Disparity Estimation using Belief Propagation for View Interpolation

Cheon Lee and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)
1 Oryong-dong, Buk-gu, Gwangju, 500-712, Republic of Korea
E-mail: {leecheon, hoyo}@gist.ac.kr

Abstract: In this paper, we propose a modified disparity estimation method using belief propagation (BP) for view interpolation. The view interpolation method generates an arbitrary intermediate view image with disparity. To extract disparities, we modified the BP based stereo matching algorith which solves the problem using Bayesian belief propagation. We propose two methods to improve accuracy of disparity. The first is about the matching cost computation. Instead of using pixel differences between corresponding pixels, we use block cost computation. The second is quad-tree region dividing for disparity refinement. Since narrow and accurate search range can guarantee accuracy of estimation, we divide the image into four distinct regions and refine disparities. By experiments, we improved quality of the interpolated images over 1~4 dB on average.

1. Introduction

Three-dimensional video (3DV) or free viewpoint TV (FTV) provides realistic feelings at arbitrary viewpoint with a wide range of viewing angles. Multi-view video data is input for these systems to offer free viewpoint selection and 3D display. Since the multi-view video data is captured by multiple cameras, techniques are different from single viewpoint video. Compression of a huge amount of data and interpolation between views are examples of issues.

The standardization of 3DAV (3 dimension audio-video) has been worked since December 2001 by moving pictures experts group (MPEG). After exploration of 3DAV, the standardization of multi-view video coding (MVC) has been progressed since August 2004 [1]. Many techniques are proposed such as prediction structure, illumination compensation, motion skip mode, and view interpolation prediction, etc. Recently, the joint draft 6.0 of MVC has been released in January 2008 [2]. For an extension of MVC, the standardization of FTV including multi-view depth data is working on since 2007 [3].

One possible approach to exploit view interpolation is the view interpolation prediction method for multi-view video coding. According to the prediction structure of MVC, some intermediate viewpoint frames are coded by referring to the adjacent view frames as well as temporal direction frames. The view interpolation prediction method in MVC is proposed to exploit the interpolated image using coded two adjacent view images as an additional reference frame [4]. If the generated frame is effective, it is selected for coding. Quality of the interpolated image is directly dependent on the coding performance.

Stereo matching is another terminology of disparity estimation. It is one of main research issues in computer vision. Especially, belief propagation (BP) based stereo matching algorithm is the best technique among various algorithms as a global method [5], [6]. However, it is sensitive about noise or similar pixel values because it employs the maximum disparity value and pixel-level matching cost. As increase the maximum disparity value, the disparity error rate is also increase. In this paper, we propose a disparity estimation method using the BP algorithm to improve the disparity estimation accuracy and consistency for view interpolation.

2. View Interpolation in Disparity Domain

Generation of a virtual viewpoint image with multiple images is the main target of image based rendering (IBR). To obtain an arbitrary viewpoint image, we can define a function for real world in terms of position, possible angle, wavelength and time. This is the plenoptic function proposed by Adelson and Bergen [7]. Even though it can give a perfect arbitrary viewpoint image, the dimension is too high to define. Many researchers have proposed to reduce the complexity limiting dimension of the function.

View interpolation is proposed by Chen and Williams, which can reconstruct arbitrary intermediate viewpoint image using only optical flow between views [8]. They exploit correspondence of pixels of objects. Disparity of the stereoscopic images is good correspondence information in 1D parallel camera rig. Because the multi-view video system uses 1D parallel camera rig, it is useful for intermediate view generation.

Disparity can be defined as a distance in horizontal coordinate of two corresponding pixels. This relationship is described in Eq. (1).

$$I_L(x,y) = I_R(x+d,y) \tag{1}$$

where d ($d \geq 0$) stands for disparity of two matching pixels between stereoscopic images, and it can be factorized into two disparities by α ($0 \leq \alpha \leq 1$), which is a new viewpoint between two images. The relationship for cameras is given as Eq. (2), where $\lfloor\ \rfloor$ is a rounding. $I(x,y)$ stands for the intensity value of the sample at position (x,y). $I\alpha(x,y)$ indicates the intermediate view point image.

$$\begin{aligned} I_\alpha(x,y) &= I_L(x+\lfloor(\alpha-1)d\rfloor, y) \\ &= I_R(x+\lfloor \alpha d \rfloor, y) \end{aligned} \tag{2}$$

3. Disparity estimation using BP Algorithm

Sun *et al.* proposed a stereo matching algorithm using belief propagation formulating the stereo matching problem as a Markov network and solved it using Bayesian belief propagation [9]. They defined Markov random fields with three coupled fields: D is the disparity field, L is a spatial line process of the depth discontinuity, and O is a spatial binary process to indicate occlusion regions. The solution of those fields is determined by solving belief propagation.

3. 1 Basic Streo Model

Using Bayes' rule, the joint posterior probability over D, L, and O given a pair of stereo images $I = \{I_L, I_R\}$, where I_L *and* I_R are the left (reference) and right images, respectively, which is represent as Eq. (3)

$$P(D,L,O \mid I) \propto P(I \mid D,L,O)P(D,L,O) \tag{3}$$

They assume that the likelihood $P(I|D,O,L)$ is independent of L, so it can be represented by the likelihood $P(I|D,O)$. It is defined as Eq. (4)

$$P(I \mid D,O) \propto \prod_{s \notin O} \exp(-F(s,d_s,I)) \tag{4}$$

where $F(x,d_s,I)$ is the matching cost function of pixels s with disparity d_s given observation I. For the matching cost, they use Birchfield and Tomasi's pixel dissimilarity as described in Eq. (5).

$$F(s,d_s,I) = \min\left\{\left|I_L(s) - I_R^-(s')\right|, \left|I_L(s) - I_R(s')\right|, \left|I_L(s) - I_R^+(s')\right|\right\} \tag{5}$$

where s' is the matching pixel of s in the right view with disparity d_s, and $I^-_R(s')$ is the linearly interpolated intensity halfway between s' and its neighboring pixel to the left.

For the prior probability $P(D,O,L)$, they ignored the statistical dependence between O and $\{D,L\}$, then it can be factorized as $P(D,O,L)=P(D,L)P(O)$. Following the Markov property, they defined $P(D,L)$ and $P(O)$ using joint clique potential function and user-customized function, respectively.

3.2 Loopy Belief Propagation

Three Markov Fields can be modeled by Markov network having an undirected graph as illustrated in Fig. 1. Nodes $\{x_s\}$ are hidden valriables representing the disparity value, and nodes $\{y_s\}$ are observed variables representing the pixel intensity value. By denoting $X = \{x_s\}$ and $Y = \{y_s\}$, the posterior $P(X|Y)$ can be factorized as Eq. (6).

$$P(X \mid Y) \propto \prod_s \varphi_s(x_s,y_s) \prod_s \prod_{t \in N(s)} \varphi_{st}(x_s,x_t) \tag{6}$$

where $N(s)$ is neighboring pixels of s, and $\varphi_{st}(x_s,x_t)$ is called the compatibility matrix between nodes x_s and x_t, and $\varphi_s(x_s,y_s)$ is called the local evidence for node x_s. This Eq. (6) is the same with Eq. (3) substituting $\varphi_{st}(x_s,x_t)$ for $P(D,O,L)$ and $\varphi_s(x_s,y_s)$ for $P(I|D,O,L)$.

There are two kinds of BP algorithms with different message update rules: "max-product" and "sum-product" which maximize the joint posterior $P(X|Y)$ of the network and the marginal posterior of each node $P(x_s|Y)$, respectively. The standard "max-product" algorithm is shown below:

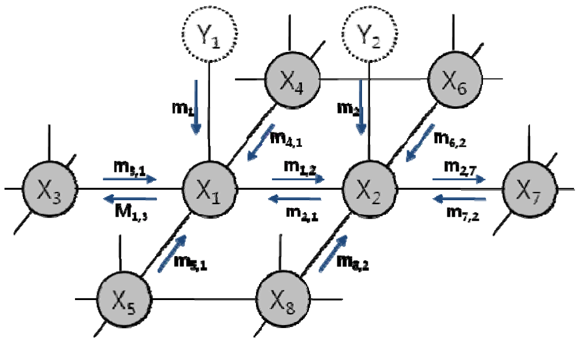


Fig. 1 Markov Network

1. Initialization: Initialize all messages $m_{st}(x_t)$ as uniform distributions and messages $m_s(x_s)$ = $\varphi_s(x_s,y_s)$.
2. Updating messages: Update messages $m_{st}(x_t)$ iteratively for i=1:T

$$m_{st}^{i+1} \leftarrow k \max_{x_s} \varphi_{st}(x_s,x_t) m_s^i(x_s) \prod_{x_k \in N(x_s)/x_k} m_{ks}^i(x_s)$$

3. Belief calculation:

$$b_s(x_s) \leftarrow k m_s(x_s) \prod_{x_k \in N(s_k)} m_{ks}(x_s)$$

$$x_x^{MAP} = \underset{x_k}{\operatorname{argmax}} b_s(x_k)$$

In the initialization process, the matching data cost is computed for every pixels with respect to every possible disparities. Every message from a certain node propagates to every other hidden node updating messages m^i_{st} by sufficient iterations. After T iterations, the beliefs become globally optimized function according to disparities. The x_k having maximum belief is the disparity value for the pixel s.

3.3 Effects of Disparity Range and Data Cost

The decision rule of disparity is choosing a disparity having the maximum belief. However, ambiguity increases by the similar colors in the same object and setting large disparity ranges. Noise of *Lambertian* surface can increase ambiguity as well. One of possible solution is to use the window based matching cost computation. Since the cost using window can represent the average color of the window, the effect of noise can be reduced. However, selection of appropriate window size is following problem.

Another problem is improper disparity range setup. As increase the disparity range, the probability of decision error getting increase. Most of disparity estimation algorithms assumed that the range is already known and used a fixed parameter. We want to point out that the assumption is not effective because the contents provider hard to know the actual disparity range for every scene. The proposed algorithm overcomes those two problems using initial disparity estimation and block-based matching cost for the observation nodes.

4. Proposed Disparity Estimation Algorithm

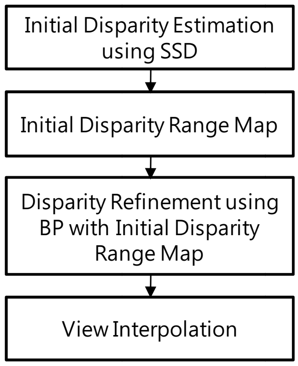


Fig. 2 Proposed disparity estimation procedure

In figure 2, we described our disparity estimation algorithm of view interpolation. Disparity estimation method using sum of squared differences (SSD) is a simple method in stereo matching. It performs quite stable in homogeneous object area, but it doesn’t in disparity discontinuous area. It is because that the estimator of SSD considers all pixels have the same disparity value. On the other hand, the BP based algorithm is very accurate around boundary region.

We have combined those two characteristics. We used the SSD method with gradient operator to obtain an initial disparity map, and then we set a disparity range map according to the initial disparity map. Refinement using BP is followed to refine the disparity values by referring to the initial disparity range map.

4.1 Initial Disparity Estimation

The objective of initial disparity estimation is generating a coarse disparity map to guide maximum disparity range without the pre-defined maximum disparity value. By the ordering constraint of stereo matching problem, we can reduce search range using region dividing. We find a disparity for a block starting from the most distinguishable block using SSD, and divide the region into left and right with respect to the corresponding block position. The subsequent block estimates the disparity using reduced search range. We already proposed this method for view interpolation prediction of multi-view video coding [10].

4.2 Block-level Data Cost

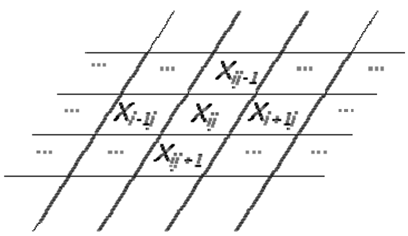


Fig. 3 Block-level Disparity Estimation

Sun *et al.* used pixel based matching cost. If reference images contain noise, the accuracy of disparity getting worse. Block-based cost calculation can be a solution. We used this solution since we deal with multi-view video sequence. We divide the image into blocks and calculate the matching cost using Eq. (7), which is the alternative cost function.

$$F(\tilde{s}, d_{\tilde{s}}, I) = \sum_{k\in W(\tilde{s})} \left| I_L(k) - I_R(k') \right| \qquad (7)$$

where $W(\tilde{s})$ is a set of pixels of block $\tilde{s}$ **.** The pixel *k* and *k’* are pixels in a block of the left image and its corresponding pixel by $d_{\tilde{s}}$ in the right image. The relationship of two corresponding pixels is Eq. (1). The number of observation nodes reduced by the block size, thus ambiguity of estimation decreases consequently.

4.3 Local Refinement using Quad-tree Region Dividing

The proposed block cost can guarantee accuracy of estimation than that of pixel based. However, it is not pixel-level method since estimator determines the same disparity for every pixel in a block. If the variation of disparity is low, this block based estimation method is effective, but it is hard to determine the variance of the scene in advance. After all, we need further step to refine the pixel-level disparity values.

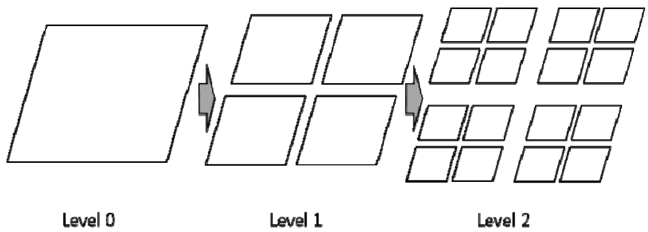


Fig. 4 Quad-tree based disparity refinement

For refinement, we use the quad-tree based region dividing method. As illustrated in Fig. 4, we set the image into several levels. The *level0* employs the basic block size. In the next level, we divide the image into 3 parts and perform the estimation process using reduced block size. Each divided partition uses different disparity range. If the partition is the upper left, the disparity range is determined by the minimum and maximum values of estimation results of *level0*. By this quad-tree based region dividing, we can enhance consistency and accuracy of disparity values

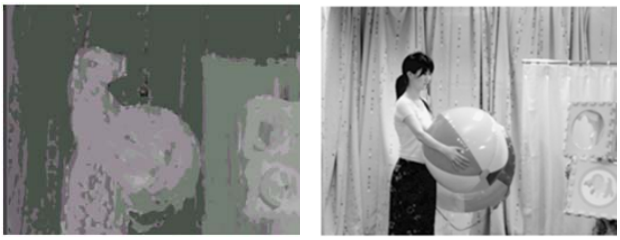
5. Experimental Results

5.1 Experiment conditions

The proposed disparity estimation algorithm for view interpolation is a kind of hybrid method using SSD and belief propagation algorithm. Actually, the SSD method used in our scheme employs the initial disparity estimation with gradient operator. The fundamental block size is 16x16 and the search range for the refinement is [-8, 8]. The number of iteration for BP is 60 which are enough iteration to maximize the performance.

In order to evaluate the proposed scheme, we compared quality of synthesized images in terms of PSNR. We use multi-view video sequences as input data such as ‘Akko&Kayo’ and ‘Rena’. Those are test sequences of the multi-view video coding. For example, if we generate a center view like 'view1' referring to two views 'view0' and 'view2', we set the actual image 'view1' as a ground truth image. We calculated the PSNRs between the generated image and the ground truth image.

5. 2 Results of view interpolation



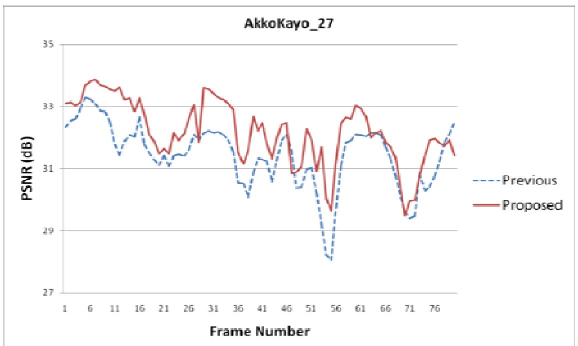
(a) Results of the previous method

(b) Results of the proposed method

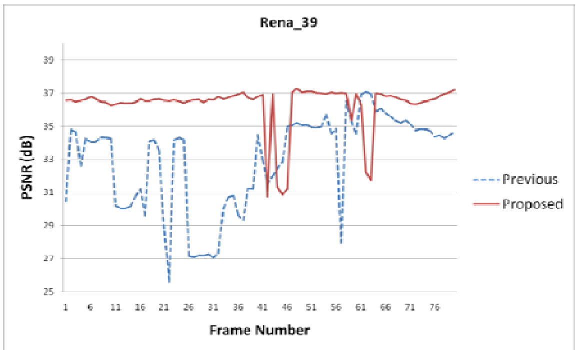
Fig. 5. Disparity map and synthesized images

The previous method is the stereo matching algorithm proposed by Sun *et al*. After generating the disparity map, we synthesized the intermediate view image using Eq. (2). The target views are 'Akko&Kayo_27', and 'Rena_39'. In case of 'Akko&Kayo', we used two reference frames (view 26 and 28) to synthesize view 27. Similarly, view 39 of 'Rena' sequence is interpolated using view 38 and 40. After generating the target view images, we calculated PSNRs between original and synthesized images. Figure 3 shows the visual results. Fig. 5(a) is the result of the initial disparity based method, and Fig. 5(b) is the result of the proposed algorithm. We can easily notify the visual consistency of the disparity maps has been improved. The disparity map of the proposed method is more consistent than the previous method. Quality of the synthesized image is also improved as well.

We also compared the PSNR values for 80 frames as shown in Fig. 6. Overall PSNR values are greater than that of the previous method. The average PSNRs are improved by 1~4dB. PSNR values are more stable than the previous method in 'Rena' sequence.



(a) PSNRs of 'Akko&Kayo_27' sequence



(b) PSNRs of 'Rena_39' sequence

Fig. 6. Comparison of PSNR values

6. Conclusion

The view interpolation method is the key technique in free viewpoint TV (FTV) or 3DV technology. In this paper, we proposed an efficient disparity estimation method using the block based matching cost computation and quad-tree based refinement to improve the belief propagation algorithm. In addition, we used the initial disparity estimation method to reduce ambiguity due to the large disparity range, and set the matching cost using blocks. We used the disparity refinement method using the quad-tree region dividing. The disparity map improved the disparity consistency and accuracy, and the average PSNRs are improved by 1~4dB.

Acknowledgement

This work was supported in part by ITRC through RBRC at GIST (IITA-2008-C1090-0801-0017).

References

[1] ISO/IEC JTC1/SC29/WG11, "Requirements on Multi-view Video Coding v.7," N8218, July 2006.

[2] JVT of ISO/IEC MPEG & ITU-T VCEG, "Joint Draft 6.0 on Multiview Video Voding," JVT-Z209, 2008.

[3] ISO/IEC JTC1/SC29/WG11, "Preliminary FTV Model and Requirements," N9168, Lausanne, Switzerland, July 2007.

[4] JVT of ISO/IEC MPEG & ITU-T VCEG, "View Interpolation for Multi-view Video Coding," JVT-U102, Oct. 2006.

[5] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient Belief Propagation for Early Vision," *Computer Vision and Pattern Recognition*, 2004.

[6] http://vision.middlebury.edu/stereo/eval/

[7] E. H. Adelson and J. Bergen, "The Plenoptic Function and the Elements of Early Vision," *in Computational Models of Visual Processing, MIT Press*, 1991.

[8] S. Chen and L. Williams, "View Interpolation for Image Synthesis," *Proc. ACM*, 1993.

[9] J. Sun, N.N. Zheng, and H.Y. Shum, "Stereo Matching Using Belief Propagation," *IEEE Trans. PAMI,* vol. 25, no. 7, July 2003.

[10] C. Lee, K.J Oh, and Y.S. Ho, "View Interpolation Prediction for Multi-view Video Coding," *Picture Coding Symposium (PCS)* , Nov. 2007.