MRI Mode Invariant Feature Descriptors

Samantha Horvath
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh PA
shorvath@cmu.edu

Abstract

Multi-modal medical image descriptors are a very interesting avenue of research. Medical images (such as MRI, CT) have very different properties compared to the camera images that are the subject of most of computer vision research. Interest point descriptors such as SIFT may work well for medical images, but customization of the descriptors could be beneficial. This project report presents the results of my work concerning the design and testing of custom image descriptors on multi-modal MRI images.

1. Introduction

This project concerns the design of feature descriptors for medical image analysis. Specifically, I aim to design interest point descriptors that are invariant to the type of MRI (magnetic resonance imaging) mode used to acquire an image. I am looking at 3 MRI modes: T1-weighted imaging, T2-weighted imaging, and fractional anisotropy (FA). Each mode captures different information about the underlying tissue. T1 images are useful for differentiating between fat and water. T2 images can show edema. Fractional anisotropy images are derived from diffusion tensor images, and show the amount of water diffusion in a voxel location. Multi-modal interest point descriptors can be useful for feature detection, and for landmark-based non-rigid registration. However, multi-modal feature descriptors have not been studied before in depth.

Standard descriptors such as SIFT [1] have been designed to work on camera images. However, Brown and Winder [2] present a framework for designing custom descriptors that may work well for multi-modal medical images. There are a number of challenges inherent to this project. The required information overlap may not even exist, making any attempt to match features between the images impossible. This would only be true for very small scale image patches, and it has been shown that enough information overlap exists between entire images for alignment purposes [3].

1.1. Previous work

Multimodal interest point based registration has been done before, however the matching was not descriptor based. Wong and Bishop [4] use prior knowledge from successful alignments (in the form of stored matched image patches) in order to predict geometric transformations of the current query images. Previous work has also been done with SURF based interest point matching, but not in a multimodal setting [5]. The combination of interest point matching with multimodal feature points provides a novel aspect to this project.

2. Methods

This project incorporates some of methods in [2] to help design a descriptor capable of matching interest points across MRI modalities. The modular framework for the custom descriptors was implemented to help find an optimal descriptor type. Descriptors were tested for each of the possible pairings of the MRI modalities (T1-FA, T1-T2, T2-FA).

A major design decision concerns how interest points will be used in this formulation. Different interest point detectors (i.e Difference of Gaussian , Harris corners, etc.) recognize different feature types in images. These feature types may not be constant across MRI modes.

2.1. Dataset

The dataset consists of sets of T1, T2, and fractional anisotropy image volumes. A set of the three image volumes was taken for each patient, in roughly the same position. The three volumes are then aligned to each other using the ITK (Insight Toolkit) registration framework. The registration parameters for each volume set were hand tuned to ensure optimal alignment. This registration allows for trivial determination of ground-truth voxel correspondence. Each voxel in one volume corresponds directly to the voxel in the same location in the other volumes in the set. 2D images frames are then extracted from the volumes to serve as our test images and source of matchable interest points (see Figure 1).

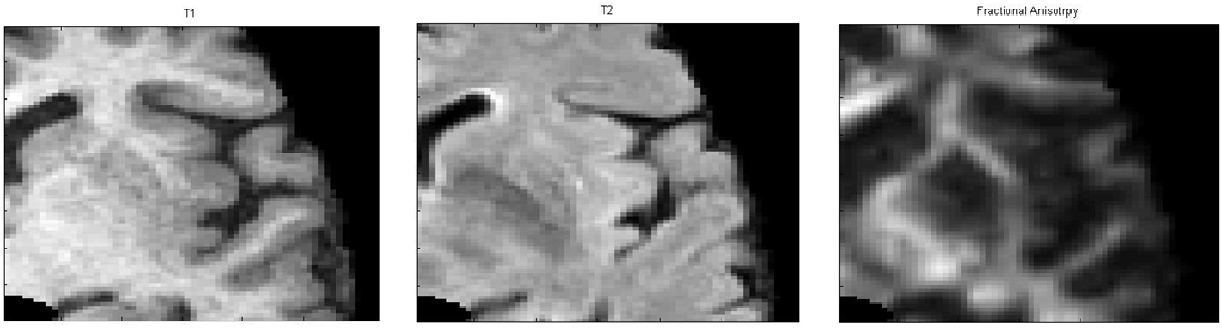

Figure 1: T1, T2, and Fractional Anisotropy image slices

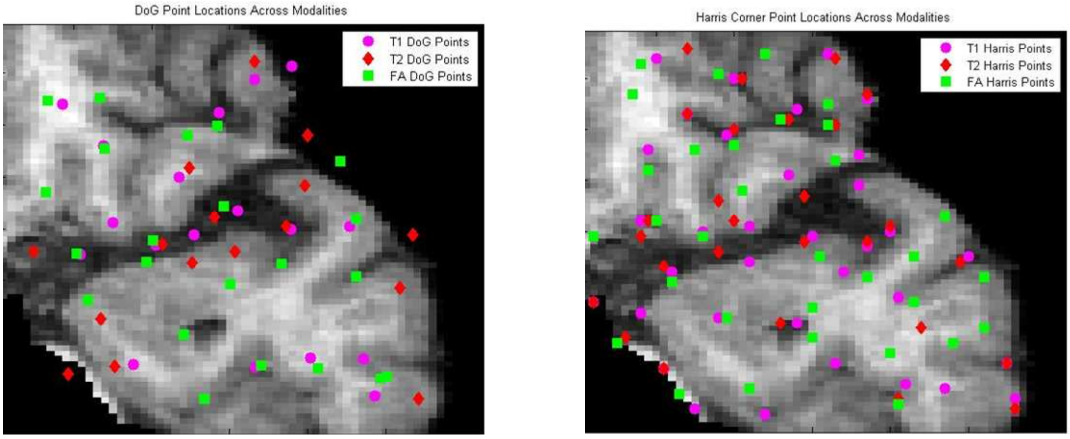

Figure 2: Interest Point Co-occurrences plotted over the T1 image

2.2. Interest point issues

The custom descriptor pipelines are tested on image patches extracted at interest points that could be matched across the three image types. Harris corners and Difference of Gaussian (DoG) points were tested to determine how well they could select true correspondences across the image types. However, both of these methods had difficult detecting interest points at the same locations across the modalities. Figure 2 shows the locations of interest points detected in T1, T2, and fractional anisotropy images with a Harris corner detector and a DoG detector. Matchable points are indicated by locations that have been detected as “interesting” in multiple image types. It can be seen that few “matchable” points are found by the DoG interest point detector. Considerably more are found by the Harris corner detector. Figure 3 shows the ratio of matchable points to total points found by each detector for each possible image type pair.

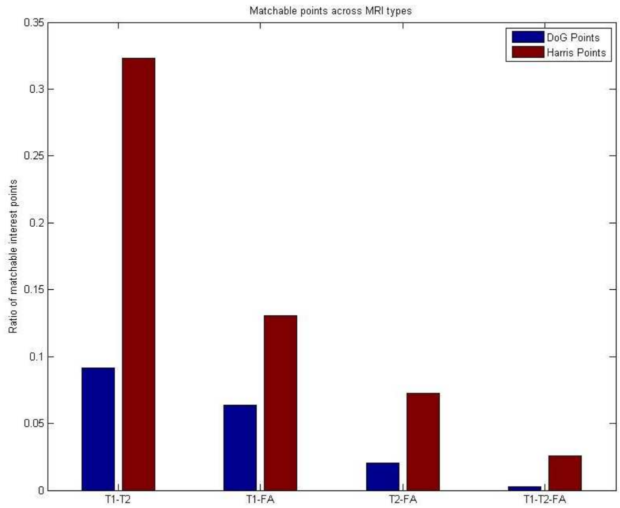

Figure 3. Ratio of matchable points. Note that the Harris points result in a better matchable fraction

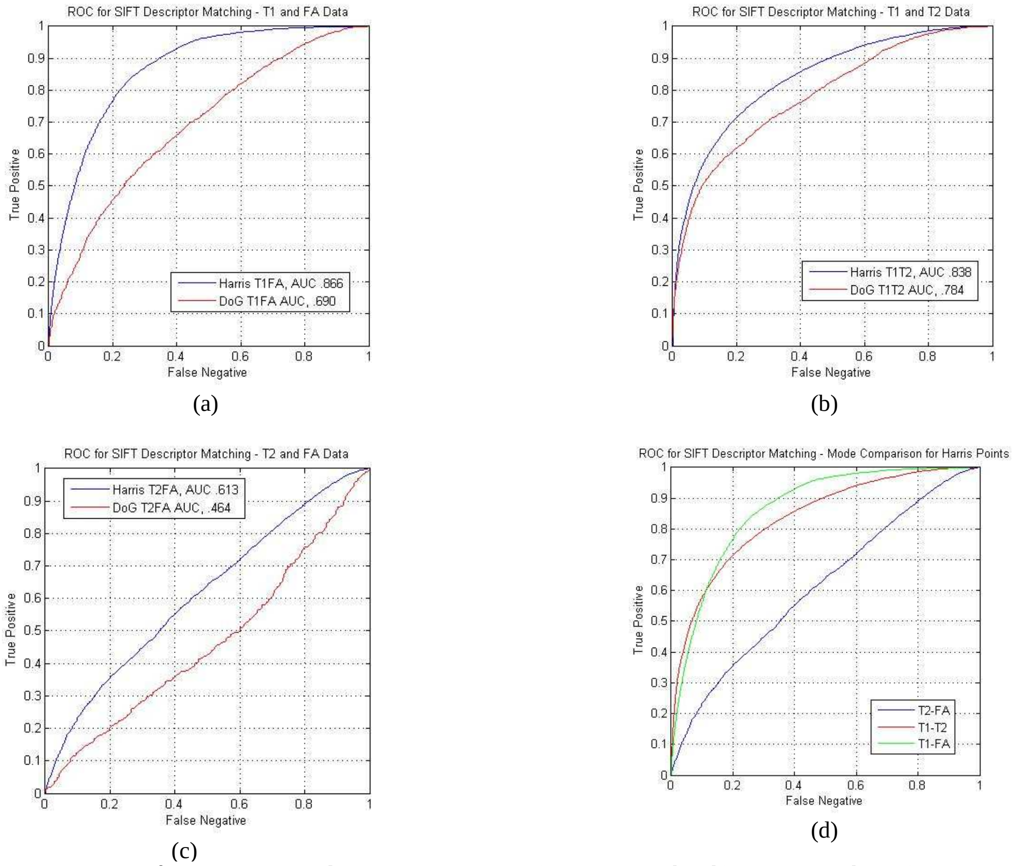

Figure 4: ROC curves for interest point detector comparisons a) T1-FA matches b) T1-T2 matches c)T2-FA matches d)all mode comparisons for Harris corners. (Standard SIFT implementation)

Sets of true and false matches were extracted for each pair of modalities (T1-T2, T1-FA, T2-FA). True match patches were extracted for Harris interest points that were successfully located in both image modalities. False matches were constructed from randomly associated interest points. These image patches were then used to test the custom descriptor pipelines

2.3. Custom descriptors

Custom descriptors were designed in the manner of [2], with each portion of the descriptor process represented by a modular "block." A transformation (TR-block) represents an operation (usually gradient based) applied at each pixel. The transformed image patch is then processed with a spatial pooling (S-block) method that creates a linear descriptor for the image patch. A number of TR-block/S-block combinations were tested, and ROC curves were computed. TR-blocks corresponding to the SIFT style gradient binning (TR1, 4 or 8 bins) and rectified gradients (TR2, 4 or 8 bins) were used. Spatial pooling methods corresponding to the SIFT style grid (S1) and DAISY [2] Gaussian weighted pooling centers (S2). Eight separate pipelines (all combinations of the TR and S blocks) were constructed and optimized. Normalization was performed on all pipelines.

3. Results

ROC curves for the interest point descriptors were constructed for both the standard SIFT implementation and for the custom pipelines.

3.1. Standard SIFT descriptor ROCs

Figure 4 shows ROC curves for matching points using the standard SIFT descriptor computed at Harris corners and DoG interest points. The SIFT descriptors extracted at Harris corners can be seen to outperform the DoG descriptors for all modality pairs. Figure 4 (d) shows that the T1-FA match pair performs the best with the standard SIFT descriptor. The T2-FA performs the worst, barely above chance. This is understandable, given how few true matches interest point matches could be found between the two image types.

3.2. Custom descriptor ROCs

Figure 5 shows ROC curves for matching using the eight custom pipelines for all three match combinations. The optimized custom pipelines gave poor performance compared to the standard SIFT descriptor, including the TR1-8, S1 pipeline, which mirrors the SIFT design. I

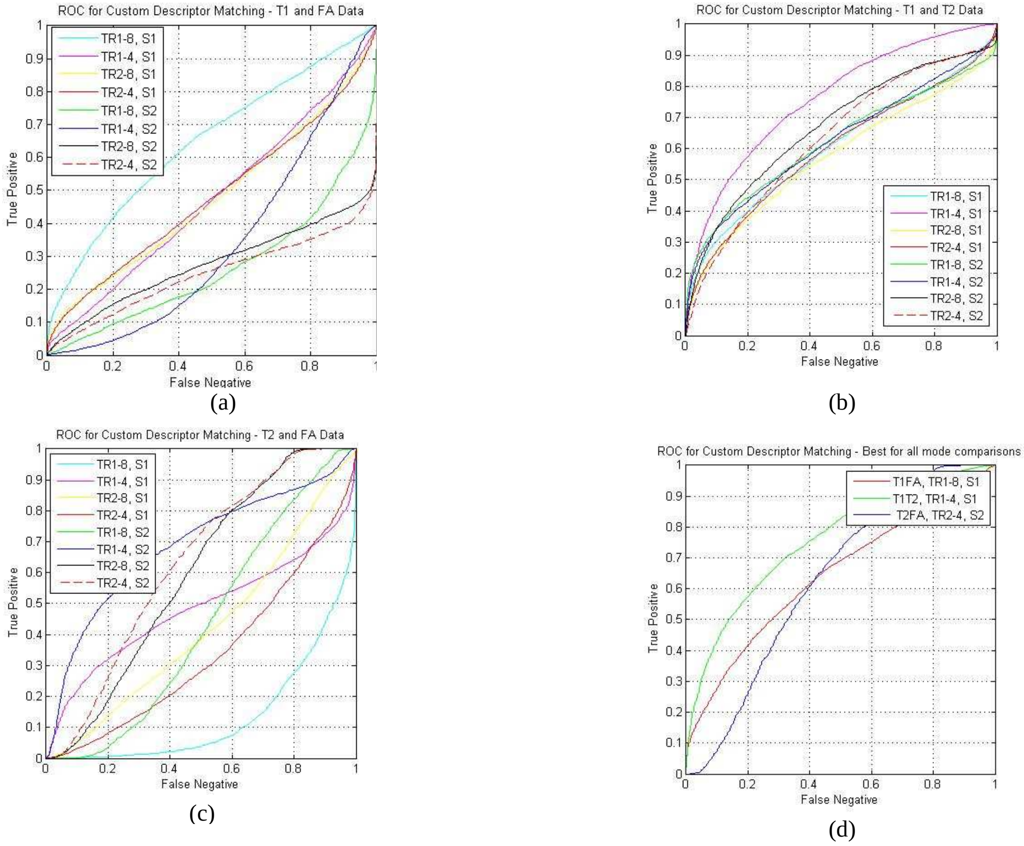

Figure 5: ROC curves for custom descriptors a) T1-FA matches b) T1-T2 matches c) T2-FA matches d)Best pipelines for all modes

have been continuing to investigate the source of this large error. However, I still feel that the custom pipeline results are valid for comparison to each other. For instance, the comparative performance of the SIFT clone pipeline (T1-FA > T1-T2 > T2-FA) was the same as that of the standard SIFT.

For the T1-FA pairing, the best pipeline was the SIFT clone, which is to be expected given the good performance given by the standard SIFT. For T1-T2 the best pipeline was the TR1-4, S1 combination (i.e. coarsely binned gradient version of SIFT). For T2-FA, the results were much more varied, however the best performers used the S2 block (DAISY style spatial pooling).

4. Conclusions

The overall results were underwhelming due to the as yet unknown errors in the custom descriptor pipelines. However, the one interesting result was the performance of the DAISY based descriptors for matching between the T2 and FA interest points.

Future work for this project will include designing the custom descriptors using image patches extracted at arbitrary match points (i.e. not restricting to point that are discovered by the interest point detector). A dense formulation is also being considered.

References

[1] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*. Vol. 20, no. 6, Nov. 2004.

[2] M. Brown, H. Hua and S. Winder, “Discriminative learning of local image descriptors,” *IEEE Trans. Pattern Anal. Mach. Intell.* ,vol. 33, no. 1, pp.43-57, Jan. 2011.

[3] W.M. Wells et al, “Multi-modal volume registration by maximization of mutual information,” *J. Med. Image Anal*, vol. 1, no. pp. 135-51. Mar. 1996.

[4] A. Wong and W. Bishop, “Indirect knowledge-based approach to non-rigid multi-modal registration of medical images,” *CCECE 2007*. pp.1175-1178, 22-26 April 2007.

[5] A. Wang et al, “Research on a novel non-rigid registration for medical image based on SURF and APSO,” *Image and Signal Processing (CISP), 2010 3rd International Congress on* , vol.6, no., pp.2628-2633, 16-18 Oct. 2010.