An Intensity-Based Method for the 3-D Motion and Structure Estimation from Binocular Image Sequences

A. Luo and H. Burkhardt
Technische Informatik I der TU Hamburg-Harburg
Postfach 90 10 52, 21071 Hamburg 90, Germany
email: luo@tu-harburg.d400.de or burkhardt@tu-harburg.d400.de

Abstract

This paper presents a new algorithm for estimating the 3-D motion and structure from stereo image sequences. In order to overcome the inherent ambiguities of the motion and structure estimation from a monocular image sequence and to improve the performance of the binocular stereo methods, both the intensity-based methods are directly integrated. Being different from other methods computing the structure from stereo and motion, this method does not need to match separate monocular optical flows, so that the high complexity of the algorithm is avoided. Instead of simplified models that are often unreasonable, new models for 3-D piecewise smooth structure and occlusion are put forward for Bayesian estimation. The experiments show that the algorithm is effective and robust to improve the 3-D motion and structure estimation.

Keywords: shape from motion, stereo and integration, motion estimation, piecewise smooth model of 3-D surfaces and Bayesian estimation

1 Introduction

The information about 3-D structure and motion is very important to many visual tasks, e.g. 3-D object recognition and robot navigation. 3-D structures can be estimated from stereo image pairs by a stereo matching method, or from image sequences by a method of "shape from motion" (SFM), whereas the cues of 3-D motion must be extracted from an image sequence, i.e. by SFM methods [Luo93].

The 3-D depth estimation from binocular stereo image pairs has been extensively studied in the past and is an essential technique of 3-D computer vision. In [LB92] an intensity-based stereo method is presented with simultaneous detection of discontinuities and occlusion. It has many advantages, i.e. direct dense depth estimation from grey images, and elimination of errors caused by occlusion and over-smoothening. As there discussed, however, there still exist large ambiguities for all binocular methods with matching between both stereo images somewhere due to the absence of intensity variations on the epipolar line. This can be solved only by using more visual cues.

Generally, the 3-D structure and motion estimation using feature-based or intensity-based SFM methods are divided into two steps, i.e. first computing the optical flow of an image sequence and then extracting the structure and motion parameters from it [Adi89]. Various methods about the optical flow estimation are put forward by many researchers. A method in [LB93] can estimate both the optical flow with discontinuities and the related occlusion effectively, where the main error sources of computing the optical flow can be satisfactorily eliminated. From the estimated optical flow or only few point-correspondences one can further get the 3-D structure and motion [AN88]. Feature-based methods [Ull79, TH84] estimate the 3-D motion by matching as few points or lines as possible, which need the accurate locations of features and their matches, and thus is sensitive to noise in real scenes. In intensity-based methods various a priori assumptions about 3-D surfaces are used for simplifying the 3-D estimation [LHP80, WKS84, Sub89, DB86, NH87], where surfaces are assumed as smooth and even high-order differentiable or planar, or only a pure rotational or translational motion is assumed. Besides, the motion estimation from noisy optical flows (especially near occluded regions and discontinuities) is inherently ambiguous for a real scene. Moreover, no absolute solution of 3-D motion and structure can be obtained from a monocular image sequence by any above method.

In order to avoid scalar ambiguities and improve the estimation performance, the dynamic stereo is thus introduced by the integration of stereo and SFM methods [Mit84, Ric85, WS86, WD86]. However, they need matching two monocular optical flows, which is very complex.

In this paper we develop a new algorithm which integrates the methods of stereo and SFM, where monocular optical flows are not needed to match and the above difficulties of the integration are avoided. The 3-D structure estimate that is obtained by a binocular method can greatly simplify the complexity of 3-D motion estimation. Using this initial information of the structure, the complete motion parameters can be obtained without a scalar ambiguity by a direct closed solution. In this method the strict assumptions about 3-D motion, optical flow and 3-D surface are not needed, where only a reasonable model for the piecewise smooth depth field and optical flow is used for their estimation. The new methods in [LB92, LB93] guarantee a satisfactory solution of estimating the depth field and optical flow with occlusion information, such that accurate motion parameters can be estimated. After obtaining the motion parameters, the initial depth estimates can be further greatly improved by the integration of binocular and motion stereo, which are much better than that of these both individual methods. In the integration a new a priori piecewise smooth model and a similar method to the trinocular stereo vision of [LB92] are applied, where the first stereo pair and one image of the second pair are used. The other image of the second pair is not used, because in spite of higher complexity no obvious improvement is obtained, as discussed in [Aya91]. In the new algorithm estimating 3-D structure and motion from a binocular image sequence the occlusion information is especially considered to be detected and used. The algorithm has a moderate complexity and is also easy to parallel implement.

This paper is organized as following: After introducing some motion models and models of the depth field and occlusion in next section, a closed solution of estimating the motion parameters from binocular image sequences is first given in section 3 and an improvement of the depth estimate by the integration of binocular and motion stereo in section 4. Then follows a conclusion.

2 Models of motion, depth field and occlusion

Using a usual viewer centered coordinate system, the optical flow field exists due to an arbitrary camera motion relative to a rigid scene (the original form in [LHP80]):

$$\mathbf{u}(\mathbf{x}) = \frac{1}{Z(x,y)}\mathbf{P}(\mathbf{x})\mathbf{T} + \mathbf{Q}(\mathbf{x})\mathbf{W} \tag{1}$$

with

$$\mathbf{P}(\mathbf{x}) = \begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix} \quad \text{and} \quad \mathbf{Q}(\mathbf{x}) = \begin{pmatrix} xy/f & -f - x^2/f & y \\ f + y^2/f & -xy/f & -x \end{pmatrix} \tag{2}$$

where f is the focal length of a camera, $\mathbf{u}$, Z, $\mathbf{x} = (x,y)^T$, $\mathbf{T} = (T_X, T_Y, T_Z)^T$ and $\mathbf{W} = (W_X, W_Y, W_Z)^T$ are the optical flow, depth, image coordinate, translational and rotational motion parameters relative to a camera system.

Based on the works in [LB92] a new a priori model of Markov's random fields with the following local potential function can be satisfactorily used for a piecewise smooth depth field or optical flow, where $Z_{(med)}(\mathbf{x})$ is the local median value of a neighborhood of $Z(\mathbf{x})$:

$$V_{\mathbf{x}}(\mathbf{Z}) = (Z(\mathbf{x}) - Z_{(med)}(\mathbf{x}))^2 \tag{3}$$

Suppose that $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ are the first and second image of a sequence. If $\mathbf{x}$ is outside occluded regions of $g_1(\mathbf{x})$, i.e. the occlusion map $O_1(\mathbf{x}) = 1$, then we get:

$$g_2(\mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}/Z_1(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}) - g_1(\mathbf{x}) = n_1(\mathbf{x}). \tag{4}$$

and a similar constraint exists also, if $\mathbf{x}$ is outside occluded regions of $g_2(\mathbf{x})$, i.e. the occlusion map $O_2(\mathbf{x}) = 1$:

$$g_1(\mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}'/Z_2(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}') - g_2(\mathbf{x}) = n_2(\mathbf{x}) \tag{5}$$

where $\mathbf{T}' = -\mathbf{T} - \mathbf{W} \times \mathbf{T}$ and $\mathbf{W}' = -\mathbf{W}$ are the parameters of the virtual inverse motion process from $g_2(\mathbf{x})$ to $g_1(\mathbf{x})$.

Both the depth maps $Z_1(\mathbf{x})$ and $Z_2(\mathbf{x})$ in different camera systems should represent a consistent 3-D structure of a scene with the following constraints:

$$\begin{aligned} Z_2(\mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}/Z_1(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}) &= -T_Z + (1 + (xW_Y - yW_X)/f)Z_1(\mathbf{x}) & (6)\\ \text{or} \quad Z_1(\mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}'/Z_2(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}') &= -T'_Z + (1 + (xW'_Y - yW'_X)/f)Z_2(\mathbf{x}) & (7)\end{aligned}$$

with the help of a simple relation $\delta Z = -T_Z + XW_Y - YW_X$.

In order to compute the occlusion maps from the estimated depth information, we present a transformation $\mathcal{OT}$, e.g. $O_1(\mathbf{x}) = \mathcal{OT}_{T'W'}(Z_2(\mathbf{x}))$ and $O_2(\mathbf{x}) = \mathcal{OT}_{TW}(Z_1(\mathbf{x}))$ with known motion parameters:

$$\begin{aligned} \{\mathbf{x}'|O_1(\mathbf{x}') = 1, \mathbf{x}' \in \mathcal{B}\} &= \{\mathbf{x}'|\mathbf{x}' = \mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}'/Z_2(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}', \mathbf{x} \in \mathcal{B}\} & (8)\\ \text{and} \quad \{\mathbf{x}'|O_2(\mathbf{x}') = 1, \mathbf{x}' \in \mathcal{B}\} &= \{\mathbf{x}'|\mathbf{x}' = \mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}/Z_1(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}, \mathbf{x} \in \mathcal{B}\}. & (9)\end{aligned}$$

One must notice that the occlusion information can easily be obtained only from the depth map in the other camera system.

About the intensity constraints of a stereo image pair and the consistent constraints of the depth maps in individual stereo systems one can find similar results in [LB92], where the similar transformations $O_r(\mathbf{x}) = \mathcal{OT}_{LR}(Z_l(\mathbf{x}))$ and $O_l(\mathbf{x}) = \mathcal{OT}_{RL}(Z_r(\mathbf{x}))$ can also be obtained. Actually, it is a special example of the above models with $\mathbf{T} = (B, 0, 0)^T$ and $\mathbf{W} = (0, 0, 0)^T$. Here B is the baseline length of stereo cameras.

3 Estimation of motion parameters by dynamic stereo

The new method of directly estimating the 3-D motion of cameras from a stereo image sequence is composed of two steps:

1. Computing the depth maps of the related camera systems from the first stereo image pair with the binocular stereo method [LB92] and the associated optical flows between the first and second frame from both individual monocular image sequences of the stereo image sequence with the new method in [LB93].

2. Solving a linear equation system to further estimate the motion parameters of the cameras from the results of the first step based on the criteria of the least mean square (LMS) error of optical flows.

From the first stereo image pair $\mathbf{g}_{1l}$ and $\mathbf{g}_{1r}$ one can estimate the depth maps $\mathbf{Z}_{1l}$ and $\mathbf{Z}_{1r}$ with simultaneous detection of the occlusion maps $\mathbf{O}_l$ and $\mathbf{O}_r$ in the associated coordinate systems by using the intensity-based cooperative bidirectional stereo matching [LB92]. Besides, the optical flows $\mathbf{u}_l$ from $\mathbf{g}_{1l}$ to $\mathbf{g}_{2l}$ and $\mathbf{u}_r$ from $\mathbf{g}_{1r}$ to $\mathbf{g}_{2r}$ can be respectively obtained by the new algorithm of computing optical flow [LB93], where the occlusions $\mathbf{O}_{1l}$ and $\mathbf{O}_{1r}$ relative to the next images are detected also.

Suppose that the left camera moves with parameters $\mathbf{T}$ and $\mathbf{W}$. Then the right camera of an epipolar system with a known baseline of B has the motion parameter:

$$\mathbf{T}_r = \mathbf{T} + \mathbf{W} \times B\vec{\imath} = \mathbf{T} + \mathbf{SW} \quad \text{and} \quad \mathbf{W}_r = \mathbf{W} \tag{10}$$

with $\vec{\imath} = (1,0,0)^T$ and

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & B \\ 0 & -B & 0 \end{pmatrix}. \tag{11}$$

According to the criteria of the least mean square error one can apply the following object function through Eqs. (1) and (10) to estimate the motion parameter:

$$\begin{aligned} E(\mathbf{T},\mathbf{W}) \;=\; & \sum_{\mathbf{x}} (O_{1l}(\mathbf{x})O_l(\mathbf{x})(\mathbf{u}_l(\mathbf{x}) - \frac{1}{Z_{1l}(\mathbf{x})}\mathbf{P}(\mathbf{x})\mathbf{T} - \mathbf{Q}(\mathbf{x})\mathbf{W})^2 + O_{1r}(\mathbf{x}) \\ & O_r(\mathbf{x})(\mathbf{u}_r(\mathbf{x}) - \frac{1}{Z_{1r}(\mathbf{x})}\mathbf{P}(\mathbf{x})(\mathbf{T} + \mathbf{SW}) - \mathbf{Q}(\mathbf{x})\mathbf{W})^2) \end{aligned} \tag{12}$$

For decreasing the complexity, one can only estimate a optical flow $\mathbf{u}_l$ from the left monocular sequence and use it in the above object function with $O_{1r}(\mathbf{x}) = 0$ for all $\mathbf{x}$.

Because this object function is quadrant and its minimum has a zero differentiation value, we can obtain the optimal estimate of the motion parameters by solving a linear equation system:

$$\begin{aligned} \mathbf{H}_1\mathbf{T} + \mathbf{H}_2\mathbf{W} &= \mathbf{H}_3 \\ \mathbf{H}_4\mathbf{T} + \mathbf{H}_5\mathbf{W} &= \mathbf{H}_6 \end{aligned} \tag{13}$$

with the matrices $\mathbf{H}_1$ to $\mathbf{H}_6$ that can easily be obtained directly from the estimated depth fields, optical flows and the related occlusions.

With the above algorithm one get the estimate of motion parameters directly from the monocular optical flows and depth maps without correspondence. As the occluded regions are detected, the wrong estimates there can be avoided to use. Beside the property of weak smoothness, no more strict assumptions for the estimated fields are necessary to use in the algorithm.

For illustrating the effectiveness of this algorithm we give a example as follows. Fig. 1 shows a stereo image sequence, where actually the cameras with the focal length of $f = 8mm = 225\text{pixels}$ and the baseline length of $B = 2\text{cm}$ have a known motion of $\mathbf{T} = (1.0, 1.65, -1.8)^T$ cm and $\mathbf{W} = (0.029, -0.016, 0.0)^T$ between the neighboring frames. We can use the new algorithm to estimate the motion parameters of cameras only from the grey stereo image sequence of Fig. 1. At first we directly estimate the necessary depth maps and optical flows with the methods in [LB92, LB93]. Based on them the following estimates of motion parameters are obtained by solving the above linear equation system:

$$\mathbf{T} = (1.0222, 1.4982, -1.7899)^T cm \quad \text{and} \quad \mathbf{W} = (0.0290, -0.0168, 0.0009)^T$$

From this results one can see that the motion parameters have been well estimated by the new algorithm, which are almost consistent to the actual motion.

4 Improvement of 3-D depth estimation by dynamic stereo

As in [LB92] discussed, there still exist large ambiguities for all binocular methods with matching between both stereo images somewhere due to the absence of intensity variations on the epipolar lines. There the triangulation geometry of binocular stereo fails to detect accurate disparities and only smooth values of their neighborhoods can be assigned to them. In order to further improve the depth estimates of binocular stereo and to decrease their ambiguities, one can integrate the motion stereo and static binocular stereo with dynamic stereo methods, where more epipolar geometry constraints can be used. The motion parameters are known or can be estimated as in the previous section.

Based on the new a-priori Bayesian model of piecewise smooth fields (3) and the triangulation geometry of grey images $\mathbf{g}_{1l}$, $\mathbf{g}_{1r}$ and $\mathbf{g}_{2l}$, one can get the a-posteriori energy functions of the depth fields in the individual coordinate systems:

$$\begin{aligned}
U_{p1l}(\mathbf{Z}_{1l}) &= \sum_{\mathbf{x}}(O_l(\mathbf{x})(g_{1r}(\mathbf{x}_r) - g_{1l}(\mathbf{x}))^2 + O_{1l}(\mathbf{x})(g_{2l}(\mathbf{x}_2) - g_{1l}(\mathbf{x}))^2 + \lambda V_{\mathbf{x}}(\mathbf{Z}_{1l})) \\
U_{p1r}(\mathbf{Z}_{1r}) &= \sum_{\mathbf{x}}(O_r(\mathbf{x})(g_{1l}(x + Bf/Z_{1r}(\mathbf{x}), y) - g_{1r}(\mathbf{x}))^2 + \lambda V_{\mathbf{x}}(\mathbf{Z}_{1r})) \\
U_{p2l}(\mathbf{Z}_{2l}) &= \sum_{\mathbf{x}}(O_{2l}(\mathbf{x})(g_{1l}(\mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}'/Z_{2l}(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}') - g_{2l}(\mathbf{x}))^2 + \lambda V_{\mathbf{x}}(\mathbf{Z}_{2l}))
\end{aligned}$$

with $x_r = x - Bf/Z_{1l}(\mathbf{x})$, $y_r = y$, $\mathbf{x}_2 = \mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}/Z_{1l}(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}$,
occlusions $O_l(\mathbf{x}) = \mathcal{OT}_{RL}(Z_{1r}(\mathbf{x}))$, $O_{1l}(\mathbf{x}) = \mathcal{OT}_{T'W'}(Z_{2l}(\mathbf{x}))$, $O_r(\mathbf{x}) = \mathcal{OT}_{LR}(Z_{1l}(\mathbf{x}))$
and $O_{2l}(\mathbf{x}) = \mathcal{OT}_{TW}(Z_{1l}(\mathbf{x}))$.

Similarly to the trinocular method in [LB92], one can minimize all the three a posteriori energy functions with the consistent constraints of depth maps simultaneously. Based on a two-stage algorithm, the convex approximations with $V_{\mathbf{x}}(\mathbf{Z}) = (Z_x^2(\mathbf{x}) + Z_y^2(\mathbf{x}))$ are first minimized by the descent method. Beginning with these estimates as initial value, the global optimal estimates are then obtained by using a descent method to minimize the original non-convex functions:

$$\begin{aligned}
Z_{1l}^{k+1}(\mathbf{x}) &= Z_{1l}^{k*}(\mathbf{x}) + \frac{O_l(\mathbf{x})}{\lambda}(g_{1r}(\mathbf{x}_r^+) - g_{1l}(\mathbf{x}))\frac{-Bfg_{1rx}(\mathbf{x}_r^+)}{(Z_{1l}^{k*}(\mathbf{x}))^2} + \\
&\quad \frac{O_{1l}(\mathbf{x})}{\lambda}(g_{2l}(\mathbf{x}_2^+) - g_{1l}(\mathbf{x}))\frac{(\mathbf{P}(\mathbf{x})\mathbf{T})^T \cdot \nabla g_{2l}(\mathbf{x}_2^+)}{(Z_{1l}^{k*}(\mathbf{x}))^2} \qquad (14) \\
Z_{1r}^{k+1}(\mathbf{x}) &= Z_{1r}^{k*}(\mathbf{x}) + \frac{O_r(\mathbf{x})}{\lambda}(g_{1l}(\mathbf{x}_l^+) - g_{1r}(\mathbf{x}))\frac{Bfg_{1lx}(\mathbf{x}_l^+)}{(Z_{1r}^{k*}(\mathbf{x}))^2} \qquad (15) \\
Z_{2l}^{k+1}(\mathbf{x}) &= Z_{2l}^{k*}(\mathbf{x}) + \frac{O_{2l}(\mathbf{x})}{\lambda}(g_{1l}(\mathbf{x}_1^+) - g_{2l}(\mathbf{x}))\frac{(\mathbf{A}(\mathbf{x})\mathbf{T}')^T \cdot \nabla g_{1l}(\mathbf{x}_1^+)}{(Z_{2l}^{k*}(\mathbf{x}))^2} \qquad (16)
\end{aligned}$$

with $x_r^+ = x - Bf/Z_{1l}^{k*}(\mathbf{x})$, $x_l^+ = x + Bf/Z_{1r}^{k*}(\mathbf{x})$ and $y_r^+ = y_l^+ = y$,
$\mathbf{x}_2^+ = \mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}/Z_{1l}^{k*}(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}$ and $\mathbf{x}_1^+ = \mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}'/Z_{2l}^{*(n)}(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}'$,
where the following predicted values are used, such that the stability of the algorithm is improved and all depth maps are enforced to represent a consistent 3-D structure:

$$\begin{aligned}
O_l(\mathbf{x}) &= \mathcal{OT}_{RL}(Z_{1r(op)}(\mathbf{x})) \qquad (17) \\
O_{1l}(\mathbf{x}) &= \mathcal{OT}_{T'W'}(Z_{2l(op)}(\mathbf{x})) \qquad (18) \\
O_r(\mathbf{x}) &= \mathcal{OT}_{LR}(Z_{1l(op)}(\mathbf{x})) \qquad (19) \\
O_{2l}(\mathbf{x}) &= \mathcal{OT}_{TW}(Z_{1l(op)}(\mathbf{x})) \qquad (20)
\end{aligned}$$

$$Z^*_{1l}(\mathbf{x}) = \frac{Z_{1l(op)}(\mathbf{x}) + O_r(\mathbf{x}_r)Z_{1r(op)}(\mathbf{x}_r) + O_{2l}(\mathbf{x}_2)\frac{Z_{2l(op)}(\mathbf{x}_2)+T_Z}{1+(xW_Y-yW_X)/f}}{1 + O_r(\mathbf{x}_r) + O_{2l}(\mathbf{x}_2)} \tag{21}$$

$$Z^*_{1r}(\mathbf{x}) = \frac{Z_{1r(op)}(\mathbf{x}) + O_l(\mathbf{x}_l)Z_{1l(op)}(\mathbf{x}_l)}{1 + O_l(\mathbf{x}_l)} \tag{22}$$

$$Z^*_{2l}(\mathbf{x}) = \frac{Z_{2l(op)}(\mathbf{x}) + O_{1l}(\mathbf{x}_1)(Z_{1l(op)}(\mathbf{x}_1) + T'_Z)/(1 + (xW'_Y - yW'_X)/f)}{1 + O_{1l}(\mathbf{x}_1)} \tag{23}$$

with the abbreviations $x_r = x - Bf/Z_{1l(op)}(\mathbf{x})$, $x_l = x + Bf/Z_{1r(op)}(\mathbf{x})$ and $y_r = y_l = y$, $\mathbf{x}_2 = \mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}/Z_{1l(op)}(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}$ and $\mathbf{x}_1 = \mathbf{x} + \mathbf{P}(\mathbf{x})\mathbf{T}'/Z_{2l(op)}(\mathbf{x}) + \mathbf{Q}(\mathbf{x})\mathbf{W}'$. In the two-stage optimizing process, $(op) = (ave)$ is used in the first steps and then $(op) = (med)$.

This algorithm is implemented with the hierarchical multilevel structure. The cooperative bidirectional relaxation processes are progressively carried out from the coarsest to the finest level, and the optimal results in a coarser level provide the good initial estimates for the finer level, such that local optimal estimates the local properties of intensity variations lead to are best avoided and the global optimal estimates are guaranteed with the greatest possibility.

To illustrate the improvement of depth estimates using this integration algorithm, a example is given also with the stereo image sequence in Fig. 1 as follows. The scene contains a poster as background, a carton and a cup as objects, which are respectively 122cm, 82cm and 52cm far away from the camera at the beginning. One can estimate the depth field from the first stereo pair by the binocular stereo or from the left monocular sequence by the motion stereo with the known camera motion, and the profiles of both estimated depth fields are shown in Fig. 2. When the motion parameters are unknown, we can estimate them by the method in the previous section. With these estimated parameters $\mathbf{T} = (1.0222, 1.4982, -1.7899)^T$cm and $\mathbf{W} = (0.0290, -0.0168, 0.0009)^T$, we further apply the integration algorithm to estimate the 3-D structure of a scene. The estimated results of $\mathbf{Z}_{1l}$, $\mathbf{Z}_{1r}$, $\mathbf{O}_l$, $\mathbf{O}_r$, $\mathbf{O}_{1l}$, $\mathbf{O}_{2l}$ and $\mathbf{Z}_{2l}$ in different coordinate systems are shown in Fig. 3. The estimated 3-D structure is almost completely consistent with the actual scene. All occluded regions are qualitatively well detected and die absolutely occluded regions are greatly decreased. In order to directly compare with the other methods, the corresponding profile of this estimate is also shown in Fig. 2, which is much better. The average relative error of the depth map within the background and objects is only 3.5%, whereas the errors of the estimated depth maps with binocular or motion stereo is ca. 5%.

5 Conclusion

In this paper a new dynamic stereo algorithm for the 3-D motion and structure estimation is put forward, which directly integrates the stereo and SFM methods, such that the associated estimation ambiguities of the individual methods are greatly eliminated. Being different from other dynamic stereo methods, this method does not need the correspondence of separate monocular optical flows and other simplified unreasonable assumptions beside the piecewise smoothness. The absolute motion parameter can be well estimated by solving a linear equation system. Besides, the direct depth estimates from a grey stereo sequence using the Bayesian method and occlusion detection are strongly improved by the integration, because the absolutely occluded regions are greatly decreased, and outside these regions one can use more grey information and more epipolar lines as matching constraints [Luo93]. This intensity-based method of integration has a moderate complexity and can be easily parallel implemented.

References

[Adi89] G. Adiv,Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field, IEEE Trans. Pattern Anal. Mach. Intelligence, Vol. 11, 1989, pp. 477-489

[Aya91] N. Ayache, Artificial vision for mobile robots: stereo vision and multisensory perception, MIT Press, 1991

[AN88] J.K.Aggarwal and N.Nandhakumar, On the computation of motion from Sequences of images - a review, Proceedings of IEEE, Vol.76, 1988, pp. 917-935

[DB86] N. Diehl and H. Burkhardt, Planar motion estimation with a fast converging algorithm, Proc. of 8th IC on Pattern Recognition, Paris, 1986, pp. 1099-1102

[LB92] A. Luo and H. Burkhardt, An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions, accepted by Int. Journal of Computer Vision.

[LB93] A. Luo and H. Burkhardt, Bestimmung des optischen Flusses aus Bildfolgen unter Berücksichtigung von Diskontinuität and Okklusion, in S.J.Pöppl and H.Handels, Hrsg., 15. DAGM-Symposium: "Mustererkennung 1993", S. 59-66, Lübeck, 1993. Reihe Informatik aktuell, Springer.

[Luo93] A. Luo, Helligkeitsbasiertes Rechnersehen zur direkten Ermittlung räumlicher Eigenschaften, Ph.D. Dissertation, Technische Universität Hamburg-Harburg, December 1993.

[LHP80] H.C.Longuet-Higgins and K.Prazdny, The interpretation of a moving retinal image, Proc. Royal Society London B, Vol.208, 1980, pp. 385-397

[Mit84] A. Michiche, On combining stereopsis and kineopsis for space perception, in Proceedings of the 1st Conf. on Artificial Intelligence, Denver, 1984, pp. 156-160

[NH87] S. Negahdaripour and B. K. P. Horn, Direct passive navigation, IEEE Trans. Pattern Anal. Mach. Intelligence, Vol. 9, 1987, pp. 168-176

[Ric85] W. Richards, Structure from stereo and motion, Journal of Optical Society of America A, Vol. 2, 1985, pp. 343-349

[Sub89] M. Subbarao, Interpretation of image flow: a spatio-temporal approach, IEEE Trans. Pattern Anal. Mach. Intelligence, Vol. 11, 1989, pp. 266-278

[TH84] R. Y. Tsai and T. S. Huang, Uniqueness and estimation of 3-D motion parameters of rigid objects with curved surface, IEEE Trans. Pattern Anal. Mach. Intelligence, Vol. 6, 1984, pp. 13-26

[Ull79] S. Ullman, The interpretation of visual motion, MIT Press, 1979

[WKS84] A. M. Waxman, B. Kamgar-Parsi and M. Subbarao, Closed form solutions to image flow equations, in Proceedings of the 1st Conf. on Artificial Intelligence, Denver, 1984, pp. 156-160

[WS86] A. M. Waxman and S. Sinha, Dynamic stereo: Passive ranging to moving objects from relative image flow, IEEE Trans. Pattern Anal. Mach. Intelligence, Vol. 8, 1989, pp. 406-412

[WD86] A. M. Waxman and J. H. Duncan, Binocular image flow: Step toward stereo-motion fusion, IEEE Trans. Pattern Anal. Mach. Intelligence, Vol. 8, 1986, pp. 715-729

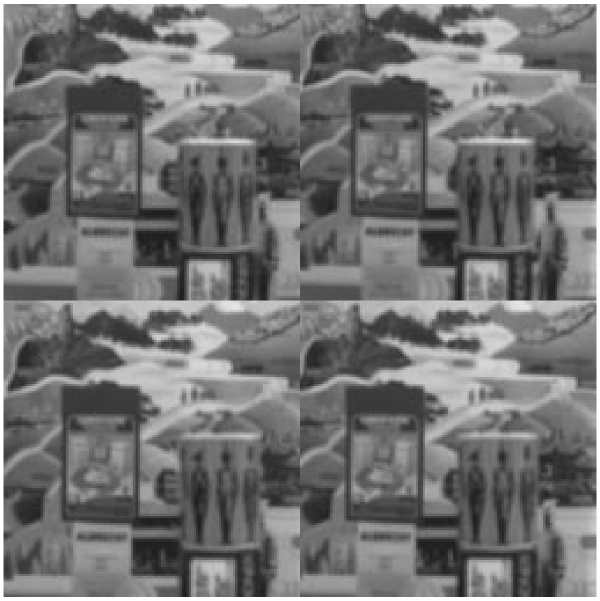

Figure 1: A stereo image sequence with known parameters of $\mathbf{T} = (1.0, 1.65, -1.8)^T$ cm and $\mathbf{W} = (0.029, -0.016, 0.0)^T$: (a) and (b): left and right image of the second stereo pair, (c) and (d): left and right image of the first stereo pair

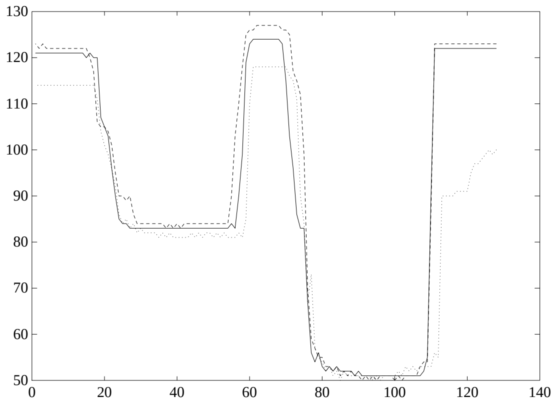

Figure 2: Profiles of three depth estimates $\mathbf{Z}_{1l}$, which are obtained respectively with the Integration (real line), motion stereo (thick dotted line) and binocular stereo (thin dotted line),in $\frac{2}{5}$ height of the field

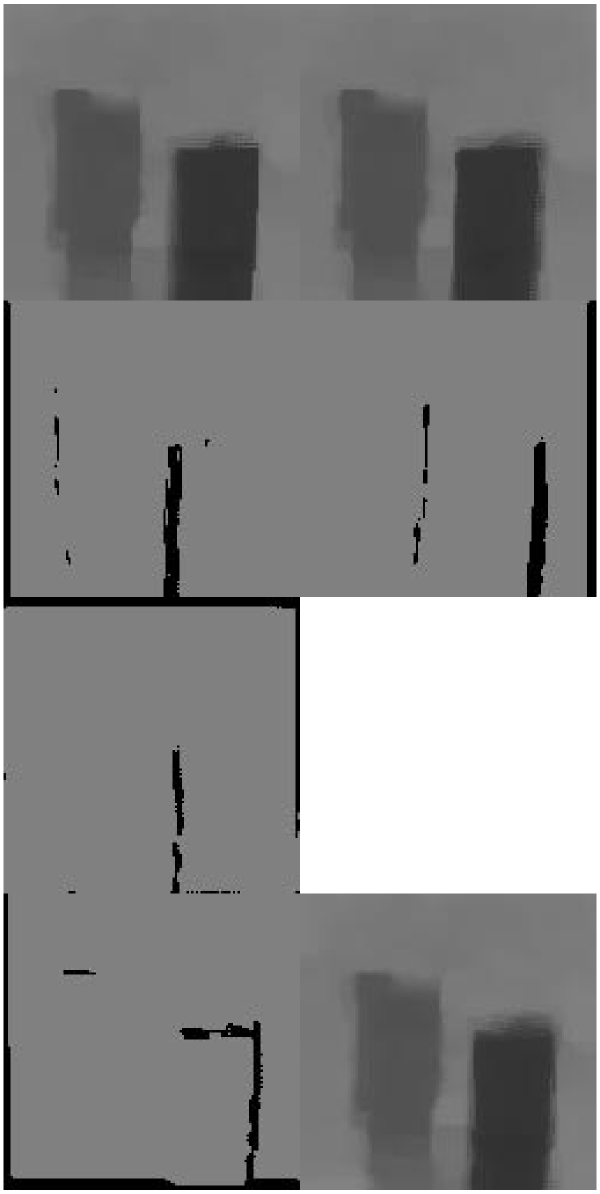

Figure 3: The results from the real stereo image sequence in Fig. 1 with the estimated motion parameters of $\mathbf{T} = (1.0222, 1.4982, -1.7899)^T$cm and $\mathbf{W} = (0.0290, -0.0168, 0.0009)^T$: (a) $\mathbf{Z}_{1l}$ and (b) $\mathbf{Z}_{1r}$, (c) $\mathbf{O}_l$ and (d) $\mathbf{O}_r$, (e) $\mathbf{O}_{1l}$, (f) $\mathbf{O}_{2l}$ and (g) $\mathbf{Z}_{2l}$