Comparison of Rigid Gradient-Based 2D/3D Registration Using Projection and Back-Projection Strategies

Roman Schaffert[1,2], Jian Wang[1,2], Anja Borsdorf[2], Joachim Hornegger[1,3], Andreas Maier[1,3]

[1]Pattern Recognition Lab, FAU Erlangen-Nuremberg, Erlangen
[2]Siemens Healthcare GmbH, Forchheim
[3]Erlangen Graduate School in Advanced Optical Technologies, Erlangen
roman.schrom.schaffert@fau.de

Abstract. In this paper, a comparison of single-view gradient-based 2D/3D rigid registration methods is presented. To achieve dimensional correspondence between the images, projection and back-projection strategies have been proposed in the literature. Two similarity measures that are applicable for both strategies are involved in the comparison. Extensions of the similarity measure are proposed and compared to the original proposals. It is demonstrated that the projection strategy achieves a median accuracy up to 0.8 mm, which outperforms the back-projection strategy with a median accuracy of up to 1.1 mm. Our extension of the covariance-based similarity measure in combination with the back-projection strategy achieves the highest convergence range (up to 34.0 mm), while the the maximum achieved convergence range for the projection strategy is 31.3 mm.

1 Introduction

In many interventional procedures, C-arm systems are used to acquire live X-ray images, which are used as a guidance during the procedure. However, some important structures are not visible in these images. Pre-operative 3D images can be used to create an overlay showing the structures. An accurate alignment of 2D and 3D images is crucial. In order to achieve this, 2D/3D registration methods are frequently used. A thorough overview of different methods is given in [1].

Gradient-based methods achieve high accuracy [2] without the need of artificial markers. A projection-based method is described in [3], where the gradients of the volume are projected and compared to the gradients of the 2D image. A back-projection-based method is introduced in [4], where bone surfaces are extracted and the gradient direction is compared in 3D to back-projected gradients. This method is extended in [5], where 3D gradients are reconstructed from a set of 2D images. A rough registration step is also introduced which is less accurate but has a higher capture range (CR). In [6], covariance matrices are

T. Tolxdorff et al. (Hrsg.), *Bildverarbeitung für die Medizin 2016*,
DOI 10.1007/978-3-662-49465-3_26, © Springer-Verlag Berlin Heidelberg 2016

used to incorporate the neighborhood information into the similarity measure. The similarity is evaluated at a set of voxels with the highest gradient values.

The remaining paper is organized as follows: the details of the compared registration methods and the similarity measure extensions are discussed in Sec. 2 as well as the evaluation methodology and the used data sets. The results are presented in Sec. 3, followed by a discussion in Sec. 4.

2 Materials and methods

The registration is performed by defining a similarity measure which is high for correct alignment of images and decreases with higher misalignment. The rigid transformation $\mathrm{T_{reg}}$ is searched which maximizes the measure. For the projection strategy, the gradients of the 3D image $\mathcal{V}$ are projected onto the image plane as derived in [4]. Ray casting on the GPU is used to obtain the projection image. For the back-projection strategy, the gradients in the 2D image $\mathcal{I}$ are back-projected onto an auxiliary plane $\mathcal{P}$ orthogonal to the ray direction $\mathbf{e}_\mathrm{r}$ [4].

2.1 Similarity measures

In [4], a similarity measure based on the angle between two gradients is proposed. It is based on the intuition that if the images are aligned, the gradients should point into the same direction. The measure is computed as

$$\mathrm{DS} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathbf{g}_{\mathrm{a,i}}| \cdot |\mathbf{g}_{\mathrm{b,i}}| \cdot \mathrm{f}(\alpha)}{\sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathbf{g}_{\mathrm{a,i}}| \cdot \sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathbf{g}_{\mathrm{b,i}}|} \quad \text{where } \mathrm{f}(\alpha) = \begin{cases} \cos^n(\alpha) & \text{if } |\alpha| \leq 90^\circ \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $\mathbf{g}_{\mathrm{a,i}}$ and $\mathbf{g}_{\mathrm{b,i}}$ denote the compared gradients and α the angle between these. As in [4], $\mathrm{n} = 2$ is used in all experiments.

However, this measure is not invariant to the number of points used. This can be demonstrated for the case where the compared gradients have identical direction and identical magnitude c. For N point pairs, the measures is

$$\mathrm{DS} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{c} \cdot \mathrm{c} \cdot 1}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{c} \cdot \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{c}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{c}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{c} \cdot \sum_{\mathrm{i}=1}^{\mathrm{N}} 1} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{c}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{c} \cdot \mathrm{N}} = \frac{1}{\mathrm{N}} \tag{2}$$

Therefore, in addition to the original DS measure, the measure DS_P is introduced, which is the weighted mean of $\mathrm{f}(\alpha)$ at the points. It is expressed as

$$\mathrm{DS\_P} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathbf{g}_{\mathrm{a,i}}| \cdot |\mathbf{g}_{\mathrm{b,i}}| \cdot \mathrm{f}(\alpha)}{\sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathbf{g}_{\mathrm{a,i}}| \cdot |\mathbf{g}_{\mathrm{b,i}}|} \tag{3}$$

In [6], a similarity measure is introduced which uses covariance matrices containing the gradient information in the neighborhood of a point. Both the covariance matrix from the volume as well as the 2D image are projected onto the auxiliary plane $\mathcal{P}$ described before. The resulting covariance matrices are denoted $\mathbf{C}_\mathcal{V}$ and $\mathbf{C}_\mathcal{P}$. The covariance matrices are computed for the neighborhood Ω_p. We choose the size of $\Omega_\mathbf{p}$ as three times the pixel size in $\mathcal{I}$ for both the projection and back-projection strategies. The similarity measure is computed using the normalized tensor scalar product. Contrary to [6], the number of points is variable. To account for this, the resulting similarity measure is computed as

$$\mathrm{CS} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{Trace}(\mathbf{C}_{\mathrm{a,i}}\mathbf{C}_{\mathrm{b,i}})}{\mathrm{Trace}(\mathbf{C}_{\mathrm{a,i}})\mathrm{Trace}(\mathbf{C}_{\mathrm{b,i}})} \tag{4}$$

For the projection strategy, the covariance matrices are compared for every pixel and no projection onto an auxiliary plane is necessary.

The covariance matrices preserve the gradient orientation, but not the direction. To force the gradients to point into the same direction, the similarity measure is extended by setting the similarity to 0 if the direction of the gradients at the central points varies by more than 90°. The modified measure is

$$\mathrm{CS\_O} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{Trace}(\mathbf{C}_{\mathrm{a,i}}\mathbf{C}_{\mathrm{b,i}})}{\mathrm{Trace}(\mathbf{C}_{\mathrm{a,i}})\mathrm{Trace}(\mathbf{C}_{\mathrm{b,i}})} \cdot \mathrm{d} \text{ where } \mathrm{d} = \begin{cases} 1 & \text{if } |\alpha| \leq 90^\circ \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

2.2 Point selection

A set of points in $\mathcal{V}$ is selected for the back-projection strategy. The points should correspond to structures which are also present in $\mathcal{I}$. Therefore, points on bone surfaces are extracted as in [4] using the canny edge detector. The gradient magnitude and hysteresis thresholds are denoted $\mathrm{t_g}$ and $\mathrm{t_h}$. Additionally, minimum and maximum intensity thresholds $\mathrm{t_m}$ and $\mathrm{t_M}$ are introduced in this work to select specific tissue classes. A minimum connected component size $\mathrm{t_N} = 1000$ is introduced to exclude small surfaces which are present due to noise.

If the points are selected only based on the 3D structures in $\mathcal{V}$, there will be points for which no large gradient is observed in the 2D image, e.g. points on surfaces parallel to the image plane. Generally, the correspondence of the gradients is good if the points are located on the occluding contour of the 3D surface, i.e. if the ray going through the point is tangential to the surface. In this case, the ray is perpendicular to the gradient in the volume. Therefore, surface points are selected which satisfy $\angle(\mathbf{g}_\mathrm{i}, \mathbf{e}_\mathrm{r}) \geq 90^\circ - \mathrm{t}_\gamma$ [7]. For the experiments t_γ is set to 4°. It was found that on sharp edges (e.g. the top and bottom of ribs) many points are missing due to high gradient direction changes and the discrete nature of $\mathcal{V}$. Therefore, the surface points are classified as edge points or not based on the covariance matrix $\mathbf{J}$ of the neighborhood $\Omega_\mathbf{J}$ around the point. If $\lambda_{\mathrm{J},2}/\lambda_{\mathrm{J},1} \geq \mathrm{t_e}$ holds ($\lambda_{\mathrm{J},1}$ and $\lambda_{\mathrm{J},2}$ are the largest and second largest eigenvalues of $\mathbf{J}$), the point belongs to an edge and a higher threshold $\mathrm{t}_{\gamma,\mathrm{e}} = 20^\circ$ is used.

2.3 Optimization

The similarity measure is optimized using gradient descent and a multi-scale approach with 5 scale levels. Both $\mathcal{V}$ and $\mathcal{I}$ are scaled by the factor 0.5 for each scale level. As high frequencies are lost when the images are down-sampled, the used gradient thresholds are also scaled by the factor 0.5 for each resolution level. The neighborhood $\Omega_{\mathbf{p}}$ and $\Omega_{\mathbf{J}}$ are scaled by the factor 2 to cover a larger region of the image. The threshold $\mathrm{t_N}$ is multiplied by the factor $0.5 * 0.5$, as the surface is a 2-dimensional structure. If occluding contour point selection is used, the angle threshold is scaled by the factor 2. This is based on the intuition that on lower resolution levels, coarser structures are involved and therefore thicker contours are used. $\mathrm{t}_{\gamma,\mathrm{e}}$ is not modified as it reflects the threshold needed on sharp edges to avoid missing points.

2.4 Evaluation setup

Error computation Two crucial properties of a registration method are the accuracy and the ability to find the correct alignment for a given initial misalignment. The accuracy can be measured by defining a set of points in $\mathcal{V}$ and measure the distance of that points for the estimated $\mathrm{T_{reg}}$ and the ground truth transformation $\mathrm{T_{gt}}$. Evaluating the mean distances in 3D space leads to the mean target registration error (mTRE) [8]. As for overlays, the error in the image plane is relevant, the points are projected before computing the error, which leads to the mean projection error (mPE) [7]. Points are uniformly distributed in $\mathcal{V}$ at 50 mm intervals and only used if inside $\mathcal{I}$ for $\mathrm{T_{gt}}$.

The CR indicates up to which misalignment a correct registration can be expected. For a correct registration, a maximum resulting error threshold of 5 mm is defined. The CR is defined as the minimum initial error for which less than a given fraction of registration with smaller or equal initial errors are successful. The mTRE error is often used as the initial error, e.g. in [6]. However, as mPE is used as the resulting error in this work, it is also used for the initial error computation. This has the advantage that the effect of projecting the error, which decreases the error in most cases, does not bias the evaluation. The accuracy is given as the median error for all successful cases below the CR. The median is used as it is insensitive to low accuracy cases which are introduced near the CR due to the cumulative computation of the CR.

Initial transformation generation To compute the CR, the initial errors have to be uniformly distributed. Similar to [8], a set of initial transformations is created which satisfies this. For each interval of 1 mm in the initial error, three transformations are generated. The CR is computed on a per-mm basis.

Data A thorax phantom and a head phantom are used in the experiments. The thorax data set contains a CT volume and three fluoroscopic images. The values $\mathrm{t_m} = 300$, $\mathrm{t_M} = 5000$, $\Omega_{\mathbf{J}} = 2\,\mathrm{mm}$ and $\Omega_{\mathbf{p}} = 0.76\,\mathrm{mm}$ are used. The ground truth is obtained using an Optotrak stereo vision system. The head data set contains a C-arm-CT volume and three X-ray images with contrasted vessels in one hemisphere. The values $\mathrm{t_m} = 300$, $\mathrm{t_M} = 10000$, $\Omega_{\mathbf{J}} = 1\,\mathrm{mm}$ and $\Omega_{\mathbf{p}} = 0.73\,\mathrm{mm}$ are used. A higher $\mathrm{t_M}$ compared to the thorax phantom is used

Table 1. Convergence Range (CR) / median accuracy for successful cases below CR. Bold values denote the best results for given strategy to achieve dimensional correspondence and similarity measure type (DS-based or CS-based, all values in mm).

	Thorax Phantom			Head Phantom		
	Surface	Occ. Cont.	Projection	Surface	Occ. Cont.	Projection
DS	17.0 / 1.8	14.0 / 1.9	20.3 / **1.1**	11.7 / 1.9	**12.3** / **1.4**	**24.3** / **0.8**
DS_P	12.0 / **1.6**	**18.7** / 1.7	**23.3** / 1.2	12.0 / 1.6	10.0 / 1.7	22.0 / **0.8**
CS	16.7 / 1.8	17.7 / **1.2**	18.3 / 1.5	5.7 / 4.4	19.3 / 2.5	**31.3** / 1.4
CS_O	**26.7** / 1.7	23.7 / 2.1	**22.0** / **1.1**	**34.0** / **1.3**	25.3 / **1.3**	28.0 / **0.8**

to include the contrasted vessels. A reference method was used to estimate T_{gt}. The given results are the mean values for the used fluoroscopic images of one data set.

3 Results

Tab. 1 shows the CR and median accuracy values. The projection strategy provides higher accuracy (error down to 0.8 mm for the head phantom) compared to the back-projection strategy (down to 1.2 mm for the thorax phantom). The accuracy is comparable for all measures except CS, which leads to an error increase of about 0.3 mm for the thorax and 0.6 mm for the head phantom. Considering only projection-based methods, the DS_P measure achieves the highest CR of 23.3 mm for the thorax phantom. For the head phantom, the highest CR of 31.3 mm is achieved using the CS measure. The extended measures lead to an increased CR for the thorax and decreased CR for the head phantoms.

For the back-projection strategy, the CS_O measure with selecting all surface points leads to the highest CR for both data sets, which is up to 34.0 mm for the head data set. The selection of occluding contours leads to better results in case of the CS measure in both data sets, while being outperformed by the extended CS_O measure selecting all points.

For two of the used X-ray images of the head data set, a CR of 0 mm to 3 mm is often achieved, leading to an overall low CR. It was observed that this is both due to the fact that vessels in $\mathcal{I}$ are aligned to bony structures in $\mathcal{V}$ and that only structures containing many points are aligned correctly. Both errors are caused by the fact that structures containing many points have an over-proportionally high influence on the similarity measure.

4 Discussion

Using the extended CS_O measure and the back-projection strategy, the highest CR is achieved. This leads to the conclusion that using neighborhoods of selected points is advantageous for the registration and that the measure has to be designed carefully not to neglect a part of the available gradient information.

This is further illustrated by comparing the CS-based measures. While the use of contour points is of advantage for the CS measure, it leads to a lower CR for the CS_O measure. This indicates that occluding contour selection can discard some of the wrong matches of gradients with opposite directions by considering only object outlines. However, the CS_O measure handles these cases more effectively and therefore is able to make use of points not located on a contour.

The projection strategy achieves higher accuracy compared to the back-projection strategy, while the measure leading to the highest CR depends on the data set. One explanation is that as every pixel is used, no advantage can be gained from using neighborhoods in the general case. Possible reasons for the increased accuracy are the noise suppression without blurring the image by the integration of gradients along a ray and a more uniform weighting of different regions compared to the selected points in the back-projection strategy.

As the highest CR is achieved using the back-projection strategy and misregistration is often caused by the distribution of points, improved point-selection methods and weighting strategies for the points are subject to future research with the main goal to improve the CR.

Acknowledgement. This work was partly supported by the Research Training Group 1773 “Heterogeneous Image Systems”, funded by the German Research Foundation (DFG).

References

1. Markelj P, Tomaževič D, Likar B, et al. A review of 3D/2D registration methods for image-guided interventions. Med Image Anal. 2012;16(3):642–61.
2. Gendrin C, Markelj P, Pawiro SA, et al. Validation for 2D/3D registration II: the comparison of intensity- and gradient-based merit functions using a new gold standard data set. Med Phys. 2011;38(3):1491–502.
3. Wein W, Roeper B, Navab N. 2D/3D registration based on volume gradients. Proc SPIE. 2005;5747:144–50.
4. Tomaževič D, Likar B, Slivnik T, et al. 3-D/2-D registration of CT and MR to x-ray images. IEEE Trans Med Imaging. 2003;22(11):1407–16.
5. Markelj P, Tomaževič D, Pernuš F, et al. Robust gradient-based 3-D/2-D registration of CT and MR to x-ray images. IEEE Trans Med Imaging. 2008;27(12):1704–14.
6. Špiclin Ž, Likar B, Pernuš F. Fast and Robust 3D to 2D image registration by backprojection of gradient covariances. Biomed Image Regist. 2014;8545:124–33.
7. Wang J, Borsdorf A, Heigl B, et al. Gradient-based differential approach for 3-D motion compensation in interventional 2-D/3-D image fusion. Proc Int Conf 3D Vis. 2014;1:293–300.
8. van de Kraats EB, Penney GP, Tomaževič D, et al. Standardized evaluation methodology for 2-D-3-D registration. IEEE Trans Med Imaging. 2005;24(9):1177–89.