

Robust Dense Matching using Local and Global Geometric Constraints

Maxime Lhuillier

Long Quan

CNRS-GRAVIR-INRIA

ZIRST – 655 avenue de l'Europe

38330 Montbonnot, France

Maxime.Lhuillier@inrialpes.fr Long.Quan@inrialpes.fr

Abstract

A new robust dense matching algorithm is introduced in this paper. The algorithm starts from matching the most textured points, then a match propagation algorithm is developed with the best first strategy to densify the matches. Next, the matching map is regularised by using the local geometric constraints encoded by planar affine applications and by using the global geometric constraint encoded by the fundamental matrix.

Two most distinctive features are a match propagation strategy developed by analogy to region growing and a successive regularisation by local and global geometric constraints. The algorithm is efficient, robust and can cope with wide disparity. The algorithm is demonstrated on many real image pairs and applications on image interpolation and creating novel views are also presented.

1 Introduction

Matching techniques are always one of the most important and most difficult task in computer vision. Dense matching consists of establishing a maximum number of pixel-to-pixel correspondences in two images. There have been many dense matching algorithms developed (e.g. [3, 7, 2, 1, 10, 5, 11]). The most common approach for dense matching consists of correlating small image windows along the one dimensional epipolar line within the maximum disparity. This method is often computationally costly, suitable only for small base line stereo images, and also depends heavily on the epipolar geometry which has to be provided, often by off-line calibration. Most recently from researchers working on uncalibrated vision, the epipolar geometry could be estimated on-line with robust techniques [14, 13] which tolerates the initial mismatches. However, this modern approach still fails frequently for widely separated image pairs since the initial matches are fragile and the fundamental matrix so estimated fits often to

subsets of images, not the whole image. For instance, for a clear background/foreground scene, the fundamental matrix often fits either to the cluster of points on the background or that on the foreground.

In this paper, we introduce a new approach of robust dense matching algorithm. It starts by constructing a dense matching map using a growing/propagation schema from a list of seed matches which may contain bad matches (Section 2). The matching is then regularised using the local geometric constraints encoded by planar affine applications (Section 3). After local regularisation of dense matching, the global geometric constraint encoded by the fundamental matrix is recovered and used to constraint the final propagation (Section 4).

This new approach has considerable advantages over the existing ones. It can cope with wide disparity between image pairs. The initial seed matches are more tolerant for mismatches as we do not use them to fit the global fundamental matrix, we use it only for match propagation. In some extreme cases, only one good seed match is sufficient to provoke an avalanche for the whole textured images while keeping bad seeds undeveloped. The algorithm is also computationally efficient. The final fundamental matrix estimation is also reliable as the input matches are evenly spread over the whole image space. Finally this algorithm can still applies for the non-rigid scene by dropping the final and global validation.

2 Initial match and propagation

The basic principles of the algorithm is to start from matching some points of interest which have the highest texture as seed points to bootstrap a region growing type algorithm, which then propagates the matches in the neighborhood of seed points from the most textured pixels to less textured ones. The algorithm therefore consists of two steps: seed selection and propagation.

2.1 Seed Selection and Initial Matching

Points of interest [6, 12] are naturally good seed point candidates, as points of interest are by its very definition image points which have the highest texture (i.e. the local maxima of the auto-correlation function of the signal). The ZNCC (zero-mean normalized cross-correlation) correlation measure is used for matching seeds as it is invariant to linear radiometric changes. The $ZNCC_{\mathbf{x}}(\Delta)$ at point $\mathbf{x} = (x, y)^T$ with the shift $\Delta = (\Delta_x, \Delta_y)^T$ is defined to be

$$\frac{\sum_i (I(\mathbf{x} + \mathbf{i}) - \bar{I}(\mathbf{x})) (I'(\mathbf{x} + \Delta + \mathbf{i}) - \bar{I}'(\mathbf{x} + \Delta))}{(\sum_i (I(\mathbf{x} + \mathbf{i}) - \bar{I}(\mathbf{x}))^2 \sum_i (I'(\mathbf{x} + \Delta + \mathbf{i}) - \bar{I}'(\mathbf{x} + \Delta))^2)^{1/2}}$$

where $\bar{I}(\mathbf{x})$ and $\bar{I}'(\mathbf{x})$ are the means of pixel luminances for the given window centered at \mathbf{x} .

2.2 Propagation

After obtaining the initial seed matches, it comes our central idea of match propagation from the initial seed matches. The idea is similar to the classic region growing method for image segmentation [9] based on the pixel homogeneity. Instead of using the homogeneity property, a similarity measure based on the correlation score is used. This propagation strategy could also be justified as the seed matches are the points of interest which are the local maxima of the texture, so the matches could be extended to its neighbors which have still strong texture though not a local maxima.

All initial seed matches are starting points of concurrent propagations. At each step, a match (a, A) with the best ZNCC score is removed from the current set of seed matches. Then we look for new matches in its 'match neighborhood' (see Figure 1 for its definition) and simultaneously add all new matches to the current set of seeds and to the set of accepted matches—under construction. The neighbors of pixels a and A are taken to be all pixels within the 5×5 window centered at a and A to enforce the continuity constraint of the matching results. For each neighboring pixel in the first image, we construct a list of tentative match candidates consisting of all pixels of a 3×3 window in the neighborhood of its corresponding location in the second image. Thus the displacement gradient limit should not exceed 1 pixel.

The unicity constraint of the matching and the termination of the process is guaranteed by choosing only new matches not yet accepted. Since the search space is reduced for each pixel, we use small 5×5 windows for ZNCC. Therefore, minor geometric changes are allowed and artefacts at occluding contours are limited.

We can notice that the risk of bad propagation is greatly diminished by the best first strategy over all matched seed points. Although seed selection step seems very similar to

Neighborhood of pixel a in view 1 Neighborhood of pixel A in view 2

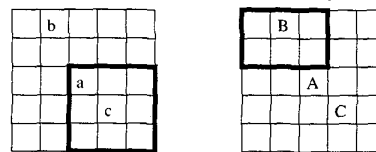


Figure 1. Definition of neighborhood $\mathcal{N}(a, A)$ of pixel match (a, A) . It is a set of matches included in the two 5×5 -neighborhood $\mathcal{N}_5(a)$ and $\mathcal{N}_5(A)$ of pixels a and A . Possible matches for b (resp. C) are in the 3×3 black frame centered at B (resp. c). The complete definition of $\mathcal{N}(a, A)$ is $\{(b, B), b \in \mathcal{N}_5(a), B \in \mathcal{N}_5(A), (B - A) - (b - a) \in \{-1, 0, 1\}^2\}$.

many existing methods [14, 13] for matching points of interest using correlation, the crucial difference is that propagation needs only to take the most reliable ones rather than taking a maximum of them. This makes our algorithm much less vulnerable to the presence of bad seeds in the initial matches. In some extreme cases, only one good match of points of interest is sufficient to provoke an avalanche of the whole textured images.

This propagation algorithm can be described as follows. The input of the algorithm is the set *Seed* of the current seed matches, the set is implemented by a heap data structure for fast selection of the best match. The output is an injective displacement mapping *Map*.

Let $s(\mathbf{x}) = \max\{|I(\mathbf{x} + \Delta) - I(\mathbf{x})|, \Delta \in \{(1, 0), (-1, 0), (0, 1), (0, -1)\}\}$ be an estimate of the luminance roughness for the pixel at \mathbf{x} , which is used to stop propagation into insufficiently textured areas a with $s(a) < t$ where $t = 0.01$ and $0 \leq I(a) \leq 1$.

Input: *Seed*

Output: *Map*

```

while Seed  $\neq \emptyset$  do
  pull the best match  $(a, b)$  from Seed
  Local  $\leftarrow \emptyset$ 
  (Store in Local new candidate matches)
  for each  $(c, d)$  in  $\mathcal{N}(a, b)$  do
    if  $(c, *)$  and  $(*, d)$  not in Map and
        $s(c) > t, s(d) > t$  and  $ZNCC(c, d) > 0.5$ 
    then store match  $(c, d)$  in Local
    end-if
  end-for
  (Store in Seed and Map good candidate matches)
  while Local  $\neq \emptyset$  do
    pull the best match  $(c, d)$  from Local
    if  $(c, *)$  and  $(*, d)$  not in Map
    then store match  $(c, d)$  in Map and Seed
    end-if
  end-while
end-while

```

The complexity of the algorithm is $O(n \log(n))$, where n is the final number of matched pixels. Notice that it is output sensitive, only dependent on the number of final matches and independent of disparity bound.

3 Check using local geometric constraints

The disparity map obtained from the propagation may still be corrupted and irregular. We assume that the scene surface is smooth enough to be approximated by small planar patches. Thus, the dense matching can be regularised by locally fitting planar patches encoded by homographies. The construction of the matched planar patches is described as follows.

The first image is initially subdivided into small regular grid. For each square patch, we obtain all matched points of the square from the dense disparity map. A plane homography should be tentatively fitted to these matched points of the square to look for potential planar patches. Because a patch is rarely a perfect planar facet except for manufactured objects, the putative homography for a patch can not be estimated by standard least squares estimators. The Random Sample Consensus (RANSAC) method [4] is used for robust estimation.

In practice, the stability of the homography fitting decreases with the patch size. Our compromise between patch grid resolution and stability fitting is to fit a planar affine application (which counts only 6 d.o.f instead of 8 d.o.f of homography) in 8×8 -pixel squares.

4 Estimate the global geometric constraints

So far, only local geometric constraints are used for matching. The global geometric constraint encoded by the fundamental matrix for the rigid scene should also be integrated. The most popular strategy is to recover it at the very beginning of the sparse matching [14, 13] within a random sampling framework. There are two ways of integrating the global constraint for our approach. The first is constrained propagation which consists in growing matches if they satisfy the epipolar constraint, while the second is unconstrained propagation. The advantage of constrained propagation is that the bad propagation might be stopped earlier, but the domain of propagation is considerably reduced; Even more seriously, the fundamental matrix estimated at this step with a robust method often tend to fit points in a subset of images. We adopt therefore the unconstrained propagation and the global constraint is only imposed by the following constrained propagation, after the local regularisation which gives widely spread corresponding points over the whole image range.

The most direct approach to estimate the fundamental matrix is from a list of pixel matches (x, y) obtained from

each local planar affine application A by $y = Ax$ where x is the middle of the corresponding square. These matches are evenly spread in image space (contrarily to interest points) and the fundamental matrix is robustly estimated by RANSAC from them. The obtained matching consensus and fundamental matrix are finally improved by applying the M-Estimator proposed in [13] and implemented by an iterated re-weighted least square [15].

5 Experimentations

The matching algorithm described above is experimented on many image pairs.

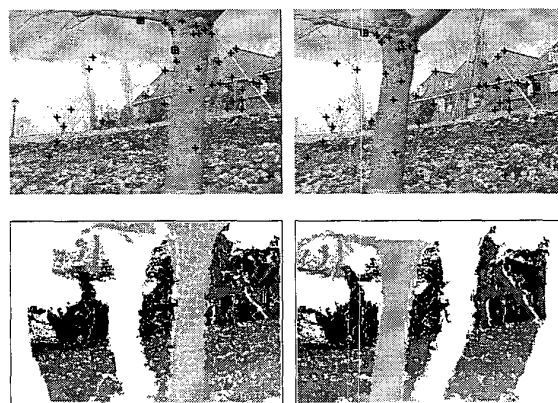


Figure 2. Top: initial seed matches. Bottom: disparity after unconstrained propagation.

5.1 Stability of region growing matching

We first compare the dense matching map by different seed selections for the first and twentieth image of the mpeg “flower garden” sequence (our matching use only two images of the sequence). The pair of “flower garden” images shown on the top row in Figure 2 is very difficult to match by classical correlation or dynamic programming as the disparity is too strong and the ordering constraint along the epipolar lines is obviously violated.

The seeds are Harris points [12] matched by 11×11 correlation ($ZNCC > 0.8$) and bidirectional consistency [5]. They are superimposed on the original pair as shown in Figure 2. Naturally, there exists still false matches marked by a square instead of a cross for good matches. The disparity map of dense matching after region growing is shown on the bottom of Figure 2. It takes 2.5 seconds by an UltraSparc 333MHz for image size of 360×240 .

We then manually select four seed matches from the previous ones and show them on the left of Figure 3. The resulting match growing is display on the right of Figure 3.

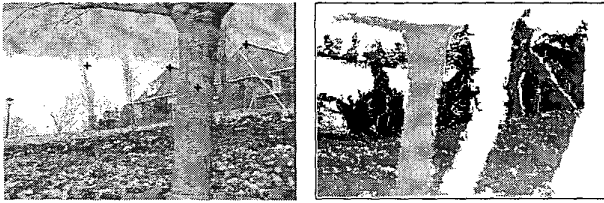


Figure 3. Left: Four manually selected seed matches. Right: the resulting disparity map.

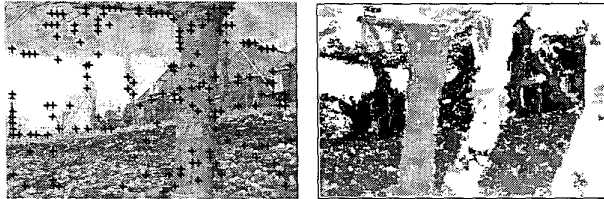


Figure 4. Left: 4 selected seeds and 158 false matches. Right: the resulting disparity map.

Each seed match is sufficient to provoke an avalanche of correct matches in each of the four isolated and textured areas. The matched areas cover roughly the same surface as that obtained with the automatic seeds in Figure 2 and 78% of matched areas are common between two methods.

To these manually selected matches, 158 false matches with good normalized correlation score ($ZNCC > 0.9$) are added. Still 70% of the matched areas are common with the automatic seeds in Figure 2.

From the above experiments, we see that the match propagation is robust w.r.t. the initial seed selection.

5.2 Comparing with Image-Matching

We show the advantage of unconstrained propagation to calculate the fundamental matrix for the previous big camera movement, by comparing our results with those of Image-Matching [14]. Image-Matching estimates a fundamental matrix F_i from least median square using seed matches by correlation and relaxation. The default parameters of Image-Matching are used. In particular, its correlation step is more conservative than ours because big 15×15 windows are used with the same ZNCC threshold (0.8). It follows that there are no seed matches in the flowers because the geometric distortion between images is too strong (top of Figure 5), and the resulting F_i is too bad to allow constrained propagation in flowers on the bottom right of the trunk (bottom left of Figure 5) in spite of sub-pixel accuracy for seeds. Our dense matching is only pixel accuracy, but unconstrained propagation from the same seeds fills the full and common regions in the two images. It follows that

our resulting fundamental matrix F_p is better in these areas than F_i and it is shown by allowing constrained propagations everywhere (bottom right of Figure 5).

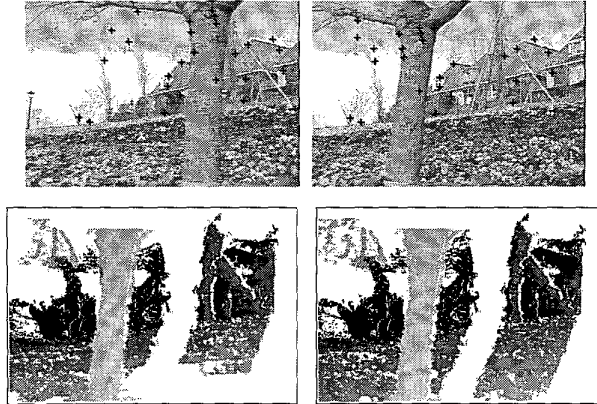


Figure 5. Top: Image-Matching seed matches. Bottom left: F_i -constrained propagation. Bottom right: F_p -constrained propagation.

5.3 Applications

We have used the result of our dense matching algorithm and fundamental matrix estimation for image interpolation [8] (left Figure 6 and Figure 8) and surface reconstruction (right Figure 6 and Figure 7). In both cases, the rendering is improved with a mesh which explicitly models half occluded areas and gradient edges (automatic building). As expected, the precision of the reconstruction decreases with the depth (Figure 6). In Figure 7, the face profile seems to be correct but the hairs are difficult to match. The algorithm parameters are the same for all examples, except for the different faces in Figure 8 under similar conditions (provided by [11]): the regions of interest for seed matches are reduced to limit bad seeds. Many examples are available at our Web site <http://www.inrialpes.fr/movi/pub/Demos>.

6 Conclusion and future work

We have presented a new method of dense matching algorithm. The main contributions are twofold. A match region growing technique inspired from the region growing for image segmentation allows to match important areas across two images. Both local and global geometric constraints are integrated. Unlike the existing methods, the global constraint is recovered and used after local regularisation. The method has been experimented on many real image pairs including wide disparity and the application for image interpolation and novel view synthesis is convincing. Actually we are investigating different multiresolution

strategies, smoothness constraints for propagation, and also work on multiples images.

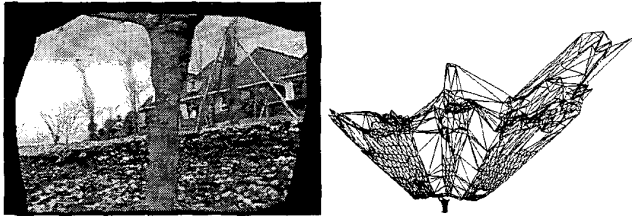


Figure 6. Interpolation and surface reconstruction for the "flower garden" image pair.



Figure 7. Herve image pair and the reconstructed surface.

References

- [1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *IJCV*, 2:283–310, 1989.
- [2] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.
- [3] U.R. Dhond and J.K. Aggarwal. Structure from stereo – a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.
- [4] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *GIP*, 24(6):381 – 395, 1981.
- [5] P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve discontinuities. In *IJCAI*, 1991.

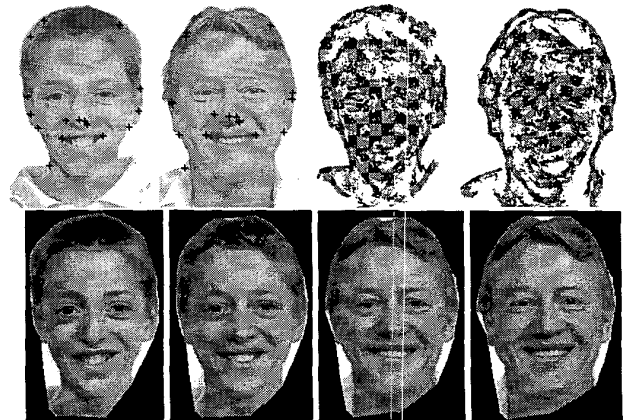


Figure 8. Top: seeds and propagation for two faces. Bottom: the resulting morphing.

- [6] C. Harris and M. Stephens. A combined corner and edge detector. In *AVC*, pages 147–151, 1988.
- [7] A. Koschan. What is new in computational stereo since 1989 : A survey on stereo papers. Technical report 93-22, University of Berlin, 1993.
- [8] M. Lhuillier and L. Quan. Image Interpolation by Joint View Triangulation. In *CVPR99*, V.2, pages 139–145.
- [9] O. Monga. An optimal region growing algorithm for image segmentation. *IJPRAI*, 1(3):351–375, 1987.
- [10] G.P. Otto and T.K. Chau. A region-growing algorithm for matching of terrain images. *IVC*, 7(2):83–94, 1989.
- [11] G.M. Quénot. The orthogonal algorithm for optical flow detection using dynamic programming. In *International Conference on Acoustics, Speech and Signal Processing*, 1992.
- [12] C. Schmid, R. Mohr and C. Bauckhage. Comparing and Evaluating Interest Points. In *ICCV98*, 1998.
- [13] P.H.S. Torr and D.W. Murray. The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix In *IJCV*, 24(3):271–300, 1997.
- [14] Z. Zhang, R. Deriche, O. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. In *AI*, 78(1-2):87–119, 1995.
- [15] Z. Zhang. Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting. Technical Report 2676, INRIA, 1995.