An Enhancement to SIFT-Based Techniques for Image Registration

Md. Tanvir Hossain, Shyh Wei Teng, Guojun Lu
Gippsland School of Information Technology
Monash University
Churchill, Victoria 3842, Australia

Martin Lackmann
Department of Biochemistry & Molecular Biology
Monash University
Clayton, Victoria 3800, Australia

e-mail: {Tanvir.Hossain, Shyh.Wei.Teng, Guojun.Lu, Martin.Lackmann}@monash.edu

***Abstract*—Symmetric-SIFT is a recently proposed local technique used for registering multimodal images. It is based on a well-known general image registration technique named Scale Invariant Feature Transform (SIFT). Symmetric SIFT makes use of the gradient magnitude information at the image's key regions to build the descriptors. In this paper, we highlight an issue with how the magnitude information is used in this process. This issue may result in similar descriptors being built to represent regions in images that are visually different. To address this issue, we have proposed two new strategies for weighting the descriptors. Our experimental results show that Symmetric-SIFT descriptors built using our proposed strategies can lead to better registration accuracy than descriptors built using the original Symmetric-SIFT technique. The issue highlighted and the two strategies proposed are also applicable to the general SIFT technique.**

Keywords-multimodal image registration; SIFT; histogram weighting

I. INTRODUCTION

Image registration is a very important image analysis technique used in a wide range of domains including remote sensing, computer vision and medical imaging. It is a process of aligning two images which have most likely been acquired in different imaging conditions. The imaging conditions may vary based on a number of factors, such as time, viewpoint, illumination, type of sensing device, noise, cluttering and occlusion. A good image registration technique should be able to correctly identify the corresponding regions and figure out the appropriate geometric transformation required to map the sensed image on the reference image, despite the presence of any (or all) of the above mentioned factors.

A. Image Registration and Image Description

To register two images it is necessary to measure the amount of dissimilarity, misalignment or lack of correspondence between the two input images. The lower the dissimilarity, the better is the alignment. To make this comparison possible, different registration techniques employ different sorts of descriptors. Descriptors may be broadly categorized into two types: global and local. Global descriptors describe the entire image as a whole and, therefore, there would usually be only one such descriptor per image. Local descriptors, on the other hand, represent prominent and stable parts of an image. Thus a single image may have more than one local descriptor each describing one of the stable parts in the image.

Unlike local techniques, global techniques are generally less affected by local deformations. However, there are a number of problems associated with the global techniques. One of the major problems is that they tend to overlook local variations in images. Though the local variations do not affect the overall result significantly, local errors in registration become quite difficult to avoid using global techniques. Again, global techniques perform poorly in presence of additional objects (or image contents) in one of the images. They also cannot produce good results if the overlap between the images is low.

In spite of the fact that the initial miss-registration in local registration techniques may sometimes lead to incorrect results, using local techniques have a number of strong advantages. In comparison to global techniques, keypoint-based local registration techniques are more invariant to similarity and affine transformations. Local techniques are also less affected by the presence of outliers, occlusion and clutter. This is why local techniques are also less affected by truncation or low overlap of images.

B. Multimodal Image Registration

Multimodal image registration has received sufficient research focus over the past decade and a good number of techniques for such registration have also been proposed [1-4]. Two or more images are called multimodal if each of them is captured by a different sensor (imaging device or modality). In addition to possibly having totally different combination of intensities between multimodal image pairs, other variations in imaging conditions (such as different scales and viewpoints) can still be present. Therefore, multimodal image registration is far more complicated and challenging as compared to basic registration problems.

There are a lot of global techniques [1-7] found in the literature that can be used for multimodal image registration. These techniques are mostly based on entropy [8] and other correlation-like statistical measures. Besides the common problems associated with global techniques, these techniques usually need to repeatedly compute the global measure for all possible transformations in the search space which can be computationally very expensive.

On the contrary, despite the strength of using local description methodologies, very few modality invariant local description techniques [9,10] are found in the literature. In this paper we concentrate on a recently proposed local

978-0-7695-4271-3/10 $26.00 © 2010 IEEE
DOI 10.1109/DICTA.2010.39

description technique named Symmetric-SIFT, which is in essence a variation of a widely used local description technique – SIFT [11,12]. Unlike SIFT, Symmetric-SIFT has been adapted to be used for multimodal image registration.

C. Contribution of the Paper

In this paper, we identify a problem associated with the feature descriptors of Symmetric-SIFT and propose solutions to the problem. Our experimental results show that our proposed solution strategies improve registration accuracy.

The problem that we try to solve in this work is associated with the way the orientation histograms are built in Symmetric-SIFT. We shall see in Section II that these histograms are the basic building blocks of the final descriptor. The existing approach may produce the same or similar histograms for image patches even though they visually appear totally different. We propose a modified orientation histogram that is more effective in fairly distinguishing between visually different image contents. As Symmetric-SIFT is basically a variant of SIFT, our proposed approach is also applicable to SIFT and other SIFT-like techniques that follow the same principle in building the orientation histogram. Therefore, in our research, besides multimodal images we also have carried out experiments with mono-modal images.

The rest of the paper is organized as follows. In Section II we describe the fundamental operations involved in Symmetric-SIFT. Then, we identify issues with the existing technique and propose our solution strategies in Section III and Section IV respectively. In Section V we present our experimental results and finally we conclude in Section VI.

II. An Overview of Symmetric-SIFT

Both SIFT and Symmetric-SIFT are basically keypoint based image description techniques. According to the similar analogy as stated in [13], most keypoint based image registration techniques can be decomposed into the following five phases – keypoint detection, keypoint description, similarity matching, transformation estimation and image transformation. Keypoints are those points in an image that are believed to survive in a wide range of geometric and photometric transformations. The purpose of the detection phase is to identify such stable points from a given image. In the description phase, each identified keypoint is described numerically for them to be used for comparison in the similarity matching phase. The next phase is to compute similarity of the descriptors from the two images. The set of best-match descriptors indicates the corresponding parts between the images. This information is used for deriving a transformation function that maps one image onto the other and that becomes the final phase of registration. The accuracy of registration, therefore, depends on the accuracy of the match set which in turn depends on the distinctiveness of the descriptors themselves.

As majority of the operations in SIFT and Symmetric-SIFT are common, we first describe how SIFT works.

A. Steps in SIFT

Scale Invariant Feature Transform (SIFT), originally introduced by Lowe [11,12], is a widely used descriptor in both image registration and retrieval. Fig. 1 outlines the major steps involved in SIFT and maps them to the first three phases of image registration. SIFT is basically a keypoint based local descriptor. The keypoints are identified by applying Difference of Gaussian (DoG) in scale space. DoG is a close approximation to Laplacian of Gaussian (LoG), yet much quicker to compute. The repeatability of SIFT keypoints is high. Once the keypoints are identified, the dominant orientation O of the gradients is computed within a region R around each keypoint. The size of R is determined based on the scale at which the keypoint was identified. This makes the descriptor (to be built later on) invariant to scale change. Rotation invariance, on the other hand, is achieved by building the descriptor relative to the dominant orientation O. The final descriptor is built on a 4 by 4 spatial grid where each cell in the grid consists of an 8-bin orientation histogram. All gradients within a cell are

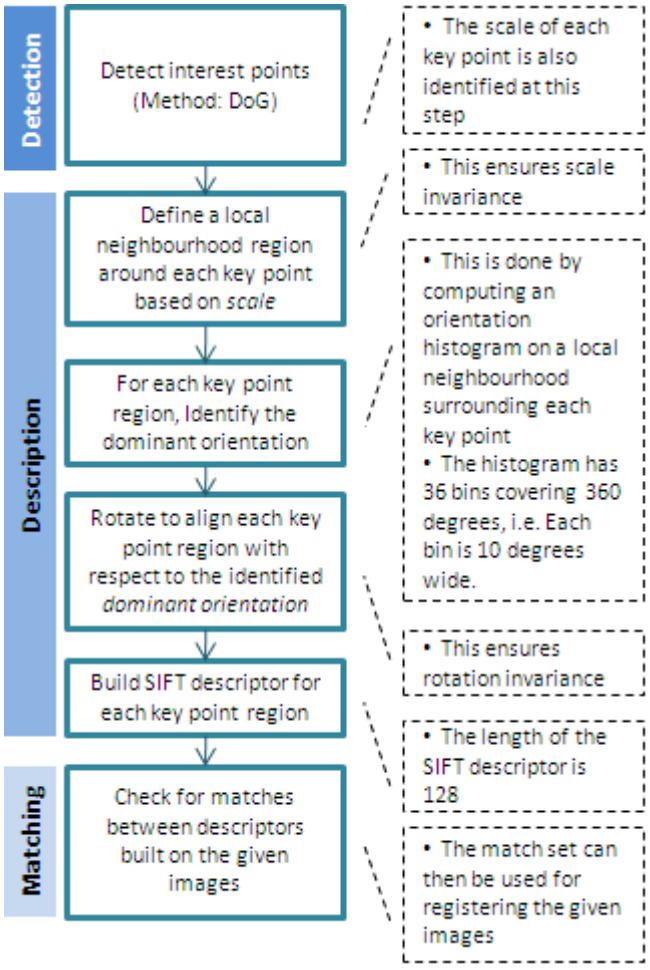

Figure 1. Steps in the keypoint detection, description and matching phases of SIFT.

quantized into one of these 8 bins. Thus we get a descriptor

of (4*4*8 or) 128 dimensions. The histogram values within the region are weighted with a Gaussian window in order to give more importance to gradients near the centre of the region. Descriptors computed in this way can then be used in subsequent steps of image registration.

B. *Properties of Multimodal Images*

There are certain properties of multimodal images that hold for most general cases. Firstly, the same portion of an object may be represented by different intensities in different modalities [5]. This is because different sensors may have different levels of sensitivity to a particular part of an object. Secondly, portions of an image may appear or vanish across modalities. In other words, portions of an object may remain invisible to some sensors and visible to others as some sensors cannot realize their presence whereas the others can. In fact, the second property can be thought as a special consequence of the first property.

It is very common in multimodal images that the gradients of corresponding parts of the images will change their direction by exactly 180˚ [5,14]. Let us call this property as 'Gradient Reversal' for future reference. Gradient reversal is one of the main reasons that causes SIFT to fail with multimodal images. This is because the gradient reversal also reverses the direction of the dominant orientation. Therefore, even if two visually similar regions are rotation normalized, they can still be totally out of phase. As a result, descriptors built on these regions will not match.

C. *Steps in Symmetric-SIFT*

Symmetric-SIFT, which was recently proposed by Chen and Tian [10], is capable of handling gradient reversal. Being a variant of SIFT, it naturally inherits all the strengths of SIFT and applies them to the multimodal domain.

According to Fig. 1, Symmetric-SIFT differs from SIFT in the following two steps:

1) *When the dominant orientation is identified, and*
2) *When the descriptor is built.*

It uses averaging squared gradients to determine the orientation of each keypoint. Moreover, at the descriptor building step, Symmetric-SIFT compensates the reversal of keypoint regions by accumulating information from both normal and reversed regions and forming a single descriptor.

III. Issues Associated with Building the Orientation Histograms in Symmetric-SIFT

We have seen in Section II that the final (SIFT/Symmetric-SIFT) descriptor is basically a combination of 16 orientation histograms. Now, let us take a closer look at how these orientation histograms are built.

As described in Section II-A, each cell in the 4 by 4 spatial grid of a keypoint region has its own orientation histogram. The orientation histogram captures and summarizes the overall distribution of gradients within a particular cell. To be more specific, the histogram is built based on gradient directions and magnitudes. Fig. 2 shows the key elements used in the process.

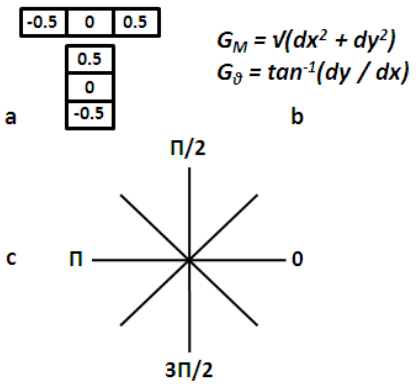

Figure 2. (a) The horizontal and vertical derivative kernels used to compute dx and dy respectively. (b) Shows how the derivatives are used to derive the gradient's magnitude and direction. (c) Shows a sample orientation histogram with 8 bins each having a range of 45˚. When a gradient is observed, it's direction is used to decide which of the 8 bins should be incremented.

As we can see from Fig. 2, the direction and magnitude values are determined based on the intensities of the horizontally and vertically opposing pixels around every pixel within a cell. In Symmetric-SIFT, the orientation histogram consists of 8 orientation bins which divide the entire range [0, Π) into 8 equal ranges. Note that, reflecting all orientations greater than Π to the range [0, Π) is one of the measures that enables Symmetric-SIFT to be invariant to gradient reversal. Original SIFT, on the other hand, divides the entire range [0, 2Π) into 8 equal bins. Whenever a pixel is examined, the direction of gradient at that point is used to determine the appropriate orientation bin. The value of this bin is incremented by the magnitude of the associated gradient. However, this way of building the histogram definitely does not fairly represent the profile of the occurrences of gradients in different directions. Consequently, during matching, the difference between two histograms will not be proportional to the difference between the actual visual appearances of the image patches they represent. Fig. 3 illustrates this issue in building the orientation histogram with an example which shows that it is not unusual to get very similar histograms from two image patches having very different visual appearances.

In Fig. 3, we see two example image patches. The image patches are visually quite dissimilar from one another. However, orientation histograms built from these patches will be exactly the same if gradient magnitudes are used to increase the histogram bins. This may lead to the false conclusion that the corresponding image patches are similar.

IV. Improvement to Orientation Histograms

In the previous section we have shown that, adding magnitudes to the orientation histograms of SIFT may not appropriately correspond to the actual visual appearance of an image patch. On the other hand, for some images the gradient magnitudes may contain important visual information. However, the method of incorporating the gradient magnitude information into the descriptor should be carefully designed.

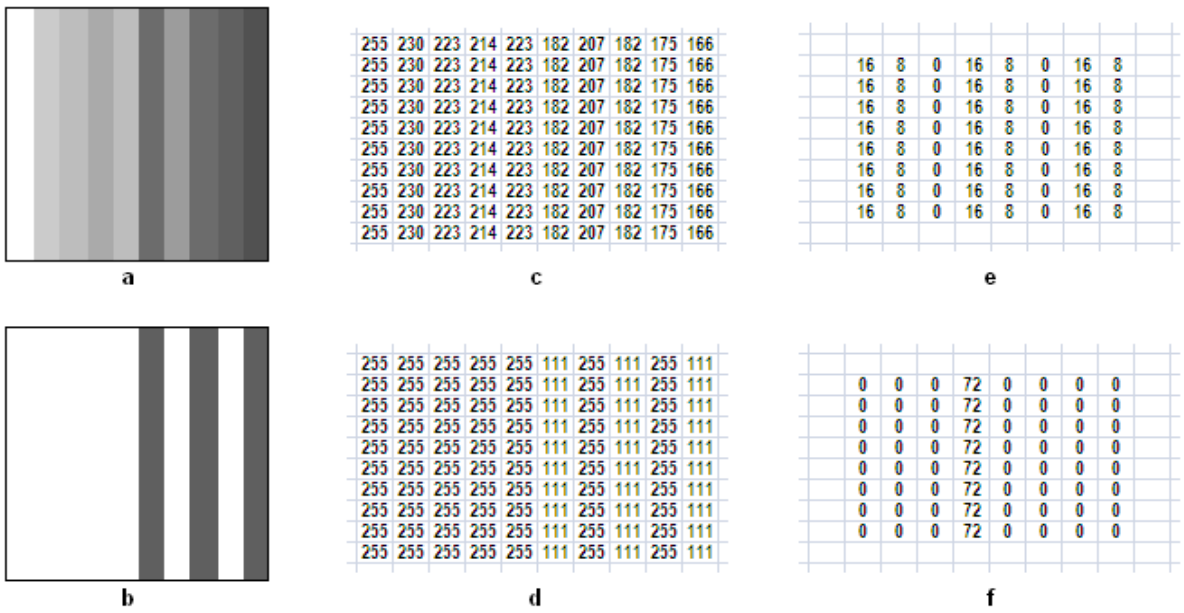

Figure 3. Shows how two very different image patches can produce identical orientation histograms. (a) and (b) are the two example image patches. (c) and (d) are their corresponding intensity maps. (e) and (f) are the corresponding horizontal derivative (magnitude) maps. As both images have change in horizontal direction only, only one bin in their corresponding orientation histograms will be populated. Note that, the sum of magnitudes in both of the derivative maps is 576. Thus the orientation histograms will also be identical though the original image patches were visually quite different.

To address the above mentioned issue we propose the following two strategies of incrementing the orientation histogram bins:

1) Increment by 1. In this case the histogram will represent the number of occurrences of gradients in different directions.

2) Increment by Average of Squared Difference (ASD) of magnitudes.

The second strategy falls somewhere in between the two extremes. (One extreme is to add the magnitudes directly and the other is to totally discard them). The term ASD has been explained below.

Let H be the orientation histogram having 8 bins b_n where n = 1, 2, 3, …, 8. Let G_{ni} be the i-th gradient sample encountered to be in bin b_n and M_{ni} be its corresponding magnitude. Rather than adding each M_{ni} to bin b_n we propose to add the ASD of these magnitudes about their mean. Mathematically,

$$ASD_n = \Sigma (M_{ni} - Avg(M_n))^2 / N \quad (1)$$

where, ASD_n is the ASD of bin b_n, $Avg(M_n)$ is the arithmetic mean of all magnitudes encountered in bin b_n and N is the number of gradients encountered in that bin.

In Section V, we present our experimental results which show that, in most of the cases, both of the proposed strategies provide better accuracy as compared to the original method of building orientation histograms.

V. Performance Study

We have used eight pairs of images [9,15] for our experiments. These include five pairs of multimodal and three pairs of natural images. A couple of image pairs are presented in the appendix highlighting the matching points. Symmetric-SIFT was used to build the descriptors. The only change that we made to this technique is the way the orientation histograms are built, i.e. we have avoided adding the magnitude terms. Instead, we have examined the effect of adding ASDs. In a separate setup, we also examined how it performs if simply 1 is added to the bins. When we do so, each bin in the orientation histogram will only represent the number of occurrences of pixels where the gradients are oriented within a predefined range of angles. The magnitudes of the gradients, in this case, are not taken into account.

A. Evaluation Measures

The accuracy of an image registration technique highly depends on the accuracy of the keypoint match set. The higher the proportion of the identified true matches, the better the accuracy of the final registration will be. Therefore, we consider the accuracy of the match set as our evaluation criteria where accuracy is defined as,

$$\text{Accuracy} = (\text{Number of True Matches} / \text{Number of Total Matches}) * 100\% \quad (2)$$

The ground truths for the image pairs used in the experiments were determined manually. A maximum of 4 pixel error was considered to be accepted as a true match. This is consistent with the acceptable threshold described in existing literature [16].

B. Results

The following table summarizes our experimental results.

TABLE I. EXPERIMENTAL RESULTS

Image Pair	Increment Strategy	# of True Matches	# of False Matches	Accuracy (%)
Horse	Magnitude	153	8	95.03
	1	**144**	**3**	**97.96**
	ASD	154	7	95.65
Brain1	Magnitude	128	8	94.12
	1	**127**	**4**	**96.95**
	ASD	105	6	94.59
Brain2	Magnitude	124	15	89.21
	1	116	11	91.34
	ASD	**108**	**6**	**94.74**
Brain3	Magnitude	11	29	27.50
	1	6	16	27.27
	ASD	**10**	**18**	**35.71**
Brain4	Magnitude	55	18	75.34
	1	33	7	82.50
	ASD	**45**	**8**	**84.91**
Streets	Magnitude	4	9	30.77
	1	2	4	33.33
	ASD	**3**	**0**	**100**
City	Magnitude	50	14	78.13
	1	**55**	**5**	**91.67**
	ASD	35	7	83.33
Ubc	Magnitude	854	22	97.49
	1	**708**	**9**	**98.74**
	ASD	807	15	98.18

The first column gives the names of the image pairs. The second column says which strategy was used to increment the histogram bins to build the descriptor. The third and fourth columns specify the number of true and false matches respectively. Finally, the rightmost column gives the accuracy of the detected match set.

The experimental results show that, on the average, the accuracy of the match set increases by 16.94% when ASD is used to build the orientation histograms instead of using magnitudes in the traditional SIFT way. On the other hand, on the average, the accuracy increases by 5.47% if we simply count the number of occurrences in the histogram bins and discard any magnitude weighting.

Now, let us first examine why adding 1 gives better accuracy. This approach will produce totally different histograms for cases that are similar to the one illustrated in Fig. 3. Thus the chance of getting false matches is reduced significantly and the overall accuracy is improved. ASD, on the other hand, embeds some magnitude information in the histogram. However, the extent of this weighting is far less than weighting by actual gradient magnitudes calculated using G_M (Fig. 2). Two ASD values differ from one another if their associated visual contents are too different from one another. This is why the use of ASD also results in decreased number of false matches and the overall accuracy is improved. In many medical imaging applications including multimodal image registration and 3D image reconstruction, accuracy of the descriptor match set is crucial to get an acceptable end result.

None of the proposed approaches affect the size of the final descriptor. The complexity of computing ASD is *O(N)* which also does not impose any significant overhead to the technique as a whole.

VI. CONCLUSION

In this paper, we propose two alternate strategies of weighting the orientation histograms. These strategies are applicable not only to Symmetric-SIFT but also to SIFT and other SIFT-based techniques. It is observed that the proposed strategies always help in getting improved accuracy which implicitly justifies our observation that, "weighting orientation bins by gradient magnitudes cannot be the most appropriate way to build the histogram."

REFERENCES

[1] P. Viola, and W.M. Wells Iii, "Alignment by maximization of Mutual Information," International Journal of Computer Vision, vol. 24, pp. 137-154, 1997.

[2] J. Orchard, "Efficient least squares multimodal registration with a globally exhaustive alignment search," IEEE Transactions on Image Processing, vol. 16, pp. 2526-2534, 2007.

[3] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The Correlation Ratio as a new similarity measure for multimodal image registration," MICCAI'98, pp. 1115, 1998.

[4] P.E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using Fast Fourier Transform techniques," IEEE Transactions on Geoscience Electronics, vol. 8, pp. 353-368, 1970.

[5] J. Pluim, J.B. Maintz, and M. Viergever, "Image registration by maximization of combined Mutual Information and gradient information," MICCAI 2000, pp. 103-129, 2000.

[6] R. Xu, and Y. Chen, "Wavelet-based multiresolution medical image registration strategy combining mutual information with spatial information," Int. J. Innov. Comput. Inf., vol. 3, pp. 285-296, 2007.

[7] D. Russakoff, C. Tomasi, T. Rohlfing, and C. Jr Image, "Similarity using Mutual Information of regions," ECCV2004, pp. 596-607, 2004.

[8] S. E, and W. Weaver, "A mathematical theory of communication," Bell Syst. Tech. J., vol. 27, pp. 379-423, 1948.

[9] A. Kelman, M. Sofka, and C. Stewart, "Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations," CVPR'07, vol. 3, 2007.

[10] J. Chen, and J. Tian, "Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor," Progress in Natural Science, vol. 19, pp. 643-651, 2009.

[11] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, pp. 91-110, 2004.

[12] D.G. Lowe, "Object recognition from local scale-invariant features," The Proc. of the Seventh IEEE International Conf. on Computer Vision, vol. 1152, pp. 1150-1157, 1999.

[13] B. Zitová, and J. Flusser, "Image registration methods: a survey," Image and Vision Computing, vol. 21, pp. 977-1000, 2003.

[14] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal, "3D multi-modality medical image registration using feature space clustering," Springer, pp. 195, 1995.

[15] K. Mikolajczyk, and C. Schmid, "A performance evaluation of local descriptors," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 1615-1630, 2005.

[16] Y. Gehua, C.V. Stewart, M. Sofka, and T. Chia-Ling, "Registration of challenging image pairs: initialization, estimation, and decision," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, pp. 1973-1989, 2007.

APPENDIX

The following two sample image pairs are chosen from our test images. A line connecting two points from one image to the other indicates a match.

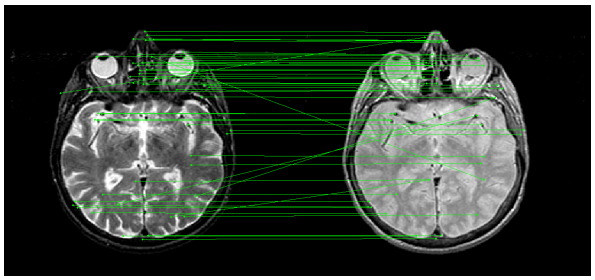

Figure 4. Brain4. Brain PD and T2 images.

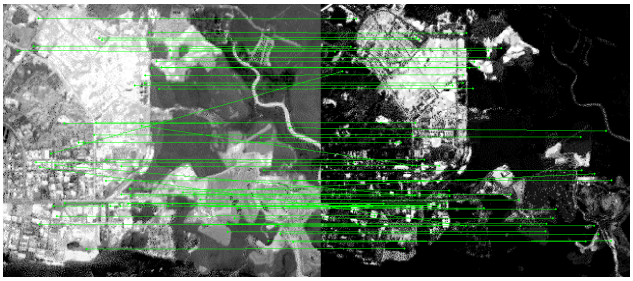

Figure 5. City. Areal image pair.