VALIDATION OF DEFORMABLE REGISTRATION IN ADAPTIVE RADIATION THERAPY WITH SCALE INVARIANT FEATURE TRANSFORM

Paganelli Chiara[1], Peroni Marta[1], Baroni Guido[1,2], Riboldi Marco[1,2]

[1] Department of Bioengineering, Politecnico di Milano, Milano, Italy
[2] Bioengineering Unit, Centro Nazionale di Adroterapia Oncologica, Pavia, Italy

ABSTRACT

A critical issue in Deformable Registration (DR) is its validation due to the lack of a gold standard in clinical practice. We analyze Scale Invariant Feature Transform (SIFT) to quantify its potential as a tool for DR assessment. For this purpose we preliminary verify SIFT properties on both rigid and non-rigid phantoms and then apply it to evaluate DR in Head&Neck Adaptive Radiation Therapy. Results demonstrate SIFT capabilities as a performance assessment metric which is operator-independent and not limited to the analysis of specific structures.

Index Terms— Scale Invariant Feature Transform (SIFT), feature extraction, validation of deformable registration.

1. INTRODUCTION

Along the course of a radiotherapy treatment, the accuracy of dose delivery may be compromised by anatomical and physiological changes as well as therapy response. A possible strategy to compensate for changes and improve delivery is Adaptive Radiation Therapy (ART), where patients are re-imaged and re-planned several times during the entire treatment course. ART mainly relies on Deformable Registration (DR), whose modeled transformation can be useful to analyze target motion as well as to measure intra-patient changes. Several algorithms have been developed for DR, but the lack of a gold standard in clinical practice still compromises its validation. The most common method of DR performance evaluation is based on the analysis of similarity measures between outlined structures, such as Dice Similarity Coefficient (DSC), Center of Mass distance (COM), and Root Mean Square Error (RMSE). An alternative approach is to compare the anatomical or external landmarks position after the registration. Most clinical applications still rely on the manually identification of anatomical landmarks, a time-consuming and operator-dependent procedure. Therefore, automatic or semi-automatic landmarks extraction methods have increasingly gained popularity. First operators for the automatic extraction of points, such as the corner detectors by Kitchen and Rosenfeld [1], Forstner [2] and Harris [3], examined an image at a single scale, thus reducing the accuracy and stability in feature detection. Furthermore, the invariance properties of a feature extraction method are crucial to find points in images acquired with different modalities, or scans of different patients, or temporal series of the same patient. Scale Invariant Features Transform (SIFT) is a method that can extract and match stable and prominent points at different scales between two images. Lowe [4,5] initially designed it for 2D images and Cheung and Hamarneh extended it to nD images [6]. Preliminary studies analyzed its applicability for DR and contour propagation based on Thin Plate Splines (TPS) [7,8,9].

In this work, we analyze SIFT performance as DR validation method. As a preliminary test, we check SIFT invariance properties on a phantom by applying different rigid transformations and, for the first time, non-rigid transformations. Finally, we apply SIFT to pairs of registered images of twenty Head&Neck (H&N) patients, to provide a new matching metric for assessing DR performance on one hand, and for facilitating the implementation of ART in a clinical environment on the other.

2. MATERIALS AND METHODS

2.1. Scale Invariant Feature Transform

2.1.1. Scale-space extrema detection

To prevent aliasing due to noise, we pre-smoothed the input image $I(x,y,z)$ with a Gaussian kernel featuring a proper standard deviation $\sigma_{anti-aliasing}$. Then, a scale space was built by convolving $I(x,y,z)$ with a Gaussian kernel $G(x,y,z,\sigma)$. The output was a stack (or octave) of blurred Gaussian Images $L(x,y,z,k\sigma)$ separated by a constant multiplicative factor $k = 2^{1/s}$, where s is the number of intervals. The value of the Gaussian kernel σ was updated by moving from one scale to another, based on the sum of sequential independent filters variances.

Once a complete octave has been processed, the Gaussian Image that had σ equal to double of the initial value was subsampled by a factor of 2 and used as a source for the next octave. Finally, Difference of Gaussian Images ($DoGs$) were obtained by subtracting each Gaussian Image from the adjacent image in each octave:

$$DoG(x,y,z,k^i\sigma) =$$
$$= \left(G(x,y,z,k^{i+1}\sigma) - G(x,y,z,k^i\sigma)\right) * I(x,y,z) =$$
$$= L(x,y,z,k^{i+1}\sigma) - L(x,y,z,k^i\sigma)$$

Candidate keypoints were detected as local extrema by comparing a voxel to its 80 neighbors in 3x3x3 regions in the current and adjacent $DoGs$ [5,6]. Therefore, $s+2$ $DoGs$ (i.e. $s+3$ Gaussian Images) were required for the whole procedure.

2.1.2. *Keypoint localization*

We defined a point as stable if it was well contrasted and not located on plate-like or tubular structures in the DoG. The first requirement was satisfied if the candidate landmark $\hat{x}$ had a value of $|DoG(\hat{x})|$ greater than a threshold T_c. Among well contrasted features, we selected those in which principal curvatures had the same sign (i) and magnitude order (ii), according to the following formulas [9]:

(i) $tr(H)\det(H) > 0$ and $\sum det_2^P(H) > 0$

(ii) $\frac{tr(H)^3}{\det(H)} < \frac{(2t_{max}+1)^3}{(t_{max})^2}$

where H is the Hessian matrix of the image, $tr(H)$ its trace, $\det(H)$ the determinant and $\sum det_2^P(H)$ the sum of principal second-order minors of H.

2.1.3. *Orientation assignment and Keypoint descriptor*

To properly describe each feature, a local orientation was assigned to each voxel, by computing gradient magnitude and orientations for each first Gaussian Image $L(x,y,z,\sigma)$ in each octave. We considered $r \times r \times r$ voxel regions around the keypoints as their local domain. Each region was weighted by a circular Gaussian window with sigma σ_w equal to one half of the region width, to avoid sudden descriptor's changes and give less emphasis to gradients far from the keypoint. An orientation histogram with $b \times b$ bins was then built in $m \times m \times m$ sub-regions of the local feature domain, as described in [6,9]. The descriptor was defined as a vector containing the values of all orientation histogram entries. Chosen values of r, m and b are defined in 2.3.1.

2.1.4. *Matching*

We defined the Euclidean distance $S_{n,n'}$ as

$$S_{n,n'} = \sqrt{\sum_{\alpha=1}^{k} |(\nabla I_n)_\alpha - (\nabla I_{n'})_\alpha|^2}$$

where $(\nabla I_n)_\alpha$ and $(\nabla I_{n'})_\alpha$ are the orientation histograms of feature points n on one image and of a potential association n' on another image and α is the bin index. Amongst all the possible $S_{n,n'}$, the two closest points n'_1 and n'_2 were identified. If the ratio $S_{n,n'_1}/S_{n,n'_2}$ was below a matching threshold T_m, the point having the lowest $S_{n,n'}$ value was chosen as the correspondent of landmark n. Otherwise, no association was identified for the keypoint. To further increase the accuracy of feature point association, the bi-directionality of each landmark association was verified.

Table I. Rigid transformations.

Transformation	Parameters
Translation (x,y,z)	[0,1,2,3,4,5,6] mm
Translation (x,y)	[0,1,2,3,4,5,6] mm
Rotation (z)	[0,1,2,3,4,5] °
Rototranslation	[0,1,2,3] mm, [0,1,2,3]°
Scaling (x,y,z)	[0.8,0.9,1.1,1.2]

Table II. Non-rigid transformations.

	Maximum deformation [mm]		
	x	y	z
vf_1	3.29	5.72	0.06
vf_2	3.29	5.75	6.91
vf_3	15.42	5.72	4.16

2.2. Testing datasets

We tested our approach on an image of a RANDO® phantom acquired on a clinical CT scanner, using supine setup and clinical acquisition protocols. The volume acquired was 512x512x123 voxels and element spacing was [0.94, 0.94, 3] mm.

Twenty H&N cancer patients were enrolled in this study. Treatment planning CT (CTsim), two Cone Beam CT (CBCT) scans (On-Board Imager, Varian Palo Alto, U.S.A. at 95KVP) after 40Gy and 50Gy dose delivery and re-planning CT (CTrepl) were acquired. CBCT and CTrepl were acquired less than three days apart.

2.3. Experiments

2.3.1. *Parameters definition*

For phantom, parameters were chosen as in the literature. After an anti-aliasing filtering with $\sigma_{anti-aliasing} = 0.5$mm [5], the Gaussian blur with $\sigma = 1.5$mm and $s = 3$ [5,6] was performed and 5 $DoGs$ per 3 octaves were generated. The so obtained feature candidates were refined using $T_c = 0.03$ [5,9] and $t_{max} = 20$ [9], assuming image values in the range [0,1]. For the keypoint descriptor, we considered $r = 16$ (i.e. $\sigma_w = 8$), $m = 4$ and $b = 8$ and matching threshold T_m was 80% [5,6].

For patients scans, we used $\sigma = 2$mm and enlarged the descriptor to $r = 24$ (i.e. $m = 6$ and $\sigma_w = 12$), to be more robust to artifacts that affect points detection. Finally T_m was set to 90% as in [9].

2.3.2. *Phantom study*

To test SIFT invariance properties, we applied rigid (Table I) and non-rigid (Table II) transformations on RANDO®, creating a synthetic image. SIFT matches are then evaluated in terms of error, computed as 3D residual distance between

SIFT feature location and its ground truth position in the synthetic image. This latter is obtained by direct application of the transformation to the points in the original RANDO®.

2.3.3. Patient study

In H&N patients (Fig. 1) to update the treatment including temporal-anatomical changes, we generated a virtual planning CT (VirtualCT) by means of multi stage B-spline DR (www.plastimatch.org) between CTsim and CBCT. We also performed a rigid registration between CTrepl and CBCT (CTrepl_rig), as well as between CTsim and CBCT (CTsim_rig). The SIFT method was then applied to image pairs and the matching error was computed as 3D distance between matches in the couple. CTsim/CBCT were analyzed to verify the presence of macroscopic changes, while the comparison between CTsim_rig/CBCT and VirtualCT/CBCT error provided an estimate of the actual need of DR to compensate for the detected variations. CTrepl_rig/CBCT error represented the level of confidence of our algorithm. Then we performed a Friedman test between the error populations obtained from VirtualCT/CBCT, CTsim_rig/CBCT and CTrepl_rig/CBCT, to study the significance of the analysis. Finally, the association error was compared with classical indices proposed in literature based on contours (DSC, COM, RMSE). Image pre-processing included a pre-alignment of CTsim with CBCT to cope with different element spacing and histogram cutting at the 99.9% percentile to compensate metal artifacts.

3. RESULTS AND DISCUSSION

3.1. Rigid and Non-Rigid Transformations

Fig. 2 shows the boxplots of the point distance distributions for different transformations. For translations along three dimensions (Panel A), we notice a decrease in the number of keypoints and a higher error for values that are not multiple of the superior-inferior (SI) voxel dimension (i.e. 3mm), due to uncertainty in the attribution of the SI coordinate. Given the anisotropic dimensions of the voxel, this uncertainty depends on two approximations. In fact, on one side, a translation of an intra-voxel quantity implies the integration of the information between adjacent slices, and on the other side, extrema are detected by means of finite difference derivative, concentrating the voxel information in its center. On the other hand, the residual distance is negligible, if the translation is a multiple of the SI voxel dimension. If the image content is translated along anterior-posterior (AP) and right-left (RL) directions (Panel B), the error is lower than 1mm with a little increase as a function of translation parameters. This is due to the translation not being exactly multiple of the voxel dimension in these directions (i.e. 0.94mm). For rotation (Panel C), the median of point distance is below the voxel dimension, confirming the method invariance to this transformation. We note few

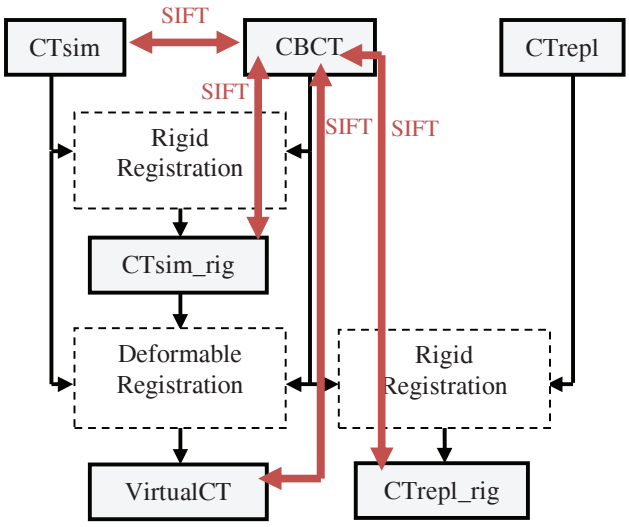


Fig.1. Clinical application workflow.

outliers around 3mm, due to a wrong association of keypoints in the SI direction. Rototranslation (Panel D) and scaling (Panel E) demonstrate the above described invariance properties. Also for growing levels of deformations (Panel F) we notice that the error is below the voxel dimension.

3.2. Patients

In Table III, we report the median and the variability of median distribution in all patients for different image pairs. The application of SIFT to the pair CTsim/CBCT shows the largest error, reflecting the presence of macroscopic anatomical changes. The lower residual distance in VirtualCT/CBCT, compared to the rigid case (CTsim_rig/CBCT), confirms the better performance of non-rigid registration. Furthermore, CTrepl_rig/CBCT and VirtualCT/CBCT resulted significantly different from CTsim_rig/CBCT.

In all these cases, outliers are due to an incorrect attribution in the SI direction and to the fact that feature matching is based just on the image content (i.e. descriptor similarity), without further distinguishing or clustering the landmarks (e.g. image relative location). Therefore, ICP-like robust association might further improve the SIFT algorithm.

Finally, by comparing results obtained from SIFT (Table III) and those obtained from classical indices (Table IV) between VirtualCT and CTrepl_rig for all patients, we see that SIFT is an accurate method, providing an index which quantifies the error as a physical length, in contrast with DSC, and is not subjected to contours variability (e.g. manual identification), such as COM and RMSE are. Further work will be focused to eliminate the normalization of the initial image, which results in a lower identification probability of soft tissues in favor of bony structures.

Table III. Median and variability (computed as difference between 75th and 25th percentiles) in all patients for different image pairs obtained with SIFT.

	median [mm]	75th-25th [mm]
CTsim/CBCT	31.52	15.95
CTsim_rig/CBCT	1.35	0.99
VirtualCT/CBCT	0.66	0.27
CTrepl_rig/CBCT	0.66	0.38

Table IV. Median of DSC, COM and RMSE between VirtualCT and CTrepl_rig for different structures in all patients.

	Mandible	Right Parotid	Left Parotid	GTV
DSC	0.85	0.78	0.79	0.56
COM[mm]	1.09	2.24	1.60	3.24
RMSE[mm]	1.4	2.21	2.23	2.55

4. CONCLUSION

We showed that SIFT can be useful in a clinical scenario such as ART, providing a new tool for DR assessment. For this purpose, we evaluated invariance properties of the method to synthetic rigid and non-rigid transformations on a comprehensive phantom study, to simulate the stages of a possible ART registration process. The main limits of the algorithm are connected to the initial normalization and the finite difference derivative computation for extrema identification. Nonetheless, the patient study demonstrate the SIFT capabilities to provide a new matching metric for DR evaluation. Even if the proposed method may not be the definitive answer to the lack of a gold standard in DR validation, it is operator-indipendent and not related to specific structures, in contrast with classical indices based on contours.

5. ACKNOWLEDGEMENT

The authors acknowledge AIRC, Italian Association for Cancer Research, for the support. The authors would like to thank G.C. Sharp, PhD, for the help during this work.

6. REFERENCES

[1] L. Kitchen and A. Rosenfeld, "Gray-level Corner Detection", *Pattern Recognition Letters*, North-Holland, pp. 95-102, 1982.
[2] W. Förstner, "A fast operator for detection and precise location of distinct points, corners and centres of circular features", *ISPRS Intercommission Workshop*, Interlaken, 1987.
[3] C. Harris and M. Stephens, "A combined corner and edge detector", *Plessey Research Roke Manor*, pp. 147-152, United Kingdom, 1988.
[4] D.G. Lowe, "Object recognition from local scale-invariant features", *proc. of the 7th IEEE Int. Conf. on Comp. Vis. (ICCV)*, pp. 1150–1157, Greece, 1999.
[5] D.G. Lowe, "Distinctive image features from scale-invariant keypoints". *Int. J. of Comp. Vis.*, pp. 91-110, Canada, 2004.

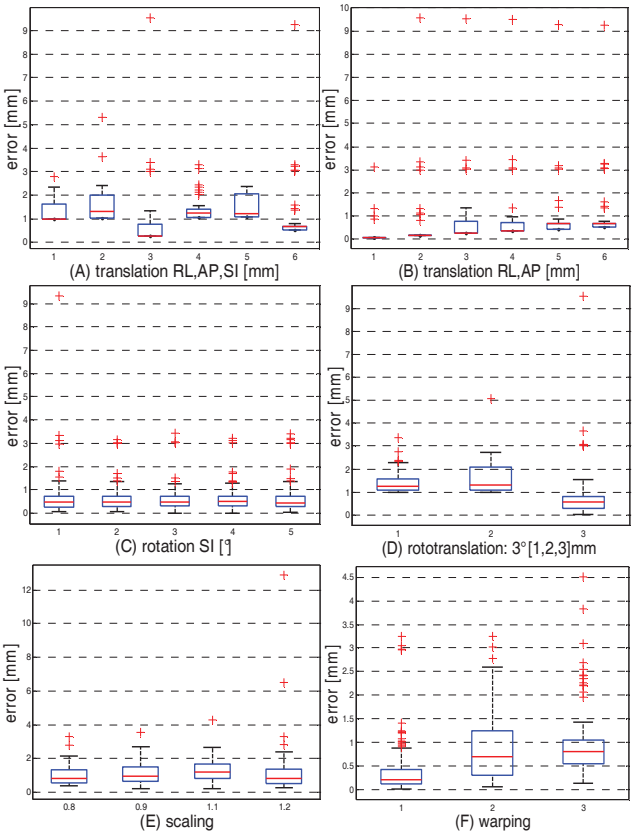


Fig. 2. Boxplots of the error distribution for different transformations. Rigid transformations: (A), translation along RL,AP,SI directions; (B), translation along RL and AP directions; (C), rotation around SI direction; (D), rototranslation; (E), scaling. Non-Rigid transformations: (F), growing levels of deformations shown from left to right. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

[6] W. Cheung and G. Hamarneh, "n-SIFT: n-Dimensional Scale Invariant Feature Transform", *IEEE Transaction on Image Processing*, pp. 2012-2021, Washington, 2009.
[7] Y. Xie, M. Chao, P. Lee and L. Xing, "Feature-based rectal contour propagation from planning CT to cone beam CT", *Medical Physics*, pp. 4450-4459, California, 2008.
[8] M. Chao, Y. Xie, E. G. Moros, Q. Le and L. Xing, "Image-based modeling of tumor shrinkage in head and neck radiation therapy", *Medical Physics*, pp. 2351-2358, California, 2010.
[9] S. Allaire et al., "Full Orientation Invariance and Improved Feature Selectivity of 3D SIFT with Application to Medical Image Analysis", *IEEE*, pp.1-8,Canada, 2008.