Bootstrap Resampling for Image Registration Uncertainty Estimation Without Ground Truth

Jan Kybic, *Senior Member, IEEE*

***Abstract*—We address the problem of estimating the uncertainty of pixel based image registration algorithms, given just the two images to be registered, for cases when no ground truth data is available. Our novel method uses bootstrap resampling. It is very general, applicable to almost any registration method based on minimizing a pixel-based similarity criterion; we demonstrate it using the SSD, SAD, correlation, and mutual information criteria. We show experimentally that the bootstrap method provides better estimates of the registration accuracy than the state-of-the-art Cramér–Rao bound method. Additionally, we evaluate also a fast registration accuracy estimation (FRAE) method which is based on quadratic sensitivity analysis ideas and has a negligible computational overhead. FRAE mostly works better than the Cramér–Rao bound method but is outperformed by the bootstrap method.**

***Index Terms*—Accuracy estimation, bootstrap, Cramér–Rao bound, image registration, motion estimation, performance limits, uncertainty estimation.**

I. Introduction

IMAGE registration [1], [2] finds a geometric transformation relating coordinates of corresponding points in two given images. Image registration is used for motion analysis, video compression and coding, object tracking, image stabilization, segmentation, stereo reconstruction, and super-resolution [3]. Biomedical applications [4]–[8] include intrasubject, intersubject, and intermodality analysis, registration with atlases, quantification and qualification of feature shapes and sizes, elastography, distortion compensation, motion detection and compensation.

Most image registration algorithms return just a single, deterministic answer, a point-wise estimate of the unknown geometric transformation. However, in practice, there is always some associated uncertainty, the registration accuracy is limited. Knowing this uncertainty is useful to determine whether and to what extent the registration results can be trusted and whether the input data is suitable. It can be used to give more weight to more reliable image pairs or spatial locations, for example, in sequence registration, group-wise registration, flow-inpainting, or recovering elastography parameters from the displacement.

Manuscript received August 12, 2008; revised July 27, 2009. First published August 25, 2009; current version published December 16, 2009. This work was supported by the Czech Ministery of Education under Project 1M0567. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pier Luigi Dragotti.

The author is with the Center for Applied Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic (e-mail: kybic@fel.cvut.cz).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2009.2030955

This paper presents a general method to estimate the uncertainty of area based (or pixel based, as opposed to landmark or feature based) image registration algorithms on a particular pair of images. This method (Section II) uses bootstrap resampling [9]–[11] and performs well at the cost of increasing the computational complexity 10 ~ 100 times with respect to the original algorithm. The key feature of our approach is that the uncertainty is estimated from the input images only, under very weak assumptions about the registration problem—no ground truth and no explicit model for the transformation, the noise, or the images is needed. Also, we aim to estimate the absolute uncertainty (in pixels), not a dimensionless confidence measure with only a relative interpretation.

There are two main limitations. (i) Only the variability of the returned transformation can be estimated, not the bias. Fortunately, the bias of image registration algorithms is often quite small, as can be seen experimentally (Section III-A). (ii) We need to assume some form of ergodicity of the image generating processes, so that their behavior across realizations can be deduced from their behavior in space.

The bootstrap method is compared experimentally with the Cramér–Rao bound method [12], [13] and also with a fast registration accuracy estimation (FRAE) method, which is based on Gaussian approximation and quadratic sensitivity analysis ideas [14] (Section I-E).

A. Problem Definition I—Image Registration

Most area based image registration algorithms can be cast into the following framework: We are given two images $f, g : \mathbb{R}^m \to \mathbb{R}^n$, with $n = 1$ for grayscale images. The images are considered to be random realizations of an image-generating process (e.g., sensor noise) and are related by an unknown geometrical transformation $T_{\boldsymbol{\theta}^*} : \mathbb{R}^m \to \mathbb{R}^m$, so that pixel $f(\mathbf{x})$ corresponds to pixel $g'(\mathbf{x}) = g\big(T_{\boldsymbol{\theta}^*}(\mathbf{x})\big)$ and their values are dependent. For simplicity of exposition, we consider here especially the case of a 2-D translation ($m = 2$)

$$T_{\theta^*}(\mathbf{x}) = \mathbf{x}' = \mathbf{x} + \boldsymbol{\theta}^* \tag{1}$$

which is fully determined by a parameter vector $\boldsymbol{\theta}^* \in \mathbb{R}^d$, $d = 2$.

The quality of the registration is measured by a criterion

$$J(\boldsymbol{\theta}) = J_R(\boldsymbol{\theta}) + J_D(\boldsymbol{\theta}) \tag{2}$$

where J_R is a regularization part of the criterion, often penalizing unsmooth deformations. The data part J_D measures the similarity of the image f and the warped image g', using an image similarity measure. Again for simplicity we shall use

the sum of square differences (SSD) similarity criterion and no regularization

$$J(\boldsymbol{\theta}) = J_D(\boldsymbol{\theta}) = \sum_{\mathbf{x}\in\Omega} \left(f(\mathbf{x}) - g'(\mathbf{x})\right)^2$$
$$\text{with} \quad g'(\mathbf{x}) = g\big(T_{\boldsymbol{\theta}}(\mathbf{x})\big) = g\big(\mathbf{x} + \boldsymbol{\theta}^*\big) \tag{3}$$

where $\Omega \subset \mathbb{Z}^m$ is a set of pixels of a suitable window.

The transformation parameters are estimated as a minimizer of J

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}). \tag{4}$$

We expect the criterion to be relevant, so that the estimated transformation parameters are close to the true ones, $\hat{\boldsymbol{\theta}} \approx \boldsymbol{\theta}^*$.

Our choice of the transformation T and the criterion J makes the registration algorithm equivalent to the well-known block matching algorithm [15]. In our implementation, image g is interpolated using cubic B-splines [16], [17], its derivative is calculated analytically, and the minimization (4) is performed using the BFGS (Broyden-Fletcher-Goldfarb-Shanno) pseudo-Newton algorithm [18], which incrementally updates the estimate of the Hessian matrix from the gradient.

B. Problem Definition II—Uncertainty Estimation

Since images f, g are random (across realizations) due to the stochastic nature of the image generation process (measurement noise), the criterion $J(\boldsymbol{\theta})$ is also random, and, hence, the estimate $\hat{\boldsymbol{\theta}}$ from (4) is random, too. The problem addressed in this article is to characterize the uncertainty of $\hat{\boldsymbol{\theta}}$. In particular, we shall evaluate the covariance matrix

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathrm{E}\left[(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^T\right] \quad \text{with} \quad \bar{\boldsymbol{\theta}} = \mathrm{E}\big[\hat{\boldsymbol{\theta}}\big] \tag{5}$$

and a mean displacement variance

$$\varepsilon^2 = \mathrm{E}\Big[\operatorname*{mean}_{\mathbf{x}\in\Omega} \left\| T_{\hat{\boldsymbol{\theta}}}(\mathbf{x}) - T_{\bar{\boldsymbol{\theta}}}(\mathbf{x}) \right\|^2\Big]. \tag{6}$$

For $T(\mathbf{x}) = \mathbf{x} + \boldsymbol{\theta}$, the expression simplifies to

$$\varepsilon^2 = \operatorname{tr} \mathbf{C}_{\hat{\boldsymbol{\theta}}}. \tag{7}$$

The mean displacement variance ε^2 is equal to the mean squared geometric error (MSE) provided that the estimator (4) is unbiased, $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}^*$. MSE is in turn closely related to the warping index [19]. We also define the root mean squared error $\mathrm{RMSE} = \mathrm{MSE}^{1/2}$.

C. Related Work on Image Registration Accuracy Evaluation

Evaluation of image registration method is most often done via simulations, generating the data artificially and comparing the recovered results with the known true transformation [20]–[22]. More realistic but less widely applicable 'gold standard' approach is to use some independent and sufficiently accurate method to determine the true deformation, such as using special markers for validation which are not used for registration [23]–[25]. A "bronze standard" [26], [27] uses a robust mean of several registration algorithms as a reference. The registration accuracy can also be estimated indirectly, from ground truth segmentations [28], [29] or by its ability to create good generative models [30]. An *a posteriori* estimate is possible for low-rank transformations and a large number of corresponding features [31], [32]. Confidence measures for block matching [33], [34] and optical flow estimation [35]–[38] are based either on the data part of the criterion (such as preferring high correlation) or on the regularization part of the criterion (penalizing unlikely deformations); they can be derived from the image derivative covariance matrices [39], [40], or from *a posteriori* probabilities [41] assuming a specific noise model. However, note that confidence measures typically do not attempt to recover absolute values of registration errors, only relative ordering between errors in different spatial positions within one image.

In some special cases, typically assuming i.i.d. Gaussian noise statistics, the expected accuracy can be evaluated analytically [42]–[46].

D. Cramér–Rao Bound

The most relevant prior art is based on estimating the Cramér–Rao bound [12], [13] for $\boldsymbol{\theta}^*$, which we review here briefly using our notation for coherence. For tractability, the following observation model is assumed:

$$f(\mathbf{x}) = u'(\mathbf{x}) + w_f(\mathbf{x})$$
$$g(\mathbf{x}) = u(\mathbf{x}) + w_g(\mathbf{x})$$
$$\text{with} \quad u'(\mathbf{x}) = u\big(T_{\boldsymbol{\theta}^*}(\mathbf{x})\big) = u(\mathbf{x} + \boldsymbol{\theta}^*)$$

where w_f, w_g are zero mean i.i.d. Gaussian additive measurement noises with variance σ^2; f, g are the input images and u is a fixed but unobservable 'true' image. The corresponding log-likelihood is

$$-\log p(f, g \mid \boldsymbol{\theta}) = \text{const} + \frac{1}{2\sigma^2}\left(\sum_{\mathbf{x}\in\Omega} \big(f(\mathbf{x}) - u'(\mathbf{x})\big)^2 + \big(g(\mathbf{x}) - u(\mathbf{x})\big)^2\right). \tag{8}$$

The elements of the Fisher information matrix (FIM) $\mathbf{F}$ are

$$F_{ij}(\boldsymbol{\theta}) = -\mathrm{E}\left[\frac{\partial^2 \log p(f, g \mid \boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right]. \tag{9}$$

The second quadratic term in (8) is constant with respect to $\boldsymbol{\theta}$ and the expected value of $w_f = f(\mathbf{x}) - u'(\mathbf{x})$ is zero. Hence

$$F_{ij}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{\mathbf{x}\in\Omega} \frac{\partial u'}{\partial\theta_i} \frac{\partial u'}{\partial\theta_j} \tag{10}$$

and using the chain rule yields

$$\frac{\partial u'}{\partial\theta_i} = \sum_{k=1}^{m} \frac{\partial u'}{\partial x_k} \frac{\partial x_k}{\partial\theta_i}. \tag{11}$$

In accordance with [13], we estimate the partial derivatives $\partial u'/\partial x_k$ using first order differences.

The Cramér–Rao bound gives us a lower bound on the covariance of any unbiased estimator of $\boldsymbol{\theta}^*$, including $\hat{\boldsymbol{\theta}}$ (4)

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \geq \mathbf{F}^{-1} \tag{12}$$

in the sense of positive-semidefiniteness.

The estimate $\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{\mathrm{CRBi}} \approx \mathbf{F}^{-1}$ is described in [12] and [13]. In practice, neither σ nor u' is available. We, therefore, first perform a registration as defined by (4) to obtain $\hat{\boldsymbol{\theta}}$ and then plug-in the following ML estimates:

$$u'(\mathbf{x}) \approx \frac{1}{2}\Big(f(\mathbf{x}) + g\big(\mathbf{x}+\boldsymbol{\theta}\big)\Big) \tag{13}$$

$$\sigma^2 \approx \frac{1}{2N}\sum_{\mathbf{x}\in\Omega}\Big(f(\mathbf{x}) - g\big(\mathbf{x}+\boldsymbol{\theta}\big)\Big)^2 \tag{14}$$

into (10) to obtain a realizable CRB estimate $\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{\mathrm{CRBr}} \approx \mathbf{F}^{-1}$.

E. Fast Registration Accuracy Estimation (FRAE)

The second method which we will review here briefly and later use for comparison is the fast registration accuracy estimation method (FRAE) [14], which is based on quadratic sensitivity analysis ideas. It is a fast method, incurring only a negligible computational overhead. Given a similarity criterion which can be written as a sum of pixel contributions

$$J(\boldsymbol{\theta}) = J_D(\boldsymbol{\theta}) = \sum_{\mathbf{x}\in\Omega} e(\mathbf{x};\boldsymbol{\theta}) \tag{15}$$

we start by determining a confidence interval of the criterion value $J(\boldsymbol{\theta})$ at $\hat{\boldsymbol{\theta}}$ around a noiseless value $J^*(\boldsymbol{\theta})$

$$\mathrm{P}\big[J^*(\boldsymbol{\theta}) - \gamma \le J(\boldsymbol{\theta}) \le J^*(\boldsymbol{\theta}) + \gamma\big] = 1-\alpha. \tag{16}$$

Assuming that $J(\boldsymbol{\theta})$ is normally distributed with a standard deviation σ_J, then for $\alpha = 0.05$

$$\gamma = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sigma_J \approx 1.96\,\sigma_J \tag{17}$$

where Φ^{-1} is the inverse normal cumulative distribution function. The standard deviation can be estimated as

$$\sigma_J^2 \approx \mathrm{Var}\Big[\sum_{\mathbf{x}\in\Omega} e(\mathbf{x})\Big] = \sum_{\mathbf{x}\in\Omega}\sum_{\mathbf{y}\in\Omega}\mathrm{Cov}\big[e(\mathbf{x}), e(\mathbf{y})\big]. \tag{18}$$

For uncorrelated $e(\mathbf{x})$, a practical estimator is

$$\sigma_J^2 \approx \sum_{\mathbf{x}\in\Omega}\big(e(\mathbf{x}) - \bar{e}\big)^2 \tag{19}$$

to which we might add the effect of quantization noise [14].

As the true criterion function $J^*(\boldsymbol{\theta})$ is known with a limited accuracy the position of its minimum $\boldsymbol{\theta}^*$ is, therefore, also known only with a limited accuracy. From the confidence interval (16) and properties of minimum we get an inequality for the true value of $\boldsymbol{\theta}^*$ based on observable quantities

$$P\big[J(\boldsymbol{\theta}^*) \le J(\hat{\boldsymbol{\theta}}) + 2\gamma\big] \ge (1-\alpha)^2. \tag{20}$$

We approximate $J(\boldsymbol{\theta})$ quadratically around $\hat{\boldsymbol{\theta}}$

$$J(\boldsymbol{\theta}) = J(\hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T\mathbf{H}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) \tag{21}$$

an estimate of the Hessian $\mathbf{H}$ is available for free as a by-product of the BFGS optimization procedure. This yields

$$\mathrm{P}\big[(\boldsymbol{\theta}^*-\hat{\boldsymbol{\theta}})^T\mathbf{H}(\boldsymbol{\theta}^*-\hat{\boldsymbol{\theta}}) \le 4\gamma\big] \ge (1-\alpha)^2 \tag{22}$$

from which we can get an equivalent covariance matrix that a normally distributed $\hat{\boldsymbol{\theta}}$ would have for (22) to hold as an equality

$$\mathbf{C}_{\hat{\theta}}^{\mathrm{FRAE}} = \frac{4\gamma}{F^{-1}\big((1-\alpha)^2, d\big)}\mathbf{H}^{-1} = \lambda\sigma_J\mathbf{H}^{-1} \tag{23}$$

where F^{-1} is the inverse cumulative χ^2_{d} distribution function. The value of λ can be precomputed, for example for $\alpha = 0.05$ and $d = 2$ we get $\lambda \approx 1.6839$.

F. Bootstrap—Introduction and Related Work

Bootstrap resampling [9]–[11], [47]–[50] is a powerful and versatile computational technique for assessing the accuracy of a parametric estimator in small sample situations. Let us have N i.i.d. samples $\mathsf{X} = \{x_1, \ldots, x_N\}$ of a random variable X with a probability distribution p_X. A bootstrap resample is constructed by randomly selecting N points from X with replacement. This is repeated B times, forming B multisets[1] $\mathsf{X}^{(b)}, b = 1 \ldots B$. The bootstrap resamples $\mathsf{X}^{(b)}$ are conditionally independent given X and follow the same distribution as X.

Let us further have a continuous statistics $\vartheta(X)$ (e.g., a mean) and its estimator $\varphi(\mathsf{X}) \approx \vartheta$. We are interested in assessing the reliability of $\hat{\vartheta} = \varphi(\mathsf{X})$, as measured for example by its variance or its confidence interval. We apply the estimator φ to the bootstrap resamples $\mathsf{X}^{(b)}$, obtaining B values $\hat{\vartheta}^{(b)} = \varphi(\mathsf{X}^{(b)})$. The desired reliability measure is then evaluated using the empirical distribution of the B bootstrap values $\hat{\vartheta}^{(b)}$.

Bootstrap resampling was used in image processing to evaluate the performance of detection and classification algorithms [51], [52] and edge detectors [53], to compensate the bias in estimation of ellipse parameters [54] and to improve image segmentation [55], [56]. Bootstrap was also used to assess the accuracy of a rigid motion estimation algorithm based on 3-D key points [57], [58].

II. Bootstrap Accuracy Estimation

Bootstrap resampling accuracy estimation [59] is a general but computationally intensive method. Its inputs are a registration algorithm and the two input images f and g. In contrast to FRAE and CRB (Sections I-D and I-E), the bootstrap method can provide a nonparametric estimate of the probability density $p(\hat{\boldsymbol{\theta}})$ and any desired statistics on $\hat{\boldsymbol{\theta}}$, such confidence intervals. However, for an easy comparison with the CRB and FRAE, we will concentrate on using bootstrap to obtain a covariance matrix estimate $\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{\mathrm{boot}}$, and consequently ε^2 from (7), which has the additional advantage of requiring only a small number of bootstrap resamples B and thus being computationally tractable (see Section III-B).

A. Bootstrap Covariance Estimation

To determine the variability of $\hat{\boldsymbol{\theta}}$ from (4) we will use bootstrap to "simulate" the behavior of the criterion function $J(\boldsymbol{\theta})$ across realizations. Bootstrap can be applied to a criterion written as a sum of pixel contributions (15). However, we use

[1] A multiset is a generalization of a set, which can contain each element several times.

a more general form, anticipating its use in Section II-D. We replace the sum by a more general operation, describing the data criterion as a function Ξ of a multiset of pairs of pixel intensities of corresponding pixels

$$J_D(\boldsymbol{\theta}) = \Xi\big(\mathsf{X}(\boldsymbol{\theta})\big), \quad \mathsf{X}(\boldsymbol{\theta}) = \big\{\big(f(\mathbf{x}), g'(\mathbf{x})\big); \mathbf{x} \in \Omega\big\}. \quad (24)$$

Following the bootstrap methodology (Section I-F), we take the pixel coordinates Ω and make a set of B bootstrap resamples $\Omega^{(b)}, b = 1 \dots B$ by sampling from Ω with replacement. We get a set of B bootstrap versions of the data criterion

$$\begin{aligned} J_D^{(b)}(\boldsymbol{\theta}) = &\quad \Xi\big(\mathsf{X}^{(b)}(\boldsymbol{\theta})\big), \\ \text{with} \quad \mathsf{X}^{(b)}(\boldsymbol{\theta}) = &\quad \big\{\big(f(\mathbf{x}), g'(\mathbf{x})\big); \mathbf{x} \in \Omega^{(b)}\big\}. \end{aligned} \quad (25)$$

For example for SSD (3), the bootstrap version is

$$J_D^{(b)}(\boldsymbol{\theta}) = \sum_{\mathbf{x} \in \Omega^{(b)}} \big(f(\mathbf{x}) - g'(\mathbf{x})\big)^2 \quad (26)$$

$$\text{with} \quad \Xi(\mathsf{X}) = \sum_{(f_i, g_i) \in \mathsf{X}} (f_i - g_i)^2. \quad (27)$$

Finally, by minimization of each $J_D^{(b)}$ we get B bootstrap versions of $\hat{\boldsymbol{\theta}}$

$$\hat{\boldsymbol{\theta}}^{(b)} = \arg\min_{\boldsymbol{\theta}} \big(J_R(\boldsymbol{\theta}) + J_D^{(b)}(\boldsymbol{\theta})\big) \quad (28)$$

which can be used to estimate any desired statistics on $\hat{\boldsymbol{\theta}}$, such as the covariance matrix

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{\text{boot}} = \frac{1}{B} \sum_{b=1}^{B} \big(\hat{\boldsymbol{\theta}}^{(b)} - \mu_{\boldsymbol{\theta}}^{\text{boot}}\big)\big(\hat{\boldsymbol{\theta}}^{(b)} - \mu_{\boldsymbol{\theta}}^{\text{boot}}\big)^T \quad (29)$$

$$\text{with} \quad \mu_{\boldsymbol{\theta}}^{\text{boot}} = \frac{1}{B} \sum_{b=1}^{B} \hat{\boldsymbol{\theta}}^{(b)}. \quad (30)$$

B. Practical Bootstrap

Algorithm 1 describes a practical implementation of bootstrap resampling. At each bootstrap run, a multiset S is constructed containing pixels from Ω, some several times, some not at all, by repeatedly drawing a random number k from the uniform distribution $1 \dots N$. This induces a bootstrap version of the criterion function (25) which is then optimized. The minimization(28) is repeated B times. We have observed that $B = 10 \sim 100$ is normally sufficient to estimate $\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{\text{boot}}$ [9]. See also Section III-B. The starting point for each minimization (Algorithm 1, line 7) can be chosen randomly around the original starting point $\boldsymbol{\theta}_0$ (used to find $\hat{\boldsymbol{\theta}}$) to detect potential local minima.

Algorithm 1: Bootstrap registration uncertainty estimation

Input: Images f, g, set of pixels $\Omega = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$.
Output: Parameter $\hat{\boldsymbol{\theta}}$, covariance matrix $\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{\text{boot}} \approx \mathbf{C}_{\hat{\boldsymbol{\theta}}}$.

1 $\hat{\boldsymbol{\theta}} \leftarrow \arg\min_{\boldsymbol{\theta}} \; J_R(\boldsymbol{\theta}) + \Xi\big(\big\{\big(f(\mathbf{x}_i), g'(\mathbf{x}_i)\big)\big\}; i = 1 \dots N\big)$
2 **for** $b = 1$ to B **do**
3 $\quad S \leftarrow$ empty multiset
4 $\quad$ **for** $i = 1$ to N **do**
5 $\quad\quad S \leftarrow S \cup \{k\}; k \sim_{\text{i.i.d}} \{1 \dots N\}$
6 $\quad \hat{\boldsymbol{\theta}}^{(b)} \leftarrow \arg\min_{\boldsymbol{\theta}} \; J_R(\boldsymbol{\theta}) + \Xi\big(\big\{\big(f(\mathbf{x}_i), g'(\mathbf{x}_i)\big)\big\}; i \in S\big)$
8 Calculate $\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{\text{boot}}$ from $\{\hat{\boldsymbol{\theta}}^{(b)}; b = 1 \dots N\}$ using (29)

C. Block Bootstrap

In reality, samples $\big(f(\mathbf{x}_i), g'(\mathbf{x}_i)\big)$ are not independent — they are based on different positions in the same images which are spatially correlated and also the measurement noise can be correlated. A possible approach is to decorrelate the samples by fitting an appropriate model before bootstrapping the residuals [9], [48], [49]. A more robust technique is a *moving block bootstrap* [9], [49], [60] which we extend here to N-D. Its essence is to sample from Ω not element by element but by spatially consecutive blocks. This way, the spatial dependency is preserved if the block size Q is chosen large enough. However, choosing Q too large decreases the randomness of the sampling; we use $Q = 5$. Algorithm 2 is a modified version of Algorithm 1 using block bootstrap. The only difference is that pixel indices are added to S one block of size $(2Q+1) \times (2Q+1)$ at a time. Alternatively, a different (not rectangular) neighborhood could be used by changing the norm at line 6 of Algorithm 2.

Algorithm 2: Block bootstrap uncertainty estimation

Input: Images f, g, set of pixels $\Omega = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, block size Q.
Output: Parameter $\hat{\boldsymbol{\theta}}$, covariance matrix $\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{\text{bboot}} \approx \mathbf{C}_{\hat{\boldsymbol{\theta}}}$.

1 $\hat{\boldsymbol{\theta}} \leftarrow \arg\min_{\boldsymbol{\theta}} \; J_R(\boldsymbol{\theta}) + \Xi\big(\big\{\big(f(\mathbf{x}_i), g'(\mathbf{x}_i)\big)\big\}; i = 1 \dots N\big)$
2 **for** $b = 1$ to B **do**
3 $\quad S \leftarrow$ empty multiset
4 $\quad$ **repeat**
6 $\quad\quad S \leftarrow S \cup \big\{i \in \{1 \dots N\}; \|\mathbf{x}_i - \mathbf{x}_k\|_\infty \le Q\big\}$
7 $\quad\quad\quad$ with $k \sim_{\text{i.i.d}} \{1 \dots N\}$
8 $\quad$ **until** $\|S\| < N$
9 $\quad \hat{\boldsymbol{\theta}}^{(b)} \leftarrow \arg\min_{\boldsymbol{\theta}} \; J_R(\boldsymbol{\theta}) + \Xi\big(\big\{\big(f(\mathbf{x}_i), g'(\mathbf{x}_i)\big)\big\}; i \in S\big)$
10 Calculate $\mathbf{C}_{\hat{\boldsymbol{\theta}}}^{\text{bboot}}$ from $\{\hat{\boldsymbol{\theta}}^{(b)}; b = 1 \dots N\}$ using (29).

D. Bootstrap for Different Similarity Criteria

To demonstrate the bootstrap generality, we show its application to several commonly used image similarity criteria besides SSD (3). The *sum of absolute differences* (SAD) criterion is written as follows [compare with (27)]:

$$\Xi(\mathsf{X}) = \sum_{(f_i, g_i) \in \mathsf{X}} \big|f_i - g_i\big|.$$

Similarly, the (negative) *normalized correlation* criterion (NCC) is obtained as follows:

$$\begin{aligned} \Xi(\mathsf{X}) &= -\sum_{(f_i, g_i) \in \mathsf{X}} \frac{(f_i - \bar{f})(g_i - \bar{g})}{(s_f s_g)^{1/2}} \\ \bar{f} &= \frac{1}{N} \sum_{(f_i, g_i) \in \mathsf{X}} f_i \qquad \bar{g} = \frac{1}{N} \sum_{(f_i, g_i) \in \mathsf{X}} g_i \\ s_f &= \sum_{(f_i, g_i) \in \mathsf{X}} \big(f_i - \bar{f}\big)^2 \qquad s_g = \sum_{(f_i, g_i) \in \mathsf{X}} \big(g_i - \bar{g}\big)^2. \end{aligned} \quad (31)$$

The *mutual information* (MI) has no readily identifiable pixel contributions, nevertheless it fits well into the formulation (24)

$$J(\boldsymbol{\theta}) = J_D(\boldsymbol{\theta}) = \Xi(\mathsf{X}) = -\sum_{k=1}^{L_F} p_k \log p_k - \sum_{l=1}^{L_G} q_l \log q_l + \sum_{k=1}^{L_F}\sum_{l=1}^{L_G} r_{kl} \log r_{kl} \tag{32}$$

where r_{kl} is the smooth joint histogram [61] with $L_F \times L_G$ bins and parameters $f_{\min}$, $g_{\min}$, h_f, h_g, and p_i, q_j are the corresponding marginal histograms

$$r_{kl} = \sum_{(f_i,g_i)\in\mathsf{X}} \underbrace{\phi\left(\frac{f_i - f_{\min}}{h_f} - k\right)}_{\phi_k^F(f_i)} \underbrace{\phi\left(\frac{g_i - g_{\min}}{h_g} - l\right)}_{\phi_l^G(g_j)}$$

$$p_k = \sum_{(f_i,g_i)\in\mathsf{X}} \phi_k^F(f_i), \quad q_l = \sum_{(f_i,g_i)\in\mathsf{X}} \phi_l^G(g_j)$$

where ϕ is the chosen windowing function; we are using a linear B-spline, i.e., P1 or linear interpolation.

The bootstrap algorithm (Algorithm 1) works unchanged for all four presented similarity criteria. Care must be taken when evaluating the criterion for the minimization on line 7 that it is calculated over the bootstrap multiset S instead of the original set of pixels. The bootstrap samples are not spatially independent, especially for the NCC and MI criteria, but in spite of that, the bootstrap works well and it is not even necessary to use the block bootstrap (see the experimental results in Section III-A).

III. Experiments

A. Block Matching Accuracy Prediction

The purpose of the first experiment is to measure the true root mean squared geometrical error (RMSE) of the block matching algorithm (Section I-A) and to compare it with the predicted ε (6), (7) by the Cramér–Rao bound method CRB (Section I-D), the FRAE method [14] (Section I-E) and the bootstrap method (Section II).

We took the gray-scale 8-bit Lena image of size 512×512 pixels and selected three rectangular regions of interest (ROI) of size 61×61 containing high, medium, and low amount of texture and detail, respectively (Fig. 1). In each run, we have displaced the ROI with a randomly selected displacement $\mathbf{t}$ uniformly distributed in the range $[-2,2]^2$ pixels. We have perturbed both the original ROI and the displaced ROI with one of three types of noise: (i) uncorrelated zero-mean i.i.d. Gaussian (white) noise with varying standard deviation σ; (ii) correlated Gaussian noise obtained by convolving the i.i.d. noise by a Gaussian kernel with standard deviation 0.8 pixels; (iii) salt & pepper noise obtained by changing with probability p the value of each pixel to either 0 or 255 (chosen randomly); p was between 10^{-3} and 0.3. The block matching registration was run with a small (up to ± 0.05 pixels) random initial displacement $\boldsymbol{\theta}_0$. A constrained BFGS optimization was used with the maximum displacement set to ± 10 pixels to detect divergence. The experiment was performed 10^3 times for each method, noise type and noise level. We are reporting the root mean squared geometrical error (RMSE) $\left(\text{mean} \left\|\mathbf{t} - \hat{\boldsymbol{\theta}}\right\|^2\right)^{1/2}$ in pixels and comparing it with the mean displacement ε (6), (7) estimated by the evaluated methods. Bias is negligible in all cases. To eliminate the influence of outliers (the optimization program failing to converge) and thus distorting the statistics, we used a trimmed mean, discarding 5% of the highest and lowest values. This influences only in minor ways the reported results and only for the highest noise levels. We only report results for ROI size 61×61 because results for other ROI sizes were similar, the error slowly decreases with increasing ROI size for all methods; this is because only translational motion is considered.

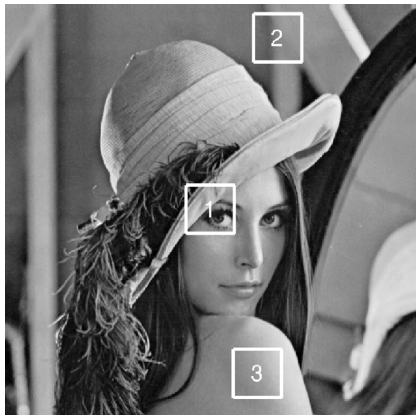


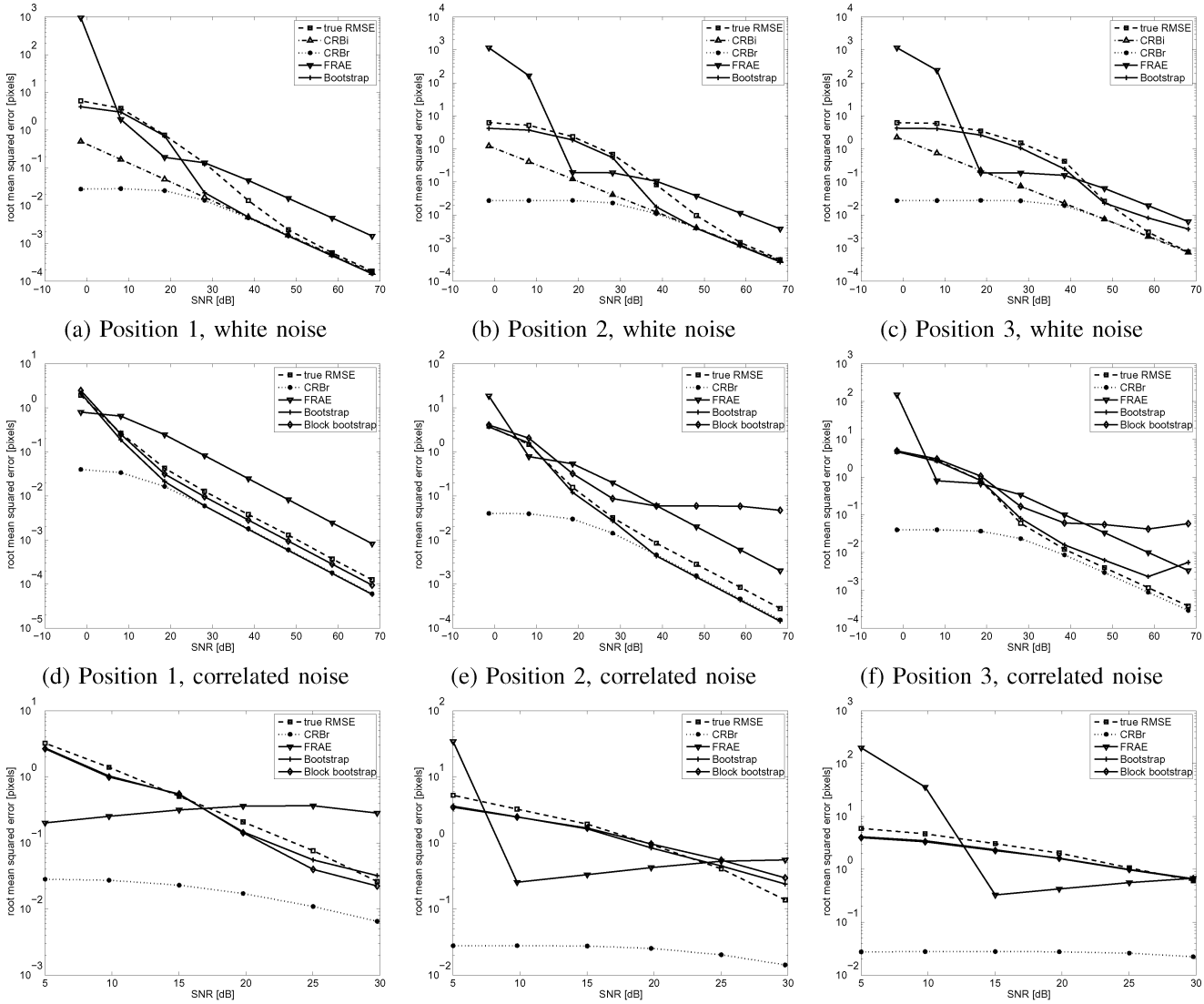
Fig. 1. Lena test image with three rectangular test areas 1,2,3 (ROIs) with progressively decreasing level of detail.

Fig. 2 shows selected results. We can see that the Cramér–Rao bound (CRBi) gives a good estimate of the accuracy, especially for higher SNR [Fig. 2(a)–(c)]. It nevertheless consistently underperforms the bootstrap and often also the FRAE method. Bear in mind, however, that under practical conditions, CRBi cannot be evaluated because it depends on unknown quantities. We can calculate only CRBr (Section I-D) which gives exceedingly optimistic estimates, especially for low SNR, being the worst of the methods tested. The advantage of CRBr is its minimal computational cost. However, the results show that it is usable only for Gaussian noise and high SNR.

For medium to high SNR and Gaussian noise, the FRAE method (Section I-E) gives usable estimates that correctly follow the trend of the true error, even though the error is often overestimated [Fig. 2(a)–(f)]. The FRAE method fails for low SNR (worse than $10 \sim 20$ dB) because the Hessian estimate is unreliable in this case. The FRAE method also fails for the salt & pepper noise at the SNR levels tested [Fig. 2(g)–(i)].

A clear winner is the bootstrap method (Section II-A). The estimated error follows the true error for both uncorrelated and correlated noise, as well as for the salt & pepper noise [Fig. 2(a)–(i)]. Most of the time the ratio between the two values is less than 2.

On the other hand, the benefit of the block bootstrap method (Section II-C) has not been demonstrated. In some cases block bootstrap performs better than normal bootstrap, such as for position 1 and correlated noise [Fig. 2(d)]. Most of the time there is no clear improvement, such as for the salt & pepper noise [Fig. 2(g)–(i)] or for uncorrelated noise (not shown). And there are also cases when block bootstrap is inferior to standard bootstrap [Fig. 2(e)-(f)].



(a) Position 1, white noise (b) Position 2, white noise (c) Position 3, white noise

(d) Position 1, correlated noise (e) Position 2, correlated noise (f) Position 3, correlated noise

(g) Position 1, salt & pepper noise (h) Position 2, salt & pepper noise (i) Position 3, salt & pepper noise

Fig. 2. Geometrical RMSE in pixels as a function of the SNR in decibels. We compare the true error with the error ε estimated by CRB, FRAE, and bootstrap methods. CRBi values are only available for the white noise case (a)–(c). Block bootstrap results are only shown in images (d)–(i) to reduce clutter. Positions 1, 2, 3 refer to Fig. 1. Each data point is an average of 10^3 experiments.

B. Number of Boostrap Resamples

The second experiment studies the effect of varying the number of bootstrap resamples B using the same setup as above (Section III-A) with uncorrelated additive Gaussian noise with standard deviation σ and region 1 (see Fig. 1). The graph in Fig. 3 shows the dependence of the coefficient of variation

$$\mathrm{cv}\big[\varepsilon_B^{\mathrm{boot}}\big] = \big(\mathrm{var}\ \varepsilon_B^{\mathrm{boot}}\big)^{1/2} / \mathrm{mean}\ \varepsilon_B^{\mathrm{boot}}$$

of the bootstrap estimate $\varepsilon_B^{\mathrm{boot}}$ on the number of bootstrap resamples B for several different noise levels. We observe that the coefficient of variation decreases with B but the decrease is slow and diminishes even further with increased noise level σ. This is in rough agreement with the theoretical formula [9]

$$\mathrm{cv}\big[\varepsilon_B^{\mathrm{boot}}\big] = \sqrt{\mathrm{cv}\big[\varepsilon_\infty^{\mathrm{boot}}\big]^2 + \frac{c}{B}} \qquad (33)$$

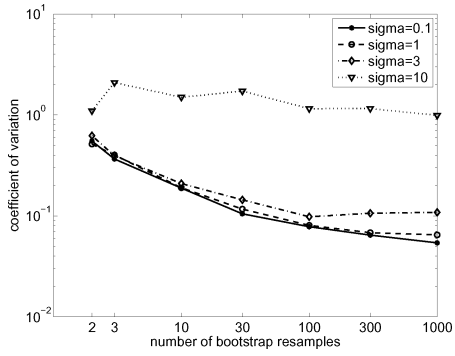


Fig. 3. Coefficient of variation (the ratio of standard deviation and mean) of the bootstrap estimate of the registration error ε as a function of the number of bootstrap resamples B for images with normal additive noise with standard deviation $\sigma = 0.1, 1, 3, 10$, corresponding to SNR $53 \sim 13$ dB.

where $\varepsilon_\infty^{\mathrm{boot}}$ represents an ideal bootstrap estimate with $B \to \infty$ (which still has some residual variance because it is based on

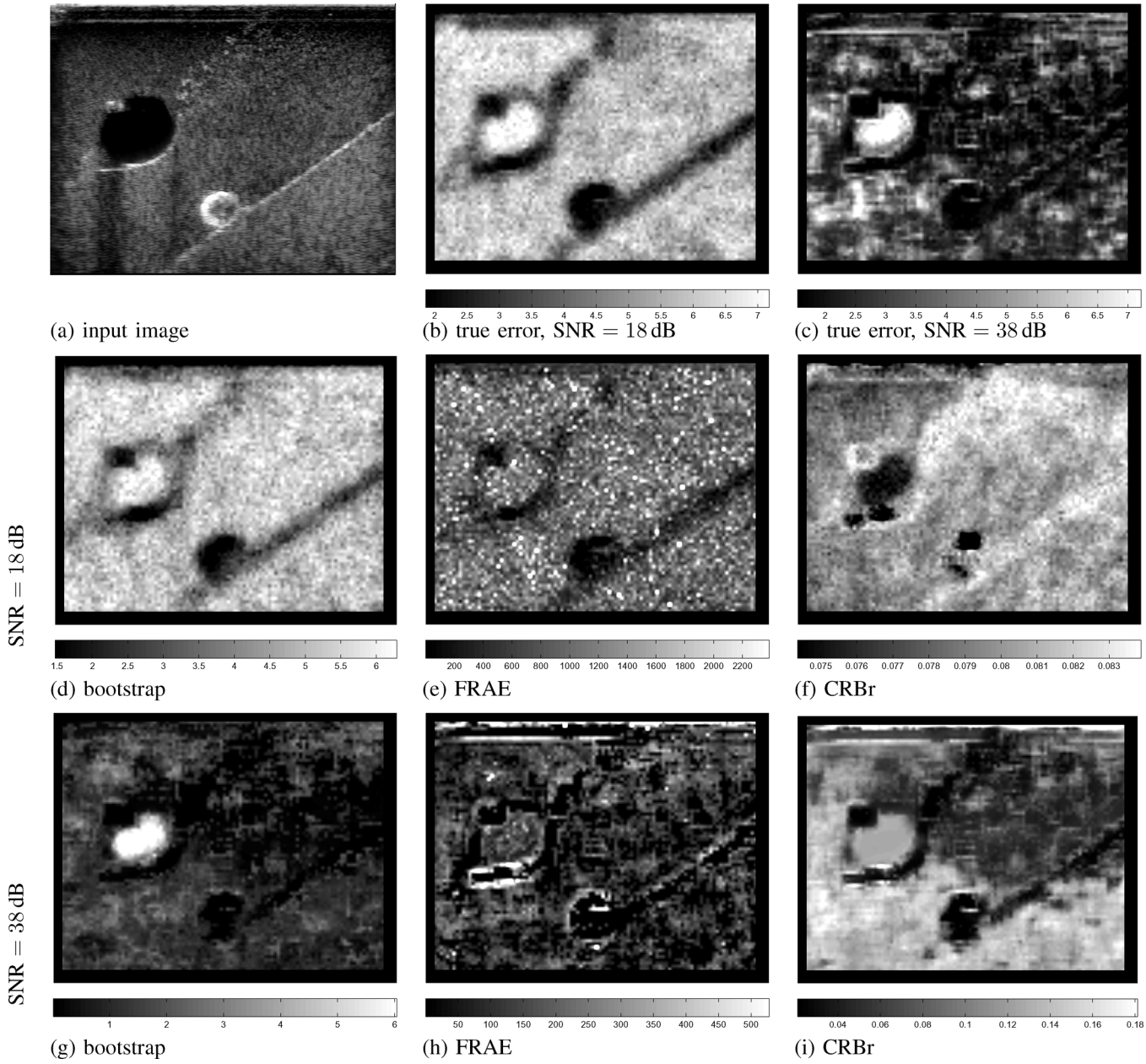


Fig. 4. Predicting spatial dependency of the block matching registration error for an ultrasound image (a) onto which we have applied a random translation. We show the geometrical RMSE in pixels using gray scale, dark tones correspond to high accuracy and vice versa. Compare the low SNR results (d)–(f) with the true error (b) and the high SNR results (g)–(i) with the true error (c). Note that the gray scale differs between images. Mean values over 10 repetitions for the bootstrap (d)–(g) and 100 repetitions for other cases (b), (c), (e), (f), (h), (i) are shown.

a finite number of noisy pixel values) and c is a constant depending on the kurtosis of ε. Both the experiment results and the literature [9] confirm our empirical observation that B as low as 10 is often enough and that it is rarely useful to increase B over 100.

C. Spatial Dependency

The third experiment evaluates the ability of the tested methods to determine the spatial dependency of the registration error. We have taken an ultrasound image [Fig. 4(a)] and translated it randomly as in Section III-A. After having added an i.i.d. Gaussing noise, we have applied the block matching algorithm to determine the displacement between the original and the translated image. We have used a window of size 29×29, its center was moved over the entire image with a step of 5 pixels. The geometrical RMSE is shown, averaged over 100 experiments, except for the bootstrap method where we have used ten experiments and $B = 10$ bootstrap repetitions. The deformation field was reinterpolated to the original resolution, the boundary regions are ignored (black in the images).

If the noise level is higher than the speckle amplitude, the deformation can be determined accurately only along the image edges [Fig. 4(b)]. This is very well captured in the bootstrap result [Fig. 4(d)] which also estimates the amplitude of the RMSE reasonably well. FRAE [Fig. 4(e)] correctly shows higher registration accuracy along some of the stronger edges but fails to

point out the low accuracy in the big textureless region (dark "cyst" on the left) and also the amplitude scale is incorrect. The CRBr results are incorrect, low and high accuracy regions are partly interchanged.

For a noise level lower than a speckle amplitude, the registration results are accurate almost everywhere except the textureless regions [Fig. 4(c)]. This is well recovered by bootstrap [Fig. 4(g)] including the amplitude scale, only the loss of accuracy in the shadows is less pronounced. The FRAE result [Fig. 4(h)] is less noisy than before [Fig. 4(e)] but the low accuracy textureless region is not identified in either. The spatial distribution shown by CRBr [Fig. 4(i)] is now more similar to the truth [Fig. 4(c)]. Nevertheless, both FRAE and CRBr are inferior to bootstrap and their amplitude scale is incorrect.

D. Motion Field Estimation, Window Size, and Sparsification

In the next set of experiments, we have applied block matching to a pair of consecutive images from the *Street* sequence, which is one of the standard synthetic sequences used in optical flow algorithm testing [62]. Fig. 5(a) and (b) shows a grayscale version of an image number 109 and an x-component of the ground truth motion field. For block matching, rectangular windows of varying size were centered at all image pixels (except borders) and a low amplitude i.i.d. Gaussian noise (SNR $\approx$ 40 dB) was added to both images. Window size is an important parameter: for too small windows the results are noisy [results with 9×9 window are shown in Fig. 5(c)], for large windows the spatial resolution is decreased [41×41 window in Fig. 5(d)]. We have measured the root mean square geometrical error (RMSE) as a function of the windows size [Fig. 5(e)], each measured point is an average of 100 registration. We have than compared the measured RMSE with the predicted ε by CRB, FRAE and bootstrap. While none of the prediction methods is able to capture the full shape of the dependence, only bootstrap is capable of successfully identifying a range of usable window sizes. FRAE fails to indicate the loss of precision for small windows and CRB exhibits no dependence at all.

The performance of confidence measures for motion field estimation is often evaluated using sparsification [38], [39], [41]: a given percentage of the 'worst' points according to a confidence measure under test is discarded and the sum of the true geometrical error of the remaining points is calculated. Fig. 5(f) shows the dependence of the total error on the relative proportion of discarded pixels for one realization of the *Street* images as described above, using block matching with window size 41×41. We see that bootstrap performs the best for most sparsification levels, the performance of FRAE and CRB is similar. All three methods are far from the best achievable results using the true error as a criterion.

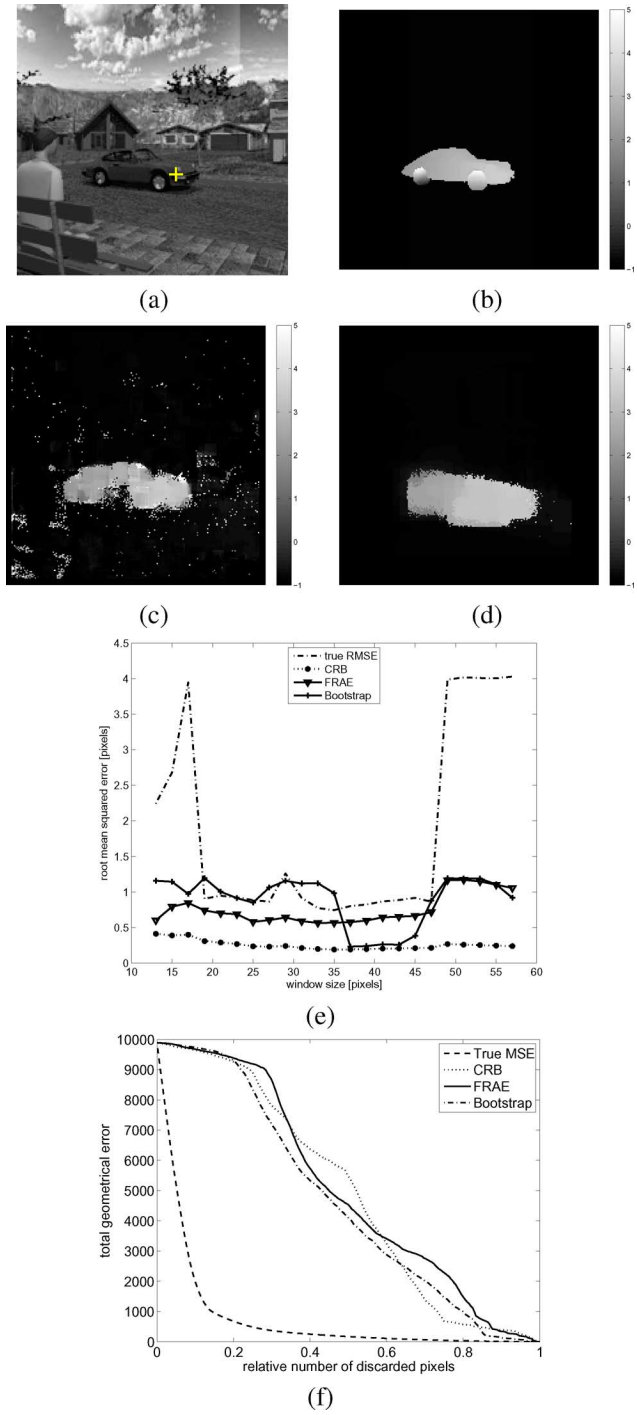


Fig. 5. Image from the *Street* synthetic sequence (a). The true x-component of the motion field (b) and the fields recovered by block matching for window sizes 9×9 (c) and 41×41 (d). The gray scale mapping is the same in images (b)–(d). The dependence of the true and predicted error on the window size (e) around the point marked by a cross in image (a). Total geometrical error over the image for one realization as a function of the relative proportion of discarded pixels ranked according to bootstrap, FRAE and CRB error estimates, and the true error (f).

E. Alternative Similarity Criteria

We have used the same experimental setup as in Section III-A to evaluate the performance of the bootstrap method for the SAD, NCC, and MI criteria. The experimental results (Fig. 6) show that the bootstrap estimation of ε (6), (7) follows the true RMSE error very well for low and medium SNR for all three criteria and for high SNR for the SAD and MI criteria. The difference for higher SNR for the correlation criterion only occurs at very high accuracy levels which are unlikely to appear in practice.

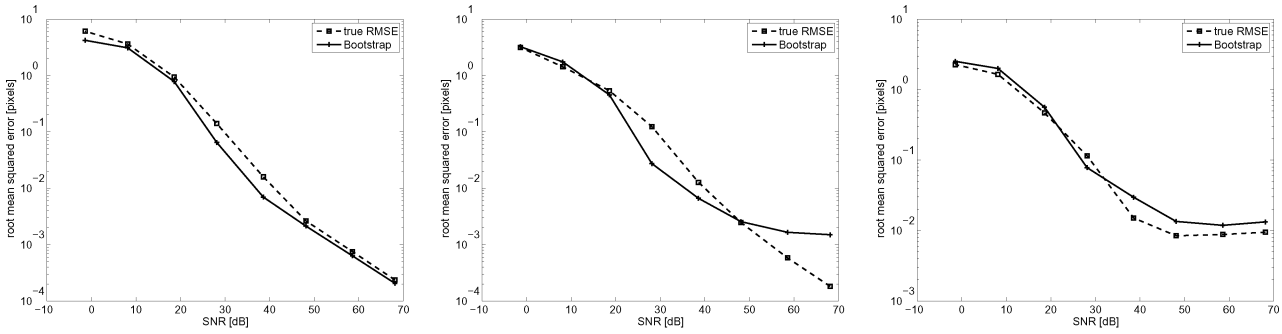


Fig. 6. Geometrical RMSE in pixels as a function of the SNR in dB for position 1 in Fig. 1, as in Fig. 2. We compare the true error with the error ε estimated by the bootstrap method for the (a) SAD, (b) normalized correlation, and (c) mutual information image similarity criteria. Each data point is an average of 100 experiments.

IV. Conclusion

We have shown a general bootstrap based technique to estimate image registration uncertainty, applicable to a wide class of minimization-based pixel registration algorithms. Given nothing more than the two images to be registered, we can predict the registration error in a quantitative way. Only very weak assumptions have to be made. Although for the sake of simplicity we have considered only 2-D translations, the presented accuracy estimation techniques are directly usable for other registration methods that find transformation with more degrees of freedom.

Of the examined methods, bootstrap performed better than both FRAE and CRB and should be used if we can afford the its computational complexity. We expect the described methodology to find many uses in all practical situations whenever an estimate of a registration uncertainty is needed.

References

[1] L. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 326–376, Dec. 1992.

[2] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, no. 21, pp. 977–1000, 2003.

[3] D. Robinson, S. Farsiu, and P. Milanfar, "Optimal registration of aliased images using variable projection with applications to super-resolution," *The Comput. J.*, Apr. 2007.

[4] H. Lester and S. R. Arridge, "A survey of hierarchical non-linear medical image registration," *Pattern Recognit.*, vol. 32, no. 1, pp. 129–149, Jan. 1999.

[5] J. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, no. 1, pp. 1–36, 1998.

[6] J. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[7] P. A. van den Elsen, E.-J. D. Pol, and M. A. Viergever, "Medical image matching—A review with classification," *IEEE Eng. Med. Biol.*, pp. 26–39, Mar. 1993.

[8] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Phys. Med. Biol.*, no. 46, pp. R1–R45, 2001.

[9] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993, vol. 57, Monographs on Statistics and Applied Probability.

[10] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*, no. 7, pp. 1–26, Jul. 1979.

[11] A. M. Zoubir and B. Boashash, "The bootstrap and its applications in signal processing," *IEEE Signal Process. Mag.*, no. 1, pp. 56–76, Jan. 1998.

[12] D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *IEEE Trans. Image Process.*, vol. 13, pp. 1185–1199, 2004.

[13] I. S. Yetik and A. Nehorai, "Performance bounds on image registration," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1737–1748, May 2006.

[14] J. Kybic, "Fast no ground truth image registration accuracy evaluation: Comparison of bootstrap and Hessian approaches," in *Proc. IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, New York, May 2008, pp. 792–795.

[15] J. Jain and A. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, pp. 1799–1808, Dec. 1981.

[16] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Process. Mag.*, vol. 16, no. 6, pp. 22–38, Nov. 1999.

[17] I. J. Schoenberg, "Contributions to the problem of approximation of equidistant data by analytic functions," *Quart. Appl. Math.*, vol. 4, pp. 45–99, 1946.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[19] P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Trans. Image Process.*, vol. 7, pp. 1–15, Jan. 1998.

[20] R. M. Haralick, H. Joo, C. N. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim, "Pose estimation from corresponding point data," *IEEE Trans. Syst., Man, Cybern.*, vol. 6, no. 19, pp. 1426–1446, 1989.

[21] J. L. Barron, S. S. Fleet, and D. J. nad Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, no. 12, pp. 43–77, 1994.

[22] S. Kiebel, J. Ashburner, J. Poline, and K. Friston, "MRI and PET coregistration—A cross validation of statistical parametric mapping and automated image registration," *Neuroimage*, no. 5, 1997.

[23] C. J. Maurer, J. Fitzpatrick, M. Wang, R. J. Galloway, R. Maciunas, and G. Allen, "Registration of head volume images using implantable fiducial markers," *IEEE Trans. Med. Imag.*, vol. 16, 8, Aug. 1997.

[24] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. J. Maurer, R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawkes, C. Studholme, J. B. A. Maintz, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, G. Q. J. Maguire, M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *J. Comput. Assist. Tomogr.*, vol. 21, no. 4, pp. 554–568, 1997.

[25] P. Jannin, J. Fitzpatrick, D. Hawkes, X. Pennec, R. Shahidi, and M. Vannier, "Validation of medical image processing in image-guided therapy," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1445–1449, Dec. 2002.

[26] T. Glatard, X. Pennec, and J. Montagnat, "Performance evaluation of grid-enabled registration algorithms using bronze-standards," in *Proc. MICCAI—Medical Image Computing and Computer-Assisted Intervention*, 2006, vol. 4191, pp. 152–160.

[27] S. Nicolau, X. Pennec, L. Soler, and N. Ayache, "Evaluation of a new 3D/2D registration criterion for liver radio-frequencies guided by augmented reality," in *Proc. Int. Symp. Surgery Sim. Soft Tissue Model*, 2003, pp. 270–283.

[28] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1451–1461, Nov. 2006.

[29] P. Hellier, C. Barillot, I. Corouge, B. Gibaud, G. Le Goualher, D. Collins, A. Evans, G. Malandain, N. Ayache, G. Christensen, and H. Johnson, "Retrospective evaluation of intersubject brain registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1120–1130, Sep. 2003.
[30] R. Schestowitz, C. J. Twining, T. F. Cootes, V. S. Petrovic, C. J. Taylor, and W. R. Crum, "Assessing the accuracy of non-rigid registration with and without ground truth," in *Proc. IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, Apr. 2006, pp. 836–839.
[31] X. Pennec and J.-P. Thirion, "A framework for uncertainty and validation of 3d registration methods based on points and frames," *Int. J. Comput. Vis.*, vol. 25, no. 3, pp. 203–229, 1997.
[32] M. H. Moghari and P. Abolmaesumi, M. I. Miga and K. R. Cleary, Eds., "Maximum likelihood estimation of the distribution of target registration error," in *Proc. Medical Imaging: Visualization, Image-Guided Procedures, and Modeling*, Apr. 2008, vol. 6918, pp. 69180I–69180I-12.
[33] M. Chan, Y. Yu, and A. Constantinides, "Variable size block matching motion compensation with applications to video coding," *IEE Proc.*, no. 4, Aug. 1990.
[34] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *Int. J. Comput. Vis.*, vol. 2, no. 3, pp. 283–310, Jan. 1989.
[35] D. J. Heeger, "Optical flow using spatiotemporal filters," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 279–302, 1988.
[36] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger, "Probability distributions of optical flow," in *Proc. Conf Computer Vision and Pattern Recognition*, Mauii, HI, 1991, pp. 310–315.
[37] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods," *Int. J. Comput. Vis.*, vol. 61, no. 3, pp. 211–231, 2005.
[38] A. Bruhn and J. Weickert, "A confidence measure for variational optic flow methods," in *Geometric Properties for Incomplete Data*, R. Klette, R. Kozera, L. Noakes, and J. Weickert, Eds. New York: Springer, 2006, pp. 283–298.
[39] A. Dev, B. J. A. Kröse, and F. C. A. Groen, "Confidence measures for image motion estimation," in *Proc. Real World Computing Symp.*, 1997, pp. 199–206.
[40] U. Neumann and S. You, "Adaptive multi-stage 2D image motion field estimation," in *Proc. SPIE Conf. Applications of Digital Image Processing*, 1998, vol. 3460, pp. 116–123.
[41] I. Patras, I. A. Hendriks, and R. L. Lagendijk, "Confidence measures for block matching motion estimation," in *Proc. Int. Conf. Image Processing*, 2002, pp. 277–280.
[42] C. Q. Davis and D. M. Freeman, "Statistics of subpixel registration algorithms based on spatiotemporal gradients or block matching," *Opt. Eng.*, vol. 4, no. 37, pp. 1290–1298, Apr. 1998.
[43] K. Kanatani, *Geometric Computation for Machine Vision*. New York: Oxford Univ. Press, 1993.
[44] K. Kanatani, "Analysis of 3-D rotation fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 543–549, May 1994.
[45] M. A. Snyder, "The precision of 3-D parameters in correspondence based techniques: The case of uniform translational motion in rigid environment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 5, no. 11, pp. 523–528, Nov. 1998.
[46] M. Steele and C. Jaynes, "Parametric subpixel matchpoint recovery with uncertainty estimation: A statistical approach," in *Proc. Computer Vision and Pattern Recognition—Workshop on Statistical Analysis in Computer Vision*, Los Alamitos, CA, 2003, vol. 8, p. 90.
[47] A. M. Zoubir and D. R. Iskander, *Bootstrap Techniques for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[48] A. M. Zoubir and R. Iskander, "Bootstrap methods and applications," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 10–19, 2007.
[49] D. N. Politis, "Computer-intensive methods in statistical analysis," *IEEE Signal Process. Mag.*, vol. 15, no. 1, pp. 39–55, Jan. 1998.
[50] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
[51] A. K. Jain, R. C. Dubes, and C. C. Chen, "Bootstrap techniques for error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 5, pp. 628–633, May 1987.
[52] B. C. Jiang, Y. M. Wang, and C. C. Wang, "Bootstrap sampling technique applied to the PCB golden fingers defect classification study," *Int. J. Production Res.*, vol. 39, no. 10, pp. 2215–2230, Jul. 2001.
[53] K. Cho, P. Meer, and C. Javier, "Performance assessment through bootstrap," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 11, pp. 1185–1198, Nov. 1997.
[54] J. Cabrera and P. Meer, "Unbiased estimation of ellipses by bootstrapping," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 7, pp. 752–756, Jul. 1996.
[55] F. Ghorbel and C. Banga, "Bootstrap sampling applied to image analysis," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1994, vol. 6, pp. 81–84.
[56] C. Banga and F. Ghorbel, "Optimal bootstrap sampling for fast image segmentation: Application to retina image," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 1993, vol. 5, pp. 638–641.
[57] B. Matei, P. Meer, and D. Tyler, "Performance assessment by resampling: Rigid motion estimators," in *Proc. IEEE Conf. Empirical Evaluation Techniques in Computer Vision*, K. Bowyer and P. Phillips, Eds., 1998, pp. 72–95.
[58] B. Matei and P. Meer, "Optimal rigid motion estimation and performance evaluation with bootstrap," in *Proc. Computer Vision and Pattern Recognition Conf.*, 1999, vol. 1, pp. 339–345.
[59] J. Kybic and D. Smutek, "Image registration accuracy estimation without ground truth using bootstrap," in *CVAMIA: Computer Vision Approaches to Medical Image Analysis*, R. Beichel and M. Sonka, Eds. New York: Springer, 2006, vol. 3117, Lecture Notes in Computer Science, pp. 61–72.
[60] H. R. Künsch, "The jackknife and the bootstrap for general stationary observations," *Ann. Statist.*, no. 17, pp. 1217–1241, 1989.
[61] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Process.*, vol. 9, no. 12, pp. 2083–2099, Dec. 2000.
[62] B. McCane, K. Novins, D. Crannitch, and B. Galvin, "On benchmarking optical flow," *Comput. Vis. Image Understand.*, vol. 84, no. 1, pp. 126–143, 2001.

Jan Kybic (M'99–SM'08) was born in Prague, Czech Republic, in 1974. He received the Bc. (BSc.) and Ing. (MSc.) degrees with honors from the Czech Technical University, Prague, in 1996 and 1998, respectively, and the Ph.D. in biomedical image processing from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2001, for his thesis on elastic image registration using parametric deformation models.

From 2002–2003, he held a postdoctorate research position in INRIA, Sophia-Antipolis, France. Since 2003, he has been a Research Fellow with the Center for Machine Perception, Czech Technical University, Prague.

Dr. Kybic is an Associate Editor for the IEEE TRANSACTIONS ON MEDICAL IMAGING. His research interests include signal and image processing, medical imaging, image registration, splines and wavelets, inverse problems, elastography, computer vision, numerical methods, algorithm theory, and control theory. He has authored or co-authored 14 articles in prestigious peer-reviewed scientific journals, one book, two book chapters, and over 50 conference papers.