem remains with the Taylor expansion method is that the shift is estimated from a set of approximations, not equations. The solution (17) is also derived for noise-free signals only. As a result, there is a systematic bias that depends on the image content and the displacement itself. In fact, the bias is a combined effect of multiple modeling errors such as truncation of Taylor series and intensity noise. The following subsections identify these biases to serve as a starting point for bias correction.

3.1. Bias due to truncation of Taylor series

One most noticeable displacement bias is due to the truncation error of the Taylor series in (15):

$$\varepsilon = \frac{1}{2!} v_x^2 \frac{\partial^2 s_1}{\partial x^2} + \frac{1}{3!} v_x^3 \frac{\partial^3 s_1}{\partial x^3} + \dots \quad (18)$$

Putting $s_2(x) - s_1(x) = v_x \frac{\partial s_1}{\partial x} + \varepsilon$ into the least-squares estimate in (17), the truncation bias can be computed as:

$$bias_T = v_x^{LQ} - v_x = \frac{\sum_S \varepsilon \frac{\partial s_1}{\partial x}}{\sum_S \left(\frac{\partial s_1}{\partial x} \right)^2} = \frac{1}{3!} v_x^3 \frac{\sum_S \frac{\partial^3 s_1}{\partial x^3} \frac{\partial s_1}{\partial x}}{\sum_S \left(\frac{\partial s_1}{\partial x} \right)^2} + \frac{1}{5!} v_x^5 \frac{\sum_S \frac{\partial^5 s_1}{\partial x^5} \frac{\partial s_1}{\partial x}}{\sum_S \left(\frac{\partial s_1}{\partial x} \right)^2} + \dots \quad (19)$$

where the even-order terms vanish due to Parseval's theorem⁸:

$$\sum_S \frac{\partial^{2n} s}{\partial x^{2n}} \frac{\partial s}{\partial x} \approx \int_{-\infty}^{+\infty} f g^* dx = \frac{1}{2\pi} \int_{-\pi}^{+\pi} F G^* d\omega = \frac{j(-1)^n}{2\pi} \int_{-\pi}^{+\pi} \omega^{2n+1} |S(\omega)|^2 d\omega = 0 \quad (20)$$

where

$$\begin{aligned} s(x) &\Leftrightarrow S(\omega) \\ f(x) = \frac{\partial s(x)}{\partial x} &\Leftrightarrow F(\omega) = j\omega S(\omega) \\ g(x) = \frac{\partial^{2n} s(x)}{\partial x^{2n}} &\Leftrightarrow G(\omega) = (-1)^n \omega^{2n} S(\omega) \end{aligned} \quad (21)$$

3.2. Bias due to noise

The formula for the truncation bias is applicable to shift estimation between two noiseless images. In reality, noise is always present and is a source of much more significant registration error. We assume additive Gaussian noise, uncorrelated with the original image and uncorrelated between realizations:

$$s_1^*(x) = s_1(x) + n_1(x) \quad s_2^*(x) = s_1(x + v_x) + n_2(x) \quad (22)$$

Under Gaussian noise, shift estimation by minimizing the MSE in (16) still holds, so does the solution (17) with an exception that the image gradient must be derived from the noiseless signal s_1 . Since s_1 is unavailable, the noisy signal s_1^* is used in (17). Although noise does not affect the truncation bias, it introduces another bias to the estimated shift:

$$bias_N = v_x^* - v_x = \frac{\sum_S (s_2^*(x) - s_1^*(x)) \frac{\partial s_1^*}{\partial x}}{\sum_S \left(\frac{\partial s_1^*}{\partial x} \right)^2} - v_x = \frac{\sum_S (s_2(x) - s_1(x)) \frac{\partial s_1}{\partial x}}{\sum_S \left(\frac{\partial s_1}{\partial x} \right)^2 + \sum_S \left(\frac{\partial n_1}{\partial x} \right)^2} - v_x = \frac{N}{S + N} v_x \quad (23)$$

where v_x^* is the estimated shift computed from the noisy signals s_i^* , $S = \sum \left(\frac{\partial s_1}{\partial x} \right)^2$ and $N = \sum \left(\frac{\partial n_1}{\partial x} \right)^2$ are gradient energies of the signal and noise respectively. It is interesting to see that the Taylor shift estimator always underestimates the displacement in the presence of noise. The bias due to noise is proportional to the shift amount. As a result, it is a more severe bias source than the Taylor series truncation, which is proportional to the cube of the shift.

4. OPTIMAL REGISTRATION ESTIMATORS

When an estimator reaches the Cramer-Rao bound, it is said to be optimal. Least-squares estimator is often optimal when it is unbiased. Although the gradient-based shift estimator is biased, its bias can be corrected in an iterative way. The resulted shift estimator is therefore optimal. Experimental results will also be carried out in this section to validate the claim.

4.1. Iterative gradient-based shift estimator

From (19) and (23), the total bias of the gradient-based shift estimator can be written as:

$$bias = bias_N + bias_T = v_x \frac{\sum_S \left(\frac{\partial n_1}{\partial x} \right)^2}{\sum_S \left(\frac{\partial s_1}{\partial x} \right)^2 + \sum_S \left(\frac{\partial n_1}{\partial x} \right)^2} + \frac{1}{3!} v_x^3 \frac{\sum_S \frac{\partial^3 s_1}{\partial x^3} \frac{\partial s_1}{\partial x}}{\sum_S \left(\frac{\partial s_1}{\partial x} \right)^2} + \frac{1}{5!} v_x^5 \frac{\sum_S \frac{\partial^5 s_1}{\partial x^5} \frac{\partial s_1}{\partial x}}{\sum_S \left(\frac{\partial s_1}{\partial x} \right)^2} + \dots \quad (24)$$

where the first term is due to noise and all other terms are due to truncation error.

Although the biases due to Taylor truncation and noise can be derived analytically, we do not favor a direct bias correction. One reason is due to an inaccurate initial shift estimation. Since the bias is derived from this initial estimate, the correction is also not accurate. Another reason is that formula (24) requires the derivatives of the uncorrupted signal s_1 , which is not available. Besides, the bias is derived from many assumptions such as an infinitely large region of interest or Gaussian noise that are not correct in the first place. As a result, we prefer an indirect bias correction method. (24) reveals that the relative bias $bias/v_x \approx N/(S + N)$ is positive and smaller than one. The bias is therefore guaranteed to reduce at an exponential rate if shift estimation is performed iteratively on a translation-corrected image pair. Since the bias can be reduced to practically zero, the iterative shift estimation converges to an unbiased solution.

The algorithm for iterative shift estimation is described as follows (integer displacement should be corrected first since this algorithm only handles sub-pixel shift):

1. Initialize a counter $i = 0$, a temporary signal $s_2(0) = s_2$, and an accumulated displacement v_x to 0
2. Perform a Taylor shift estimation on s_1 and $s_2(i)$ that results in a displacement $v_x(i)$
3. Accumulate the latest resolved displacement: $v_x = v_x + v_x(i)$
4. Translate the second signal by the accumulated displacement: $s_2(i + 1) = shift(s_2, v_x)$
5. Exit loop if the incremented displacement $v_x(i)$ is within a desirable accuracy or if a maximum number of iterations is reached, otherwise increment the counter $i = i + 1$ and go back to step 2.

The iterative algorithm is guaranteed to terminate, either due to accuracy condition or maximum number of iterations, whichever is reached first. With practical SNRs ($SNR = 10\log_{10}(S/N)$), the algorithm converges so fast that the registration bias reduces to an order of $O(1e-3)$ after only 2 iterations:

$$\begin{aligned} bias(1^{st} \text{ iter}) &\approx \frac{N}{S+N} v_x \leq \frac{1}{11} \times 0.5 \approx 4.5e-2 \\ bias(2^{nd} \text{ iter}) &\approx \frac{N}{S+N} bias(1^{st}) \approx \left(\frac{N}{S+N}\right)^2 v_x \leq \left(\frac{1}{11}\right)^2 \times 0.5 \approx 4.1e-3 \end{aligned} \quad \text{for } SNR \geq 10dB \quad (25)$$

The iterative solution is also very efficient since the image derivatives in (17) are reusable across iterations. In fact, an extra iteration of shift estimation requires less computation than the bias correction method.²

4.2. Normal distribution and optimality

To examine the influence of noise on the variance of the iterative shift estimator, we apply shift estimation to a shifted pair of 8-bit 256×256 images. The $[0.5 \ 0.5]$ -shifted pair is generated from a large Pentagon image (1028×1028) by Gaussian smoothing ($\sigma = 2$) followed by 4-times subsampling. Gaussian noise of the same variance is added to both images before the estimation. For each noise level, 100 noise generalizations are tested and the estimated shifts are plotted in figure 1a. The point clusters around $[0.5 \ 0.5]$ suggests that the estimated shift is unbiased. However, the estimated shifts are not exactly $[0.5 \ 0.5]$ because of a stochastic variation that is proportional to the variance of input noise.

The variation of the estimated shift due to input noise is explainable from the solution of the 2D gradient-based shift estimation (derived in Appendix A):

$$\begin{aligned} v_x &= \left(\sum_S I_x I_t \sum_S I_y^2 - \sum_S I_y I_t \sum_S I_x I_y \right) / Det \\ v_y &= \left(\sum_S I_y I_t \sum_S I_x^2 - \sum_S I_x I_t \sum_S I_x I_y \right) / Det \end{aligned} \quad (26)$$

where $\mathbf{v}=[v_x \ v_y]^T$ is the displacement of an image I_2 against a reference image I_1 . $Det = \sum_S I_x^2 \sum_S I_y^2 - \left(\sum_S I_x I_y\right)^2$ is the determinant of a Hessian matrix that characterize the reference image I_1 (same matrix in (9)) and I_x, I_y, I_t are shortcuts for $\partial I_1/\partial x, \partial I_1/\partial y, I_2 - I_1$, respectively. Since the iterative solution converges, the precision of the overall shift equals to that of the final estimation. At the final iteration, the estimated shift is already quite accurate such that the translated image I_2 only differs from the reference frame I_1 by a noise term: $I_t = I_2 - I_1 \approx n_2 - n_1$. Since I_t is pure noise, $E[\sum I_x I_t] = E[\sum I_y I_t] = 0$ and v_x, v_y are therefore normally distributed around zero.

The precision of the iterative shift estimation is compared against the Cramer-Rao bound in figure 1b. The continuous line represents the theoretical CRB, the dash lines plots the standard deviation of v_x over 100 trials (the error bar is constructed from 10 of such $std(v_x)$). The plot of precision for v_y is similar and is therefore omitted. It can be seen that the shift estimate could even get a slightly smaller variance than the CRB for practical SNRs ($SNR \geq 0$ or $\sigma_{noise} \leq 20$). This is because the CRB is computed from noisy image derivatives (equation 10), which must be estimated at a certain scale. Also drawn in figure 1b is a dash-dot line, which show a better shift variance at high noise thanks to signal pre-smoothing. As can be seen, the variance of shift estimation without image pre-smoothing (dash line) soars above the CRB for noise with $\sigma_{noise} \geq 20$ ($SNR \leq 0$). Gaussian pre-smoothing with an adaptive $\sigma_s = \sigma_{noise}/20$ really matches the shift standard deviation curve (dash-dot line) with the optimal bound.

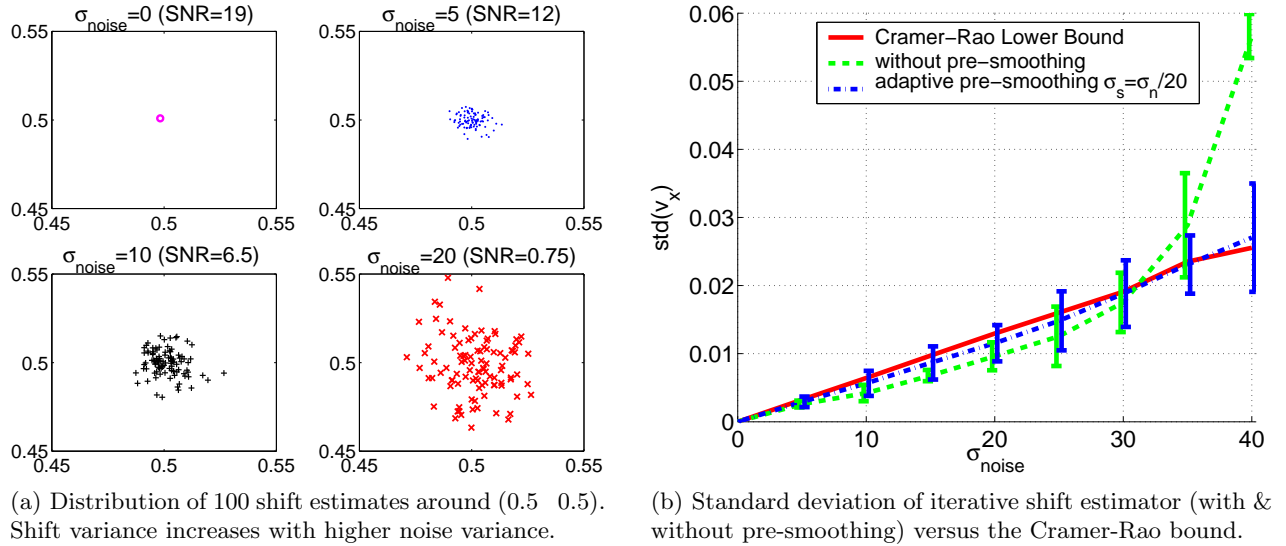


Figure 1. Influence of Gaussian noise on the precision of the iterative gradient-based 2D shift estimator

4.3. Optimal 2D projective registration

The presented iterative shift estimation has actually been used by many^{4,7} without knowing its optimality. In fact, the iterative solution is a gradient descent process that converges to a local minimum. Similarly, 2D projective registration is solved by an iterative Levenberg-Marquardt optimization process.¹⁰ At each iteration, the registration parameters \mathbf{m} (see section 2.3) are incremented by a small amount computed from a linear system whose matrix is identical to the the 8×8 Fisher information matrix derived in (7) and (14). The similar gradient descent nature suggests that the Levenberg-Marquardt 2D projective registration estimator is also optimal. This is verified in an experiment whose results are shown in figure 2.

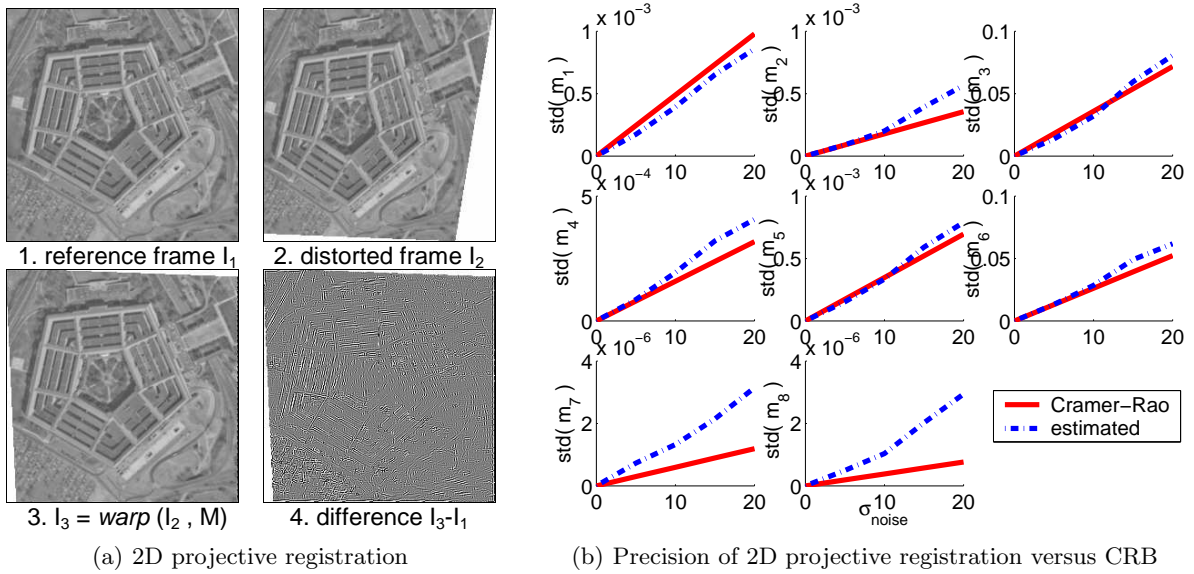


Figure 2. Precision of the gradient-based 2D projective estimator using Levenberg-Marquardt optimization.

In this experiment, a 256×256 reference image (figure 2a-1) is projectively distorted by a random 3×3 homography matrix (equation 27), resulting in image 2a-2. The registration matrix M is estimated directly

from the two input images. The distorted image is then warped onto the first image, resulting in image 2a-3. Note that the registration is estimated correctly since image 2a-3 looks very similar to the reference image. The difference between the reference and the warped image in figure 2a-4 also shows no perceptible difference.

$$M = \begin{bmatrix} 1.087094 & 0.041690 & -0.897298 \\ 0.039365 & 0.911578 & -1.471319 \\ 0.000313 & -0.000490 & 1.000000 \end{bmatrix} \quad (27)$$

By adding different noise generalizations to input images 2a-1 and 2a-2 before estimating registration, we perform the same variance analysis as done in the previous subsection for shift estimation. Figure 2b plots the precision of the estimated parameters for different input noise levels. The estimated precision (dash-dot line) matches the Cramer-Rao bound (continuous line), which implies the optimality of the Levenberg-Marquardt registration.

Since eight parameters are involved in the optimization process, the 2D projective registration requires many more iterations than the simple shift estimation. To speedup the process and reduce the chance of convergence to a wrong local minimum, a multi-scale strategy is used. The registration is first estimated at a coarse scale, the registration parameters are then used as initialization to another estimation at a finer scale. With a three-scale scheme (Gaussian smoothing $\sigma = 4, 1, 0$), it takes 10 to 15 iterations for the Levenberg-Marquardt registration to convergence.

4.4. Beating the limit of registration

The Cramer-Rao bound dictates that there is a fundamental limit to the precision of any image registration. The CRB derived in this paper, however, only applies to pairwise registration of noisy images. One way to improve the precision of registration is to register a new image against a high-resolution mosaic reconstructed from previous images. Since the mosaic image is of higher resolution and is less noisy, its gradient information is closer to the noiseless ground truth. The gradient-based solution computed from these gradients is therefore more reliable. Another way to improve the performance of registration is to involve multiple images in a single estimation. Bundle adjustment³ ensures registration consistency amongst several frames to reduce the registration error due to noise of a single pairwise estimation. The shift estimated in the next section, for example, is the average result of a forward and a backward shift estimation. Unfortunately, there is also a limit in the number of images used in bundle adjustment. Due to many factors such as reducing overlapping areas, motion in the scene, perspective or lens distortion, registration consistency might not hold for all images in a sequence.

5. SUPER-RESOLUTION APPLICATION

Registration with sub-pixel accuracy is desirable in many applications such as image mosaicking and Super-Resolution (SR). In fact, with enough overlapping frames, SR can be done simultaneously with image mosaicking. Figure 4 shows such an example of constructing a super-resolved panorama view from a shifted image sequence. The wide view of a highway in figure 4d is reconstructed from 200 128×128 Low-Resolution (LR) frames. The LR frames are captured by a panning camera. Frames number 0, 122 and 199 can be seen in figure 4a-c. The camera also slightly tilts over 200 frames to cause sub-pixel shifts in the vertical direction for SR. Since the scene is captured by a panning camera from afar, motion between frames are roughly translational. The displacement between every adjacent frames are computed in both directions and the results are averaged to obtain a better shift estimation. New frame is fused onto an existing mosaic by nearest neighbor pixel stamping. Despite many poor data conditions (small frame size, low contrast, high noise, intensity flickering, moving cars and human, dead sensors, ...) the estimated shifts are quite good and the reconstructed mosaic shows no blurring or jitter artifact due to mis-registration.

The result of two-time SR can be compared against a single LR input in figure 3a-b. It is clear that the SR image has much lower noise and higher dynamic range. These greatly improve the visibility of the scene. For example, the hexagonal pattern on the concrete structure along the highway is only visible in the SR image. Smaller details and reduced aliasing are also observable in a zoomed-in version of the SR image (figure 3d compared to the same LR region in figure 3c). Notice that the fence is now visible on both side of the highway crossing. Tree branches are also distinguishable from the background in the SR image.

6. CONCLUSION

In summary, we have derived a Cramer-Rao lower bound for any unbiased parametric image registration. The minimum variance bound of shift estimation is proportional to input noise variance and inversely proportional to the total gradient energy in the estimation direction. Although the gradient-based shift estimator is biased, the bias can be eliminated by an iterative bias correction process. This iterative shift estimator is optimal in a sense that its variance approaches the Cramer-Rao lower bound. The iterative shift estimator is also very efficient, requiring only two iterations to achieve an accuracy of one-hundredth of a pixel for 256×256 images at normal signal-to-noise ratio ($SNR > 10dB$). Similar to shift estimation, iterative 2D projective registration also reaches the Cramer-Rao bound after 10 to 15 iterations of Levenberg-Marquardt optimization. Finally, registration with sub-pixel accuracy has been applied to the application of image mosaicking and super-resolution of distant surveillance videos.

APPENDIX A. TWO-DIMENSIONAL SHIFT ESTIMATION

A.1. Gradient-based 2D shift estimation

We start off with a local Taylor expansion for a two-dimensional signal:

$$I_2 = I_1 + v_x I_x + v_y I_y + \varepsilon \quad (28)$$

where image I_2 is a shifted version of image I_1 by a small displacement (v_x, v_y) , $I_t = I_2 - I_1$ is the discrete temporal derivative, I_x and I_y are image derivatives in x and y dimension respectively, and ε is the truncation error:

$$\varepsilon = \frac{1}{2}v_x^2 I_{xx} + v_x v_y I_{xy} + \frac{1}{2}v_y^2 I_{yy} + \frac{1}{6}v_x^3 I_{xxx} + \frac{1}{2}v_x^2 v_y I_{xxy} + \frac{1}{2}v_x v_y^2 I_{xyy} + \frac{1}{6}v_y^3 I_{yyy} + \dots \quad (29)$$

where the number of occurrences of x and y in the suffix of I determine the order of derivative in that dimension. Minimizing the MSE for v_x and v_y results in a set of equations in matrix form:

$$\begin{bmatrix} \sum_S I_x^2 & \sum_S I_x I_y \\ \sum_S I_x I_y & \sum_S I_y^2 \end{bmatrix} \cdot \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} \sum_S I_x I_t \\ \sum_S I_y I_t \end{bmatrix} \quad (30)$$

which gives solution:

$$\begin{aligned} v_x &= \frac{1}{Det} \left(\sum_S I_x I_t \sum_S I_y^2 - \sum_S I_y I_t \sum_S I_x I_y \right) \\ v_y &= \frac{1}{Det} \left(\sum_S I_y I_t \sum_S I_x^2 - \sum_S I_x I_t \sum_S I_x I_y \right) \end{aligned} \quad (31)$$

where $Det = \sum_S I_x^2 \sum_S I_y^2 - \left(\sum_S I_x I_y \right)^2$ is the determinant of the Hessian matrix in (30).

A.2. 2D shift bias

The truncation bias for two-dimensional Taylor shift estimator can be derived analytically by putting $I_t = v_x I_x + v_y I_y + \varepsilon$ to (31):

$$\begin{aligned} bias_x^T &= \frac{1}{Det} \left(\sum_S \varepsilon I_x \sum_S I_y^2 - \sum_S \varepsilon I_y \sum_S I_x I_y \right) \\ bias_y^T &= \frac{1}{Det} \left(\sum_S \varepsilon I_y \sum_S I_x^2 - \sum_S \varepsilon I_x \sum_S I_x I_y \right) \end{aligned} \quad (32)$$

where the even-order terms in the expression of $\sum_S \varepsilon I_x$ and $\sum_S \varepsilon I_y$ vanish due to Parseval's theorem: (formulae for $\sum_S \varepsilon I_y$ is similar and is therefore omitted):

$$\sum_S \varepsilon I_x = \frac{1}{6}v_x^3 \sum_S I_{xxx} I_x + \frac{1}{2}v_x^2 v_y \sum_S I_{xxy} I_x + \frac{1}{2}v_x v_y^2 \sum_S I_{xyy} I_x + \frac{1}{6}v_y^3 \sum_S I_{yyy} I_x + \dots \quad (33)$$

The bias due to noise is more complicated for the two-dimensional case (shown here for $bias_x^N$ only):

$$bias_x^N = v_x^* - v_x = \frac{\sum_S I_x I_t \left(\sum_S I_y^2 + \sum_S n_y^2 \right) - \sum_S I_y I_t \sum_S I_x I_y}{\left(\sum_S I_x^2 + \sum_S n_x^2 \right) \left(\sum_S I_y^2 + \sum_S n_y^2 \right) - \left(\sum_S I_x I_y \right)^2} = \frac{\sum_S I_x I_t \sum_S n_y^2}{Det + N} + v_x \frac{N}{Det + N} \quad (34)$$

where v_x^* is the estimated shift computed from (31) using derivatives of the noisy images $I_1^* = I_1 + n_1$; n_x and n_y are derivatives of the noise n_1 in x and y dimension respectively. N is the over-estimation of the Hessian matrix's determinant if the gradient information of the noisy image is used:

$$N = \sum_S n_x^2 \sum_S I_y^2 + \sum_S n_y^2 \sum_S I_x^2 + \sum_S n_x^2 \sum_S n_y^2 \quad (35)$$

A.3. Aperture problem and convergence

Accurate shift estimation is not always possible using a computational approach due to a phenomenon called *the aperture problem*.¹ In the aperture problem, the solution of shift estimation is not unique or is undefined due to a homogeneous structure in the signals. For example, a translation is undefined for 1D signals if the signal is constant over the estimated region. This causes a division-by-zero error in solution (17) for 1D shift estimation. In 2D cases, solution (31) can also be undefined if the determinant of the Hessian matrix in (30) equals zero. This happens when I_y/I_x is constant over the supported region. In other words, the aperture problem occurs for 2D signals when there is a single orientation over the entire neighborhood.

Like all other shift estimators, the iterative gradient-based shift estimator fails under the aperture problem. However, it is also unstable for conditions close to the aperture problem. Under low SNR and high orientation congruency, the absolute relative bias may get larger than one, which prevents the convergence of the iterative solution:

$$\frac{bias_x}{v_x} = \frac{1}{v_x} \frac{\sum_S I_x I_t \sum_S n_y^2}{Det + N} + \frac{N}{Det + N} + \frac{av_x^2 + bv_x v_y + cv_y^2 + d \frac{v_y^3}{v_x} + \dots}{Det} \quad (36)$$

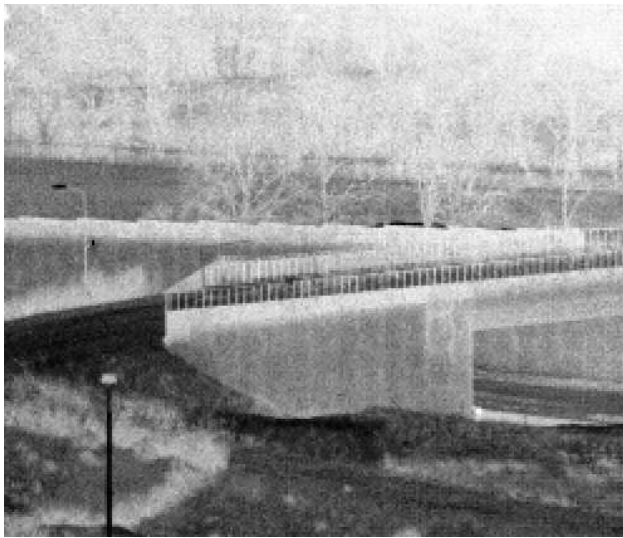
If noise level is high or local orientation is uniform, $N \gg Det$ and the second term in (36) is close to one. As soon as $|v_x|$ is smaller than a certain threshold while $|v_y|$ is still high, the $1/v_x$ and v_y^3/v_x terms might become large enough to amplify the absolute relative bias to more than one. This stops the convergence of the 2-D iterative shift estimator and the solution thus ends up far away from the ground truth. However, this is not surprising at all because the Cramer-Rao bound (equation 10) is then also very high.

There is a simple solution to alleviate this instability problem: pre-smoothing of images before shift estimation.⁹ Pre-smoothing causes an attenuation of all gradient-dependent terms in (36). Since N reduces much faster than Det , the scalar term in (36) becomes much smaller than one and so do other terms in the formulae. The relative bias is then more likely to stay well below one, and the iterative solution would thus converge. As a more accurate registration is obtained, the $1/v_x$ term in (36) will not cause an overshoot because $\sum I_x I_t$ approaches zero due to the pure noise in $I_t \approx n_2 - n_1$.

REFERENCES

1. J. Barron, D. Fleet and S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.
2. M. Bezuijen. *Improved sub-pixel shift estimation in 2D gray-value images*. Bachelor Thesis, Pattern Recognition Group, Delft University of Technology, the Netherlands, 2003.
3. D. Capel and A. Zisserman. Computer vision applied to super resolution. *IEEE Signal Processing Magazine*, May 2003. pp. 75–86.
4. M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53(3):231–239, 1991.
5. S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

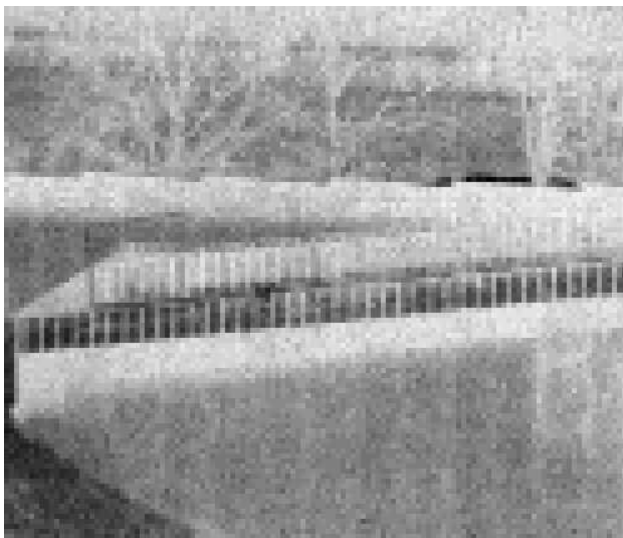
6. W. Li and H. Leung. A Maximum Likelihood Approach for Image Registration Using Control Point and Intensity. In *IEEE Trans. Image Processing*, 13(8):1115–11127, 2004.
7. B.D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proc. of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, 1981. pp. 674–679.
8. A.V. Oppenheim, A.S. Willsky and I.T. Young. *Signals and Systems*. Prentice-Hall, 1983.
9. M.D. Robinson and P. Milanfar. Fundamental Performance Limits in Image Registration. In *IEEE Trans. Image Processing*, 13(9):1185–1199, 2004.
10. R. Szeliski. Video mosaics for virtual environments. In *IEEE Computer Graphics and Applications*, 16(2):22–30, 1996
11. K. Zuiderveld. Contrast Limited Adaptive Histogram Equalization in P.S. Heckbert (ed), *Graphics Gems IV*. Boston Academic Press, 1994. pp. 474–485.



(a) a region of interest in LR frame no. 183



(b) two-time SR from a shifted sequence

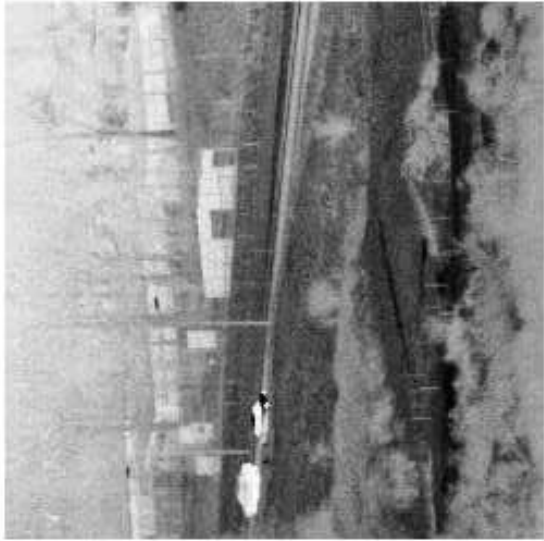


(c) zoom-in of the LR image

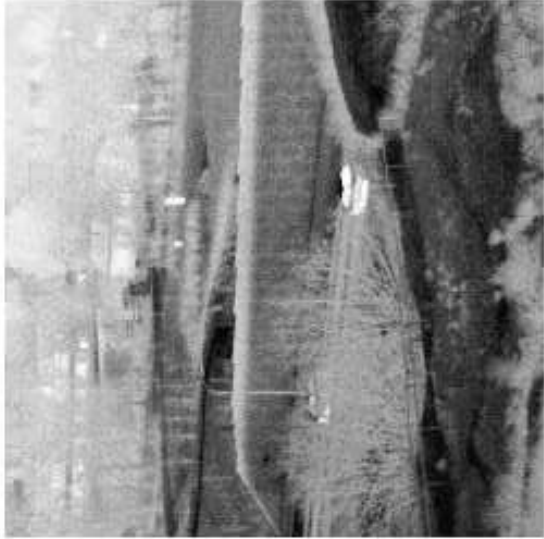


(d) zoom-in of the SR image

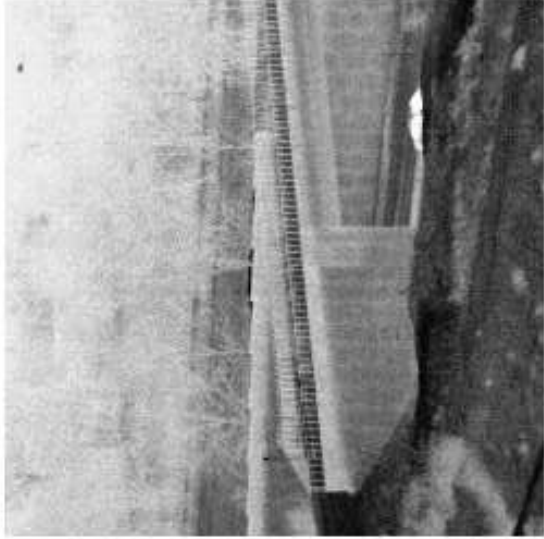
Figure 3. Two-time super-resolution image versus a low-resolution image. All images are stretched.¹¹



(a) LR frame no. 0



(b) LR frame no. 122



(c) LR frame no. 199



(d) $2\times$ super-resolved mosaic view from 200 LR frames. Ghost effect on the highway is due to moving cars.

Figure 4. Super-resolution of an image sequence captured by a pan-and-tilt infra-red camera. All images are stretched.¹¹