An Empirical Study of Some Feature Matching Strategies

Étienne Vincent and Robert Laganière

School of Information Technology and Engineering
University of Ottawa, Ottawa, Canada, K1N 6N5
evincent,laganier@site.uottawa.ca

Abstract

This paper proposes an empirical evaluation of different matching strategies that have been proposed in the literature to solve the problem of feature point correspondence between images. They will be evaluated in terms of their ability to reduce the number of false matches in given match sets, while preserving the good matches. The validation process determines the number of good matches and the proportion of good matches in a given match set, and this for the different parameter values of a matching constraint.

1 Introduction

Recently, great advances were made in establishing correspondence between views generated by uncalibrated systems of cameras. Most matching schemes share a common structure [11, 15, 16]. They apply correlation between automatically detected feature points to obtain a set of candidate matches. Then, a robust estimation method is used to find the epipolar or trifocal geometry of the camera system. This estimated geometry can then be used to reject some incompatible candidate matches, and to guide the search for more matching points.

The efficiency and accuracy of this scheme depends greatly on the quality of the candidate match set initially obtained. Indeed, robust estimators require candidate match input sets with many correct matches to find an accurate solution, and with few mismatches to perform efficiently. Thus, candidate match sets should be filtered before camera system geometry estimation. This is generally done by introducing basic constraints that aim to eliminate mismatches. These additional constraints are basic in the sense that, at the stage where they are applied, the epipolar or trifocal geometry of the camera system geometry is not yet known. Thus, neither image rectification or guided matching are possible at this point.

In this paper, we propose to empirically compare and validate the effectiveness of different matching strategies. They will be evaluated in terms of their ability to reduce the number of false matches in given match sets, while preserving the good ones. The match sets obtained by these matching strategies are intended to serve as input to robust estimators of the epipolar geometry which can thereafter be used in further improving the sets. It is important to note that the constraints and strategies studied here would not be sufficient, by themselves, to find match sets of sufficient quality. Instead, they should be used within more elaborate matching schemes. The objective of the present study is therefore to validate the constraints used inside matching algorithms, not to study these algorithms as a whole. These algorithms have been surveyed and empirically compared in several other works [3, 5, 9, 12].

Many authors use iterative processes in the steps preceding robust estimation. Relaxation is such an iterative process [16]. In this case, an energy function, corresponding to some aggregate value of a constraint applied to the pairs in a candidate set, is iteratively minimized. Testing the same constraint outside of such an iterative scheme represents a good measure of its effectiveness. This is why we have chosen to limit the scope of this study to the direct application of constraints.

The next section describes our scheme for evaluating matching methods. Then, section 3 studies the role of feature point detection in matching. Section 4 looks at the way in which correlation is applied. Section 5 is concerned with matching constraints that require corresponding features to have similar properties. Section 6 looks at matching constraints that require matches to have similar disparities as their neighbors. Finally, section 7 justifies the use of some matching constraints when the goal is fundamental matrix estimation.

2 Validating Point Correspondences

The feature point matching problem consists in finding pairs, among many candidate feature points, that correspond to the same scene element. To evaluate and compare matching strategies, we will use image pairs on which all possible good matches were identified, among fixed sets of detected feature points. Results of various matching schemes will then be compared against these exact solutions.

The image pairs shown in Fig. 1 were selected. The pairs have varying levels of change in the translation, rotation, zoom, and illumination between their images. Feature points have been detected on each image (see section 3), and all correct matches between these points were determined. To determine these matches, all possible pairs must be considered, a laborious task if it had to be done entirely manually.

Fortunately, to build this *ground truth* set, many pairs can be automatically discarded. First, by visual inspection, the horizontal and vertical disparity ranges of each image pair can be determined. All matches having a disparity outside these ranges can be rejected. Secondly, the matches that do not agree with the epipolar geometry of the image pair can be automatically eliminated. To this end, the epipolar geometry of each image pair was estimated using the method described in [11]. Following this pruning, we are left with a smaller set of image point pairs from which all the good matches can be extracted manually in reasonable time[1].

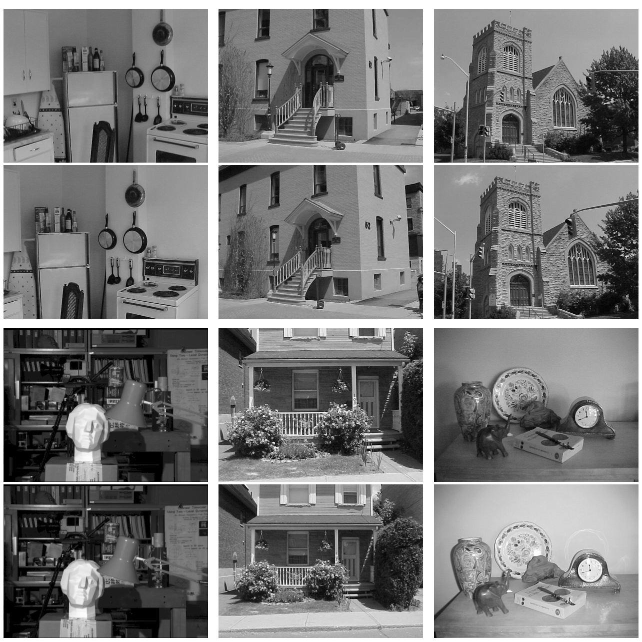


Figure 1: The six test image pairs with feature points extracted (approximatively 500 per image), from top to bottom: *Kitchen, Building, Church, Lab, House, Objects*.

Having identified the set of all possible good matches between detected points in an image pair, it becomes possible to evaluate the effectiveness of different strategies used for matching: a matching constraint is considered useful if it filters out many mismatches found in an input matching set, while preserving most good matches.

A given method will use different parameters or thresholds towards accepting or rejecting a given candidate match. There will usually be a tradeoff in the selection of these parameters. In order to appreciate the effectiveness of an approach, results will be shown on a graph showing the number of good matches in the resulting match set (on the Y-axis) versus the proportion of good matches in that set (on the X-axis). For each image a curve is generated representing results obtained for different values of a control parameter associated with the method under study. In such a graph, a perfect method would be one producing a horizontal line, i.e. all points eliminated are false matches. Conversely, a useless technique would be one that produces a vertical line, i.e. one that randomly eliminates points, thus keeping the good match proportion constant. Note that, in practice, we might expect that an effective method would produce a nearly horizontal curve until some point where the curve will start to drop vertically, when more severe thresholding cannot further improve the quality of the match set.

3 Corners as Points of Interest

The choice of feature point detector has a definite impact on the results produced by a matching scheme. Among the most popular feature point detectors is the Harris detector[6]. The Intel OpenCV library[2] proposes an implementation of this operator.

When feature points are detected for the purpose of matching, the key property of the detector is repeatability: in different views of the same scene, the detector should extract the same points, despite the variations due to a change in perspective or lighting conditions. The Harris detector is considered the most stable, with regards to this property [13]. To confirm this, we compared it to the SUSAN corner detector [14], another well-known detector which looks at the characteristics of the area inside small windows that have similar brightness to their center points.

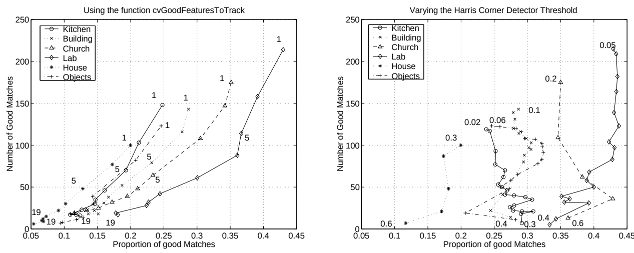
These corner detectors were used to extract around 500 points in the image pairs of Fig. 1. Then, all good matches among these corners were found using the technique described in Section 2. From this, the number of scene corners that have been correctly detected in both images (the repeatability of the corner detector), was obtained. Comparing these numbers between Harris and SUSAN results in a repeatability approximatively 3 times superior for Harris.

The OpenCV library includes a second function[3] for corner detection, which filters the corners found by the first one. It ensures that the detected corners are far enough from one another. This is done by iteratively finding the strongest feature point, and throwing out all other feature points that are

[1] These images, the detected feature points and the correct match set are available at www.site.uottawa.ca/research/viva/projects/imagepairs/

[2] freely available at developer.intel.com

[3] This second function is called cvGoodfeatureToTrack.



(a) (b)

Figure 2: (a) Eliminating corners close to stronger ones. The numbers shown represent the minimal acceptable distance between corners. (b) Modifying the threshold value of the Harris corner detector.

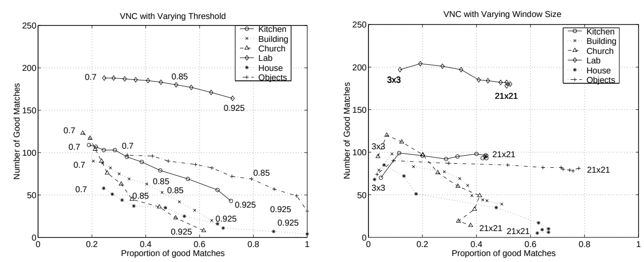
closer than a threshold distance from it. In order to determine if this method brings an increase in the quality of the candidate match set, the corners detected in both images of each pair were again counted, for different distance thresholds. The resulting graph, shown in Fig. 2 (a), demonstrates that this 'cleaning' of the corner set significantly worsens the set of candidate matches. Thus, this function should not be used when corners are detected for the purpose of matching between different views.

The OpenCV-Harris operator will therefore be the one used in our experiments. However, it should be noted that, when using it, it's control parameter directly influences the number of corners detected. To determine the effect of modifying this threshold on the quality of the candidate set, corners were detected using different thresholds on the test image pairs, and the number of corners detected in both images was determined. Results are shown in Fig. 2 (b). This graph shows that, within a reasonable range, the proportion of detected corners remain relatively constant. It therefore follows that one can increase the number of good matches just by accepting more corners. However, this is done at the price of a proportional increase of the total number of corners to analyze. The corner detector's threshold should therefore be set so that the number of matches found is suitable, but not much greater than the amount needed for the considered application.

4 Correlation

Correlation is the basic mean by which interest points on different images are matched. Variance normalized correlation (VNC) is a commonly used correlation function. It offers the advantage of producing stable and reliable results over a wide range of viewing conditions. The fact that VNC scores are normalized, is an advantage over other correlation functions, as it makes the choice of a threshold much easier.

Two basic parameters influence the performance of the correlation: the size of the window (the neighborhood) used to correlate point pairs, and the threshold value on which the decision to accept or reject a match is based. The results shown in Fig. 3(a) and (b), where VNC is applied to the image pairs of Fig. 1, illustrate how these parameters affect the quality of the resulting match set. As expected, tightening the threshold increases the proportion of good matches, but at the same time, decreases the number of good matches quite rapidly. The experiment also shows that increasing the size of the window is an effective mean to identify and reject more false matches (an observation also made in [10]), but this is only true up to a certain size (until about 11×11, in our experiments).



(a) (b)

Figure 3: Correlating corners of Fig. 1 using VNC. (a) 0.7 to 0.925 threshold, on 11×11 windows. (b) window size between 3×3 and 21×21, with 0.8 threshold.

4.1 Unicity

So far, all pairs having a correlation score above some threshold value were considered. Thus, a feature point could be matched with several others. Imposing unicity means that for each feature point in one image, only its strongest match in the other image is considered. A generalization of unicity was studied, where the n strongest matches are kept.

Fig. 4(a) shows the results of applying VNC to our image pairs while imposing unicity of different orders. Unicity proved beneficial, as it rejected many mismatches. This important improvement obtained in the proportion of good matches is at the expense of a fairly small loss in the absolute number of good matches. The resulting large improvement in the quality of the match sets should justify the general use of unicity of order 2 or 1.

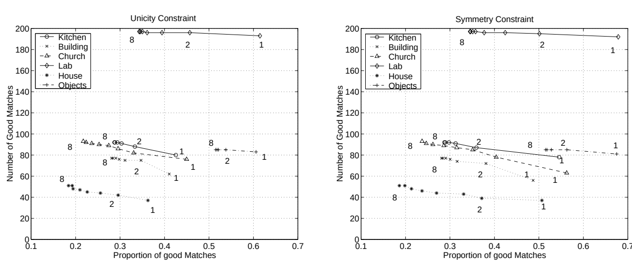
4.2 Symmetry

When unicity is imposed, VNC becomes asymmetric, a situation which is physically impossible. Thus, a right image point, which gives the highest correlation score, when paired with a certain left image point, can itself be paired with a different left image point with a higher score.

Imposing symmetry means keeping only pairs in which each point is the other's strongest match [4]. This increases the chances that the two points in the matched pairs correspond to projections of the same physical scene point. Fig. 4(b) shows the results of the same experiment as in 4(a), but where the symmetry constraint was applied in addition to unicity. It shows that imposing symmetry is clearly advantageous as it eliminates many mismatches while affecting only few good ones. Note that symmetry is generally imposed after unicity even if, in fact, symmetry can hold for matches that violates a unicity of a given order. This is often the case for scenes where several occlusion boundaries can be found; the *church* image pair is such an example where more than 20 good matches are lost when symmetry is applied after unicity of order 1 rather than order 8.

Nevertheless, imposing unicity and symmetry constitutes a very effective way of improving the quality of a match set. The fact that these constraints have a relatively low computational costs reinforces this statement.

For this reason, the experiments presented in the remainder of this work will use match sets obtained from VNC on 9×9 windows and using a threshold of 0.8 with first order unicity and symmetry applied. Table 1 summarizes the characteristics of the resulting match sets. The question now is to determine how the quality of these match sets can be further improved.



(a) (b)

Figure 4: (a) Imposing the unicity constraint with varying order while applying VNC to the corners of Fig. 1. (b) Imposing symmetry on the sets obtained in (a). VNC is applied to 9×9 windows with a threshold of 0.8.

5 Using Feature Point Properties

Many properties can be used to describe image points. Some should remain relatively unchanged in different views, and hence could be used to constrain matching. However, many properties such as corner strength and orientation, which are useful in describing corners, will not improve candidate match sets, as they are indirectly accounted for during correlation.

Image pair	Proportion of good matches	Number of good matches
Kitchen	54.5%	78
Building	48.7%	56
Church	56.2%	63
Lab	67.8%	192
House	50.7%	37
Objects	67.5%	81

Table 1: Characteristics of match sets obtained from VNC correlation using 9×9 windows, and thresholds of 0.8, with unicity of order 1 and symmetry.

5.1 Corner Shape Similarity

One possible strategy is to require that the corners in a pair have similar shapes. A corner shape is defined here as a small area around the feature point, belonging to the same scene object as this feature point. A method to extract the corner from its background is therefore required. Two such simple methods were investigated.

The first method uses univalue segment assimilating nuclei (USANs), as described in [14]. The idea is to extract the portion of feature point neighborhoods that is of similar intensity values. A USAN is then assumed to belong to the same scene object as the feature point.

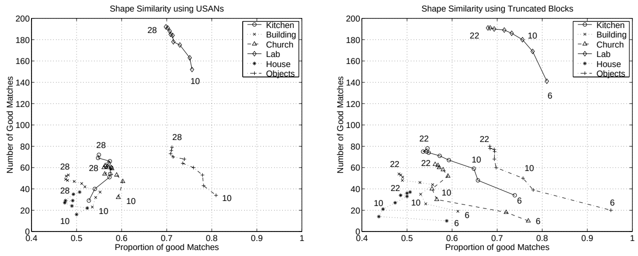
The other method is inspired from rudimentary block truncation coding [2]. The correlation window is separated into two regions according to the window's average intensity value. The foreground consists of the pixels within the same region as the feature point.

Once corner shapes have been extracted, the Hamming distance between the obtained binary foreground/background maps is computed and pairs for which this distance is above some threshold are eliminated. Results are shown in Fig. 5(a) and (b). While foreground extraction using USAN does not seem to be very effective, results based on truncated blocks show a certain improvement in the proportion of good matches, but the corresponding reduction in the number of good matches might appear excessive for some applications.

5.2 Eliminating the Background

Some correlation functions, such as [1], attempt to consider only the scene objects on which the feature points lie, to establish correspondence. The foreground extraction methods of subsection 5.1 can be used to determine the region to which correlation should be restricted.

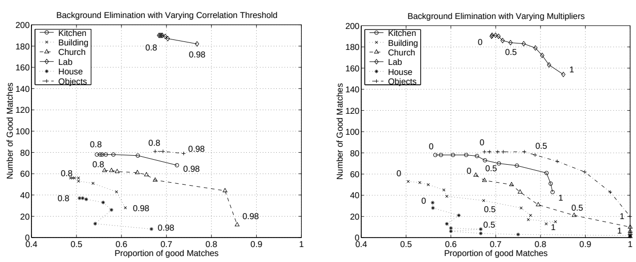
Results obtained when using this kind of selective correlation (where VNC is applied directly to the foregrounds of corners, but is weighted by a multiplying factor when applied to the background) are shown in Fig. 6(a) and (b). These graphs show that performing a correlation on the fore-



(a) (b)

Figure 5: Applying the shape similarity constraint on 9×9 windows, for different choices of threshold. (a) using USANs. (b) using truncated blocks.

grounds gives better results than using the simple shape similarity criterium of the previous subsection. Higher proportions of good matches can be achieved with higher multiplying factors, but at the cost of reducing their number.



(a) (b)

Figure 6: Applying the background elimination constraints on 21×21 windows. (a) with a varying threshold, and a multiplier of 0 for the background. (b) with a varying multiplier for the background, and a correlation threshold of 0.92.

6 Enforcing Disparity Consistency

It is reasonable to assume that, in most cases, the disparity of a match should be similar to the ones of its neighbors. Hence, constraints could be established that ensure that matches behave as their neighbors.

6.1 Confidence Measure

Based on the principle that each point of a match pair should have a neighborhood with similar properties, a confidence measure was proposed in [16]. It is defined for a pair of points and uses the feature points belonging to their neighborhoods. All candidate matches, found in the neighborhood, having a relative position similar to the pair being considered, are counted by the measure.

Fig. 7 shows the results of constraining the confidence measure with $\varepsilon_r = 10$ and 61×61 neighborhoods. A drawback of this measure is that it cannot be estimated if a point does not have close neighbors in the candidate set. Also, the method was found difficult to tune because several parameters have to be adjusted. Although this constraint achieves positive results on some of the image pairs, it is seen that it does not perform as well as the other constraints of this section.

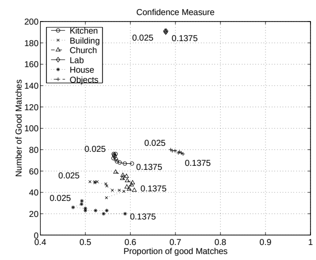


Figure 7: Constraints on the confidence measure, with varying threshold.

6.2 Disparity Gradient

The disparity gradient is a measure of the compatibility of two pairs [8]. If two pairs (m, m') and (n, n'), have disparities $d(m, m')$ and $d(n, n')$ respectively, then their cyclopean separation, $d_{cs}(m, m'; n, n')$, can be defined as the vector joining the midpoints of the line segments $\overline{mm'}$ and $\overline{nn'}$, and, their disparity gradient is defined as:

$$\Delta d(m, m'; n, n') = \frac{|d(m, m') - d(n, n')|}{|d_{cs}(m, m'; n, n')|} \qquad (1)$$

Compatibility measures, such as the disparity gradient can be used in a iterative process, as in [11], where incompatible matches are iteratively removed until all pairs have a similar disparity gradient. Here, the disparity gradient as well as the constraints of the next subsection are used in a new way, in a local constraint that enforces that a match's disparity be similar to those of it's closest neighbors.

This measure was used in a constraint that accepts pairs that share a disparity gradient below some threshold value with at least 2 of its 5 closest neighbors. Fig. 8 (a) shows the results of applying this constraint with varying threshold, and demonstrates that it can eliminate a significant number of outliers while eliminating few good matches.

Fig. 8 (b) shows the effect of a change in the proportion of the neighbors that must be compatible with a match, in order for it to be considered valid. The constraint was applied with a 0.4 threshold, but where 2 out of n neighbors must be compatible, for different ns. This shows that this constraint is most effective when applied to the 3 to 5 closest neighbors.

6.3 Relative Positions of the Neighbors

In a similar way to what may be done using disparity gradients, the relative position of two pairs can be constrained. For good matches, the vectors $\overrightarrow{mm'}$ and $\overrightarrow{nn'}$ would be similar. Thus, constraints were used, which require these vectors to have have similar direction and magnitude disparities as at least 2 of their 5 closest neighbors.

Fig. 8 (c) shows results of applying the constraint on angles, and (d) shows the constraint on magnitudes. As long as the image pair's baseline is relatively small, these simple constraints give similar results as the constraint on the disparity gradient. However, it is more difficult to select a good thresholds on the disparity angle and magnitude.

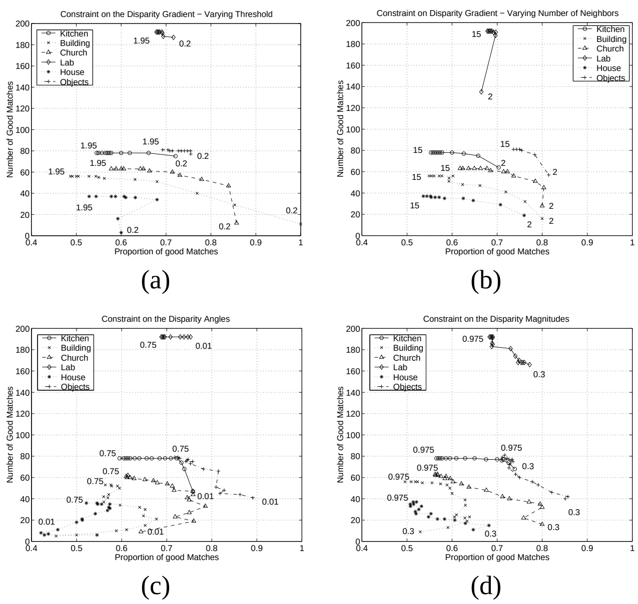


Figure 8: Constraints on disparity. (a) Constraint on disparity gradients with varying threshold. (b) with varying number of neighbors. (c) Constraint on disparity angles, with varying threshold. (d) Constraint on disparity magnitudes, with varying threshold.

7 Estimating the Epipolar Geometry

The matching strategies presented in this work aim at improving match sets by filtering out bad matches. The legitimacy of this objective will now be demonstrated by showing how a good initial match set can greatly improve the efficiency of fundamental matrix estimation (step 3, mentioned in Section 1).

The fundamental matrix is usually found using a RANSAC scheme, in which random selections of 8 matches are iteratively considered [7]. A fundamental matrix is computed for each selection, and its accuracy is assessed by

Match set	N	P	E
Kitchen, VNC 0.8	480	20.0%	1 170 207
Kitchen, VNC 0.9	129	45.7%	1 564
VNC 0.8 + constraints	117	59.8%	181
Building, VNC 0.8	497	16.7%	4 951 418
Building, VNC 0.9	133	36.1%	10 407
VNC 0.8 + constraints	72	65.3%	90
Church, VNC 0.8	911	12.3%	55 414 078
Church, VNC 0.9	220	19.1%	1 697 813
VNC 0.8 + constraints	55	76.8%	24
Lab, VNC 0.8	734	27.4%	93 705
Lab, VNC 0.9	363	53.7%	431
VNC 0.8 + constraints	272	71.7%	42
House, VNC 0.8	545	13.2%	33 246 450
House, VNC 0.9	159	22.6%	433 770
VNC 0.8 + constraints	36	63.9%	107
Objects, VNC 0.8	292	29.8%	48 240
Objects, VNC 0.9	62	90.3%	6
VNC 0.8 + constraints	100	80.0%	17

Table 2: Characteristics of different match sets and the theoretical expectation of the number of iterations required to find the exact fundamental matrix using RANSAC.

considering the cardinality of the subset of the candidate matches that support it. After a sufficient number of random selection, it is expected that an accurate estimate of the fundamental matrix will be uncovered. The number of iterations required is basically a function of the proportion, p, of good matches in the considered set. The number of iterations, n, needed to obtain a correct fundamental matrix with 95% probability is expressed theoretically as

$$n = \frac{\log(0.05)}{\log(1 - p^8)} \qquad (2)$$

Table 7 presents the characteristics of different match sets which could be used for fundamental matrix estimation, for the 6 image pairs of Fig. 1. The first two lines of each row correspond to the sets obtain using only VNC, with thresholds of 0.8 and 0.9 respectively. The last line corresponds to the match set obtained with the 0.8 VNC threshold, on which the additional constraints of unicity and symmetry were imposed, as well as a background elimination constraint using truncated blocks with a threshold of 0.9 and a background multiplier of 0.25, and a disparity gradient constraint with a threshold of 0.4.

It is seen that simply using VNC with a low threshold can yield poor results, and that the most advantageous way of improving the match set is to filter it, rather then simply increasing the VNC threshold.

We also used Roth's software[4], described in [11], to estimate fundamental matrices from the match sets of table

[4]available at www2.vit.iit.nrc.ca/~gerhard

7. This experiment illustrates how the use of matching constraints makes the process of robust fundamental matrix estimation more efficient. The solution shown in Fig. 9, where the match set on which several constraints have been applied was used, is an accurate one. This solution was found in less than 500 RANSAC iterations. This is in contrast with the solutions found using a non-filtered match set that required, in this experiment, about 10000 iterations before finding an acceptable, but nevertheless inferior solution.

Figure 9: Epipolar geometry estimated from a filtered match set: the epipolar lines corresponding to selected points on the left image.

8 Conclusion

It was seen that many methods are useful in improving the quality of sets of candidate matches. The Harris corner detector was confirmed as a stable feature point extractor. VNC produces a good first candidate match set when unicity and symmetry are imposed. Constraints that use simple models of the corners were found to be beneficial. Finally, comparing matches with their neighbors to ensure that they have similar disparities allows the rejection of many mismatches.

A disadvantage of combining many constraints is that it results in more thresholds, and thus a greater need for tuning. However, as seen in the example of Table 7, good results can be obtained by combining constraints and using conservative thresholds.

This work was restricted to the study of matching in the case of image pairs with narrow baselines. An interesting extension would be the investigation of strategies for the case of image pairs that are more difficult to match.

References

[1] T. Darell, A Radial Cumulative Similarity Transform for Robust Image Correspondence, *CVPR*, 1998.

[2] E. Delp, O. Mitchell, Image Compression Using Block Truncation Coding, *IEEE Trans. on Communication* vol. 27, no. 9, pp. 1335-1342, 1979.

[3] O.D. Faugeras, P. Fua, B. Hotz, R. Ma, L. Robert, M. Thonnat, Z. Zhang, Quantitative and Qualitative Comparison of some Area and Feature-based Stereo Algorithms, *Robust Comp. Vision*, pp. 1-26, 1992.

[4] P. Fua, Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserve Depth Discontinuities. *ijcai*, 1991.

[5] E. Guelch, Result of Test on Image Matching of ISPR wg iii, *Int. Archives of Photogrammetry and Remote Sensing*, 27 (III), pp. 254-271, 1988.

[6] C. Harris, M. Stephens, A Combined Corner and Edge Detector, *Alvey Vision*, pp. 147-151, 1988.

[7] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, *Cambridge University Press*, 2000.

[8] R. Klette, K.Schluns, A. Koschan, Computer Vision: three-dimensional data from images. Sprigner, 1996.

[9] A. Koschan, A Framework for Area Based and Feature Based Stereo Vision. *MGV*, vol. 2, no. 4, 1993.

[10] H. Nishihara, T. Poggio, Stereo vision for Robotics, *ISRR*, 1983.

[11] G. Roth, A. Whitehead, Using Projective vision to find Camera Positions in an Image Sequence, *VI*, pp.225-232, 2000.

[12] D. Scharstein, R. Szeliski, R. Zabih, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, *CVPR*, *Workshop on Stereo and Multi-Baseline Vision*, 2001.

[13] C. Schmid, R. Mohr, C. Bauckhage, Comparing and Evaluating Interest Points, *ICCV*, pp. 230-235, 1998.

[14] S.M. Smith, J.M. Brady, SUSAN - A New Approach to Low Level Image Processing *Technical Report*, Oxford University, 1995.

[15] P. Torr, A. Zisserman, Robust Computation and Parameterization of Multiple View Relations, *ICCV*, pp. 727-732, 1998.

[16] Z. Zhang, R. Deriche, O. Faugeras, Q.-T. Luong, A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geomerty, *Technical Report*, INRIA, 1994.