H.U. Lemke, M.W. Vannier, K. Inamura, and A.G. Farman (eds.), *Proc. Computer Assisted Radiology and Surgery 1999 (CARS'99), Paris, France, 23-26 June 1999*, pp. 233–237.
Elsevier Science, Amsterdam, 1999.

Validating Point-based MR/CT Registration Based on Semi-automatic Landmark Extraction

S. Frantz[a], K. Rohr[a], H.S. Stiehl[a], S.-I. Kim[a], and J. Weese[b]

[a]Universität Hamburg, Fachbereich Informatik, Arbeitsbereich Kognitive Systeme, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany

[b]Philips Forschung, Technische Systeme, Röntgenstr. 24–26, 22335 Hamburg, Germany

Manually extracting 3D anatomical point landmarks from images is generally time-consuming and usually lacks reproducibility. To improve on this, we developed a semi-automatic procedure for landmark extraction. In this paper, we report on the validation of our procedure for the case of point-based rigid registration of MR/CT head images. The results of our semi-automatic procedure are compared with those of a manual procedure for landmark extraction. Five observers participated in the study. Five MR/CT image pairs were used from which 7–9 landmarks were extracted using both procedures. Our experiments show that (a) the elapsed time spent for landmark extraction can significantly be reduced with the semi-automatic procedure, (b) the registration results of both procedures generally have similar quality, and (c) the inter-observer variability in the localized landmark positions is smaller with the semi-automatic procedure.

1. INTRODUCTION

One general, intuitive, and efficient approach to registering 3D medical images employs corresponding image points to compute a spatial transformation that maps the different images onto each other. For this purpose, 3D anatomical point landmarks are often used. However, manually extracting such landmarks from images is generally time-consuming and usually lacks reproducibility. To improve on this, we developed a semi-automatic procedure for landmark extraction. A semi-automatic procedure enables the physician to interactively control the results, which we consider crucial in clinical routine. In this paper, we report on the validation of our procedure for the case of point-based rigid registration of MR/CT head images. The results of the semi-automatic procedure are compared with those of a manual procedure for landmark extraction. By contrast, Strasters *et al.* [1] validated MR/CT registration based on manual landmark extraction only.

2. VALIDATION STRATEGY

To assess the performance of our semi-automatic procedure for landmark extraction, we used the following criteria: (a) the elapsed time spent for landmark extraction, (b) the registration results, and (c) the reproducibility of the results for different observers. Five observers participated in the study. Although none of the observers can be considered a clinical expert, each observer had sufficient anatomical knowledge and experience with the task at hand.

In the experiments, landmarks were simultaneously extracted from both modalities. For this purpose, two workstations were arranged side by side, each of which showing a single-modality image in orthogonal views. The user-interfaces for semi-automatic and manual landmark extraction were nearly identical. The elapsed time spent for landmark extraction was automatically recorded by the system. After extracting corresponding landmarks from both modalities, each observer subjectively ranked his confidence in the localized positions. Then, the images were registered using a least-squares rigid transformation [2]. After inspecting the registration result, each observer was allowed to reject up to two outlier landmarks and to redo registration using the reduced landmark set. The basis for this were the transformed MR image fused with edges of the CT image, the landmark rankings, and the landmark registration error. To ensure independence, the interval between the experiments of each observer was at least four weeks. To prevent a bias to the semi-automatic procedure, each observer started with this procedure.

After completing all experiments, we assessed the results of both procedures. However, drawing conclusions w.r.t. the possible reduction in the elapsed time spent for landmark extraction is problematic. The hypothesis we use is that the measured relative reduction, $\triangle\tau$, in the elapsed time is comparable with that of a clinical expert. Since for the data used in the study a 'gold standard' determining the correct transformation was not available to us, we chose a more subjective approach to evaluating the registration results: first, we visually assessed the registration results of both procedures, comparing the transformed MR image fused with edges of the CT image. Second, we inspected the root-mean-squared (RMS) landmark registration error, $e_{\mathrm{RMS}} = \sqrt{1/n \sum_{i=1}^{n} ||\mathbf{e}_i||^2}$. Here, n is the number of landmarks used for registration, and $\mathbf{e}_i$ denotes the residual error at the i-th landmark. Third, in the case of one particular image pair, we used 22 landmarks determined independently by an expert ([1]) to estimate the registration error. To study the reproducibility of the results of both procedures, we analyzed the inter-observer variability in the localized landmark positions on the basis of the RMS distances, $d_{\mathrm{RMS,sem}}$ and $d_{\mathrm{RMS,man}}$, to the respective mean landmark positions.

3. SEMI-AUTOMATIC LANDMARK EXTRACTION

Our interaction scenario for semi-automatic landmark extraction is as follows: (i) the observer interactively determines the landmark position coarsely, (ii) an algorithm is applied within a region-of-interest (ROI) to automatically detect landmark candidates, and (iii) the observer selects one candidate. Next, we briefly describe the constituents of the used algorithm for landmark extraction. For details we refer to [3–5].

Selecting a suitable ROI. The ROI should be large enough to enable reliable detection of the landmark at hand. However, the ROI should not include neighboring structures to avoid additional false detections. Addressing this problem, we introduced a statistical differential approach to selecting a suitable ROI size automatically [4,5]. The approach tries to scale a cubic ROI at the interactively determined position so that the landmark at hand, e.g., a tip, is isolated.

Detecting landmarks. We use a computationally efficient 3D differential operator introduced in [3]. The operator is applicable to different types of landmarks and is relatively robust w.r.t. noise (see [6] for a comparative performance study of various differential operators for landmark detection). The operator reads $Op3 = det(\mathbf{C})/trace(\mathbf{C})$, where $\mathbf{C}$ denotes the averaged dyadic product of the intensity gradient.

Reducing false detections. We use prior knowledge of the intensity structure at a landmark to impose additional constraints on the candidates [5]. Using curvature properties of the isointensity surfaces at the detected candidates, we distinguish between saddle points and tips as well as between tips of dark and bright structures w.r.t. the background. Candidates with an inconsistent intensity structure are rejected automatically.

Building on the work in [7], the partial derivatives of the intensity function are estimated using a scheme based on cubic B-spline image interpolation and Gaussian smoothing. Anisotropic image resolution is correctly taken into account in estimating derivatives [5].

4. USED DATA AND LANDMARKS AND PARAMETER SETTINGS

In our study, we used five MR/CT image pairs from different patients: one (*C06*) acquired at Utrecht University Hospital and four (*V101, V104, V107,* and *V109*) acquired at Vanderbilt University. The voxel sizes in the *C06* data are $0.86 \times 0.86 \times 1.2\text{mm}^3$ (MR) and $0.63 \times 0.63 \times 1.0\text{mm}^3$ (CT). The voxel sizes in the remaining original data are about $0.85 \times 0.85 \times 3.0\text{mm}^3$ (MR) and $0.42 \times 0.42 \times 3.0\text{mm}^3$ (CT) (however, we used up-sampled data with a slice thickness of 1.0mm based on cubic B-spline image interpolation [7]).

As landmarks, we used visually salient anatomical features located on the skull and within the brain: the saddle points at the zygomatic bones (MC15), the tip of the external protuberance (MC5e), the topmost concavity of fourth ventricle roof (MC2), the junction at the upper end of pons (MC18), the tips of the frontal (MC6) and occipital horns (MC7) of the ventricular system. Depending on the field-of-view, lesions, and the image quality, we ended up with 7–9 landmarks for each image pair.

The minimum and maximum width of the cubic ROI was set to 7 and 21 voxels, resp. In between, the optimal width was automatically selected for each landmark. Depending on the scale of the respective landmark, we used two different scales for the Gaussian derivative filters: $\sigma = 1.0$mm for MC15 and MC18 and $\sigma = 1.5$mm for MC5e, MC2, MC6, and MC7 (however, these scales were adapted according to the voxel size, see [5]). Averaging the gradient was done within a 5^3 window. Local maxima of the operator responses were determined in 3^3 neighborhoods. To suppress detections with insignificant operator responses, we applied a dynamic threshold (10% of the maximum operator response).

5. VALIDATION RESULTS AND DISCUSSION

Elapsed time spent for landmark extraction. In Tab. 1, the relative difference, $\triangle\tau$, between the elapsed time needed with the semi-automatic procedure, τ_S, and the manual procedure, τ_M, to extract the suggested landmarks is given for each observer ($\triangle\tau = (\tau_M - \tau_S)/\tau_M \cdot 100\%$). Additionally, for each observer the relative difference between the total elapsed time spent for all image pairs is given ($\triangle\tau_{\text{total}}$). In all cases, the elapsed time was shorter with the semi-automatic procedure (in part significantly shorter for different observers). Since the strategies for exploring the data and the extent of mastering the user-interface were different for the different observers, the measured elapsed time and

Table 1
Relative differences $\triangle\tau$ and $\triangle\tau_{\text{total}}$ between the elapsed time needed with the semi-automatic procedure and the manual procedure to extract the suggested landmarks.

Data	*Obs. 1* $\triangle\tau$	*Obs. 2* $\triangle\tau$	*Obs. 3* $\triangle\tau$	*Obs. 4* $\triangle\tau$	*Obs. 5* $\triangle\tau$
C06	66%	57%	28%	8%	25%
V101	50%	34%	36%	19%	3%
V104	59%	29%	44%	30%	56%
V107	59%	34%	21%	35%	53%
V109	34%	26%	34%	24%	37%
$\triangle\tau_{\text{total}}$	53%	35%	33%	24%	38%

hence $\triangle\tau$ varied. The mean elapsed time needed with the semi-automatic procedure was 11'30 minutes, while manual landmark extraction took in the mean 18'28 minutes (mean of the corrected elapsed time needed to extract 9 corresponding landmarks from both modalities). Thus, the achieved mean relative reduction in the elapsed time was 38%.

Registration results. In most cases, the different observers used all extracted landmarks for registration. Comparing the registration results of both procedures visually, we found that the results had similar quality. Fairly good registration results were obtained for *C06* (with the exception of one observer) and *V104*. The registration results for *V107* and *V109* in part showed a small error for both procedures. The registration results for *V101* in part showed a larger registration error. Here, the results of the manual procedure were slightly better than those of the semi-automatic procedure. In Tab. 2, the RMS registration errors e_{RMS} and the respective maximum error $e_{\max}$ (both in mm) are given for each observer and each MR/CT image pair. The mean of these values, $\overline{e}_{\text{RMS}}$, is also given. Note that in the case of *C06* the error measures were computed based on 22 landmarks determined independently by an expert ([1]). In the other cases, the error measures were computed based on the landmarks used for registration. In the case of *C06*, for three observers

Table 2
RMS registration error e_{RMS} and maximum error $e_{\max}$ (both in mm). See text.

Data	*Landmark extraction*	*Obs. 1* $e_{\text{RMS}}/e_{\max}$	*Obs. 2* $e_{\text{RMS}}/e_{\max}$	*Obs. 3* $e_{\text{RMS}}/e_{\max}$	*Obs. 4* $e_{\text{RMS}}/e_{\max}$	*Obs. 5* $e_{\text{RMS}}/e_{\max}$	$\overline{e}_{\text{RMS}}$
C06	Semi-automatic	2.09/4.57	2.41/5.20	2.09/4.57	2.09/4.57	3.56/6.44	2.45
	Manual	2.26/4.70	2.09/4.79	2.14/5.13	2.26/5.07	2.67/5.80	2.28
V101	Semi-automatic	2.45/3.45	3.24/3.96	2.80/4.54	2.73/3.70	3.40/6.51	2.92
	Manual	1.44/2.04	3.61/6.30	2.59/4.53	2.21/2.96	1.55/2.74	2.28
V104	Semi-automatic	1.69/3.46	1.99/3.70	3.54/5.40	1.99/3.70	1.68/2.93	2.17
	Manual	1.66/2.39	1.20/1.94	3.04/4.75	1.75/2.39	1.51/2.34	1.83
V107	Semi-automatic	2.81/4.78	1.97/2.78	2.92/5.09	1.54/2.19	2.81/4.41	2.41
	Manual	1.96/2.72	2.22/3.19	2.44/3.57	2.13/2.87	1.81/2.76	2.11
V109	Semi-automatic	2.67/3.64	3.42/5.74	2.34/3.97	2.56/3.68	2.99/5.16	2.79
	Manual	1.81/2.66	1.97/2.87	2.83/4.24	1.78/2.85	2.10/2.96	2.09

the registration error is smaller for the semi-automatic procedure. However, the mean of these values is slightly smaller for the manual procedure due to the results of Observer 5.

The relatively large maximum error, which occurred in all cases at the same landmark, indicates that the 'ground truth' positions for this landmark are possibly erroneous. The computed values for the other images convey reasonable registration results for both procedures, which was additionally confirmed by visual inspection. In the mean, the landmark registration error is smaller for the manual procedure. Note, however, that conclusions based on this measure are generally critical (see [8] for an analysis of error measures for rigid registration).

Inter-observer variability in localized landmark positions. In total, each observer extracted up to 76 landmarks. For 58 landmarks (76%) we observed a smaller variability in the landmark positions localized with the semi-automatic procedure, and for 47 of these 58 landmarks all five observers selected the same candidate (i.e., in 62% of all cases we have $d_{\mathrm{RMS,sem}} = 0\mathrm{mm}$). For 7 of these 58 landmarks only one observer selected a different candidate. For 10 landmarks of the remaining 18 landmarks where the semi-automatic procedure showed a larger variability in the landmark positions only one observer selected a different candidate. The variability was similar in MR and CT: in MR for 30 out of 38 landmarks the variability was smaller with the semi-automatic procedure, while in CT for 28 out of 38 landmarks the variability was smaller with the semi-automatic procedure.

ACKNOWLEDGMENT

Support of Philips Research Hamburg, project IMAGINE (IMage- and Atlas-Guided Interventions in NEurosurgery) is gratefully acknowledged. Images were provided as part of the AIM project COVIRA and as part of the project "Evaluation of Retrospective Image Registration", National Institutes of Health, Project Number 1 R01 NS33926-01, Principal Investigator, Dr. J. Michael Fitzpatrick, Vanderbilt University, Nashville, TN, USA. We thank Dr. Derek L. G. Hill, Guy's Hospital, London, UK, for providing the 'ground truth' positions for the *C06* data.

REFERENCES

1. K.C. Strasters *et al.* Anatomical Landmark Image Registration: Validation and Comparison. In *Proc. CVRMed-MRCAS'97*, pp. 161–170. Springer, 1997.
2. K.S. Arun *et al.* Least-Squares Fitting of Two 3-D Point Sets. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 9(5):698–700, 1987.
3. K. Rohr. On 3D differential operators for detecting point landmarks. *Image and Vision Computing*, 15(3):219–233, 1997.
4. S. Frantz *et al.* Multi-step Procedures for the Localization of 2D and 3D Point Landmarks and Automatic ROI Size Selection. In *Proc. ECCV'98*, vol. I, pp. 687–703. Springer, 1998.
5. S. Frantz *et al.* Reducing False Detections in Extracting 3D Anatomical Point Landmarks. In *Proc. 3. Workshop Bildverarbeitung für die Medizin*, pp. 54–59. Springer, 1999.
6. T. Hartkens *et al.* Performance of 3D differential operators for the detection of anatomical point landmarks in MR and CT images. In *Proc. SPIE's Internat. Symposium Medical Imaging: Image Processing*. SPIE, 1999 (in press).
7. M. Unser *et al.* B-Spline Signal Processing: Part I—Theory. *IEEE Trans. on Signal Processing*, 41(2):821–833, 1993.
8. J.M. Fitzpatrick *et al.* Predicting Error in Rigid-Body Point-Based Registration. *IEEE Trans. on Medical Imaging*, 17(5):694–702, 1999.