Proceedings of the 6th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 16-18, 2006 (pp44-49)

Intermediate view synthesis and interpolation based on disparity calibration and region segmentation

PING AN[1], ZHAOYANG ZHANG[1,2], RAN MA[1]
[1] School of Communication and Information Engineering
[2] Key Laboratory of Advanced Displays and System Application, Ministry of Education
Shanghai University
149 Yanchang Road, Shanghai 200072
CHINA

Abstract: - Since image quality of the synthesized intermediate view depends mainly on precise and dense disparity field and on detection of occlusion in multi-view applications, this paper proposes an intermediate view creation and interpolation algorithm taking advantages of transmitted and decoded disparity fields to avoid complex disparity calculation. This algorithm also combines region segmentation and disparity calibration techniques together to correct those mismatched disparities. What's more, in the intermediate view creation, region segmentation technique is used to judge the contents of regions with occlusion and blur, and fill them depending on the disparity characteristics. Experimental results show that the constructed intermediate view images have favorite view effect. This algorithm can be used in 3D videoconference systems in which the intermediate view images have to be synthesized in real-time.

Key-words: - Intermediate view synthesis, Stereo image, 3D videoconference, Region segmentation, Disparity calibration, Decoded disparity field

1 Introduction

In recent years, many researches focus on telepresence videoconference that raise the demand for new interactive communication capabilities between remote partners.[1-6] Telepresence videoconferencing systems give the user an illusion of true contact, bringing participants together in a virtual space. When he moves his head, a user should be able to acquire corresponding views, and thus "see around" the conferees or 3D objects. One main principle of this technique is the synthesis of intermediate views not acquired by physical cameras. The image quality of the synthesized intermediate view depends mainly on accuracy of the disparity estimates. Therefore, many algorithms for intermediate synthesis focus on estimating accurate disparity field to improve image quality. However, the process for disparity estimation usually consumes large amount of calculation that is disadvantageous for some real time applications. [3~6]

This paper proposes an algorithm for intermediate view creation and interpolation by taking advantages of the decoded disparity fields to avoid mass calculation in disparity estimation. However, these kinds of disparity fields usually involve a lot of mismatched disparities that are not usable for intermediate view creation. Therefore, we propose a disparity calibration technique combined with region segmentation to correct those mismatched disparities. The region segmentation is further used in intermediate view creation to judge the contents of regions with occlusion and blur, and fill them depending on the disparity characteristics. Since the intermediate view images are directly created at the decoder, the proposed algorithm can be used in applications emphasizing real time and not the high image quality such as videophone conference, telecommunication and visual mobile communication.

The structure of this paper is organized as follows. In section 2, we introduce a region growing method for region segmentation. Section 3 proposes a method for disparity calibration. Section 4 describes the intermediate view creation procedure. Experimental results are presented in section 5 and final is the conclusion.

2 Region Segmentation

Region growing is a common used region segmentation approach that groups pixels into regions based upon two criteria: proximity and homogeneity. According to the criteria, image can be split into smaller regions, and small regions can be merged into larger ones. We use a causal 4 neighborhood as shown in figure 1, which can obtain better region comparing effect. [7]

2	3	4
1	X	

Fig.1 Neighboring regions of region X

The second criterion, homogeneity, is satisfied by an implementation specific function that quantifies the similarity between regions. Suppose each region is modeled as N pixels sampled from a normally distributed region with mean μ and standard deviation σ. Each pair of neighboring region is compared as follows.

1) Homogeneity comparison 1: The difference ($\Delta\mu$) between the two means (μ_1 and μ_2) is calculated and compared with a parameter specifying the maximum allowed difference ($\Delta\mu_{\max}$).

$$\Delta\mu = |\mu_1 - \mu_2| \quad , \quad \Delta\mu \le \Delta\mu_{\max} \,? \tag{1}$$

2) Region homogeneity comparison 2: If the comparison 1 is satisfied, the new mean (μ_{new}) formed by merging the two regions is calculated. The difference between the first mean (μ_1) and the new mean (μ_{new}) is divided by the first standard deviation (σ_1) to produce the number of standard deviations between the first mean and the new mean ($\sigma_{1,new}$). This is compared with a second parameter defining the maximum number of standard deviations (σ_{max}) by which the first mean and new mean may vary.

$$\mu_{\text{new}} = \frac{N_1\mu_1 + N_2\mu_2}{N_1 + N_2} \,, \sigma_{1,\text{new}} = \frac{|\mu_{\text{new}} - \mu_1|}{\sigma_1} \,, \sigma_{1,\text{new}} \le \sigma_{\max} \,? \tag{2}$$

Where, N_1 and N_2 are pixels contained in the two compared regions respectively. The first comparison prevents merging between regions of very different tissues regardless of their size and standard deviation. The second comparison, while allowing absorption by large regions of small neighbors, ensures that significantly different regions of comparison size are preserved. If both comparisons are satisfied, the two regions are considered to be homogeneous.

By using the region-growing algorithm, we first split each image into initial seed regions made up of square pixel blocks, such as 2 pixels on a side. The algorithm then iteratively scans the image, comparing each region (shown as 'X' in figure 1) with its causal 4 neighborhood. After the image finishes iteratively growing, regions with area smaller than a certain minimum size can be merged with the neighbor region with the closest mean. At last, the image is segmented into lots of regions with homogeneous pixels within each one.

3 Error detection and correction for disparity field

Gray scale image segmentation usually depends on two characteristics of the pixels' intensities: discontinuity and similarity. Pixels within one region usually are similar in intensity, but discontinuous at boundaries between regions. In stereo image pairs, this discontinuity often implies belonging to different object backgrounds, so that result in discontinuity in depth and disparity. Thus the segmentation technique can be used in disparity error detection and correction in order to avoid errors due to wrong mapping.

Regard left image as the reference image, and right image as the target image. Applying hierarchical overlapped block disparity estimation algorithm we proposed in reference [8] to evaluate the disparities, the disparity field of left-to-right image is obtained. Used for stereo video coding, algorithm proposed in reference [8] can improve the coding effect. The disparity field can be restored at the decoder after coding and transmitting. If decoded disparity field is directly used in view synthesis, inferior intermediate view image will be created due to much mismatching. In order to improve the disparity accuracy, the decoded disparity field should be calibrated with region segmentation before view synthesis.

The region segments of left image are obtained with the region-growing algorithm described in section 2. Then, pixels belonging to the same region should belong to the same object or background, thus their disparities or depths are similar.

Let L and R denote the left and right stereo images, b is the baseline distance between the two camera's focus. According to imaging geometry theory in parallel light axes, disparity d^{LR} between corresponding pixels respectively located at the same scan line in the two images is

$$d^{LR} = \frac{bf}{z} \tag{3}$$

Where, f is the focus, z denotes the distance from the lens plane to some point P in real world, i.e. the depth of object.

Given the disparity field and the segmentation regions, stating the average $\bar{d}_i$ of the disparity vectors belonging to the same region in left image, we have

$$\bar{d}_i = \frac{1}{\Omega_i} \sum_{\Omega_i} d_k^{LR} \qquad d_k^{LR} \in \Omega_i \tag{4}$$

Where, i is the region index, Ω_i is the total disparity vectors number in the region, d_k^{LR} denotes a disparity value in the region. When calibrating the disparities, set a smooth window Ws for each scan line and move the window from left to right. If pixels within the window belong to the same segmentation region i, they will belong to the same object or background. Calculate average of the disparities within the window, noted as $\bar{d}_i^W$. Suppose disparity of the center pixel in the window is $\bar{d}_i^C$, we have

$$\bar{d}_i^C = \begin{cases} \bar{d}_i^W & \left\|\bar{d}_i^W - \bar{d}_i\right\| \le \lambda \\ \bar{d}_i & else \end{cases} \tag{5}$$

The first item of above equation indicates that the matching is correct if the disparities mean within window satisfy $\left\|\bar{d}_i^W - \bar{d}_i\right\| \le \lambda$ (λ is the error threshold permitted). Then replace disparity of the center pixel in the window with $\bar{d}_i^W$, so that disparities within the window Ws are smoothed. In many disparity estimation methods based on block, there is distinct disparity jump between adjacent blocks due to illumination change or surface slant even if they belong to the same object or background. Replace disparity of the center pixel in the window with $\bar{d}_i^W$, which make a smooth transition between disparities of adjacent blocks, so that a more smooth disparity field can be obtained.

The second item in equation (5) indicates that there are distinct errors for disparities within the window if the disparities mean doesn't satisfy $\left\|\bar{d}_i^W - \bar{d}_i\right\| \le \lambda$. Then replace disparity of the window center pixel with the region disparity mean $\bar{d}_i$.

4 Intermediate view creation

To create an intermediate view, for each matched pixels pair in left and right images, position of their correspondence in intermediate view image should be searched according to disparity field. Thus the disparity field should be projected to the intermediate view image. Suppose there is a virtual camera located between the left and right cameras, and the distance from left camera to the virtual one is x, disparity from projection point P^L to the intermediate projection point P^I is d^{LI}. Base on equation (3), we have

$$d^{LI} = \frac{xf}{z} = \left(\frac{x}{b}\right) d^{LR} = \alpha d^{LR}; \quad 0 \le \alpha \le 1 \tag{6}$$

Where, α is the interpolation factor. Let X^L and X^R denote positions of the corresponding points P^L and P^R respectively. Then, position of the corresponding point P^I in intermediate image can be predicted as follows.

$$X^I = \alpha X^L + (1-\alpha) X^R = X^L - (1-\alpha) d^{LR} = X^R + \alpha d^{LR} \tag{7}$$

For all points un-occluded in left and right images, their correspondence in intermediate image can be predicted directly from points in left and right images and their disparities. For the occluded points without specific disparity definition, we apply region segmentation technique to judge and fill. Steps of the intermediate view synthesis algorithm based on region segmentation are listed as follows.

Step 1: Perform the above region-growing algorithm on left and right image to get respective segmentation regions.

Step 2: Applying disparity error detection and correction, calibrate the two disparity fields of left-to-right and right-to-left, and obtain two new disparity fields denoted as d^{LR} and d^{RL}.

Step 3: Scanning the image from left to right, for each point in intermediate image, respectively calculate positions of the two correspondences in left and right images by using equation (7).

Step 4: Create the intermediate image according to the region segments of left image firstly. If the correspondence in left image locates at the same region in right image by disparity mapping, which means intensity of the corresponding point in intermediate image is similar as its correspondences in left or right image, which can be interpolated from the weighted mean of pixel intensity in left and right images. On the contrary, if the disparity doesn't map the correspondences in the same region, which indicates the pixel is only visible in one image and directly replaced by intensity value in left image. Thus, the pixel intensity value $I^{LI}(X + d_i^{LI}, Y)$ in intermediate image can be estimated as follows.

$$I^{LI}(X + d_i^{LI}, Y) = \begin{cases} \alpha * I^L(X,Y) + (1-\alpha) * I^R(X + d^{LR}, Y) & \Omega_i^L = \Omega_i^R \\ I^L(X,Y) & else \end{cases} \tag{8}$$

Where, $I^L(X,Y)$ is the intensity value of a pixel in left image, and $I^R(X+ d^{LR},Y)$ is the intensity value of the correspondence in right image.

Similar processing is performed on the right-to-left disparity field. The synthesized intermediate view image is

$$I^{RI}(X - d_i^{RI}, Y) = \begin{cases} (1-\alpha) * I^L(X + d^{RL}, Y) + \alpha * I^R(X,Y) & \Omega_i^L = \Omega_i^R \\ I^R(X,Y) & else \end{cases} \tag{9}$$

Step 5: According to characteristics of stereoscopic imaging, the scene projected in left image is on the right side, and on the left side in right image. Therefore, for area with position $X < (1-\alpha)*W$ (W denotes the image width) in the synthesized image, its content mainly refers the left image, and other image content is mainly created from the right image. Thus the final intensity field of the synthesized image is calculated as follows.

$$I^I(X,Y) = \begin{cases} \alpha * I^{LI}(X,Y) + (1-\alpha) * I^{RI}(X,Y), & \left\| I^{LI}(X,Y) - I^{RI}(X,Y) \right\| < \delta \\ I^{LI}(X,Y), & I^{RI}(X,Y) \text{ is occluded, } X < \alpha * W \\ I^{RI}(X,Y), & I^{LI}(X,Y) \text{ is occluded, } X \ge \alpha * W \end{cases} \tag{10}$$

For each created intermediate image point, record its intensity, region segment index, and the two disparities d^{LR} and d^{RL}.

Step 6: Mark all pixels without prediction in intermediate image. If depth at the left side of the mismatched pixel is smaller than the one at right side, its disparity is regarded as the same of the left side pixel, and its intensity can be calculated by equation (8) from the left image. Otherwise, copy the disparity of its right side pixel, and the intensity is estimated by equation (9) from the right image.

If the mismatched pixel locates at the right side of the scan line, disparity of its left side pixel is copied, and its intensity can be calculated by equation (8) from the right image. Contrarily, estimated by equation (9). This bases on hypothesis that the occluded areas are parts of the nearest un-occluded background.

Repeat step 6, until all occluded pixels are estimated.

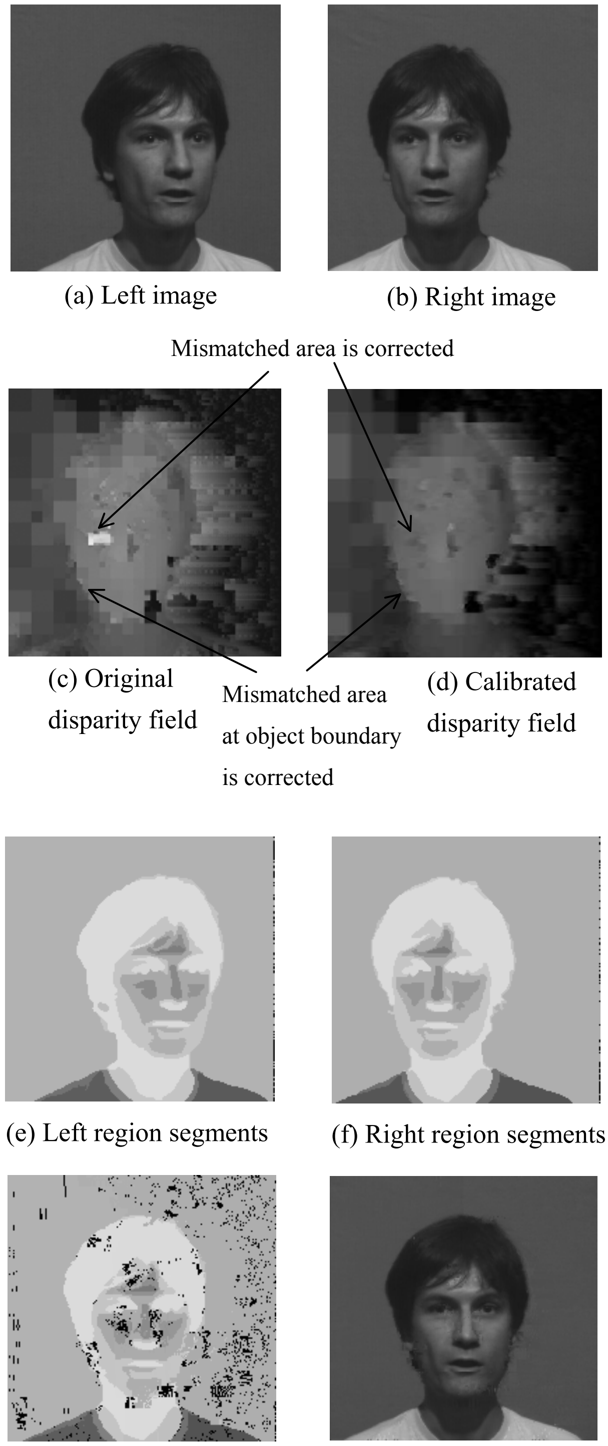
5 Experimental results

The proposed algorithm has been tested with several stereo image pairs. Figure 2 shows the results of stereo pair "Man". Where, figure 2(a) and (b) are the left and right images, and figure 2(c) is the decoded disparity field with the method in reference [8], figure 2(d) is the calibrated disparity field. Compare the two disparity maps, we find that some mismatched areas in the original disparity field are corrected and the disparity field is obviously improved and becomes much more smooth. Figure 2(e) and (f) respectively show the segmentation regions of left image and right image. Figure 2(g) illustrates the region segments of intermediate image after step 5, where the black areas denote the occlusions that will be filled in step 6. Shown as figure 2(h), quality of the created intermediate image is accepted.

Test results for another stereo pair "Tree" are shown as figure 3. Since nature scene image with complicated texture has better error concealment performance for our eyes can't perceive minor distortion in complicated texture, such that we obtain an intermediate view with high quality.

6 Conclusion

A procedure to synthesize the intermediate views based on disparity calibration, region segmentation and linear interpolation is presented for real time stereo video system. The proposed algorithm doesn't depend on disparity accuracy as many synthesis methods do, and its computation burden is lower. Experimental results illustrate that the proposed algorithm synthesizes intermediate view with reasonable image quality. The algorithm can be used in applications needing coding and transmission, and requiring real time synthesis.



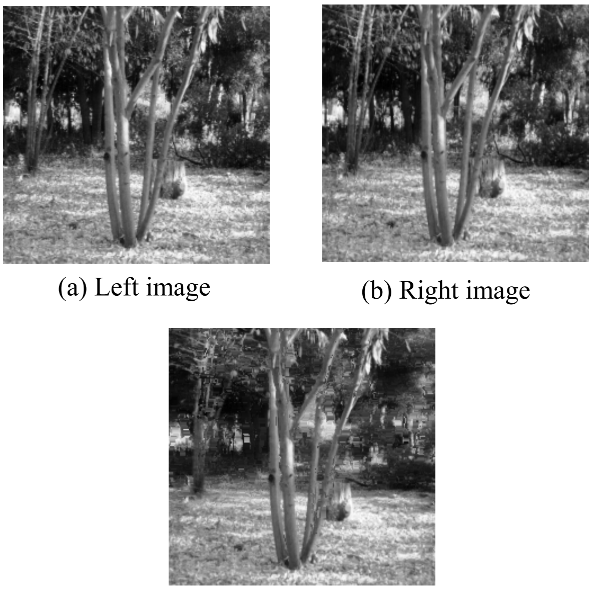
(a) Left image (b) Right image

(c) Original disparity field (d) Calibrated disparity field

(e) Left region segments (f) Right region segments

(g) Intermediate image region (h) Synthesized image

Fig.2 Experimental results of Man



(a) Left image (b) Right image

(c) Synthesized image

Fig.3 Experimental results of Tree

Acknowledgements

This project was supported by the Nature Science Foundation of Shanghai (04ZR14056), Development Foundation of Shanghai Municipal Commission of Education (04AB56) and Natural Science Foundation of China (60572127).

References:

[1] J. Park, S.Inoue. Arbitrary view generation from multiple cameras. *In Proceedings of the IEEE International Conference on Image Processing* 1997, 1: pp.149-152

[2] Mansouri, A. R., Konrad, J.: Bayesian Winner-take-all Reconstruction of Intermediate Views from Stereoscopic Images. *IEEE Trans. on Image Processing*, Vol. 9, 2000, pp.1-13

[3] M. Ott, J. Lewis, and I. J. Cox. Teleconferencing eye contact using a virtual camera. *In INTERCHI '93*, 1993, pp. 109-110

[4] D. Scharstein. Stereo vision for view synthesis. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'96)*, San Francisco, CA, June 1996, IEEE Computer Society Press, pp.852-858

[5] Jens-Rainer O, Ebroul Izquierdo M. An object-based system for stereoscopic viewpoint synthesis. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.7, No.5, 1997, pp.801-811

[6] A. R. Mansouri, J. Konrad. Bayesian winner-take-all reconstruction of intermediate views from stereoscopic images. *IEEE Transactions on Image Processing*, Vol.9, No.10, 2000, pp.1-13

[7] I. Pitas, Digital Image Processing Algorithms, *Prentice Hall*, 1993

[8] An Ping , Zhang Zhaoyang, Hierarchical Overlapped Block Disparity Estimation and Adaptive Compensation for Stereo Images, *Journal of China Institute of Communications (in Chinese)*, Vol.24, No.3, 2003, pp.1-7