Disparity Estimation in Stereo Sequences using Scene Flow

Fang Liu
Fang.F.Liu@philips.com
Philips Germany

Vasanth Philomin
Vasanth.Philomin@philips.com
Philips Germany

Abstract

This paper presents a method for estimating disparity images from a stereo image sequence. While many existing stereo algorithms work well on a single pair of stereo images, it is not sufficient to simply apply them to temporal frames independently without considering the temporal consistency between adjacent frames. Our method integrates the state-of-the-art stereo algorithm with the scene flow concept in order to capture the temporal correspondences. It computes the dense disparity images and scene flow in a practical and unified process: the disparity is initialized by a hybrid stereo approach which employs the over-segmentation based stereo and pixelwise iterative stereo; then the scene flow, estimated via a variational approach, is used to predict the disparity image and to compute its confidence map for the next frame. The prediction is modeled as a prior probability distribution and is built into an energy function defined for stereo matching on the next frame. The disparity can be estimated by minimizing this energy function. Experimental results show that the algorithm is able to estimate the disparity images in an accurate and temporally consistent fashion.

1 Introduction

3D cinema is making a third comeback in history and it looks like it is here to stay around this time with over a 1000 theatres (just in the USA) already equipped with screens to show movies created in stereo. Set makers and the major Hollywood studios are scrambling to standardize a 3D format in order to bring the same experience to the home. Autostereoscopic displays (no-glasses 3D displays) hold the biggest potential for bringing 3D to the home in the long term. However, a pragmatic format [22] that contains parallax or disparity at its core is crucial for the success of these displays.

The problem of estimating the disparity images for a stereo image sequence has been an active research topic in the vision community for many years. It is a challenging problem for several reasons: first, the difficulties that arise from just a single pair of stereo frames, such as the presence of textureless areas and occlusions, image noise and errors, and different radiometric properties of multiple cameras; and second, the difficulties that arise from the sequence, such as fast object movement, motion blurring, etc. Some existing stereo match algorithms can achieve good results on a single pair of stereo frames [1,4,5,6]. They formulate the problem within an energy minimization framework and the energy is then optimized using one of the popular optimization methods such as graph cuts [7,8] or belief propagation [3,4,5,6].

BMVC 2009 doi:10.5244/C.23.55

However, these methods are not sufficient for depth estimation on a stereo image sequence, without considering the temporal consistency between adjacent frames. Previous research shows that better and more consistent disparity maps can be achieved by incorporating temporal constraints into stereo models [16,17,18,19,20,21].

In this paper, we propose a method to estimate the disparity maps for a stereo image sequence. The approach takes the advantages of an over-segmentation based stereo algorithm and integrates it into a pixel based algorithm, which leads to good performance on a single pair of stereo images. Furthermore, we model the temporal consistency of disparity maps using scene flow [14] (also known as disparity flow), which captures the dense 3D motion in the scene from a given view. The scene flow is built into the stereo models for the next frame as a soft constraint, which helps to resolve stereo ambiguity from the temporal domain, while at the same time reducing the error propagation in disparity maps.

1.2 Related Work

Recent research shows that grouping pixels with similar color into segments can reduce the depth ambiguity within textureless regions and allow for precise delineation of object boundaries corresponding to depth discontinuities [10,11,12]. Zitnick et al. [11] use an over-segmentation scheme to represent a scene as a collection of small fronto-parallel planar segments. This approach is robust to noise and intensity bias by computing match values over entire segments rather than single pixels. Color-based segmentation also helps to more precisely delineate object boundaries and reduce boundary artifacts. The depths of the segments are computed using loopy belief propagation within a Markov Random Field framework.

However, segment based approaches have a common drawback in that image segmentation based on only color information is not consistent with object boundaries and may span depth discontinuities. The input images are segmented in a separate pre-processing step and one cannot recover from any errors caused by the segmentation process. Taguchi et al. [12] further improve on this by jointly estimating image segmentation and depth information. The segment shapes and depths are updated alternatively and iteratively. This technique has difficulties with handling surfaces that are not fronto-parallel since it estimates disparity for segments with the assumption that these segments are fronto-parallel. In our method described in section 2, we use a different technique to account for the segmentation errors.

For stereo over image sequences, some techniques have been proposed to obtain accurate disparity maps by utilizing consistency in the temporal domain [16,17]. Some of them assume either that the scenes are static/quasi-static or that the motion is negligible compared to the sampling frequency [17]. These approaches have difficulty handling dynamic scenes or scenes with constant lighting.

Temporal consistency has been explored for dynamic scenes in [18,19,20,21]. Tao et al.'s approach [19] segments the input images into homogeneous color regions and models each segment as a 3D planar surface patch. The projections of a given planar patch on two adjacent frames are interpreted by a temporal homography. The temporal homography, together with the spatial homography, is then used to estimate the parameters of the planar patch. Since their approach is segmentation-based, the accuracy of the results are limited by the image segmentation algorithm. In Leung et al.'s approach [18], the temporal consistency is enforced by minimizing the difference between the

disparity maps of adjacent frames. This approach may have difficulties in handling scenes that contain large motions by penalizing disparity changes always, which may also be a problem for Larsen et al.'s approach [21]. Gong's method [20] models temporal consistency in disparity space using the concept of scene flow. But the computation is optimized for real-time online stereo using the local area method.

Variational methods have been exploited to compute optical flow in a lot of research work. The best results in terms of accuracy were obtained by Brox et al. [13]. The global energy is only linearized inside the minimization algorithm after warping the image at time *t+1* on to the image at time *t*. [14] further extends optical flow into scene flow estimation. The method computes scene flow by joint estimation of the disparity maps and the motion field from a calibrated stereoscopic image sequence within a unified variational framework. In [15], the depth and 3D motion are decoupled because the nature of motion estimation and disparity estimation are very different and the problems can be solved more efficiently.

This paper adopts the approach in [14,15] to estimate the 3D motion, but for a different purpose. A prior probabilistic model is built from the scene flow estimated and used in stereo estimation at time *t+1*. In this case, the scene flow doesn't enforce the hard constraint on the disparity maps at time *t+1*, but instead uses them as a soft constraint, which has two advantages: first, the errors in previous disparity maps (time *t*) and scene flow can be corrected and hence the error propagation can be reduced; second, the temporal consistency constraint is effectively added to the temporal disparity estimation.

The rest of this paper is organized as follows. In section 2, we describe the algorithms in detail. Section 3 shows the experimental results including the evaluations on Middlebury datasets and a real world image sequence. In section 5, we finally conclude the paper with a summary of our algorithms and some comments on the improvements.

2 Algorithm Descriptions

To simplify the stereo matching and scene flow estimation, we first rectify the two image streams so that the stereo disparity is along the horizontal direction in the images. The outline of the presented algorithm is shown in Figure 1. In this workflow, assuming that the disparity maps (D_t^L, D_t^R) (in left and right views) at time *t* are estimated, we then compute the scene flow between time *t* and *t+1*, and use them to predict the disparity maps at time *t+1* (P_{t+1}^L, P_{t+1}^R). The scene flow is also used to compute the confidence maps that model the strength of *temporal links*. These are further employed in the iterative stereo estimation to generate the disparity maps at time *t+1*.

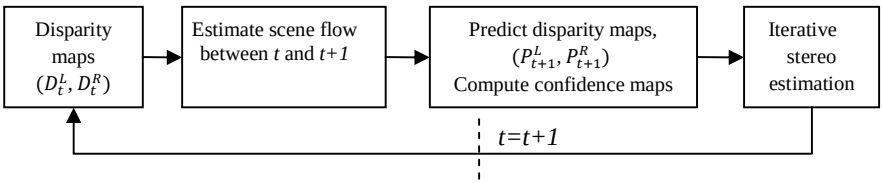

Figure 1. Diagram of the proposed stereo algorithm

We adopt a hybrid approach to generate the initial disparity maps for the first frame, which includes two steps: first, we use the over-segmentation based stereo algorithm similar to [11] to estimate the preliminary disparity maps. Then we add the constraints

based on these disparity maps to the pixel-wise iterative stereo estimation. The reasons are twofold: First, the over-segmentation based approaches prove to be very robust and insensitive to image noise and color bias between the left and right views. It is well known that the difficulties and ambiguities caused by textureless or occluded regions can be handled well by segmenting the images. Second, over-segmentation may not be correct in cluttered scenes since it is only based on color. Any segmentation artifacts around the object border will directly affect the accuracy of the disparity maps. One solution is to update the small segments during the iterative stereo estimation process [12]. In our implementation, we adopt a more straightforward, yet effective way to improve on this problem. Instead of updating the over-segmentation results, we first do stereo estimation at the segment level and then use the estimated disparity maps to guide the estimation at the pixel level. We do this by incorporating the estimated disparity of each segment into the stereo model as a soft constraint. It penalizes the disparity difference between the segment disparity and the pixel disparity within that segment. At the pixel level, the estimation can maintain the robustness of the segment-based approach and at the same time correct the segmentation errors using stereo match measurements. This is different from some stereo methods which employ the plane or segment constraint as a soft constraint [6], where initial disparities are first estimated at pixel level and are then used to fit disparity planes for every segment. If most of the initial disparities are wrong within one segment, the fitting plane is not reliable or useful. Though we incorporate the segment constraint in a similar way, we don't rely on the disparity plane fitting.

2.1 Iterative Stereo Estimation

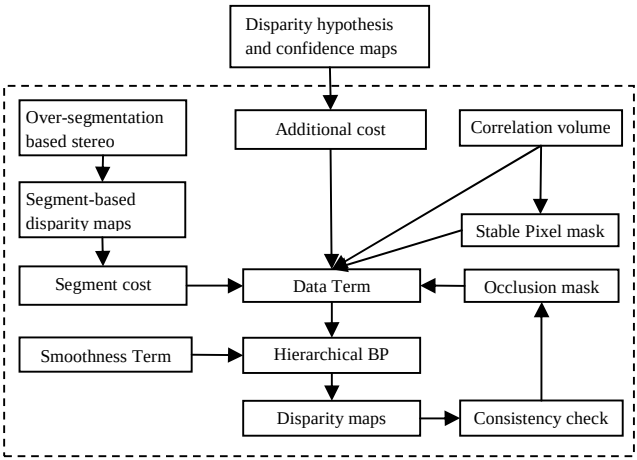

Figure 2. Diagram of the iterative stereo algorithm

The iterative stereo algorithm comprises of several modules as shown in Figure 2. In the over-segmentation based stereo algorithm, we first apply mean shift segmentation [9] to the left and right images and split the large regions into small segments. Then we compute the data cost from the greatest percentage of pixels within a segment that match over a range of possible offset differences [11]. We construct a pairwise Markov Random Fields model and compute the segment disparities using graph cuts [7] with a smoothness cost of a truncated

quadratic model. For the second step, we apply the pixelwise iterative algorithm similar to [5], where the occlusions are explicitly handled in a symmetric stereo match model with the segment disparity maps encoded as a soft constraint.

We first compute the correlation volume [6] using the locally adaptive support weight method of Yoon and Kweon [2]. The stable pixel mask is generated from the match cost cube. The pixel is marked as stable if the best peak in the match cost is distinct from the second best peak. We also estimate the occluded pixel mask from the disparity maps estimated during the last iteration using the left-right consistency check (the occlusion mask is empty initially). Then we merge the correlation volume and segment cost into the final data term, using the adaptive weights based on these two mask images. The additional temporal cost, when available, is also coalesced into the data term with the adaptive weights. The unstable or occluded pixels are given more regularization from the additional constraints (segment cost and temporal cost).

The smoothness cost is defined as a truncated linear model and is tuned based on color edges because the depth edges are likely to coincide with color edges. The smoothness cost will decrease for strong color edges and increase for uniform regions. This will help ensure that the disparity maps are consistent with the object borders.

Hierarchical loopy belief propagation [3] is employed to realize the optimization of the total energy. Hierarchical BP adopts a coarse-to-fine strategy: first performing BP at the coarsest scale and then using the messages from the coarser scale to initialize the input for the next scale. This strategy significantly speeds up the convergence. In our implementation, we use 5 scales and 5 iterations for each scale.

Once the optimization is done, we use the current solutions to update the occlusion masks and the data cost. We then optimize again with the new data cost and this process is stopped once the solutions converge.

2.2 Scene Flow Estimation

Scene flow describes the motion of each 3-D point between two time steps. We use the view-dependent representation of the scene flow similar to the work by Huguet [14]. The optical flow and disparity maps are jointly estimated from the two rectified image sequences. Let $I_l(x,y,t)$ and $I_r(x,y,t)$ be the two stereo images at time t. Assuming that the disparity maps at time t have been estimated, to predict the disparity maps at $t+1$, the scene flow field is defined as (u,v,w), where (u,v) is the optic flow in the left view, and w is the change in disparity. Each element of the scene flow is the scalar function defined in the reference view, i.e., frame reference of $I_l(x,y,t)$. Given the initial disparity map and the scene flow for each pixel (x,y) in the reference view, we can find its correspondences in all the other three images.

The global energy functional is defined as a sum of data term which makes use of the color constancy assumption between the correspondences and the smoothness term, which enforces flow fields to be smooth.

$$E(u,v,w) = E_{Data} + \alpha E_{Smooth} \tag{1}$$

where α is regularization parameter; the data cost consists of three data terms from the left and right optic flows and stereo at time $t+1$:

$$E_{Data} = E_{fl} + E_{fr} + E_{st} \tag{2}$$

where each data term is defined over the image domain Ω as follows:

$$E_{fl} = \int_{\Omega} m_{fl}(x,y) \sum_{c=1}^{3} \Psi((I_l^c(x+u, y+v, t+1) - I_l^c(x,y,t))^2) dxdy \quad (3)$$

$$E_{fr} = \int_{\Omega} m_{fr}(x,y) \sum_{c=1}^{3} \Psi((I_r^c(x+d+u+w, y+v, t+1) - I_r^c(x+d,y,t))^2) dxdy \quad (4)$$

$$E_{st} = \int_{\Omega} m_{st}(x,y) \sum_{c=1}^{3} \Psi((I_r^c(x+d+u+w, y+v, t+1) - I_l^c(x+u,y+v,t+1))^2) dxdy \quad (5)$$

where m_{fl}, m_{fr} and m_{st} are the mask images of non-occluded pixels for left optic flow, right optic flow and the stereo match at time $t+1$. They are computed from image warping techniques similar to the method described in [13]. Three color channels R, G and B are used in the data term and c denotes one of 3 color channels. Ψ is a robust function: $\Psi(x^2) = \sqrt{x^2 + \varepsilon^2}$, where ε=0.001 to make the robust function differentiable. The robust function helps to reduce the influence of the outliers on the solutions.

The smoothness term is defined to penalize the local variations in the flow fields.

$$E_{Smooth} = \int_{\Omega} \Psi(\| \nabla u \|^2 + \| \nabla v \|^2 + \lambda \| \nabla w \|^2 + \gamma \| \nabla d \|^2) dxdy \quad (6)$$

where the same robust function Ψ is applied to the sum of the gradient norms to help preserve the discontinuities of the flows since the discontinuities likely appear simultaneously in the scene flow fields. Parameter λ adjusts the relative weight between disparity flow and optic flow, and γ scales initial disparity versus optic flow. Though the disparity field $\boldsymbol{d}$ is known, we still add its gradient norm to help keep the boundary of the scene flow image sharp.

According to calculus of variations, an extreme of the energy functional E can be achieved by solving its Euler-Lagrange equations:

$$\sum_{c=1}^{3} [m_{fl} \Psi'((I_{lz}^c)^2) I_{lz}^c I_{lx}^c + m_{fr} \Psi'((I_{rz}^c)^2) I_{rz}^c I_{rx}^c + m_{st} \Psi'((I_{dz}^c)^2) I_{dz}^c (I_{rx}^c - I_{lx}^c)] - \alpha div(\Psi_S' \nabla u) = 0 \quad (7)$$

$$\sum_{c=1}^{3} [m_{fl} \Psi'((I_{lz}^c)^2) I_{lz}^c I_{ly}^c + m_{fr} \Psi'((I_{rz}^c)^2) I_{rz}^c I_{ry}^c + m_{st} \Psi'((I_{dz}^c)^2) I_{dz}^c (I_{ry}^c - I_{ly}^c)] - \alpha div(\Psi_S' \nabla v) = 0 \quad (8)$$

$$\sum_{c=1}^{3} [m_{fr} \Psi'((I_{rz}^c)^2) I_{rz}^c I_{rx}^c + m_{st} \Psi'((I_{dz}^c)^2) I_{dz}^c I_{rx}^c] - \alpha\lambda div(\Psi_S' \nabla w) = 0 \quad (9)$$

where the $\Psi'(x^2)$ denotes the derivative of Ψ with respect to x^2, and I_{lx}^c and I_{ly}^c mean the gradients of the warped left image $t+1$. The terms I_{rx}^c and I_{ry}^c are defined similarly. And

$$I_{lz}^c = I_l^c(x+u, y+v, t+1) - I_l^c(x,y,t); \quad (10)$$

$$I_{rz}^c = I_r^c(x+d+u+w, y+v, t+1) - I_r^c(x+d,y,t); \quad (11)$$

$$I_{dz}^c = I_r^c(x+d+u+w, y+v, t+1) - I_l^c(x+u,y+v,t+1); \quad (12)$$

In our implementation, by assuming that $I_l^c(x,y,t) = I_r^c(x+d,y,t)$, we get that $I_{lx}^c = I_{rx}^c$ and $I_{ly}^c = I_{ry}^c$, and we can then simplify equations (7) and (8) by removing the linearized third constraint in these equations.

To solve these non-linear equations, we adopt the strategy of two nested fixed point iteration loops as in [13]. In the outer iteration loop, a first order Taylor expansion is applied to the Euler Lagrange equations, specifically to the expressions I_{lz}^c, I_{rz}^c and I_{dz}^c. In each iteration, the second image is warped according to the current estimated flow and an increment of the flow vectors is estimated. In the inner iteration loop, the nonlinear terms Ψ' are further linearized and the resulting sparse linear system of equations can now be solved using common numerical methods, such as SOR iterations. The solution will then be used to update Ψ'.

These fixed point iterations are combined with a coarse-to-fine strategy to better approximate the global optimum of the energy. The stereo image pyramids are constructed with a down sampling factor, and when the fixed point iterations are conducted at a given pyramid level, the solution is scaled and upsampled to the next finer level. This process is repeated until the full resolution is reached. In our implementation, we use 0.8 for the down sampling factor, and as in [14], 0.05 is used as the stopping condition for the inner fixed point iterations and 0.01 for the outer fixed point iterations.

2.3 Disparity maps prediction

From the scene flow definitions, the predicted disparity maps can be computed as:

$$P^L_{t+1}(x+u, y+v) = D^L_t(x,y) + w(x,y) \quad (13)$$

$$P^R_{t+1}(x+u+d+w, y+v) = -D^L_t(x,y) - w(x,y) \quad (14)$$

where the superscript L(R) denotes left(right) view. Let $D^{L\prime}_t(x,y) = D^L_t(x,y) + w(x,y)$; we warp the disparity map $D^{L\prime}_t$ to the left and right views at time *t+1* using Z buffering techniques. In other words, if two or more pixels in the current frame are warped to the same pixel in the next frame, the largest disparity will be used as it represents the most frontal 3D point and therefore should be visible.

The predicted disparity maps may have some mismatches. For instance, the errors in disparity maps at time *t* may be propagated and the scene flow may be over-smoothed at object boundaries. To prevent error propagation, we compute the confidence image together with the predicted disparity maps. The confidence of a pixel is measured by the color similarity between that pixel and its correspondences according to the scene flow vector. Similar to the data term defined in scene flow estimation, we use I^c_{lz}, I^c_{rz} and I^c_{dz} as the color similarity measures. The color differences are summed up and converted to the weight image using an exponential function:

$$\omega = m_{fl} m_{fr} m_{st} exp\left(-\frac{||I_{lz}||^2 + ||I_{rz}||^2 + ||I_{dz}||^2}{\sigma_c^2}\right) \quad (15)$$

where ω denotes the confidence of the correspondences specified by the scene flow field, and σ_c is the normalization parameter for the color difference (set to 6.0 in the experiments). In scene flow, the color differences in the 3 color channels are treated separately, but here we use the Euclidean color distance. $\|I_{lz}\|$ is the L2 norm of I_{lz}, a 3-dimensional vector corresponding to 3 color channels. The visibility masks used in scene flow estimation also appear here since we only want to propagate the disparity which can be verified in all the four frames involved. The confidence image is then warped to the left and right frames at time *t+1* with the estimated scene flow vectors.

The prediction and confidence is then incorporated into the stereo models for time *t+1* as a soft constraint. The temporal cost is defined to penalize the differences between the predicted disparity and the disparity to be estimated based on the confidence:

$$E_{temporal}(D) = \sum_i \omega_i \min\left(|d_i - P_i|, T_{temporal}\right) \quad (16)$$

where for each pixel *i*, d_i is the disparity to be estimated. P_i and ω_i are the predicted disparity and confidence respectively. $T_{temporal}$ is a truncation threshold for temporal cost (set to 10 in our experiments). Basically the confidence controls the contribution of the temporal consistency term into the overall energy for stereo estimation. If the confidence is too small, the prediction has little influence on the disparity. The temporal cost is fed into the iterative stereo estimation module as an additional cost and merged into the final

data term as explained in Section 2.1.

3. Experimental Results

We first test the hybrid stereo algorithm on the Middlebury stereo datasets [1] and the results rank 8th overall in the evaluation as of April 2009. The overall performance is much improved compared with the results in [11], and very close to the results in [12] where the over-segmentation and segment disparity are updated iteratively.

Datasets	Non-occluded	All	Discontinuities
Tsukuba	1.01 $_{8}$	1.34 $_{4}$	5.46 $_{8}$
Venus	0.28 $_{17}$	0.58 $_{16}$	3.62 $_{24}$
Teddy	6.67 $_{15}$	12.1 $_{17}$	16.8 $_{18}$
Cones	2.87 $_{5}$	9.00 $_{16}$	7.44 $_{3}$

Table 1. Evaluation of the hybrid stereo algorithm on the Middlebury datasets. The numbers are the percentage of pixels whose absolute disparity error is greater than 1. The subscript of each number is the rank of that score.

To evaluate the scene flow algorithm, we use these datasets and the above initial disparity map. Each dataset consists of 8 views of the same static scene. The images are captured from equally-spaced viewpoints along the x-axis from left to right and are also rectified. Similar to [14], we took images 2 and 6 of the Venus, Teddy and Cones datasets as the stereo pair at time t, and images 4 and 8 as the stereo pair at time $t+1$. In this special configuration, the optic flow part is strictly horizontal (v=0), and the disparity maps are the same (w=0), but our algorithm doesn't know anything about these. Ground truth is given as the disparity from 2 to 6, and the optic flow is half the disparity.

In the evaluation, we set the smoothness parameters α=60, λ=2, γ=0.5. We calculate the root mean square (RMS) and the absolute angular error (AAE) on the optic flow and disparity maps without occluded areas. They are all measured in pixels. Although the initial disparity map is estimated outside of the scene flow algorithm, we still include it in the comparison. The results are summarized in Table 2. Compared with Huguet's method [14], we achieve much lower error rates for the initial disparity map and this helps to improve the accuracy of the scene flow fields.

Dataset	RMS d (Initial disparity)		RMS w		RMS (u,v)		AAE (u,v) (mean, standard deviation)	
Venus	**0.38**	0.97	**0.22**	1.48	**0.32**	0.31	**(1.38, 0.76)**	(0.98, 0.91)
Teddy	**1.17**	2.27	**0.49**	6.93	**1.01**	1.25	**(0.37, 0.48)**	(0.51, 0.66)
Cones	**1.13**	2.11	**0.43**	5.24	**0.99**	1.11	**(0.56, 0.74)**	(0.69, 0.77)

Table 2. RMS and AAE errors in pixels on the Middlebury datasets. We show the errors (the number in bold) from the initial disparity map, the scene flow maps obtained with our approach, and the results (the other number in each item) from Huguet's result[14].

We further use the disparity maps and scene flow maps estimated above to predict the disparity maps at time $t+1$ and to compute the confidence images. The predictions and confidence images are translated into the temporal cost and fed into the stereo models at time $t+1$. In Figure 3, we show the intermediate results on the Cones dataset. The predicted disparity maps have some artifacts, for instance, the bent thin structure and the blurring edges. These artifacts can be fixed by our approach.

The algorithm is also tested on a real world stereo image sequence. The test sequence we show here is challenging since first, the background curtain behind the girl contains rich and repeated textures which cause problems in over-segmentation and stereo estimation, and second, there are obvious color variations between left and right views, which makes it difficult to match the walls behind the lamp. As shown in Figure 5, we compare the disparity maps obtained by applying our hybrid approach to each stereo pair independently, and the disparity maps from applying the same approach with the temporal constraint. In this example, the temporal constraint helps to remove the artifacts in the background walls (uniform region) and curtains (repeated textures).

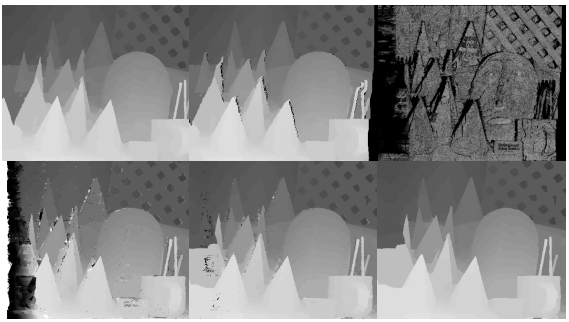

Figure 3. Top row: from left to right are the initial disparity map, the predicted disparity map, and the confidence image; Bottom row: winner-take-all disparity map from LASW match cost, winner-take-all disparity map from the data cost integrating LASW match cost and temporal cost, and the disparity maps at time *t+1*.

Figure 4. Top Row: 5 consecutive Frames (only left view); Middle Row: the disparity maps estimated frame-independently using the hybrid stereo approach; Bottom Row: the corresponding disparity maps estimated with temporal constraints.

4. Conclusions

In this paper, we proposed a method to estimate the disparity maps for a stereo image sequence. It follows the principle of a hierarchical approach by unifying the over-segmentation based stereo model and a pixel based stereo model, which can achieve a good performance on a single pair of stereo images. Furthermore, we use the scene flow, which captures the dense 3D motion in the scene from a given viewpoint, to model the

temporal consistency of disparity maps. The scene flow is built into the stereo estimation for the next frame as a soft constraint, which help to resolve stereo ambiguity from the temporal domain, while at the same time reducing the error propagation in disparity maps. Experiments on real world data show that our temporal modeling can indeed improve the temporal consistency in the disparity maps.

As for the limitations, we observed that the error propagation in disparity maps cannot be avoided always. For instance, some errors in disparity maps could survive in the prediction and confidence evaluation steps, and then could be propagated to next frames, such as the errors in uniform regions and repeated textures. Generally, our approach can only achieve the suboptimal solution compared to the bigger optimization problem of estimating disparity maps over a whole sequence (where the global optimal solution is almost impossible for a long sequence) and the initial disparity map has some influence on the later disparity maps. In future work, we expect to look in to a more extensive MRF framework that includes temporal and stereo constraints to approximate the optimal solution better.

References

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

[2] K.-J. Yoon and I.-S. Kweon, Locally Adaptive Support-Weight Approach for Visual Correspondence Search, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. II:924-931, 2005.

[3] P. F. Felzenszwalb and D. P. Huttenlocher, Efficient Belief Propagation for Early Vision, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. I:261-268, 2004.

[4] J. Sun, N.-N. Zheng and H.-Y. Shum, Stereo Matching Using Belief Propagation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 7, July 2003.

[5] J. Sun, Y. Li, S. B. Kang and H.-Y. Shum, Symmetric Stereo Matching for Occlusion Handling, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. II:399-406, 2005.

[6] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister. Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. In CVPR, volume 2, pages 2347–2354, 2006.

[7] Y. Boykov, O. Veksler and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 11, 2001.

[8] V. Kolmogorov and R. Zabih, Computing Visual Correspondence with Occlusions using Graph Cuts, IEEE International Conference on Computer Vision, Vol. I:508-515 2001.

[9] D. Comaniciu and P. Meer, Mean shift: A Robust Approach Toward Feature Space Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 4, May 2002.

[10] H. Tao and H. Sawhney, Global Matching Criterion and Color Segmentation Based Stereo, IEEE Workshop on Applications of Computer Vision, pp. 246-253, 2000.

[11] C. L. Zitnick and S. B. Kang, Stereo for image-based rendering using image over-segmentation. IJCV, 75(1):49–65, 2007.

[12] Y. Taguchi, B. Wilburn, and L. Zitnick. Stereo reconstruction with mixed pixels using adaptive over-segmentation. CVPR 2008.

[13] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In European Conference on Computer Vision (ECCV), pages 25–36, 2004.

[14] F. Huguet and F. Devernay. A variational method for scene flow estimation from stereo sequences. In IEEE Eleventh International Conference on Computer Vision, ICCV 07, Rio de Janeiro, Brazil, October 2007.

[15] Wedel, A., Rabe, C., Vaudrey, T., Brox, T., Franke, U., and Cremers, D. 2008. Efficient Dense Scene Flow from Sparse or Dense Stereo Data. In Proceedings of the 10th European Conference on Computer Vision: Part I (Marseille, France, October 12 - 18, 2008), pages 739-751.

[16] D. Min and K. Sohn. Edge-preserving simultaneous joint motion-disparity estimation. In ICPR ’06: Proc. 18th International Conference on Pattern Recognition, pages 74–77, Washington, DC, USA, 2006.

[17] Davis, J., Nehab, D., Ramamoothi, R., and Rusinkiewicz, S., Spacetime stereo: a unifying framework for depth from triangulation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 27 (2005)

[18] Leung, C., Appleton, B., Lovell, B. C., and Sun, C.: An energy minimisation approach to stereo-temporal dense reconstruction. Proc. International Conference on Pattern Recognition. Cambridge, UK. (2004) 72-75.

[19] Tao, H., Sawhney, H. S., and Kumar, R.: Dynamic depth recovery from multiple synchronized video streams. Proc. IEEE Conference on Computer Vision and Pattern Recognition. Kauai, Hawaii, USA. (2001)

[20] Gong, M., Enforcing Temporal Consistency in Real-Time Stereo Estimation. In European Conference on Computer Vision (ECCV) 2006, pages 564-577.

[21] S. Larsen, P. Mordohai, M. Pollefeys, H. Fuchs, Temporally Consistent Reconstruction from Multiple Video Streams using Enhanced Belief Propagation, Proc. ICCV'07. Rio de Janeiro, Brazil, October 2007.

[22] http://en.wikipedia.org/wiki/2D-plus-depth