Match Propagation for Image-Based Modeling and Rendering

Maxime LHUILLIER and Long QUAN

Department of Computer Science, HKUST, Clear Water Bay, Kowloon, Hong Kong
maxime,quan@cs.ust.hk

Abstract

This paper presents a quasi-dense matching algorithm between images based on match propagation principle. The algorithm starts from a set of sparse seed matches, then propagates to the neighboring pixels by the best-first strategy, and produces a quasi-dense disparity map. The quasi-dense matching aims at broad modeling and visualization applications which rely heavily on matching information. Our algorithm is robust to initial sparse match outliers due to the best-first strategy; It is efficient in time and space as it is only output sensitive; It handles half-occluded areas because of the simultaneous enforcement of newly introduced discrete 2D gradient disparity limit and the uniqueness constraint. The properties of the algorithm are discussed and empirically demonstrated. The quality of quasi-dense matching are validated through intensive real examples.

Key words: quasi-dense matching, stereo vision, image-based modeling and rendering.

1 Introduction

One natural approach to IBMR (Image-based modeling and rendering) is classical computer vision based 3D reconstruction with rendering by texture mapping. Arbitrary views of the scene can be synthesized by reprojecting the reconstructed 3D model. Typical examples of this approach are [11, 33, 29, 18] among others. More recent work has shown that multi-image matching constraints —the fundamental matrix for two views and the trifocal tensor for three can be used to synthesize new views without explicit 3D reconstruction. Laveau and Faugeras [20] use fundamental matrices [41, 42, 10, 16] and Avidan and Shashua [2] use trifocal tensors or 3-view geometry [38, 15, 34]. These matching tensor methods are essentially equivalent to implicit 3D reconstruction methods, and as the explicit reconstruction methods, they require rigid 3D scenes. In computer graphics, image-based rendering is often viewed as a problem of interpolation from a collection of images, inspired by techniques which generate smooth transitions between reference images by simply interpolating each pixel from the first to the second image value. Chen et al. [7] popularized the idea of direct pixel by pixel interpolation, however they originally assumed that the pixel correspondences in the basis images were given, as the basis images were computer rendered. Seitz and Dyer [37] investigate view interpolation, but are mainly concerned with physically-valid view generation via rectification of a perspective image pair following the linear combination method developed for object recognition of affine images. Like the reconstruction-based methods, they also aim at rendering rigid scenes. A more abstract formulation on which a large amount of work has been based [27, 21, 13] is light field or plenoptic function. This models all sets of rays seen from all points, considering each image as a set of rays. Image-based rendering is then about reconstructing this plenoptic function from the available images. The major challenge is to establish pixel correspondences between images particularly in sparse sampling cases.

In this paper we describe a new matching algorithm which produces a quasi-dense pixel matching between images. Our algorithm is targeted for automatic image-based rendering applications as the traditional sparse matching is insufficient and general dense stereo algorithms are not sufficiently robust except for pre-calibrated cameras in pre-settled environment [29, 18]. The basic idea of the new matching algorithm is to start from a set of sparse matches as seed points, then propagate the seed points to neighboring pixels using a region growing technique. The result is a quasi-dense disparity map. Early versions of this work appeared in the conference papers [22, 23, 25]. The paper is organized as follows. Section 2 describes the motivation and rationale of quasi-dense matching. Section 3 develops the algorithm based on match propagation. Section 4 discusses and empirically demonstrates the nice properties of the propagation. Section 5 presents the related work. Application examples are shown in Section 6. Finally, some concluding remarks and future directions are given in Section 7.

2 Motivation and rationale

Establishing correspondences is one of the fundamental problems of computer vision. There has been a great amount of matching algorithms developed in different contexts. Traditionally, abundant work has been devoted to dense matching techniques in carefully calibrated stereo case (e.g. [32, 8, 31, 12, 19, 30, 17, 9, 4]). Stereo matching algorithms reduce the search space to 1D along epipolar lines, but mismatches are still frequent around occluding and texture-less areas. Dense matching based on optical flow computation (e.g. [1, 3, 40, 35]) is restricted to closer spaced images and assumes a smooth and well-behaved intensity function. Dense matching is generally ill-posed due to the aperture problem and regularization based on smoothness constraint is always necessary [3]. It is therefore practically very difficult to expect reliable dense matching. Dense matching is also sometimes unnecessary for view synthesis as discussed by [36] for computational efficiency: homogeneous areas are difficult to match, but usually do not create visual artifacts if their boundaries are correctly matched.

Recently developed sparse matching of interest points [14] using robust techniques [10, 16, 41, 42] based on the geometric constraint encoded by the fundamental matrix has been very successful. It is becoming a standard for camera geometry estimation, self-calibration for both calibrated and uncalibrated images [42, 41]. However sparse matching is insufficient for visualization purpose as it merely consists of a sparsely distributed cloud.

We propose an in between quasi-dense pixel matching approach in which dense matches are computed where possible, i.e. in sufficiently textured areas. The key observation is that sparse matching has been shown to be efficient due to its highly discriminant nature of the points of interest [26, 14], which are by definition the local maxima of auto-correlation. Even though they are not local maxima, the neighborhood around the points of interest still shows high auto-correlation, i.e. high textureness. Therefore these areas could still be reliably matched as there are sufficient discriminant information in the images. Match propagation takes advantage of this observation to densify the matching by expanding the actual established sparse matching to their immediate neighboring areas simultaneously in two images. This principle is similar to a region-growing segmentation algorithm in an individual image, but with propagation operating in two images. It is suitable for image-based rendering application as it densely matches the textured areas for reducing rendering artifacts, but avoids computing unreliable uniform and occluding areas. We believe that quasi-dense matching is the most appropriate choice for broader modeling and visualization applications by filling the gap between sparse and dense matching to overcome the insufficiency of sparse matching for visualization based application and fragile dense matching. We will first develop our quasi-dense matching algorithm based on match propagation, then discuss its interesting properties and related work.

3 Quasi-Dense Matching by Best-First Propagation

3.1 Seed selection and initial matching

We start with a traditional sparse matching algorithm between two images for the points of interest detected in each image [41, 42]. Points of interest are naturally reliable two-dimensional point features [14, 39], therefore they can handle cases of large disparity.

The Zero-mean Normalized Cross-Correlation (ZNCC) is used to match points of interest in two images, as it is invariant to local linear radio-metric changes. The correlation at point $\mathbf{x} = (x, y)^T$ with shift $\Delta = (\Delta_x, \Delta_y)^T$ is defined to be

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

where $\bar{I}(\mathbf{x})$ and $\bar{I}'(\mathbf{x})$ are the means of pixel luminances for the given windows centered at $\mathbf{x}$. After the correlation step, a simple cross-consistency check [12] is used to retain a one-to-one matching between two images. The cross-consistency check consists of correlating pixels of the first to the second and inversely by correlating those of the second to the first image, only best matches consistent in both directions are retained. From many examples, we have found that a good compromise is to reject definitively a match if $ZNCC < 0.8$ using 11×11 windows. The choice of the region of interest will be described in the experiment Section 6.

3.2 Propagation

All initial cross-checked sparse matches are sorted by decreasing correlation score as seed points for concurrent propagations. At each step, the match $(\mathbf{x}, \mathbf{x}')$ composed of two corresponding pixels $\mathbf{x}$ and $\mathbf{x}'$ with the best ZNCC score is removed from the current set of seed matches. Then we search for possible new matches in the immediate spatial neighborhood $\mathcal{N}(\mathbf{x}, \mathbf{x}')$ precisely defined below. The ZNCC is still used for match propagation as it is more conservative than others such as sum of absolute or square differences in uniform regions, and is more tolerant in textured areas where noise is important. We also *simultaneously* enforce a smoothness matching constraint called "discrete 2D disparity gradient limit" precisely defined below, the uniqueness constraint and a confidence measure defined below on these neighboring pixels to be considered as potential new matches. The matching uniqueness and termination of the process are guaranteed by choosing only new matches that have not yet been selected.

Discrete 2D disparity gradient limit The usual 1D disparity gradient limit along the epipolar lines has been widely used in rectified stereo matching for disambiguation [32, 5]. We propose here a 2D extension, the discrete 2D disparity gradient limit, to deal with uncalibrated pair of images including rigid scenes with inaccurate epipolar constraint and non-rigid scenes. This limit is said to be "discrete" because only integer values are allowed for disparities. This allows us to impose directly the uniqueness constraint while propagating.

Let $\mathcal{N}(\mathbf{x}) = \{\mathbf{u}, \mathbf{u} - \mathbf{x} \in [-N, N]^2\}$, $\mathcal{N}(\mathbf{x}') = \{\mathbf{u}', \mathbf{u}' - \mathbf{x}' \in [-N, N]^2\}$ denote all $(2N+1) \times (2N+1)$ neighboring pixels of pixels $\mathbf{x}$ and $\mathbf{x}'$. The possible matches limited by the discrete 2D disparity gradient are given as

$$\mathcal{N}(\mathbf{x}, \mathbf{x}') = \{(\mathbf{u}, \mathbf{u}'), \mathbf{u} \in \mathcal{N}(\mathbf{x}), \mathbf{u}' \in \mathcal{N}(\mathbf{x}'), ||(\mathbf{u}' - \mathbf{u}) - (\mathbf{x}' - \mathbf{x})||_\infty \leq \epsilon\}.$$

and are illustrated in Figure 1 for $N = 2, \epsilon = 1$. Our choice is to impose the most conservative non-zero disparity gradient limit, i.e., $\epsilon = 1$ for integer pixel coordinates $\mathbf{u}, \mathbf{u}', \mathbf{x}, \mathbf{x}'$. In the case of rectified images of a rigid scene, the usual 1D gradient disparity limit for two neighboring pixels $\mathbf{x}$ and $\mathbf{u} = \mathbf{x} + (1, 0)$ in the same line is directly deduced from our 2D one: writing disparities $d(\mathbf{x}) = \mathbf{x}' - \mathbf{x}$ and $d(\mathbf{u}) = \mathbf{u}' - \mathbf{u}$ of matches $(\mathbf{x}, \mathbf{x}')$ and $(\mathbf{u}, \mathbf{u}')$, we obtain $|d(\mathbf{x}) - d(\mathbf{x} + 1)| \leq 1$. The newly introduced discrete 2D disparity gradient limit therefore naturally extend the usual 1D disparity gradient limit for uncalibrated image pairs with inaccurate epipolar constraint and for non-rigid scenes.

A minimal size of neighborhood $\mathcal{N}$ should be used to limit bad matches at the occluding contours. The smallest neighborhood size that makes sense of the definition of gradient limit is therefore $N = 2$, i.e. the 5×5 neighborhood.

Confidence measure There is a wide variety of definitions of confidence measure in the literature and its choice depends on the applications. We use a simple difference-based confidence measure as

$$s(\mathbf{x}) = max\{|I(\mathbf{x} + \Delta) - I(\mathbf{x})|, \Delta \in \{(1, 0), (-1, 0), (0, 1), (0, -1)\}\}.$$

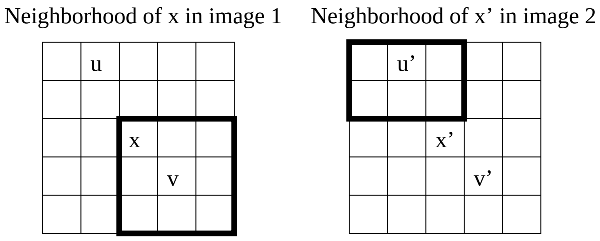

Figure 1: Possible matches $(\mathbf{u}, \mathbf{u}')$ and $(\mathbf{v}, \mathbf{v}')$ around a seed match $(\mathbf{x}, \mathbf{x}')$ come from its 5×5-neighbor $\mathcal{N}(\mathbf{x})$ and $\mathcal{N}(\mathbf{x}')$ as the smallest size for discrete 2D disparity gradient limit. The match candidates for $\mathbf{u}$ (resp. $\mathbf{v}'$) are within the 3×3 (black framed) centered at $\mathbf{u}'$ (resp. $\mathbf{v}$).

This is less restrictive than Moravec operator [28] and allows propagation to walk and match along edges in spite of the aperture problem while avoiding matching uniform areas. We forbid propagation in too uniform areas with $s(\mathbf{u}) \leq t, t = 0.01$ assuming that $0 \leq I(\mathbf{u}) \leq 1$. More conservatives measures such as those suggested in optical flow [3] are also possible.

3.3 Propagation algorithm

The 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 with a heap data structure for both fast selection of the ZNCC-best match and incremental additions of seeds. The output is an injective displacement mapping *Map*.

Input: *Seed*
Output: *Map*
$Map \leftarrow \emptyset$
while $Seed \neq \emptyset$ **do**
  pull `the ZNCC-best match` $(\mathbf{x}, \mathbf{x}')$ `from` *Seed*
  $Local \leftarrow \emptyset$
  (Store in Local new candidate matches enforcing the disparity gradient limit)
  for each $(\mathbf{u}, \mathbf{u}')$ **in** $\mathcal{N}(\mathbf{x}, \mathbf{x}')$ **do**
    if `s(u)>t and s(u')>t and ZNCC(u,u')>z`
    then `store (u,u') in` *Local*
    end-if
  end-for
  (Store in Seed and Map final matches enforcing the uniqueness constraint)
  while $Local \neq \emptyset$
    pull `the ZNCC-best match (u,u') from` *Local*
    if `(u,*) and (*,u') are not in` *Map*
    then `store (u,u') in` *Map* `and` *Seed*
    end-if
  end-while

end-while

The parameter values by default of t and ϵ have been suggested when defined. The complexity of this algorithm in time is $O(nlog(n)))$ with n the final number of matched pixels, assuming that the number of initial seeds is negligible. Our algorithm depends only on the number of final matches and is independent of any disparity bound: it is therefore output sensitive. The memory complexity is linear in the image size.

As the search space for potential correspondents for a given pixel reduces to a very small region of 3×3 as shown in Figure 1, the matching criterion is relaxed to a smaller window and weaker correlation score than the seed points. We found by experiments that a rejection takes place for a match having $ZNCC < z, z = 0.5$ within a 5×5 window. There are two benefits for smaller ZNCC window size: minor perspective distortions are accepted and matching artifacts are limited around occluding contours.

4 Match propagation properties

The match propagation algorithm has many nice properties that we examine. For illustrative purpose, we use the 1st and 20th frames of the public domain *flower garden* MPEG sequence. This image pair is difficult for many algorithms as there is a large occluding tree in the foreground of the scene and disparities might be large. Many matching algorithms produce high rates of outliers in the half-occluded area and it is particularly difficult for Dynamic Programming-like approaches as the ordering constraint is not observed.

Robustness w.r.t. false seed matches The robustness and stability of this algorithm based on best-first propagation are considerably improved by the global best-first strategy w.r.t. the sparse matching. Though the seed selection step seems very similar to many existing methods [41, 42] for matching points of interest using correlation, the key difference is that the propagation can rely on only a few most reliable ones rather than taking a maximum of them. This makes the algorithm much less vulnerable to the presence of match outliers in the seed selection stage. The risk of bad propagation is significantly diminished for two reasons: the bad seed points have no chance to be developed if they are not ranked on top of the sorted list; the propagation by bad seed points is stopped very quickly due to lack of consistency in its neighborhood even if the bad seed points might occasionally ranked high in the list. In some extreme cases, only one single good match of points of interest is sufficient to provoke an avalanche of the whole textured images while keeping all other seeds points including bad ones undeveloped.

Figure 2 shows the disparity maps obtained from different selections of seed points to illustrate the stability of propagation. The first example on the left column is produced using automatic seed points as sparse matching of points of interest described above. The set of automatic seed points still contains match outliers marked as a square instead of a cross for a good

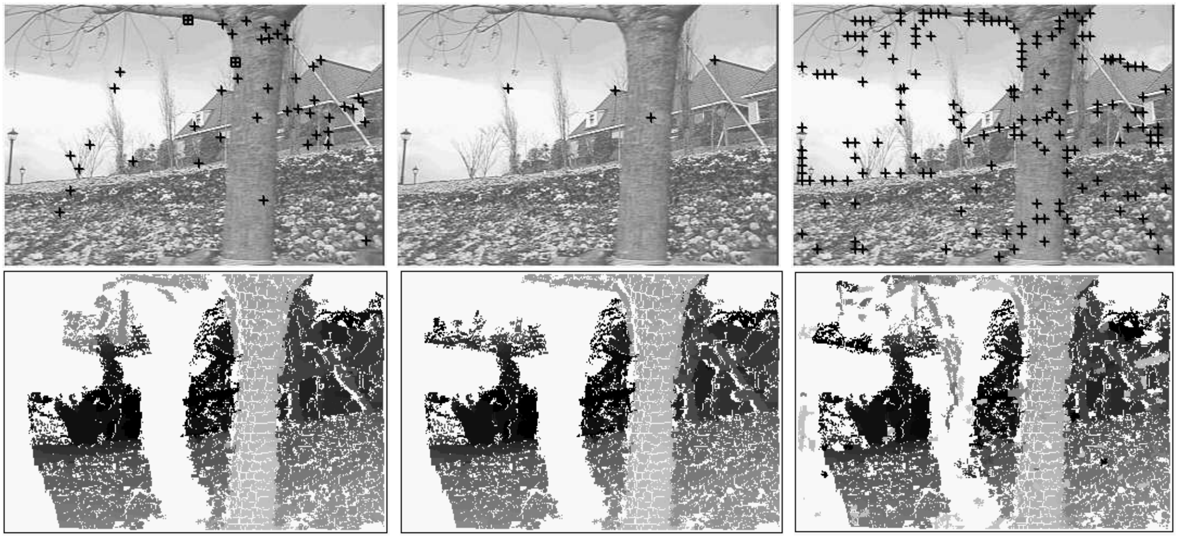

Figure 2: The disparity maps produced by propagation with different seed points and without the epipolar constraint. Left column: automatic seed points with the match outliers marked with a square instead of a cross. Middle column: four seed points manually selected. Right column: four seed points manually selected plus 158 match outliers with strong correlation score $ZNCC > 0.9$).

match. The second example in the middle is produced with four seed points, manually selected from the set of automatic seed points. Each seed match is sufficient to provoke an avalanche of correct matches in each of the four isolated and textured areas. The whole matched areas cover roughly the same surface as that obtained with the automatic seeds and 78% of matched areas are common between the two results. The third example on the right is produced with four seed points as in the second example plus 158 more seed points which are all match outliers but still have strong correlation score ($ZNCC > 0.9$). This very severely corrupted set of seed points still gives 70% of the matched areas common to the automatic case. This robust scenario of good seeds developing faster than bad ones is more dominant in matched areas than occluded regions.

Further we notice from these examples that minor geometric distortion is well tolerated in the flowers thanks to the small 5×5 ZNCC-window size during propagation. In the untextured sky area, the propagation is stopped by the confidence measure. The near-periodic thin unmatched gaps on the trunk illustrate the effect of the uniqueness constraint.

Stability in low textured scenes For low textured images such as typical polyhedric scenes, one could have expected that the matched areas are to be reduced to a small neighborhood of seeds due to poor textures, but very interestingly, the propagation grows nicely along the gradient edges and we may show that the distance along edges covered by propagation in both images is similar if the perspective distortion is moderate. Figure 3 illustrates an interesting consequence of this property by using two views of a wooden house. A single manually selected seed propagates globally to the whole images and covers the most part of the normally

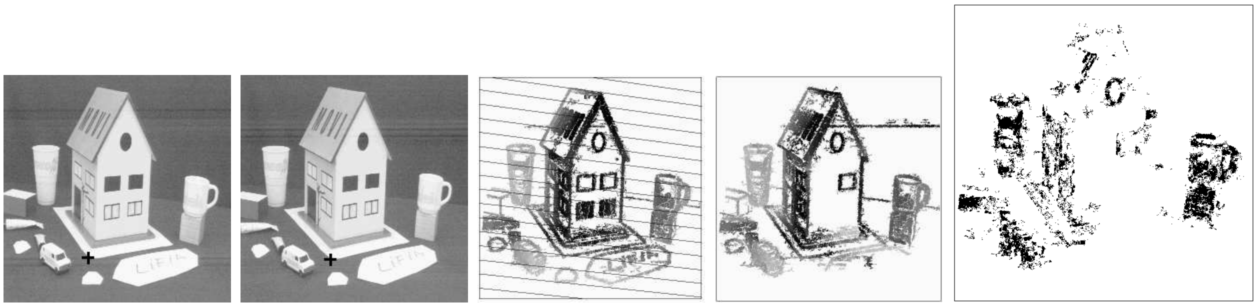

Figure 3: Examples of propagation for low textured images. From left to right: (1) two images with low geometric distortion of a small wood house, (2) the disparity map automatically produced with the epipolar constraint, (3) the disparity map produced by a single manual seed on the bottom without the epipolar constraint, (4) the common matching areas in black between the two maps in (2) and (3).

propagated results from many seeds distributed over the whole image. This global stability of propagation for low textured images is very good for interpolation or morphing applications.

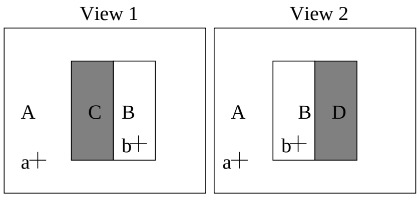

Figure 4: Two views of a scene with background A, foreground B and half occluded areas C and D. Assume that correct matches within A and B have better scores than bad ones within C and D. A and B are first filled in by propagation from seed a and b before the algorithm attempts to grow in C or D. Once A and B are matched, the procedure is stopped by the uniqueness constraint at the boundary of C in the first view (resp. D in the second view) because the corresponding boundary in the second view (resp. the first one) encloses an empty area.

Handling half-occluded areas The algorithm provides satisfactory results in half-occluded areas, mainly due to the simultaneous enforcement of the global best-first strategy and the uniqueness constraint. The principle can be illustrated with the help of Figure 4 in which we assume a foreground object B over a background A, and the half-occluded areas C and D. The global best-first propagation strategy first accepts matches with the best correlation scores before trying the majority of matches with mediocre scores. As the foreground A and background B are both visible, it is expected that they are matched before bad matches for pixels in half-occluded areas C and D are tried. Consequently, the propagation stops in the half-occluded areas because it is always stopped by the uniqueness constraint.

We show in Figure 5 that with at least 4 seed points, the disparity maps obtained are excellent in handling the occluded areas. When we remove one important seed point from the

Figure 5: The disparity maps produced by propagation with different seed points and without the epipolar constraint. Left and Middle columns: four manually selected seed points marked by a cross between the 1st and the 20th frame of the flower garden. Right column: remove one manual seed located on the front tree. It has more match outliers in the occluded regions.

front tree (which plays the role of region B in Figure 4), the tree is not matched at all as expected, many match outliers have invaded into the background area.

Another example including half occluded areas for a thin object is shown in Figure 6: a 2-3 pixel width electric post and its disparity map by propagation with and without the epipolar constraint. These fine details and their backgrounds have been successfully matched. As in the previous case, the usual fattening artifact around the occluding contours is limited because of the small 5×5 ZNCC-window and local propagation.

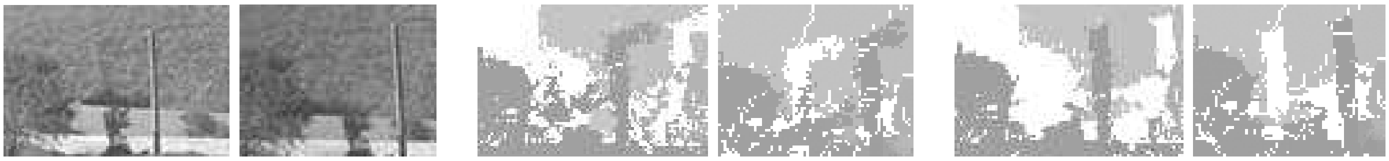

Figure 6: Left pair: two sub-images of an electric post. Middle and Right pairs: the disparity maps without (middle) and with (right) the epipolar constraint.

5 Related work

Imposing simultaneous matching constraints Matching constraints [8, 19] such as uniqueness, limit on disparity, disparity gradient and ordering constraints are always necessary to reduce matching ambiguity. These constraints are often implemented in stereo algorithms along the corresponding epipolar lines.

Uniqueness constraint is often imposed by cross-consistency check i.e. by correlating pixels of the first to the second image and inversely by correlating those of the second to the first image; only the best matches consistent in two ways are retained. The error rate by cross-consistency check is low [12], but the resulting disparity map is less dense [18] unless multiple resolutions or additional images are used [12]. An alternative consists of evaluating a set of possible correspondents in the second image for each pixel in the first image, final correspondences are established using relaxation techniques like the PMF algorithm [32] or searching for disparity components [4]. However, these methods sequentially impose limits on disparity gradients and uniqueness constraint. In our approach we impose them simultaneously. This considerably improves the matching results and allows the efficient handling of half-occluded areas. Unlike stereo matching algorithms working along 1D epipolar lines, we extended the definition of 1D disparity gradient limit to the discrete 2D disparity gradient limit which naturally handles uncalibrated images including rigid scenes with inaccurate epipolar geometry and non-rigid scenes.

Reducing search space by adding direct limit on disparity greatly improves the performance of the majority of existing methods, but the result and complexities of our method are respectively much less sensitive and independent to this constraint.

Using a best-first match-growing strategy A related region-growing algorithm was previously introduced in the photogrammetry domain by [31] with the "Gotcha" (Gruen-Otto-Chau) ALSC—Adaptative Least Square Correlation, algorithm designed to match two SPOT images. Deformable windows and patch-to-patch propagation are used instead of the disparity constraints and pixel-to-pixel propagation strategy. The main advantage of this approach is that the matching can reach sub-pixel accuracy from ALSC patch optimization. However, this patch based optimization and propagation are the sources of two drawbacks: first, a uniqueness constraint can no longer be defined for the overlapping patches and second, large window sizes than ours is unavoidable for stable adaptive least squares, especially if gray level distortions are considered.

Also, the patch propagation could not deal with fine texture details (like electric posts) unless optimization is done for each pixel. The optimization process suffers from over-parameterization when gray level distortions are considered and is hardly workable for matching different scenes. The most serious shortcoming is poor performance around occluding contour points due to lack of a uniqueness constraint and the larger window size.

Independent to our paper [22, 23], another pixel-to-pixel propagation presented as volumetric matching was proposed by [6]. However, this approach does not use the best-first strategy and the uniqueness constraint. In addition, the occlusion problem is ignored and the seed points are not localized in the highest texture areas. A progressive scheme for stereo matching was also introduced more recently by [43]. It starts from robustly matched interest points [42], then densifies the matching by using a growing principle. It considers simultaneously multiple current matches and propagates in a larger area instead of one seed match in a very small predefined area in our approach. This tends to produce smoother disparity map, but more outliers for half-occluded areas. It is also unknown of its performance for non-rigid scenes.

6 Application examples

We show in this section some applications of creating in-between views by interpolation based on the quasi-dense matching algorithm presented above, with the setting of propagation parameters proposed in Section 3.

Triangulating in image planes is always necessary not only to remedy matching sparseness, but also primarily to approximate images for rendering efficiency. Traditional independent Delaunay triangulation performed on each individual image gives a good approximation when occluded areas are negligible in the rendering view-field, but becomes insufficient when the occlusions are apparent as illustrated in the garden flower sequence. To overcome the shortage of the independent triangulation, a Joint View Triangulation that triangulates simultaneously two images while keeping matching information coherence in images has been introduced and implemented [23, 24]. It is built on the re-sampled quasi-dense matches. Also a pseudo-painter's algorithm based on the joint view triangulation has also been described [23] to render novel view. It first draws unmatched triangular patches before matched ones and an heuristic drawing order based on triangle distortion for unmatched triangles and the disparity for matched ones is also used. We show four examples of rigid scenes in Figure 7 and two examples of non-rigid scene in Figure 8.

For all rigid scenes in Figure 7, the search area for seed matches is big, 40% left and right and 20% top and bottom of image size. The fundamental matrix is estimated from a first and epipolar unconstrained propagation using robust statistics to cope with sparse outliers [25, 41, 42], and a second and epipolar constrained propagation is finally applied. The rigid scenes all contain quite important half-occluded areas. The rendering results are globally convincing mainly due to the good behavior of the propagation at occlusions and the strength of the disparity gradient limit, though there are still small visible artifacts due to finite approximation of the joint triangulation around the occluding contours. The third example is the most difficult, especially on the top left corner where the texture drastically changes for different layers of branches. The computation time for each epipolar constrained propagation are respectively of 4.3, 8.1, 2.6 and 1.3 seconds with a Pentium III Mobile 500Mhz to obtain 177000, 168000, 61370 and 42835 matches for these examples of image size of $768 \times 512, 768 \times 512, 562 \times 450, 352 \times 240$. One example of creating novel views by extrapolation is given in the right column of Figure 7. A circular trajectory centered at the middle of both camera positions and whose radius is the half of the distance between two camera positions is defined for the garden flower images. We can notice more geometric distortion on rendered images as the camera trajectory deviates more from the in-between path.

For non rigid objects or scenes in Figure 8, the search space for seed matches is reduced to 5% left, right, top and bottom of image size, instead of the fundamental matrix encoding the global rigidity, a compensation 2D translation vector is robustly estimated from the first propagation

Figure 7: Left: Four sequences from two images (first and last row) composed of three automatically generated images. Right: Some extrapolated views of the garden flower sequence with a circular camera trajectory.

with the big search space. We used Achermann's image database of different faces taken under similar lighting and pose conditions. The eyes, mouth and nose are usually correctly matched for the morphing application, but artifacts appears at the hairs and the shoulders when the changes are too abrupt. The last example is particularly difficult because of the big difference between the two faces, and its main artifacts are localized at the shoulders and the eyebrows on the left. For theses examples, many uniform areas inside the face silhouette are not matched, but their borders are as the propagation tends to match edges by covering the same distance in both images. The computation time for each unconstrained propagation is approximately 0.7 second for 512×342 image to obtain 29296 and 23796 matches.

Many other examples have been experimented, including both positive and negative results. We have noticed that the most notable matching blunders for image based rendering are due to periodic textures: if a seed match is shifted from some periods, then the resulting propagation has the same shift. It is expected that these blunders are reduced using a more conservative matching method for match seeds as the relaxation step proposed by [42] for the fundamental matrix estimation.

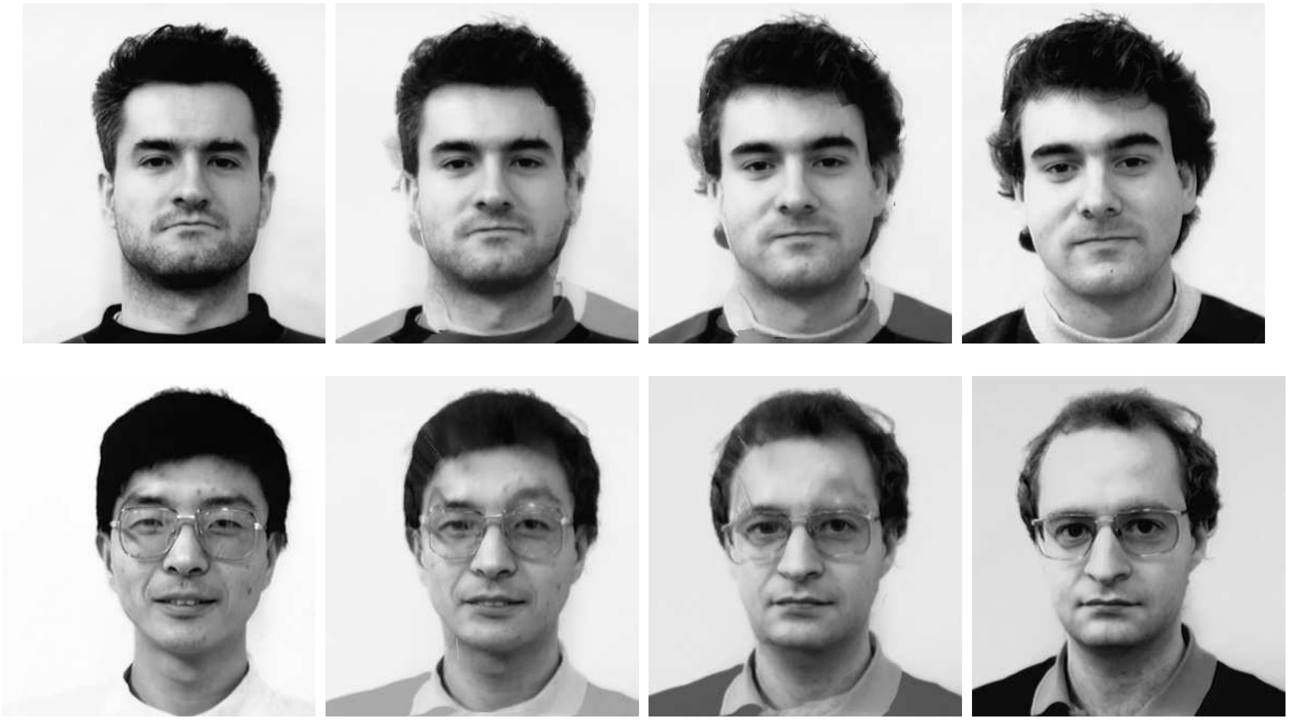

Figure 8: Each row shows four images of an automatically generated sequence from two faces of different persons, using our matching propagation algorithm.

7 Conclusion and future work

We developed a new quasi-dense matching algorithm based on the best-first match propagation strategy. The algorithm has many desirable properties: robustness to initial seed match outliers, efficiency in time and space, and possibility of handling occluded areas. We believe that the quasi-dense matching is an intermediate solution between sparse and dense matching appropriate for a broad range modeling and visualization based applications. These properties have been empirically demonstrated over various modeling and view synthesis applications from real images. We are continuing work in 3D modeling and rendering quasi-dense models based on the algorithm presented in this paper.

Acknowledgements

We would like to thank D. Taylor for providing us the images of the man in New York, and B. Achermann for the face image database.

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[2] S. Avidan and A. Shashua. Novel view synthesis in tensor space. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 1034–1040, June 1997.

[3] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[4] Y. Boykov, O. Veksler, and R. Zabih. Disparity component matching for visual correspondence. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, USA*, pages 470–475, 1998.

[5] P. Burt and B. Julesz. A gradient limit for binocular fusion. *Science*, 208:615–617, 1980.

[6] Q. Chen and G. Medioni. A volumetric stereo matching method: Application to image-based modeling. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pages 29–34, Fort Collins, Colorado, 1999.

[7] S.E. Chen and L. Williams. View interpolation for image synthesis. In SIGGRAPH *1993*, pages 279–288. Computer Graphics, 1993.

[8] U.R. Dhond and J.K. Aggarwal. Structure from stereo – a review. IEEE *Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, November 1989.

[9] L. Falkenhagen. Depth estimation from stereoscopic image pairs assuming piecewise continuous surfaces. In Y. Paker and S. Wilbur, editors, *Proceedings of the Workshop on Image Processing for Broadcast and Video Production, Hamburg, Germany*, Series on Workshops in Computing, pages 115–127. Springer-Verlag, November 1994.

[10] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[11] A.W. Fitzgibbon, G. Cross, and A. Zisserman. Automatic 3d model construction for turn-table sequences. In *3D Structure from Multiple Images of Large-scale Environments SMILE'98*, pages 154–169. Springer Verlag, 1998.

[12] P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve discontinuities. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence, Sydney, Australia*, August 1991.

[13] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. In SIGGRAPH *1996, New Orleans, USA*, pages 43–54, 1996.

[14] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[15] R.I. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140, 1997.

[16] R.I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[17] S.S. Intille and A.F. Bobick. Disparity-space images and large occlusion stereo. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 179–186. Springer-Verlag, 1994.

[18] E. Izquierdo and S. Kruse. Image analysis for 3d modeling, rendering and virtual view generation. *Computer Vision and Image Understanding*, 71(2):231–253, August 1998.

[19] A. Koschan. What is new in computational stereo since 1989 : A survey on stereo papers. Technical report 93-22, Department of Computer Science, University of Berlin, August 1993.

[20] S. Laveau and O. Faugeras. 3D scene representation as a collection of images and fundamental matrices. Technical report 2205, INRIA, February 1994.

[21] M. Levoy and P. Hanrahan. Light field rendering. In SIGGRAPH *1996, New Orleans, USA*, pages 31–42, 1996.

[22] M. Lhuillier. Efficient dense matching for textured scenes using region growing. In *Proceedings of the ninth British Machine Vision Conference, Southampton, England*, pages 700–709, 1998.

[23] M. Lhuillier and L. Quan. Image interpolation by joint view triangulation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, volume 2, pages 139–145, 1999.

[24] M. Lhuillier and L. Quan. Edge-constrained joint view triangulation for image interpolation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, volume 2, pages 218–224, 2000.

[25] M. Lhuillier and L. Quan. Robust dense matching using local and global geometric constraints. In *Proceedings of the 15th International Conference on Pattern Recognition*, volume 1, pages 968–972, 2000 (Best Student Paper Award).

[26] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 1981.

[27] L. McMillan and G. Bishop. Plenoptic modelling: an image-based rendering system. In SIGGRAPH *1995, Los Angeles, USA*, pages 39–46, 1995.

[28] H.P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, USA*, page 584, August 1977.

[29] P.J. Narayanan, P.W. Rander, and T. Kanade. Constructing virtual worlds using dense stereo. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 3–10, 1998.

[30] M. Okutomi and T. Kanade. A multiple-baseline stereo. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, April 1993.

[31] G.P. Otto and T.K. Chau. A region-growing algorithm for matching of terrain images. *Image and Vision Computing*, 7(2):83–94, 1989.

[32] S.B. Pollard, J.E.W. Mayhew, and J.P. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient constraint. *Perception*, 14:449–470, 1985.

[33] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 90–95, January 1998.

[34] L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(1):34–46, January 1995.

[35] G.M. Quénot. The orthogonal algorithm for optical flow detection using dynamic programming. In *International Conference on Acoustics, Speech and Signal Processing*, 1992.

[36] D. Scharstein. Stereo vision for view synthesis. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 852–858, June 1996.

[37] S.M. Seitz and C.R. Dyer. View morphing. In SIGGRAPH *1996, New Orleans, USA*, pages 21–30, 1996.

[38] A. Shashua. Trilinearity in visual recognition by alignment. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 479–484. Springer-Verlag, May 1994.

[39] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pages 593–600, 1994.

[40] R. Szeliski and H.Y. Shum. Motion estimation with quadtree splines. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 757–763. IEEE Computer Society Press, June 1995.

[41] P.H.S Torr and D.W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.

[42] Z. Zhang, R. Deriche, O.D. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78(1-2):87–119, 1994. Appeared in October 1995, also INRIA Research Report No.2273, May 1994.

[43] Z. Zhang and Y. Shan. A progressive scheme for stereo matching. In *SMILE'2, Second Workshop on Structure from Multiple Image of Large Scale Environnements, ECCV Workshop, Dublin, Ireland*, pages 26–31, July 2000.