Vision, Video and Graphics (2005)
E. Trucco, M. Chantler (Editors)

SVD-matching using SIFT features

Elisabetta Delponte, Francesco Isgrò, Francesca Odone, Alessandro Verri

INFM - DISI, Università di Genova, Italy

Abstract

The paper tackles the problem of feature points matching between pair of images of the same scene. This is a key problem in computer vision. Among the many possible applications of feature matching we are motivated for helping in the initialisation of a 3D registration procedure. The method we discuss here is a version of the SVD matching proposed by Pilu, modified in order to cope with large scale variations. We detail the algorithm and present experimental evidence of the improvement in performance. The main contribution of this work is in showing that this compact and easy algorithm can be used for large-baseline matching.

1. Introduction

Finding correspondences between feature points is one of the keystones of computer vision, with application to a variety of problems. Automatic feature matching is often an initialisation procedure for more complex tasks, such as fundamental matrix estimation, image mosaicing, object recognition, and three-dimensional point clouds registration.

The last of the tasks mentioned above is the one that motivated our work, and is probably the less known. A 3D data acquisition system returns measures of the same object coming from different poses, therefore the various 3D scanning need to be registered. Even if methods for 3D data registration do exist [BM92], they do need a good initialisation, that might be provided by point matched in the images, providing that an appropriate calibration between an image and a 3D point cloud is given. We aim at obtaining a system to compute sparse correspondences between image pairs and use them to initialise a 3D registration procedure. From the feature matching standpoint our goal is to devise a procedure that allows us to obtain a reasonably high number of accurate matches from image pairs acquired by a large baseline system.

As mentioned before, point matching is a corner-stone for a large class of computer vision algorithms, and for this reason has been tackled since the old days of computer vision research [vdBR77, RvdB77]. For this reason we focused our presentation on general images, and not on images related to our 3D scanning problem. The problem of matching points between images is covered by a very rich literature. We do consider here the case when the epipolar geometry is not known, and then the corresponding point can be anywhere in the image. A classical approach, for the case of short baseline, is the one presented in [DZLF94], that adopts a standard correlation step, followed by a relaxation step. Correlation between grey-level values is also used in [Pil97]. The performance of all these algorithms are usually poor in the case of large scene differences.

In this paper we discuss a modification of the algorithm proposed in [Pil97], that uses meaningful local descriptors to deal with a larger baseline and scale changes. We claim that the not great performance of the algorithm in the case of wide-baseline is due to the feature descriptor adopted, more than to a limit of the algorithm. To this end our method uses difference of Gaussians to find interesting points that are tolerant to change in scale and rotation, and represents them with the SIFT descriptor [Low04]. Finally it computes correspondences among the key-points with a mixed geometric and appearance based approach, as in [Pil97]. The algorithm proposed in [Pil97] shows that a reasonably good solution can be achieved simply by singular value decomposition of an appropriate correspondence strength matrix. The method — building on top of a work by Scott and Longuet-Higgins [SLH91] for pattern matching, and applying it to intensity images — is a simple and elegant matching procedure that takes into account both intensity and geometric relationships.

Scale Invariant Features (often referred to as SIFT) were first proposed in [Low99] and attracted the attention of the computer vision community for their tolerance to scale, illumination and pose variations. A comparative study of many local image descriptors [MS03], shows the superiority of SIFT with respect to other feature descriptors for the case of several local transformations.

We report promising results, discussing how the matching behaves while the baseline grows. A comparison with the original work shows how SIFT features make the system more tolerant to the effects of a larger baseline. We also report a comparison to the matching criterion for SIFT keypoints proposed by Lowe, showing that our approach is a better compromise between accurate results and a high number of correspondences.

The paper is organised as follows. The next Section gives a brief overview of matching algorithms for large scale variations. In Section 3 we review the SVD matching algorithm. In Section 4 the scale invariants features and their associated descriptor are briefly reviewed. Our modified SVD matching is described in Section 5, and experimental results are shown and discussed in Section 6. Section 7 is left to the final remarks.

2. Previous work on matching with large scene variations

The state-of-the art on matching algorithms is vast. In the remainder of the section we sketch some of the most interesting approaches in dealing with substantial scene variations, either in scale or in the view-point.

The great part of matching algorithms address these issues through the so-called invariant regions constructed around salient points, typically corners, in such a way to try and keep the characterisation of that area insensitive to viewpoint and illumination changes. For instance, scale invariant feature extraction can be achieved by using the Harris detector [HS88] at several scales, or by considering local extrema in a pyramidal difference of Gaussians [Low04]. Once relevant regions are detected, the actual feature matching takes place. Finding appropriate region descriptors may facilitate feature comparisons. The SIFT descriptor, better outlined later in the paper, has been shown [MS03] to be one of the most efficient to date.

SIFT descriptors are used in combination with multiscale Harris by Mikolajczyk and Schmid [MS02], while Baumberg [Bau00] propose a matching technique based on a similar feature and a description based on the Fourier-Mellin transform to achieve invariance to rotation. Harris corners are also used in [ADJ00], where rotation invariance is obtained by a hierarchal sampling that starts from the direction of the gradient. Matas et al. [MCUP02] introduce the concept of maximally stable extremal region to be used for robust matching. Essentially, these regions are connected components of pixels which are brighter or darker than pixels on the region's contour. They are invariant to affine and perspective transform, and to monotonic transformation of image intensities. Tuytelaars [TG00] deals with wide-baseline matching extracting image regions around corners, where edges provide orientation and skew information, while scale variation is addressed computing the extrema of a 2D affine invariant function; as a descriptor they use generalised colour moments, while the actual matching is done with a Malahnobis distance. In a more recent work [FTG03] they establish wide-baseline correspondences among unordered multiple images, by first computing pairwise matches, and then integrating them into feature tracks each representing a local patch of the scene. They exploit the interplay between the tracks to extend matching to multiple views. A method based on automatic determination of local neighbourhood shape is presented in [GSBB03], but it only works for image areas where stationary texture occurs.

An alternative approach to determining feature correspondences relies on prior knowledge on the observed scene, for instance in knowing the epipolar geometry of two or more views [SZ97, FR96]. Georgis et al. [GPK98] assume that projections of four corresponding non coplanar points at arbitrary positions are known. Pritchett and Zissermann [PZ98] use local homographies determined by parallelogram structures or from motion pyramids. Lourakis et al [LTAO03] present a method based on the assumption that the viewed scene contains two planar surfaces and exploits the geometric constrains derived by this assumption. The spatial relation between the features in each images, together with appearance, is used in [TC02]. Recently simple ordering constraint that can reduce the computational complexity for wide baseline matching, for the only case of approximately parallel epipolar lines, has been proposed in [LM04].

3. SVD matching

This section summarises the matching algorithm proposed in [Pil97], upon which we base our key-points matching technique. The algorithm builds on top of a method for points matching [SLH91] and adapts it to deal with pixel correspondences. In [SLH91] it was shown that, in spite of the well-known combinatorics complexity of feature correspondence, a reasonably good solution can be achieved through the singular value decomposition of an appropriate correspondence strength matrix. In [Pil97] this matrix is adapted to take into account image intensity as well as geometric properties.

Let A and B be two images, containing m and n features respectively (A_i, $i = 1, \ldots, m$, and B_j, $j = 1, \ldots, n$), each of them represented by a simple $w \times w$ image patch. The goal of the algorithm is to put the two sets of features in one-to-one correspondence.

The algorithm consists of three steps:

1. Build a correspondence matrix $\mathbf{G}$ that models both geometric proximity and similarity; each element G_{ij} is com-

puted as follows

$$G_{ij} = \frac{C_{ij}+1}{2} e^{-r_{ij}^2/2\sigma^2}. \quad (1)$$

$r_{ij} = ||A_i - B_j||$ is the Euclidean distance between the two features, if we imagine them in the same reference plane, and C_{ij} is the normalised correlation between them. The parameter σ controls the degree of interactions between features, where a small σ enforces local correspondences, while a bigger σ allows for more distant interactions. The elements of $\mathbf{G}$ range from 0 to 1, with higher values for more correlated features.

2. Compute the Singular Value Decomposition for $\mathbf{G}$: $\mathbf{G} = \mathbf{VDU}^\top$.
3. Compute a new correspondence matrix $\mathbf{P}$ by converting diagonal matrix $\mathbf{D}$ to a diagonal matrix $\mathbf{E}$ where each element D_{ii} is replaced with a 1: $\mathbf{P} = \mathbf{VEU}^\top$. It is shown in [SLH91] that $\mathbf{P}$ carries similar information of $\mathbf{G}$, with the interesting property of enhancing good pairings.

In [Pil97] experimental evidence is given that the proposed algorithm performs well on short baseline stereo pairs. In fact the performance falls when the baseline increases. It is our target to show that the reason for this behaviour is in the feature descriptor chosen and is not an intrinsic limit of the algorithm.

4. Features detection in scale-space

Lowe [Low04] presented a method for extracting and representing local features (*Scale invariant features transform* key-points, also known as SIFT) tolerant to scale changes, illumination variations, and image rotations. These features are also claimed robust to affine distortion, change of viewpoints and additive noise. Recently it has been shown [MS03] that SIFT descriptors are more stable than other state of the art interest point descriptors. In the remainder of this section we will briefly introduce the key-points and a possible description. In Figure 1 we show the feature points that have been extracted from two images used in our experiments.

Scale-space and interest points detection in images When approaching to computer vision one of the first remarks is that every object in an image assumes a different significance if observed at a different scale. It has been demonstrated [Lin96] that *scale-space* is a good framework to handle objects in images at different scale. Indeed, scale-space is a representation of the image which is seen at different resolution levels while its fine-scale structures are deleted. The description obtained is not a simple random suppression of details, but it is a well defined process that guarantees linearity and spatial shift invariance. A foremost aspect of the scale-space approach is that there are methods [Lin98] that allow to automatically choose the appropriate resolution level discarding unuseful information.

Scale invariant features The process of building SIFTs [Low04] is heavily inspired by the scale-space framework. The process can be sketched in two phases: the first is key-points detection in scale-space pyramid and the second is key-points description using the image gradient at the right level of resolution.

Key-points are detected in a structure which is a pyramid of *difference-of-Gaussians* (DoG) filtering of the original image. Given an image I, it is convolved twice with a Gaussian function to obtain I_{σ_1} and I_{σ_2}: the difference of these two images will be the first level of the pyramid. Afterwards the image is sub-sampled and the process to obtain the DoG is repeated until the sub-sampled image keeps some useful information. Once the pyramid of DoGs is completed, maxima and minima are located. The feature detection phase ends with a cleaning procedure for discarding low contrast features and for filtering out edges.

Regions detected by DoG extrema are mainly blob-like structures. There are no significant signal changes in the centre of the blob and therefore the Gaussian filter-based descriptors perform better in larger point neighbourhood [MS03]. The key-point position is defined in the scale-space to gain invariance to scale change. To achieve also invariance with respect to image rotation another feature is attributed to the key-point: its orientation. The orientation of a key-point is defined using a histogram for the gradient direction in a circular neighbourhood of the pixel.

Following [Low04], once key-points are selected the associated descriptors are computed as a composition of direction histograms in the neighbouring regions of the scale-space, shifted according to the dominant orientation of the feature. It is remarkable to notice that this description keeps scale information since the histogram is evaluated at the proper level of the scale in the pyramid.

5. SVD matching using SIFT

In this section we discuss the use of the SIFT descriptor in the SVD matching algorithm. As mentioned in the introduction the SVD matching presented in [Pil97] does not perform well when the baseline starts to increase. The reason for this behaviour is in the feature descriptor adopted. The original algorithm uses the grey level values in a neighbourhood. It is now well known that image neighbour grey level values is a descriptor too sensitive to changes in the view-point, and more robust descriptor have been introduced so far (see, for instance, [ZW94, TG00, FA91]).

Results of a comparative study, performed on a set of planar scenes, of the performance of various features descriptors have been reported in [MS03], where it is shown that the SIFT descriptor is better than the other descriptors with respect rotation, scale changes, view-point change, and local affine transformations. The quality of the results decrease in the case of changes in the illumination. In the same work,

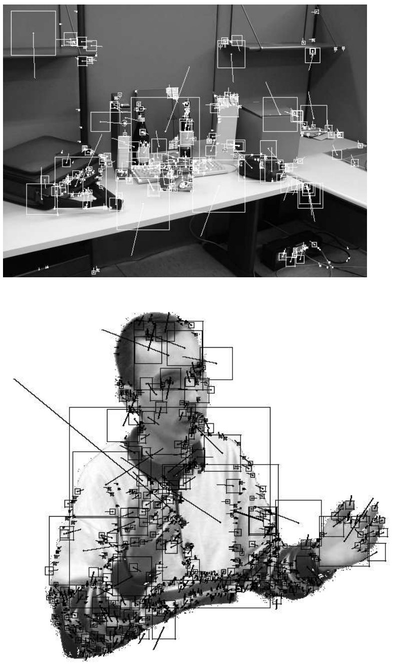


Figure 1: *Example of SIFT features extracted from some of the images used in our experiments. The square around the feature points represent the support area of the feature.*

cross-correlation between the image grey levels returned not stable performance, depending on the accuracy of the point, that depends strongly on the kind of transformation considered. The considerations above suggested the use of a SIFT descriptor, instead of grey levels. We left the matrix **G** in equation (1) unchanged in its form, but C_{ij} is now the cross-correlation between SIFT descriptors. As it will be shown in the next Section, this straightforward modification improves the performance of the SVD matching, and also gives better results, in terms of number of points correctly matched, with respect the SIFT distance used for the experiments reported in [MS03]. We do plan to experiment with different SIFT distances in the SVD matching, which might require to modify the form of the **G** matrix.

6. Experimental results

In this section we report some experiments carried out on different image pairs and sequences. First we show some of the matches returned by our algorithm on few image pairs. Then we attempt a more quantitative analysis of the performance of our algorithm on short image sequences, compared against other two matching algorithms.

Experiments on image pairs The first lot of experiments that we show refers to results of our algorithm on image pairs of two different scenes. In Figures 2 (a) and 2 (b) we show all the matches determined on two pairs of images of a desk scene. The first one presents a reasonable level of scene variation, whereas the latter is a synthetic rotation of the first image. We spotted only a wrong match in Figure 2 (a). The last image pair is relative to a studio scene with scale variation. The result is shown in Figures 2 (c). Our visual inspection of the results determined only few wrong matches between points on the border of the table.

Comparative experiments Now we report experiments carried on different types of short images sequences. For the first type of image sequences the camera was moving around a *complex* indoor scene (i.e., with several objects), increasing the baseline with respect to the camera pose for the first frame in the sequence. The second type of images sequences we considered are stereo image sequences, and in particular we focused our experiments on input sequences for immersive video-conferencing systems. For reason of space we show here only results on a sequence of 5 frames of the first type, and a 30 frames stereo sequence for the second class of data.

For each frame in the sequence we extracted a set of interest points, using the DoG points detector described in Section 4, that proved invariant to rotation and scale changes [Low99, MS03]. We remind here that the points detected are local scale-space extrema of the difference of Gaussians. The size of the support region, the area used for computing the associated descriptor, is determined from the selected scale. We compare the performance of the SIFT based SVD matching, henceforth S-SVD, against the performance of the correlation based one (C-SVD), and a SIFT based point matcher proposed by Lowe in [Low99], and used in [MS03] for measuring the SIFT performance, which uses the Euclidean distance between SIFTs. We will address to this last matching method as S-DIST. More formally the correspondence are established as

- **S-SVD**: point matches are established following the algorithm of Section 3

$$C_{ij} = \sum_t \frac{(S^i_t - mean(S^i))(S^j_t - mean(S^j))}{stdv(S^i)stdv(S^j)}$$

  where S^i and S^j are the SIFT descriptors;
- **C-SVD**: point matches are determined as above but with

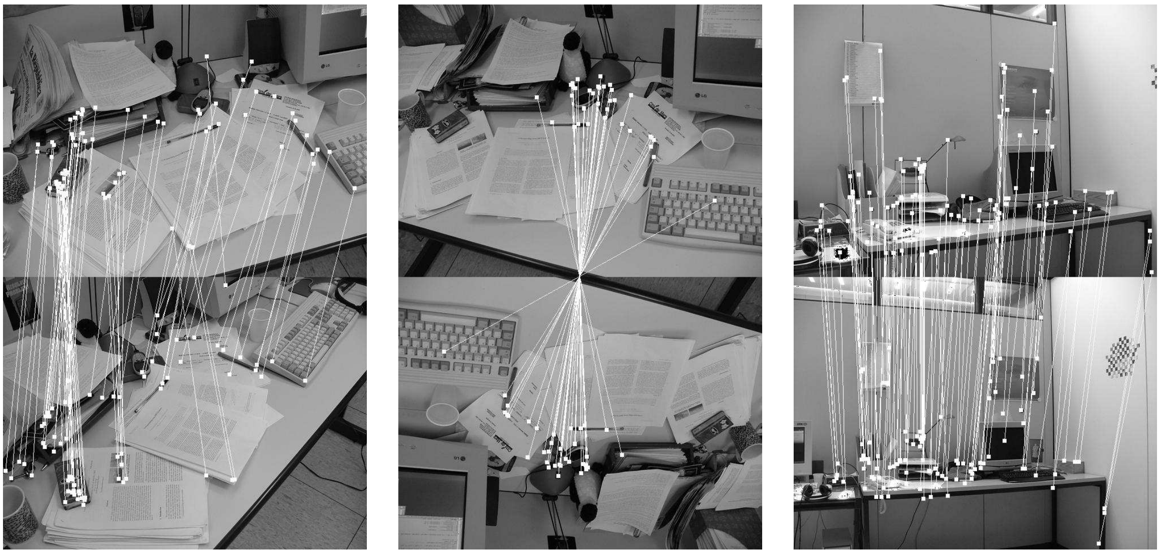
$$c_{ij} = \sum_t \frac{(I^i_t - mean(I^i))(I^j_t - mean(I^j))}{stdv(I^i)stdv(I^j)}$$

  where I^i and I^j are the two grey-levels neighbour;
- **S-DIST**: two features i and j matches if

$$d_{ij} = \min(D_i) < 0.6\min(D_i - \{d_{ij}\})$$

  where $D_i = \left\{ d_{ih} = ||S^i - S^h|| \right\}$.

In order to discriminate between a correct match and a



(a) (b) (c)

Figure 2: *a) Matches determined for a stereo pair of a desk, with a reasonable level of scene variation. We could notice only one wrong match between the wall and the corner of the screen. b) Matches determined for a stereo pair of a desk, the second image is a synthetic rotation of the first one. No wrong matches have been determined. c) Studio stereo pair with scale variation.*

false match we decided to compute the fundamental matrix [HZ00] between the two frames using a statistically robust method, that allows us to identify the wrong matches. In particular our implementation adopted the Least Median of Squares regression method [MMRK91]. Needless to say that our measure is reliable only on the assumption that the wrong matches are less than 50% of the all matches. It is worth to point out that with this measure the correctness of a match differs from the common understanding, as they only need to be *consistent* with the epipolar geometry. It is therefore possible to have a wrong match that is passed as correct (see Figure 3 for an example).

For evaluating the performance of the three point matching methods used for this work we computed: a) the total number of matches detected; b) the number of correct matches; c) the accuracy, defined as the ratio between number of correct matches and the total number of matches detected. The plot of these three value relative to the 5-frames sequence for which we show results, can be found in Figure 4. In Figure 4 and 3 we show results of the matching between the first frame of the sequence and the sixth and the eightieth frame respectively. Overall we can say that S-SVD outperforms C-SVD in all the cases. In general S-SVD returns the largest number of correct point matches, and total number of point matches. In terms of accuracy the best one is S-DIST, but S-SVD has an accuracy of more than 0.5 for almost half the length of each sequence, that makes the quality of the matches good enough for any state of the art robust estimator [MMRK91]. On the other hand, the number of correct matches detected by the S-DIST for the cases when the accuracy of the S-SVD is below 0.5, is very often too small (2-6) for being useful for any task.

The quality of the matches decreases for all the three methods as the baseline starts to be too large, meaning that none of the methods can be feasible for wide-baseline matching, and some more work needs to be done in attempting to make the S-SVD algorithm more robust to the baseline variation. In Figure 6 and 5 we show the results for the videoconferencing stereo sequence. Again the S-SVD returns the largest number of correct matches. In this case the accuracy of the three methods are comparable. The better performance in terms of accuracy are probably due to the fact that for this case, being the baseline constant, it was possible a better tuning of the parameters of the SVD-matching.

As last experiment we show in Figure 7 all the matches detected by the three algorithms considered for a pair with very large scene variation. It is evident that even if our algorithms finds more correct matches than the other two, in its present form it cannot still cope with very large scene variation, as the other two methods considered. Work is in progress to try and modify the similarity measure to cope with such configurations.

7. Conclusions

In this paper we described an improved version of the SVD matching presented in [Pil97] that is capable to deal with stereo pairs with reasonably large baseline. The improve-

Figure 3: *Results for a 5-frames image sequence. Correct matches between first (top) and third (bottom) frame. Left: S-SVD. Centre: C-SVD. Right: S-DIST. Note that one of the matches in the left image is a wrong match that appears to be consistent with the epipolar geometry.*

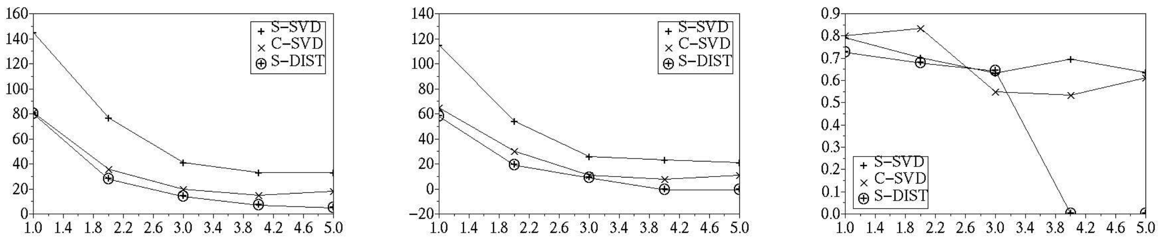


Figure 4: *Results for a 5-frames image sequence. Correspondences are computed between the first frame and each other frame in the sequence. The baseline is increasing along the sequence. Left: total number of matches detected. Centre: number of correct matches. Right: accuracy of the method.*

ment is obtained by using a more robust feature descriptor (SIFT) than the one used in the former version of the algorithm. Experimental evidence shows the better performance of the proposed version with respect the original one, and with respect a standard SIFT based point matcher.

More work is still necessary in trying to make the algorithm feasible for wide baseline matching when the images can look too much different. In our view what should be tried is: the use of a different interest point detector as the *improved* Harris point detector discussed in [MS03], and the use of SIFT distance measures different form the cross-correlation used in the current version of the S-SVD.

Acknowledgements

The authors thank Maurizio Pilu for useful discussions. This work is partially supported by European FP6 NoE aim@shape grant 506766, the INFM Advanced Research Project MAIA, the FIRB Project ASTA RBAU01877P. The video-conferencing sequences used in this work have been provided by the EU Framework-V grant (IST-1999-10044) VIRTUE (VIRtual Team User Environment).

References

[ADJ00] ALLEZARD N., DHOME M., JURIE F.: Recognition of 3d textured objects by mixing view-based and model based representations. In *Proceedings of ICPR* (2000).

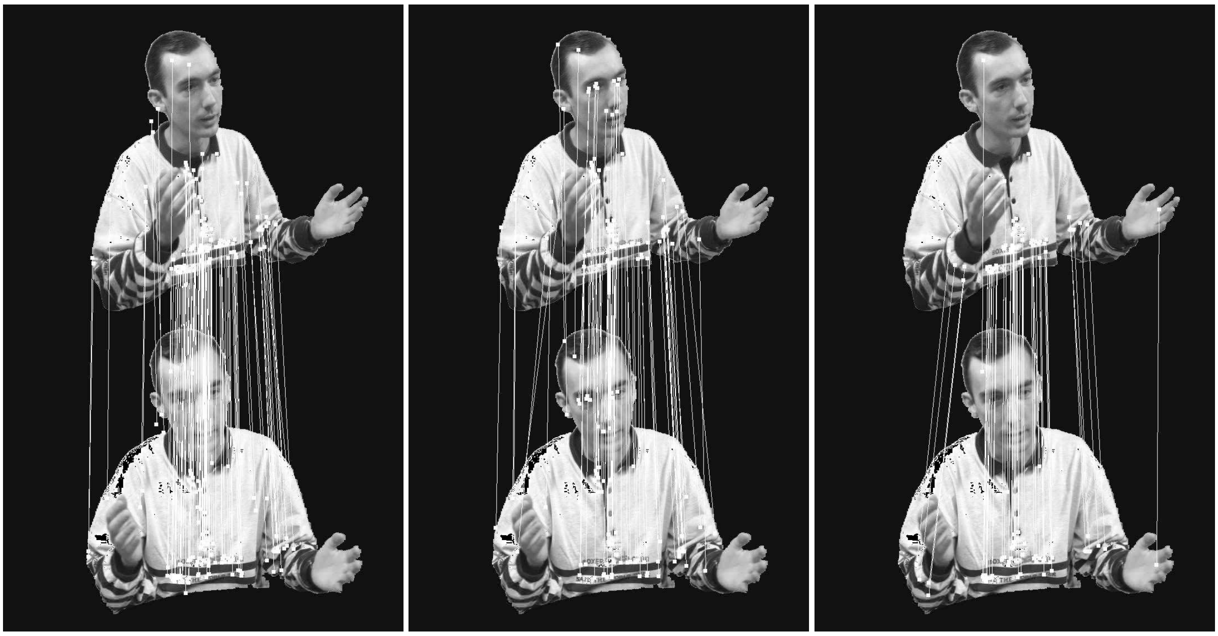


Figure 5: *Results for a 29-frames stereo sequence. Correct matches between the left (top) and right (bottom) fist frames. Left: S-SVD. Centre: C-SVD. Right: S-DIST.*

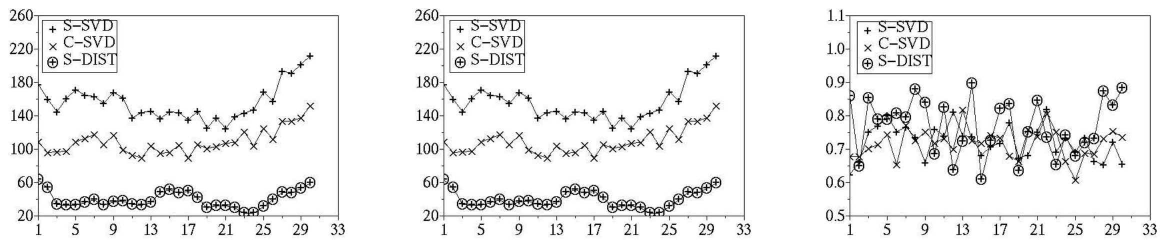


Figure 6: *Results for a 29-frames stereo sequence. The baseline is fixed for all the stereo pairs, and the correspondences are computed for each stereo frame of the sequence. Left: total number of matches detected. Centre: number of correct matches. Right: accuracy of the method.*

[Bau00] BAUMBERG A.: Reliable feature matching across widely separated views. In *Proceedings of CVPR* (2000), pp. 774–781.

[BM92] BESL P., MCKAY N.: A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 14*, 2 (1992), 239–256.

[DZLF94] DERICHE R., ZHANG Z., LUONG Q. T., FAUGERAS O.: Robust recovery of the epipolar geometry from an uncalibrated stereo rig. In *Proceedings of ECCV* (1994), Eklundh J. O., (Ed.), vol. 800 of *LNCS*, pp. 567–576.

[FA91] FREEMAN W., ADELSON E.: The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence 13*, 9 (1991), 891–906.

[FR96] FAUGERAS O., ROBERT L.: What can two images tell us about a third one? *International Journal on Computer Vision 18*, 1 (1996), 5–20.

[FTG03] FERRARI V., TUYTELAARS T., GOOL L. V.: Wide-baseline muliple-view correspondences. In *Proceedings of CVPR* (2003).

[GPK98] GEORGIS N., PETROU M., KITTLER J.: On the correspondence problem for wide angular separation of non-coplanar points. *Image and Vision Computing 16* (1998).

[GSBB03] GALUN M., SHARON E., BASRI R., BRANDT A.: Texture segmentation by multiscale aggregation of filter responses and shape elements. In *Proceedings of ICCV* (2003), pp. 716–723.

[HS88] HARRIS C., STEPHENS M.: A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference* (1988), pp. 147–151.

[HZ00] HARTLEY R. I., ZISSERMAN A.: *Multiple view geometry*. Cambrige University Press, 2000.

Figure 7: *All the matches returned for a pair with very large scene variation. Our algorithm finds more correct matches than the other two, but it is evident that it cannot still cope with so large scene variations. Left: S-SVD. Centre: C-SVD. Right: S-DIST.*

[Lin96] LINDEBERG T.: Scale-space: A framework for handling image structures at multiple scales, 1996.

[Lin98] LINDEBERG T.: *Principles for automatic scale selection*. Tech. Rep. ISRN KTH/NA/P-98/14-SE, CVAP Department of numerical analysis and computing science KTH, S-100 44 Stockholm, Sweden, August 1998.

[LM04] LU X., MANDUCHI R.: Wide baseline feature matching using the cross-epipolar ordering constraint. In *Proceedings of CVPR* (2004), vol. I, pp. 16–23.

[Low99] LOWE D. G.: Object recognition from local scale-invariant features. In *Proceedings of ICCV* (Corfú, Greece, 1999), pp. 1150–1157.

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110.

[LTAO03] LOURAKIS M., TZURBAKIS S., ARGYROS A., ORPHANOUDAKIS S.: Feature transfer and matching in disparate stereo views through the use of plane homographies. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 2 (2003).

[MCUP02] MATAS J., CHUM O., URBAN M., PAJDLA T.: Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of BMVC* (2002), pp. 384–393.

[MMRK91] MEER P., MINTZ D., ROSENFELD A., KIM D. Y.: Robust regression methods for computer vision: a review. *International Journal of Computer Vision 6*, 1 (1991), 59–70.

[MS02] MIKOLAJCZYK K., SCHMID C.: An affine invariant interest point detector. In *Proceedings of ECCV* (2002), vol. I, pp. 128–142.

[MS03] MIKOLAJCZYK K., SCHMID C.: A performance evaluation of local descriptors. In *Proceedings of CVPR* (2003), vol. II, pp. 257–263.

[Pil97] PILU M.: A direct method for stereo correspondence based on singular value decomposition. In *Proceedings of CVPR* (Puerto Rico, 1997), pp. 261–266.

[PZ98] PRITCHETT P., ZISSERMAN A.: Wide baseline stereo matching. In *Proceedings of ICCV* (1998), pp. 754–760.

[RvdB77] ROSENFELD A., VAN DER BRUG G.: Coarse-fine template matching. *IEEE Transactions on Systems, Man and Cybernetics 7* (February 1977), 104–107.

[SLH91] SCOTT G., LONGUET-HIGGINS H.: An algorithm for associating the features of two images. *Proc. Royal Society London B244* (1991), 21–26.

[SZ97] SCHMID C., ZISSERMANN A.: Automatic line matching across views. In *Proceedings of CVPR* (1997), pp. 666–671.

[TC02] TELL D., CARLSSON S.: Combining appearance and topology for wide baseline matching. In *Proceedings of ECCV* (2002), vol. I, pp. 68–81.

[TG00] TUYTELAARS T., GOOL L. V.: Wide baseline stereo matching based on local, affinely invariant regions. In *Proceedings of BMVC* (200), pp. 412–425.

[vdBR77] VAN DER BRUG G., ROSENFELD A.: Two-stage template matching. *IEEE Transactions on Computers 26*, 4 (April 1977), 384–393.

[ZW94] ZABIH R., WOODFILL J.: Non-parametric local transforms for computing visual correspodsence. In *Proceedings of ECCV* (1994), pp. 151–158.