Quasi-Dense Wide Baseline Matching Using Match Propagation

Juho Kannala and Sami S. Brandt
Machine Vision Group
University of Oulu, Finland
{jkannala,sbrandt}@ee.oulu.fi

Abstract

In this paper we propose extensions to the match propagation algorithm which is a technique for computing quasi-dense point correspondences between two views. The extensions make the match propagation applicable for wide baseline matching, i.e., for cases where the camera pose can vary a lot between the views. Our first extension is to use a local affine model for the geometric transformation between the images. The estimate of the local transformation is obtained from affine covariant interest regions which are used as seed matches. The second extension is to use the second order intensity moments to adapt the current estimate of the local affine transformation during the propagation. This allows a single seed match to propagate into regions where the local transformation between the views differs from the initial one. The experiments with real data show that the proposed techniques improve both the quality and coverage of the quasi-dense disparity map.

1. Introduction

Automatic three-dimensional model acquisition for an unknown scene from multiple images is a classical problem in computer vision. For uncalibrated image sequences, the standard approach is based on sparse points of interest which are tracked across the sequence and reconstructed together with the camera motion [5, 3, 6]. Eventhough the sparse approach is often sufficient for calibrating both the internal and external parameters of the cameras it is not sufficient for the accurate three-dimensional description of the scene. However, the sparse approach is often used to simplify the problem whereafter dense matching methods are applied for the calibrated views [16, 17].

So far there are many dense matching algorithms available. The number of traditional two-frame stereo correspondence algorithms is large [18]; there are also several volumetric approaches which are voxel-oriented [19, 8] or based on graph cuts [7] or level sets [2]. In addition, probabilistic [20] and PDE-based formulations [21] have been used for dense matching of wide baseline images.The problem with traditional stereo algorithms is that they are designed for view pairs with a small baseline and for almost frontoparallel planes. Hence, they can not be easily used for wide baseline views for which the epipolar lines are not parallel. On the other hand, the volumetric approaches use a discretized volume or restrict the possible depth values to a predefined accuracy. Thus, these methods are expensive in terms of time and memory when high accuracy is needed. However, the quasi-dense approach [10] is more efficient than the dense approaches but sufficient also for 3D surface reconstruction.

This paper is hence built on the previous works [9, 10] where a quasi-dense approach to surface reconstruction from image sequences was proposed. In this approach, sparse interest point matches are used as seed points for the match propagation algorithm which produces denser, but not completely dense, pixel correspondences that are called quasi-dense pixel correspondences. The quasi-dense pixel correspondences are computed for successive image pairs in the sequence and they are also used in the image geometry estimation. For uncalibrated image sequences, the quasi-dense approach provides a full automatic geometry estimation like the standard sparse approach but it also provides a high density of 3D points on which a surface representation can be reconstructed. Moreover, it has been reported that the quasi-dense approach provides more robust and accurate geometry estimation than the sparse approach [10]. Hence, even if the final goal would be a completely dense reconstruction the quasi-dence approach may still be useful since most of the dense methods, such as [21], require reliable pixel correspondences for initialization.

There are certain issues in the match propagation algorithm which limit its applicability in the wide baseline case. At each step of propagation, small image patches are extracted around the current seed point in both images and the new candidate matches are scored according to the zero-mean normalized cross-correlation (ZNCC) [9]. Hence, it is implicitly assumed that the local transformation between the patches is effectively a translation, so the algorithm is

1-4244-1180-7/07/$25.00 ©2007 IEEE

not suitable for wide baseline matching. To cope with this problem, we use a general affine model for the local transformation between the patches. Moreover, we will show that it is possible to determine the local affine transformation adaptively during the propagation. This is achieved by using the second order intensity moments locally together with the epipolar geometry. The experiments show that the proposed extensions improve the performance of match propagation in the wide baseline setting.

The structure of this paper is as follows. First, there is a brief description of the original match propagation algorithm in Section 2, and the proposed improvements are then described in Sections 3 and 4. The experiments are presented in Section 5.

2. Background

The match propagation algorithm [9] starts from a set of sparse matches between two images, I and I', and produces a quasi-dense disparity map which contains a large number of pixel correspondences. In [9], the initial matching of sparse interest points was done by using the ZNCC score together with a simple cross-consistency check. The propagation itself starts from the set of initial matches $\{(\mathbf{x}_i, \mathbf{x}'_i)\}_i$, where $\mathbf{x}_i$ and $\mathbf{x}'_i$ are corresponding pixels in the two images. The initial matches are called *seed points*. The propagation proceeds by iterating the following three steps:

(i) the seed point $(\mathbf{x}_i, \mathbf{x}'_i)$ with the best ZNCC score is removed from the list of seed points

(ii) new candidate matches are searched from the spatial neighborhood of the current seed $(\mathbf{x}_i, \mathbf{x}'_i)$

(iii) the candidate matches exceeding a ZNCC threshold and a difference-based confidence limit are stored in the disparity map and added to the list of seed points

In this way the number of pixel correspondences in the disparity map increases until the list of seeds becomes empty.

More precisely, in step (ii) above the spatial neighborhood, $\mathcal{N}(\mathbf{x}, \mathbf{x}')$, of a match $(\mathbf{x}, \mathbf{x}')$ is defined so that the matches in this neighborhood satisfy a constraint called *discrete 2D disparity gradient limit*. The precise definitions are as follows. The $(2N+1) \times (2N+1)$ neighborhoods of pixels $\mathbf{x}$ and $\mathbf{x}'$ are

$$\mathcal{N}(\mathbf{x}) = \{\mathbf{u} \mid (\mathbf{u} - \mathbf{x}) \in [-N, N]^2\},$$
$$\mathcal{N}(\mathbf{x}') = \{\mathbf{u}' \mid (\mathbf{u}' - \mathbf{x}') \in [-N, N]^2\},$$

as usual, and the neighboring matches satisfying the disparity gradient constraint are given by

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

where ϵ is the disparity gradient limit. With integer pixel coordinates, the smallest nonzero value for ϵ is 1 which is also the value used in [9] (together with $N = 2$). Enforcing the disparity gradient constraint implies that the 2D vectors from the current seed point to the new candidate point must have approximately the same direction in both images.

In step (iii) of the algorithm, when the candidate matches $(\mathbf{u}_l, \mathbf{u}'_l)$ are given for the current seed, a difference-based confidence measure s is computed for the candidate pixels,

$$s(\mathbf{u}_l) = \max\{|I(\mathbf{u}_l + \delta) - I(\mathbf{u}_l)|, \delta \in \{(\pm 1, 0), (0, \pm 1)\}\},$$

and $s'(\mathbf{u}'_l)$ is computed in the corresponding way. In order to prevent propagation into too uniform areas, those candidate matches are discarded for which $s(\mathbf{u}_l) < t$ or $s'(\mathbf{u}'_l) < t$, where t is a predefined threshold. The remaining candidate matches are sorted using the ZNCC measure, which is computed using the windows of size $(2W+1) \times (2W+1)$ around the candidate pixels. Those matches which are not yet in the disparity map (either one of the matching pixels) and exceed a predefined threshold z for the ZNCC score are stored in the disparity map and added to the list of seeds.

The match propagation algorithm is efficient. When the ordering of seed points according to the ZNCC scores is maintained in a heap data structure the time complexity of the algorithm is $O(n \log n)$ where n is the final number of matched pixels. The algorithm is also robust to outliers lying in the set of initial matches due to the best-first strategy in the propagation [9]. In addition, if the fundamental matrix for the view pair is known, the epipolar constraint can be used as an additional matching constraint during the propagation. In [10], in fact, the unconstrained propagation is first used to generate quasi-dense point correspondences for the fundamental matrix estimation and thereafter the constrained propagation is used to produce the final matches.

The major problem with the match propagation algorithm is that it requires conventional stereo image pairs. In particular, the orientation of the images should be similar due to the usage of the direct ZNCC similarity measure and the aforementioned disparity gradient limit. However, in the following we will show that using an affine model for the local transformation between the images makes the match propagation approach applicable in the wide-baseline case.

3. Match propagation for wide baseline views

The first step towards quasi-dense wide baseline matching is to use the sparse wide baseline matching approaches [15, 14] for initial matching. Hence, we suggest using affine covariant regions [15] as seed matches for the match propagation. Such seed matches can provide an estimate for the local geometric transformation between the images, in addition to the position of the match. In order to properly utilize this information, we propose an extension to the match

propagation algorithm which allows adequate modeling of geometric transformations between image patches. But first, we briefly describe the computation of seed matches.

3.1. Initial matches

The affine covariant region detectors [15] together with a viewpoint invariant descriptor [12] have shown good performance in sparse wide baseline matching. Hence, it is a natural choice to use this approach in computing the initial matches for match propagation. In fact, the idea of propagating matches from few initially matched regions to nearby ones has been used in the context of object recognition [4]. The detector that we used in our experiments was the Hessian-Affine detector proposed in [13]. The detected regions were matched using the SIFT descriptor [12]. In principle, any other region detector could be used as well, as long as the detected regions can be represented by ellipses so that the ellipses of corresponding regions, given as symmetric 2×2 matrices $\mathbf{C}$ and $\mathbf{C}'$, are related by

$$\mathbf{C}' = \mathbf{A}^{-\top}\mathbf{C}\mathbf{A}^{-1}, \tag{1}$$

where $\mathbf{A}$ is the local affine transformation between the matching regions [13]. Equation (1) provides a basis for estimating the affine transformation $\mathbf{A}$ given the measurement regions $\mathbf{C}$ and $\mathbf{C}'$. In fact, from (1) we get

$$\mathbf{A} = \mathbf{C}'^{-1/2}\mathbf{R}\mathbf{C}^{1/2}, \tag{2}$$

where $\mathbf{R}$ is an arbitrary orthogonal matrix. Additional information is needed to determine $\mathbf{R}$ and here we used the orientation of local image gradients as in [12, 13]. In summary, the result of the initial matching step is a set of point matches $\{(\mathbf{x}_i, \mathbf{x}'_i)\}_i$ (the centroids of the matched elliptical regions) accompanied with the estimates $\mathbf{A}_i$ for the corresponding local transformations.

Giving an example, Figs. 1(a) and 1(b) illustrate the matched Hessian-Affine regions in two images of a planar scene. The images are from [22] where also the known homographies between the views are available together with the implementations of region detectors [15] and descriptors [14]. For the descriptor matching we used a modified nearest neighbor strategy. We required both that the matching descriptors are the nearest neighbors of each other and the distance to the second nearest neighbor is above a certain threshold. In our experiments, we found that such matching strategy significantly reduces false matches for images with repeating texture patterns as opposed to using just a single threshold for the nearest neighbor distance. In Fig. 1, the 50 best matching regions are shown.

3.2. Propagation with affine normalization

In this section we describe our first extension to the match propagation algorithm. We append the seed matches

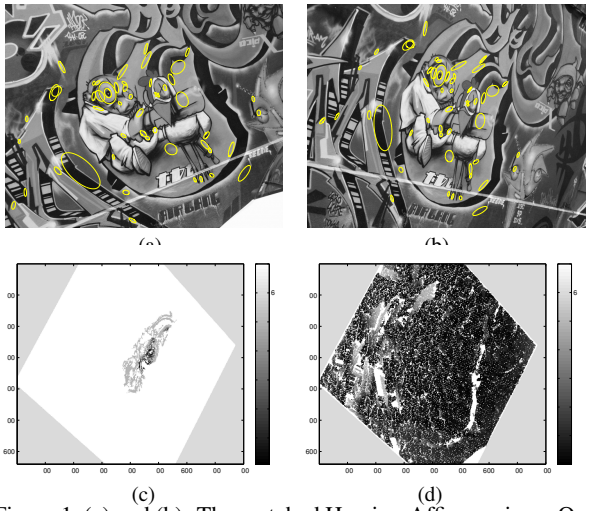

Figure 1. (a) and (b): The matched Hessian-Affine regions. One of them, denoted by the cross, is used as a seed match for the match propagation algorithm. The pixel matches computed with the original (c) and the modified (d) propagation approach are illustrated in the second view by coloring them according to their Sampson distance from the ground truth homography. (The distances over 5 are suppressed to 5, the grayvalue for the noncommon area is 6.)

with an affine transformation matrix which is used to normalize the local image patches before enforcing the 2D disparity gradient limit and computing the ZNCC scores. In addition, we use a subpixel level of accuracy for the position of seed points whereas the disparity map is stored at the pixel level resolution. Although the proposed improvements seem to be minor, they significantly improve the matching result when carefully implemented.

As before, the initial seed matches, consisting of points $\{(\mathbf{x}_i, \mathbf{x}'_i)\}_i$ and matrices $\mathbf{A}_i$, are first scored using the ZNCC measure. The ZNCC scores are computed from geometrically normalized image patches where the normalization is done as follows: (a) a $(2W+1)\times(2W+1)$ square patch centred on the seed point is extracted from the image which locally has a lower resolution (the local magnification factor is given by $|\det \mathbf{A}_i|$), (b) the corresponding area in the other image is determined by affinely mapping the square patch on the other image, (c) the patches from both images are transformed to $(4W+1)\times(4W+1)$ windows by interpolation, (d) the interpolated windows are decimated to the size $(2W+1)\times(2W+1)$. This process is illustrated in Fig. 2 and, for a correct match, it should ideally result in two identical image patches when the affine transformation model is tenable.

The scoring of seeds is not the only place where geometric normalization is needed during the propagation. At each iteration, the whole neighborhood from which new matches are searched needs to be normalized. Therefore, in order to reduce the number of interpolations per iteration, we ac-

tually compute the normalized patches corresponding to a window of size $(2(W+N)+1)\times(2(W+N)+1)$ in the lower resolution image. The normalized neighborhoods are illustrated in Fig. 3 and they can be treated as in the original algorithm, i.e., the disparity gradient limit and the ZNCC measure can be well used *after* the normalization. However, the pixels of the normalized patch do not correspond to integer pixel coordinates in the higher resolution image. Hence, we use subpixel level of accuracy for the seed matches. The coordinates are rounded to integer pixel values only when used to index the disparity map, for example when storing new matches. Importantly, when new matches are added to the list of seeds they inherit the affine transformation matrix from the current seed. This implies that a seed match always contains an estimate of the local affine transformation and it is the basis for the geometric normalization.

In the normalization process, as shown in Fig. 2, the image patch of the lower resolution image is first upsampled and then downsampled. This is necessary in order to cope with small scale changes between the images, i.e., the upsampling is needed to prevent aliasing effects in the interpolation of the higher resolution image and both image patches have to be processed in a similar way to assure comparability. In addition, the decimation of the transformed patches can be performed efficiently with two one-dimensional anti-aliasing filters. In fact, despite the inevitable interpolations, the extended match propagation algorithm is efficient. It is clear that the overall complexity remains the same and depends only on the number of final matches. In fact, we save some time by skipping the check of the confidence measure limit described in Section 2. We additionally found that the geometric normalization allows using higher threshold for the ZNCC score and this threshold alone is often sufficient for preventing the propagation into too uniform areas.

The necessity for the affine normalization is made clear in Fig. 1, where the poor performance of the original propagation approach is illustrated with the disparity map in Fig. 1(c). In this example the propagation was started from a single seed match, denoted by the cross in Figs. 1(a) and 1(b), using the default parameter values ($N=2$, $\epsilon=1$, $t=0.01$, $W=2$, $z=0.5$) [9]. The number of pixel correspondences in the final disparity map is 10726 but most of them are not correct as their Sampson distance [6] to the ground truth homography is large. On the contrary, the propagation with affine normalization performs well as shown Fig. 1(d). There are 261572 matches and the parameter values used were $N=2$, $\epsilon=1$, $W=2$ and $z=0.8$. Despite the higher correlation threshold the number of matches is much larger than in Fig. 1(c) and most of them are correct. The computation time per match was $1.7\,ms$ for Fig. 1(c) and $3.5\,ms$ for Fig. 1(d) when Matlab implementations of both algorithms were used.

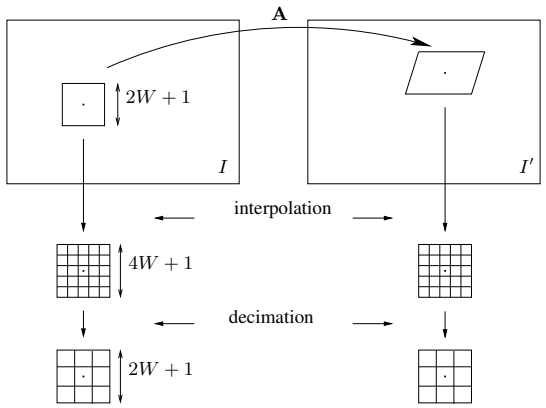

Figure 2. The geometric normalization of local image patches.

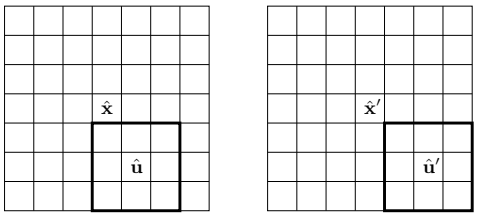

Figure 3. The normalized image neighborhoods of size $(2(W+N)+1)\times(2(W+N)+1)$ for a seed point $(\mathbf{x},\mathbf{x}')$. Here $N=2$, $W=1$, and $\hat{\cdot}$ refers to the symbols in the normalized frame. The ZNCC score for the candidate match $(\mathbf{u},\mathbf{u}')$ is computed using the black framed windows.

4. Adaptive propagation

In this section we propose our second extension to the match propagation algorithm which allows the adaptation of the local affine transformation estimates during the propagation. The adaptation is based on the second order intensity moments which determine the local affine transformation up to a rotation. The remaining rotational degree of freedom is determined from the epipolar geometry which is assumed to be known.

The motivation for proposing an adaptive propagation technique is to allow propagation into regions where the local transformation between the views differs from the initial one. In fact, in the affine normalization scheme, proposed in Section 3, the estimate of local affine transformation is inherited from the current seed to its descendants. This implies that each initial seed match can propagate only into such regions where the estimated transformation, originally computed from the affine covariant interest regions, is valid. In some cases this may be sufficient, for example, when the global transformation between the views is close to affine, as in Fig. 1, or when the initial seed regions are dense enough so that the local transformations in different parts of the scene can be modelled. However, in some cases a more powerful propagation strategy would be useful. In Fig. 5 we have two images of a scene containing two planes,

a map on a table and a calibration object. Here the affine transformation recovered from one region match, located in the front and denoted by yellow ellipses, does not adequately model the transformation of the whole plane since the perspective distortion is significant. In order to address this kind of situations, we propose the principle of adaptive propagation in the following.

4.1. Local shape from intensity moments

Let $f, f' : \mathbb{R}^2 \rightarrow [0, 1]$ denote the image intensity functions so that $f'(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{x})$ where $\mathbf{A}$ is a nonsingular 2×2 affine transformation matrix. Here we have dropped the translational part as we assume that the coordinate systems are centred to the points under consideration. Given two positive window functions g and g' so that $g'(\mathbf{x}) = g(\mathbf{A}^{-1}\mathbf{x})/|\det \mathbf{A}|$, we define the windowed second moment matrix of f by

$$\mathbf{S}_{f,g}(\mathbf{x}) = \int \mathbf{v}\mathbf{v}^\top f(\mathbf{v}) g(\mathbf{x} - \mathbf{v}) d\mathbf{v}, \tag{3}$$

and the matrix $\mathbf{S}_{f',g'}$ of f' is defined in a similar way. Since the affine transformation between the window functions is the same as between the intensity functions we get

$$\mathbf{S}_{f',g'}(\mathbf{x}) = \mathbf{A}\mathbf{S}_{f,g}(\mathbf{A}^{-1}\mathbf{x})\mathbf{A}^\top. \tag{4}$$

By introducing a simpler notation at the origin, $\mathbf{S}' = \mathbf{S}_{f',g'}(\mathbf{0})$ and $\mathbf{S} = \mathbf{S}_{f,g}(\mathbf{0})$, we get from (4) that

$$\mathbf{A} = \mathbf{S}'^{1/2}\mathbf{R}\mathbf{S}^{-1/2}, \tag{5}$$

where $\mathbf{R}$ is an orthogonal matrix which can be determined from a pair of corresponding directions in the two images (i.e. from one unit vector correspondence pair). Hence, the equation (5) implies that the moment matrices could be used to determine the local affine transformation in a similar way as the elliptical interest regions were used in Section 3.1.

In general, the problem of using (5) is that if g is fixed and $\mathbf{A}$ is unknown, g' is also unknown. However, in our application we usually have a relatively good guess for the local transformation in the neighborhood of the initial matches. Hence, at each propagation step, we use the affine estimate of the current seed in forming the window functions and thereafter the estimate is updated for the new seeds using (5).

In detail, let us consider one step of the adaptive propagation process with the illustrations in Figs. 2 and 3. Assume that the current seed is $(\mathbf{x}, \mathbf{x}')$ and the corresponding transformation estimate is $\mathbf{A}_0$. The normalization process is carried out as described in Section 3.2 and the resulting normalized image patches are shown in Fig. 3. Now, consider that the normalized cross-correlation of the black framed subwindows in Fig. 3 exceeds the threshold so that $(\hat{\mathbf{u}}, \hat{\mathbf{u}}')$, corresponding to a match $(\mathbf{u}, \mathbf{u}')$ in the original images, is a new seed match. Instead of performing the integration (3) in the original images we integrate in the normalized domain and thereafter transform the obtained moment matrices to the image coordinate frames. If the characteristic function of a $(2W + 1) \times (2W + 1)$ window, centred on $\hat{\mathbf{u}}$ and $\hat{\mathbf{u}}'$, is used as a window function in the normalized patches the integration in (3) reduces to a simple summation. Then the obtained moment matrices $\hat{\mathbf{S}}$ and $\hat{\mathbf{S}}'$ are transformed back to the original frames by $\mathbf{S} = \hat{\mathbf{S}}$ and $\mathbf{S}' = \mathbf{A}_0\hat{\mathbf{S}}'\mathbf{A}_0^\top$. The new transformation estimate for the new seed match is then obtained by using (5). Instead of uniform weighting of the square window by its characteristic function a different weight function can also be used. For example, a gaussian window function of size $(2W + 1) \times (2W + 1)$ was found to give good results. The advantage of performing the integration in the normalized domain is that the values of the windowing function can be computed in advance since the domain of integration is always the same. This significantly improves the efficiency.

The idea behind our approach, presented above, is similar in spirit to those presented in [11] and [13]. However, here the context of application and the implementation are different. Importantly, the shape adaptation is not based on the windowed second moment matrix of the intensity gradient as in [11, 13] but on the windowed second moment matrix of the intensity function itself. In our experiments we found that, instead of the gradient, the intensity function provides a more stable basis for estimating the local affine transformation between a pair of views. This is also intuitively reasonable since the gradient is more sensitive to noise. In addition, computing the windowed gradients during the propagation would increase the computational load of the quasi-dense approach.

4.2. Photometric normalization

There is a one important issue related to (3) that should be taken in to account in a careful implementation of the algorithm. If there additionally is an affine photometric transformation between the images, $f'(\mathbf{x}) = \lambda f(\mathbf{A}^{-1}\mathbf{x}) + b$, the transformation rule (4) does not hold. This is a drawback since the ZNCC measure is invariant to affine photometric transformations and, hence, the adaptive extension might be a bottleneck in applying the algorithm in varying illumination conditions. Therefore, instead of applying (4) directly to the intensity functions we first remove the effect of photometric transformations by the following local normalizations

$$\tilde{f}(\mathbf{x}) = \frac{f(\mathbf{x}) - \mu(f)}{\sigma(f)} + C, \tag{6}$$

$$\tilde{f}'(\mathbf{x}) = \frac{f'(\mathbf{x}) - \mu'(f')}{\sigma'(f')} + C, \tag{7}$$

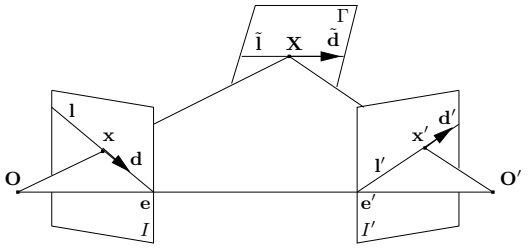

Figure 4. The local relative orientation is obtained from epipolar lines. The two-fold ambiquity is solved using the jointly oriented epipoles [1].

where C is a positive constant, the mean $\mu(f) = \int f(\mathbf{x})g(\mathbf{x})d\mathbf{x} / \int g(\mathbf{x})d\mathbf{x}$, the standard deviation $\sigma(f) = (\int (f(\mathbf{x}) - \mu(f))^2 g(\mathbf{x})d\mathbf{x} / \int g(\mathbf{x})d\mathbf{x})^{1/2}$. The corresponding quantities for f' are defined in a similar way using the window function g'. Here the constant C is added in order to preserve the positive (semi)definiteness of matrices $\mathbf{S}_{\tilde{f},g}$ and $\mathbf{S}_{\tilde{f}',g'}$. The value $C = 2$ was used in all of our experiments. After the normalization we proceed as above using the normalized intensity functions $\tilde{f}$ and $\tilde{f}'$ instead of the original ones.

4.3. Local orientation from the epipolar constraint

As described above, the windowed second order intensity moment matrices allow to recover the local affine transformation up to a rotation. The unknown rotation in (5) can be determined from one pair of corresponding unit vectors [13]. Such corresponding unit vectors can be determined from corresponding epipolar lines using the concept of *joint orientation of epipoles* which determines the mutual position of cameras [1]. The details are as follows.

Consider the two view setting illustrated in Fig. 4. There a 3D point $\mathbf{X}$ on a smooth surface is imaged with two pinhole cameras placed at $\mathbf{O}$ and $\mathbf{O}'$. The tangent plane of the surface at $\mathbf{X}$ is Γ and its intersection with the epipolar plane, defined by points $\mathbf{X}$, $\mathbf{O}$ and $\mathbf{O}'$, is the line $\tilde{\mathbf{l}}$. Hence, under the assumption of locally planar surfaces, the epipolar lines $\mathbf{l}$ and $\mathbf{l}'$ are images of the same 3D line in the neighborhoods of the corresponding image points $\mathbf{x}$ and $\mathbf{x}'$. Hence, the epipolar lines give a line correspondence from which corresponding directions may be determined up to sign, i.e., the corresponding direction for $\mathbf{d}$ in Fig. 4 is either $+\mathbf{d}'$ or $-\mathbf{d}'$. The correct choice is of course such that the directions in both images correspond to the same direction $\tilde{\mathbf{d}}$ on the scene plane. Given the fundamental matrix alone one may now compute the oriented epipoles $\mathbf{e}$ and $\mathbf{e}'$ so that the corresponding epipolar lines $\mathbf{l} = \mathbf{e} \times \mathbf{x}$ and $\mathbf{l}' = \mathbf{x}' \times \mathbf{e}'$ have proper joint orientation, i.e., the vectors $\mathbf{d} = (l_2, -l_1)^\top$ and $\mathbf{d}' = (l'_2, -l'_1)^\top$ indicate corresponding directions in the images. Here we assume that the image coordinate bases of both cameras have equal handedness [1].

5. Experiments

We experimented the proposed extensions to the match propagation algorithm with real images. In the first experiment (Section 5.1) we had two views of a scene containing planar surfaces. In the second experiment we used three views of a complex 3D scene as described in Section 5.2.

5.1. A scene with planar surfaces

In Fig. 5 there are two images of a scene which contains two planes, a map on a table and a calibration plane orthogonal to the plane of the map. This image pair is suitable for evaluating the algorithms since the calibration plane allows to compute the homographies which describe the mappings between the planes; using the homographies the found matches can be verified. In addition, it can be seen that there is a clear projective distortion between the views so that affine mapping is not a good approximation for the homographies. In order to assure that the recovered homographies are accurate we calibrated the intrinsic parameters of the camera beforehand and removed the radial distortion from the images.

The results of match propagation, started from a single seed match, using both the original approach and the proposed modifications are illustrated in Fig. 5. The results indicate that the affine normalization improves the matching in the neighborhood of the initial match as far as the affine transformation estimate is valid. Utilizing the epipolar constraint reduces the number of false matches, as expected. The best results are achieved using the adaptive propagation approach. The results with two different window functions, uniform and gaussian, are shown and it can be seen that the gaussian window performs slightly better. The largest difference between the last two disparity maps is on relatively uniform areas where the matching is more unreliable. In fact, if the reliability of matches was critical, one could prevent propagation into too uniform areas by using an additional threshold for the variance of the correlation windows.

The coverage of the last two disparity maps in Fig. 5 is extensive verifying that the adaptive propagation works in practice. The matching is not completely dense, and there are quite many unmatched pixels in the upper part of the image. However, this is mostly due to the different local magnification factors in the images, especially in the back regions of the ground plane. The improvement, illustrated in Fig. 5, is achieved without significantly reducing the efficiency of the approach. The computation times per match for the subfigures (c)-(h) were 7.2, 5.9, 3.8, 3.3, 4.9 and 3.6 ms, respectively, using our current Matlab implementations. Here the relatively long computation times of the original approach are probably due to the fact that the ZNCC measure is computed many times without obtaining

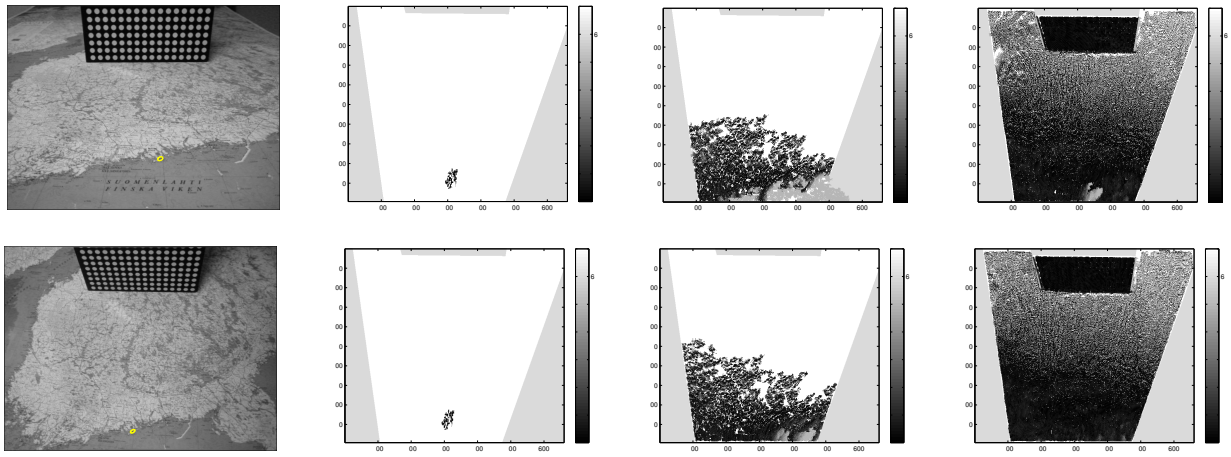

Figure 5. Quasi-dense matching of images starting from a single seed. The images of two planes and one detected region match, first column. The matching result of the original algorithm without and with the epipolar constraint, second column. The matching result using affine normalization without and with the epipolar constraint, third column. Finally, the result using the adaptive propagation algorithm with a uniformly weighted window and a gaussian window, last column. The parameter values used for all figures were $N = 3$, $\epsilon = 1$, $W = 3$ and $z = 0.8$, and for the original algorithm we had $t = 0.01$. The matching pixels are here colored according to their Sampson distance from the known homographies. The values over 5 are suppressed to 5, and the noncommon image area has grayvalue 6.

new matches since the threshold z is high (0.8).

5.2. A complex 3D scene

The algorithms were also tested with a more complex scene, shown in Fig. 6. Again, we used the calibration plane for computing the camera poses accurately so that the known geometry can be used to evaluate the matches. Here we matched the middle image with the first and third one using the original and the adaptive approach, as it showed the best performance in the previous experiment. The initial region matches were required to satisfy the epipolar constraint and they are illustrated in the top row of Fig. 6. Those pixels in the middle image for which a match was found both in the first and third image are illustrated in the bottom row of Fig. 6. The three-view matches are colored according to their Sampson distance from the known trifocal tensor [6]. Since the three-view constraint is stronger than the two-view constraint the false matches fitting to the correct geometry are relatively improbable and, hence, this constraint can be used to evaluate the matches.

In Fig. 6, the number of matches is 69522 for the original algorithm and 141859 for the adaptive algorithm. The median values of the distance from the trifocal tensor are 2.0 and 0.72, respectively. Hence, it can be clearly seen that the adaptive approach provides matches whose coverage and accuracy are better. Moreover, the original approach fails in matching the repeating texture on the calibration plane correctly. In addition, it does not find matches from the ground plane (the plane of the table) since the viewpoint changes affect significantly on its appearance. Finally, in order to demonstrate the quality of the matches, we computed the 3D reconstruction of pixels having a Sampson distance below 0.5 and the top view of the 42353 reconstructed points is shown in the last subfigure in Fig. 6.

6. Conclusion

In this paper we have proposed two extensions to the match propagation algorithm which make the algorithm applicable in wide baseline matching. The first extension incorporates an affine normalization step into the basic propagation algorithm. The propagation is initialized by using affine covariant regions as seed matches. The epipolar constraint can be additionally utilized in matching if it is known. In the second extension, the adaptive affine normalization step is introduced for the match propagation in cases where the epipolar geometry is known. In the experiments it was shown that the proposed extensions significantly improve the matching result in wide baseline cases. These improvements are achieved so that the efficiency of the approach is preserved.

References

[1] O. Chum, T. Werner, and T. Pajdla. On joint orientation of epipoles. *Research Report, CTU-CMP-2003-10, Czech Technical University*, 2003.

[2] O. Faugeras and R. Keriven. Variational principles, surface evolution, PDE's, level set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, 1998.

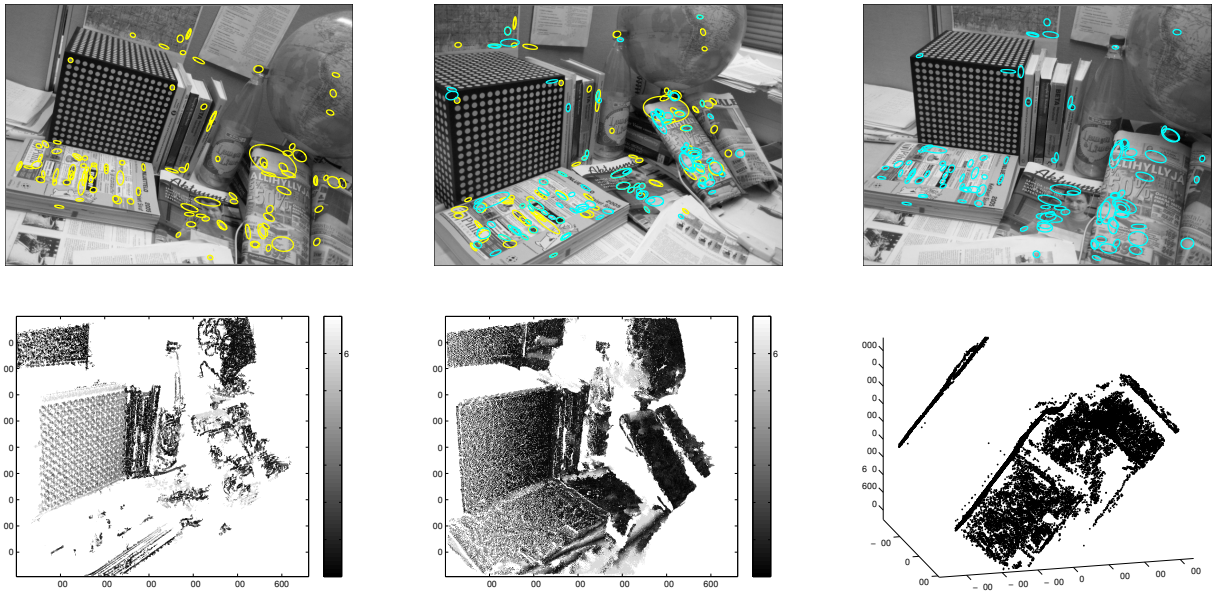

Figure 6. Top: Three views of a 3D scene. The 100 best matching interest regions for the first and second pair of views are plotted in yellow and cyan, respectively. Down: The pixels matched between both view pairs are illustrated in the middle image for the original (left) and the adaptive (middle) approach. The epipolar constraint was utilized with both approaches. The default parameters were used for the original algorithm and the parameters for the adaptive approach were the same as in Fig. 5. Here the matching pixels are colored according to their Sampson distance from the known trifocal tensor so that the values over 6 are suppressed to 6. Finally in the last figure in the right, there is a top view of the reconstructed 3D points; only those points are shown for which the distance from the trifocal tensor is below 0.5.

[3] O. Faugeras and Q. T. Luong. *The Geometry of Multiple Images*. The MIT Press, 2001.

[4] V. Ferrari, T. Tuytelaars, and L. Van Gool. Simultaneous object recognition and segmentation by image exploration. In *ECCV*, 2004.

[5] A. W. Fitzgibbon and A. Zisserman. Automatic 3D model acquisition and generation of new images from video sequences. In *EUSIPCO*, 1998.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, 2nd edition, 2003.

[7] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *ECCV*, 2002.

[8] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. In *ICCV*, 1999.

[9] M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1140–1146, 2002.

[10] M. Lhuillier and L. Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):418–433, 2005.

[11] T. Lindeberg and J. Gårding. Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure. *Image and Vision Computing*, 15:415–434, 1997.

[12] D. Lowe. Distinctive image features from scale invariant keypoints. *IJCV*, 60:91–110, 2004.

[13] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60:63–86, 2004.

[14] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[15] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *IJCV*, 65:43–72, 2005.

[16] D. Nister. *Automatic dense reconstruction from uncalibrated video sequences*. PhD thesis, KTH, 2001.

[17] M. Pollefeys. *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. PhD thesis, K.U.Leuven, 1999.

[18] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47:7–42, 2002.

[19] S. M. Seitz and C. R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *CVPR*, 1997.

[20] C. Strecha, R. Fransens, and L. Van Gool. Wide-baseline stereo from multiple views: a probabilistic account. In *CVPR*, 2004.

[21] C. Strecha, T. Tuytelaars, and L. Van Gool. Dense matching of multiple wide-baseline views. In *ICCV*, 2003.

[22] `www.robots.ox.ac.uk/∼vgg/research/affine/`.