

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Temporally Consistent Depth Maps Recovery from Stereo Videos

^{1,2}Li Yao, ^{1,2}Dong-Xiao Li and ^{1,2}Ming Zhang

¹Institute of Information and Communication Engineering, Zhejiang University, Hangzhou 310027, China

²Zhejiang Provincial Key Laboratory of Information Network Technology, Hangzhou 310027, China

Abstract: Dense depth maps provide significant geometry information for 3D video and free viewpoint video systems. Traditional stereo correspondence methods usually deal with each stereo image pair separately. As a result, the generated depth sequence is temporally inconsistent. This paper presented a novel approach to recover spatio-temporally consistent depth maps. The proposed method first applied sequential belief propagation algorithm to achieve an approximate minimum of spatial energy on Markov random fields. Then in temporal domain, a smoothness cost along optical flow was incorporated between consecutive frames. The combined cost which determined the disparity value was passed forward and temporal consistency was enforced during the process. In addition, the streamlined implementation was time and memory efficient. In experimental validation, quantitative evaluation as well as subject assessment was performed on several test datasets. The results showed that the proposed method yielded temporally consistent depth sequence and reduced flickering artifacts in the synthesized view while maintaining visual quality.

Key words: Depth map estimation, multi-view stereo, Markov random fields, belief propagation, optical flow, temporal consistency

INTRODUCTION

The concept of three-dimensional display has attracted human being for several decades. Seeking for a real impression of natural world, people have made many attempts to exploit stereopsis in 3D display (Konrad and Halle, 2007). Glass-based stereoscopic displays use filters or shuttles to provide virtual 3D visual experience (Urey *et al.*, 2011). Obviously, it is inconvenient for the viewers to wear glasses or other special device. Autostereoscopic techniques which directly present stereoscopic images seem more popular (Dodgson, 2005). In autostereoscopic displays, more than two views are required so that the viewers can observe different corresponding scenes in different positions (Zwicker *et al.*, 2007). However, the vast raw data of multi-view video is considered as a severe conflict with the transmission bandwidth (Meesters *et al.*, 2004).

Recently, Depth Image-Based Rendering (DIBR) has been considered as one of the most significant technologies for 3DTV (three-dimensional television). The 3D content is typically represented by regular 2D video and associated gray-scale depth map (Fehn, 2004). On the other hand, the topic of stereo correspondence is a fundamental issue in computer vision (Brown *et al.*, 2003; Zigh and Belbachir, 2010; Yu *et al.*, 2011; Shuchun *et al.*, 2011). A large number of methods have been proposed to solve this ill-posed problem which suffers from image

noise, border mismatch, textureless regions and occlusions. Scharstein and Szeliski (2002) showed a survey of taxonomy and categorization for dense stereo correspondence algorithms. The existing techniques for dense stereo correspondence roughly fall into two categories: local methods and global methods. Local methods determine the disparity value of a concerned pixel depending on a local surrounding area, e.g., block based method (Bae *et al.*, 2008) and variable windows approach (Veksler, 2003). They have very efficient implementations but lead to mismatch in boundary regions. On the contrary, global methods make explicit smoothness assumption and solve the optimization problem in a global framework. The main distinction among these methods is the optimization procedure used, such as simulated annealing (Barbu and Zhu, 2005), graph cuts (Boykov *et al.*, 2001), belief propagation (Yang *et al.*, 2006) and so on. In general, these methods deal with each single image pair and disregard of the correlation between consecutive frames in a video sequence (Scharstein and Szeliski, 2002). They do not explicitly distinguish moving objects with static background. As a result, the depth value of static objects and background may fluctuate in a different time. Furthermore, it causes critical artifacts in the synthesized virtual view and discomforts the viewers because human vision is sensitive to the frequent flickering (Lee and Ho, 2010).

Motivated by the demand of enhancing temporal consistency, the stereo correspondence procedure incorporates several improvements in current research. Bleyer and Gelautz (2009) applied smoothing operation on disparity map sequence to decrease flickering artifacts. Smirnov *et al.* (2010) proposed a filtering method to achieve high-quality and temporally consistent depth maps. Commonly these methods enforce temporal coherence with filters, so the results often get blurred in object boundary. Tao *et al.* (2001) addressed the problem of extracting depth information of non-rigid dynamic 3D scenes from multiple synchronized video streams and the temporal consistency is improved during the process. Zhang *et al.* (2009) proposed a bundle optimization framework to incorporate geometric coherence constraint of multiple frames in a video. Also a general form of scene flow can be utilized to estimate depth maps (Vedula *et al.*, 2005) but it increases computation cost dramatically and fails to realize in real-time application.

In contrast with 3D MRF model which treats all the frames simultaneously (Zhaozheng and Collins, 2007), this framework adopts streamlined implementation. First in spatial domain, Belief Propagation (BP) algorithm is applied for each frame to find a minimization of MRF energy. Then in temporal domain, a recursive function combines all frames and computes total energy in a linear structure. After BP based matching of each frame accomplished, the temporal algorithm allows the computational belief messages to be released. Only an aggregated cost is transferred forward for the next frame. Moreover, temporal consistency is refined during the process. Based on the streamlined framework, a novel method was proposed to recover consistent depth maps from stereo video sequences in this paper. Temporal smoothness function was employed along motion path decided by optical flow. The framework of traditional stereo correspondence based on Markov Random Fields (MRF) was extended to whole video sequence.

MATERIALS AND METHODS

Overview of framework: To facilitate the work of stereo correspondence, several necessary preprocess are required. First, epipolar geometry constraint is imposed. Unlike motion estimation in video compression which only cares about the data redundancy, disparity vector in stereo correspondence relates to a pair of pixels which are exactly from a same position in 3D scene. In order to reduce the number of potential correspondences and increase matching reliability, original stereo video is calibrated in an epipolar line (Hartley and Zisserman, 2004). In epipolar geometry, the search range of disparity is limited in a horizontal scan-line and the disparity value

can be easily transformed to a depth value when camera parameters are known. Second, image quality of stereo video need to be corrected. The intrinsic property of image such as color, or intensity, will influence the measure of similarity and further impact the computation of matching cost.

In general, the methods are formulated in an energy-minimization framework based on Markov Random Fields (Geman and Geman, 1984). MRF model provides a convenient and powerful foundation for many intractable problems in early vision involved with gridded image-like data (Bai *et al.*, 2008; Izabatene and Rabahi, 2010). A regular algorithm is designed by terms of Maximization A Posteriori (MAP) estimation based on Bayesian network. A typical MRF model of N_4 neighborhood system is shown in Fig. 1. The white circles $f(i, j)$ denote unknowns to be inferred while the dark circles $f^*(i, j)$ denote input data. The black boxes $d(i, j)$ denote elemental data penalty terms and $s(i, j)$ denote interact potentials between connected nodes in the random fields. The data term together with the smoothness term make up the total energy function. In stereo correspondence problem, the goal is to find corresponding points between two rectified images. The label of each pixel indicates discrete variable of disparity value. The data term measures how well the pixel in source image matches the one in reference image. While the smoothness term imposed by MRF model indicates that pixels in neighboring areas should have similar disparity value (Boykov *et al.*, 2001). Altogether, the energy function of stereo correspondence problem can be defined as:

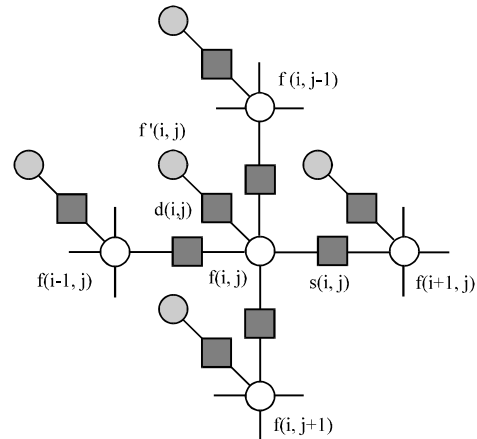


Fig. 1: Graphic model for an N_4 neighborhood Markov random field. Note: The white circles denote the unknowns $f(i, j)$ while the dark circles denote the input data $f^*(i, j)$. The black boxes $d(i, j)$ denote data penalty and $S(i, j)$ denote interaction potentials between adjacent nodes

$$E = E_d + \lambda E_s \quad (1)$$

where, E_d is the data energy and E_s is the smoothness energy. λ gives the relative weight of the smoothness penalty.

The overall flow chart of the proposed method can be depicted as follows: After input stereo video stream is rectified (Du *et al.*, 2004), raw matching cost is computed first. A robust measure of truncated AD (absolute difference) of intensity is used which is specified in Eq. 3, where, p and p' are the corresponding pixels with the label of disparity l_p and S is the set of pixels in the image. A Disparity Space Image (DSI) is created for storing raw costs within all possible disparity. Then, BP based stereo correspondence algorithm is applied for each single frame. If current frame is the first of the video sequence, the disparity map can be immediately obtained. Otherwise, temporal smoothing is enforced by calculating an aggregated cost along motion path. The temporal accumulated energy and the spatial energy together form the total energy. Then a Winner-Take-All (WTA) strategy is used to determine the output depth maps. The raw cost and temporary memory storage will be released before the next round starts.

To sum up, the modified energy function from Eq. 1 of each frame can be defined as:

$$E = E_d + \lambda E_s + E_t \quad (2)$$

$$E_d = \sum_{p \in S} D_p(l_p) = \sum_{p \in S} \min(|I(p) - I(p')|, \text{Trunc}) \quad (3)$$

$$E_s = \sum_{(p,q) \in N} V_s(l_p, l_q) \quad (4)$$

where, (p, q) denotes a pair of neighboring pixels. And E_t is the temporal aggregation energy which will be specified in the following section.

The smoothness energy term still need to be detailed. A simple case of Potts model (Boykov *et al.*, 2001) assumes that labeling should be piece-wise constant. This model considers only two conditions. For equal labels the cost is zero and for different labels the cost is set to a constant. It is unsuitable for some complicated situations. Here a general form of multi-pass jump smoothness costs function is applied (Felzenszwalb and Huttenlocher, 2004), in which different pairs of adjacent labels can lead to different costs:

$$V_s(l_p, l_q) = \min(\rho_s |l_p - l_q|, \tau_s) \quad (5)$$

where, ρ_s controls the rate of increase in the cost and τ_s is the truncation value.

Spatial optimization: There are two basic assumptions to produce dense disparity map: uniqueness and continuity. That is, the disparity maps have a unique value per pixel and are continuous almost everywhere. In addition, the MRF model adopts spatial and contextual constraints which are ultimately necessary and significant in low-level vision.

Belief propagation which is employed in this method, can be used as an effective way of solving inference problems in pair-wise MRF models. The main advantage of BP is to solve MAP problem of labeling on MRF model and reduce the computation time from exponential level to linear level. The original standard BP proposed by Pearl (1988) takes the idea of passing local messages around the nodes through edges and guarantees the convergence for any tree-structured graphic model. Recently, it is extended to loopy belief propagation so that it can deal with graphs with loops (Ihler *et al.*, 2005).

The loopy BP algorithm works by passing message to neighboring nodes along the four-connected image grid. Each message is a vector of dimension with the number of possible labels. It is initialized to zero in form of negative log probabilities. Each node uses the message received from neighboring nodes to compute new message for other neighboring nodes. Let $m_{pq}(l_q)$ be the message that node p sends to a neighboring node q on label l_q (Fig. 2a). It is updated in the following way:

$$m_{pq}(l_q) = \min_b (V_s(l_p, l_q) + D_p(l_p) + \sum_{r \in N(p), r \neq q} m_{pr}(l_r)) \quad (6)$$

where, the data cost $D_p(l_p)$ and the smoothness cost $V_s(l_p, l_q)$ are defined in Eq. 3 and 5, respectively.

$$\sum_{r \in N(p), r \neq q} m_{pr}(l_r)$$

denotes the message calculated from neighboring nodes of p except node q .

After several iterations of message passing in all directions, the final belief of node q is calculated as:

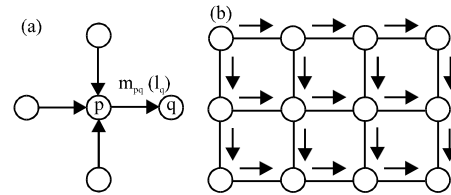


Fig. 2(a-b): Graphic illustration of the proposed BP algorithm. (a) Message computed from node p to q and (b) Messages propagate in forward scan-line direction

$$b_q(I_q) = D_q(I_q) + \sum_{p \in N(q)} m_{pq}(I_q) \quad (7)$$

An involved issue during BP message passing is how to arrange the updating schedule (Tappen and Freeman, 2003). The message updating schedule determines when a node uses the messages received from neighbors to compute the new message. In parallel implementation, the messages are passed along rows and then along columns and used to compute the next round of messages. It starts at the first node and passes messages in one direction until reach the end. Then the messages are passed backward in a similar way.

However, the convergence time of parallel schedule is always quite long. An alternative schedule is to propagate messages in one direction and sequentially update each node. It means when a node sends message to its neighboring node, the neighboring node would use it to compute message for next node immediately. Our sequential implementation is inspired from TWR-S algorithm (Kolmogorov, 2006). It processes nodes in scan-line order, with forward and backward passing (Fig. 2b). Such asynchronous updating scheme allows the message to propagate much more quickly across the image. Thus it is preferred in this framework.

Temporal optimization: Video processing and analysis takes into account not only the pixel values in a single static frame but also the temporal relations between frames. Like the intra-frame continuity, temporal smoothness constraint is exploited for an analytic framework. A direct way is to connect regular location of continuous frames in a 6-neighborhood 3D grid model. But the basic assumption of continuity in temporal domain is violated when the object is moving. Therefore, motion information is exploited in this method. Motion source refers to the temporal variations in image sequences (Stiller and Konrad, 1999; Xu and Zi-Shu, 2011). The 3D motion of object induces 2D motion on the image plane. In order to make the problem tractable, it is assumed that the motion of object is finite. And the disparity value varies continuously along the motion path between consecutive frames.

Motion estimation has found various applications in motion-compensated video compression as well as video summarization and video stabilization, for the reason that the temporal relation of intensity is high and useful (Korah and Perinbam, 2006; Ren *et al.*, 2010; Iffa *et al.*, 2011). The temporal constraint is now applied in the stereo correspondence problem for a video sequence. To make explicit estimate of motion at each independent pixel, a general form of optical flow is exploited. Estimating the optical flow requires inferring a dense field of

displacement vectors which map all points in the first image to the corresponding locations in the second image. The basic technique for computation of optical flow is based on the optical flow constraint equation:

$$I_x \cdot u + I_y \cdot v + I_t = 0 \quad (8)$$

It relates the spatio-temporal intensity changes to the velocity (u, v). However, the solution of (8) can not be determined because there are two unknown variables in one linear equation. A smoothness constraint was introduced by Horn to form a global function (Horn and Schunck, 1981):

$$E_{opt} = \iint (\lambda(I_x \cdot u + I_y \cdot v + I_t) + (u_x^2 + u_y^2 + v_x^2 + v_y^2)) dx dy \quad (9)$$

The optical flow is computed by the non-linear diffusion algorithm (Proesmans *et al.*, 1994). Then each pixel of current frame derives a motion vector pointed to the previous one. The computational flow is illustrated in Fig. 3a. The final cost is decided by both the cost of current frame and the consistency cost between consecutive frames. It is calculated as follows for node q in frame n :

$$C_q^n(I_q) = \begin{cases} b_q^n(I_q) & (n=1) \\ b_q^n(I_q) + \lambda_t (\min(V_t(I_q, I_q) + C_{nor, q}^{n-1}(I_q))) & (n>1) \end{cases} \quad (10)$$

where,

$$V_t(I_q, I_q) = \min(\rho_t |I_q - I_q|, \tau_t) \quad (11)$$

$C_q^n(I_q)$ normally consists of three components. $V_t(I_q, I_q)$ represents the temporal smoothness cost function which has a similar definition to Eq. 5. $b_q^n(I_q)$ denotes the spatial belief cost and $C_{nor, q}^{n-1}(I_q)$ denotes the normalized aggregated cost of frame $n-1$ at node q' which points to the node q along the motion path. λ_t is the temporal weighting coefficient. Basically, this function can be also interpreted as a pair-wise dynamic programming procedure (Fig. 3b). The optimal path between consecutive frames with minimum cost is computed to generate the final cost. The label l which minimizes the final cost is selected as the output disparity value at that pixel. The aggregated process is unidirectional so that any latter frame is not used for reference. Note that the enforcement of temporal consistency may deteriorate the results when optical flow estimation fails. The temporal weighting coefficient λ_t is defined in Eq. 12. When the brightness constancy is violated, λ_t will be decreased adaptively to reduce the effect by such errors:

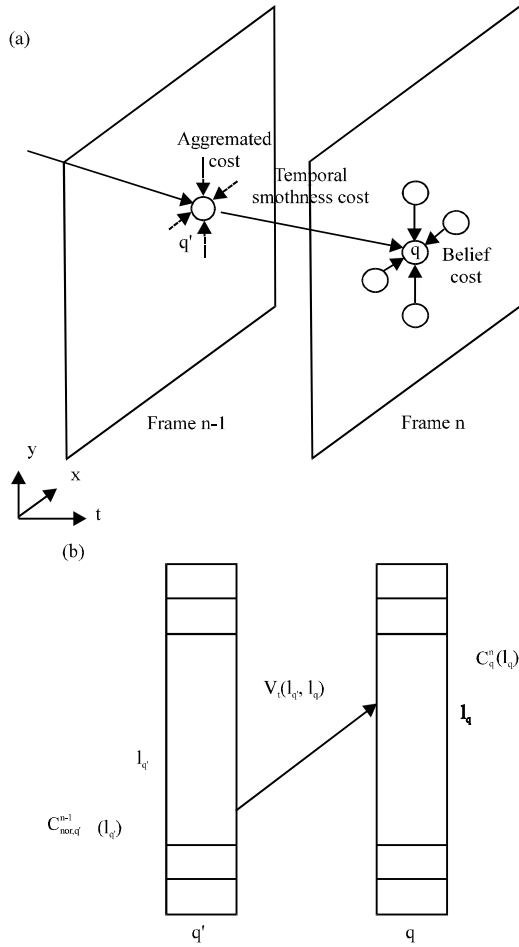


Fig. 3 (a-b): The final cost computation in spatio-temporal domain. (a) Computational flow; (b) Equivalent pair-wise dynamic programming process for solving Eq. (10). Note: In (b), for each label l_q of pixel q , the optimal path is determined by $l_{q'a}$ which minimizes $V_t(l_{q'}, l_q) + C_{nor,q'}^{n-1}(l_{q'})$. Then the term $C_q^n(l_q)$ can be updated recursively

$$\lambda_t = ke^{\frac{|l_a - l_q|}{\gamma}} \quad (12)$$

where, k and γ are the parameters determined empirically.

In the end, the cost was normalized as follows:

$$\sum_{i=0}^{l-1} C_{nor,q}^n(l_i) = 0 \quad (13)$$

EXPERIMENTAL RESULTS AND DISCUSSION

To demonstrate the effective performance of the proposed method, experiments are conducted on PC

(personal computer) platform with Intel Core2 Duo 3.16 GHz CPU (central processing unit). Several stereo video test sequences from FhG-HHI (2011) (Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute) 3DV (three-dimensional video) data base (FhG-HHI, 2011) are used, including 'Book Arrival', 'Alt Moabit', 'Door Flowers' and 'Leaving Laptop'. The resolution of some sequences is re-sampled to reduce the candidates of possible disparity values and enhance the reliability. In traditional stereo correspondence problem, a quantitative way of measuring the quality of the computed result is concerned. Two general approaches are mentioned (Scharstein and Szeliski, 2002). First, compare the resultant disparity map with ground truth data. Second, evaluate the synthetic image rendered by the original source image and the computed disparity map.

Additional measures are required to evaluate the quality in temporal domain. In fact, most of the ground-truth maps for the test sequences are unavailable. Meanwhile, the ground truth data should not be the unique criterion when considering the whole 3DTV system. The disparity map can be recognized as an intermediate product which will be discarded after virtual view synthesized. Not only the intrinsic quality of single depth map but also the consistency of intra and inter frames is critical for the DIBR process. A smooth depth map can improve the synthesized view with less hole regions. To sum up, an overall assessment of the experimental results is taken in the following two aspects.

Subject evaluation: At first, a comparative result of 'Book Arrival' sequence is shown in Fig. 4. Figure 4a shows the source images from five consecutive frames which contain apparent motion objects. The result of the proposed method (Fig. 4d) is compared with other existing methods in the related literature. Fig. 4b is obtained from the Hierarchical BP algorithm (Yang *et al.*, 2006) which has no temporal improvement. Another local method based on adaptive windows (Veksler, 2003) is temporally improved by smoothing filter (Smirnov *et al.*, 2010) and the result is shown in Fig. 4c.

Some consequence can be inferred during the comparison. The regions close to left border of the depth map can be ignored since they are invisible in the reference view. The observed results show that the proposed method has a greater stability than other BP method without temporal improvement. Results in Fig. 4d also perform a conspicuous insusceptibility in the static scenes where the disparity value should not vary. With respect to the temporal method using filter, the proposed method is also outperformed. No blur can be detected and motion objects still maintain their outline.



Fig. 4 (a-d): Disparity map results of different methods of frame 27 to 31 (from top to bottom) on the ‘book arrival’ sequence. (a) Source sequence; (b) Hierarchical BP method without temporal improvement; (c) adaptive window-based method with temporal smoothing filter and (d) Proposed method

Since, it is difficult to present the visual results of whole sequence directly, only a few frames are selected to show the improvement in temporal domain (Fig. 4). Actually, in the proposed method, smooth disparity map sequence and rendered virtual view sequence can be observed obviously during playing. On the contrary, the result of Hierarchical BP (Yang *et al.*, 2006) dealing with each independent frame has fluctuant disparity value and finally causes flickering artifacts.

The improvement by adopting motion information in the proposed method is also illustrated. In the regions of motion objects between adjacent frames, the method based on regular 3D grid model is prone to generate

mismatch (Fig. 5a). After incorporating the motion information of optical flow, the matching errors are reduced (Fig. 5b).

Quantitative evaluation: Here, two basic measures are introduced focusing on the objective evaluation of the depth map sequences. First, the temporal consistency is checked between consecutive frames. Second, the quality of each separated disparity map is tested.

To check the temporal consistency, motion objects need to be separated from the test sequence. It is evident that only the background and static objects maintain the

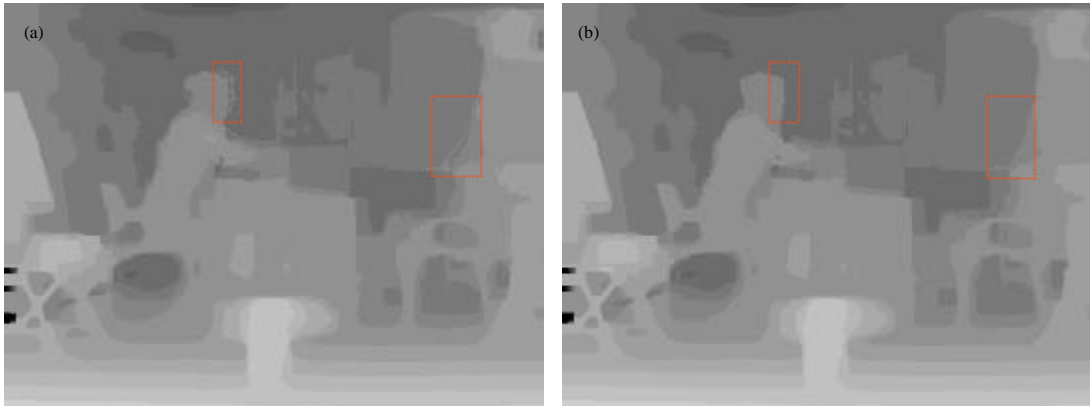


Fig. 5 (a-b): Improvement of disparity map at the boundary of motion objects. (a) Result of regular 3D grid model and (b) Result of the proposed model with motion compensation

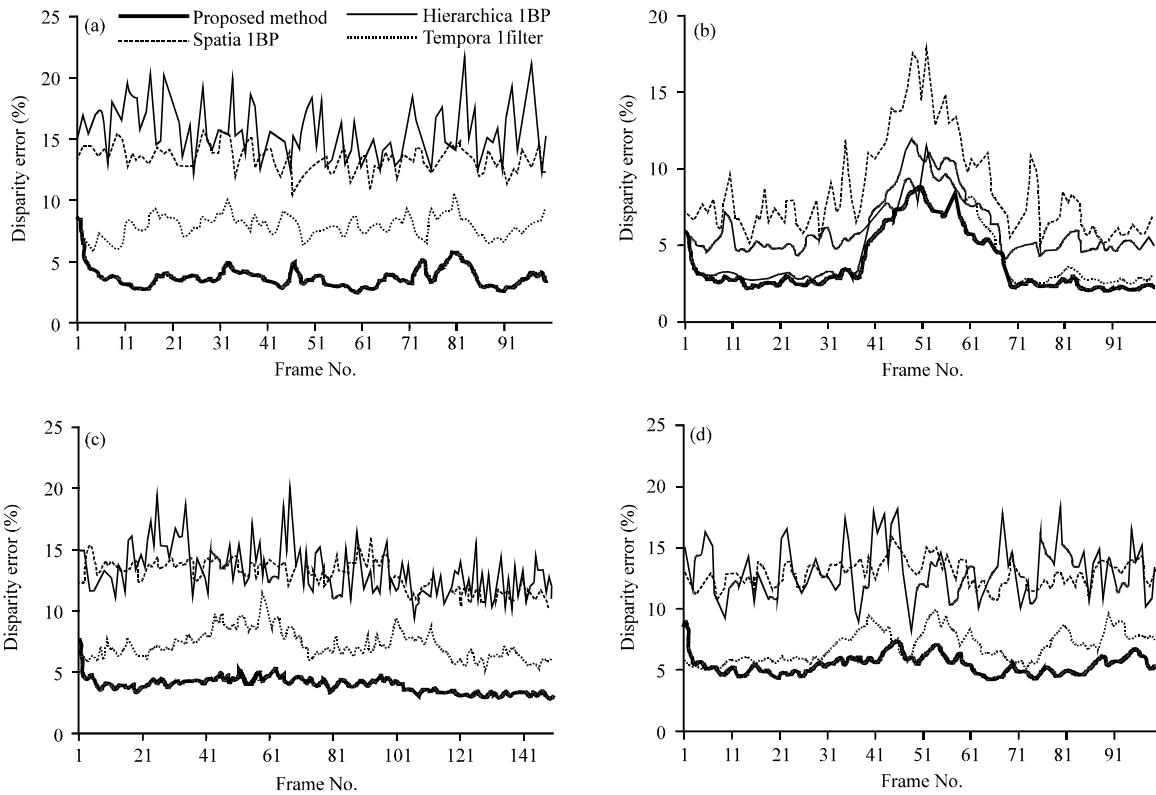


Fig. 6 (a-d): Comparative results of temporal consistency. (a) Book arrival, (b) Alt moabit, (c) Door flowers and (d) Leaving laptop. The pixel whose disparity value is inconsistent with the one in the next frame is identified as an error pixel. The total error percentage of each frame is compared among different methods marked by different colors

disparity value and viewers mostly feel flickering artifacts at these regions. The static regions are manually segmented for each test sequence. Then each disparity map of current frame is compared with the following one.

If the difference of any pair of pixels in the static regions exceeds a threshold, the pixel is marked as an inconsistent pixel. The percentage of total error pixels is calculated and the results are compared with other methods in Fig. 6:

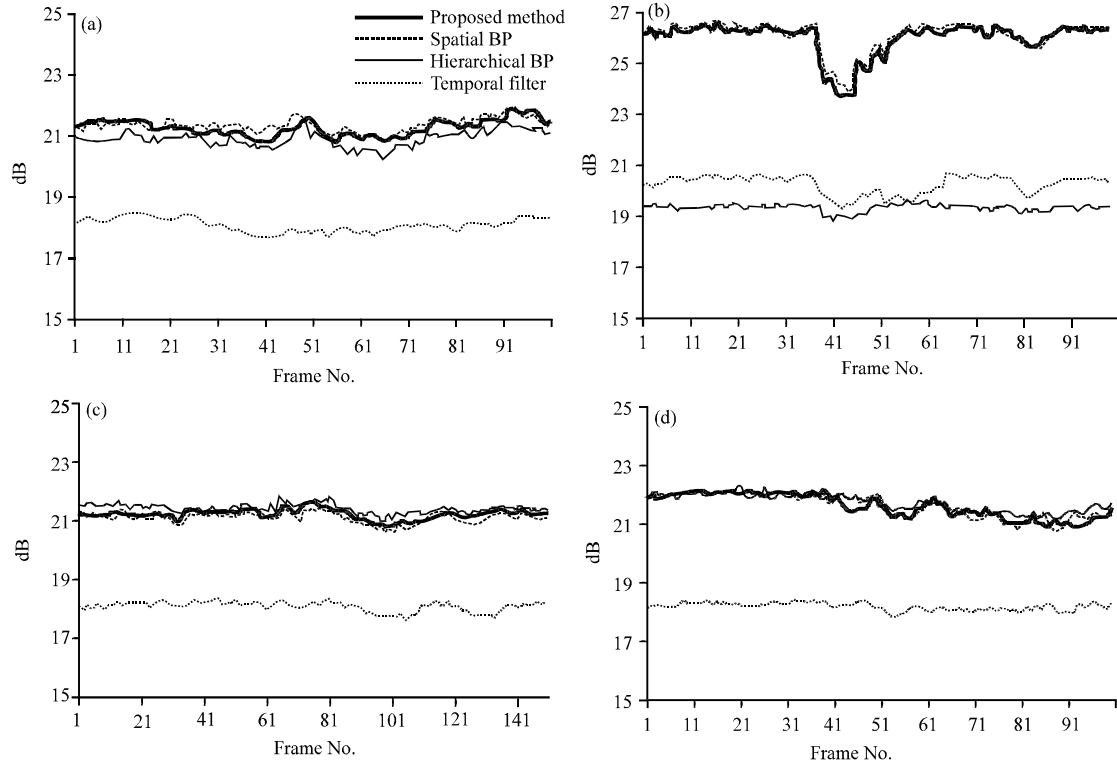


Fig. 7 (a-d): Comparative results of PSNR of synthesized virtual views. (a) Book arrival, (b) Alt moabit, (c) Door flowers and (d) Leaving laptop. Note: The virtual view is synthesized by the source view and the disparity map. PSNR of Y component is computed by comparing it with the original view taken at that location. The PSNR result of each frame is compared among different methods marked by different colours

$$\text{Error} = \frac{100}{N_{\text{total}} \rho_{\text{max}}} \sum_{\rho} \langle |d(p) - \hat{d}(p)| > \delta \rangle \quad (14)$$

On the other hand, accurate disparity map straightly brings on a good-quality synthesized image. Since the ground truth maps for these test sequences are unavailable, the experiments are made by means of evaluating the synthesized view. The left source view is projected to the corresponding position in right view using the disparity map. Then the obtained view is measured via PSNR of Y component in contrast with the original view. The pixels in the exposed regions with void value are set to 255. The regions close to right border of the image are ignored since they are invisible in the source view. Fig. 7(a-d) shows the comparative result of different methods:

$$\text{PSNR}_Y = 10 \times \log \left(\frac{255^2 \times N}{\sum_p |I(p) - \hat{I}(p)|^2} \right) \quad (15)$$

The results of objective measures are analyzed. The figures of the first experiment distinctly illustrate the temporal consistency in different sequences (Fig. 6a-d).

The error percentage on static regions is about 5% in the proposed method while the “spatial BP” and “Hierarchical BP” methods present more than 10% errors. As it can be inferred, by employing temporal coherency in the BP optimization framework, the proposed method presents lower error percentage in most of the datasets than other BP methods without temporal improvement. The temporal smoothing method with filter (Smirnov *et al.*, 2010) also has a relatively low error percentage which is between 5% and 10%. However, it fails to gain high performance in the experiment of image PSNR (Fig. 7). The result is 3 to 6 dB lower than the proposed method. It is because that the smoothing filters such as Gaussian filter deteriorate the intrinsic features of images. The proposed method is also compared with spatial BP method without temporal component in the same framework. Consequently they have comparable PSNR results with expectation. The proposed method also performs better than the Hierarchical BP method on ‘Book Arrival’ and ‘Alt Moabit’ sequences in the second experiment. And they achieve similar PSNR results on the rest sequences. As shown in Fig. 7, the PSNR result of the proposed method is about 21.24, 25.81, 21.20 and 21.50 on average,

respectively on the test sequences. While the Hierarchical BP method achieves about 20.89, 19.27, 21.35 and 21.52, respectively.

In brief, the experimental results demonstrate that the purpose of enhancing temporal consistency can be achieved at no loss of image quality and less additional operation is required in the implementation. The generated depth maps results are promising to be applied for the depth-image based rendering in 3DTV (three-dimensional television) system as well as other applications that require depth information for three-dimensional reconstruction.

CONCLUSIONS

Depth estimation plays a crucial role in three-dimensional video systems. This paper have presented a novel method for generating dense depth maps from stereo video sequences. Not only the quality of separate depth map but also the temporal consistency is concerned. The proposed method first applies BP algorithm for each frame. The obtained belief message is utilized as a measure of matching cost. Then, optical flow is used to connect consecutive frames and address the correspondence problem in form of a joint spatio-temporal function. It is defined as a recursive process. The cost is accumulated forward along the time axis so that any unnecessary information of the former frames can be released. The experimental results manifest that the proposed method can convincingly produce temporally consistent disparity maps without degradation of synthesized image quality.

ACKNOWLEDGMENTS

This stud was supported in part by the National Natural Science Foundation of China (Grant No. 60802013, 61072081), the National Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 2009ZX01033-001-007), Key Science and Technology Innovation Team of Zhejiang Province, China (Grant No. 2009R50003) and China Postdoctoral Science Foundation (Grant No. 20110491804).

REFERENCES

Bae, K.H., J.H. Ko and J.S. Lee, 2008. Errata: Stereo image reconstruction using regularized adaptive disparity estimation. *J. Electron. Imag.*, 10.1117/1.2976106
Bai, L., F. Chen and X. Zeng, 2008. Application of markov random field in depth information estimation of microscope defocus image. *Inform. Technol. J.*, 7: 808-813.

Barbu, A. and S.C. Zhu, 2005. Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27: 1239-1253.
Bleyer, M. and M. Gelautz, 2009. Temporally consistent disparity maps from uncelebrated stereo videos. *Proceedings of 6th International Symposium on Image and Signal Processing and Analysis*, Sep.16-18, 2009, Salzburg, pp: 383-387.
Boykov, Y., O. Veksler and R. Zabih, 2001. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23: 1222-1239.
Brown, M.Z., D. Burschka and G.D. Hager, 2003. Advances in computational stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25: 993-1008.
Dodgson, N.A., 2005. Autostereoscopic 3-D displays. *Computer*, 38: 31-36.
Du, X., H.D. Li and W.K. Gu, 2004. A simple rectification method for linear multi-baseline stereovision system. *J. Zhejiang Univ. Sci.*, 5: 567-571.
Fehn, C., 2004. Depth-Image-Based Rendering (DIBR), compression and transmission for a new approach on 3D-TV. *Proc. SPIE Stereoscopic Displays Virtual Reality Syst. XI*, 5291: 93-104.
Felzenszwalb, P.F. and D.R. Huttenlocher, 2004. Efficient belief propagation for early vision. *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, 1: I-261-I-268.
FhG-HHI, 2011. Mobile 3DTV content delivery optimization over DVB-H system. *Mobile 3DTV, Solideyesight*, <http://sp.cs.tut.fi/mobile3dtv/stereo-video/>.
Geman, S. and D. Geman, 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6: 721-741.
Hartley, R.I. and A. Zisserman, 2004. *Multiple View Geometry*. Cambridge University Press, Cambridge, UK.
Horn, B.K.P. and B.G. Schunck, 1981. Determining optical flow. *Artif. Intell.*, 17: 185-203.
Iffa, E.D., A.R.A. Aziz and A.S. Malik, 2011. Gas flame temperature measurement using background oriented schlieren. *J. Applied Sci.*, 11: 1658-1662.
Ihler, A.T., J.W. Fischer and A.S. Willsky, 2005. Loopy belief propagation: Convergence and effects of message errors. *J. Mach. Learn. Res.*, 6: 905-936.
Izabatene, H.F. and R. Rabahi, 2010. Classification of remote sensing data with markov random field. *J. Appl. Sci.*, 10: 636-643.
Kolmogorov, V., 2006. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28: 1568-1583.

- Konrad, J. and M. Halle, 2007. 3-D Displays and signal processing. *IEEE Signal Process Mag.*, 24: 97-111.
- Korah, R. and J.R.P. Perinbam, 2006. A novel coarse-to-fine search motion estimator. *Inform. Technol. J.*, 5: 1073-1077.
- Lee, S.B. and Y.S. Ho, 2010. View-consistent multi-view depth estimation for three-dimensional video generation. *Proceedings of IEEE 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, June 7-9, 2010, IEEE Xplore, pp: 1-4.
- Meesters, L.M.J., W.A. IJsselsteijn and P.J.H. Seuntjens, 2004. A survey of perceptual evaluations and requirements of three-dimensional TV. *IEEE Trans. Circuits Syst. Video Technol.*, 14: 381-391.
- Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1st Edn., Morgan Kaufmann Public Inc., San Francisco, CA., USA., ISBN: 0-934613-73-7.
- Proesmans, M., L. Van Gool, E. Pauwels and A. Oosterlinck, 1994. Determination of optical flow and its discontinuities using non-linear diffusion. *Comput. Vision*, 801: 294-304.
- Ren, G., P. Li and G. Wang, 2010. A novel hybrid coarse-to-fine digital image stabilization algorithm. *Inform. Technol. J.*, 9: 1390-1396.
- Scharstein, D. and R. Szeliski, 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47: 7-42.
- Shuchun, Y., Y. Xiaoyang, S. lina, Z. Yuping and S. Yongbin *et al.*, 2011. A reconstruction method for disparity image based on region segmentation and RBF neural network. *Inform. Technol. J.*, 10: 1050-1055.
- Smirnov, S., A. Gotchev and K. Egiazarian, 2010. A memory-efficient and time-consistent filtering of depth map sequences. *Proc. SPIE*, Vol. 7532, 10.1117/12.855811
- Stiller, C. and J. Konrad, 1999. Estimating motion in image sequences. *IEEE Signal Process. Mag.*, 16: 70-91.
- Tao, H., H.S. Sawhney and R. Kumar, 2001. Dynamic depth recovery from multiple synchronized video streams. *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, 1: I-118-I-124.
- Tappen, M.F. and W.T. Freeman, 2003. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. *IEEE Conf. Computer Vision*, 2: 900-906.
- Urey, H., K.V. Chellappan, E. Erden and P. Surman, 2011. State of the art in stereoscopic and autostereoscopic displays. *Proc. IEEE*, 99: 540-555.
- Vedula, S., P. Rander, R. Collins and T. Kanade, 2005. Three-dimensional scene flow. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27: 475-480.
- Veksler, O., 2003. Fast variable window for stereo correspondence using integral images. *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, 1: I-556-I-561.
- Xu, W. and H. Zi-Shu, 2011. Target motion analysis in three-sensor TDOA location system. *Inform. Technol. J.*, 10: 1150-1160.
- Yang, Q., L. Wang, R. Yang, S. Wang, M. Liao and D. Nister, 2006. Real-time global stereo matching using hierarchical belief propagation. *Proceedings of the British Machine Vision Conference*, January 16, 2006, Edinburgh, UK.
- Yu, S., D. Yan, Y. Dong, H. Tian, Y. Wang and X. Yu, 2011. Stereo matching algorithm based on aligning genomic. *Inform. Technol. J.*, 10: 675-680.
- Zhang, G., J. Jia, T.T. Wong and H. Bao, 2009. Consistent depth maps recovery from a video sequence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31: 974-988.
- Zhaozheng, Y. and R. Collins, 2007. Belief propagation in a 3D spatio-temporal MRF for moving object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 17-22, 2007, Minneapolis, MN., pp: 1-8.
- Zigh, E. and M.F. Belbachir, 2010. A neural method based on new constraints for stereo matching of urban high-resolution satellite imagery. *J. Applied Sci.*, 10: 2010-2018.
- Zwicker, M., A. Vetro, S. Yea, W. Matusik, H. Pfister and F. Durand, 2007. Resampling, antialiasing and compression in multiview 3-D displays. *IEEE Signal Process Mag.*, 24: 88-96.