Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition
Kauai, Hawaii, December 2001 · Volume 2 · Pages 331-336

Image Registration by Aligning Entropies

Clark F. Olson
University of Washington, Bothell
Computing and Software Systems
18115 Campus Way NE, Box 358534, Bothell, WA 98011-8246

Abstract

Maximization of mutual information is a powerful method for registering images (and other data) captured with different sensors or under varying conditions, since the technique is robust to variations in the image formation process. On the other hand, the high level of robustness allows false positives when matching over a large search space and also makes it difficult to formulate an efficient search strategy for this case. We describe techniques to overcome these problems by aligning image entropies, which are robust to illumination variation and can be applied to multi-sensor registration. This results in a lower rate of false positives and a more efficient method to search an image for the matching position. The techniques are applied to real imagery and compared to methods based on mutual information and gradients to demonstrate their effectiveness.

1 Introduction

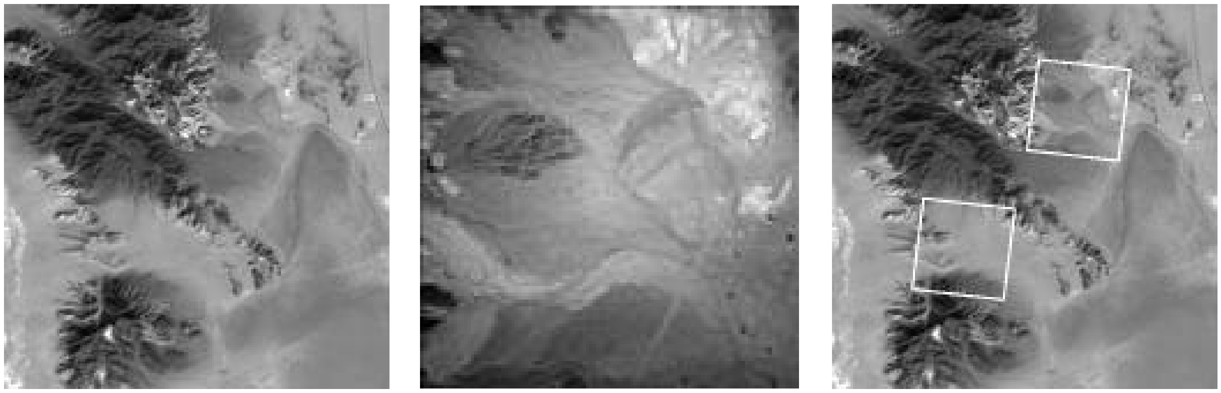
Maximization of mutual information [3, 9] is a technique commonly used for image registration when the images are from different modalities, such as in medical imaging. This technique is very robust to the changes caused through the use of varying sensors, allowing registration even when the images have differing appearances, as long the intensities of overlapping pixels between the images are statistically correlated when the images are registered.

In some cases, this high level of robustness can be a drawback, rather than a feature. When the search space is large, allowing such matches makes false positives more likely than when additional constraints can be placed on the relationship between the pixel intensities. An example of such a situation can be seen in Fig. 1. In this example, we wish to locate the position of an aerial terrain image in an orbital image encompassing the same location. When mutual information is used, an incorrect location is selected as the best match, even when the search is restricted to translation only.

There are two additional areas where the use of mutual information for image registration is less than ideal. First, mutual information is not able to match images that have smooth shading changes due to differences in illumination when only one reference image is used. Matching can be performed in this case with two reference images [9]. Second, matching using mutual information is computationally expensive. Most implementations sample the image data in order to reduce the computation time and use iterative optimization techniques that can fail if a good starting location is not known [3, 5, 8, 9].

In this paper, we describe an alternative to mutual information that improves upon these areas. The basic idea of our method is to align the entropy at each location in a template image to the entropy at corresponding locations in a reference image according to some transformation of the template. An efficient search strategy has been developed combining fast search over translations using the FFT, coarse search over the remaining parameters (similarity or affine), and refinement using iterative optimization. Multi-resolution techniques can be used for the coarse search for additional speed, but are not necessary in many cases. We have applied this technique to several applications, including registration of overhead and forward-looking terrain images captured with different sensors.

Our method is related to previous work based on matching using image gradients (see, for example, [2]). We use image entropy, since it retains more of the information content of the image than gradients, and has a lower rate of errors in practice. Our search strategy is not limited to matching entropy images; it can be used to determine the position that maximizes the normalized correlation between any template and reference image. A trivial modification allows matching between vector-valued images, such as entropies or



(a) (b) (c)

Figure 1: Motivating example. (a) Orbital image of the Avawatz Mountains and Silurian Valley in California. (b) Aerial image showing a detail of (a). (c) Annotated image. The upper right box shows the location of (b) in the orbital image. The lower left box shows the best matching position found using mutual information.

gradients at multiple resolutions.

2 Entropy alignment

For a discrete random variable A, with marginal probability distribution $p_A(a)$, the entropy is defined as:

$$H(A) = -\sum_a p_A(a) \log p_A(a). \tag{1}$$

Note that $0 \cdot \log 0$ is taken to be zero, since:

$$\lim_{x \to 0} x \log x = 0. \tag{2}$$

We apply an entropy transformation to both the template and the reference image as follows. For each image location (x, y), we examine the intensity values in an image window (with some specified size $k \times k$) centered (x, y). The intensities are histogrammed and the entropy for the window is computed according to Eq. (1) above. In practice, it is useful to smooth the histograms prior to using Eq. (1). For efficiency, we use a histogram with 64 bins (6 bits of information) and smooth the histogram using a Gaussian ($\sigma = 1.0$ bins).

Figure 2 shows an example of the entropy images generated at a variety of scales for the image in Fig. 1. It can be observed that the entropy transformation captures the amount of local variation at each location in the image. As the size of the window increases, the local entropies are spread and smoothed over a larger area. This property is useful in the generation of a multi-resolution search strategy. While these images are not invariant to all illumination changes and sensor characteristics, they are insensitive to many issues in multi-sensor registration while retaining much image information. The entropy images are invariant to bias in the original images and the correlation peaks are invariant to gain in the entropy images. Of course, illumination changes and different sensors cause more complex transformations than these, but we have found that the locations of the correlation peaks change relatively little.

Now, we want to maximize the normalized correlation between the entropy images over the search space. In general, this involves rotating and rescaling the template entropy image according to the pose parameters. The shape and scale of the template will, thus, vary over the search space and this must be accounted for in the normalization. In order to deal with this in the search for the best position, whenever we generate a rescaled template, we generate another template of the same shape storing the weight of each position in the correlation. Most positions will have a weight of one, but the edges of the template will have lower values, since bilinear interpolation does not have four pixels to interpolate from in this case. This template can be used in the Fast Fourier Transform (FFT) to quickly determine the normalization values for each translation of the template in the search strategy described below.

3 Efficient search

In determining the best match between the template image and the reference image, there are a number of strategies that we could use. Most previous

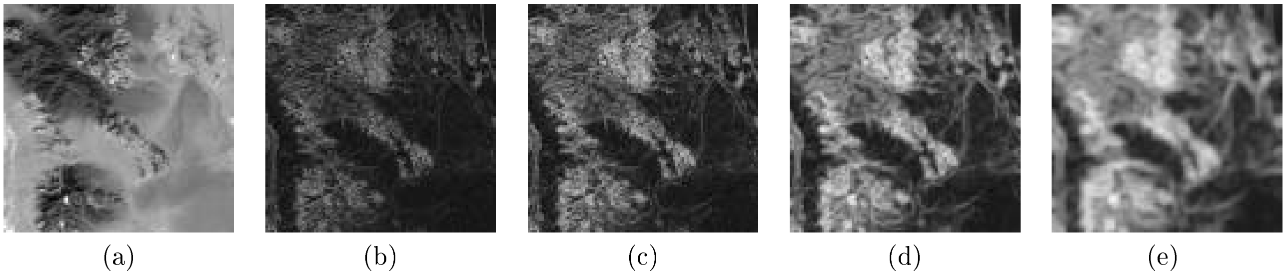


Figure 2: Entropy images at different scales. (a) Original image. (b) Entropy image for 3×3 window. (c) Entropy image for 5×5 window. (d) Entropy image for 11×11 window. (e) Entropy image for 21×21 window.

techniques for similar problems using mutual information have used iterative search techniques that require an initial estimate of the template position to converge to the global minimum. We are interested in methods that can determine matches in a large search space without an initial pose estimate. This is a problem that arises in many situations. For example, we are interested in the application of these techniques to matching imagery from the descent of a spacecraft onto a planetary surface to orbital imagery encompassing the landing location. For this case, we may have initial estimates of the spacecraft orientation from inertial sensors. However, the precise location for landing on a planet such as Mars is not known, thus requiring a search for the correct landing location.

We note that the efficient guaranteed search method of Rucklidge [7] cannot be used for this problem, since the normalization value depends upon the location of the template with respect to the reference image. It is possible to place a bound on this value. However, in our experiments with a divide-and-prune search strategy similar to the strategy used by Rucklidge [7] and in our own work [4], we have found that an insufficient amount of pruning can be performed to decrease the computation time of the search.

Of course, a brute force search is possible, but the search space has six degrees of freedom for affine matching of planar surfaces. A brute force search would, thus, require much time. A technique that is very useful for performing correlation operations when the search space is restricted to translations is the FFT, since it is well known that cross-correlation can be performed efficiently in the frequency domain.

For a more complex search space (such as similarity or affine transformations), we must have a mechanism for examining the additional parameters. In this case, we sample the additional parameters coarsely and consider each sample point separately. (The following section examines how coarse the sampling can be and still achieve good results.) For each sample point, we use the FFT method to locate peaks in the entropy alignment. A peak is then considered further if the score is at least K standard deviations above the average value for that sample point. K may be chosen arbitrarily, but we have found that values ranging from 2.5 to 3.0 work well. When a peak is selected for further consideration, we use Powell's iterative optimization method [6] to converge to the locally optimal solution in the full, continuous search space. We then select the best such solution found among those tested, or all solutions that surpass a pre-determined threshold.

This procedure results in a fast search strategy since we are able to sample relatively few points from the non-translational parameters, each of which can be examined quickly using the FFT. At present, the most time consuming portion of the search is the iterative optimization of each peak selected. We believe that this can be improved considerably through the use of more advanced techniques.

4 Capture range

It is important in the search strategy described above to have a good estimate of the capture range of the method. We want to be able to sample the space sparsely and still find the correct position. Thus, we must determine how far a sample can be from the correct position and still find a large enough peak to examine with the iterative method and, furthermore, the iterative method must converge to the correct location. In this section, we study these issues for several sample problems.

We tested six sample problems using real imagery. The problems can be seen in Fig. 1 and Fig. 4-8. The problems using overhead imagery (Fig. 1 and Fig. 6) used a six-dimensional affine transformation space:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \quad (3)$$

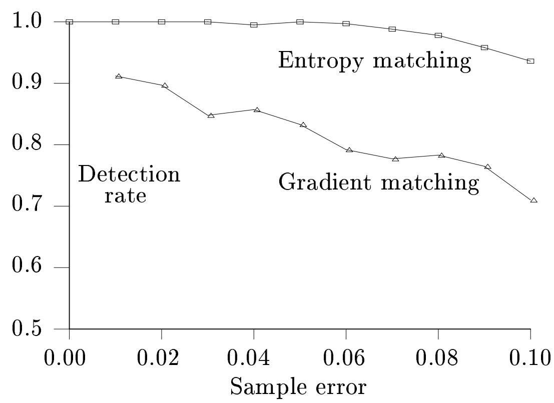


Figure 3: Experiment determining effective capture range. The upper plot is for entropy matching. The lower plot is for gradient matching.

The other four problems used a four-dimensional similarity transformation space. This case was treated the same as the affine case, except that we maintained the following constraints:

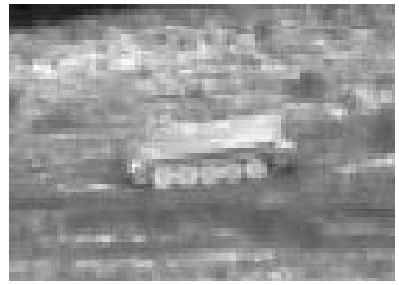
$$a = d \tag{4}$$

and

$$b = -c. \tag{5}$$

For each of the sample problems, we tested initial sample points that were in error according to some bounded amount for each of the scaling parameters: a, b, c, d. We then determined whether the highest scoring location using our search strategy was the correct position of the template. Figure 3 shows the results as a function of the bounded error. The figure also shows the results of using the same techniques, but applied to gradient images rather than entropy images. Each data point represents 100 trials for each of the test problems (600 total). It can be observed that nearly perfect results are obtained up to an error of 0.05 in the parameters. This represents a 5% scale change in each parameter. We are thus able to obtain excellent results by sampling the parameters at every 0.10 interval, since this will result in a sample point within 0.05 in each dimension of every point in the space.

Note that the performance of the technique is considerably worse when gradients are used rather than entropies. While matching with gradient images is also robust to variations in image formation, the gradient images do not capture as much information about the original images. For this reason, the best match found is less likely to be correct. The entropy-based method performed better than the gradient-based method for each of the test problems.



(a)

(b)

Figure 4: Registration between FLIR and CCD camera images. (a) FLIR image of a military vehicle. (b) Registered location in a CCD image.

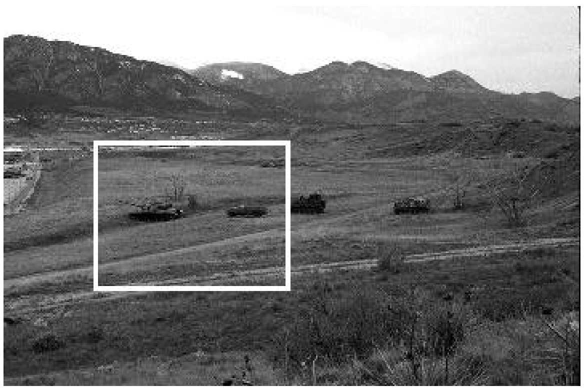
5 Results

Figures 4-8 show examples of the application of our method to real images. The correct location was determined in each example using the techniques described above.

Figure 4 shows FLIR and CCD images of a military vehicle taken from close-by locations [1]. The template was generated by cropping the FLIR image to contain the vehicle and some nearby terrain. Matching was then performed against the CCD image using similarity transformations in 22.5 seconds on a 333 MHz Sun UltraSPARC, with a 138×100 pixel template and a 720×480 pixel reference image. Note that the entropy matching technique succeeds in detecting the correct location, despite the camouflage on the vehicle that is not visible in the infrared image and the highly different appearance of the vehicle in the two images. The experiment shown in Fig. 5 is similar in nature to that of Fig. 4.

In Fig. 6, the template is an image captured using

(a)



(b)

Figure 5: Registration between FLIR and CCD camera images. (a) FLIR image of military vehicles. (b) Registered location in a CCD image.

(a)

(b)

Figure 6: Registration between aerial and orbital images. (a) Aerial image of California desert. (b) Registered location in an orbital image.

a helicopter at an elevation of approximately 800 meters and the reference image is an orbital image that encompasses the same location. The correct location of the template is difficult to detect manually, unless prior knowledge of the area is used. Successful matching was performed with the affine transformation and the correct location was determined in 13.8 seconds with a 112×112 pixel template and a 500×500 pixel reference image. Figure 1 shows a similar example.

The final problem domain (Fig. 7 and Fig. 8) uses images from Mars. Both the template and the reference image were captured with the Imager for Mars Pathfinder (IMP) cameras. However, they were captured with different lens filters and at different times during the day, so the illumination and the shadowing of the terrain is different between the images. Furthermore, the reference image has undergone a non-linear warping operation to remove lens distortion, while the template has not. The matching algorithm was able to succeed despite these differences and required 54.14 seconds with a 164×127 pixel template and a 776×400 pixel reference image for the problem in Fig 7.

6 Summary

We have described a new method for performing registration between images captured with different sensors or with different illumination. This method is robust and is able to succeed in cases where matching with mutual information and gradient alignment fail. The basic idea is to transform the images into a representation that contains the image entropy at each pixel location. The best registration is determined using normalized correlation. In order to perform this operation efficiently over a large search space, such as similarity or affine transformations, we have developed a search strategy that combines search over translations using the Fast Fourier Transform, coarse search over the remaining parameters by sampling the search space, and refinement using iterative optimization. Experiments examining the capture range of the method were performed to determine how sparsely to

(a)

(b)

Figure 7: Registration between images taken with different filters and with different illumination. (a) CCD image of a Martian rock. (b) Registered location of rock in a CCD image with a different lens filter.

sample the search space. The resulting method is able to quickly and robustly register images from different sensors.

Acknowledgements

Two of our experiments used data from a publicly released data set collected at Fort Carson [1]. The research described in this paper was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] J. R. Beveridge, D. P. Panda, and T. Yachik. November 1993 Fort Carson RSTA data collection final report. Technical Report CS-94-118, Colorado State University, 1994.

[2] L. M. G. Brown and T. E. Boult. Registration of planar film radiographs with computed tomography. In *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 42–51, 1996.

[3] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, April 1997.

(a)

(b)

Figure 8: Registration between images taken with different filters and with different illumination. (a) CCD image of a Martian rock. (b) Registered location of rock in a CCD image with a different lens filter.

[4] C. F. Olson. Maximum-likelihood template matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 52–57, 2000.

[5] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual information matching in multiresolution contexts. *Image and Vision Computing*, 19:45–52, 2001.

[6] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Plannery. *Numerical Recipes in C.* Cambridge University Press, 1988.

[7] W. J. Rucklidge. Efficient guaranteed search for gray-level patterns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 717–723, 1997.

[8] P. Thévenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099, December 2000.

[9] P. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.