Learning-based Neuroimage Registration

Leonid Teverovskiy and Yanxi Liu[1]

October 2004
CMU-CALD-04-108, CMU-RI-TR-04-59

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Neuroimage registration has been a crucial area of research in medical image analysis for many years. Aligning brain images of different subjects in such a way that same anatomical structures correspond spatially is required in many different applications, including neuroimage classification, computer aided diagnosis, statistical quantification of human brains and neuroimage segmentation. We combine statistical learning, computer vision and medical image analysis to propose a multiresolution framework for learning-based neuroimage registration. Our approach has four distinct characteristics not present in other registration methods. First, instead of subjectively choosing which features to use for registration, we employ feature selection at different image scales to learn an appropriate subset of features for registering a specific pair of neuroimages. Second, we use interesting-voxel selection to identify image voxels that have the most distinct image feature vectors. These voxels are then used to estimate the deformation field for registration. Third, we iteratively improve our choice of features and interesting voxels during registration process. Fourth, we create and take advantage of a statistical model containing information on image feature distributions in each anatomical location.

[1]**This work is supported in part by NIH grants AG05133 and DA015900-01**

Keywords: feature vectors, feature selection, interesting voxels, deformable registration, image pyramid, thin plate splines, RANSAC.

Contents

1 Introduction

Neuroimage registration is an essential problem in medical image analysis. Transforming neuroimages so that their corresponding anatomical structures become aligned is essential for statistical quantification of human brain, computer aided diagnosis, neuroimage segmentation and study of normal aging.
Neuroimage registration, particularly cross-subject registration, presents a number of challenges.

1. Corresponding anatomical brain structures of different subjects may differ in shape and topology. While some structures have relatively simple form that does not change much from subject to subject, others, like sulci, have a complex shape that varies significantly between people. These variations generally increase with aging, and can be further intensified by the presence of neurological disorders.

2. In the cases where registration involves neuroimages of pathological brains, some lesions, like tumors, can be present in one brain but not in the other. In such cases the problem of aligning corresponding anatomical structures becomes ill-posed since not every structure in one brain has a corresponding one in the other brain.

In this work we propose a new learning-based method for neuroimage registration. Our method addresses the challenges by learning which features to use for each pair of images and which voxels of the images to use in order to drive the registration process.
This paper is organized as follows. In section 2, we outline strength and weaknesses of the existing methods for neuroimage registration. Then we present our approach in section 3. Experimental results follow in section 4; section 5 contains a brief discussion of the future work, and we conclude the paper in section 6.

2 Existing Approaches

Existing methods for neuroimage registration can be divided into three groups:

1. Deformation model driven registration [18, 20, 14]. Methods in this group maximize some similarity metric between two images as a function of transformation parameters. Most popular similarity metrics are mean sum of squared differences (MSSD) between the reference and registered input image and mutual information (MI). Advantages of these methods are that

   (a) they are fully automatic;

   (b) they do not require computation of features.

   However, these methods also have a number of disadvantages:

   (a) they are susceptible to converging to suboptimal solutions because currently used similarity measures, like MI and MSSD are not convex functions of transformation parameters [20, 21];

   (b) they require a parametrized deformation model to be chosen beforehand;

(c) their performance depends on the number of parameters in a deformation model [14]. The fewer parameters the deformation model has, the easier it is for a method to find a near optimal solution in the space of transformations defined by the model. However, spaces induced by deformation models with few parameters are often not rich enough to contain an adequate transformation for cross-subject registration. On the other hand, more descriptive deformation models with many degrees of freedom are computationally expensive and make it harder to escape local extrema because of the need to estimate a large number of parameters. [21, 20];

(d) their performance depends on the initial orientation of the images [20].

2. Landmark-based registration. Methods in this group compute registering transformation based on the user-specified correspondences between certain voxels (landmarks) in the reference image and the input image. Advantages of the methods in this group are that

   (a) deformation does not have to be parametrized in advance. The registering transformation is computed based on the given correspondences;

   (b) user-specified correspondences are accurate.

   Shortcomings of these methods include human intervention and time-consuming landmark specification.

3. Feature-vector based registration methods [27] . Methods in this group compute a feature vector for each voxel. Correspondences between voxels in the reference image and voxels in the input image axe estimated based on the similarity of their feature vectors. Attractive characteristics of such methods are that

   (a) they are automatic;

   (b) they are less prone to converging to suboptimal transformations;

   (c) they do not require parametrized deformation model, but estimate the registering transformation based on the computed correspondences.

   However, such methods

   (a) use pre-selected set of features;

   (b) involve many hand-tuned parameters.

Our approach builds on feature-based registration methods. As does the method presented in [23], we utilize various features to construct an attribute vector describing a voxel in a neuroimage. Also, as in [27], we use gaussians to model distribution of features belonging to a given voxel. However, our method has a number of distinguishing traits:

1. We estimate parameters of gaussians based on affinely transformed copies of the reference image.

2. We utilize decision theoretic framework to find interesting voxels in the reference image. These interesting voxels are far from other voxels in the features space, i.e. they are different from the others. The same framework is then used to find matches for the interesting voxels among the voxels in the input image.

Table 1: Existing approaches

	Deformation model driven	Landmark based	Feature-vector based
Require human intervention	no	yes	no
Increasing degrees of freedom of the deformation model makes the method more prone to converge to local extrema	yes	no	no
Depend on the initial orientation of the reference and input images.	yes	no	no
Select driving voxels	no	yes	yes
Learn features	no	no	no

3. We do not select features manually in advance but learn which feature subset to use for every pair of neuroimages automatically. This gives our method the power to adapt to particularities of the specific pair of neuroimages.

3 The algorithm

3.1 Introduction

In this section we describe our algorithm for neuroimage registration. Characteristics of our algorithm are:

1. automatic;
2. multiscale;
3. adaptive: for every pair of images the algorithm learns which features and which voxels to use for registration.
4. independent of the initial orientation of the images;

As a preprocessing step for our algorithm we must estimate a statistical model that describes how feature vectors for every anatomical location are distributed. The model is estimated based on a reference image. The model does not have to be reestimated every time the algorithm is run as long as the registration is done to the image of the same subject. Since we can register any two images to each other by registering each of them to the third image, in most cases we can avoid relearning the model.
The algorithm contains the following major components (Figure 3):

1. Feature selection: this mechanism selects a subset of features to use for registration. It uses a feature pool - set of all features available to the algorithm.
2. Correspondence matching: this mechanism uses the model to find correspondences between a voxel in the reference image and a voxel in the input image.

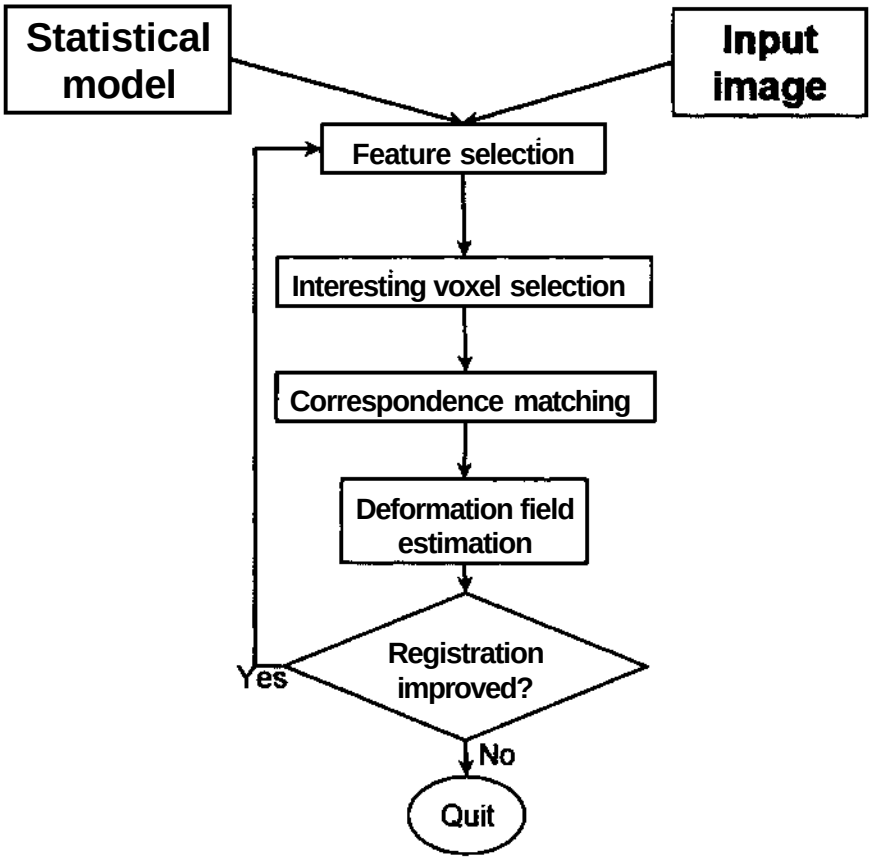

Figure 1: Block diagram of the proposed algorithm. Inputs to the algorithm are statistical model computed based on the reference image and an input image to be registered to the reference image. The algorithm starts by choosing a feature subset at random. Then we find interesting voxels of the reference image, i.e. voxels that can be matched correctly and with high coitfidence in this feature subspace to the corresponding voxels of the model. Under the same feature space we find corresponding voxels in the input image. Transformation between input and reference images is computed based on these correspondences and registration error is calculated. Then a new feature is added to the feature subset and the registration is repeated until the addition of any feature does not reduce registration error. Once the best feature subset is determined, we use it to find a potentially different set of interesting voxels, and repeat the above procedure again. The algorithm terminates if registration error is not decreasing any longer.

Table 2: Affine transformation parameters. A transformation is formed by composing rotation, skewing and scaling in the direction of X-axis, 18 x 11 x 9 = 1782 affine transformation are constructed and applied to the reference image

	Minimum	**Maximum**	Step
Angle, °	0	340	20
Skew	0	0.5	0.05
Scale	0.8	1.2	0.05

3. Interesting voxel selection : it determines which voxels in the reference image are used as landmarks.

4. Deformation field estimation: this component fits a thin plate spline (TPS) transform [3] to the set of candidate correspondences.

5. Registration evaluation: this component computes the similarity measure between the reference image and the registered input image. We used mutual information and mean sum of square intensity differences as similarity measures.

3.2 Statistical model for image features

The statistical model for image features consists of two components. First component is a reference image, which prescribes what image coordinates each anatomical location should have in the registered image. The second component is a set of image features distributions. It contains information about how feature vectors are distributed for voxels in the reference image.

For a reference image we simply need to select an MR neuroimage of a healthy individual. Then image coordinates of, say, left tip of the corpus callosum in the input image after the registration should be the same as those of left tip of the corpus callosum in the reference image.

As for the second component, we have to estimate probability density function *f(X\vi)* which tells us the likelihood of observing feature vector *X* at the voxel at anatomical location *V{*. Under the assumption that the components *Xj* of the feature vector *X* are independent we can factorize $f(X|v_i)$:

$$f(X|v_i) = \prod_{j=1}^{m} f(x_j|v_i), \tag{1}$$

where *m* is the dimensionality of *X*. Now we face an easier task of estimating *f(xj\v{)* for every individual component of *X*. An ideal training set for this task would be a set of neuroimages of different subjects, where we know voxel by voxel correspondences between our selected reference image and every neuroimage in the set. However, such correspondences are sometimes semantically ambiguous [25] and are time consuming to produce by hand. Instead, we create our training set by applying 1782 different affine transformations to our reference image (see Table 2). We have an advantage of knowing exact voxel to voxel correspondences between the reference image and each of the transformed images. After we calculate feature vectors for every voxel in every image, we obtain a sample of 1782 feature vectors for each anatomical location *V{* (see Figure 2).

These samples enable us to perform estimation of the distribution *f(xj\vi)* of feature vector components *Xj* at each anatomical location *V{*. In our experiments we chose to use parametric

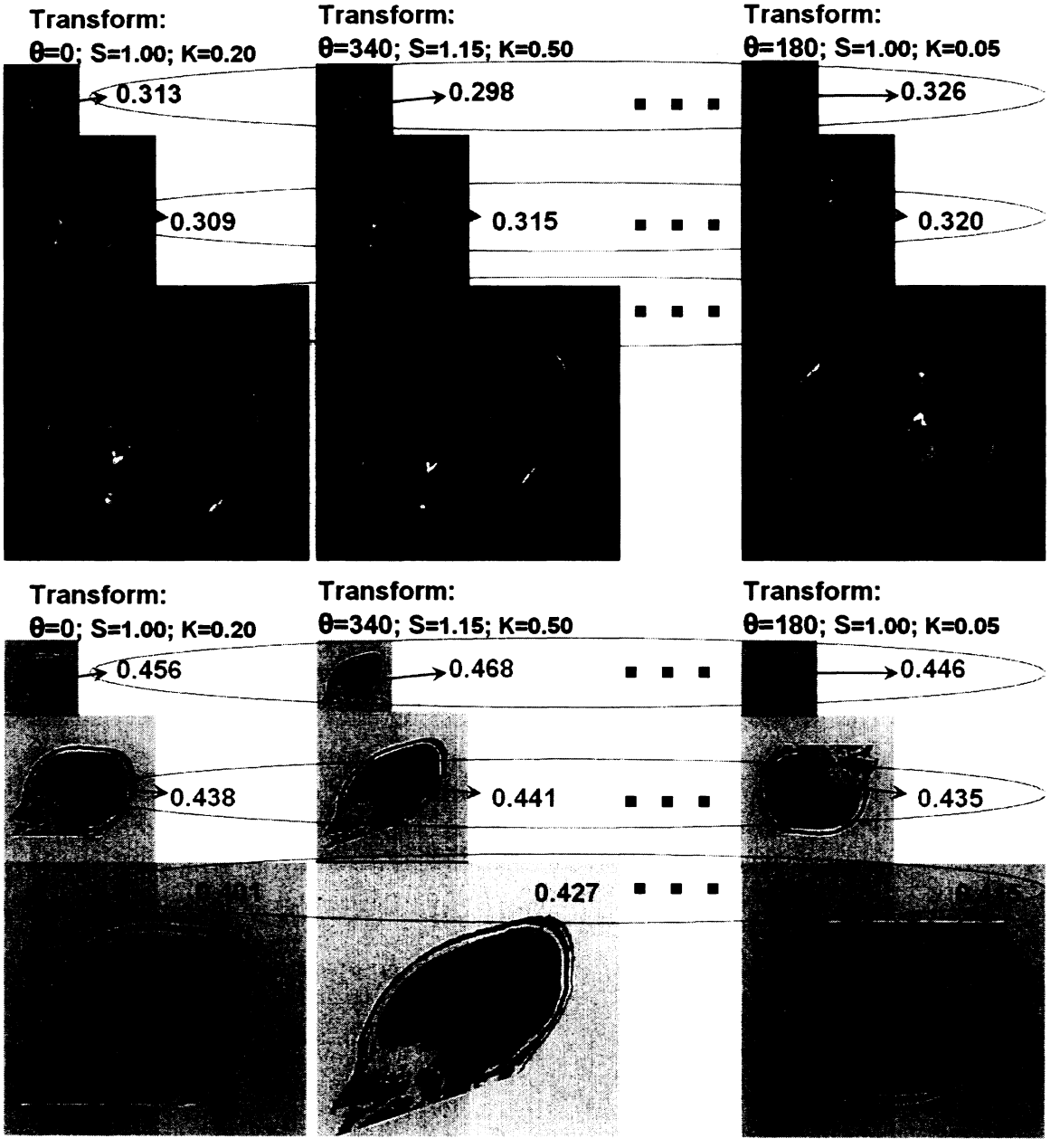

Figure 2: For each combination of angle θ, scaling S and skewing K shown in the Table 2 we compute an affine transform and apply it to the reference image. Features are computed for every transformed copy at three image scales. Values of *intensity_2_mean* (top) and *gabor_0_3* (bottom) features for a particular voxel are shown

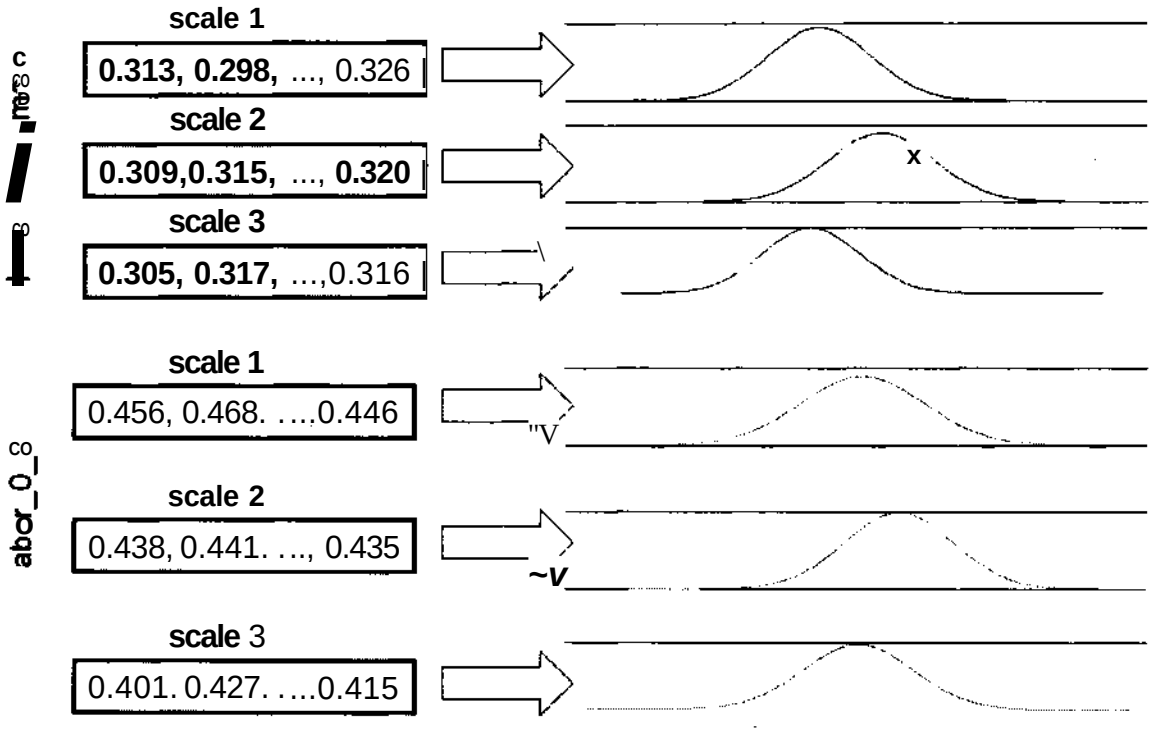

Figure 3: For every voxel, feature and scale we fit a Gaussian to the values obtained from affinely transformed copies of the reference image (see Figure 2). Gaussians of *intensityJ2jmean* (top) and *gabor*_0_3 (bottom) features for a particular voxel are shown

estimation with gaussian family of distributions [27]. Non-parametric density estimation is another alternative, but it is more expensive computationally. As Figure 4 illustrates, Gaussian family is a reasonable parametric model to use for estimating feature distributions at specific image voxels. By using formula (1) we find distribution $f(X|v_i)$ of feature vectors X for every anatomical location v_i in the reference image (see Figure 3).

3.3 Feature selection

In our approach we extract a feature vector for each voxel. To a voxel in an image we can apply a number of neighborhood operators, such as Gabor filters, Laplacian operators, Harris detector, etc [9, 4, 10]. Each such operator corresponds to a dimension in a feature space, and the responses of these operators define the coordinates of the voxel in that feature space. To make our features rotationally invariant we apply neighborhood operators at several orientation and select maximum of the responses for each operator.
Feature pool contains features available to the algorithm. Only subset of these features is automatically selected and used for the actual registration. The feature pool used in our experiments is summarized in the Table 3. Note that our feature pool included only generic commonly used features. The distinct property of our learning based system is that the feature pool can contain any classical or novel features deemed appropriate by a researcher.

We utilize wrapper approach [12] to feature selection and sequential feature selection strategy [19, 16, 15]. At each step of the sequential feature selection we perform registration using currently selected feature subsets (one to select interesting voxels, the other one - to estimate correspondences). We evaluate quality of registration using sum of squared differences between reference image and registered input image. If quality of the registration improves compared to the previous step of the sequential selection, we record current subsets and continue the feature selection process. Otherwise we go back to the subsets selected during previous step and try to add (remove) different feature. If the addition(removal) of the remaining features does not improve registration, we stop the iteration.

3.4 Correspondence matching

Suppose we know which feature subspace is a good one to use for the registration of MR neuroimages. Then for each voxel W_k in the input image we can compute a feature vector X_k. We need to estimate a probability mass function $m(v_i|X_k)$ which returns a probability that feature vector X_k belongs to a voxel v_i in the *reference* image.
Using Bayes rule we can write $m(v_i|X_k)$ as

$$m(v_i|X_k) = \frac{f(X_k|v_i)p(v_i)}{\sum_{j=1}^{n} f(X_k|v_j)p(v_j)}, \tag{2}$$

where $f(X_k|v_i)$ is the probability density function (see equation 1) which tells us the likelihood of observing feature vector X_k at the voxel at the anatomical location v_i; $p(v_i)$ is prior distribution for voxels in the reference image; n is number of voxels in the reference image.
Prior $p(v_i)$ lets us incorporate prior knowledge into the distribution estimation. In the multiscale registration, for example, if we have estimated correspondences on a coarser scale, then a priori on a finer scale we would like to decrease probabilities of correspondences that are inconsistent with

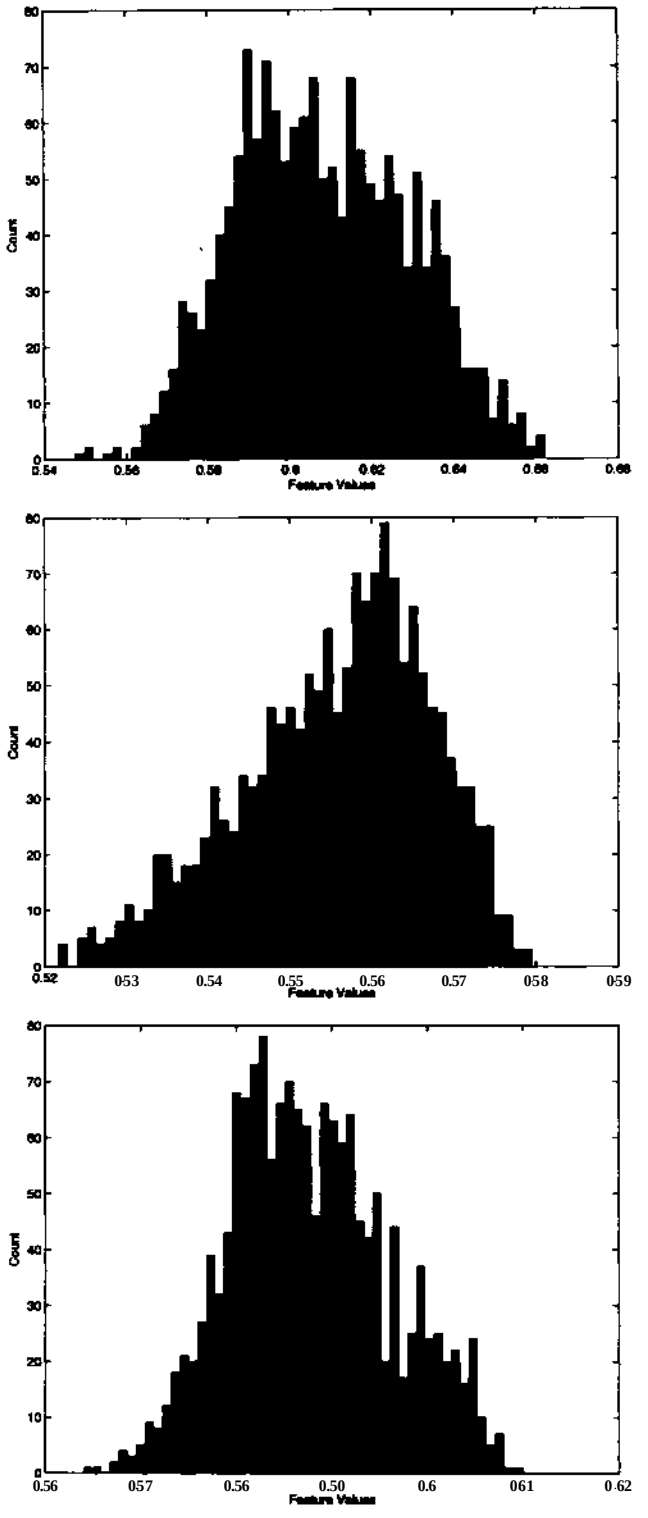

Figure 4: Histograms of values of the *INTENSITY£JAEAN* feature for the same voxel at 3 different scales indicate Gaussian-like distributions

Table 3: Image Feature Pool

Feature	Description
First derivative (Dl)	Maximum absolute value of the response of the first derivative operator applied at 8 equally spaced orientations from -90° to 90°
Second derivative (D2)	Maximum absolute value of the response of the second derivative operator applied at 12 equally spaced orientations from -90° to 90°
Third derivative (D3)	Maximum absolute value of the response of the third derivative operator applied at 16 equally spaced orientations from -90° to 90°
Fourth derivative (D4)	Maximum absolute value of the response of the fourth derivative operator applied at 20 equally spaced orientations from -90° to 90°
Fifth derivative (D5)	Maximum absolute value of the response of the fifth derivative operator applied at 24 equally spaced orientations from -90° to 90°
Gabor.0.3 (Gl)	Maximum absolute value of the response of the Gabor filter with scale 0 and spatial frequency 3 applied at 8 equally spaced orientations from -90° to 90°
Gabor.0.5 (G2)	Maximum absolute value of the response of the Gabor filter with scale 0 and spatial frequency 5 applied at 12 equally spaced orientations from -90° to 90°
Gabor_2_7 (G3)	Maximum absolute value of the response of the Gabor filter with scale 2 and spatial frequency 7 applied at 16 equally spaced orientations from -90° to 90°
Gabor_3_7 (G4)	Maximum absolute value of the response of the Gabor filter with scale 3 and spatial frequency 7 applied at 12 equally spaced orientations from -90° to 90°
Gabor.4.9 (G5)	Maximum absolute value of the response of the Gabor filter with scale 4 and spatial frequency 9 applied at 16 equally spaced orientations from -90° to 90°
Laplacian (L)	Response of the Laplacian operator
Harris Detector (H)	Response of the Harris detector
Intensity_1.mean (Ml)	Mean of the voxel intensities inside a ring with inner radius 0 and outer radius 1
Intensity_l_std (SI)	Standard deviation of the voxel intensities inside a ring with inner radius 0 and outer radius 1
Intensity_2_mean (M2)	Mean of the voxel intensities inside a ring with inner radius 1 and outer radius 2
Intensity_2_std (S2)	Standard deviation of the voxel intensities inside a ring with inner radius 1 and outer radius 2
Intensity_4_mean (M3)	Mean of the voxel intensities inside a ring with inner radius 2 and outer radius 4
Intensity_4_std (S3)	Standard deviation of the voxel intensities inside a ring with inner radius 2 and outer radius 4
Intensity_8_mean (M4)	Mean of the voxel intensities inside a ring with inner radius 4 and outer radius 8
Intensity_8jstd (S4)	Standard deviation of the voxel intensities inside a ring with inner radius 4 and outer radius 8
Intensity*A* 6_mear (M5)	Mean of the voxel intensities inside a ring with inner radius 8 and outer radius 16
Intensity_16jstd (S5)	Standard deviation of the voxel intensities inside a ring with inner radius 8 and outer radius 16

the coarser level. If only one scale is used, or when we work at the coarsest scale in a multiscale pyramid, we use uninformative uniform distribution as our prior. In this case, equation (2) becomes

$$m(v_i|X_k) = \frac{f(X_k|v_i)}{\sum_{j=1}^{n} f(X_k|v_j)} \tag{3}$$

3.5 Interesting voxel selection

Selecting interesting voxels in the model improves both the performance of the algorithm and accuracy of the proposed correspondences. Suppose that a voxel in the reference image is similar to many other voxels in the reference image. Then its corresponding voxel in the input image must be similar to many other voxels in the input image. Therefore, our soft matching process will find that the model voxel under consideration corresponds, with high probability, to *many* voxels in the input image. As a consequence, even if we pick the correspondence with highest probability, there is a larger chance for a mistake.

In order to quantitatively evaluate the quality of an estimated correspondence, we define a risk measure as follows

$$R(X_k) = \sum_{i=1}^{N} m(v_i|X_k) D(v_i, v_k), \tag{4}$$

where $D(v_i, v_k)$ is the geometric distance between voxel v_k corresponding to the feature vector X_k, and voxel v_i. The value of $R(X_k)$ is high if there are many voxels that correspond to X_k with high probability, and these voxels are far away from each other. $R(X_k)$ has small values when there are few voxels that correspond to feature vector X_k with high probability, and these voxels are clustered in the neighborhood of v_k.

For each voxel v_i in the reference image we form a feature vector $X_i = \bar{x}_{i1}, \bar{x}_{i2}, ..., \bar{x}_{im}$, where $\bar{x}_{ij} = mean(f(x_j|v_i))$. Then we match X_i to the voxels in the reference image. interesting voxels are those that can be matched correctly and with low risk.

We match interesting voxels with the voxels in the input image using similar process. Now feature vectors X_i are computed for voxels in the input image and matched to the *interesting* voxels in the reference image. However, in this case, we cannot compute the true risk because we do not know the correct correspondence. Therefore we use the correspondence that has highest probability as an estimate of the correct correspondence. Because of that, low risk now only assures that correspondences with high probabilities are clustered together, but it does not guarantee that the cluster is in the correct place anymore. Correspondences with low risk are selected as candidate correspondences based on which we compute the deformation field.

3.6 Deformation field estimation

Since we are estimating the deformation field based on the correspondences between voxels in two images, we have to deal with the problem of incorrect matches. We use a randomized algorithm, RANSAC [8], to fit an affine transform to a set of candidate correspondences. Candidate correspondences that are inconsistent with the estimated affine transform are removed as outliers. The inliers, or driving correspondences, are used to estimate a thin plate spline transform which registers input image to the reference image.

Work of the algorithm is illustrated in the Figure 5. Figure 3.7 demonstrates improvement of reg-

istration results as feature selection process progresses.

3.7 Registration evaluation

We use two different ways to evaluate the quality of a registration:

1. mean sum of square differences between intensities of the reference image and registered input image;
2. mutual information between intensities of the reference image and registered input image.

4 Experiments

4.1 Feature selection strategy

The goal of this experiment is to compare different feature subset selection strategies and to determine how feature selection affects the quality of registration.
We design the following strategies (see Table 4):

1. Select a random set of interesting voxels among the voxels that lie on the edges of the reference image; select a random subset of features to find voxel correspondences.
2. Select a random set of interesting voxels among the voxels that lie on the edges of the reference image; use forward feature selection to find a subset of features to be used for estimating voxel correspondences.
3. Select a random subset of features to find interesting voxels among the voxels that he on the edges of the reference image; select a random subset of features to find voxel correspondences.
4. Select a random subset of features to find interesting voxels among the voxels that he on the edges of the reference image; use forward selection to choose subset of features used for determining voxel correspondences.
5. Select a random subset of features to find interesting voxels among the voxels that he on the edges of the reference image; use forward selection to choose subset of features used for determining voxel correspondences. This time start from the subset used to find interesting voxels without one feature.
6. Select a random subset of features to find interesting voxels among the voxels that he on the edges of the reference image. Then employ forward selection for choosing a subset of features to be used for determining voxel correspondences. Find a new set of interesting voxels using this subset of features and iterate.
7. Select a random subset of features to find interesting voxels among the voxels that he on the edges of the reference image. Then employ forward selection for choosing a subset of features to be used for determining voxel correspondences. This time start from the subset used to find interesting voxels without one feature. Find a new set of interesting voxels using this selected subset of features and iterate.

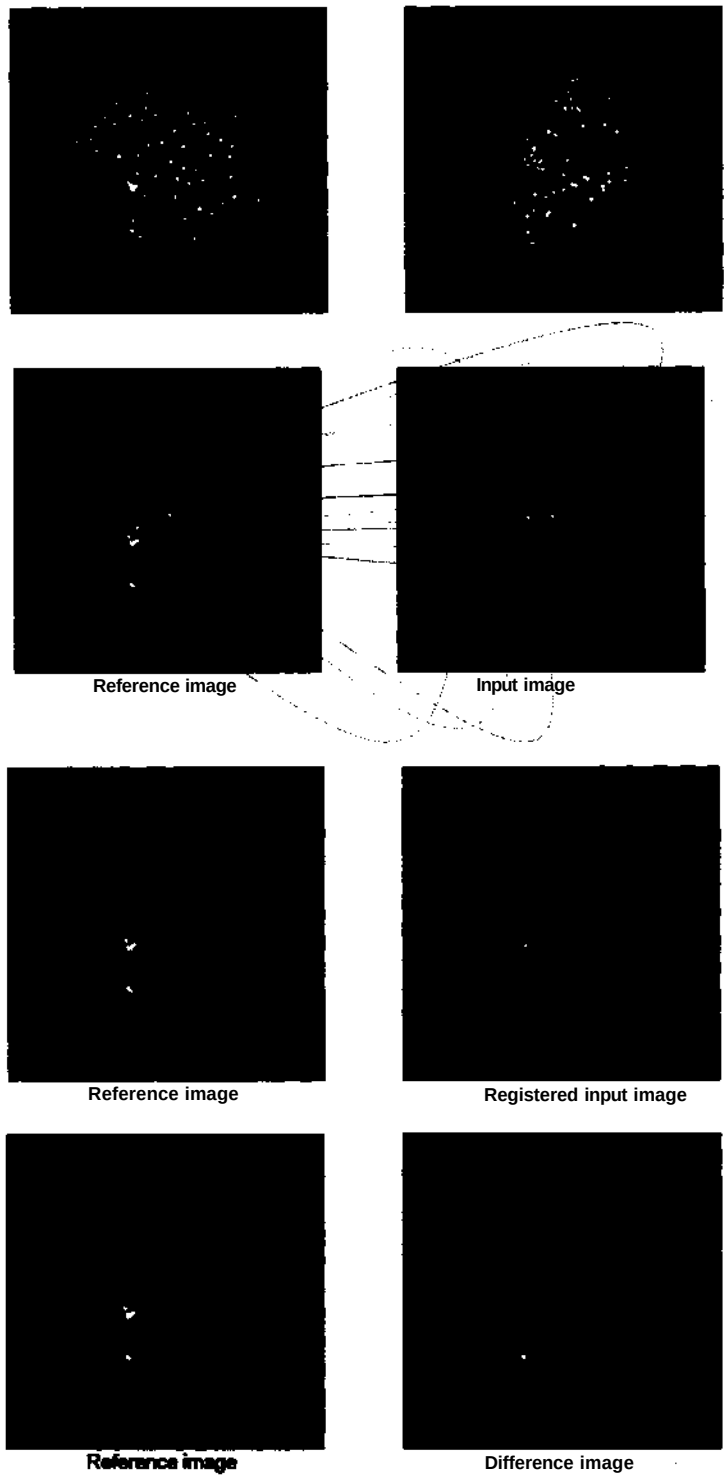

Figure 5: Illustration of the work of the algorithm. Top row: reference image (left) with interesting voxels marked; input image (right) with voxels that match interesting voxels of the reference image. Second row: the same with incorrect correspondences removed automatically. Third row: registration results. Reference image is on the left; registered input image is on the right. Yellow number is registration error measured as mean sum of square differences of intensities between reference and registered input images. Fourth row: reference image (left); difference image(right)

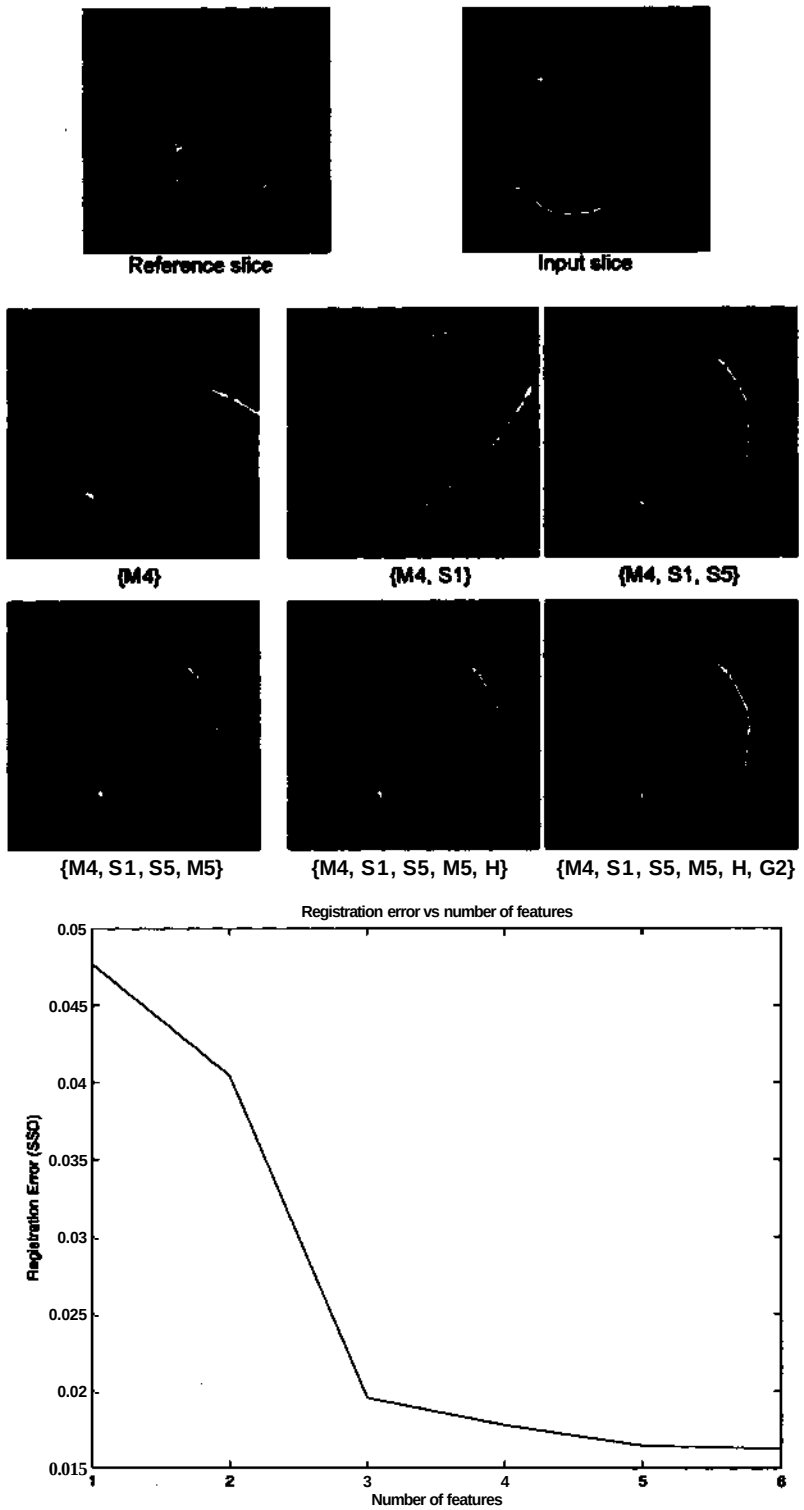

Figure 6: Feature selection at work. Top row: reference and input images. Second and third row: results of the registration as feature selection progresses. Currently selected feature subset is shown below each image (see Table 3 for feature codes). Numbers in yellow show registration error, measures as mean sum of square differences between reference and registered input image. The plot on the bottom illustrates that error reduction rate flattens as the number of selected features increases

Table 4: Feature selection strategies. **A** is a subset of features selected for correspondence estimations; **B** is a subset of features used for interesting voxel selection.

	Interesting voxel selection			Feature selection for correspondence estimation		
	random interesting voxels	using random feature subset	using random feature subset followed, then using **A**	using random set	forward selection	forward selection starting from **B** without one feature
Strategy 1	✓			✓		
Strategy 2	✓				✓	
Strategy 3		✓		✓		
Strategy 4		✓			✓	
Strategy 5		✓				✓
Strategy 6			✓		✓	
Strategy 7			✓			✓

Since voxels that lie on the edges are generally more distinct, in all the strategies above we select interesting voxels only from the edge voxels in order to speed the algorithm up. For each feature selection strategy we run the registration algorithm eight times, each time restarting at a different random point. Each run continues for 20 iterations. Results of the experiments are summarized in Figures 7, 8, 9, 10, 11, 12. The results show that feature selection reduces the registration error in half. We also see that interesting voxel selection increases the number of driving correspondences. Since in this set of experiments we always selected 100 interesting voxels, this means that the number of incorrectly estimated correspondences decreases because of the interesting voxel selection.

4.2 Effect of the number of interesting voxels on the registration quality

The goal of these experiments is to determine the minimal number of interesting voxels used with little sacrifice to the registration quality.
The following feature selection strategies were used:

1. Select a random set of interesting voxels among the voxels that lie on the edges; use forward feature selection to find a subset of features to be used for estimating correspondences.

2. Select a random subset of features to find interesting voxels among the voxels that lie on the edges. Starting from this subset without a feature, employ forward selection for choosing subset of features to be used for determining the correspondences. Find a new set of interesting voxels using this selected subset of features and iterate.

Since voxels that lie on the edges are generally more distinct, in the strategies above we select interesting voxels only from the edge voxels in order to speed the algorithm up.

The numbers of interesting voxels and numbers of the top correspondences for each feature selection strategy we have tried are shown in the Table 5.
Top correspondences are the matches with the lowest risk. We have ran the registration algorithm eight times for each feature selection strategy and every pair of interesting points and top

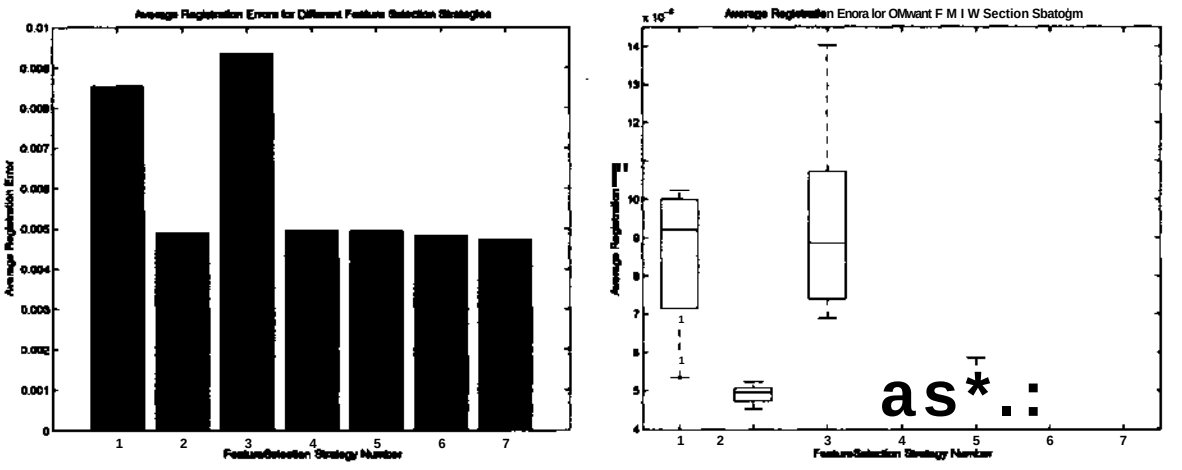

Figure 7: Bar and box plots of the registration error for different feature selection strategies. Strategies 1 and 3, where we randomly select feature subset used for finding correspondences, are significantly worse than the strategies where sequential feature selection is employed. For example, the t-test that mean errors of strategy 1 and 3 are the same as the mean error of strategy 7 yields p-values of 0.00031 and 0.00059 respectively.

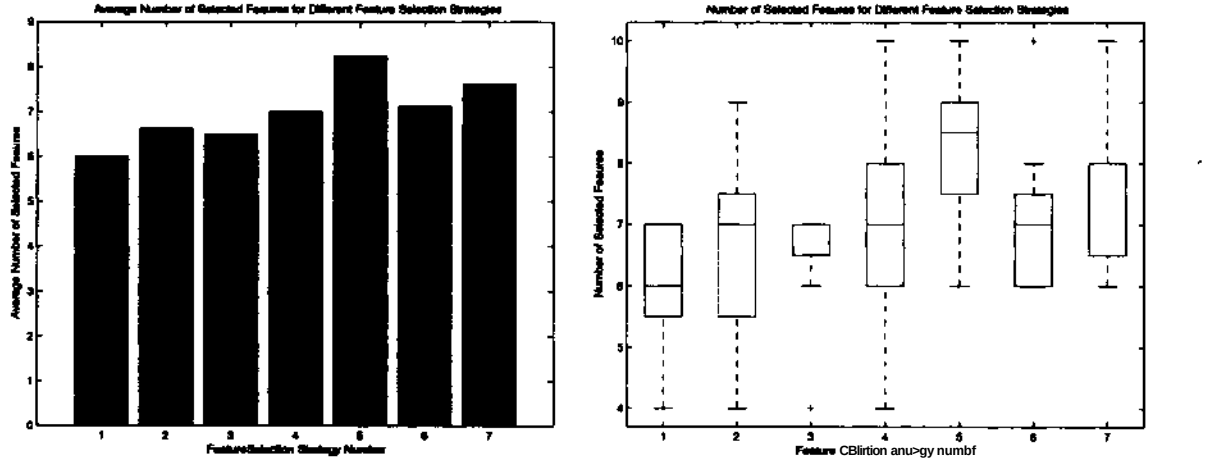

Figure 8: Bar and box plots of the number of selected features for each feature selection strategy. "+" sign indicates outliers

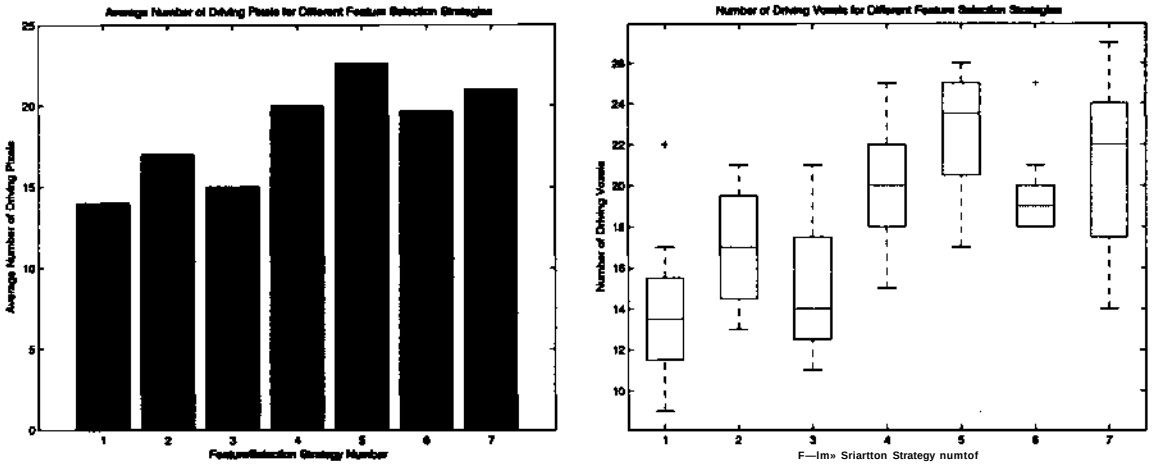

Figure 9: Bar and box plots of the number of driving voxels for each feature selection strategy. First two feature selection strategies, where no interesting voxel selection is performed have fewer driving voxels than the last four. So does feature selection strategy number 3, where feature subset to select interesting voxels and feature subset to find correspondences are chosen at random. "+" sign indicates outliers

Table 5: Number of interesting voxels and top correspondences used in the experiments

Number of IP	80	60	40	20	10	5
Number of Top	60	40	30	15	7	4

correspondences shown in the Table 5, each time restarting at a different random point. Each run continued for 10 iterations. The results are in Figures 13, 14, 15, 16.

For the two feature selection strategies mentioned above, registration error increases as number of interesting voxels decreases. There is an abrupt deterioration of registration quality when the number of interesting voxels drops from 40 to 20 (see Figure 13).
The registration quality does not seem to significantly differ between the two strategies, as shown in Figures 13 and 15. This can be explained by the following factors:

1. Even though we do not select interesting voxels for the second strategy, we still rank correspondences that we estimate according to their risk. We select only the top correspondences and pass them to RANSAC. Therefore, even if there are more incorrect matches for the case with no interesting voxel selection, many of these mistakes are filtered out during the selection of top correspondences.

2. When we choose interesting voxels at random, we still select only voxels that lie on edges. Edge voxels are generally more distinct.

However, we can see that for the same number of interesting points, the number of driving voxels is higher for the strategy with interesting voxel selection (see Figures 14 and 16). This means that interesting voxel selection allows us to find correspondences more accurately.

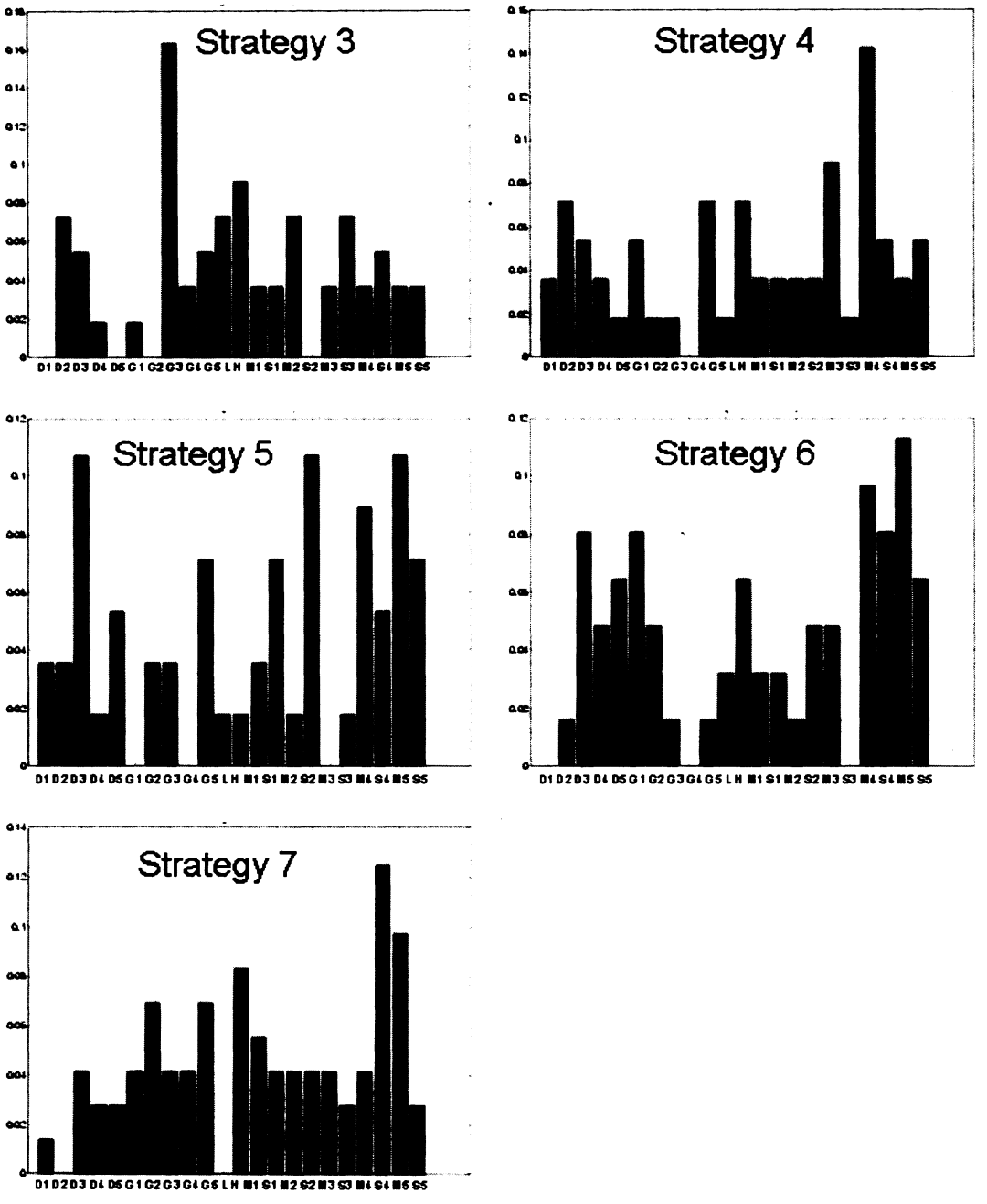

Figure 10: Normalized histograms of the features selected for choosing interesting voxels. Strategies 1 and 2 select interesting voxels at random. Most frequently selected features: strategy 3 - *gabor_3_7*; strategy 4 - *intensity_8_mean*; strategy 5 - *third derivative*, *intensity_2_std*, *intensity_4_mean*; strategy 6 - *intensity_16_mean*; strategy 7 - *intensity_8_std*. Least frequently selected features: strategy 3 - *first and fifth derivatives*, *gabor_0_5*, *intensity_2_std*; strategy 4 - *gabor_3_7*; strategy 5 - *gabor_0_3*, *gabor_2_7*, *intensity_4_mean*; strategy 6 - *first derivative*, *gabor_3_7*, *intensity_4_std*; strategy 7 - *second derivative*, *laplacian* (see Table 3 for feature codes)

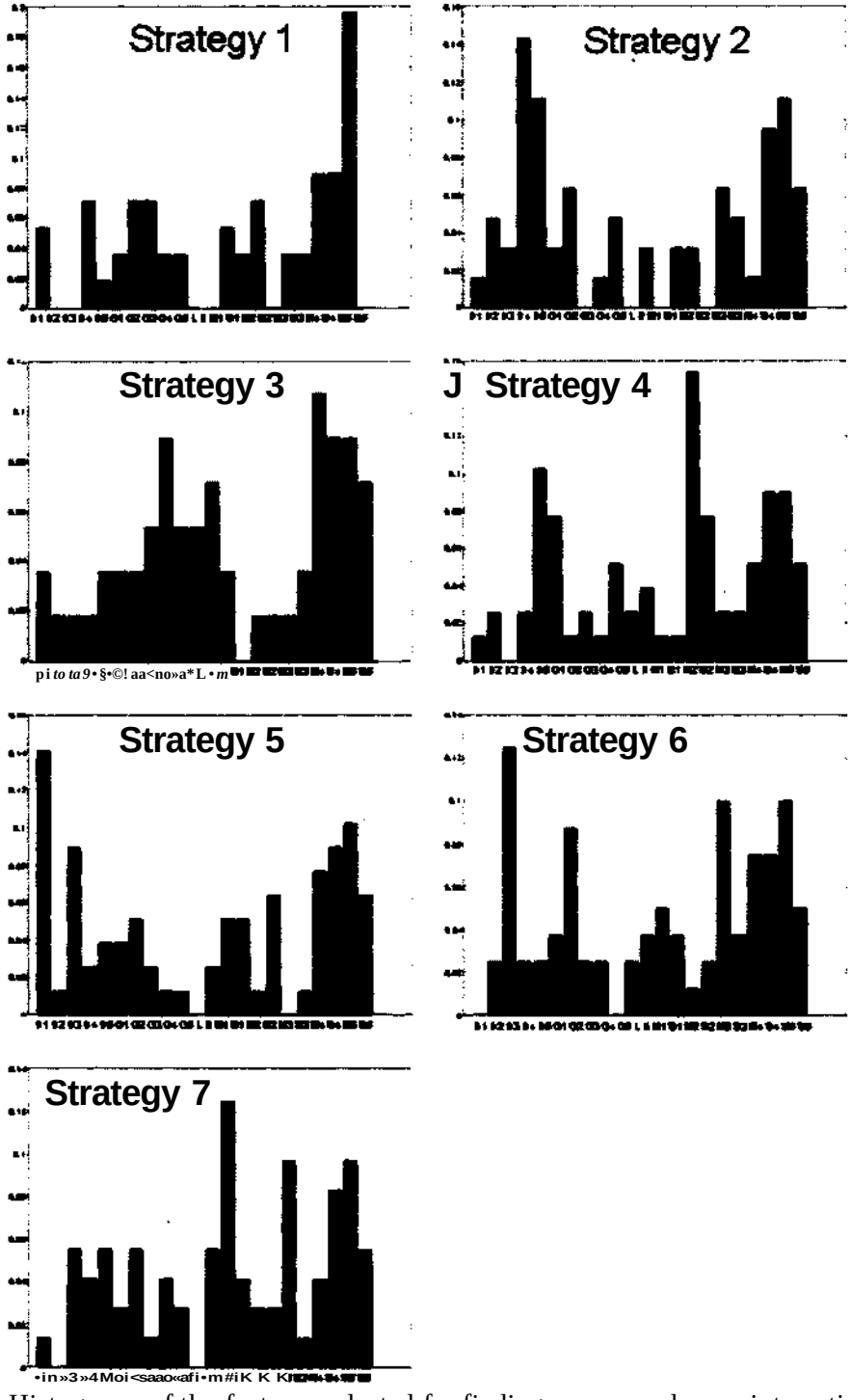

Figure 11: Histograms of the features selected for finding correspondences interesting voxels. Most frequently selected features: strategy 1 - *intensity*_16*jmean;* strategy 2 - *fourth derivative;* strategy 3 - *intensitySjmean;* strategy 4 - *intensity'J2jmean;* strategy 5 - *first derivative;* strategy 6 - *third derivative;* strategy 7 - *intenityAjmean.* Least frequently selected features: strategy 1 - *second and third derivatives, laplacian, Harrisdetector, intensityJLstd;* strategy 2 - pa6or_2_7, *laplacian, intensity*_1*jmean, intensityJlstd;* strategy 3 - *intensityA std;* strategy 4 - *third derivative;* strategy 5 - *laplacian, intensityAjmean;* strategy 6 - *first derivative, gaborAJ);* strategy 7 - *second derivative, laplacian.* (see Table 3 for feature codes)

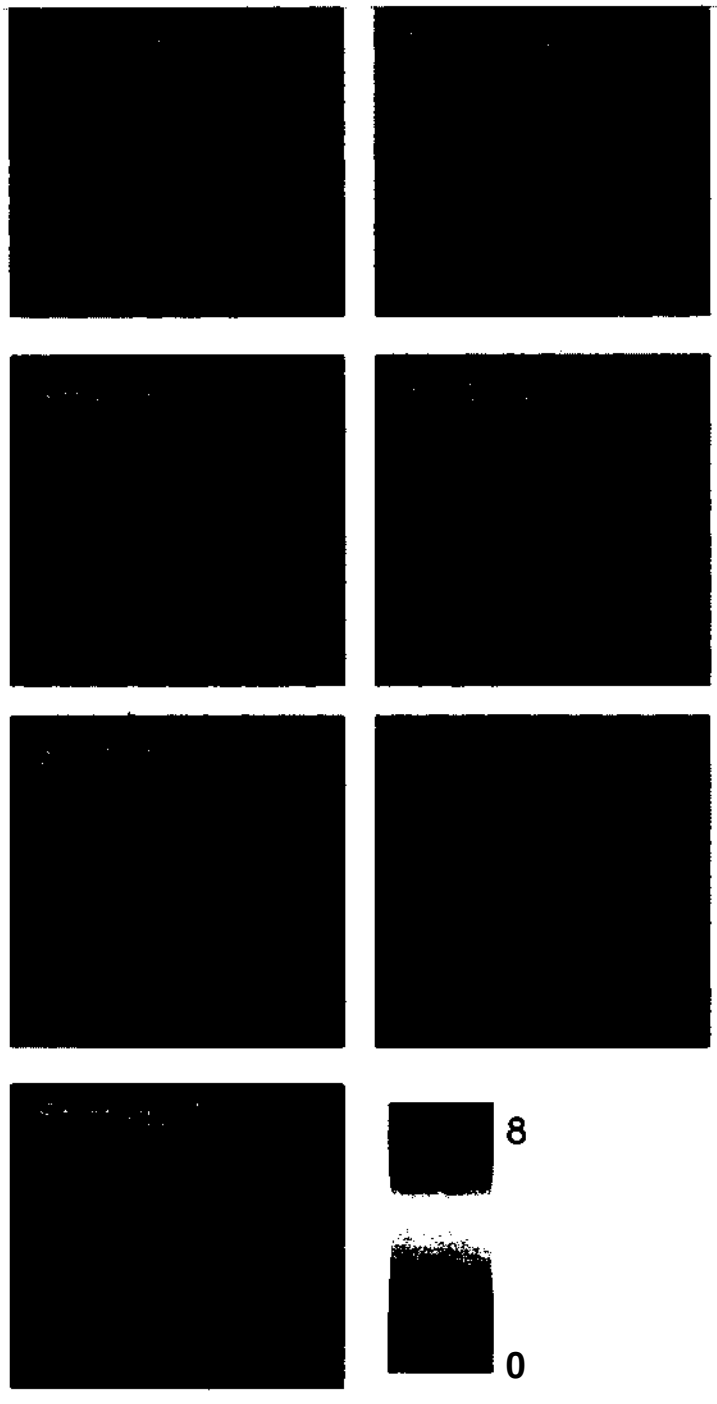

Figure 12: Histograms of the driving voxels. Top row, from left to right: strategy 1, strategy 2; second row, from left to right: strategy 3, strategy 4; third row, from left to right: strategy 5, strategy 6; last row: strategy 7. Blue means the voxel is rarely selected, red means that the voxel is selected often. For each feature selection strategy the algorithm was run 8 times each time starting from a different random point. 100 interesting voxels were selected during each run.

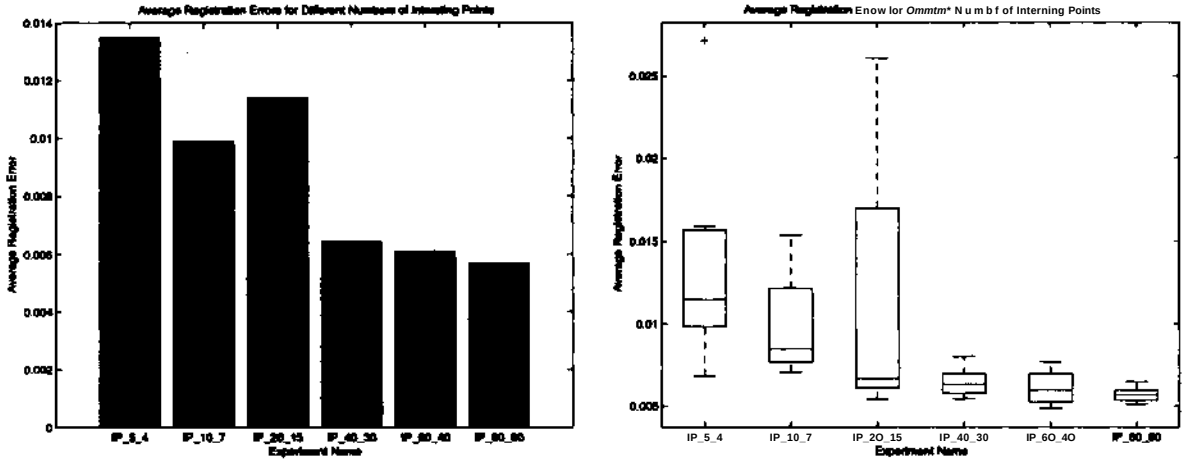

Figure 13: Interesting voxel selection. Bar and box plots of the registration error for each pair of the number of interesting voxels and the number of the top correspondences. "+" sign indicates outliers

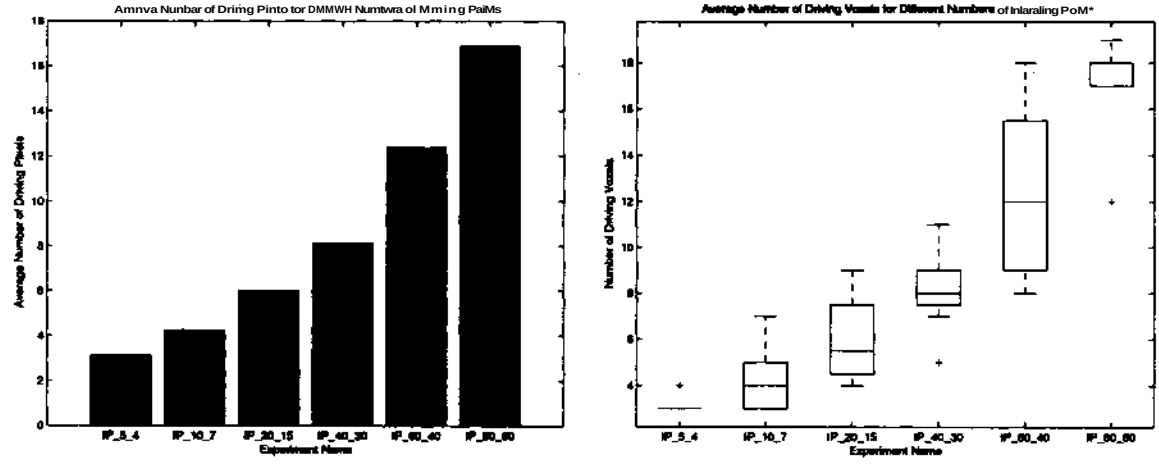

Figure 14: Interesting voxel selection. Number of driving voxels for each pair of the number of interesting voxels and the number of the top correspondences. "+" sign indicates outliers

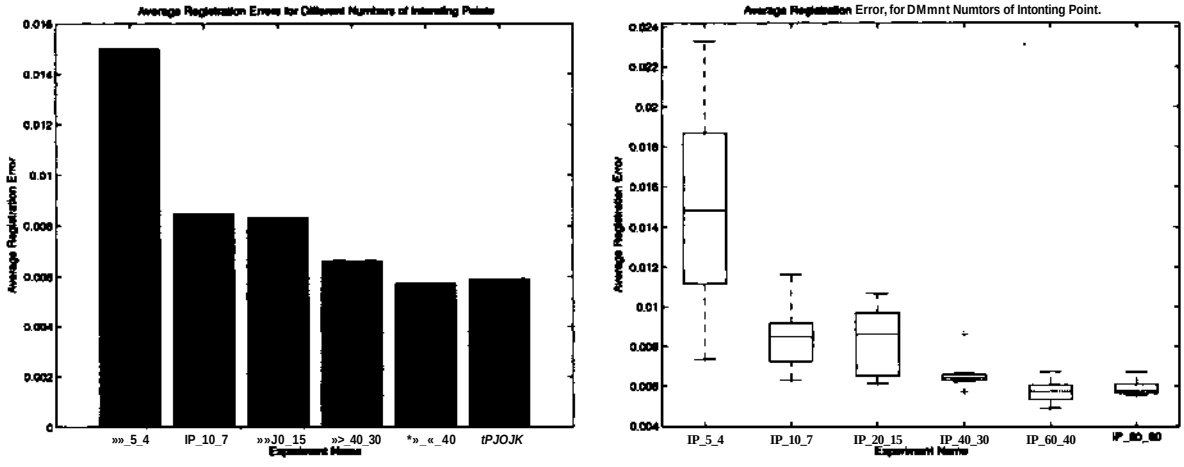

Figure 15: Random interesting voxels. Bar and box plots of the registration error for each pair of the number of interesting voxels and the number of the top correspondences. "+" sign indicates outliers

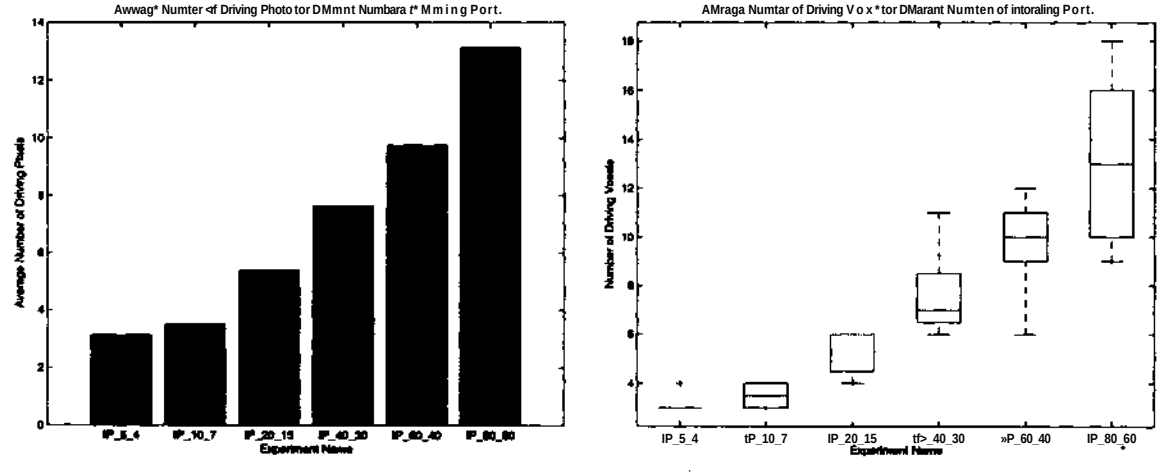

Figure 16: Random interesting voxels. Number of driving voxels for each pair of the number of interesting voxels and the number of the top correspondences

4.3 Using previously learnt interesting voxels and feature subsets

The goal of this experiment is to determine whether we can successfully use interesting voxels and feature subsets learnt during registration of one pair of neuroimages for registration of a different pair of neuroimages.
The following settings were used:

1. Use previously learnt interesting voxels and feature subsets directly.
2. Take previously learnt feature subset for finding correspondences. Use it to select interesting voxels. Perform registration using these interesting voxels and previously learnt subset for finding correspondences.
3. Take previously learnt feature subset for finding correspondences. Use it to select interesting voxels. Perform forward feature selection to learn new subset for finding correspondences.

Each of the three strategies was applied to 40 different neuroimages. Each neuroimage was affinely transformed 10 times by a random affine transformation. 50 interesting voxels were used.

Error plots are presented in the Figures 17, 18, 19.

As the experiments indicate, directly using previously learnt interesting voxels or feature subsets for future registrations gives poor results, as shown in Figures 17 and 18. If we use previously learnt interesting voxels but learn a feature subset for determining correspondences online, the results improve but are still not very good (see Figure 19). These experiments suggest that the key to our algorithm's good performance (see Tables 13 and 20) is the online selection of features *and* iterative improvement of the set of interesting voxels and the feature subsets.

4.4 Testing registration on 40 slices from different subjects

The goal of this experiment is to test how well our algorithm performs when online feature subset selection is used. The following feature selection strategy was used: select a random subset of features to find interesting voxels among the voxels that lie on the edges; starting from the subset used to find interesting voxels without one feature, employ forward selection for choosing subset of features to be used for determining the correspondences; find a new set of interesting voxels using this selected subset of features and iterate.

Our method was applied to 40 different neuroimages. Each neuroimage was affinely transformed 10 times by a random affine transformation. 50 interesting voxels were used. The algorithm was running for 10 iterations. The results are shown in the Figures 20, 21. The algorithm performs competitively well (compare with Figures 13 and 7), except in one case, where input image is significantly brighter than reference image. This happens because we estimate feature distributions for each voxel based only on the reference image, and most of the features in the feature pool are intensity based. Therefore, the algorithm was not able to find correspondences for most of the interesting voxels.

5 Future Work

5.1 Selecting different subspaces of features for each voxel

Selecting single subset of features for all voxels, while relatively straightforward, has a serious limitation. It does not recognize that voxels in a neuroimage can be unique each in its own way. To

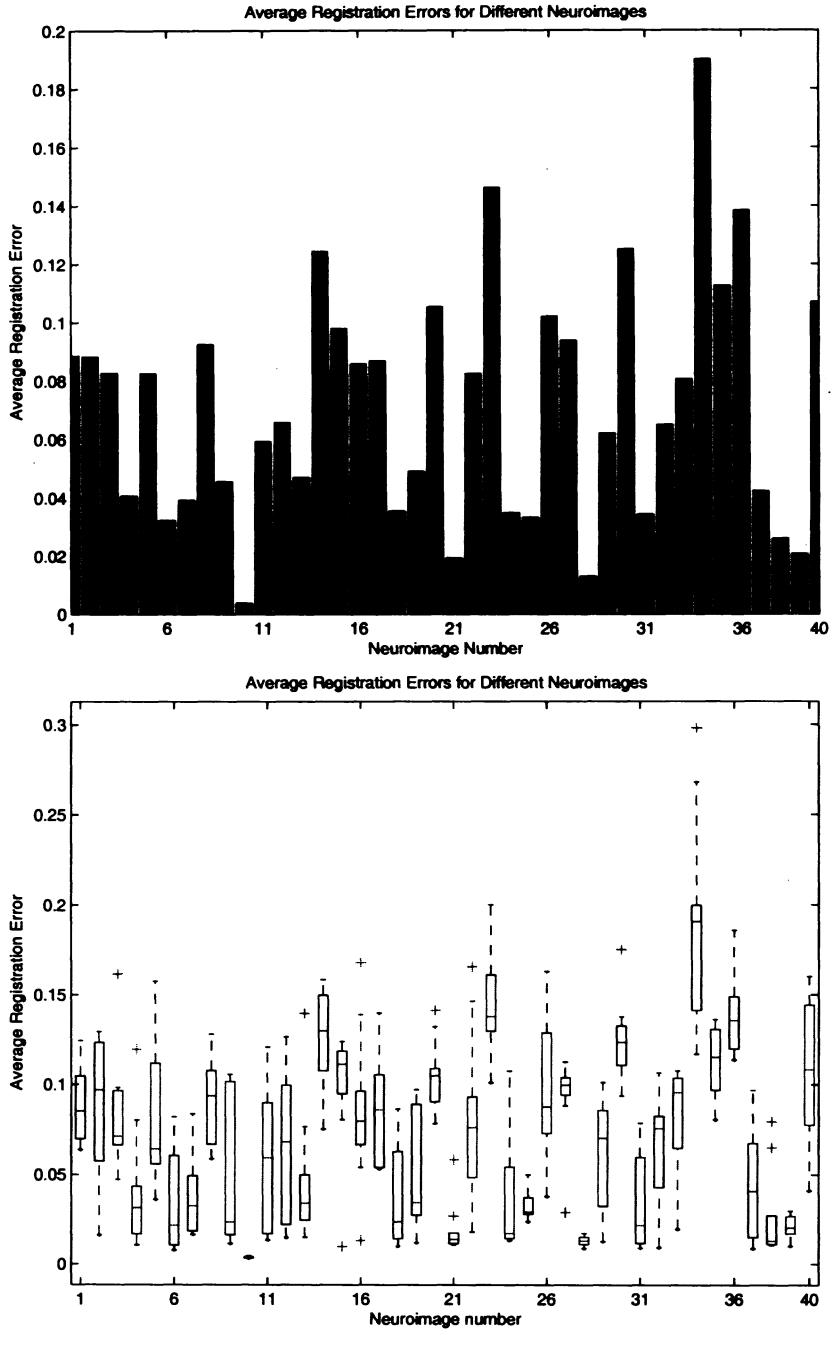

Figure 17: Registration errors when previously learnt interesting voxels and feature subsets are used directly. "+" sign indicates outliers

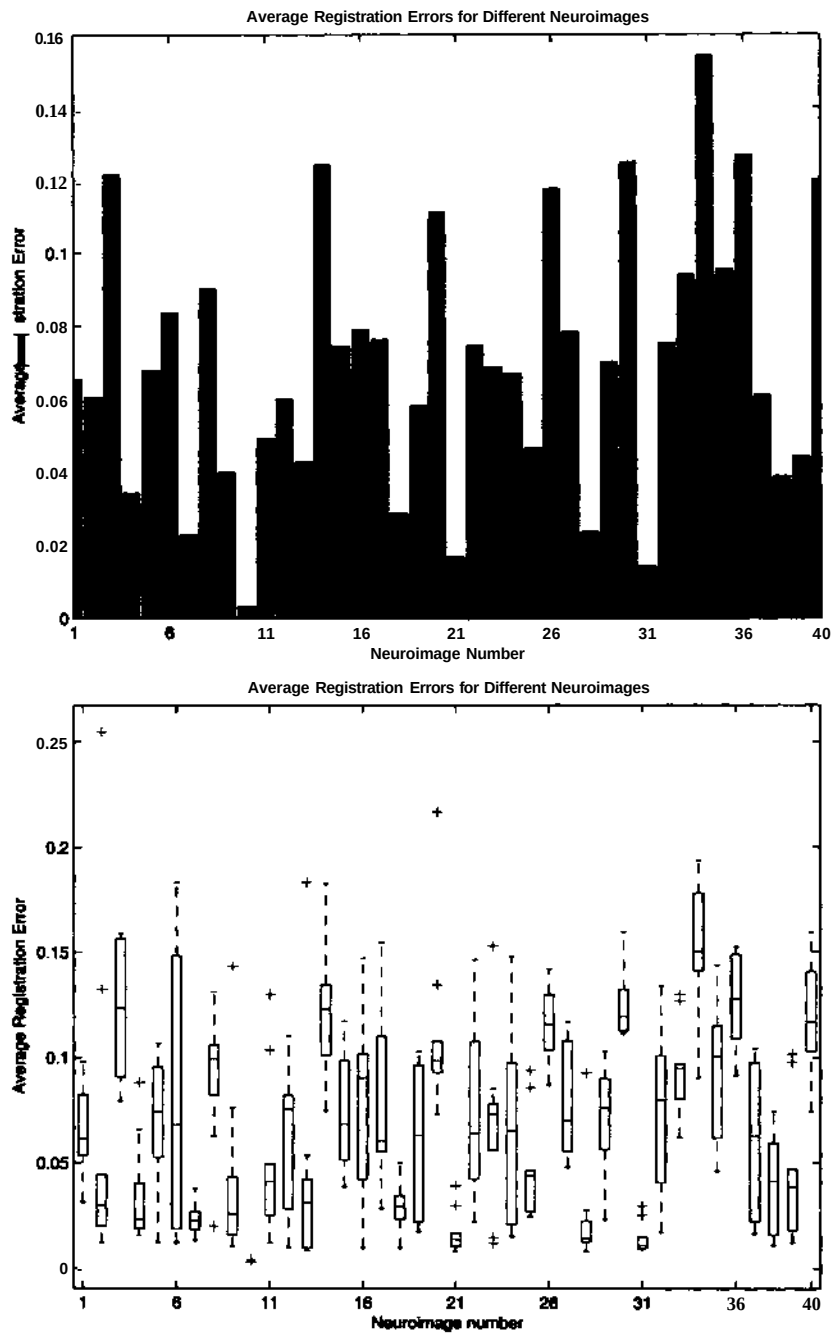

Figure 18: Registration errors when previously learnt feature subsets are used directly. "+" sign indicates outliers

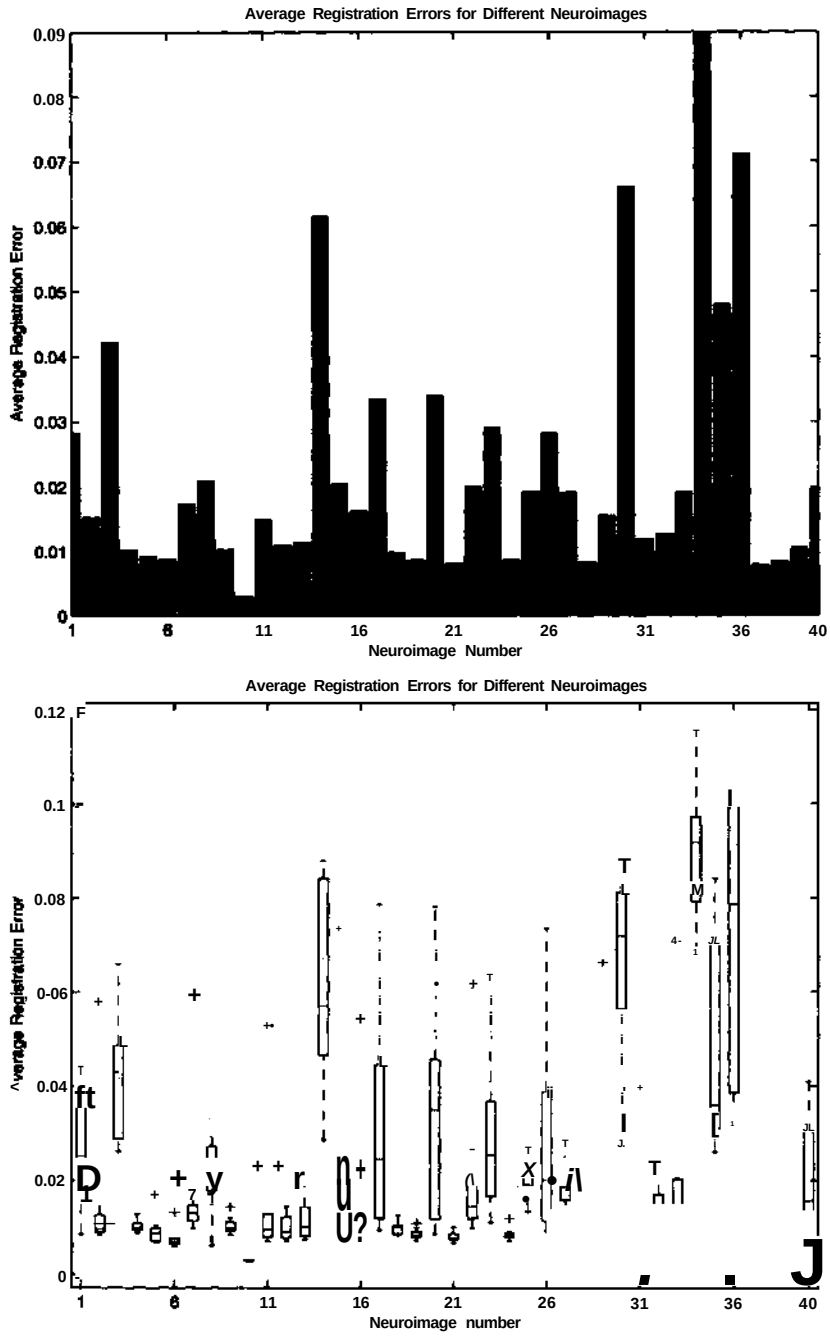

Figure 19: Registration errors when previously learnt feature subset for estimating correspondences is used to find interesting voxels. Forward selection is then used to find new feature subset for correspondences. "+" sign indicates outliers

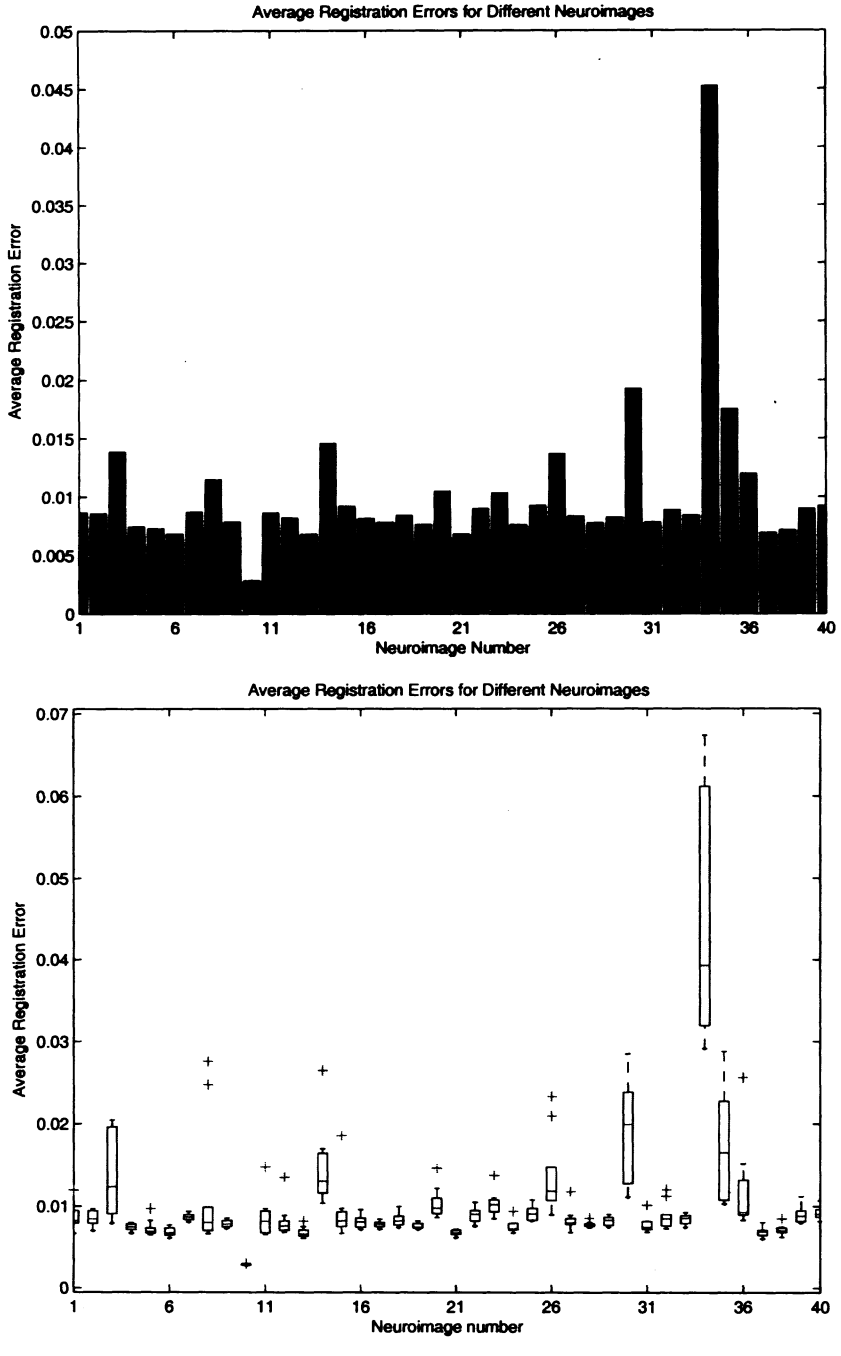

Figure 20: Bar and box plots of the registration errors for the algorithm with online learning. The smallest error corresponds to the registering transformed reference image to itself. "+" sign indicates outliers

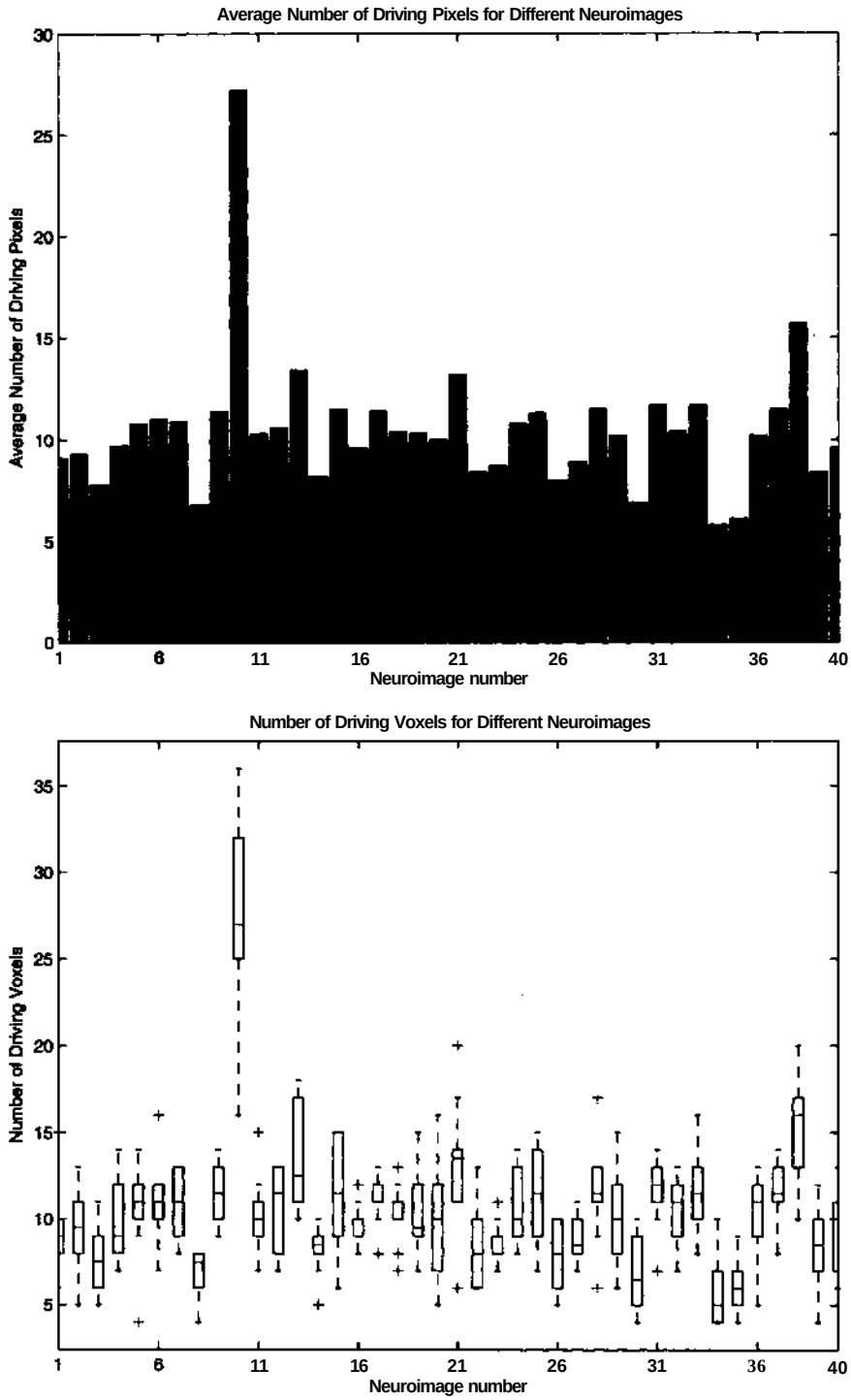

Figure 21: Bar and box plots of the number of driving voxels for each input image. We can see direct relationship between number of driving voxels and quality of the registration. The largest number of driving voxels corresponds to the case where reference image is registered to transformed copies of self. Smallest number of driving voxels corresponds to the case with the largest registration error.

give a simplistic example, suppose that we have an image that contains a circle and a small square. Assume that we have two features available: one that responds strongly to corners, the other one - to the center of a circle. In the subspace induced by corner-detecting feature, only corners of the square would be selected as interesting voxels. In the subspace induced by center-of-a-circle detecting feature only the center of the circle will be selected as an interesting voxel. Which voxels are selected as interesting when we use subspace consisting of these two features depends on how corner-detecting feature responds to non-corners and center-of-a-circle-detecting feature responds to non-centers. However, if we select subspace consisting of just the first feature for corners of the rectangle, and subset consisting of just the second feature for center of a circle, then both center of the circle and the corners will be selected as interesting.

To overcome these limitations we select different subsets of features for every voxel. In order to accomplish this, we need to propose a measure different from registration quality as the criteria for feature selection. The reason for this is that we cannot evaluate registration quality until we have chosen feature subsets for all the voxels. Therefore, a step of sequential feature selection would consist of adding/removing a feature to/from subsets associated with every voxel. Consequently, increase or decrease in registration quality at any given step of sequential selection cannot be attributed to any specific feature of a specific voxel, but rather to the group of features added/removed in the previous step. On the other hand, risk associated with finding correspondence for a given voxel is computed independently from the risk of the other voxels. Risk is a measure of quality of a given match and thus an indirect measure of registration quality. Hence, we use sequential feature selection to find a feature subset which minimizes risk associated with finding a match for a given voxel. The algorithm now has the following steps:

1. for every voxel in the neighborhood of a voxel *V{* in the reference image find its feature subspace in which *V{* can be matched to itself with least risk.
2. select voxels with risk smaller than a threshold as interesting voxels.
3. use the same subset of features as in step 1 to find correspondences for each interesting voxel and perform rough registration based on these correspondences. Note that for each voxel these subsets are different.
4. for each interesting voxel use sequential selection to find a subset of features using which we can match this interesting voxel and a voxel in the input image so that registration quality improves. Note that since the input image is already coarsely aligned with the reference image, improvement in the registration quality that we obtain by moving one voxel at a time is meaningful.

5.2 "Likelihood of the registration" as a similarity measure

Lets define likelihood of a correspondence to be the likelihood that feature vector of a voxel in the input image matches to some corresponding voxel in the model. This likelihood, *F(X\v{),* under the assumption of the independence of the feature vector components can be factorized according to the formula (1). The likelihood of the registration is the product of likelihoods of correspondences for all voxels in the input image. In practice, for numerical reasons, we can replace likelihood of a correspondence by a loglikelihood of a correspondence. In a very degenerate form, when feature subsets for *all* the voxels consist only of intensity value, and all Gaussians in the model have the same variance, loglikelihood of registration amounts to sum of square differences of intensities.

6 Conclusions

We have presented a novel learning-based method for deformable image registration. Our approach has 3 distinguishing characteristics:

1. It employs feature selection to automatically choose features that are best for registering a given pair of images.
2. It uses risk associated with finding correspondence for a given voxel as a way to automatically select landmarks to be used during registration.
3. It estimates a model of how features are distributed at each anatomical location and uses this model to find correspondences between voxels.

References

[I] Ashburner, J., Priston, K. *Nonlinear Spatial Normalization Using Basis Functions.* Human Brain Mapping 7:254 - 266, 1999

[2] Blum A., and Langley P. *Selection of relevant features and examples in machine learning.* Artificial Intelligence, 97(1-2):245271, 1997.

[3] Bookstein, F.L., *Principal Warps: Thin-Plate Splines and the Decomposition of Deformations* PAMI(ll), No. 6, June 1989, pp. 567-585.

[4] Canny, J. *A Computational approach to edge detection.* IEEE Pattern Analysis and Machine Intelligence, 8(6):679-698, 1986.

[5] Chen, M. *3-D Deformable Registration Using a Statistical Atlas with Applications in Medicine* doctoral dissertation, tech. report CMU-RI-TR-99-20, Robotics Institute, Carnegie Mellon University, October, 1999.

[6] Collins, R., and Liu, Y. *On-Line Selection of Discriminative Tracking Features.* Proceedings of the 2003 International Conference of Computer Vision (ICCV '03), October, 2003.

[7] Fischer, B., Modersitzki, J. *Intensity-based Image Registration with a Guaranteed One-to-one Point Match.* Methods Inf Med. 2004;43(4):327-30.

[8] Fischler, M.A. and Bolles, R.C., *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,* RCV87, 1987

[9] Gabor, D. *Theory of Communications.* Proc. IEE, vol.93, pp. 429-459, 1946.

[10] Harris, C, Stephens, M *A Combined Corner and Edge Detector.* Proceedings of $4^{t/l}$ Alvey Vision Conference, Manchester, 189-192, 1998.

[II] Hastie, T., Tibshirani, R., Friedman, J. *The Elements of Statistical Learning; Data Mining, Inference, and Prediction.* Springer series in statistics, 2001. ISBN 0-387-95284-5

[12] R. Kohavi, G.H. John, *The Wrapper Approach,* in Feature Selection for Knowledge Discovery and Data Mining, H. Liu and H. Motoda (Ed.), Kluwer, 33-50, 1998.

[13] Hellier, P., and Barillot, C. *Coupling Dense and Landmark-based Approaches for Non-rigid Registration.* IEEE Transactions on Medical Imaging, 22(2) (2003), pp. 217227.

[14] Ibanez, L., Schroeder, W., Ng, L., Cates, J. *ITK Software Guide* Kitware, Inc., 2003.

[15] Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M., Carter, C, Stenger, A., Davis, S., Aizenstein, H., Becker, J., Lopez, O., Meltzer, C. *Discriminative MR Image Feature Analysis for Automatic Schizophrenia and Alzheimer's Disease Classification.* MICCAI (1) 2004: 393-401

[16] Liu, Y., Zhao, T., and Zhang, J., *Learning Multispectral Texture Features for Cervical Cancer Detection.* Proceedings of 2002 IEEE International Symposium on Biomedicallmaging: Macro to Nano, July, 2002.

[17] Maintz, J., and Viergever, M. *A Survey of Medical Image Registration.* Medical Image Analysis, 2 (1998), pp. 136.

[18] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens P., *Multimodality image registration by maximization of mutual information* , IEEE transactions on Medical Imaging, vol. 16, no. 2, pp. 187-198, April 1997

[19] Mitra, S., and Liu, Y. *Local Facial Asymmetry for Expression Classification* Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), June, 2004.

[20] Periaswamy, S. *General-Purpose Medical Image Registration.* PhD thesis, Dartmouth College, April 2003.

[21] Pluim, J., Maintz, A., Viergever, M. *Mutual-Information-Based Registration of Medical Images: A Survey.* IEEE Transactions on Medical Imaging, Vol. 22, No. 8, August 2003.

[22] Shen, D. *Image Registration by Hierarchical Matching of Local Spatial Intensity Histograms.* Medical Image Computing and Computer-Assisted Intervention, Prance, Sept 26-30, 2004

[23] Shen, D., Davatzikos, C. *HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration.* IEEE Transactions on Medical Imaging, Vol 21, No 11, November 2002.

[24] Talairach, J. and Tournoux, P. *Co-Planar Steriotaxic Atlas of the Human Brain* Thieme Medical Publishers, 1988.

[25] Toga, A., Mazziotta, J. *Brain Mapping: The Methods.* Academic Press, 1996.

[26] Wells, W., Viola, P., Atsumi, H., Nakajima, S. and Kikinis, R. *Multi-modal volume registration by maximization of mutual information.* Medical Image Analysis, 1(1):35-51, March 1996.

[27] Zhong, X., Dinggang, S., and Davatzikos, C. *Determining Correspondence in 3D MR Brain Images Using Attribute Vectors as Morphological Signatures of Voxels* IEEE Transactions on Medical Imaging, 23(10): 1276-1291, Oct 2004