

Medical image registration using machine learning-based interest point detector

Sergey Sergeev, Yang Zhao, Marius George Linguraru* and Kazunori Okada,
sergeev@sfsu.edu, zhaoyang@sfsu.edu, milingura@cmmc.org, kazokada@sfsu.edu

Department of Computer Science, San Francisco State University

*Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center,
Washington DC

ABSTRACT

This paper presents a feature-based image registration framework which exploits a novel machine learning (ML)-based interest point detection (IPD) algorithm for feature selection and correspondence detection. We use a feed-forward neural network (NN) with back-propagation as our base ML detector. Literature on ML-based IPD is scarce and to our best knowledge no previous research has addressed feature selection strategy for IPD purpose with cross-validation (CV) detectability measure. Our target application is the registration of clinical abdominal CT scans with abnormal anatomies. We evaluated the correspondence detection performance of the proposed ML-based detector against two well-known IPD algorithms: SIFT and SURF. The proposed method is capable of performing affine rigid registrations of 2D and 3D CT images, demonstrating more than two times better accuracy in correspondence detection than SIFT and SURF. The registration accuracy has been validated manually using identified landmark points. Our experimental results shows an improvement in 3D image registration quality of 18.92% compared with affine transformation image registration method from standard ITK affine registration toolkit.

Keywords: Image registration, machine learning, neural network, interest point detection, SIFT, SURF

1. MOTIVATION

Interest point detection (IPD) is crucial for feature- and landmark-based medical image registration¹ which requires accurate localization of corresponding anatomical features across target images. Well-known IPD methods, such as SIFT³, SURF⁴, and Harris Corner Detector⁵, have previously been applied to detect these features automatically in medical image applications¹²⁻¹⁵. Despite their promise, these methods suffer from relatively low detection rate of corresponding features when applied to real clinical applications, such as abdominal CT scan registration. The above mentioned standard IPD methods are designed to select interest points (IPs) that conform to a certain first principle such as scale invariance¹⁶ and cornerness⁵. These principles are not designed to explicitly improve accuracy for point-correspondence detection so that they can increase the chance for point-correspondence failures due to domain-specific data and noise structures, leading to lower registration accuracy. Therefore, choosing an IPD principle that directly maximizes performance of correspondence detection would improve the registration accuracy. In this paper, we propose a novel machine learning (ML)-based IPD method that is designed to maximize the performance of the feature-based registration by using the same principle for feature selection and IPD. The target application of this work is registration of clinical abdominal CT scans.

2. METHODOLOGY

The overall structure of the registration framework is depicted in Figure 1 and it can be divided into three successive phases:

Phase 1: Feature Selection

- Exhaustively building point-wise ML detectors for every data point
- Performing feature selection according to trained detectors' cross validation (CV) performance

Phase 2: Feature Detection and Matching

- Establishing correspondences by applying the selected ML detectors to target images

Phase 3: Image Registration

- Calculation of transformation model using RANSAC⁹ algorithm
- Transform the floating image using the estimated transformation model

Suppose a pair of images I^F (*fixed*) and I^M (*moving*) with $N_x * N_y * N_z$ voxels are given, where spatial variables x_F for I^F and x_M for I^M are related by a coordinate transform $T: x_F = T(x_M)$. A feature-based registration first detects a set of K corresponding points in I^F and I^M , $\{ (x_F^k ; x_M^k) / k = 1, \dots, K \}$ and then estimates the transformation T from these correspondences.

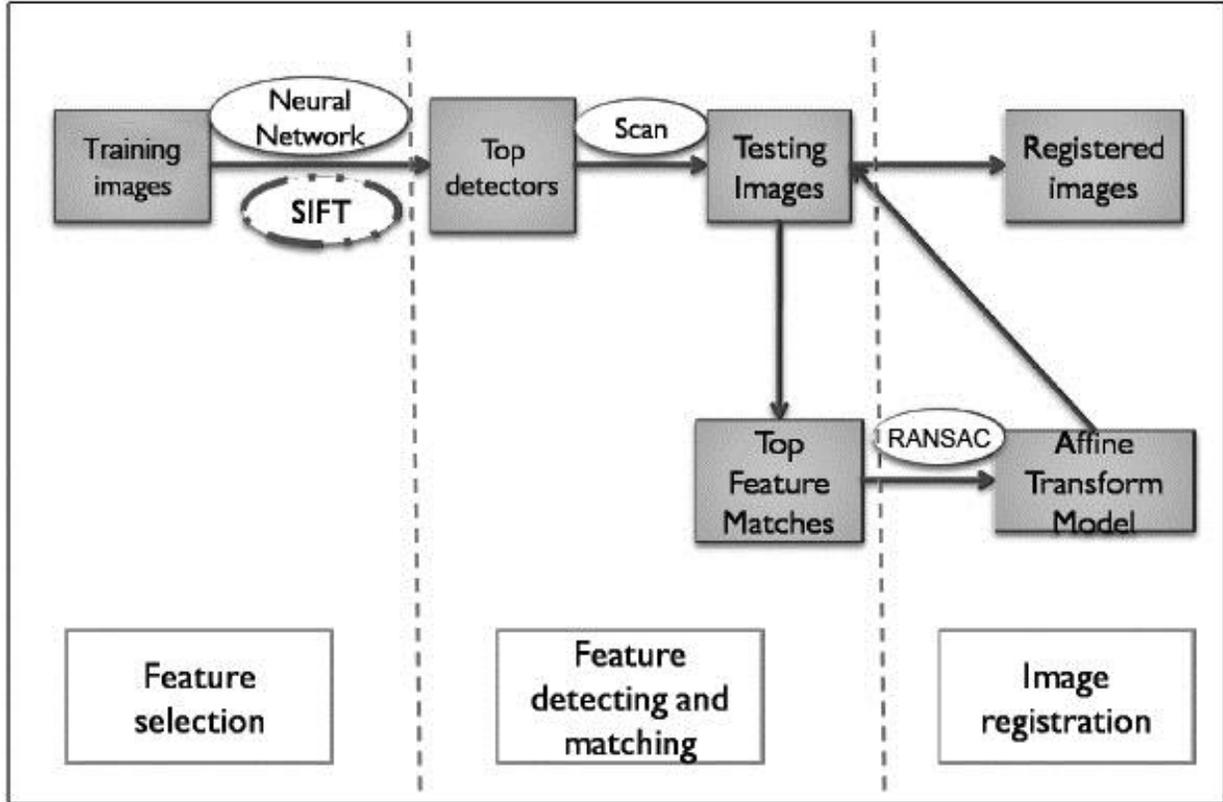


Figure 1. Overview of ML-based Registration Framework

Our IPD consists of K feed-forward neural network (NN) classifiers $\{ f_k(y(c)) = p \}$, each of which is trained to detect a distinct easy-to-find feature point x_k . The classifier maps an intensity feature vector y derived from a local intensity window centered at c to a likelihood p of c being the feature point x_k . NN classifier has a static structure consisting of 2 hidden layers with 60-neurons each. We use two features as input vectors: 1) raw intensity values normalized to $[-1, 1]$ and 2) SIFT descriptor³. The standard sigmoid function¹¹ is used as a transfer function in each neuron. The output layer of NN consists only of a single neuron whose value in $[0, 1]$ indicates the IP likelihood p . Higher p indicates higher likelihood of a window center c being the interest point (IP) sought after.

A trained NN is scanned by sliding the input window within an image. The window center c that yields the maximum likelihood defines the detected IP,

$$x_k = \operatorname{argmax}_{c \in \Omega} f_k(y(c)) \quad (1)$$

where Ω is the set of all pixel/voxel locations in the image. This is repeated to another image, yielding point-correspondences $(x_F^k ; x_M^k)$ without any post point matching process.

A set of K NN detectors $\{f_k(y(c))\}$ is constructed in three successive steps. First, we characterize each *valid* data point with *detectability index*. This index is computed as the leave-one-out CV (LOOCV) score of a NN classifier trained to detect an IP, given a set of M training CT scans. As a pre-process, the training scans are non-linearly registered to a fixed reference image by using B-spline based IRTK registration tool⁸. Next, background in each image is removed by intensity thresholding. *Valid data points* (VDP) are defined by the points that are a part of the patient's abdomen for at least one of the training images in the pre-registered coordinate frame. For computational efficiency and redundancy reduction, we uniformly sub-sample these data points, yielding R VDPs. At each VDP $r = 1, \dots, R$, LOOCV is repeated M times. At each run, $M-1$ local windows centered at a VDP are used as positive samples (Figure 2 gives examples of image patterns used as “positive” input vectors for NN training), while $2M$ negative samples are collected randomly at various locations and scans (Figure 3 shows randomly chosen “negative” inputs). Conjugate gradient descent¹⁷ is used for optimization. The resulting LOOCV score (e.g. a detection rate with M trials) is associated with a VDP as its detectability index.

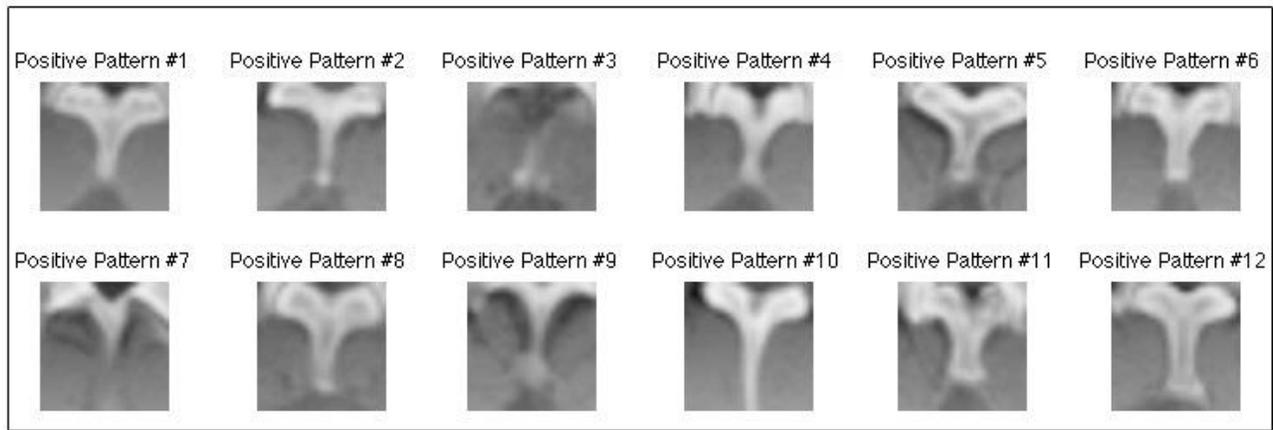


Figure 2. “Positive” input patterns from central vertebrae region for training of neural network detector.

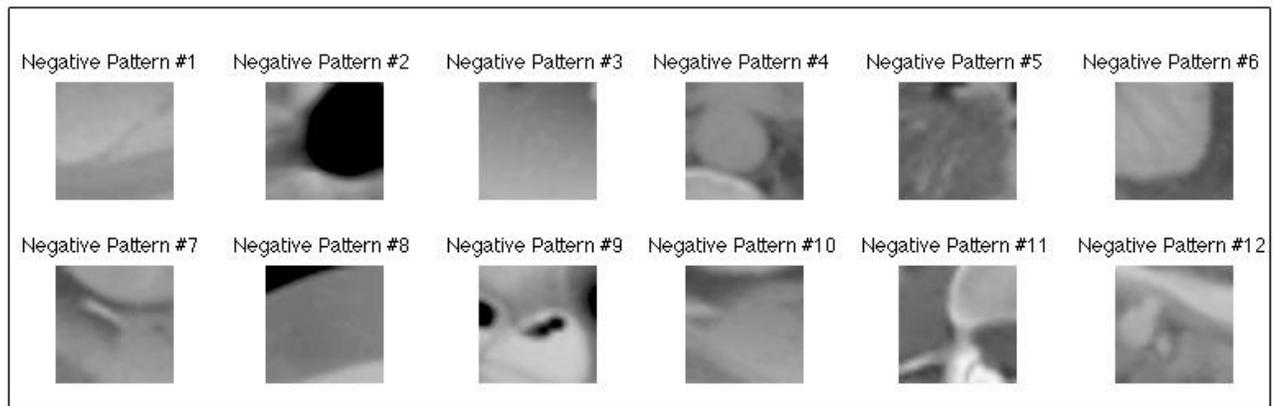


Figure 3. Randomly chosen “negative” image patterns for central vertebrae region used for training of NN detector.

Second, we select the top K VDPs from the total of R VDPs according to the detectability index. Finally, we build K IPDs by training NN classifiers using all M positive samples and the same negative samples at each selected IP. When top K IPDs are trained, we run the correspondence detection on pairs of CT scans. A correspondence between points is established when IPD detects a pattern in both images. After all chosen IPDs on the pair of images are run we will have up to K point correspondences which are used as input to RANSAC⁹ algorithm. RANSAC or Random Sample Consensus is an iterative method for estimating the parameters of a mathematical model from a set of observed data which is contaminated by a significant percentage of gross errors. It internally detects and removes outliers to get a transformation

model. This is a non-deterministic algorithm and the resulting probability increases along with the number of input points and iterations. The output of RANSAC algorithm is a N^3 affine transformation matrix which is applied to a floating 3D CT scan image to be aligned with a reference image.

3. RESULTS

3.1. Data

The dataset used in the project was obtained from NIH Clinical Center. It consists of two sets: a training a testing set. Both of them contain abdominal 3D CT images displaying a abdominal area of human body and consist of 50 slices with 5mm thickness. Training set consists of 29 3D images well registered to each other, which means that various image patterns in them are well aligned. Training dataset consists of cases with normal anatomy obtained from a control population of healthy volunteers with no known pathologies while the testing set consist of 12 diseased cases with surgical resections of spleens and kidneys (e.g. splenectomy and nephrectomy). The original dimensions of 3D scans were $N_x = 512$, $N_y = 512$ and $N_z = 50$ with 12-bit intensity depth. For the project we down-sampled them using Gaussian pyramid blurring to $N_x = 256$, $N_y = 256$ and $N_z = 50$, but left the intensity of images unchanged.

3.2. Comparison of correspondence detection between ML-based IPD, SIFT and SURF in 2D

We evaluated the correspondence detection quality of the proposed method and compared it against SIFT³ and SURF⁴ algorithms. It was performed on three 2D CT image sets sampled at different axial slice-depths in order to exploit the public 2D implementations of SIFT⁷ and SURF¹⁰ with their default parameters. These public SIFT and SURF do not provide a function to rank VDPs required to perform feature selection. To compare both SIFT and SURF to our ML-based IPD, we chose the top 20 SIFT and SURF IPs by associating (e.g. ranking) them with the detectability index of the nearest VDP computed by our MI-based method. Figure 4 shows illustrative examples of these three slices.

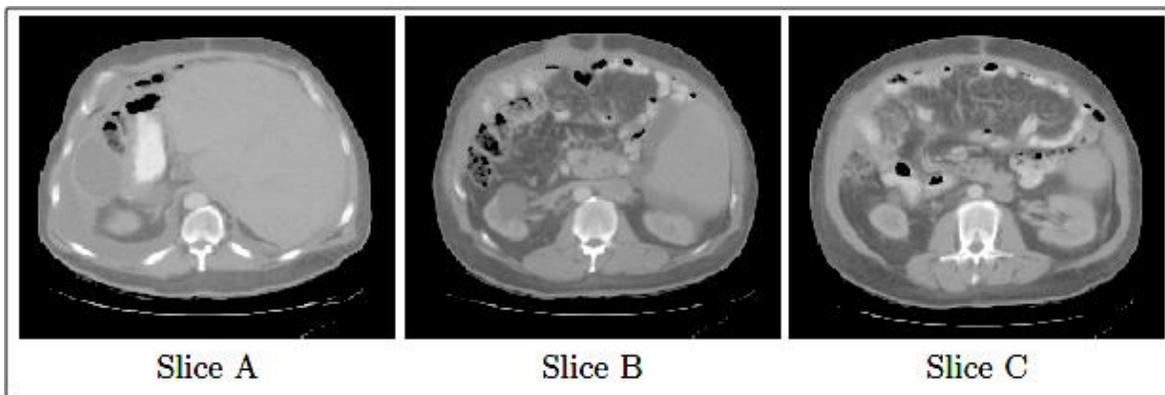


Figure 4. Three slice depths used for 2D comparative study.

	Slice A	Slice B	Slice C
ML-based	36.28	44.16	48.54
SIFT	15.27	17.54	18.47
SURF	11.18	12.79	13.26

Table 1. Average percentage (%) of correctly identified point correspondences by ML-based IPD, SIFT and SURF in 2D

The numbers in the Table 1 denote averaged percentages of correctly identified correspondences between all pairs

of CT images of a corresponding testing set (e.g. Slice A, B, C). Correspondence detection was evaluated using the top 20 features selected by NN and SIFT/SURF algorithms. Each estimated correspondence is visually inspected by an expert and classified to either correct or mismatch. The 2D sets were extracted from the original 3D training/testing sets. Each of the manually prepared 2D sets (abbreviated in Table 1 and Figure 4 as *A*, *B* and *C*) corresponds to a particular slice in a 3D CT volumetric image (e.g. between 1 and 50) . For example, set *A* was created for slice #18.

For all three slice-depths that we evaluated the proposed method performed substantially better than SIFT and SURF. Please refer to Figure 5 for examples of correspondences detected by ML-based approach, SIFT and SURF algorithms. For slices *A*, *B*, and *C*, our method was approximately 2.6, 2.3, and 2.5 times better than the second best technique, namely SIFT.

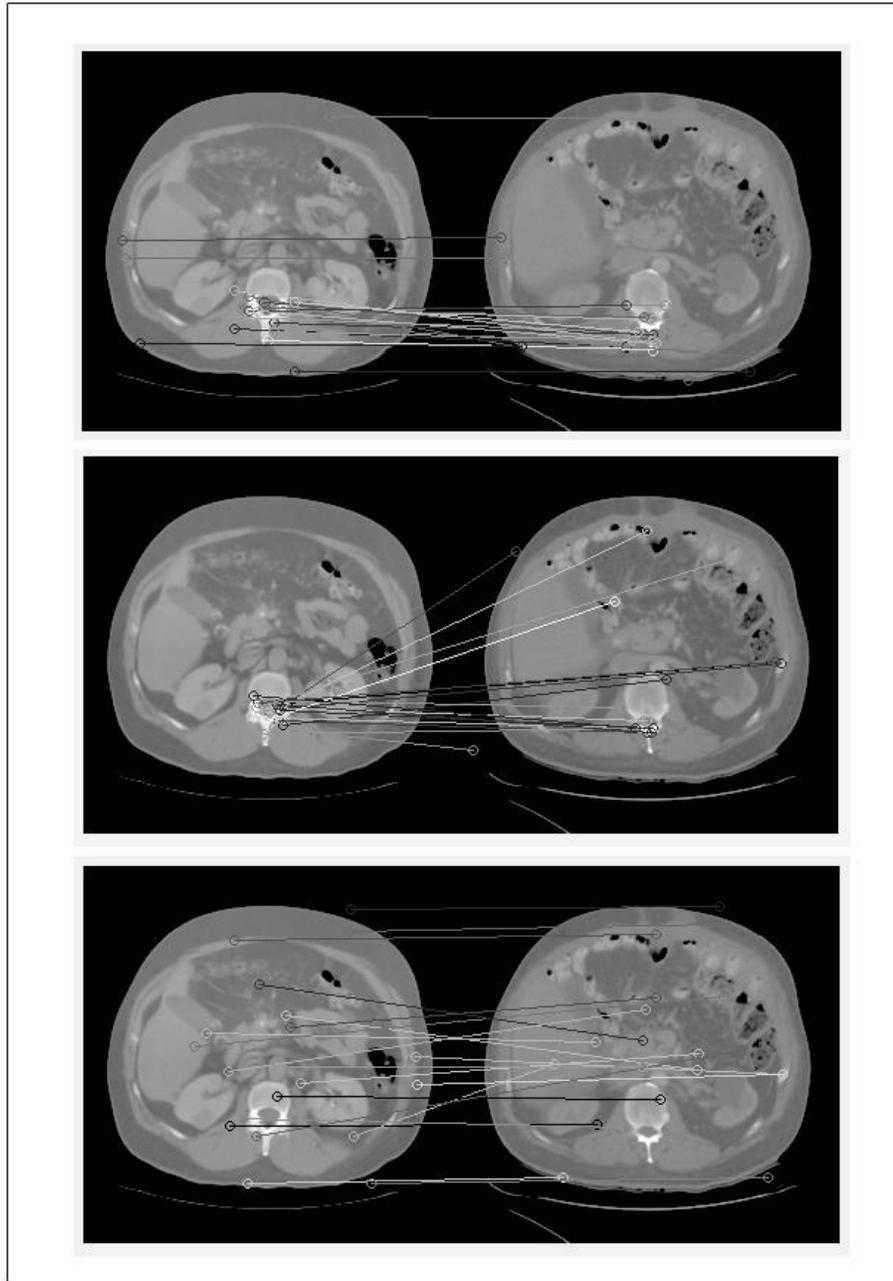


Figure 5. Illustrative examples of point-correspondences detected by ML-based IPD (top), SIFT(middle) and SURF (bottom)

3.3. Qualitative comparison of ML-based IPD with SIFT and SURF in 2D

While performing correspondence detection experiments we observed differences in the types of features selected by these three IPD methods. Figure 5 illustrates examples of the correspondence detection results by SIFT, SURF and our method. For our ML-based IPD, most of the highest ranked IPs are located around the spinal vertebrae and the outer body regions between the ribs and skin. There are just a few IPs with high CV-scores selected on the internal organs. IPs detected on rigid and/or highly structured regions seems to give good point-correspondences.

Top 20 filtered SIFT IPs were often similar (e.g. located in the same body regions) to those chosen by our ML-based IPD (e.g. middle-left image in Figure 5). SIFT, on average, produced roughly around 200 points per image, we only compared the correspondence qualities of the filtered 20 points, and it was observed that their point-correspondences were substantially worse than those identified by our ML-based IPD.

Top 20 IPs by SURF seemed to be different from those found by ML-IPD and SIFT. IPs by SURF were often located on the internal organs, spreading across the abdomen region. Due to the highly variable nature of the abdominal CT scans, these IPs resulted in more failures of point-correspondences than IPs identified by ML-IPD and SIFT.

In summary, our observations indicate i) that ML-IPD favor more rigid and structured regions while SURF favors internal organ regions more than the others, ii) that ML-IPD is better in detecting point-correspondences than SIFT for similarly identified points, and iii) the types of IPs chosen by SIFT and SURF tend to be different even though their point-correspondence detection performance is quantitatively similar.

3.4. Image registration with proposed ML-based IPD in 3D

Image registration performance with the proposed feature detection method was evaluated by using a 3D affine transformation model. To validate the registration accuracy, we placed a set of 12 anatomical landmarks on each of our testing images. The registration errors were computed using root-mean-square (RMS) distance measure between the ground-truth and transformed landmarks. We tested two window sizes: 15 and 19 in voxels with two types of input features for the NN classifier: one was using raw intensity values for input vectors and another using 3D SIFT descriptor⁶ consisting of 121 values. The raw intensity feature assigns a vector of voxel intensities within the window to NN's input layer, while the SIFT descriptor is computed using the method presented by Scovanner et al.⁶ For comparison, the standard image-based 3D affine registration tool in Insight Toolkit² is used with linear interpolation and gradient descent for optimization. Table 2 summarizes the experimental results. Window size of 19 voxels with SIFT descriptor as input feature performed the best, resulting in 43% reduction in error compared to data before registration. This parameter setup also showed an improvement of 35% over the baseline ITK affine registration error. The absolute errors were substantially higher for all cases due to the data set containing surgical resections and significant differences in organ geometry across the patients.

Type	RMS (mm)	Std-Dev
Before Registration	24.4	37.9
Size-15 Raw	21.4	50.1
Size-19 Raw	18.6	26.8
Size-19 with SIFT descriptor	12.2	14.5
ITK Affine	18.9	26.6

Table 2. Averaged registration errors (RMS and with standard deviation) by ML-based method with various window sizes, different input features and by a base ITK affine registration algorithm.

4. CONCLUSION

This paper presents a novel feature-based registration framework with new feature selection and IPD method using NN classifier and CV-based detectability score. The key contribution of this work was to show that a ML method can be used to define (feature selection) and detect (point-correspondence) interest points which can help improve the registration accuracy for clinical medical images. Our experimental results show that the proposed ML-based IPD performs better than other standards: SIFT and SURF for correspondence detection and better than ITK built-in affine registration algorithm for image registration.

For our future work, we plan to conduct experiments with the proposed method to include more valid points (R), more training and testing samples (M) and compare it with other competing IPD methods. The training phase of our method is time consuming, taking more than a week in 3D experiments. Developing efficient training algorithm is important for future work. Considering the abnormal anatomy in abdominal registration brings not only clinical relevance toward cancer staging and follow-up applications, but also additional technical challenges to the already difficult abdominal registration. Evaluating how the ML-based method performs with data from diseased population in more details is another critical future work. We also plan to extend our registration framework to work with non-rigid registration cases and test other ML algorithms such as Support Vector Machine, Random Forest and Adaboost as a base IPD.

REFERENCES

- 1 J.B. Maintz and M.A. Viergever. "A survey of medical image registration" *Med Image Anal*, 2(1): 1-36, 1998
2. Luis. Ibanez, Will. Schroeder, Lydia. Ng, Josh. Cates, and Insight Software Consortium. The ITK Software Guide. ITK.org, November 2005.
3. D. Lowe. "Object Recognition from local scale-invariant features", In Proc. ICCV, pp.1150-1157, 1999
4. H. Bay, et al. "SURF: Speeded Up Robust Features", *Computer Vision and Image Understanding*, 110(3): 346-359, 2008
5. C. Harris and M. Stephens. "A Combined corner and edge detector", In Proc. The 4th Alvey Vision Conference pp.147-151, 1980
6. P. Scovanner, S. Ali, and M. Shah. A 3-dimensional SIFT descriptor and its application to action recognition. *ACM Multimedia*, September, 2007
7. SIFT for Matlab. <http://www.vlfeat.org/vedaldi/code/sift.html>, last accessed February 2011
8. D. Rueckert et al. "Non-rigid registration using free-form deformations: Application to breast MR images" *IEEE Trans Medical Imaging*, 18(8):712-721, 1999
9. M. Zuliani. RANSAC toolbox. <http://vision.ece.ucsb.edu/~zuliani/Code/Code.html>, last accessed July 2011
10. OpenSURF <http://www.chrisevansdev.com/computer-vision-opensurf.html> last accessed February 2011
11. C.M. Bishop. "Neural Networks for Pattern Recognition" Oxford Univ Press, 1995
12. S. Allaire et al. "Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis" In Proc. CVPR Workshop, 2008
13. M. Moradi et al. "Deformable registration using scale space keypoints" In Proc. SPIE Medical Imaging, 2006
14. A. Wang et al. "Research on a novel non-rigid registration for medical image based on SURF and APSO" In Proc. Int Congress Image and Signal Processing, pp. 2628-2633, 2010
15. G. Khaissidi et al. "A fast medical image registration using feature points" *ICGST-GVIP Journal*, 9(3): 19-24, 2009
16. T. Lindeberg. "Edge detection and ridge detection with automatic scale selection", *International Journal of Computer Vision*, 30(2): 117-154, 1998
17. Hagan, M.T., H.B. Demuth, and M.H. Beale, *Neural Network Design*, Boston, MA: PWS Publishing, 1996.