

# Location Registration and Recognition (LRR) for Longitudinal Evaluation of Corresponding Regions in CT Volumes

Michal Sofka and Charles V. Stewart

Department of Computer Science  
Rensselaer Polytechnic Institute  
Troy, New York 12180-3590  
{sofka,stewart}@cs.rpi.edu \*

**Abstract.** The algorithm described in this paper takes (a) two temporally-separated CT scans,  $I_1$  and  $I_2$ , and (b) a series of locations in  $I_1$ , and it produces, for each location, an affine transformation mapping the locations and their immediate neighborhood from  $I_1$  to  $I_2$ . It does this without deformable registration by using a combination of feature extraction, indexing, refinement and decision processes. Together these essentially “recognize” the neighborhoods. We show on lung CT scans that this works at near interactive speeds, and is at least as accurate as the Diffeomorphic Demons algorithm [1]. The algorithm may be used both for diagnosis and treatment monitoring.

## 1 Introduction

This paper is concerned with volumetric registration tools to assist in longitudinal monitoring of pre-cancerous and cancerous locations in the body. Our experimental focus is lung cancer, the deadliest of all cancers [2]. The idea is that nodule locations, or any other locations of interest, having been selected within one CT scan,  $I_1$ , manually or using CAD, are automatically located in a subsequent (or earlier) scan,  $I_2$ . The regions surrounding the nodule and its corresponding location in the two scans are aligned (rigid or affine) in preparation for either display or subsequent quantitative analysis. This problem will become important with continuing increases in the number and size of CT scans and with increasing use of longitudinal studies that require alignment of scans [3].

Implicit in this problem statement is a *decision* about the neighborhood of  $I_2$  that corresponds to the chosen neighborhood of  $I_1$ . In essence, the problem requires *recognizing* the neighborhood as part of the registration process. We refer to this problem as the *Location Registration and Recognition* or LRR problem.

The obvious solution to the LRR problem is to apply a global deformable registration technique, followed by local affine registration to produce the final transformations [3]. We propose a different approach because deformable registration (a) is not completely solved, (b) it is slow, and (c) solving it is not necessary to solve the LRR problem. Moreover, a good solution to the LRR problem can produce correspondences for deformable registration. Finally, even

---

\* This article was supported by the DOD and the Medical University of South Carolina under DOD Grant No. W81XWH-05-1-0378. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Department of Defense or the Medical University of South Carolina.

if advances in deformable registration make it practical for the LRR problem, a decision component is still required to ensure correct neighborhood matching.

Our approach is feature-based, involving a pre-processing step to extract a variety of features and keypoints distributed throughout the scan volumes even in low-contrast regions. The main algorithm is applied to each pre-selected (e.g. nodule) location separately.<sup>1</sup> It involves steps of hypothesis generation, hypothesis refinement, and decision. Keypoints near the location are described using shape-contexts [4] and matched to generate initial hypotheses. Hypotheses are rank-ordered and refined in succession using a robust generalization of ICP [5]. Decisions about acceptability of the final alignment recognition are based on accuracy and stability measures combined in an SVM classifier. All components of the algorithm are designed to be robust to differences between scans caused by changes in scanning procedures, patient positioning, and physical changes in the lung.

Our experiments on lung CT scans are designed to guide algorithm design decisions and to validate the overall algorithm. Comparisons are made with the Diffeomorphic Demons [1] registration algorithm. The overall results show that LRR is in most cases in agreement with the resulting deformation fields and sometimes even provides better alignment. Our initial implementation runs in 10 seconds on average per initial location.

In summary this paper makes three main contributions: (1) the overall algorithm, (2) evaluation of shape-context methods [4] for indexing and for generating an initial transformation estimate in CT scans, and (3) a decision criteria for determining when a region of one scan has been correctly recognized and aligned in a fraction of a second.

## 2 Background

Most previous lung CT registration algorithms have been based on B-splines, demons, thin-plate splines, radial basis functions, and alignment of surfaces (see [3] for a review). Some algorithms are quantitatively compared in [6]. Other algorithms, such as in [7] work by matching and aligning nodule locations that have been specified in both volumes. The approach in [8] finds a single point-wise correspondence to a given point by relative configurations to pre-computed stable features. Neither of the two algorithms [7, 8] aligns the local regions. Finally, several applications such as aligning a volume of interest around pulmonary nodules [9] work by template matching or by minimizing mean-squared difference of the intensity volumes. These applications can use the proposed solution to the LRR problem which also includes the verification of the final alignment.

Keypoint, landmark and corner extraction algorithms [10] have been widely studied in both medical imaging and in computer vision. Recent work in computer vision has emphasized the computation of descriptor vectors at keypoint locations, followed by indexing methods to match them between images [11]. One set of these methods, shape-contexts [4], which emphasizes the distribution

---

<sup>1</sup> In future versions, neighborhood proximity will be exploited.

<p><b>Input:</b> Image volumes <math>I_1</math> and <math>I_2</math> and a set of locations <math>\mathcal{L} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}</math>, pre-selected manually or semi-automatically from <math>I_1</math>.</p> <p><b>Output:</b> Set <math>\{\hat{\boldsymbol{\theta}}_1, \dots, \hat{\boldsymbol{\theta}}_N\}</math>, where each <math>\hat{\boldsymbol{\theta}}_k</math> is a vector of affine parameters or the zero-vector <math>\mathbf{0}</math> indicating that no transformation was found.</p> <p><b>Preprocessing:</b> Extract features (Sec. 3.1), and extract keypoints and descriptors (vectors) (Sec. 3.2) in both volumes. This step may be done offline. Form a k-d tree storing all descriptors from <math>I_2</math>.</p> <p><b>Repeat</b> for each pre-selected location <math>\mathbf{x}_k</math>:</p> <ol style="list-style-type: none"> <li>1. Establish a local 3D neighborhood <math>\mathcal{N}(\mathbf{x}_k)</math> surrounding location <math>\mathbf{x}_k</math>.</li> <li>2. Gather keypoints from <math>I_1</math> within the neighborhood <math>\mathcal{N}(\mathbf{x}_k)</math> into a set <math>\mathcal{U}_k</math> (Sec. 3.2).</li> <li>3. Indexing: for keypoint in <math>\mathcal{U}_k</math>, search the k-d tree to find the keypoint in <math>I_2</math> with the closest descriptor. The two keypoints form a hypothesized match (Sec. 3.2).</li> <li>4. Order the hypothesized matches by increasing descriptor distance. Consider the best <math>M = 20</math>.</li> <li>5. <b>Repeat</b> <ol style="list-style-type: none"> <li>(a) Pick the next match and compute an initial rigid transform by aligning local keypoint coordinates.</li> <li>(b) Estimate affine transformation parameters <math>\hat{\boldsymbol{\theta}}_k</math> of the mapping function <math>\mathbf{T}_k</math> using robust generalization of ICP (Sec. 3.3). This is applied between features in 3D neighborhoods <math>\mathcal{M}(\mathbf{x}_k)</math> and <math>\mathcal{M}(\mathbf{x}'_k)</math>, <math>\mathbf{x}'_k = \mathbf{T}_k(\mathbf{x}_k)</math>.</li> <li>(c) Apply the decision classifier (Sec.3.4): If <math>\hat{\boldsymbol{\theta}}_k</math> is <i>correct</i>, save it and proceed to the next pre-selected location. Otherwise continue to the next match (Step 5a).</li> </ol> </li> <li><b>Until</b> all matches from the list have been processed.</li> <li>6. No transformation has been found for <math>\mathbf{x}_k</math>, so set <math>\hat{\boldsymbol{\theta}}_k = \mathbf{0}</math>.</li> </ol>
---

**Fig. 1.** LRR algorithm outline.

of features and their orientations, has been generalized to range data [12] and to landmark-based deformation field regularization [13].

Final decisions on alignment accuracy are often left to manual assessment. Since this is not always practical, there is a need for reliable automatic methods. The technique in [14] for aligning 2D images applies statistical tests of randomness based on the number of overall and consistent keypoint matches. The algorithm in [15] combines measures of accuracy, consistency, and stability checks. Similar methods have not yet been applied to 3D medical imaging.

### 3 Algorithm

The algorithm is outlined in Fig. 1, with important details described below.

#### 3.1 Feature Extraction

Features summarize the structural content of the image volumes and are used in all aspects of the LRR process. In order to ensure that features are available to drive the LRR algorithm anywhere in the volume, the emphasis is to extract multiscale features distributed throughout the volumes (coverage) without missing anatomical structures (completeness). We rely on robust statistical estimation to ensure that correspondences generated when a feature in one volume has no

analog in the other are ignored. During the extraction process, features are categorized as sheet-like, tube-like or corner-like according to the local intensity distribution.

The datasets are resampled to be isotropic. At each voxel location  $\mathbf{x}$ , the intensity gradient,  $\nabla I(\mathbf{x})$ , is computed and an outer product (auto-correlation) matrix  $\mathbf{M}(\mathbf{x})$  is computed as a gaussian-weighted average of  $(\nabla I(\mathbf{x}))(\nabla I(\mathbf{x}))^\top$  over a neighborhood of  $\mathbf{x}$ . A strength is assigned to each point:  $m(\mathbf{x}) = \text{trace}(\mathbf{M}(\mathbf{x}))$ . Pruning discards features that have strength below  $\mu_m + \sigma_m$ , where  $\mu_m$  and  $\sigma_m$  are median and std. dev. strength values computed in overlapping neighborhoods ( $30 \times 30 \times 30$  mm) throughout the volume. Features are classified as 1d (sheet-like), 2d (tube-like) or 3d (corner-like) based on an eigen-decomposition of  $\mathbf{M}(\mathbf{x})$ . The associated eigenvectors form a local orthonormal basis, with sign ambiguities resolved by the signs of the directional derivatives. Non-max suppression is applied according to the dimensionality of the features (e.g. all 3 directions for corners, 2 for tubes, 1 for sheets). Surviving locations are interpolated to subvoxel accuracy, so that features can be located between voxels, rather than strictly at voxel coordinates. Next, these locations are sorted by  $m(\mathbf{x})$  values. Starting with the highest strength point and continuing until the list is exhausted, points are accepted one-by-one. When a feature (at location  $\mathbf{x}$ ) is accepted, remaining points whose strength is less than  $\tau m(\mathbf{x})$  ( $\tau = 0.8$ ) and within distance  $\Omega$  of  $\mathbf{x}$  ( $\Omega = 4$  mm, currently) are eliminated. Fig. 2 shows example features as part of a checkerboard mosaic of X-Y slices taken from aligned volumetric neighborhoods.

### 3.2 Keypoints and Indexing

Corner-like features and their associated neighborhoods are used to form keypoints for indexing. These are found by repeating the foregoing process with  $\Omega = 6$  mm. A descriptor is computed at each keypoint. This is done using a shape contexts approach [4, 12]. A spherical coordinate system is formed in the keypoint’s local coordinate system. Four angular bins and four radial bins are formed, (up to  $r = 30$  mm), together with an additional central bin. In each bin, the 3d gradient directions of all features that fall into the bin are averaged without gradient magnitude weighting. The resulting directions are concatenated into a 195-component descriptor vector. This is nearly the size of 2d image descriptors and much smaller than descriptors used for range data [12]. This smaller descriptor is sufficient for 3D volumes due to more distinctive spatial distribution of points than in range data, and it results in much faster matching.

Since our whole approach hinges on effective keypoint matching, it is important to briefly examine some empirical tests that motivate the approach and guide design decisions. These experiments are based on a small set of CT volume pairs (see Section 4). For each pair, Deformable Demons algorithm [1], currently the best algorithm for global lung CT registration [6], is applied to produce a transformation  $\mathbf{T}_G$ . For any keypoint match, if, after application of  $\mathbf{T}_G$  to the keypoint from  $I_1$ , the keypoints are within 9 mm of each other and their axes are within  $20^\circ$  then this is considered a “good” match. We use this to examine

the results of keypoint match at a random sampling of 1000 locations in and around the lung in each scan. Here is a summary of the results:

- If we look only at the *closest* keypoint for each location, it forms a “good” match 40% of the time. This shows the promise of the approach, but clearly indicates that more than just the closest keypoint must be considered.
- Within 30 mm of the keypoint, a good match exists 96% of the time. This shows much greater promise.
- Since we can not test all keypoints in  $\mathcal{N}(\mathbf{x}_k)$  (on average 160 within 30 mm), we must consider the effects of rank-ordering of the matches. In this case 95% of the  $\mathbf{x}_k$  have at least one good match within the top 20. If we replace our ordering by descriptor distance with Lowe’s widely-used distance ratio [11], this percentage drops slightly to 93%. These encouraging results justify the design path taken here.<sup>2</sup>

### 3.3 Transformation Estimation

The estimation procedure finds the transformation  $\mathbf{T}_k$  with parameters  $\boldsymbol{\theta}_k$  that best aligns the neighborhoods  $\mathcal{M}(\mathbf{x}_k)$  and  $\mathcal{M}(\mathbf{x}'_k)$  (size 50 mm),  $\mathbf{x}'_k = \mathbf{T}_k(\mathbf{x}_k)$  by applying a robust form of ICP:

$$\mathcal{F}(\boldsymbol{\theta}_k) = \sum_{(\mathbf{p}_i, \mathbf{q}_j) \in \mathcal{C}} \rho(d(\mathbf{p}_i, \mathbf{q}_j; \boldsymbol{\theta}_k)/\sigma), \quad (1)$$

where  $\rho$  is the M-estimator robust function and  $\mathcal{C}$  is a set of correspondences, each established between a point  $\mathbf{p}_i$  from  $\mathcal{M}(\mathbf{x}_k)$  and a point  $\mathbf{q}_j$  from  $\mathcal{M}(\mathbf{x}'_k)$ . Distance function  $d$  is feature-type specific and is normalized by scale  $\sigma$  or it is computed using covariance of estimated parameters and feature locations [5]. The estimation alternates in two steps: 1) finding parameters  $\hat{\boldsymbol{\theta}}_k$  by minimizing the squared errors after applying robust function and weights, and 2) establishing correspondences  $\mathcal{C}$  and recomputing the weights. Estimation is symmetric and  $\mathcal{C}$  is randomly sampled to have size under 4000 for speedup.

### 3.4 Alignment Verification

Verification is used to reject incorrect hypothesized transformations so that further initializations can be processed. This step is based on an eight dimensional vector  $\boldsymbol{\Phi}_k$  obtained using forward ( $\mathbf{T}_k$ ) and backward ( $\mathbf{T}_k^{-1}$ ) transform. For each direction, the vector consists of the average error between fixed and mapped moving points and of the maximum trace of the transfer error covariance of the parameters,  $\mathbf{J}\mathbf{C}\mathbf{J}^\top$ , computed for regularly sampled point locations  $\mathbf{p}_i$ .  $\mathbf{J}(\mathbf{p}_i) = \partial\mathbf{T}_k/\partial\boldsymbol{\theta}_k$ . It also contains average difference between tangent angles of tubular features in moving and fixed neighborhoods as well as difference

<sup>2</sup> This 95% number does not represent an upper bound on the overall success rate of the algorithm since the alignment process can often overcome initial misalignments from keypoint matches that are not quite “good”.

between normal angles of sheet features. The vector  $\Phi_k$  is computed for all initializations, even those that did not provide a correct final alignment. Alignments with changes are left out. Therefore the output of the classifier is binary: aligned or misaligned.

## 4 Experiments

The data for our experiments are screening and diagnostic lung CT scans taken approximately one year apart. The original volumes were cropped and their sizes range from  $280 \times 180 \times 30$  to  $310 \times 250 \times 90$ . Their spacing ranges from  $0.70 \times 0.70 \times 2.5$  mm to  $0.89 \times 0.89 \times 5$  mm.

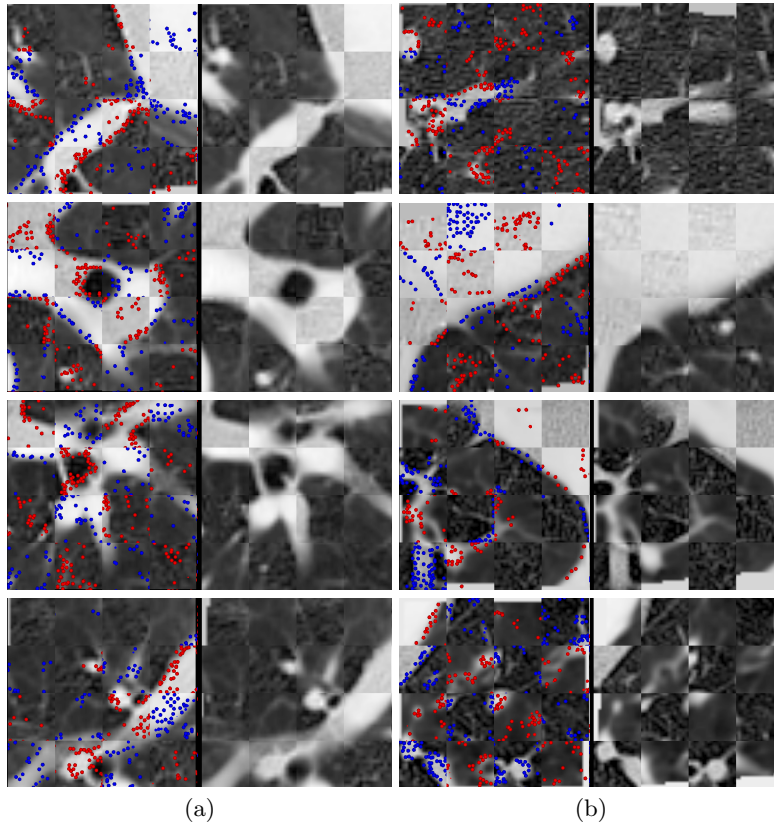
Our first experiment tests the estimation component. The results are grouped based on initialization being good (using a good match) or bad (Section 3.2). Manually marking final result as aligned/misaligned we have learnt that for two pairs all good (and several bad) initializations produced successful alignment (Table 1). Corresponding neighborhoods with partial alignment but visible changes (see Figure 2) were marked as undefined. The estimation did not produce a successful alignment for 2 cases each in three volumes and in 7 cases for one volume. Such cases are automatically identified and a different initialization tried. The overall algorithm succeeds on 97.3% of all initial locations.

Verification algorithm is tested for each pair using the leave-one-out procedure producing a false positive (FP) and false negative (FN) counts. False positives are alignments deemed aligned by the algorithm but manually graded as misaligned. False negatives are those deemed misaligned but in fact produced a successful alignment. The total number of false positives was 2 out of 574 neighborhood pairs, while the number of false negatives was 29. Therefore the algorithm is not likely to accept an incorrect alignment and rejection simply means that another initialization is tried. The most ambiguous for the algorithm (and manual grading) are pairs of local regions with changes between them (marked as undefined): out of 26 classified as aligned, the FN count was 7.

Our overall evaluation compares the location registration to the diffeomorphic demons algorithm [1]. As before (Section 3), we initialize our technique at  $M = 100$  random locations. We obtain a set of point locations by regularly sampling in  $\mathcal{M}(\mathbf{x}_k)$  and compute RMS error between the points mapped with LRR final transform and with the deformation field. On average the error is below 2.0 mm except for one pair for which it was 2.36 mm. Several examples of these are shown in Figure 2. They represent examples where the LRR is in agreement with the deformable registration but also examples where the location registration alignment appears to be more accurate. The algorithm correctly ignored the changes which tend to cause misalignments in the deformable registration.

estimation initializ.	success		failure		un- def.
	good	bad	good	bad	
Pair 1	63	10	2	16	9
Pair 2	81	5	2	9	3
Pair 3	86	3	0	11	0
Pair 4	78	6	2	11	3
Pair 5	92	6	0	2	0
Pair 6	43	17	7	22	11

**Table 1.** Estimation results using 600 initializations. Numbers indicate successful/unsuccessful alignments given a good (2nd/4th col.) and a bad (3rd/5th col.) match. Estimation succeeds in many cases where the initialization is poor (3rd col.) and the number of estimation failures is small (4th col.).



**Fig. 2.** Examples of LRR (1st and 3rd col.) vs. deformable registration (2nd and 4th col.). Agreement of both results **(a)**, and examples where LRR alignment is better **(b)**. Features detected in fixed (blue) and moving (red) images drive the registration and robust estimation handles outliers (e.g. the case with changes on the bottom right which is partially aligned in the lower half).

All steps take on average 10 seconds per location on a laptop PC with Pentium 4, 2GHz processor, 2GB of RAM, which includes speedup by discarding initializations with high rotations (10% initializations in the first 20 have rotation more than  $45^\circ$ ). Note that further speedups are possible through improved data structures, code profiling and resulting optimizations, running the algorithm on a multi-core or multi-processor machine parallelizing each component, or trying multiple initializations concurrently.

## 5 Discussion

We have presented a novel algorithm for Location Registration and Recognition (LRR) without solving the deformable registration first or simultaneously. We have shown the applicability of shape contexts for indexing and how they can be used to obtain an initial transform. Overall experiments on the longitudinal CT scans of the lung show that the algorithm is able to handle changes between

the local regions, is at least as accurate as the deformable registration result, and runs at interactive speeds. Novel verification algorithm decides whether a region of one scan has been correctly recognized and aligned.

One potential difficulty of the algorithm could be regions that do not have much structure to drive the registration. To correct this, the neighborhood might need to be chosen adaptively. This together with more in depth experimental validation and exploring other applications are the main directions in our future work.

## References

- [1] Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Non-parametric diffeomorphic image registration with the demons algorithm. In: Proc. 10th MICCAI, Brisbane, Australia (2007) 319–326
- [2] Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., Thun, M.J.: Cancer statistics, 2007. *CA: A Cancer Journal for Clinicians* **57** (January 2007) 43–66
- [3] Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: a survey. *IEEE T. Med. Imag.* **25**(4) (2006) 385–405
- [4] Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE T. Pattern Anal. Machine Intell.* **24**(4) (June 2002) 509–522
- [5] Sofka, M., Yang, G., Stewart, C.V.: Simultaneous covariance driven correspondence (CDC) and transformation estimation in the expectation maximization. In: Proc. CVPR, Minneapolis, MN (18–23 June 2007)
- [6] Urschler, M., Kluckner, S., Bischof, H.: A framework for comparison and evaluation of nonlinear intra-subject image registration algorithms. In: ISC/NA-MIC Workshop on Open Science at MICCAI 2007. (2007)
- [7] Betke, M., Hong, H., Thomas, D., Prince, C., Ko, J.P.: Landmark detection in the chest and registration of lung surfaces with an application to nodule registration. *Med. Image Anal.* **7** (2003) 265–281
- [8] Okada, K., Huang, X.: Robust click-point linking: Matching visually dissimilar local regions. In: In proc. IEEE Int. Workshop on Beyond Multiview Geometry: Robust Estimation and Organization of Shapes from Multiple Cue. (2007)
- [9] Reeves, A.P., Chan, A.B., Yankelevitz, D.F., Henschke, C.I., Kressler, B., Kostis, W.J.: On measuring the change in size of pulmonary nodules. *IEEE T. Med. Imag.* **25**(4) (2006) 435–450
- [10] Hartkens, T., Rohr, K., Stiehl, H.S.: Evaluation of 3D operators for the detection of anatomical point landmarks in MR and CT images. *Comput. Vis. Image Und.* **86**(2) (May 2002) 118–136
- [11] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vis.* **60**(2) (November 2004) 91–110
- [12] Frome, A., Huber, D., Koller, R., Buehler, T., Malik, J.: Recognizing objects in range data using regional point descriptors. In: Proc. Eighth ECCV. (2004)
- [13] Urschler, M., Zach, C., Ditt, H., Bischof, H.: Automatic point landmark matching for regularizing nonlinear intensity registration: Application to thoracic CT images. In: Proc. 9th MICCAI, Copenhagen, Denmark (2006) 710–717
- [14] Brown, M., Lowe, D.: Automatic panoramic image stitching using invariant features. *Int. J. Comp. Vis.* **74**(1) (August 2007) 59–73
- [15] Yang, G., Stewart, C.V., Sofka, M., Tsai, C.L.: Registration of challenging image pairs: Initialization, estimation, and decision. *IEEE T. Pattern Anal. Machine Intell.* **23**(11) (November 2007) 1973–1989