

Effective Intensity-Based 2D/3D Rigid Registration between Fluoroscopic X-Ray and CT

D. Knaan and L. Joskowicz

School of Computer Science and Engineering
The Hebrew University of Jerusalem, Jerusalem 91904, Israel.
josko@cs.huji.ac.il

Abstract. We present an effective intensity-based method for rigid registration of a patient preoperative CT to its intraoperative situation with a few fluoroscopic X-ray images obtained with a tracked C-arm. It improves upon existing methods and overcomes most of their intrinsic speed, robustness, and accuracy problems. For speed, we generate Digitally Reconstructed Radiographs on small, dynamically selected regions of interest from precomputed ray gray levels in expected viewing directions, and use a multiresolution hierarchy of fluoroscopic X-ray images. For robustness and accuracy, we use a two-step comparison measure: Normalized Cross Correlation followed by Variance Weighted Sum of Local Normalized Correlation. To avoid local minima, we use a genetic search method. Our experiments on simulated, in-vitro, and cadaver data show an overall mean Target Registration Error of 0.8mm (max=1.5mm), 95% of the time, computed in 20–100 seconds under realistic conditions.

1 Introduction

Registration is an essential step in most computer-aided surgery (CAS) systems. It consists of finding a transformation that aligns common features from two modality data sets taken at different times [1]. It is required in image-guided and robotic surgery to match preoperative images and plans to the intraoperative situation, to determine the relative position of surgical tools and anatomy, and to accurately position and move surgical robots. Several types of CAS systems have been developed, mostly for neurosurgery [2], orthopaedic surgery [3], and radiation therapy [4].

Most CAS systems in clinical use rely on implanted fiducials, which may require additional surgery, or on points harvested by direct contact on the anatomy surface, which may require additional anatomy exposure and is time consuming. An alternative is anatomical image-based registration, which relies on anatomical structures to match the data sets. It allows both less invasive and percutaneous procedures, is faster and less prone to human error, and does not require training. Intraoperative X-ray fluoroscopy is an ubiquitous modality which is well suited for registering preoperative CT data of bones and rigid structures.

Anatomical image-based 2D/3D rigid registration between fluoroscopic X-ray and CT is technically much harder than fiducial or contact-based registration. It requires analyzing the fluoroscopic X-ray images, which are 2D, have a small field of view, have limited resolution, have camera position-dependent geometric and intensity distortions, and may contain surgical tools, implants, and anatomy not present in the preoperative CT.

2 Previous Work

Because of its great potential, researchers have developed a variety of methods for 2D/3D rigid registration between X-ray and CT. Geometry-based algorithms match selected geometric features from each data set by finding the transformation that minimizes the sum of distances between paired features [5,6,7]. Intensity-based algorithms match the intensities of one image data set with the intensity of the other by minimizing a similarity measure between them [8,9,10,11,12,13,14]. The main advantage of intensity-based registration is that it does not require the segmentation of CT and X-ray images. By averaging out all the available information, it reduces the influence of outliers and is thus more robust to the presence of foreign objects and surrounding anatomy.

However, intensity-based registration is computationally expensive. It requires generating many Digitally Reconstructed Radiographs (DRRs) for comparison with the original X-ray images and searching a 6-dimensional space which can have many local minima and is very sensitive to the similarity measure. It requires an initial bone pose estimate close to the final one. Proposed methods to overcome these limitations include speeding up DRR generation [11], computing DRRs in small Regions of Interest (ROIs) [9], approximating or subsampling DRRs [10], defining new similarity measures [13], and guiding the search with stochastic optimization [14]. To date, the Cyberknife [4] is the only system in clinical use based on intensity-based registration [9]. The main obstacles are robustness, accuracy, efficiency, and system integration.

3 Effective Intensity-Based Registration

We address the entire registration process, which includes orientation-dependent fluoroscopic X-ray C-arm calibration [15] and optical tracking. Our approach is generic, customizable, and targeted to routine clinical use in orthopaedics. Following an in-depth analysis of the most common orthopaedic procedures, we set our goal to obtain a registration with an overall mean surface Target Registration Error (sTRE) of 1mm (max=2mm), 95% of the time or more, in less than two minutes on a CT data set with 1mm thick slices 1.5mm apart and 2–5 fluoroscopic X-ray images from different viewpoints, with foreign objects and surrounding anatomy, with viewpoint localization of 0.3mm (max=0.5mm).

We improve upon existing intensity-based registration methods and successfully combine their strengths to overcome most of the speed, robustness, and accuracy problems (see [16] for a complete description).

Table 1. The role of each technique on the effectiveness of intensity-based registration.

TECHNIQUE	SPEED	ROBUSTNESS		ACCURACY
		range	noise	
1. Dynamic ROIs	yes			some
2. Multiresolution fluoroscopy	yes			
3. Transgraph	yes			
4. NCC and VLNC measures		yes	yes	yes
5. Genetic search		yes		

The registration consists of: 1. initial bone pose estimation, 2. precomputation, and 3. intensity-based registration. The initial bone pose can be determined from the clinical setup, estimated from skin markers, or computed from a few surgeon-defined matching landmarks on the CT and fluoroscopic X-ray images. The precomputation step generates a Transgraph [11], a data structure for fast DRR generation, and a multiresolution hierarchy of fluoroscopic X-ray images. Intensity-based registration is performed in three steps: 1. generation of DRRs for each C-arm viewpoint; 2. bone pose difference measurement by comparing the DRRs with the real fluoroscopic X-ray images, and 3. computation of the new bone pose that minimizes the difference. The process is repeated with the new bone pose estimate until no further progress can be made. To make the algorithm effective, we incorporate into it five special techniques (Table 1).

To speed up the comparison between fluoroscopic X-rays and DRRs, it is performed only on small, dynamically selected ROIs. The key observation is that uniform regions with small pixel value variance, which usually correspond to background, soft tissue, and bone interior, add very little information for matching and can thus be ignored. High-variance regions usually correspond to bone exterior and interior contours and to pronounced thickness changes; they are far fewer (10–15% of the image pixels) and contain more information. High-variance regions are automatically identified once on the original fluoroscopic X-ray images by selecting small, non-overlapping rectangular windows (e.g., 7x7 pixels) whose normalized gray value variance is above a prespecified threshold (Fig. 1). In addition, pixels corresponding to the calibration grid spheres and to the image margins are not included. Dynamically selected ROIs improve upon random pixel sampling [14] and upon predefined static ROIs [9]. These ROIs increase the final accuracy because they focuses the registration on informative data. To further reduce the number of pixels, we use a multiresolution hierarchy of fluoroscopic X-ray images. The original images are downsampled and stored, and the DRR computation is performed at the appropriate resolution. We use a two-level 1:16 downsampling in our implementation.

The DRRs for the estimated bone pose of selected ROIs are generated using a Transgraph, a data structure of gray-level ray values from expected viewing directions [11]. Its indexing planes are computed from an initial bone pose estimate and CT volume size. A data structure stores the gray-level values of rays in the expected viewing directions. The gray value of each DRR pixel is computed by quadrilinear interpolation with just 45 arithmetic operations. In our

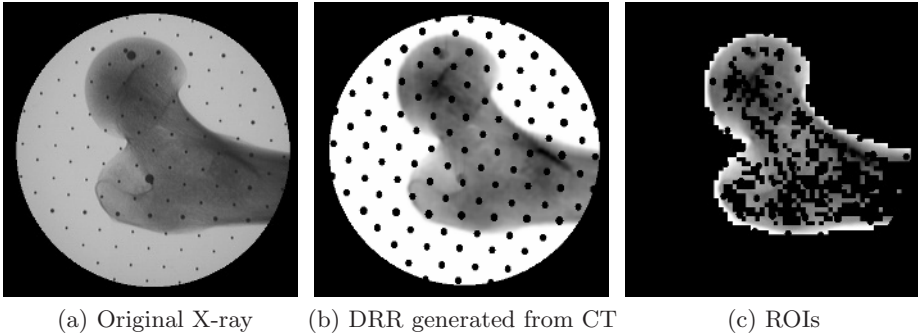


Fig. 1. Illustration of ROIs. In (c), black pixels do not belong to the ROIs.

implementation, a 800×600 DRR is generated in about 0.08secs on a 30MB Transgraph consisting of 15M rays constructed from $100 \times 70 \times 150\text{mm}^3$ CT volume, computed in about 5 minutes.

To robustly and accurately compare the DRRs and the fluoroscopic X-ray images, we use a two-step comparison. The similarity measure should emphasize rigid bone, filter out deformable soft tissue, and minimize image contrast, and sharpness differences. It should filter out surgical tools and surrounding anatomy that are present in the fluoroscopic X-ray images but not in the CT. As observed in [12] and based on our own experiments, a single similarity measure cannot simultaneously be fast, robust, accurate, and have a wide convergence range. Instead, we apply the Normalized Cross Correlation (NCC) comparison measure first and then automatically switch to an improved Variance Weighted Sum of Local Normalized Correlation (VLNC) [11] when NCC leads to convergence. The NCC measure is invariant to linear changes in image intensity, has a wide convergence range, and is robust, but is not very accurate (Fig. 3a). The VLNC measure weights more high-variance regions where the relevant information is, so it improves the accuracy (Fig. 3b). It also reduces the influence of regions containing foreign objects since their local NCC values are small when the estimated bone pose is close to the actual one.

To effectively search for the final pose, we use the iterative downhill simplex optimization method. The objective function is the sum of comparison measures (NCC or VLNC) for each C-arm viewpoint. To avoid local minima, we use a genetic search method. The search starts from the initial guess and n randomly generated guesses. For each guess, its transformation is computed by solving an optimization problem. The transformations yielding similarity values that are too far from the best one are discarded, and the process is repeated with $n/2$ new transformations obtained by random pairwise linear combinations of the best ones. The process stops when one transformation is left.

The algorithm effectively combines all the above techniques in three successive refinement steps. It starts from the initial pose guess plus n additional randomly generated poses on entire downsampled X-ray images and uses the NCC measure. It proceeds from the computed pose guess plus n additional ran-

Table 2. Summary of experimental results. Each entry shows the mean (maximum) surface target registration error (sTRE) in millimeters. Each scenario (ideal, realistic, and bad) is defined by the CT slice spacing Δh and the presence of foreign objects in the fluoroscopic X-ray images (none, some). Three fluoroscopic X-ray images were used in all cases. Computation times are 20–100 secs.

Data set	Ideal	Realistic	Bad
	$\Delta h = 0.6\text{mm}$, none	$\Delta h = 2.4\text{mm}$, some	$\Delta h = 4.2\text{mm}$, some
SIMULATION			
1. clinical pelvis	—	0.3 (0.7)	0.5 (0.7)
IN VITRO			
3. dry spine	0.3 (0.4)	0.4 (0.7)	0.7 (1.5)
2. dry femur	0.7 (1.1)	0.9 (1.8)	1.6 (3.4)
CADAVER			
4. lamb hip	1.0 (1.1)	1.1 (1.8)	1.3 (2.4)

domly generated poses on ROIs downsampled X-ray images and uses the VLNC measure. It then computes the final transformation for the computed pose guess on ROIs full-resolution X-ray images and uses the VLNC measure. In our implementation, $n = 4$ and the registration takes on average 70 seconds and requires 3,000 DRRs: 5 seconds for initialization, 10 seconds and 400 DRRs in the first step, 30 seconds and 2,400 DRRs in the second step, and 25 seconds and 200 DRRs in the third step.

4 Experimental Results

We implemented the algorithm, incorporated it in the registration process, and validated it in three situations: 1. simulation experiments with clinical CT data and simulated fluoroscopic X-rays; 2. in-vitro experiments with dry bones, and; 3. a cadaver experiment. The simulation experiments establish the algorithm accuracy with no fluoroscopic X-ray imaging and no tracking errors. The in-vitro experiments establish the overall error for real CT, fluoroscopic X-ray images, and tracking under ideal conditions. The cadaver experiment emulates the surgical situation and establishes the expected navigation error. We demonstrate the genericity of our method by applying it to four anatomical structures: human femur, spine, pelvis, and lamb hip. Fig. 2 and Table 2 summarize the results.

We used a CT scanner (Marconi, USA), a 9" BV29 C-arm (Phillips, The Netherlands), a Polaris optical tracking camera (NDI, Canada), a FluoroTrax C-arm calibration ring and active optical trackers (Traxtal, Canada), and a Matrox Meteor II digital frame grabber. Processing was on a 2.4Ghz, 1GB RAM PC under Windows XP. All CT scans were at 0.6mm slice interval and 0.5mm slice thickness except for the pelvis CT.

We performed a simulation experiment on a clinical pelvis CT. We generated DRRs at known C-arm viewpoints to simulate fluoroscopic X-ray images. We offset the bone pose by 10mm sTRE, a typical initial guess. The resulting mean sTRE is 0.4mm (max=0.7mm) on CT data sets with 2.0mm and 4.0mm slice

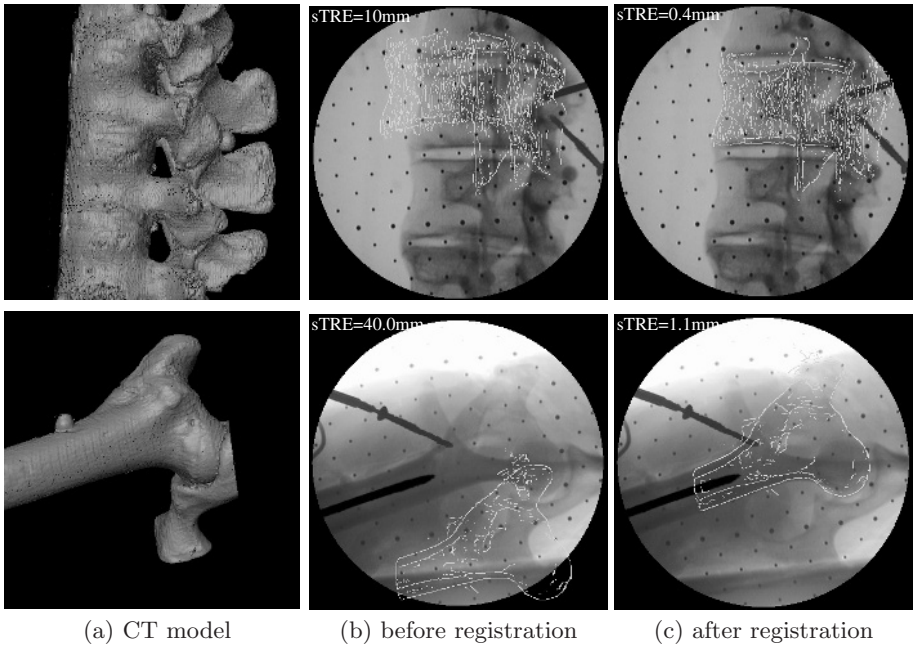


Fig. 2. Registration results: in-vitro dry spine and cadaver lamb hip. The first column shows the CT model. The second and third column show one fluoroscopic X-ray image with contours at initial and final pose superimposed on them (white lines).

interval with 1.5mm slice thickness. It is most likely the best possible simulation result obtainable at this CT resolution.

We performed in-vitro experiments on a single vertebra of a dry spine and on a dry proximal femur. First, we implanted seven 6mm aluminum spheres and CT scanned them. We extracted the sphere centers to a mean accuracy of 0.1mm (max=0.3mm). We then acquired two sets of three fluoroscopic X-ray images at various C-arm viewpoints, one with, and one without anatomy for optimal C-arm calibration. We performed C-arm calibration to a mean accuracy of 0.3mm (max=0.6mm), and obtained a ground-truth registration by fiducial contact-based registration on the spheres to a mean accuracy of 0.3mm (max=0.5mm). We then performed 2D/3D registration and compared the results. We obtained an average sTRE of 0.7mm, maximum of 1.3mm with 100% (95%) success for an initial sTRE pose offset of 0–25mm (0–40mm) for the femur and 0–10mm for the vertebra (0–15mm). The vertebra convergence range is smaller because of the presence of other vertebra. We observed a slight decrease (0.1mm for the femur) in accuracy when only two fluoroscopic X-ray images were used, an no significant increase with more than three images.

We performed a cadaver experiment on a fresh lamb hip with the same in-vitro protocol, except that we used only four fiducials. The small decrease in accuracy was due to a less accurate ground-truth registration, to the presence of surrounding anatomy, and to few salient bone features on the lamb femur. Fig. 3

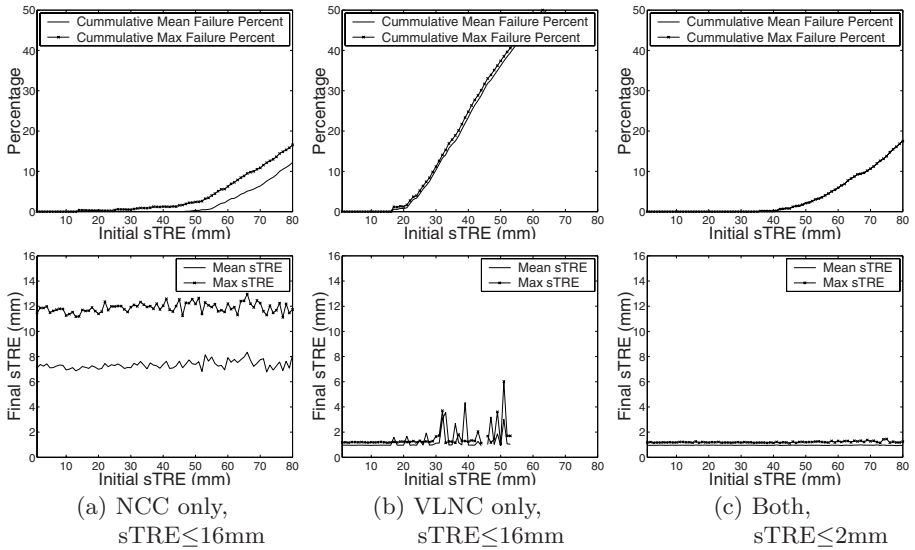


Fig. 3. Similarity comparison measures effect for the lamb hip experiment on the cumulative percentage of failures (top), and on the final accuracy (bottom) when using: (a) the NCC measure only; (b) the VLNC measure only, and; (c) both. The horizontal axis indicates the initial pose offset as the mean sTRE from the ground-truth transformation. The vertical axis indicates the cumulative percentage of failures (top) and the final mean and maximum sTRE.

shows the effect of the NCC and VLNC similarity measures on the cumulative percentage of failures and on the final accuracy. Running times are 40–60 secs for NCC only, 30–70 secs for VLNC only, and 70–90 secs for both. The combination of NCC and VLNC achieves 1mm accuracy with small deviation (max=0.3mm) and a 5% cumulative failure (an sTRE > 2mm is considered a failure) for initial sTRE poses of up to 55mm, well within the range of coarse initial pose estimation at the expense of a small time increase.

5 Conclusion

We conclude from our experimental results that anatomical image-based rigid registration between fluoroscopic X-ray and CT with an overall mean sTRE of 1mm (max=2mm), 95% of the time, with three X-ray images in less than two minutes in a realistic setup, is practically feasible. This is achieved by judiciously combining five techniques to enhance the speed, robustness, and accuracy of the basic intensity-based registration algorithm. Dynamic ROIs, two-step similarity measure, and genetic search are used for the first time in this context. The algorithm is fully automatic and does not rely on high-quality image segmentation. It proved robust in the presence of foreign objects and surrounding anatomy.

We are currently conducting further in-vitro and cadaver experiments and are incorporating the registration in an image-guided navigation system for trauma.

Acknowledgment. This research was supported in part by a grant from the Israel Ministry of Industry and Trade for the IZMEL Consortium on Image-Guided Therapy.

References

1. Maintz, J.B.A. and Viergever, M.A., A survey of medical image registration, *Medical Image Analysis* 2(1), 1998.
2. Smith, K., Frank, K., and Bucholz, R., The neurostation: a highly accurate, minimally invasive solution to frameless stereotactic neurosurgery, *Comp. Medical Imaging and Graphics* 18(1), 1994.
3. Nolte, L. P. and Ganz, R., *Computer Assisted Orthopedic Surgery*, Hogrefe&Huber Publishers, 1999.
4. Adler J. Jr., Murphy, M., Chang, S. and Hancock, S., Image-guided robotic radiosurgery, *Neurosurgery* 44(6), 1999.
5. Besl, P. and McKay, N. A method for registration of 3D shapes, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14(2), 1992.
6. Hamadeh, A., Lavallée, S. and Cinquin, P., 'Automated 3-dimensional computed tomographic and fluoroscopic image registration, *Comp. Aided Surgery* 3(1), 1998.
7. Guézic, A., Kazanzides, P., Williamson, B., and Taylor, R.H., Anatomy based registration of CT-scan and intraoperative X-ray images for guiding a surgical robot, *IEEE Transactions on Medical Imaging* 17(5), 1998.
8. Lemieux, L., Jagoe, R., Fish, R. *et al.*, A patient-to-computed-tomography image registration method based on DRRs, *Medical Physics* 21(11), 1994.
9. Murphy, M.J., An automatic six-degree-of-freedom registration algorithm for image-guided frameless stereotactic radiosurgery, *Medical Physics* 24(6), 1997.
10. Roth, M., Brack, C., Burgkart, R., and Czopf, A., Multi-view contourless registration of bone structures using a single calibrated x-ray fluoroscope, *Proc. of the Computer-Assisted Radiology and Surgery Conf.*, 1999.
11. LaRose, D.A, Bayouth, J., and Kanade, T., Transgraph: interactive intensity-based 2D/3D registration of X-ray and CT data, in *SPIE Image Processing*, 2000.
12. Penney, G.P., Weese, J., Little, J.A., Desmedt, P., Hill, D.L.G. and Hawkes, D., "A comparison of similarity measures for use in 2D/3D medical image registration," *IEEE Trans. on Medical Imaging*, 17(4), 1998.
13. Penney, G.P., Batchelor, P.G., Hill, D.L.G., Hawkes, D.J., and Weese, J., "Validation of a 2D to 3D registration algorithm for aligning preoperative CT images and intraoperative fluoroscopy images," *Medical Physics*, 28(6), 2001.
14. Zöllei, L., Grimson, W.E.L., Norbush, A., and Wells III, W.M., "2D-3D rigid registration of X-ray fluoroscopy and dCT images using mutual information and sparsely sampled histogram estimators," in *Proc. of IEEE Computer Vision and Pattern Recognition Conf.*, 2001.
15. Livyatan, H., Yaniv, Z., and Joskowicz, L. Robust automatic C-arm calibration for fluoroscopy-based navigation: a practical approach, *Proc 5th Int. Conf. on Medical Image Computing and Computer-Aided Intervention*, Tokyo, Japan, 2002.
16. Knaan, D. *Intensity-based 2D/3D rigid registration of fluoroscopic X-ray to CT*, MSc. Thesis, The Hebrew University of Jerusalem, Israel, June 2003.