Uncalibrated Dynamic Stereo Using Parallax

Francesco Malapelle, Andrea Fusiello
DIEGM - University of Udine
Via delle Scienze, 208 - Udine, Italy
Email: name.surname@uniud.it

Beatrice Rossi, Emiliano Piccinelli, Pasqualina Fragneto
AST Lab, STMicroelectronics
Via Olivetti, 2 - Agrate Brianza (MB), Italy
Email: name.surname@st.com

***Abstract*—In this paper a novel method for computing parallax maps from monocular and uncalibrated video sequences is described. Acquired frames are processed pairwise, starting from a first reference and progressively integrating information coming from subsequent frames in temporal order using a Kalman filter. In this way, temporal stabilization of the generated maps is obtained as well as more consistency with the real video content. Results and evidences coming from the benchmark of the system on both synthetic and natural images have also been reported, showing significant improvements with respect to the parallax maps obtained without temporal integration.**

I. Introduction

In this paper we address the problem of disparity estimation in a monocular sequence of images acquired by a moving camera, a.k.a. *motion stereo* [21]; in particular we confront – for the first time – the *uncalibrated* variant, where camera internal parameters as well as camera motion are both unknown.

Whereas in classical binocular stereo two cameras separated by a fixed baseline are employed, in motion-stereo a single camera moves through a static scene. As a result, over a period of time, the camera traverses a "baseline" of arbitrary length. The rationale for investigating such problem is the attempt to solve the well-known *accuracy-precision* trade-off in stereo matching, which can be summarized as follows: due to quantization errors, the estimated disparity is more precise with a longer baseline, but the matching is less accurate (i.e., more false matches). There is clearly a compromise between precision and accuracy which motion stereo and multiple-baseline stereo approaches address (see for example [10], [12]).

From the geometrical standpoint, the problem is how to integrate disparity measures in a common reference frame. The easiest one is the 3D space: ultimately, every disparity measure must translate into a depth. The computation of depth, however, can be avoided if one considers depth-proxies such as binocular disparity or planar parallax.

If camera motion is constrained such as image planes are coplanar (lateral shift with no rotation), then binocular disparities are related by a scale factor proportional to the baseline [10], hence the integration can take place at the disparity level. If camera motion is constrained on a line (camera centers are collinear), then a multiview rectification can be applied, thus falling back to the previous case. In the general motion case, however, disparity cannot be used, as it depends on the displacement between the reference and the current view (see Fig.1); in this case *planar parallax* can be profitably employed. *Parallax* is the apparent change in the position of an object caused by a change in the point of view; *Planar parallax* represents the displacement in the apparent position of objects imaged from different points of view with respect to a reference plane, and can be computed from binocular disparity.

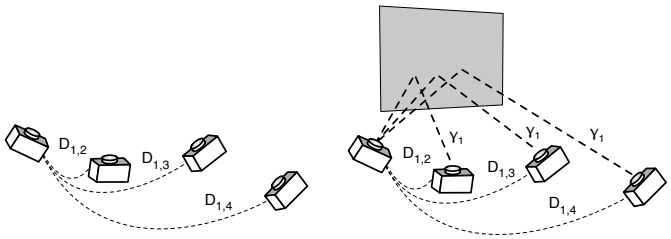

Fig. 1. Binocular disparity D depends on both reference and current views, so it is not possible to univocally merge together several disparity maps for a general, unknown motion; planar parallax γ, instead, depends only on the reference view and a reference plane, hence parallax maps coming from different current views can be integrated.

We use a *dynamic* approach, as we apply a Kalman filter based formulation for recursive estimation of parallax maps by combining several measurements along the time line. The idea of integrating disparity measures using Kalman filter is far from new; early work [19], [10], [5], [13], [17] require motion and camera parameters to be known, because the depth (or its inverse) is the quantity that supports the aggregation from different frames. Most of them restricts to lateral motion. A weakness common to these methods is that they warp the disparity map from previous to current frame: this process introduces errors and approximations that make the integration pointless (i.e. the current disparity map is likely more accurate than the prediction). In our work we avoid this problem by keeping the reference view fixed, thanks to the properties of planar parallax (instead of disparity). Indeed, whereas binocular disparity depends on both reference and current views, parallax depends only on the reference view and a reference plane, hence parallax maps coming from different current views can be integrated. The contribution of this paper can be seen as an unconstrained, uncalibrated extension of these classical works.

A related stream of work is the one denoted as "multiple-baseline" stereo, where the camera centres are collinear (equiv-

alent to lateral motion) and an aggregated matching cost is computed which considers all the images simultaneously [9], [12], [8]. More recent motion-stereo approaches require calibrated cameras and aggregate measures in a discretized 3D volume [20], [11], [23].

In the approach described in this paper, frames are processed pairwise, starting from a first reference then progressively integrating information coming from subsequent frames in temporal order, thus obtaining more stable, reliable and consistent parallax values along the video sequence. Obtained parallax maps might act as enabler for many applications in several different areas: for example novel kind of men-machine touchless interfaces, augmented or virtual reality world generation where real and synthetic objects can interact together or also high-quality view synthesis for smart frame rate up-conversion and free viewpoint 3D TV.

II. Background

In this section we review some background notions needed to understand the proposed method. A complete discussion and formulation of the planar parallax theory can be found in [16], [7]. A more general reference on the geometry of multiple views is [4].

Between two image planes, there exists a non-singular linear transformation, or homography, H that maps the projected points of the space plane in the first view onto the corresponding points in the second view. More in detail, if $(\mathbf{m}_1, \mathbf{m}_2)$ are projections of a 3-D point M belonging to some space plane Π on two reference views I_1, I_2, we have:

$$\mathbf{m}_2 \simeq H_\Pi \mathbf{m}_1 \tag{1}$$

where H_Π is the homography induced by plane Π and $\simeq$ means equality up to a scale factor.

For points M not on the space plane Π, the more general relation holds:

$$\mathbf{m}_2 \simeq H_\Pi \mathbf{m}_1 + \mathbf{e}_2 \gamma_1 \tag{2}$$

where $\mathbf{e}_2$ is the epipole in the second image and γ_1 is a quantity proportional to the distance of the point M from the plane. Observe that in this case the homography H_Π does not map the points in the first view to their corresponding points in the second view. The displacement between the homography mapped point $H_\Pi \mathbf{m}_1$ and the corresponding point $\mathbf{m}_2$ is called *planar parallax* (or, simply, *parallax* if the context is clear) of the point $\mathbf{m}_1$. With a little abuse of notation, the same term will be used also to denote the magnitude γ_1 of this displacement (the direction is toward the epipole).

It is possible to show that the parallax γ_1 depends only on the first view and the plane Π, and not on the second view, although a second view is needed to compute it. Moreover γ_1 is proportional to the inverse of the depth of points. By setting a reference view together with a fixed reference plane Π, one can thus obtain a projective proxy for the depth of a point that is consistent across several views.

Parallax values can be obtained from point correspondences and a plane homography by solving for γ_1 in Eq. 2:

$$\gamma_1 = \frac{(\mathbf{m}_2 \times \mathbf{e}_2)^T (H_\Pi \mathbf{m}_1 \times \mathbf{m}_2)}{\|\mathbf{m}_2 \times \mathbf{e}_2\|^2}. \tag{3}$$

Finally observe that, when two image planes are coplanar (i.e. up to coordinate change, motion is along X axis) and the reference plane is the one at infinity, then H_Π is the identity, the epipole is $\mathbf{e}_2 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$, and thus planar parallax in Eq. 2 results to be proportional to binocular disparity.

III. Proposed Method

The outline of the proposed method is shown in Fig. 2. The input is a monocular *uncalibrated* video sequence of N frames I_i with $i = 1, \ldots, N$, where uncalibrated means that both intrinsic parameters and camera trajectory are unknown. The only assumption made is that a portion of the first view is kept visible at all the subsequent frames of the sequence.

After sparse correspondences are matched across the video, and a projective reconstruction of the camera matrices is obtained, the core of the pipeline, i.e. parallax maps computation, is performed through an iterative procedure repeated $N - 1$ times. At each iteration only two frames are considered: the reference frame I_1 and a second frame obtained by considering subsequent frames in temporal order. Since parallax depends only on the reference frame and the reference plane, each iteration provides a new parallax map referred to I_1. Finally, all the generated maps can be combined together in order to achieve more stable and accurate parallax values.

In the following we shall analyze each step of the proposed pipeline more in detail.

A. Sparse Multi-Matching

The goal of this stage is to match sparse visual features across the video sequence. The procedure is fairly standard and mainly follows the approach of [1]. This step also includes a robust MSAC procedure [18] for the estimation of two-views geometry between pairs of matching images in order to discard bad matches. The output of this stage is a set of tracks, i.e. keypoints matching in more than three images, and a set of fundamental matrices linking pairs of views.

B. Projective Reconstruction

The projective reconstruction proceeds by chaining partial reconstructions from 3 views that are obtained using a 6-points procedure – described in [4] – inside a MSAC iteration.

Lying in a projective stratum, each of the triples of reconstructed perspective projection matrices, PPM for simplicity, is related to the correct (Euclidean) one by a collineation of the 3D space. Once a reference projective frame is fixed, e.g. the one associated to the first triple $\{P_1, P_2, P_3\}$, subsequent triples of PPMs with an overlap of two, as shown in Fig. 2, can be brought to the same frame by computing the proper collineation T as explained in the following.

Let P_i and P_i' be the same camera in two different projective frames, i.e., P_i and P_i' represents the same camera in two

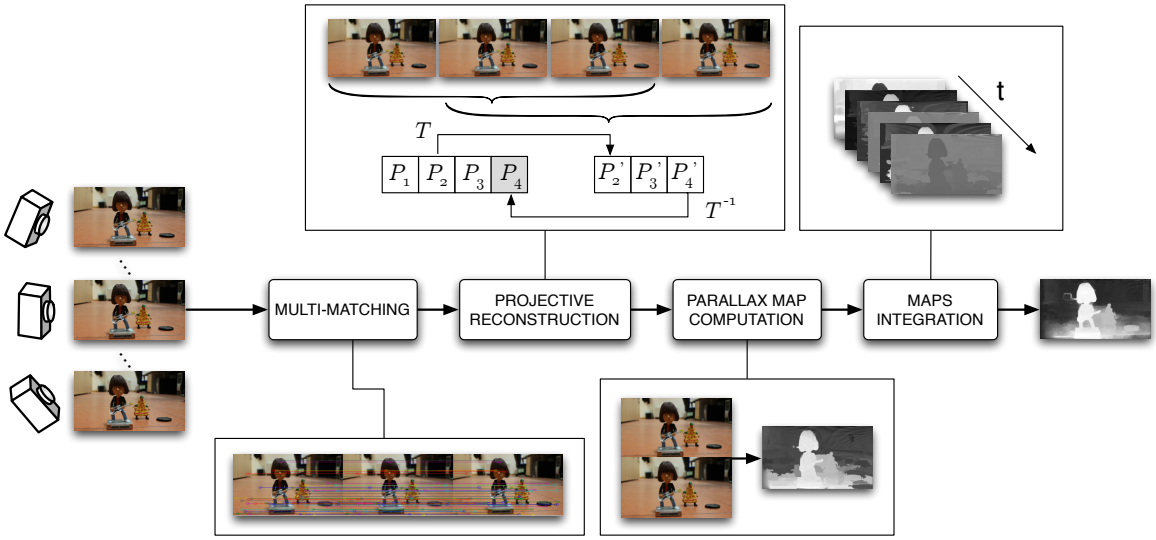

Fig. 2. Overview of the method.

different triplets. They are related by an unknown collineation T:

$$P_i T \simeq P'_i. \tag{4}$$

Introducing the vec operator that arranges the elements of a matrix in a vector column-wise, we obtain:

$$\mathrm{vec}(P_i T) \simeq \mathrm{vec}(P'_i). \tag{5}$$

As shown in appendix A, the equality of two vectors $\mathbf{a}$ and $\mathbf{b}$ of $\mathbb{R}^n$ up to a scale can be written as $[\mathbf{a}]_\times \mathbf{b} = 0$ where $[\mathbf{a}]_\times$ is a suitable $n(n-1)/2 \times n$ matrix that generalizes the external product matrix of $\mathbb{R}^3$. Hence, Eq. 5 can be rewritten as:

$$[\mathrm{vec}(P'_i)]_\times \, \mathrm{vec}(P_i T) = 0. \tag{6}$$

Using the properties of the Kronecker product, Eq. 6 is equivalent to the following linear system of equations in the unknown $\mathrm{vec}(T)$:

$$[\mathrm{vec}(P'_i)]_\times (I_{4\times 4} \otimes P_i) \, \mathrm{vec}(T) = 0. \tag{7}$$

Since the coefficient matrix has rank at most 11, at least two camera matrices are needed to stack-up the 15 equations required to compute the 4×4 matrix T up to scale. This is the reason why our projective reconstruction processes triples of cameras with an overlap of two.

Let $P_i := [Q_i | q_i]$ be the PPM associated to the i-th frame of the video sequence and let:

$$H^{\Pi}_{1i} := Q_i Q_1^{-1} \quad i = 2, \cdots, N \tag{8}$$

be the infinity plane homography between views I_1 and I_i. Observe that in a Euclidean frame H^{Π}_{1i} would be the homography induced by the *true* infinity plane. However, since our cameras are uncalibrated, the PPMs are defined in a projective frame where the infinity plane corresponds to a generic plane Π in the Euclidean frame.

Bringing all the PPMs into a common projective frame ensures that the space plane associated to homographies H^{Π}_{1i} is the same. In this way we obtain an estimate for a fixed reference plane that does not depend on a particular choice of the corresponding points which generate the projective reconstruction. This has clear advantages over other strategies such as tracking 3D points belonging to a plane along the video sequence, or by considering the dominant collineation [14], [3].

A projective bundle adjustment is run eventually over cameras and sparse triangulated 3D points in order to improve the reconstruction precision.

C. Parallax Map Computation

Each iteration of this stage is aimed to calculate several estimates of the parallax map of I_1 using each pair of views I_1, I_i for $i = 2, \ldots, N$ independently.

The input for this stage are the collineations H^{Π}_{1i} obtained from Eq. 8, and the epipoles $\mathbf{e}_i$ estimated from epipolar geometry derived from the multi-matching step. Observe that H^{Π}_{1i} and $\mathbf{e}_i$ together constitute an uncalibrated description of the motion between views I_1 and I_i, where the collineation is a proxy for the rotation and the epipole for the translation.

The computation of each parallax map is divided in the following three subsequent steps.

1) Uncalibrated Epipolar Rectification: Since camera calibration parameters are unknown, an uncalibrated rectification procedure based on sparse correspondences is applied to the pair of views I_1, I_i, see [2] for details.

2) Stereo Matching: Once the images pair I_1, I_i is rectified, dense correspondences can be obtained using any stereo matching algorithm. In our experiments we used a simple block-matching with Census transform [22] as a matching score. As a confidence measure we integrated the left-right consistency (LRC) check with an indicator of the curvature of

the matching score curve around the maximum. Namely, the confidence associated to the disparity computed at pixel x is:

$$\chi(\mathbf{m}) := \begin{cases} 0 \quad \text{if pixel } \mathbf{m} \text{ fails the LRC check} \\ \dfrac{2+2c(d_m)-c(d_m-1)-c(d_m+1)}{4} \quad \text{o/w} \end{cases} \tag{9}$$

where d_m is the disparity value assigned to pixel $\mathbf{m}$ and $c(d)$ denotes the matching score – normalized in $[0,1]$ – associated to disparity d. According to this definition $\chi(\mathbf{m})$ varies in $[0,1]$, where 0 means that pixel $\mathbf{m}$ is "totally unreliable" and 1 means "maximally confident" (in practice this value is hardly attainable).

Dense correspondences are then transferred back to the original reference images by applying the inverse of the rectifying homographies (de-rectification).

3) Parallax calculation: each parallax map is computed according to Eq. 3, where $(\mathbf{m}_1^k, \mathbf{m}_i^k)$ with $k = 1, \ldots, K$ is the dense set of correspondences on the pair of views I_1, I_i, and H_{1i}^{Π} and $\mathbf{e}_i$ are computed as explained before.

D. Multiple Maps Integration

Each of the $N-1$ independent estimates of the parallax map obtained with our iterative procedure contains errors and valuable information: the goal of integration is to enhance the latter while smoothing out the former. The rationale behind maps integration is twofold: from one side, a small baseline implies few occlusions, an easier stereo matching but raw quantization of the disparity, i.e. matching is more accurate but estimated parallax is less precise. On the other side, a large baseline implies better quantization of the disparity but more occlusions and harder matching, i.e. the estimated parallax is more precise, but the matching is less accurate.

A first possibility for maps integration is averaging, simplicity being its main attraction. However, by modeling temporal variations through a simple Bayesian filter, one expects to achieve better results. In particular, our temporal integration of parallax data is performed through a simple 1-d Kalman filter with constant state and direct measure model (which is indeed a weighted average):

State model: $x(t+1) = x(t) + w(t) \quad \text{var}[w(t)] = q(t)$

Measure model: $z(t) = x(t) + v(t) \quad \text{var}[v(t)] = r(t)$

Prediction: $x(t)^- = x(t-1)^+ \quad p(t)^- = p(t-1)^+ + q$

Update: $x(t)^+ = \dfrac{x(t)^- r + p(t)^- z}{p(t)^- + r} \quad p(t)^+ = \dfrac{p(t)^- r}{p(t)^- + r}.$

The process noise $w(t)$ models the quantization error made by extrapolating a given disparity to a larger baseline. Assuming the baseline increase from one frame to the next, the process noise variance q can be considered constant. The measurements noise $v(t)$ models the (inverse) reliability of the disparity measure, hence its variance $r(t)$ is inversely related to the confidence χ defined in Eq. 9 according to the formula: $r(t) = -\log(\chi)$.

Finally, observe that before the integration, parallax maps must be normalized, as independent estimates differ from each other by a scale factor. If one computes this ratio pixel-wise, however, results will differ due to noise and outliers. Therefore a simple robust estimation procedure is employed to estimate the correct scale scale factor for each map.

TABLE I
ERROR RATES [%] OF DISPARITY MAPS OBTAINED WITH DIFFERENT STRATEGIES (SEE TEXT FOR EXPLANATION).

Data	Optimal	Best map	Max conf	Average	Kalman
Art	30.01	48.48	42.99	34.03	32.30
Cloth1	15.13	24.40	28.14	16.27	15.72
Dolls	23.10	40.80	40.17	27.88	26.74
Moebius	28.09	41.43	44.78	31.32	30.85
Reindeer	21.82	36.42	38.20	25.33	23.01
Books	25.90	46.77	45.92	33.53	31.73

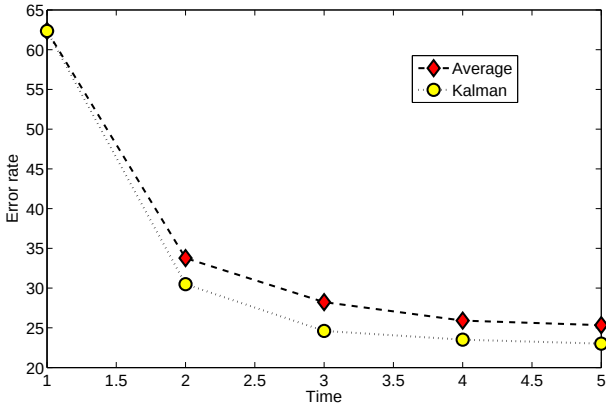

Fig. 3. Error rate vs time for the "Reindeer" sequence from the Middlebury dataset [15].

IV. Experiments and Results

We run two sets of experiments. The first one is devoted to validate the benefit of disparity integration with a simple geometry and uses image sequences from the Middlebury dataset [15]. The second one aims at testing the whole pipeline at work, including the geometric part, and uses more general sequences (from [23] and office-made) without ground truth.

A. Disparity integration in rectified context

In the experiments with the Middlebury dataset, since the images are already rectified, we work under the assumption that the camera motion is along the X axis. Thus we can compare the ground truth disparity maps directly with our parallax maps.

The error rate is defined as the percentage of retrieved parallax values whose difference with the ground truth is greater than one [15]. Pixels marked as occluded in the ground truth have been left out from the count. Figure 3 shows an example of how the error rate decreases as more parallax maps are integrated. Observe also how the Kalman filter integration performs better than averaging, as expected.

In Table I the error rate of parallax maps obtained by our method is compared against two touchstones, the error rate

of the "optimal selection" map (**Optimal** in the table) and minimum error rate among all the input parallax maps (**Best Map** in the table). The former is the error rate obtained if we could somehow choose the optimal parallax value among all the input parallax estimates for each pixel [6]. If the correct value is not achieved in (at least) one of the estimates, an error occurs for that pixel. Instead, the minimum error rate of the input parallax maps is an indicator of whether the integration is beneficial with respect to a standard two-views stereo. We compared our approach to two baseline strategies: the maximum confidence selection (**Max conf** in the table) and pure averaging (**Average** in the table). The former consists in selecting, for each pixel, the disparity that achieved the maximum confidence χ, whereas the latter consists in computing the average of the parallax values. Our strategy (**Kalman** in the table) achieves the best results with respect to averaging and maximum confidence selection strategies. We always outdistance the best input map and, in some cases, get close to the optimal.

B. Parallax map integration with general camera motion

For the second set of experiments we provide qualitative results only, since ground truth maps were not available. Tests are performed on sequences from [23] ("Road" and "Flowers") and on a *office-made* video[1] ("Milo"). In Fig.s 4, 5 and especially in Fig. 6, we can appreciate the benefits of maps integration, even when maps generated by single view-pairs are characterized by very low quality. The map resulting from the integration process shows progressive improvements in several aspects: values corresponding to surfaces are smoother, higher accuracy on object contours and *disocclusion* for many pixels.

V. Conclusion and Future Directions

This paper describes an ongoing work, which we plan to extend along several lines. Spatial consistency has to be integrated since the parallax estimate is obtained without considering any neighborhood information: more sophisticated strategies are likely to be more effective, but the goal of this paper was to investigate temporal integration in isolation. More confidence measures have to be considered and compared, and the stereo matching module can be substituted by a more sophisticated one.

Regarding stereo matching, it should be noted that being based on epipolar rectification, our current implementation does not allow forward motion, when the epipole is within the image. A more general stereo matcher that does not assume rectified images should be employed in order to avoid rectification altogether and compute planar parallax directly.

Appendix A
Equality of two vectors up to a scale

Let $\mathbf{a}$ and $\mathbf{b}$ two vectors of $\mathbb{R}^n$. Their equality up to a scale can be written as: $\mathrm{rank}[\mathbf{a},\mathbf{b}] = 1$. This is tantamount to say that all minors of $[\mathbf{a},\mathbf{b}]$ are zero. There are $n(n-1)/2$ of such order-two minors, and they can be obtained by multiplication of $\mathbf{b}$ by a suitable $n(n-1)/2 \times n$ matrix that contains the entries of $\mathbf{a}$. Let us call this matrix $[\mathbf{a}]_\times$ in analogy to the $\mathbb{R}^3$ case, where equality up to a scale reduces to $\mathbf{a}\times\mathbf{b} = 0$. Since, by construction, $\mathbf{a}$ belongs to the null-space of $[\mathbf{a}]_\times$, its rank is at most $n-1$. Hence $\mathbf{a} \simeq \mathbf{b}$ gives rise to the linear system of $n(n-1)/2$ equations $[\mathbf{a}]_\times \mathbf{b} = 0$ where only $n-1$ of them are independent. The matrix $[\mathbf{a}]_\times$ is composed by $n-1$ blocks arranged by rows. The the i^{th} block has $(n-i)$ rows and n columns $(i = 1 \dots n-1)$:

$$B_i = \begin{bmatrix} \mathbf{0}_{1\times(i-1)} & -a_{i+1} & a_i & 0 & 0 & \dots & 0 \\ \mathbf{0}_{1\times(i-1)} & -a_{i+2} & 0 & a_i & 0 & \dots & 0 \\ \mathbf{0}_{1\times(i-1)} & -a_{i+3} & 0 & 0 & a_i & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times(i-1)} & -a_n & 0 & 0 & 0 & \dots & a_i \end{bmatrix} \quad (10)$$

and

$$[\mathbf{a}]_\times = \begin{bmatrix} B_1 \\ \vdots \\ B_{n-1} \end{bmatrix}. \quad (11)$$

[1] Downloadable from www.diegm.uniud.it/fusiello/demo/dsp/

References

[1] M. Farenzena, A. Fusiello, and R. Gherardi. Structure-and-motion pipeline on a hierarchical cluster tree. In *IEEE International Workshop on 3-D Digital Imaging and Modeling*, 2009.

[2] A. Fusiello and L. Irsara. Quasi-euclidean epipolar rectification of uncalibrated images. *Machine Vision and Applications*, 22(4):663 – 670, 2011.

[3] M. Han and T. Kanade. Homography-based 3d scene analysis of video sequences. In *Proceedings of the DARPA Image Understanding Workshop*, 1998.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[5] J. Heel. Direct dynamic motion vision. In *Robotics and Automation, 1990. Proceedings., 1990 IEEE International Conference on*, pages 1142–1147 vol.2, 1990.

[6] X. Hu and P. Mordohai. Evaluation of stereo confidence indoors and outdoors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1466–1473, 2010.

[7] M. Irani and P. Anandan. Parallax geometry of pairs of points for 3D scene analysis. In *Proceedings of the European Conference on Computer Vision*, pages 17–30, 1996.

[8] S. B. Kang, J. A. Webb, C. L. Zitnick, and T. Kanade. A multibaseline stereo system with active illumination and real-time image acquisition. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 88–, 1995.

[9] H. Li and R. Hartley. Rectification-free multibaseline stereo for non-ideal configurations. In *Proceedings of the 13th international conference on Image Analysis and Processing*, pages 810–817, 2005.

[10] L. Matthies, T. Kanade, and R. Szelisky. Kalman filter based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3:209–236, 1989.

[11] R. A. Newcombe and J. Andrew. Live dense reconstruction with a single moving camera davison. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

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

[13] J. Santos-Victor and J. Sentiero. Generation of 3D dense depth maps by dynamic vision. In *British Machine Vision Conference*, pages 129–138, 1992.

[14] H. Sawhney and S. Ayer. Compact representations of videos through dominant and multiple motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):814–830, 1996.

[15] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):7–42, 2002.

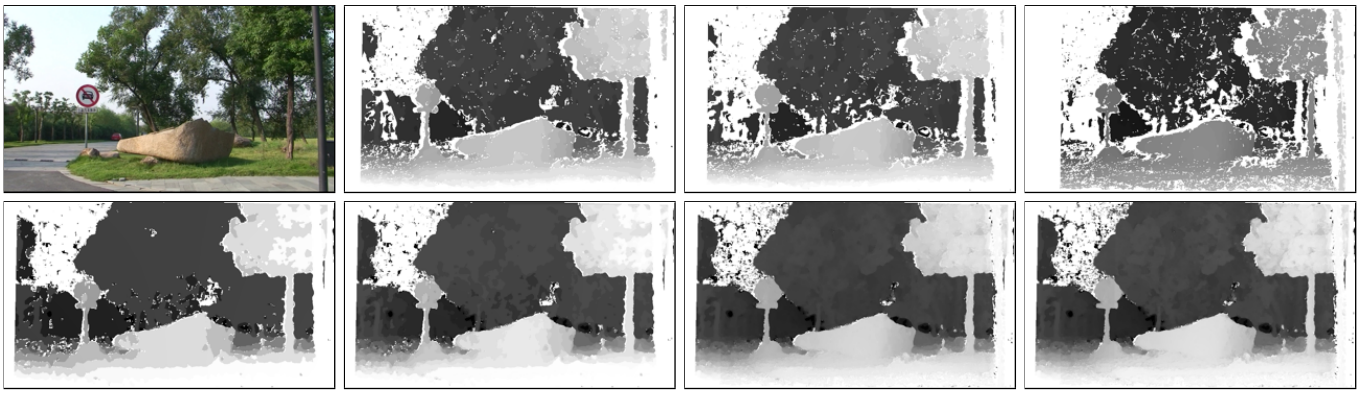

Fig. 4. Results for the "Road" sequence from [23]. Top row: reference frame, and some input parallax maps. Bottom row: incremental parallax map construction. The last picture is the final result using 20 frames.

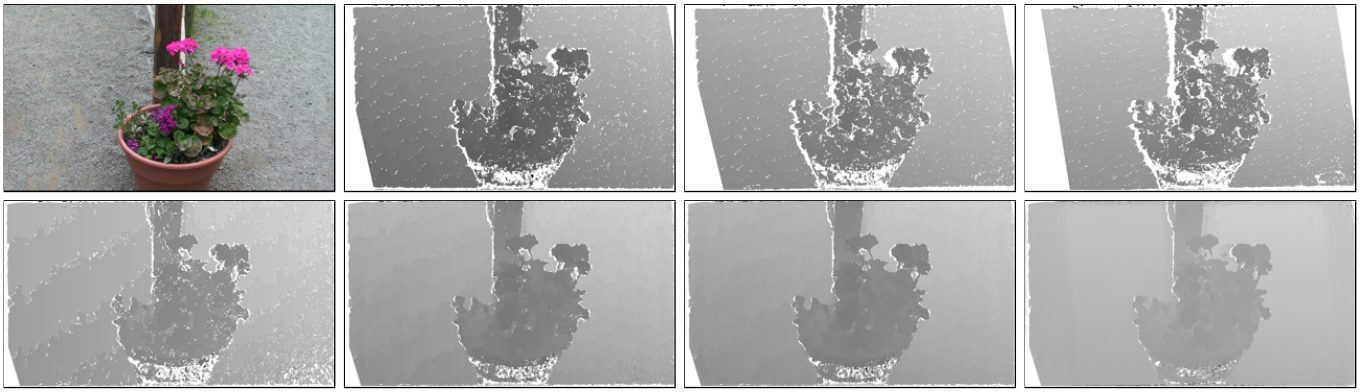

Fig. 5. Results for the "Flower" sequence from [23]. Top row: reference frame, and some input parallax maps. Bottom row: incremental parallax map construction. The last picture is the final result using 30 frames.

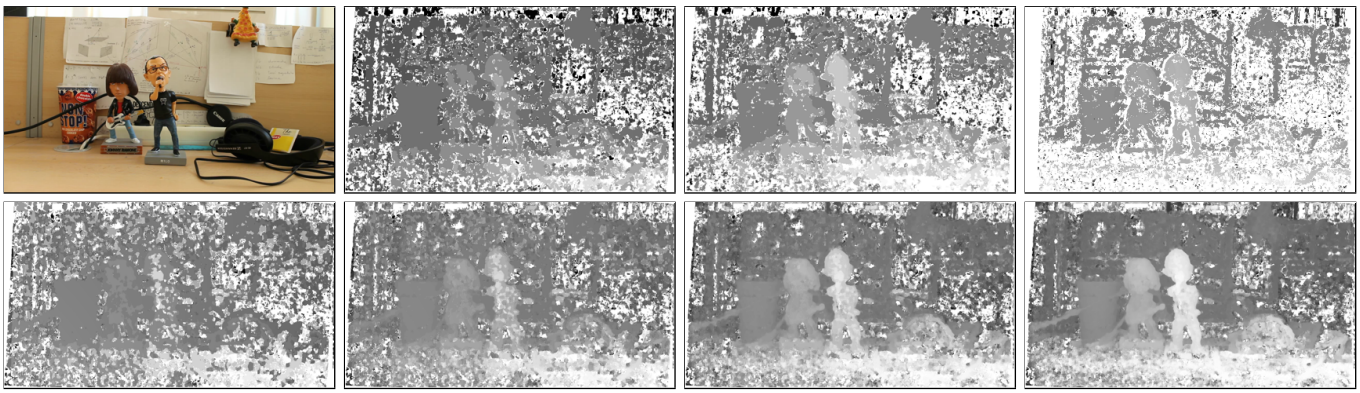

Fig. 6. Results for the "Milo" sequence. Top row: reference frame, and some input parallax maps. Bottom row: incremental parallax map construction. The last picture is the final result using 27 frames.

[16] A. Shashua and N. Navab. Relative affine structure: Canonical model for 3D from 2D geometry and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):873–883, 1996.

[17] A. P. Tirumalai, B. G. Schunk, and R. C. Jain. Dynamic stereo with self-calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(12), 1992.

[18] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:2000, 2000.

[19] E. Trucco, V. Roberto, S. Tinonin, and M. Corbatto. SSD disparity estimation for dynamic stereo. In *Proceedings of the British Machine Vision Conference*, pages 342–352, 1996.

[20] G. Vogiatzis and C. Hernàndez. Video-based, real-time multi-view stereo. *Image and Vision Computing*, pages 434–441, 2011.

[21] A. M. Waxman and S. S. Sinha. Dynamic stereo: Passive ranging to moving objects from relative image flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(4):406–412, 1986.

[22] R. Zabih and J. Woodfill. Non-parametric local transform for computing visual correspondence. In *Proceedings of the European Conference on Computer Vision*, volume 2, pages 151–158, 1994.

[23] G. Zhang, J. Jia, T.-T. Wong, and H. Bao. Consistent depth maps recovery from a video sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):974–988, 2009.