Uncalibrated Stereo Correspondence by Singular Value Decomposition

Maurizio Pilu
Digital Media Department
HP Laboratories Bristol
HPL-97-96
August, 1997

image analysis, feature correspondence, stereo, singular value, decomposition

This paper presents a new simple method for achieving feature correspondence across a pair of images which requires no calibration information and draws from the method proposed by Scott and Longuet Higgins [8]. Despite the well-known combinatorial complexity of the problem, this work shows that an acceptably good solution can be obtained directly by singular value decomposition of an appropriate image-based correspondence strength matrix. The paper includes several experiments and discusses the method and draws comparisons with a related relaxation-based method by [14]. Given its tremendous performance / complexity figure, the method is particularly suitable for research purposes where an off the shelf but reliable feature correspondence is needed. For this reason, a succinct MATLAB implementation of the method is included and a C version will be soon available on the WEB.

Internal Accession Date Only

Presented at *Computer Vision & Pattern Recognition*, Puerto Rico, June, 1997.

0.1 Introduction

The problem of feature correspondence across two or more images is well known to be of crucial importance for many images analysis tasks. Reliable inter-image feature correspondence – and its closely related problem of image registration – is needed, just to cite a few, by structure-from-stereo approaches, motion analysis and tracking, image mosaicing, object pose and self-motion estimation.

Recently, there has been a boost of interest in the correspondence estimation problem due to the development of the Fundamental Matrix theory [3] and its tremendous practical implications in the analysis of uncalibrated stereo pairs and image sequences. If the fundamental matrix is known, reliable and fast feature correspondence can be obtained in general situations. However, in order for the fundamental matrix to be computed one needs a good initial set of feature correspondences (either lines, points or both [11]).

There are two schools of thought for solving the feature correspondence problem. In the first one, features are detected in one image and then correspondences for each of them are sought for in the second image, generally via multi-scale techniques. In the second approach, which the present work addresses, features are detected independently in both images and then matched up usually by relaxation (see, e.g., the classic [6]). Incidentally, recent state-of-the-art work on the fundamental matrix estimation [14, 11] follows this latter avenue for achieving initial correspondences.

This paper proposes an new neat and simple algorithm for achieving feature correspondence across pairs of images. Despite the well-known combinatorial complexity of the problem, this work shows that an acceptably good solutions can be obtained directly by singular value decomposition of an appropriate correspondence strength matrix.

The approach is largely inspired by the clever algorithm proposed by Scott and Longuet-Higgins for finding corresponding features in planar point patterns [8], which has been inexplicably overlooked in this area. This paper shows, for the first time, the viability of their approach for general stereo correspondence and propose a new mixed geometric and intensity-based correspondence strength function. Extensive experimental evidence is presented and discussed and some future work is proposed.

0.2 The correspondence problem

Due to its inherent combinatorial complexity and ill-posedness, feature correspondence is one of the hardest low-level image analysis tasks. The problem can be stated as finding pairs of features in two (or more) perspective views of a scene such that each pair correspond to the same scene point.

Early works in this field include illustrious works notably by Ullman [12] and Marr and Poggio [4]. In particular, Ullman put forward his *minimal mapping*

theory to implement three intuitive *local* criteria for establishing good *global mapping* that are: the *principle of similarity*, *principle of proximity* (other things being equal, choose the closest) and the *principle of exclusion* (only one-to-one matchings are allowed). As Marr pointed out, by simple local interactions a good global mapping effect can often be achieved.

These early works were inspired by psychology and neurophysiology and indeed provided some new insight into our visual system too. Since then, a vast amount of work has been done on the subject (for a review see [1], Ch. 6).

Most methods have a sometime complicate algorithmic formulation. For tasks such as estimating the fundamental matrix – where only a few tens of initial matches are needed[1]– leaner methods would perhaps be more suitable.

0.3 The Scott and Longuet-Higgins algorithm

In a landmark paper [8], Scott and Longuet-Higgins proposed a neat, direct way of associating features of two arbitrary patterns. The algorithm exploits some properties of the singular value decomposition (SVD) to satisfy both the exclusion and proximity principles set forth by Ullman. A remarkable feature of the algorithm is its straightforward implementation founded on a well-conditioned eigenvector solution which involves no explicit iterations[2].

In the following, a brief description of the algorithm is given along with a simple experiment that illustrate its intrinsic usefulness. The reader is redirected to the original paper [8] for further theoretical and philosophical insights.

Let I and J be two images, containing m features I_i $(i = 1 \dots m)$ and n features J_j $(j = 1 \dots n)$, respectively, which we want to put in one-to-one correspondence. The algorithms consists of three stages.

1. Build a *proximity matrix* $\mathbf{G}$ of the two sets of features where each element G_{ij} is Gaussian-weighted distance between two features I_i and J_j:

$$G_{ij} = e^{-r_{ij}^2/2\sigma^2} \qquad i = 1 \dots m, j = 1 \dots n \tag{1}$$

   where $r_{ij} = \|I_i - J_j\|$ is their Euclidean distance if we regard them as lying on the same plane. $\mathbf{G}$ is positive definite and a G_{ij} decreases monotonically from 1 to 0 with distance. The parameter σ controls the degree of interaction between the two sets of features: a small value of σ enforces local interactions, while a larger value permits more global interactions.

2. Perform the *singular value decomposition* (SVD) of $G \in M_{m,n}$:

$$\mathbf{G} = \mathbf{TDU}^T.$$

[1] Accurate, dense correspondence can be more efficiently determined later by exploiting the epipolar constraint.

[2] Of course, numerical implementations of the SVD do actually require inner iterations.

where $\mathbf{T} \in M_m$ and $\mathbf{U} \in M_n$ are orthogonal matrices and the diagonal matrix $\mathbf{D} \in M_{m,n}$ contains the (positive) singular values along its diagonal elements D_{ii} in descending numerical order. If $m < n$, only the first m columns of $\mathbf{U}$ have any significance [8].

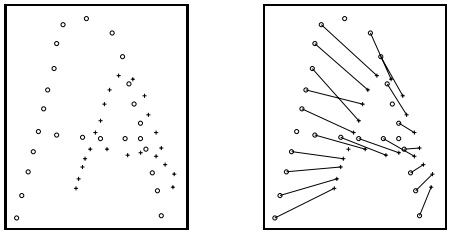
3. Convert $\mathbf{D}$ to a new matrix $\mathbf{E}$ obtained by replacing every diagonal element D_{ii} with 1 and then compute the product

$$\mathbf{P} = \mathbf{TEU}^T.$$

This new matrix $\mathbf{P} \in M_{m,n}$ has the same shape as the proximity matrix $\mathbf{G}$ and has the interesting property of sort of "amplifying" good pairings and "attenuating" bad ones: "*if P_{ij} is* both *the greatest element in its row* and *the greatest element in its column, then we regard those two different features I_i and J_j as being in 1:1 correspondence with one another; if this is not the case, it means that features I_i competes unsuccessfully with other features for partnership with J_j*" [8].

It is not difficult now to figure out that this apparently simple method embeds both the *proximity* and the *exclusion* principle: the former one is a consequence of the nature of the proximity matrix and the latter arises from the orthogonality of the matrix $\mathbf{P}$. In fact *"the fact that the squares of the elements in each row of* $\mathbf{P}$ *must add up to 1 implies that a given feature I_i cannot be strongly associated with more than one feature J_j. The mutual orthogonality of the rows tends to keep different features in the first image from becoming closely associated with the same feature in the second image"*[8]. Moreover, $\mathbf{P}$ is a matrix which effectively produces a *minimum squared distance mapping*, since by applying the algorithm the trace of $\mathbf{P}^T\mathbf{G}$ is maximized [8].

Although not mentioned in the original paper, the algorithm is rooted into the solution of the subspace rotation problem known as *orthogonal Procrustes problem*. Foundations and proofs can be readily found in, for instance, [2].



As an example, the figures above show the mapping found by this algorithm for two hand-input patterns representing two letters "A" (circles and crosses) scaled and translated; the overall mapping is extremely good and, as claimed, the proximity principle is defied in favor of a more globally consistent mapping.

Figure 1: LEFT: The SVD method applied with G as in Eqn. (1) leaves out many points due to ambiguities cause by rogue points. RIGHT: The SVD method applied with just point correlation yield non-sense results.

Scott an Longuet-Higgins show that the algorithm copes nicely with translation, shearing and scaling deformations and with moderate rotations (as in our visual system) and suggest criteria for the choice of the distance σ.

To my best knowledge, thus far only two (similar) works appears to have used the startling properties of this algorithm, namely [7] and [9], where the method was applied to matching modes of variation of finite element shapes, and in [5], where applications were suggested in eigen-shape fitting. The following section explains why the method *as is* does not fare as it could in real image matching situations and proposes a simple but key improvement on the nature of the **G** matrix.

0.4 Stereo matching by SVD

Let us now consider the case of real images pairs, and the problem of matching across points of interest (e.g. corners). Actual point feature detectors are unstable and it is reasonable to expect that some points that appear in one image do not show up in the other (*rogue points*).

Rogue points cause lots of ambiguous, equally good matching possibilities in the space of pairings, and the sole proximity used to build **G** in Equation (1) does not have enough "character" to discriminate amongst them. As a consequence, only a handful of features are safely in one-to-one correspondence with others (also pointed out in [8]).

Only a few experiments are needed to validate this claim. Figure 1-left shows how only few corners (cf. with Figure 2) are found in 1:1 correspondence; a large number of highly-correlated corners are left out because rogue points cannot be told apart from good ones just from their spatial location. This behavior can be summarized by saying that the Scott and Longuet-Higgins algorithm *does not*

embed the feature similarity principle, so dear to most stereo correspondence approaches.

Obviously, this behavior calls for the use of some local measurements to quantify feature similarity, such as the normalized (cross) correlation between gray level patches about the features.

If we represent two $W \times W$ areas centered on features I_i and J_i as two $W \times W$ arrays of pixel intensities $\mathbf{A}$ and $\mathbf{B}$, respectively, the normalized correlation is defined as $C_{ij} = \frac{\sum_{u=1}^{W}\sum_{v=1}^{W}(A_{uv}-\overline{\mathbf{A}})\cdot(B_{uv}-\overline{\mathbf{B}})}{W^2\cdot\sigma(\mathbf{A})\cdot\sigma(\mathbf{B})}$ where $\overline{\mathbf{A}}$ ($\overline{\mathbf{B}}$) is the average and $\sigma(\mathbf{A})$ ($\sigma(\mathbf{B})$) the standard deviation of all the elements of $\mathbf{A}$ ($\mathbf{B}$). C_{ij} varies from -1 for completely uncorrelated patches to 1 for identical patches.

One way of including this correlation information into the proximity matrix is to transform the elements of $\mathbf{G}$ as follows:

$$G_{ij} = [e^{-(C_{ij}-1)^2/2\gamma^2}] \cdot e^{-r_{ij}^2/2\sigma^2} \tag{2}$$

where term is bracket is a gaussian-weighted function of the correlation C_{ij} in which γ determines how quickly its values decreases with a diminishing C_{ij} ($\sigma = 0.4$ throughout the paper).

This new correspondence strength can be seen as a *correlation-weighted proximity*. It is easy to see that the elements of $\mathbf{G}$ still range from 0 to 1 and, as in Equation (1), the closer and the more correlated two features I_i and J_i are, the higher G_{ij} is going to be.

This new correspondence strength now embodies *similarity* between features and is therefore much more selective than just proximity as in Equation (1). In some ways, by applying the algorithm with the said correlation-weighted G, we obtain a minimum overall distance mapping still complying to the proximity and uniqueness principles but *under the constraint of similarity*[3].

Figure 2-left shows the matches obtained by this method for the same image as of Figure 1-left. It can be seen that a considerably higher number of 1:1 matches has been found. In other examples (not shown here) many bad pairings were also replaced by correct ones.

Next section discusses this experiment and others in further detail to show the effectiveness of the method.

0.5 Some experimental results

Several experiments have been performed on image pairs of various size and quality; some of the results are reported and discussed here.

The features were detected via the SUSAN corner detector [10]. In order for the correlation not to be too affected by noise, images were Gaussian smoothed.

[3]If only the correlation information was used (i.e. $G_{ij} = e^{-(C_{ij}-1)^2/2\gamma^2}$ the results would be extremely poor, as shown in Figure 1-right, producing a curious "maximal sum of correlation" mapping, which makes clearly no physical sense.

Figure 2: SVD matching with correspondence strength as in Eqn. (2). Disparities are overlaid onto the left image and matching corners onto the right one. Notice how the method managed to find the correct correspondences in difficult areas such as in the back of the stapler. (Images courtesy of INRIA)

The key parameter σ in Equation (2) was set to 1/8 of the image width; more about this later in the section.

It is extremely important to point out that in order to illustrate the performance of the algorithm *by itself*, no further processing for filtering out bad matches has been applied. Most of the bad pairings that will be seen in the examples could have been simply eliminated by any commonly used technique, such as coherence of disparity.

Back to the first experiment, Figure 2-left shows the disparities overlaid onto the first image and the corresponding corners on the second image (right); this method for presenting results will be used throughout. Being related by translation and rotation, this case involves non-uniform disparities but, as it can be seen, the results are extremely good and numerous 1:1 pairings have been found. It is encouraging to see how the method managed well to disentangle itself in areas where there is a large number of close-by features. Notably, a few good matches (about 10) have been missed out for some reasons such as low-correlation or simply because there were two or more equally competing alternatives.

Figure 3-left presents the case of a poor quality road scene, with a remarkable expansion. It can be seen that although there are six mismatches, an overall good mapping was obtained. Note that near the focus of expansion the disparities cannot be very accurate working at pixel resolution.

Figure 3-middle shows an image pair related mainly by translation. This case has some potential problems because of the clusters of features concentrated in the right-hand side of the image (e.g. the car wheel). The method has however performed exceedingly well, leaving just one grossly misjudged pairing.

Finally, Figure 3-right presents another case that has prevalent translation. Here too there could be some difficulties due to the high displacement and

Figure 3: Some more test images pairs. Disparities are overlaid onto the top images and matching corners onto the bottom ones. Comments in Section 0.5.

the highly repetitive and seamless features of the window and the tree leaves, respectively. Although 7 pairings are grossly wrong, the overwhelming majority is correct and would easily allow a robust next stage to operate.

The choice of the parameter σ, thanks to the better discriminating correspondence strength function, is fairly easy. The table in Figure 4-right gives the number of mismatches (found by visual inspection) for the pairs in Figure 3-middle and left with respect to changes in σ expressed as fraction of the image width.

It is clear that σ can vary within a relative large range without affecting performance too much. Having said this, in [8] it is suggested that the value of σ should reflect the average displacement of points; supportingly, our experiments also show that the best results are obtained when σ roughly matches the actual image displacement.

0.6 Comparison with a recent relaxation method

Recently a features matching method has been presented by Zhang's *et al.* in [14] which is based on maximizing the sum of a measure of support over the possible feature pairings. The method (as it can be tried out on their on-line demo) manages to disentangle itself in quite difficult situations and produces matches good enough to allow an easy recovery of the epipolar geometry.

We have adventured in re-implementing Zhang's *et al.* method and have

performed some qualitative comparisons. Amazingly the performances of the methods is remarkably similar both in the good and in the bad.

The explanation is simple. Their method relies upon a pairwise correspondence strength that uses a local measure of support weighing the straight correlation between candidate matches with a measure of "distortion" of distances to neighboring matches; the underlying principle is that the relative distance between neighboring sets of features should not change under a local affine approximation of the transformation. The relaxation stage does nothing but selecting matches with *mutual maximal strength* and that also show *little ambiguity* with other competing matches.

As explained in Section 0.4, these criteria are implicitly implemented in the method proposed here, albeit in a decisively global fashion. The globality of the proposed algorithm as opposed to the relative local-ness of Zhang's gives it a relative speed disadvantage. However, although not mentioned in their papers, a close look at the algorithm reveals that its complexity is $O(nMK^2)$, where n is the number of features in one image, M is the average number of candidates in the other image and K is the average number of neighbors (within a given radius) of a features. As a matter of fact, M and (to some extent) K both grow linearly with the number of overall features and so one should watch out before declaring it $O(n)$ (as it can be seen experimentally)!

One last thing to be said is that Zhang's method performs well when there are many features sprinkled uniformly in the images in order to give support to candidate matches. The method proposed here performs extremely well also in *sparse situations*, which might be a clear advantage in multistage approaches where just a few good features can be matched well in order to compute the epipolar geometry, rather than using hordes of them.

0.7 Discussion

This paper has proposed a new effective method for performing features-based stereo correspondence. We have seen that its quality lies not in its accuracy but in a tremendous performance/complexity ratio, that makes it particularly suitable as bootstrap for other more accurate methods.

Although the additional cost of the algorithm is only that of computing the SVD of $\mathbf{G}$ [4], the computation time should not be overlooked. The SVD is one of the stablest numerical matrix operations and its basic complexity is $O(m^2 n)$. This is not too bad for a stereo matching algorithm but for large number of features can become impractical. In numbers, by the SVD routine provided by the MATLAB package, on a HP9000/720 workstation it takes about 6s for $n = m = 200$ but just 0.2s for $n = m = 50$. A straightforward remedy would be to zero out very small G_{ij} that have no chances of becoming 1:1

[4] The overhead of feature detection and correlation computation are common to most feature-based matching approaches.

$\left(\frac{\sigma}{\text{img\_width}}\right)$	# mismatches	
	Fig.3 mid	Fig.3 left
1/6	2	5
1/7	1	4
1/8	1	6
1/9	0	6
1/10	1	7

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%          Singular Value Decomposition Stereo Matching
%
% C=svdmatch(point_corr,point_dist,sigma,corr_th);
%  point_corr: m x n matrix  of feature correlations
%  point_dist: m x n matrix  of feature interdistance
%  sigma:      adjust point interaction (expected displacem.)
%  corr_th:    min acceptable correlation for a match
%  C: Feature mapping matrix;  C(i,j)=1 when 1:1 corresp.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function C=svdmatch(point_corr,point_dist,sigma,corr_th);
% Get number of features in both images
    [m, n] = size(point_corr);
% Build correspondence strength matrix
   G =  exp(-((point_corr-1).^2)/(2*0.4^2)) .*
     exp(-(point_dist).^2/(2*sigma^2));
% Perform SVD and extract orthogonal  matrix P
    [T,D,U] = svd(G);
    D = diag(ones(min(n,m),1));  D(m,n)=0;  % D is m x n
    P = T*D*U';
% Find dominant of each line and column
    [V, I] = max(P');  [V, J] = max(P);
% Initialize correspondence matrix  C
    C = zeros(m, n);
% Set a one-to-one correspondence between Ii and Jj
% only if  P(i,j) is MAX of  both row i and column j
% AND  their correlation is  above corr_th
    for i=1:length(J)
       if I(J(i))==i
          if (point_corr(i,J(i))>corr_th)
             C(i,J(i)) = 1;
          end
       end
    end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Figure 4: LEFT: Sensitivity of results to variations of σ. RIGHT: MATLAB code of the SVD stereo correspondence method proposed in this paper.

pairings in order to make the matrix **G** sparse and allow for optimized numerical solutions. Intriguingly, due to its extreme neatness and regularity, the algorithm could lends itself to real-time hardware implementations, thanks to some general purpose, scalable SVD hardware engines such as those proposed for trajectory control of robots [13]. Other more complex algorithms can hope for real-time performance only by implementations on expensive parallel architectures.

Another interesting property of this algorithm is that it does not explicitly require a specific correlation threshold for a feature pair to be accepted as what matters is its relationship with others competing matches. In all the examples shown in the previous section the correlation threshold was set to as low as 0.4, whereas normally it is set to 0.7/0.8 (see e.g. [14]).

Arguably, the black-box nature of the algorithm may be seen as a limitation. For instance, it is not possible to embed any non-pairwise constraint, such as the disparity gradient or the geometric ordering constraint[6] . Some of these problems could be bypassed, though. If, for instance, in a two-stage approach a few *definitely good* matches are known beforehand, one could recover a very coarse epipolar geometry and embed the deviation from epipolar lines in the correspondence strength function in Equation (2) and then apply the algorithm to all the other features.

Although some standard pruning method could be easily applied to cut down rogue matches, other pair-wise similarity measures could be more interestingly

tried out in place of (or in conjunction with) correlation in Equation (2). In particular, multi-band correlation and measures of neighboring matches support – such as the *strength of the match* used in [14] – seem the most promising ones.

Lines are more stable features than corners and recently a tensor-based method has been found for uniformly use line and points to estimate the epipolar geometry (e.g., see [11]). A natural extension of the algorithm that should be easy to implement, is to apply it to properly parameterized lines.

Finally, since it is possible that the algorithm will be used by other researchers due to its simple implementation and reasonably good performance, the commented MATLAB code is given in Figure 4.

Acknowledgments

Thanks to Andrew Fitzgibbon and Stephen Pollard for useful suggestions.

Bibliography

[1] O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, 1993.

[2] G. Golub and C. V. Loan. *Matrix Computations*. North Oxford Academic, 1983.

[3] Q. Luong and O. Faugeras. The fundamental matrix: Theory, algorithms and stability analysis. *International Journal of Computer Vision*, (17):43–75, 1996.

[4] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proc. Royal Society London*, B 204:301–328, 1979.

[5] M. Pilu. *Part-based Grouping and Recognition: A Model-Guided Approach*. PhD Thesis, Department of Artificial Intelligence, University of Edinburgh, Scotland, Sept. 1996.

[6] S. Pollard, J. Mayhew, and J. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, (14), 1985.

[7] S. Sclaroff and A. Pentland. Modal matching for corrispondence and recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(6):545–561, 1995.

[8] G. Scott and H. Longuet-Higgins. An algorithm for associating the features of two patterns. In *Proc. Royal Society London*, volume B244, pages 21–26, 1991.

[9] L. S. Shapiro and J. Brady. Feature-based corresondence: An eigenverctor approach. *Image and Vision Computing*, pages 283–288, June 1992.

[10] S. Smith and J. Brady. SUSAN - A new approach to low level image processing. *International Journal of Computer Vision*, 1997. To appear.

[11] P. Torr and D. Murray. A review of robust methods to estimate the fundamental matrix. Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford, 1996.

[12] S. Ullman. *The interpretation of Visual Motion*. MIT Press, Cambridge, MA, 1979.

[13] I. Walker and J. Cavallaro. Parallel VLSI architectures for real-time kinematics of redundant robots. In *Proceedings IEEE International Conference on Robotics and Automation*, pages 870–877, 1993.

[14] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luon. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78:87–119, 1996.