STEREO-IMAGE REGISTRATION BASED ON UNIFORM PATCHES

M. Abbasi-Dezfouli
Computer Science and Statistics Dept
College of Science, Shahid Chamran University
Ahwan, Iran

T.G. Freeman
School of Computer Science
Australian Defence Force Academy
Canberra, ACT 2600, Australia

Commission III, Working Group H3002

KEY WORDS: Registration, stereo-images, image matching, uniform patches, feature-based matching, automatic registration, SPOT.

ABSTRACT:

The automatic processing of stereo-images to determine terrain height has proved to be a far from trivial problem. Methods have been developed by a number of people that rely on operator guidance in establishing ground control points or seed points to start the matching process.
In an attempt to remove the operator interaction altogether, we have developed an unsupervised feature-based matching method based on identifying patches of uniform colour. By comparing the shapes and positions of such patches, we can establish how coarse features of the two images match, giving us a method for initiating the matching process.
Details of the shape comparison method are presented, along with results when applied to SPOT stereo-pairs and close range images.

1. INTRODUCTION

The use of stereo-images for estimating distance to parts of a scene has been long established in photogrammetry and in robotics. The task of automating the matching process has been an active research area for a decade, with many published methods and some commercial systems now available (Kauffman and Wood, 1987; Otto and Chau, 1989; Day and Muller, 1989; Li, 1991; Schenk *et al*, 1991; Trinder *et al*, 1990, 1993). Almost all of these methods rely on an operator providing seed points to start the matching process.

In an effort to minimize the extent of operator involvement in this process, we have looked at the challenge of performing the matching entirely automatically. If we can succeed at this, it will find immediate application in robotics and in close-range work with fixed camera positions. Its application to photogrammetry will be of less advantage, because ground control points are needed to overcome uncertainty in the position of the viewpoints and viewing direction. Nevertheless, success could reduce the dependence in photogrammetry on so many ground control points, with fewer requiring to be located in the image, and the remainder being supplied as X,Y,Z coordinates to be reconciled with the matching process. This would alleviate the problem of precisely locating all the points in the images, an unfortunate source of error in current methods.

To automate the matching process, we have considered the human visual system and the way we would perform the task if denied the option of using our eyes concurrently, one for each view. Of course, human physiology is extremely efficient in performing this matching process in everyday life, where both eyes concurrently provide the differing views, and the brain is able to roughly estimate distance over a whole scene within a fraction of a second. The challenge of performing similar matching with computer hardware remains a very distant goal.

When humans are denied concurrent access to the two views, our visual system is still able to match them, but is much less efficient, by several orders of magnitude. When confronted with two views, the eyes rapidly search for features that can be located in the other image, and from coarse details such as these, finer comparisons are made, picking up more subtle differences and similarities in the images.

Different types of feature have been investigated by various workers seeking to mimic this behaviour within a computer. Edge features, or lines where there is a major change in light intensity across them, have received much attention (Marr and Hildreth, 1980; Harallick, 1984; Greenfeld, 1987; Otto and Chau, 1989; Schenk *et al*, 1991). They are not always easy to work with, because of difficulties in defining which direction the changes in intensity needs to be measured along, and in setting a threshold which will give continuity of edges while not swamping the matching process with too much unwanted detail. Point features, such as corners of objects, are particularly attractive because they are easy to compare (Moravec, 1977; Förstner, 1986, 1987; Trinder *et al.*, 1990), and extension to the whole area from the found features is quite natural.

The approach we have adopted is to look for features which are areas of uniform colour (Abbasi and Freeman, 1994a). This is motivated by looking at images of natural terrain, where water bodies and roofs of buildings are often the first features identified. Of course, these are not the only features that can be identified; roads of characteristic shape can be strong, easily recognized features, as can also be fences between paddocks of differing vegetation, particularly the corners.

The early details of our patch-based approach have been published elsewhere (Abbasi and Freeman, 1994a,b). In this paper, the basic method will be outlined, and then our efforts to proceed from matched patches to actual points in the images will be described, along with tests with SPOT images and close-range images.

2. PATCH-BASED MATCHING

2.1 Patch extraction

An advantage of patch-based matching is the simplicity of definition of the feature and its computer implementation. To find a uniform patch, we search in the image for an area of uniform colour or gray value that is larger than some minimum size, and expand the area to include neighbouring pixels of the same colour.

The minimum size was chosen arbitrarily as a 3x3 pixel window, this size being large enough to be visible by eye in a typical SPOT image, but small enough to not miss small patches. The notion of "uniform colour" had to be adjusted to accommodate the variation in colour typically found in areas visually perceived as uniform. Extensive testing established that a grayscale tolerance of ± 2 was most effective. Even with this tolerance, there are areas of bright colour that we perceive as uniform but which exceeds this level of variation. Perhaps a logarithmic variation of the tolerance with gray level is needed to account for this. We have not tested this yet.

In implementation, a patch is stored as a series of scanlines, with a start and end pixel address for each line; concave patches may require more than one start and end for the one line. Associated with a patch will be its bounding area, the minimum and maximum X and Y coordinates. The area can be quickly derived by summing the differences between the start and end for each scanline. Drawing the boundary of a patch is more difficult with this storage scheme, as the start for a particular scanline must be matched with its corresponding one in the next scanline. When there is no corresponding point in the next scanline, a horizontal line is needed.

2.2 Patch matching criteria

With each image having a set of patches found by the above procedure, the next task is to find how the patches match. A deliberate decision was made to ignore gray value or colour in this comparison. This was based on the observation that in images of natural scenes, water bodies often appear as quite different colours when viewed from different directions. This is understandable when the physics of light interaction with the water bodies is considered. Whereas most other objects in natural scenes exhibit Lambertian reflection, typical of matt surfaces, where the amount of light reflected by the surface is constant for all viewing directions, water bodies (and also many metallic roofs) exhibit specular reflection, where the reflected light is much brighter in some viewing directions than in others. Because water bodies are features that our method is seeking to locate, it was important that patch matching not exclude them by requiring uniformity of colour intensity. Other workers have handled colour differently, catering for overall differences in intensity level, possibly caused by differing levels of atmospheric pollution or different responses of the photo-detectors in the two views, and seeking to allow for this variation while requiring the relative colours in the two images be the same.

The characteristic of a patch that is of greatest assistance in performing the match is size. The patches are sorted by size, and the largest patches are considered first in seeking a match. There are usually very many small patches, and few large ones. By working first with the large patches, we anticipate that we will be concentrating on the easy patches before considering the more difficult ones, and find successful matches most rapidly.

Of course, the patches may not necessarily be the same size in the two images, owing to the two different viewing directions. We need to allow for the different geometry of the two views when comparing area or any other attribute of two patches. In the testing we have performed, we have used the geometry of two SPOT images, one an overhead view and the other at the maximum oblique angle, as the basis for comparing shapes. A scanline in nadir viewing covers 60km and in the oblique view 80km, so in comparing patch attributes, we have adopted the similarity test:

$$\frac{2}{3}P_2 \leq P_1 \leq \frac{4}{3}P_2,$$

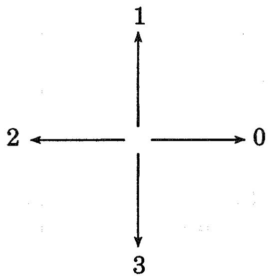
where P_1 and P_2 are an attribute of two patches being compared, one from each image.

This test could be greatly improved if we know the two viewing directions and we know in advance the extent to which the distance from the scene to the view points varies across the images.

We explored a variety of criteria for comparing patches (Abbasi, 1995), and concluded that several criteria were needed to effectively distinguish similar patches:

- area of the patch
- width of bounding rectangle
- height of bounding rectangle
- perimeter of the patch
- linearity of the patch
- concavity of the patch

The last two were developed from an analysis of chain codes for representing and comparing patches. Chain codes (Freeman, 1961) are usually a sequence of digits based on traversing the boundary of an area in a clockwise direction, with each digit representing the direction of an edge vector, such as:



Although chain codes have been understood for a long time, they have fallen into relative disuse because of difficulties

1. in establishing a starting point to enable two shapes to be compared, and
2. the difficulty in obtaining a similarity measure other than testing exact equivalence.

The first difficulty can be surmounted by using *shape number* (Bribiesca and Guzman, 1980), where the starting point is chosen to minimize the numerical value of the chain code. We were able to overcome the second difficulty by analyzing frequencies of chain code digits and frequencies of changes in chain code digits (Abbasi and Freeman, 1994c; Abbasi, 1995).

A chain code for representing a shape is represented as a set of digits:

$$c_1, c_2, c_3, \ldots c_m$$

where each c_i is a direction digit (0-3), and the differences between these codes is a related set of digits:

$$d_1, d_2, d_3, \ldots d_m$$

where $d_i = c_i - c_{i+1}$. The frequency of the digits representing the four cardinal directions in each of these codes will be independent of starting point, and will give a similarity measure that is useful even when shapes are not identical. For a particular direction digit k, these frequencies are defined as:

$$F_k = \sum_{j=1}^{m} \begin{cases} 1 & \text{if } c_j = k \\ 0 & \text{if } c_j \neq k \end{cases}$$

$$G_k = \sum_{j=1}^{m} \begin{cases} 1 & \text{if } d_j = k \\ 0 & \text{if } d_j \neq k \end{cases}$$

The perimeter m will then be:

$$m = G_0 + G_1 + G_3$$

(G_2 is always zero.)

The *linearity, L*, of a shape, or the proportion of its edge segments that are locally collinear, can be identified with G_0:

$$L = \frac{G_0}{m}$$

The *concavity, V*, of a shape, or the extent to which its exterior is indented, can be identified with G_1, assuming that the chain code traverses the shape in a clockwise direction:

$$V = \frac{G_1}{m}$$

F_0 will equal F_2, and F_1 will equal F_3. These are less useful in discriminating shape than the frequencies of chain code differences; nd the width and height of the bounding rectangle are very similar to them and are easier to calculate.

2.3 Multi-Patch process, using relative geometry

When there are many patches, we will not want to systematically compare all patches from one image with those from the other. Even when size and other criteria are taken into account, it should be possible to use information on the patches already matched to guide the selection from the available candidates. Some of these candidates will be impossible because of their relative geometry to those we know already to be matched. We call this method of accumulating relative geometry of patches and eliminating impossible candidates based on their position the *multi-patch process*.

Basically, it works in this way:

1. The patches for both images are held in a size sorted list, with the largest first.
2. Three large patches are chosen from the first image.
3. A search is conducted among the candidates for each of these three patches for a combination of candidates that form a triangle of similar size, shape and orientation to the patches in the first image.
4. If no suitable set of candidates can be found, one of the patches in step 2 is replaced, and step 3 is repeated. If necessary, more than one of the three original patches from the first image may need to be replaced in the search process.
5. After steps 3 and 4 have been repeated to establish suitable triangles, other triangles are formed. With the first two successfully matched patches, other patches are processed in turn from the first image, with patches of suitable size and shape being considered as in step 3 to find the one with the correct geometry. All patches are processed in this way using the two reference patches.

In the tests we have performed using this method, we have compared triangles by the length and orientation of edges. We have arbitrarily chosen ±5° for comparing edge orientation, and the distance criterion above for edge length. These criteria would be tightened if camera position and the nature of the scene were known more precisely.

2.4 Success of patch matching

As previously reported (Abbasi and Freeman, 1994b; Abbasi, 1995), the patch matching process has been tested on sections of SPOT images. The first tests were with images captured 45 months apart, on 400x400 sections each containing a major water body. Over one such sub-image pair, four patches were found, all correctly; in the second sub-image pair, eight patches were found with six being correct and the others being a neighbouring area of similar colour. The failures were due to the areas identified by the program as a patch not really being what we would regard as features. They were areas of vegetation of numerically uniform grayscale only slightly differing in colour from neighbouring areas, where the patch edge is non-existent to the eye, and imprecisely defined within the computer.

We tested a second SPOT image pair captured six weeks apart, using a 500x500 section that contained no water bodies. Thirty patches were found to match by our algorithm, but of these, only seventeen were correct. Again, a large number of the patches were in areas that we would not identify as a feature, areas where gray-level gradually changed, giving patches with visually indistinct edges. Although better results would have been expected from the shorter elapsed time between capture of the two images, the lack of distinct uniform areas other than fields with slightly varying colour caused the large number of errors.

Our present definition of what makes a patch is clearly inadequate. We have tried to improve it by requiring that the uniform area be surrounded by an edge, and using one of the many edge operators to locate the rapid change in gray-value. This has so far proved unworkable, because numerically defined edges are seldom continuous, and linking the segments possibly associated with a patch is not easy.

More work needs to be done in this area. An avenue worth exploring is to require some proportion of the patch boundary pixels also to be edge pixels, by some definition. Points of high interest as indicated by the Moravec or Förstner operators should be abundant on clearly visible patches. This should eliminate the within-field patches which are the main source of error at this stage of development.

3. From patches to points

The proof of the viability of a patch-based method for automating registration of a pair of images will be the success with which matching points can be found in the images, from the patches. Although we could consider the centre or centroid of the patch as such a point, any vagueness in the location of the patch boundary will translate into uncertainty in the position of the centre. A better approach will be to look for significant points along the boundary of the patch and systematically match them.

In the testing we have done, we have selected all boundary points as worth searching for. We select a point on the boundary of a patch in the left image, and search for its match in the right image, using a search window. This window is centred on a point in a roughly corresponding place, related to the centroids of the patches. We had to choose between using a least-squares iterative method (Ackermann, 1984; Gruen, 1985) and a correlation method to find the matching point (eg, Barnard and Thompson, 1980). We chose to use a correlation method, because we felt that the alternative least squares method might be too susceptible to failure in the initial stages when the geometry of the matches had not been determined with much precision, even though it would probably be faster and may give better accuracy.

Under our chosen method, the correlation between the selected point in the left image and all points in the related search area are calculated. The point in the search area with the highest correlation with the selected point is chosen, provided that this correlation is greater than some specified threshold. The match is subsequently refined to achieve sub-pixel accuracy.

We need to select the size of the search window, and the size of the area around the point upon which the correlation calculation is to be based. Although sizes of correlation areas ranging from 3x3 to 27x27 have been used by other workers, Shirai (1986) showed that large windows were generally suitable for obtaining global range information, but gave smooth changes in correlation with a broad minimum around the corresponding point, and consequent imprecision in the match. Small windows gave a sharper minimum at the corresponding point, but were more sensitive to noise.

We conducted our own tests, with the results being checked carefully by hand to find the success in each case:

Correlation area	Number of corresponding points found	Number correct
3x3	393	90
5x5	362	171
7x7	360	242
9x9	342	195

As can be seen the best results were achieved with the 7x7 correlation area, which we adopted.

The size of the search window will directly influence calculation time and success. We tested two sized areas, 11x11 and 23x23. We found the smaller area to be quite adequate, with no fewer correctly matching points, and with fewer ambiguities where more than one point had a high correlation.

3.1 Sub-pixel accuracy

The correlation method as just described should find a match with an accuracy of one pixel. In practice, the actual point will more likely be somewhere around this pixel, since the centre of each one of the corresponding pixels usually is not the image projection of the same point in the true object. To get to the point, a window of size 3x3 centred on the corresponding point is chosen. Then, a new coordinate value is calculated as a weighted average of the pixel's coordinates in the window, using as weight the correlation value for each pixel in the window:

$$x_{new} = \frac{\sum_{i=0}^{8} w_i\, x_i}{\sum_{i=0}^{8} w_i}$$

$$y_{new} = \frac{\sum_{i=0}^{8} w_i y_i}{\sum_{i=0}^{8} w_i}$$

where w_i is the correlation, and x_i and y_i are the coordinates of pixel i in the window. The new corresponding point with the coordinates of (x_{new}, y_{new}) should give sub-pixel accuracy.

4. Results

Tests have been performed on sections of SPOT images and on some close range images obtained from the Internet. Examples of the results can be seen in Figure 1 and in the following table.

Image set	A	B	C
Corresponding points found for correct patches	360	284	706
Corresponding points found from incorrect patches		0	164
Correct points	242	205	624
Overall success rate (fraction of identified points that are correct, %)	67	72	72
Success rate from correct patches (%)	67	72	89

As before, images A and B come from a single SPOT pair collected 45 months apart, and image C is from a SPOT pair captured six months apart. With the B images, in all cases where the patches were incorrectly matched, no points were identified by the search procedure as having sufficient correlation to match, so the incorrect patches did not reduce to the final success rate of the point matching procedure. With the C image pair, some of the points bordering the incorrect patches had a high enough correlation to register as matches in our point matching procedure. These unfortunately reduced the overall success rate. For our procedure to work completely unsupervised, we will need to overcome the patch matching errors for point matching to proceed.

In cases where a patch was correctly matched, there were often differences in the identified patch outline. Our search procedure was often successful in accommodating the patch outline error, and found correct point matches well away from the identified patch boundary. This can be seen on the right side of the patch in Figure 1b, and also in the lower left corner of the patch.

This gives us confidence that the method is sound in principle. When the method was applied to two pairs of close range photographs, success rates of 91% and 85% were achieved from patches, all of which were in the correct general area, but again were areas of gradual colour change rather than clearly identified features.

The present method of proceeding from patches to points attempts to find a match for every boundary pixel of the patches in the first image. In practice, this

Fig. 1(a) Section of left sub-image from C image pair

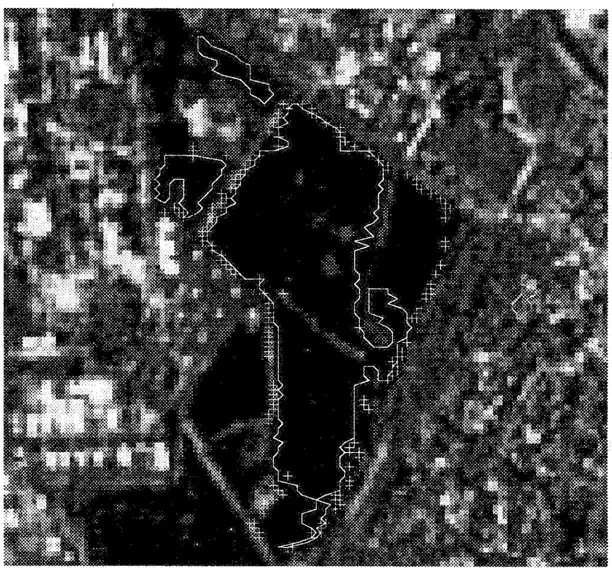


Fig. 1(b) Section of right sub-image from C image pair

Figure 1: Result of point matching with one patch, showing the success in finding matching points, despite an incorrect patch boundary having been identified. The image colours have been exaggerated by histogram equalization; the actual gray values lie in quite a narrow range, and the overall intensity of the two images is quite different.

will give too dense a set of matched points for most purposes. It also presents problems in areas where there is little colour variation along the boundary. A better approach will be to confine the search in some way, such as by identifying boundary points with high interest and finding their match, before proceeding with other boundary points or points away from the patch.

4. Conclusions

We have demonstrated the principle of operation of our proposed method and have had some encouraging success. Problems remain, mainly in patch identification. We will need to revise the definition of what makes up a patch in order to pick only those areas that clearly are features of the image, and not include areas where colour is slowly changing.

The point matching part of the procedure can be improved by first trying only significant points along the patch boundary. Then, we expect that limiting the search to epipolar lines will remove most of the remaining errors in point searching.

The method will be most useful in cases where camera position is known and manual location of ground control points in the images is not required. The method is intended to eliminate the need for seed points, and with further work, this goal seems feasible.

The work described in this paper has already been presented in more detail in a thesis by M. Abbasi-Dezfouli, submitted in mid-1995, which led to his being awarded a PhD (Abassi, 1995).

5. References

Abbasi-Dezfouli, M. (1995) "Patch-based matching method for stereo-images", PhD thesis, University College, University of New South Wales.

Abbasi-Dezfouli, M. and Freeman, T.G. (1994a) "Patch matching in stereo images based on shape", *Proc. 7th Australasian Remote Sensing Conf.*, Melbourne, 322-327.

Abbasi-Dezfouli, M. and Freeman, T.G. (1994b) "Patch matching in stereo-images based on shape", *Proc. of International Archives of Photogrammetry and Remote Sensing (ISPRS)*, Munich **30**, Commission III, 1-8.

Abbasi-Dezfouli, M. and Freeman, T.G. (1994c) *A modified chain code method for classification and comparison of shapes*, Computer Science Dept. Report CS23/94, University College, University of New South Wales.

Ackermann, F. (1984) "Digital image correlation: performance and potential application in photogrammetry", *Photogrammetric Record*, **64** (11), 429-439.

Barnard, S.T., and W.B. Thompson (1980) "Disparity analysis of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-2** (4), 333-340.

Bribiesca, E. and A. Guzman (1980) "How to describe pure form and how to measure differences in shape using shape numbers", *Pattern Recognition*, **12**, 101-112.

Day, T., and J-P. Muller (1989) "Digital elevation model production by stereomatching SPOT image-pairs: a comparison of algorithms", *Image and Vision Computing*, **7** (2), 95-101.

Förstner, W. (1986) "A feature based correspondence algorithm for image matching", *Int. Arch. for Photogrammetry and Remote Sensing*, **26** (3/3), 150-166.

Förstner, W., and E. Gulch (1987) "A fast operator for detection and precise location of distinct points, corners and circular features", *Proc. of ISPRS Workshop*, Interlaken, 281-305.

Freeman, H. (1961) "Boundary encoding and processing" in *Picture Processing and Psychopictorics*, B. S. Lipkin and A. Rosenfeld, Academic Press, New York, 1970, pp. 241-266.

Greenfeld, J.S. (1987) *A stereo vision approach to automatic stereo matching in photogrammetry*, Dept. of Geodetic Science and Surveying, Ohio State University.

Greenfeld, J.S. (1991) "An operator-based matching system", *P.E. & R.S.*, **57** (8), 1049-1055.

Greenfeld, J. S. and A. F. Schenk (1989) "Experiments with edge-based stereo matching", *P.E. & R.S.*, **55** (12), 1771-1777.

Gruen A.W. (1985) "Adaptive least squares correlation: a powerful image matching technique", *S. Afr. J. of Photogrammetry, Remote Sensing and Cartography*, **14** (3), 175-187.

Harallick, R.M. (1984) "Digital step edges from zero crossing of second directional derivatives", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6** (1), 58-68.

Kauffman, D.S., and S.A. Wood (1987) "Digital elevation model extraction from stereo satellite images", *Proc. IGRSS '87*, 349-352.

Marr, D. and E. Hildreth (1980) "Theory of edge detection", *Phil. Trans. Roy. Soc. London*, **B 207**, 187-217.

Moravec, H.P. (1977) "Towards automatic visual obstacle avoidance", *Proc. of 5th International Joint Conference on Artificial Intelligence*, 584.

Li, M. (1991) "Hierarchical multipoint matching", *P.E. & R.S.*, **57** (8),1039-1047.

Otto, G.P., and T.K.W. Chau (1989) "Region-growing algorithm for matching of terrain images", *Image and Vision Computing*, **7** (2), 83-94.

Rosenholm, D. (1987) "Multi-point matching using the least-squares technique for evaluation of three-dimensional models", *P.E. & R.S.*, **53** (6), 621-626.

Schenk, T., J. Li and C. Toth (1991) "Towards an autonomous system for orienting digital stereopairs", *P.E. & R.S.*, **57** (8), 1057-1067.

Shirai, Y. (1986) *Three-dimensional computer vision*, Springer-Verlag.

Trinder, J. C., B. E. Donnelly, P. Lartique and K. Becek (1990) "Mapping from SPOT analogue and digital images", *Proc. Fifth Australasian Remote Sensing Conference*, 1260-1267.

Trinder, J.C., and B.E. Donnelly (1993) "Tests on DEM software for SPOT images", *Proc. Advanced Remote Sensing Conference*, Sydney, 341-348.