Tracking Performance of Several Combinations of Common Evaluation Metrics and Sub-pixel Methods

J. Albinsson[1], T. Jansson[2], and M. Cinthio[1]

[1] Department of Biomedical Engineering, Faculty of Engineering, Lund University, Lund, Sweden
[2] Clinical Sciences Lund, Biomedical Engineering, Lund University, and Medical Services, Skåne University Hospital, Lund, Sweden

Abstract— **Motion estimation in a series of consecutive images is used in a variety of areas, e.g. video compression and investigation of tissue characteristics and organ function in medical images. Several methods exist both for estimating motions on a pixel level, e.g. block-matching in which two blocks in consecutive images are compared by an evaluation metric, and on a sub-pixel level. In this paper, we have evaluated the tracking performance of all combinations between three evaluation metrics and eight sub-pixel estimation methods. The tracking performance of a sub-pixel method varies depending on the evaluation metric used. This indicates that a reported tracking performance for a sub-pixel estimation method can be significantly different when combined with another evaluation metric. Also there is a large variation in the time needed for the motion estimations depending primarily on the sub-pixel method used but also on the evaluation metric.**

Keywords— **Sub-pixel estimation, block-matching, motion estimation, ultrasound**

I. INTRODUCTION

Motion estimation in a series of consecutive images is used in a variety of areas, e.g. video compression and investigating tissue characteristics and organ function in medical images. For example in ultrasound where tissue motion measurements results in functional information on the tissue-of-interest, and have attracted attention for various applications such as evaluation of cardiac [1], vascular [2], and skeletal muscle [3] function. It has been shown that tissue motion measurements can not only provide new information about the tissue-of-interest [2], but can also provide prognostic information for patients having suspected cardiovascular disease, both with cardiac [4] and vascular applications [5].

Several methods for estimating motion in a series of digital images exist, e.g. optical flow and phase correlation. In this work, we have used a basic block-matching method implemented with a small search region. In order to estimate the motion in an area, a block representing the area is chosen and compared to every block in the search region with the block most similar with the original block assumed to depict the same object. To determine the similarity between two blocks, an evaluation metric, e.g. sum of absolute difference (SAD), sum of squared difference (SSD), or two dimensional normalized cross correlation (NCC), can be used.

However, most of the time, the length of the motion of an object between two successive images is not an exact number of pixels. Therefore, an estimation of the motion on a level of pixels is not enough as the relative error between estimated movement and true movement will, at least for short distances, be larger than acceptable. Several methods exist to determine the motion on a sub-pixel level. The methods can roughly be divided into three sub-groups: 1) analytically solved using the evaluation metric values as input, 2) interpolation of evaluation metric values, and 3) image interpolation. Both the first and second group uses the evaluation metric values in order to determine the sub-pixel position; values that differ depending on the evaluation metric used.

In motion estimations, two performance criterions are vital: estimation time and tracking error. Naturally the error should be kept as small as possible in order to give a correct estimation; but for real-time applications also the estimation time is critical. Often, the size of the estimation error is negatively correlated to the estimation time.

We have found no previous study that evaluates the impact on the performance of a sub-pixel method when used together with different evaluation metrics. The aim of this work was to evaluate the performance of eight sub-pixel position methods when combined with three different evaluation metrics *in silico*.

II. MATERIAL AND METHODS

A. Image sequences

The investigations in this paper have been made on simulated ultrasound images. In ultrasound images, horizontal direction is commonly denoted lateral while vertical direction is denoted axial; a convention that has been adopted in this paper.

The cineloops (50 images each, pixel density: 8.1x4.1 pixel/mm) were simulated using Field II [6]. The lateral velocities included 0.3, 0.6, 0.9, 1.4, 2.2, and 2.8 pixels/image whereas the axial velocity was 0 pixels/image. Motion estimation was performed on 100 blocks (15x7

H. Mindedal and M. Persson (eds.), *16th Nordic-Baltic Conference on Biomedical Engineering*,
IFMBE Proceedings 48, DOI: 10.1007/978-3-319-12967-9_4

pixels each) in each image using a full-search scheme (11x11 pixels) centered on the pixel position of the searched-for block. The searched-for blocks were evenly distributed above and below the focus depth.

B. Evaluation metrics

The performance for motion estimation was evaluated for three evaluation metrics:

SAD $$\alpha = \sum_{i=1}^{i=l}\sum_{j=1}^{j=k}\left|X_{i,j} - Y_{i+m,j+n}\right| \quad (1)$$

SSD $$\beta = \sum_{i=1}^{i=l}\sum_{j=1}^{j=k}\left(X_{i,j} - Y_{i+m,j+n}\right)^2 \quad (2)$$

NCC $$\gamma = \frac{\sum_{i=1}^{i=l}\sum_{j=1}^{j=k}(X_{i,j}-\bar{X})(Y_{i+m,j+n}-\bar{Y})}{\sqrt{\sum_{i=1}^{i=l}\sum_{j=1}^{j=k}(X_{i,j}-\bar{X})^2 \sum_{i=1}^{i=l}\sum_{j=1}^{j=k}(Y_{i+m,j+n}-\bar{Y})^2}} \quad (3)$$

Here α, β, and γ denotes the calculated evaluation metric value; m and n denotes the displacement between the two blocks; l and k denotes the size of the blocks; $X_{i,j}$ and $Y_{i,j}$ denotes the pixel values at position (i, j) in the tracked block and the compared block, respectively, while $\bar{X}$ and $\bar{Y}$ denotes the average pixel values of the respective blocks.

C. Sub-pixel estimation methods

A total of eight sub-pixel position methods were evaluated. Three of the methods were analytically solved: 1. 1D parabolic interpolation (1D PI), 2. grid slope interpolation (GS) [7], and 3. 2D parabolic interpolation (2D PI). Three of the methods interpolated the calculated evaluation metric values to a factor 128:1: 4. 1D PI, 5. 2D PI, and 6. 2D cubic interpolation (2D Cubic). Two methods interpolated the image data a factor 128:1: 7. 2D PI and 8. 2D Cubic.

1. 1D PI fits a one dimensional second-degree polynomial (4), to three evaluation metric values with the center value corresponding to the block with the best similarity to the searched-for block.

   $$y = ax^2 + bx + c \quad (4)$$

   The polynomial was fitted separately laterally and axially in order to get a sub-pixel estimation in both directions. Fitting + (4) gave, when analytically solved, the sub-pixel estimation:

   $$\Delta x = \frac{\alpha_1-\alpha_3}{2(\alpha_1+\alpha_3-2\alpha_2)} \quad (5)$$

   Here α_1, α_2 (center), and α_3 were evaluation metric values and Δx was the sub-pixel part of the movement with a value less than ±0.5 pixels.
2. GS estimates the sub-pixel position separately laterally and axially by use of two evaluation metric values from the current image: the center value corresponding to the block with the best similarity to the searched-for block and the value closest to zero of the two values next to the center value (left or right for lateral or up or down for axial estimation). GS also uses an evaluation metric value calculated between the searched-for block and a block in the same image as the searched-for block at the position of the evaluation metric value used in the current image:

   $$\Delta x = 0.5\left(1 - \frac{\alpha_2-\alpha_i}{\alpha_{2,0}-\alpha_{i,0}}\right) \quad (7)$$

   Here α_2 (center) and α_i were evaluation metric values in the current image and $\alpha_{2,0}$ and $\alpha_{i,0}$ were evaluation metric values in the previous image. It should be noted that one of $\alpha_{2,0}$ and $\alpha_{i,0}$ were zero [7].
3. In 2D PI the second-degree polynomial was two dimensional.

   $$z = a + bx + cy + dx^2 + exy + fy^2 \quad (6)$$

   The polynomial was fitted to nine evaluation metric values with the value in the center corresponding to the block with the best similarity to the searched-for block. The polynomial was then solved analytically by finding the extreme point close to the center position.
4. The polynomial (4) was used for interpolating evaluation metric values in 127 evenly distributed points between each pixel position before finding the position with the best evaluation metric value. This was done separately axially and laterally.
5. The polynomial (6) was used for interpolating evaluation metric values in 127 evenly distributed points between each pixel position both axially and laterally before finding the position with the best evaluation metric value.
6. 2D Cubic used cubic spline interpolation [8] in order to interpolate nine evaluation metric values to 128:1 both axially and laterally. The optimal position was given directly by the best evaluation metric value.
7. The polynomial (6) were fitted to nine pixel values and were used to interpolate a square, ±0.5 pixels both laterally and axially, around the center pixel to 128x128 samples for a total of 15x7 original pixels before performing a full-search with the chosen evaluation metric.
8. 2D Cubic used cubic spline interpolation [8] in order to interpolate 15x7 pixel values to 128:1 both axially and laterally original pixels before performing a full-search with the chosen evaluation metric.

D. Evaluation of tracking performance

The tracking performance of each combination of evaluation metric and sub-pixel estimation method was evaluated by measuring the estimation time and by calculating the difference between the set movement and the estimated movement. The axial and lateral estimation errors were treated separately. The errors of the motion estimation (pixel/image) for each combination of evaluation metric, sub-pixel position method, and velocity are presented as boxplots in Figure 1. Values outside ±0.25 pixel are part of the statistics but not shown in the figure. The average estimation time in seconds needed for each image (100 blocks) is presented between the lateral and axial errors.

III. RESULTS AND DISCUSSION

Figure 1 shows average estimation time between the lateral and axial estimation errors. For every combination of evaluation metric and sub-pixel estimation method, the estimation errors have been combined in a boxplot. The results show an expected trade-off between estimation time and size of estimation errors. The smallest estimation error is in general obtained when the image was interpolated whereas the analytically solved methods are up to 150 times faster.

Reading the description of GS, we find nothing indicating that the method was tested or even intended to be used for determining sub-pixel positions axially. It could be hypothesized that the method are sensitive for noise when the distances between samples are short. This should explain the difference in the variance of estimation errors axially and laterally as the axially distance is a factor 2 shorter.

The large difference in tracking performance using 2D PI for image interpolation can be explained by discontinuities in the interpolated image, i.e. the interpolation will produce edges halfway between two original pixels where two polynomials meet. The edges will also cause a fluctuation of energy in the image. A great strength of NCC can be seen in Figure 1 as its tracking performance is significantly better for this sub-pixel method than the other evaluation metrics.

A tendency that can be observed in the results of several sub-pixel methods is the presence of bias in the tracking errors, i.e. the error is dependent on the size of the movement in the images.

However, several combinations give satisfying results both in terms of estimation error magnitude and estimation time. The figure also indicates, for the first time, possible combinations of methods for improved performance, e.g. using SAD and analytically solved 1D PI (axial)/GS(lateral).

Though the investigations in this paper have been made on ultrasound images, we are confident that similar results would be seen in clinical images obtained with other modalities such as MRI or CT.

IV. CONCLUSIONS

It is well known that the choice of evaluation metric can have a significant impact on both estimation time and the magnitude of the estimation error when using block-matching for motion estimation. Here, we show that a combination of evaluation metric and sub-pixel estimation method have an effect on the size of the estimation error. Thus, even if two evaluation metrics have the same tracking performance on a pixel level, the total tracking performance can differ depending on what sub-pixel estimation method is applied.

ACKNOWLEDGMENT

This study was supported by grants from the Swedish Research Council

REFERENCES

[1] W. N. McDicken, G. R. Sutherland, C. M. Moran, and L. N. Gordon, "Colour Doppler velocity imaging of the myocardium," Ultrasound Med. Biol., vol. 18, pp. 651-654, 1992.

[2] M. Cinthio, Å. R. Ahlgren, J. Bergkvist, T. Jansson, H. W. Persson, and K. Lindström, "Longitudinal movements and resulting shear strain of the arterial wall," Am. J. Physiol.-Heart. C., vol. 291, pp. H394-H402, 2006.

[3] S. Brorsson, A. Nilsdotter, M. Hilliges, C. Sollerman, and Y. Aurell, "Ultrasound evaluation in combination with finger extension force measurements of the forearm musculus extensor digitorum communis in healthy subjects," BMC Med. Imaging, vol. 8, p. 6, 2008.

[4] C. F. Madler, N. Payne, U. Wilkenshoff, A. Cohen, G. A. Derumeaux, L. A. Pierard, J. Engvall, L. A. Brodin, G. R. Sutherland, and A. G. Fraser, "Non-invasive diagnosis of coronary artery disease by quantitative stress echocardiography: optimal diagnostic models using off-line tissue Doppler in the MYDISE study," Eur. Heart. J., vol. 24, pp. 1584-1594, 2003.

[5] J. Blacher, A. P. Guerin, B. Pannier, S. J. Marchais, M. E. Safar, and G. M. London, "Impact of aortic stiffness on survival in endstage renal disease," Circulation, vol. 99, pp. 2434-2439, 1999.

[6] B. J. Geiman, L. N. Bohs, M. E. Anderson, S. M. Breit, and G. E. Trahey, "A novel interpolation strategy for estimating subsample speckle motion," Phys. Med. Biol., pp. 1541-1552, 2000.

[7] R. G. Keys, "Cubic Convolution Interpolation for Digital Image Processing," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-29, pp. 1153-1160, 1981.

[8] J. A. Jensen, "Field: A program for Simulating Ultrasound Systems," 10th Nordic-Baltic Conference on Biomedical Imaging Published in Medical & Biological Engineering & Computing, vol. 34, pp. 351-353, 1996.

Corresponding author:

Author: John Albinsson
Institute: Biomedical Engineering
Street: Ole Römers väg 3
City: Lund
Country: Sweden
Email: john.albinsson@bme.lth.se

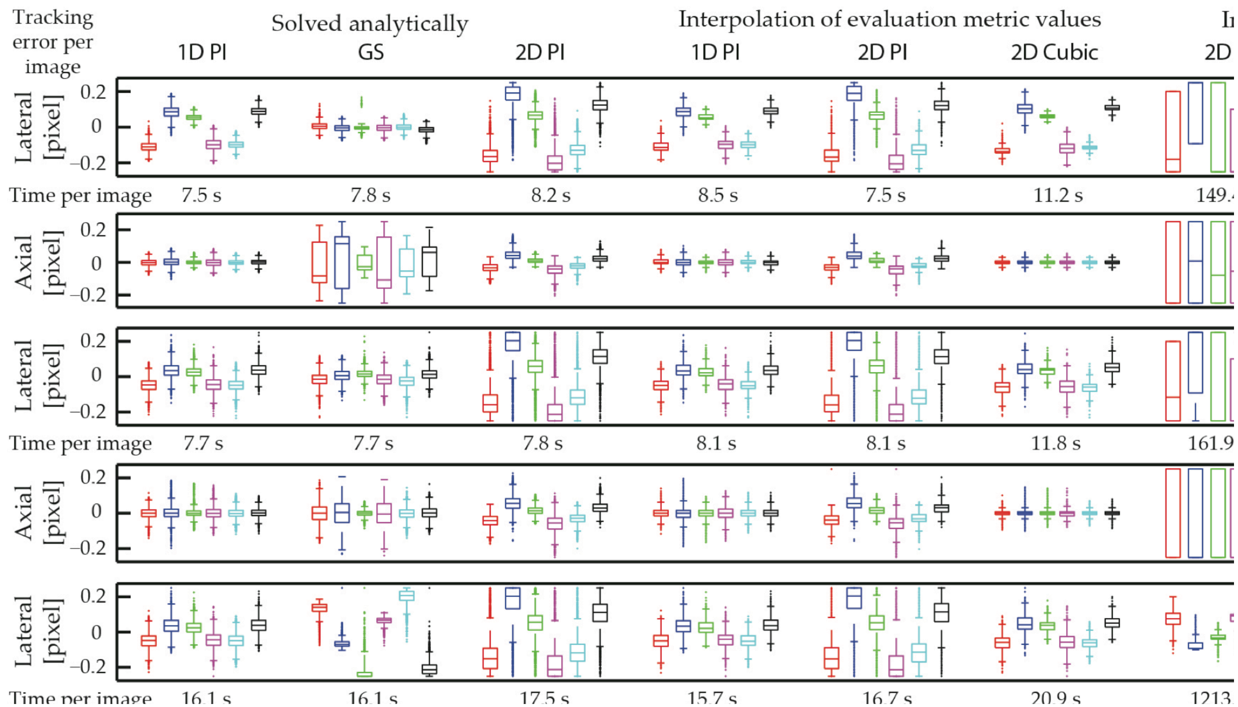

Fig. 1 Tracking error per image of the motion estimation performed on the in silico cineloops against the set movement per image of the cineloops. The boxes indicate the lower and upper quartiles and the median. The bar line indicate 99% of the values. Outliers are indicated as points. Each box is based on 4,900 estimations explaining the seemingly large numbers of outliers. No error larger than 0.25 pixels is shown but they were part of all the statistics.