34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Scale Invariant Feature Transform as feature tracking method in 4D imaging: a feasibility study*

C. Paganelli, M. Peroni, F. Pennati, G. Baroni, P. Summers, M. Bellomi and M. Riboldi

Abstract—**We propose the use of Scale Invariant Feature Transform (SIFT) as a method able to extract stable landmarks from 4D images and to quantify internal motion. We present a preliminary validation of the SIFT method relying on expert user identification of landmarks and then apply it to 4D lung CT and liver MRI data. Results demonstrate SIFT capabilities as an operator-independent feature tracking method.**

I. Introduction

Four-dimensional imaging techniques such as 4D CT and 4D MRI can depict the breathing motion of internal organs. The application of automatic motion tracking techniques to derive a quantitative description of breathing-induced motion of specific anatomo-pathological structures is still a challenge. Although extensively applied, image-based deformable registration methods across 4D CT/MRI motion phases suffer from the lack of validation data [1,2]. Potential applications of automatic tracking in 4D imaging include the optimization of external beam radiotherapy by supporting time-resolved treatment delivery methods and the customization of target volume margins [3,4]. Preliminary studies for automatic landmark extraction, such as the use of corner detectors by Kitchen and Rosenfeld [5], Forstner [6] and Harris [7], have examined an image at a single scale, thus limiting the accuracy and stability of feature detection. Furthermore, the invariance properties of feature extraction methods are crucial to detect points in temporal image series of the same patient. Scale Invariant Features Transform (SIFT) is a method that provides extraction and matching of stable and prominent points at different scales between two images. Lowe [8,9] initially designed it for 2D images and Cheung and Hamarneh extended it to nD images [10].

In this work, we investigate SIFT as a feature tracking method for quantitative motion analysis in 4D image datasets. As a preliminary study, we compared SIFT matches with manual landmarks in a lung 4D CT dataset. We then applied SIFT to volume pairs of 4D lung CT and 4D liver MRI data, to demonstrate its use for internal motion tracking.

II. Materials and Method

A. Scale Invariant Feature Transform

Scale-space extremes detection

As described by [9,11], we pre-smoothed the input image $I(x,y,z)$ with a Gaussian kernel featuring a standard deviation σ_{noise}, to prevent aliasing. Then a scale space was built by convolving the pre-smoothed $I(x,y,z)$ with a Gaussian kernel $G(x,y,z,\sigma_{prev})$. 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_{prev} was updated by moving from one scale to the next one, based on the sum of sequential independent filters variances. Once a complete octave was processed, the Gaussian Image that exhibited 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 feature points (i.e. landmarks) were detected as local extremes, by comparing each voxel to its neighbors in 3x3x3 regions in the current and adjacent $DoGs$. For a 3D image, there were a total of $(3^{3+1}-1)$ neighbors to be checked (i.e. the cube of voxels excluding the central one in the current DoG) [9,10]. As the scale above and below the current one were needed to determinate extremes, $s+2$ $DoGs$ (i.e. $s+3$ Gaussian Images) were required for the whole procedure.

Feature point localization

We classified a point as stable, when found to be well contrasted and not located on plate-like or tubular structures in the DoG. The first requirement was satisfied if the candidate point $\hat{x}$ 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 (i) sign and (ii) magnitude order, according to the following formulas [11]:

(i) $\quad \mathrm{tr(H)}\,\det(\mathrm{H}) > 0$ and $\sum \det_2^{\mathrm{P}}(\mathrm{H}) > 0$

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

where H is the Hessian matrix of the image, $\mathrm{tr(H)}$ its trace, $\det(\mathrm{H})$ the determinant, $\sum \det_2^{\mathrm{P}}(\mathrm{H})$ the sum of principal second-order minors of H and $\mathrm{t_{max}}$ the ratio between the largest magnitude eigenvalue of H and the smaller one.

Orientation assignment and feature point 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 $L(x,y,z,\sigma_{prev})$ in each octave. We considered $r \times r \times r$ voxel regions around the feature points 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 changes and to give less emphasis

*Research supported by AIRC, Italian Association for Cancer Research Foundation.

C. Paganelli, M.Peroni, F. Pennati, G. Baroni and M. Riboldi are with the Department of Bioengineering, Politecnico di Milano, Milano, Italy (e-mail: chiara1.paganelli@ mail.polimi.it).

G. Baroni and M. Riboldi are with Bioengineering Unit, Centro Nazionale di Adroterapia Oncologica, Pavia, Italy.

P. Summers and M. Bellomi are with Division of Radiology, European Institute of Oncology, Milano, Italy.

to gradients far from the feature point. An orientation histogram with $b \times b$ bins was then built in $m \times m \times m$ sub-regions of the local feature domain. The descriptor was formed from a vector containing the values of all orientation histogram entries, as described in [9,11]. Selected values of r, m and b are defined in section C.

Feature 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 corresponding to n. Otherwise, no association was identified for the feature point. To further increase the accuracy of feature point association, the bi-directionality of each landmark association was verified.

B. Testing dataset

We tested our approach on a 4D CT dataset (www.dir-lab.com) of the entire thorax acquired while an external respiratory phase signal was recorded. The images acquired were sorted into ten equally sampled phases of the respiratory cycle. The reconstructed non isotropic volumes were 512×512×128 voxels of 0.97×0.97×2.5 mm. Corresponding maximum inhale (L5) and exhale phase (L0) landmarks were manually individuated by an expert [12,13]. We also applied the method to a 4D MRI dataset of the liver (www.vision.ee.ethz.ch/4dmri/). Fourteen 3D anisotropic volumes were reconstructed from dynamic sagittal 2D images acquired during free breathing, by a multi-dimensional gating measure relying on a dedicated navigator slice [14]. Each volume consisted of 256×256×25 voxels of 1.37×1.37×4 mm.

C. Experiments

Parameters definition

Parameters were derived from an analysis of current literature on the use of SIFT in biomedical imaging. After an anti-aliasing filtering with $\sigma_{noise} = 0.5$mm [8], the Gaussian blur with $\sigma_{prev} = 2$mm and $s = 3$ [10,11] was performed and 5 $DoGs$ per 3 octaves were generated. The feature candidates were refined using $T_c = 0.03$ [9,11] and $t_{max} = 20$ [11], assuming image values in the range [0,1]. For the feature point descriptor, we found $r = 24$ (i.e. $\sigma_w = 12$), $m = 6$ and $b = 8$ to be more robust against artifacts that affect points detection. Finally T_m was set to 90% as suggested in [11].

Match validation

4D CT SIFT matches were evaluated in terms of error, computed as 3D residual distance between SIFT feature location at maximum exhale (0) and maximum inhale (5) phases. This error was compared with the one produced by manual landmarks identified on the same phases (L0 and L5). Mann-Whitney test [15] was performed between the two error distributions (SIFT 0-5 and L0-L5), in order to test the reliability of the SIFT method for feature tracking.

Feature tracking

We applied SIFT to volume pairs of 4D scans according to two approaches:

- keeping maximum exhale phase (0) as reference;
- moving consecutively along the breathing cycle, i.e. selecting the previous phase as reference for feature tracking in the current phase.

For each strategy, we computed the number and the residual distance of the matches in each volume pair. The number of common matches between all phases was computed, preserving also the matches which were present in a neighborhood of the maximum inhale phase (i.e. 5 and 6 for 4D CT) and with no more than two misses in the previous and following phases. Final trajectories were obtained by cubic interpolation both of features tracked on all phases and those affected by less than five misses. For 4D CT, image pre-processing included lung masking, in order to avoid SIFT bony structure identification in favor of vessels and bifurcations. Given the higher number of phases in the 4D MRI scan, the neighborhood of the maximum inhale phase (6) was enlarged to phases 5, 6 and 7. No further masking was needed for this dataset.

III. Results and Discussions

A. Match validation

Table I reports the number of matches identified at maximum exhale and maximum inhale phases by an operator (L0-L5) and SIFT (SIFT 0-5) in 4D CT. We also report median and variability (computed as difference between 75th and 25th percentiles) of the two error distributions. Despite the slightly higher median of our method, the Mann-Whitney test [15] confirms that the two populations were not significantly different (p-value=0.71). This results in SIFT preliminary validation and allows its application in feature tracking and subsequent internal trajectory estimation.

B. Feature tracking

Fig. 1 shows the number of SIFT matches identified by following the two proposed approaches in 4D CT and 4D MRI data. In 4D CT, the number of matches was higher when moving along the breathing cycle (Panel B) relative to keeping phase 0 as the reference (Panel A). In 4D MRI, on setting phase 0 as reference (Panel C), the number of matches was lower at maximum inhale (phase 6), and was higher for the breathing cycle tracking approach (Panel D). In 4D MRI, the number of matched features exhibits a breathing cycle trend, not seen with 4D CT. This is due to the fact that the 4D MRI re-sorting method does not assume a constant breathing depth or even strict periodicity and does not depend on an external gating signal. Table II reports the number of common and preserved matches that were included for trajectory reconstruction. The number of common matches increased when tracking was performed along the breathing cycle, whereas the number of preserved matches was larger when phase 0 was set as reference.

Results reported in Fig. 1 and Table II show how searching in two subsequent breathing phases resulted in an overall greater number of features tracked; however, this approach excluded some potential features that may have been trailed on certain phases only. Conversely, when tracking referred to a single phase, only the most stable features were trailed.

Fig. 2 presents median and variability of residual distance distributions in 4D CT and 4D MRI data for both approaches. In 4D CT, breathing cycle trailing (Panel B) looked more favorable than referring to maximum exhale phase (Panel A), due to block artifacts caused by breath to breath variation in respiratory amplitude and frequency [13]. In 4D MRI (Panel C and D), where these artifacts were not present, the two approaches were comparable.

Fig. 3 shows some common features between maximum exhale and inhale phases, represented in the maximum exhale phase for 4D lung CT (Panel A) and 4D liver MRI (Panel B). Identified features were vessels and bifurcations and liver vessels, such as the hepatic portal vein, in the 4D CT and 4D MRI, respectively.

Fig. 4 shows examples of internal trajectories (red line), obtained by SIFT feature (blue circle) interpolation. Panel A depicts a lung vessel bifurcation trailed throughout the breathing cycle. In all directions, we noted a change in the trajectory concavity near phase 5, attesting to the presence of inhale and exhale phases. Along the superior-inferior (SI) direction, feature displacements of 2.5 mm were found. Panel B reports the trajectory of a lung vessel, for which phase 4 and 7 (black circles) were obtained by means of cubic interpolation. The trajectory of hepatic portal vein feature is shown in Panel C. SI maximum and AP minimum corresponded to phase 6 (i.e. maximum inhale phase) and the trajectories looked to follow the respiratory cycle. Right-left (RL) position confirmed feature stability within the slice. Panel D reports the interpolated trajectory of a liver vessel, in which the RL position reflected a slice displacement (i.e. 4 mm) in phase 5, 6 and 7. In all trajectories the displacements were proportional to the voxel dimensions. This is due to the finite difference derivative computation for extremes identification, which involves the concentration of the voxel information in its center. As a result, an increasing accuracy in feature localization is expected by reducing the slice thickness in the image acquisition process.

Furthermore, the higher image quality of MRI images and the presence of internal gating [14] resulted in feature trajectories more similar to the theoretical breathing cycle trend. Another issue that compromises 4D CT trajectory estimation is the presence of changes in air content due to ventilation, that might affect the SIFT descriptor directly.

Finally, the application of the same main SIFT parameters to both image modalities could affect results in feature identification. Therefore, future work should be dedicated to provide an adaptive image-based quantitative selection of these parameters.

C. Computational Cost

The algorithm was derived from [10] and implemented in C++ (www.plastimatch.org). The computational performance of the application was evaluated using a 2.2 GHz Intel Core i7 processor. The CPU time required ranged between 2 min for 4D MRI and 50 min for 4D CT, proving to be highly correlated with the involved image size.

TABLE I. Number of matches, median and variability of error distributions at maximum exhale and inhale phases obtained by SIFT, compared to manual identification.

	# Matches	Median [mm]	Variability [mm]
L0-L5	300	12.98	18.22
SIFT 0-5	509	13.23	17.90

TABLE II. Number of common and preserved matches by keeping phase 0 as reference and by moving along the breathing cycle for 4D CT and 4D MRI data.

		4D CT	4D MRI
Phase 0 as reference	Common	60	37
	Preserved	117	24
	Total	**177**	**61**
Along breathing cycle	Common	77	41
	Preserved	9	1
	Total	**86**	**42**

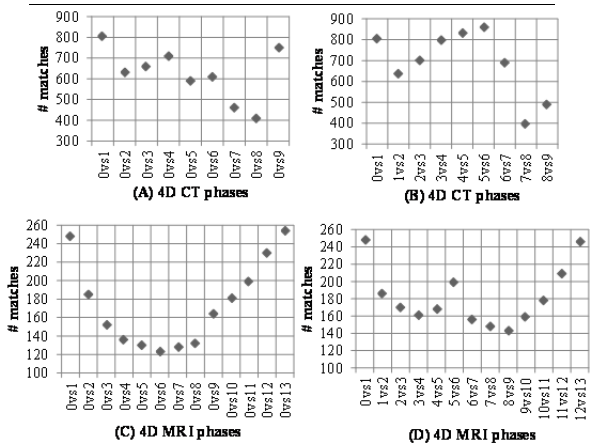

Fig.1. Number of matches for each phase. (A) 4D CT with phase 0 as reference. (B) 4D CT by moving along the breathing cycle. (C) 4D MRI with phase 0 as reference. (D) 4D MRI by moving along the breathing cycle.

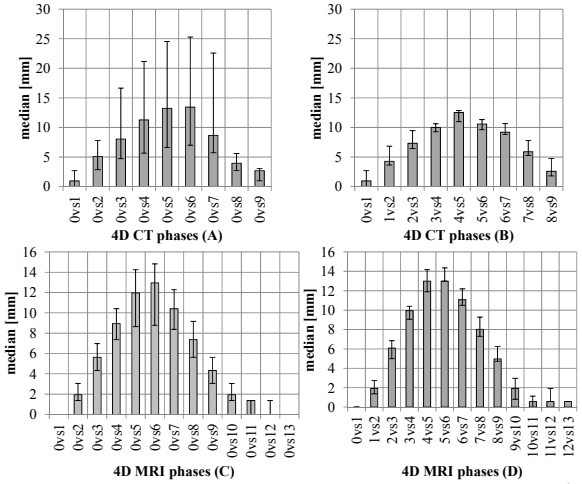

Fig. 2. Median and variability (computed as difference between 75th and 25th percentile) of residual distance distributions in 4D CT and 4D MRI data for both approaches. (A) 4D CT with phase 0 as reference. (B) 4D CT by moving along the breathing cycle. (C) 4D MRI with phase 0 as reference. (D) 4D MRI by moving along the breathing cycle.

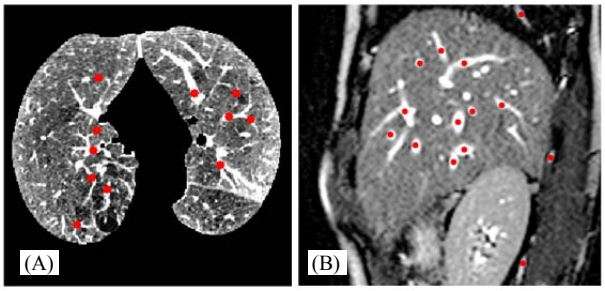

Fig.3. Feature tracking. Common features between maximum exhale and inhale phases, shown at maximum exhale for 4D lung CT (A) and 4D liver MRI (B), in trans-axial and sagittal slices, respectively.

IV. Conclusion

In this work, we showed that SIFT can be useful for automated internal motion estimation. We demonstrated the method's feasibility on a 4D lung CT and 4D liver MRI dataset, and validated SIFT matches detected on the 4D CT scan by comparing them with manually localized features. Main limits of the method are related to image and re-sorting quality and to qualitative selection of SIFT parameters. Furthermore, optimal trade-off between spatial and temporal resolution to reconcile feature localization accuracy with computational cost is needed. Future work will be focused on extensive SIFT match validation and breathing motion tracking applications for external beam radiotherapy.

Acknowledgment

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.

References

[1] K. Murphy, B. van Ginneken et al, "Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge", in *Medical Imaging, IEEE Transactions on*, 2011, vol. 30, p. 1901-1920

[2] G. Christensen, J. Kuhl, et al, "Introduction to the non-rigid image registration evaluation project (nirep)" in *Proceedings of SPIE*, 2006, p. 128-135

[3] P. J. Keall et al, "The management of respiratory motion in radiation oncology report of AAPM Task Group," in *Medical Physics*, vol. 33, 2006, p. 3874-900

[4] E. Naqa, D. A. Low et al, "Automated breathed motion tracking for 4D computed tomography", *IEEE Nuclear Science Symposium Conference Record*, 2003, vol.5, p. 3219-3222

[5] L. Kitchen and A. Rosenfeld, "Gray-level Corner Detection", *Pattern Recognition Letters*, North-Holland, 1982, p. 95-102

[6] W. Förstner, "A fast operator for detection and precise location of distinct points, corners and centres of circular features", *ISPRS Intercommission Workshop*, Interlaken, 1987

[7] C. Harris and M. Stephens, "A combined corner and edge detector", *Plessey Research Roke Manor*, United Kingdom, 1988, pp. 147-152.

[8] D.G. Lowe, "Object recognition from local scale-invariant features", *proc. of the 7th IEEE Int. Conf. on Comp. Vis. (ICCV)*, Greece, 1999, p. 1150–1157

[9] D.G. Lowe, "Distinctive image features from scale-invariant keypoints". *Int. J. of Comp. Vis.*, Canada, 2004, p. 91-110

[10] W. Cheung and G. Hamarneh, "n-SIFT: n-Dimensional Scale Invariant Feature Transform", *IEEE Transaction on Image Processing*, Washington, 2009, p. 2012-2021

[11] S. Allaire et al, "Full Orientation Invariance and Improved Feature Selectivity of 3D SIFT with Application to Medical Image Analysis", *IEEE Computer Vision and Pattern Recognition* , Canada, 2008, p.1-8

[12] R. Castillo et al, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets", in *Physics in Medicine and Biology*, vol. 54, 2009, p. 1849-1870

[13] E. Castillo et al, "Compressible image registration for thoracic computed tomography images", in *Journal of Medical and Biological Engineering*, vol. 29, 2009, p. 222-233

[14] M. von Siebenthal, "Analysis and Modelling of Respiratory Liver Motion using 4DMRI", PhD Thesis, ETH Zürich, 2008

[15] H.B. Mann, D.R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other". *Annals of Mathematical Statistics* vol. 18, 1947, p. 50-60

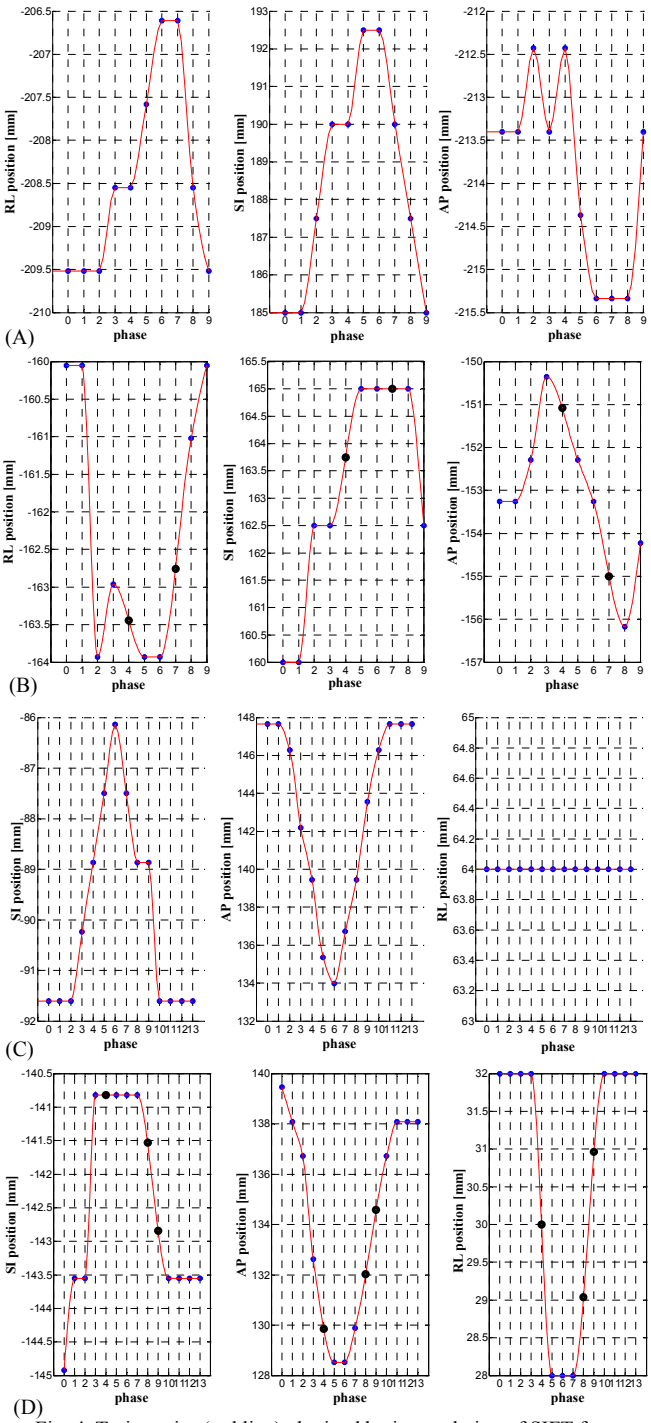

Fig. 4. Trajectories (red line) obtained by interpolation of SIFT features (blue circle). Absent features are reported (black circle). (A) Bifurcation trajectory of lung (4D CT). (B) Vessel trajectory of lung (4D CT), with absence in phase 4 and 7. (C) Hepatic portal vein trajectory (4D MRI). (D) Vessel trajectory of liver (4D MRI), with no match in phase 4, 8 and 9.