FEATURE-BASED PIECEWISE RIGID REGISTRATION IN 2-D MEDICAL IMAGES

Dirk Smeets *Johannes Keustermans* *Jeroen Hermans* *Dirk Vandermeulen* *Paul Suetens*

K.U.Leuven, Faculty of Engineering ESAT/PSI
IBBT-K.U.Leuven Future Health Department

ABSTRACT

Piecewise rigid registration is a fundamental problem in medical imaging involving intra-patient pose differences in multiple medical images, mostly due to articulated motion. In this paper, we propose a method to extract multiple rigid transformations in 2D medical images in the presence of outliers. First, points of interest in the images are extracted and matched with the SIFT algorithm. Secondly, multiple rigid motions are sampled and clustered by the mean shift algorithm in the special Euclidean group $\mathbf{SE}(2)$, a smooth manifold of 2-D rigid transformation matrices. The method proposed is evaluated for intra-subject registrations of knee fluoroscopy images, demonstrating a mean angular and translational error on the estimated motion of 0.39° and 6.65 pixels, respectively.

Index Terms— Motion estimation, image registration

1. INTRODUCTION

Medical image registration aims to find the transformation that geometrically aligns two (or more) images in the same frame of reference. Intra-subject image registration allows to extract clinical information from different time points (multi-temporal registration) or different modalities (multi-modal registration). This information can be important for surgical or radiotherapy planning, in treatment follow-ups or for drug development by small animal imaging,... Methods for intrinsic image registration have been classified as landmark-based, surface-based, or intensity-based [1]. With respect to the geometrical transformation, methods can be either rigid, affine, piecewise rigid or non-rigid. Rigid and non-rigid intensity-based methods, mostly using mutual information as similarity measure [2, 3], have been successfully used for several applications where often rigid registration is used as initialization for further non-rigid registration. However, large piecewise rigid deformations, mostly due to articulated motion, remain a challenge, since a single rigid transformation (as result of rigid registration) is typically insufficient as initialization of a non-rigid registration method [4]. Therefore, multiple rigid transformations need to be estimated for a better initialization.

In literature, some approaches have been proposed to address the problem of articulated motion. Little et al. [5] present a method in which the different rigid transformations of the articulating objects are estimated using a user-defined segmentation of the rigid structures (bone). The remainder of the image deforms due to both the rigid body transformations and the transformation of user-specified soft-tissue landmarks. Papademetris et al. [6] also assume a segmentation of the rigid structures. A transformation is then computed that blends piecewise rotations and is continuous at the interfaces. For both methods, the estimation of the rigid transformation matrices is thus obtained by user-defined segmentations.

Some contributions for automated rigid transformation estimation have been made [4, 7]. Li et al. [4] use the robust point matching (RPM) approach of Chui and Rangarajan [8] for the registration of point clouds extracted by bony surfaces using simple intensity thresholding. The surface-based method is then followed by an intensity-based algorithm to refine the transformations. However, the RPM approach is not able to handle large deformations and a large number of outliers. Martín Fernández et al. [7] propose an automated landmark-based elastic registration method in which individual rigid structures are affinely registered and soft tissues are elastically registered. An automated segmentation of the rigid structures is used to extract landmarks in the joints of the metacarpals (rigid) in radiography images. However, this procedure for landmark detection is not generally applicable.

In this paper, we propose an automated approach for feature-based piecewise rigid transformation estimation, that is robust against outliers and large deformations, and generally applicable to n-D medical images. We will present a pipeline for 2D images, which can easily be extended to n-D images. First, points of interest in the images are detected and described using the Scale Invariant Feature Transform (SIFT) algorithm [9], as explained in Sect. 2. Second, we propose an algorithm for multiple 2-D rigid motion estimation, based on [10]. The algorithm clusters rigid transformation matrices on Lie groups using mean shift and is described in Sect. 3.

2. KEYPOINT MATCHING

In a first step, keypoints are detected in the images, local feature descriptors are calculated at every such keypoint and keypoints are matched based on these local descriptors.. In the

current implementation, we use the Scale Invariant Feature Transform (SIFT) algorithm of Lowe [9] as local feature descriptor, since it is rotation-invariant and robust against noise. Note that also other local feature description methods can be in the proposed method.

2.1. Key Point Localization

First, a scale-space representation is constructed by convolving the original image signal with a Gaussian kernel creating a set of images that are smoothed versions of the original image. The idea behind a scale space representation is to separate the structures in the original image according to the their spatial scale.From this scale space, interest points are extracted as extrema of the Difference of Gaussians (DoG) that occur at multiple scales.

2.2. Feature Description

In order to obtain a rotation-invariant feature descriptor, each keypoint is assigned one or more canonical orientations based on local image gradient directions. The direction of the image gradient in a scale-dependent region around the keypoint is summarized in a histogram, weighted by the magnitude of the image gradient. The bin with maximal value and each bin with value larger than 80% of this maximal value correspond to the canonical orientations. The descriptor contains 16 8-bin gradient magnitude weighted histograms of orientation values for 4×4 regions around the keypoint. Each of these regions consist of 4×4 pixels.

2.3. Feature Matching

Matches between keypoints are defined using the cosine distance as dissimilarity criterion and filtered. Only when the ratio between the lowest and second lowest dissimilarity is smaller than 0.8, the match is retained.

3. RIGID MOTION CLUSTERING

Next, the rigid motion clustering extracts the relevant rigid motions present in between two images, by sampling rigid transformation matrices in the special Euclidean group $\mathbf{SE}(n)$ and mean shift clustering in the associated Lie algebra $\mathfrak{se}(n)$ (for 2D images, $n = 2$). In the first, sampling step the candidate rigid motions are sampled using the corresponding features pairs. In the second, clustering step the relevant rigid motions are selected out of these candidates.

3.1. Rigid Motion Sampling

Given two sets of matched key points locations $\mathbf{X}^{(1)} = \{\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \dots \mathbf{x}_m^{(1)}\}$ and $\mathbf{X}^{(2)} = \{\mathbf{x}_1^{(2)}, \mathbf{x}_2^{(2)}, \dots \mathbf{x}_m^{(2)}\}$, l rigid transformation matrices $\{\mathbf{M}_j\}_{j=1\dots l}$ are sampled in the special Euclidean group $\mathbf{SE}(n)$, i.e. the matrix Lie group of rigid motions in n-D,

$$\mathbf{M} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}, \tag{1}$$

with $\mathbf{R}$ a rotation matrix in the special orthogonal group $\mathbf{SO}(n)$, since $\mathbf{R}\mathbf{R}^T = I$ and $\det(\mathbf{R}) = 1$, and $\mathbf{t}$ a translation vector in $\mathbb{R}^n$ [11].

A sample is taken by randomly selecting n corresponding feature pairs and computing the translation vector $\mathbf{t}_j$ and rotation matrix $\mathbf{R}_j$ that best aligns the n corresponding feature pairs. In order to reduce the computational cost of the next, mean shift clustering step, a sample is retained if the sampled transformation maps at least one of the $m - n$ other features to its corresponding feature, mathematically expressed as $\min_i ||\mathbf{x}_i^{(1)} - \mathbf{R}_j\mathbf{x}_i^{(2)} - \mathbf{t}_j|| < \epsilon_v$, where ϵ_v is the validation threshold and i is not used for computing the sampled transformation. Although l samples are taken, the samples will be better by the postprocessing [10]. We choose $l = 500$ and $\epsilon_v = 3\sqrt{n}$.

3.2. Mean Shift Clustering

Given l rigid transformation matrices $\{\mathbf{M}_j\}_{j=1\dots l}$ as samples of a distribution defined on a Lie group, the modes of the distribution are found with the mean shift algorithm, which is a robust clustering technique without need for prior knowledge on the number of clusters.

In order to compute the mean shift vector at $\mathbf{M}$, the distance $d(\mathbf{M}, \mathbf{M}_j)$ must be computed from $\mathbf{M}$ to all sampled transformation matrices $\mathbf{M}_j$. Since the Lie group has a structure of a manifold, the distance $d(\mathbf{M}, \mathbf{M}_j)$ is the length of the geodesic between these matrices, i.e. the curve with the minimum length on the Lie group. This length is called the intrinsic distance and can be easily computed in the associated Lie algebra, which is the tangent space in $\mathbf{M}$, as $d(\mathbf{M}, \mathbf{M}_j) = ||\mathfrak{m}_j||$. Here, $\mathbf{M}_j$ is mapped to $\mathfrak{m}_j$. This is conceptually represented in Fig. 1.

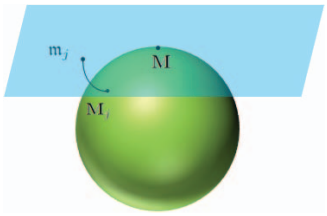


Fig. 1. Conceptual representation of the Lie group (sphere) and its associated Lie algebra in $\mathbf{M}$.

This mapping is the logarithmic map, $\log : \mathbf{M} \to \mathfrak{m}$, around I and maps transformation matrices in the special Euclidean group $\mathbf{SE}(n)$ to vectors in the Lie algebra $\mathfrak{se}(n)$. The

Lie algebra of rigid motions $\mathfrak{se}(n)$ is the set of matrices [11]

$$\mathfrak{m} = \begin{bmatrix} \mathbf{\Omega} & \mathbf{u} \\ \mathbf{0} & 0 \end{bmatrix}, \tag{2}$$

where $\mathbf{\Omega}$ is a skew-symmetric matrix, which is

$$\mathbf{\Omega} = \begin{bmatrix} 0 & \omega_z \\ -\omega_z & 0 \end{bmatrix}, \quad \mathbf{\Omega} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}. \tag{3}$$

for $n = 2$ and $n = 3$, respectively. $\mathbf{\Omega}$ can be written as $\mathbf{\Omega} = ||\omega||\mathbf{V}$ [12, p. 51], with $\omega = \omega_z$ ($n = 2$) or $\omega = [\omega_x\ \omega_y\ \omega_z]^T$ ($n = 3$). Given $\mathbf{R}$, $\mathbf{\Omega}$ is then computed using

$$||\omega|| = \cos^{-1}\left(\frac{Tr(\mathbf{R}) - n + 2}{2}\right) \tag{4}$$

$$\mathbf{V} = \frac{1}{2\sin||\omega||}(\mathbf{R} - \mathbf{R}^T). \tag{5}$$

The exponential map, $\exp : \mathfrak{m} \rightarrow \mathbf{M}$, is the inverse mapping and maps the vectors in the Lie algebra to the Lie group and is an homeomorphism near the identity element. It is found by taking the matrix exponential, which can be written for rigid transformation matrices as

$$\mathbf{R} = I_n + \sin(||\omega||)\mathbf{V} + (1 - \cos(||\omega||))\mathbf{V}^2 \tag{6}$$

$$\mathbf{t} = (I_n + \frac{(1 - \cos(||\omega||))}{||\omega||})\mathbf{V} + \frac{(||\omega|| - \sin(||\omega||))}{||\omega||})\mathbf{V}^2)\mathbf{u}, \tag{7}$$

where $||\omega||$ is extracted directly from $\mathbf{\Omega}$ and $\mathbf{V} = \mathbf{\Omega}/||\omega||$.

The mean shift vector at location $\mathbf{M}$ is then defined as [10]

$$\mathbf{m}_h(\mathfrak{m}) = \frac{\sum_{j=1}^{l} \mathfrak{m}_j g_N\left(||\frac{\mathfrak{m}_j}{h}||^2\right)}{\sum_{j=1}^{l} g_N\left(||\frac{\mathfrak{m}_j}{h}||^2\right)}, \tag{8}$$

with $\mathfrak{m}_j = \log(\mathbf{M}^{-1}\mathbf{M}_j)$, the mapping of $\mathbf{M}_j$ to the Lie algebra of $\mathbf{M}$ and $g_N(s) = -\frac{1}{2}e^{-\frac{1}{2}s}$, the derivative of a Gaussian kernel. The new location of $\mathbf{M}$ is then given by

$$\mathbf{M} = \mathbf{M}\exp(\mathbf{m}_h(\mathfrak{m})) \tag{9}$$

Eq. (8) and Eq. (9) are applied until convergence has occurred. This procedure is repeated for each sampled transformation matrix, setting $\mathbf{M} = \mathbf{M}_j$. Finally, each sample has moved to a cluster center (local mode in the distribution). If the probability of the local mode is sufficient according to a threshold P_{thr}, the cluster center corresponds to a rigid motion $\tilde{\mathbf{M}}_i$.

When a certain keypoint belongs to different rigid motion clusters (it can be sampled several times), it is assigned to the largest cluster, rejecting all other assignments.

4. EXPERIMENTAL RESULTS

The proposed approach is validated for intra-subject registrations of in-vitro knee fluoroscopy images. Due to the knee bending large deformations occur causing to fail most local, dense registration techniques. The images of implanted Smith & Nephew knees made of sawbone are captured with a Siemens Arcadis fluoroscopy device.

After motion clustering, the number of clusters is restricted to two. A visual result of the proposed sparse rigid motion clustering method is shown in Fig. 2. Two rigid motions are extracted, indicated by red and green lines for the movement of the upper and lower leg, respectively. Note that more matches are found for the lower leg, since this is completely stationary.

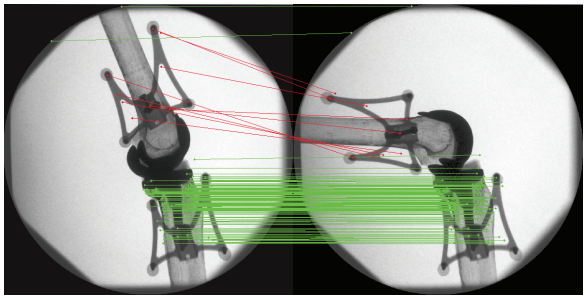


Fig. 2. Example of found extracted rigid motion (green and red) for virtual knee fluoroscopy images with $\pm 67°$ bending.

More quantitative results for the knee image registration are shown in Fig. 3 evaluating 59 fluoroscopy images (knee bending from 20° till 75°). The estimated transformations of the upper and lower leg are evaluated with respect to a gold standard, obtained by indicating 8 landmarks, that are highly repeatable due to the attached frame, per image. A boxplot of the number of features detected and assigned to a cluster are given (left) for the upper leg and lower leg motion. Also, the angular error (middle) and translation error (right) are shown for the upper and lower leg with respect to motions extracted from manual indicated landmarks. The green '$\circ$' indicates the mean error value and black '$\times$' the error for the motion extracted from a second manual indication of landmarks.

For both motions, the angular error is mostly smaller than 0.5°, while the translational error is mostly limited to 10 pixels. For both errors, the results are not influenced by the amount of knee bending. The mean angular and translational error averaged over both motions is 0.39° and 6.65 pixels, respectively. These values are close to the error for a second manual indication (0.25° and 4.25 pixels). Surprisingly, the translational error for the stationary lower leg is, on average, larger than for the moving upper leg. We believe this is due to the larger manual indication error for the landmarks in the lower leg.

5. CONCLUSION

In this paper, we have proposed a method for sparse 2D image registration in presence of large deformations. Exploiting that

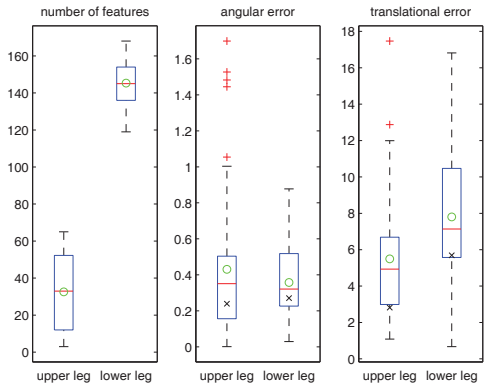


Fig. 3. Quantitative results of the registration of virtual knee fluoroscopy images. Boxplots are shown for the number of features detected and assigned, the angular and translational error (in degrees and pixels, respectively). The '$\circ$' indicates the mean and '$\times$' the error for a second manual indication of landmarks.

these deformations can be reduced to multiple rigid motions, allows to find robust matches by clustering the different rigid motions on the matrix Lie group $\mathbf{SE}(2)$.

Future work involves extending SIFT for n-D images, allowing a general framework for rigid motion estimation in n-D images. A first attempt to extend SIFT is the n-SIFT algorithm of Cheung and Hamarneh [13]. However, this algorithm is not rotation-invariant. This invariance is essential for the proposed approach.

Another useful extension would be the transformation transfer to the remainder of the image, since the k rigid motions $\{\tilde{\mathbf{M}}_i\}_{1\ldots k}$, found by mean shift clustering on the Lie group of rigid transformation matrices, are only estimated in a small set of matching feature locations $\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}_i\}_{1\ldots p}$. This transformation in each other point $\mathbf{r}$ could be computed by a linear combination of the estimated rigid motions in the Lie algebra,

$$\mathbf{M}(\mathbf{r}) = \exp\left(\sum_{i=1}^{k} \mathbf{w}_i(\mathbf{r}) \log(\tilde{\mathbf{M}}_i)\right), \qquad (10)$$

with $\mathbf{w}_i = \mathbf{q}_i / \sum_j \mathbf{q}_j$ and $\mathbf{q}_i = ||\mathbf{r} - \mathbf{r}_i||^{-\mu}$, $(\mu > 1)$.

6. ACKNOWLEDGEMENTS

This work is supported by IWT Vlaanderen, the Research Programme of FWO and the Research Fund K.U.Leuven. The authors sincerely thank Smith & Nephew for the images provided.

7. REFERENCES

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

[2] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information.," *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[3] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast mr images," *IEEE Trans. Med. Imaging*, vol. 18, pp. 712–721, 1999.

[4] X. Li, T. E. Peterson, J. C. Gore, and B. M. Dawant, "Automatic registration of whole body serial micro ct images with a combination of point-based and intensity-based registration techniques," in *ISBI '06*, 2006, pp. 454–457.

[5] J. A. Little, D. L. G. Hill, and D. J. Hawkes, "Deformations incorporating rigid structures," in *MMBIA '96*. 1996, pp. 104–, IEEE.

[6] X. Papademetris, D. P. Dione, L. W. Dobrucki, L. H. Staib, and A. J. Sinusas, "Articulated rigid registration for serial lower-limb mouse imaging," in *MICCAI '05*, 2005, vol. 3750 of *LNCS*, pp. 919–926.

[7] M. Á. Martín-Fernández, R. Cárdenes, E. Muñoz Moreno, R. de Luis-García, M. Martín-Fernández, and C. Alberola-López, "Automatic articulated registration of hand radiographs," *Image Vision Comput.*, vol. 27, pp. 1207–1222, 2009.

[8] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Underst.*, vol. 89, pp. 114–141, 2003.

[9] D. G. Lowe, "Object recognition from local scale-invariant features," in *ICCV '99*. 1999, vol. 2, pp. 1150–1157, IEEE.

[10] O. Tuzel, R. Subbarao, and P. Meer, "Simultaneous multiple 3d motion estimation via mode finding on lie groups," in *ICCV '05*. 2005, vol. 1, pp. 18–25, IEEE.

[11] B. C. Hall, *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, vol. 222, Springer, 2003.

[12] J. M. Selig, *Geometrical Methods in Robotics*, Monographs in Computer Science. Springer, 1996.

[13] W. Cheung and G. Hamarneh, "n-SIFT: n-dimensional scale invariant feature transform," *Trans. Img. Proc.*, vol. 18, pp. 2012–2021, 2009.