Block-Wise *MAP* Disparity Estimation for Intermediate View Reconstruction

Liang Zhang

Communications Research Centre Canada
3701 Carling Avenue, Ottawa, Ontario, Canada K2H 8S2
Email: liang.zhang@crc.ca

ABSTRACT

A dense disparity map is required in the application of intermediate view reconstruction from stereoscopic images. A popular approach to obtaining a dense disparity map is *maximum a-posteriori* (*MAP*) disparity estimation. The *MAP* approach requires statistical models for modeling both a *likelihood* term and an *a-priori* term. Normally, a *Gaussian* model is used. In this contribution, block-wise *MAP* disparity estimation using different statistical models are compared in terms of Peak Signal-to-Noise Ratio (*PSNR*) of disparity-compensation errors and number of corresponding matches. It was found that, among the *Cauchy*, *Laplacian*, and *Gaussian* models, the *Laplacian* model is the best for the *likelihood* term while the *Cauchy* model is the best for the *a-priori* term. Experimental results show that reconstruction algorithm with the *MAP* disparity estimation using the determined models can improve image quality of the intermediate views reconstructed from stereoscopic image pairs.

Keywords: 3-D television, stereoscopic images, disparity estimation, intermediate view reconstruction.

1. INTRODUCTION

Since three-dimensional (3-D) television can provide viewers with an impression of depth and a greater sense of presence, 3-D TV system could be the next major rung in the evolution of television[1]. As the main depth cue comes from stereovision, it seems natural to record and distribute 3-D signals as separate video streams, one for the left eye and the other for the right eye. However, the feeling of being present in the scene is limited, as the stereoscopic video is usually not adapted to every observer's viewpoint[2]. Moreover, a frequent major deficiency of stereoscopic visualization is excessive stereoscopic 3-D cue that causes visual discomfort[3]. To enable viewer-dependent viewpoint adaptation, as well as control of the strength of stereoscopic 3-D cues, a technique for intermediate view reconstruction is needed.

The reconstruction of arbitrary intermediate views can be done by interpolation from the left-eye and right-eye images based on knowledge of the depth information contained in these two images. The depth information can be obtained by estimating the disparity between the left-eye and right-eye images. A general approach to obtaining disparity maps from the two images involves locating corresponding points by measuring intensity differences between the left-eye and right-eye images[4].

A popular approach is *maximum a-posteriori* (*MAP*) disparity estimation, in which statistical models for the *likelihood term* and the *a-priori term* are required. Normally, a *Gaussian* probability distribution is used to model the *likelihood* term in *MAP* disparity estimation[5-8]. For modeling the *a-priori* term, Belhumeur used a *Gaussian* probability distribution[5]. He explained this term as a constraint of surface smoothness. Falkenhagen measured probabilities for disparity differences with natural stereoscopic image pairs[7] and did not assign a specific probability distribution to model this *a-priori* term. Among the models, *Cauchy*, *Laplacian* and *Gaussian*, Sebe, et al, found that a *Cauchy* probability distribution led to the best results for pixel-wise *maximum likelihood* (*ML*) disparity estimation[9]. We also

Stereoscopic Displays and Virtual Reality Systems XII, edited by Andrew J. Woods, Mark T. Bolas, John O. Merritt, Ian E. McDowall, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 5664 © 2005 SPIE and IS&T · 0277-786X/05/$15

found that the *Gaussian* model did not work as well as the *Laplacian* model and that pixel-wise *ML* disparity estimation is not reliable compared with the block-wise *ML* method[10].

In this contribution, statistical models are investigated for block-wise *MAP* disparity estimation. To determine the model for the *likelihood* term, block-wise *MAP* disparity estimation under the assumption, that the *a-priori* term has a uniform model, is implemented using different models, a *Cauchy*, a *Laplacian* and a *Gaussian* model. The performances of these *MAP* approaches with different models are then compared in terms of *PSNR* of disparity-compensation errors against the number of corresponding matches. To examine a model for the *a-priori* term, statistical characteristics of disparity map is measured and then modeled using a *Cauchy*, a *Laplacian* and a *Gaussian* model. Checking the model error gives the best model for the *a-priori* term.

This paper is organized as follows. Section 2 briefly describes *MAP* disparity estimation. After that, statistical models are examined for block-wise *MAP* disparity estimation with natural stereoscopic image pairs in Section 3. Section 4 gives the implementation of *MAP* estimation with the models determined in the previous section. Section 5 presents the experimental results and its application to intermediate view reconstruction. A short conclusion is drawn in the final section.

2. *MAP* DISPARITY ESTIMATION

Let s_l and s_r be the left-eye and right-eye images. $\varphi(\boldsymbol{p}_j)$ is the true disparity value at the pixel position $\boldsymbol{p}_j$ that maps the right-eye image to the left-eye image, i.e.,

$$s_l(\boldsymbol{p}_j) = s_r(\boldsymbol{p}_j - \varphi(\boldsymbol{p}_j)). \tag{1}$$

The goal of the disparity estimation is to find an estimate $d(\boldsymbol{p}_j)$ of the true disparity $\varphi(\boldsymbol{p}_j)$ that minimizes the intensity difference $s_l(\boldsymbol{p}_j)$- $s_r(\boldsymbol{p}_j$- $d(\boldsymbol{p}_j))$.

A block-wise *maximum a-posteriori* (*MAP*) disparity estimation takes neighborhood pixels into consideration by assuming that all pixels within one block share the same disparity value and is defined as[7]

$$d = arg \max_{\gamma} \left\{ f(\gamma, \bar{s}_{r,B} \middle| \bar{s}_{l,B}) \right\} \quad . \tag{2}$$

According to the *Bayesian* rule, the *a-posteriori* joint probability density $f(\gamma, \bar{s}_{r,B} \mid \bar{s}_{l,B})$ can be rewritten as

$$\frac{f(\bar{s}_{l,B} \mid \gamma, \bar{s}_{r,B})\, f(\bar{s}_{r,B} \mid \gamma)\, f(\gamma)}{f(\bar{s}_{l,B})} = f(\bar{s}_{l,B} \mid \gamma, \bar{s}_{r,B})\, f(\gamma)\ , \tag{3}$$

since $\bar{s}_{r,B}$ and γ are independent and the probability density $f(\bar{s}_{l,B})$ is assumed to be the same as $f(\bar{s}_{r,B})$. The joint probability density $f(\bar{s}_{l,B} \mid \gamma, \bar{s}_{r,B})$ is a measure of how well one right-eye image block with the disparity value γ matches the left-eye image block and is referred to as the *likelihood* term. The probability density $f(\gamma)$ is a measure of how probable a particular γ is *a-priori*, i.e., before the left-eye image is observed, and is referred to as the *a-priori* term. $f(\gamma)$ is represented as[7]

$$f(\gamma) = \prod_j f(\gamma(\boldsymbol{p}_j) - \gamma(\boldsymbol{p}_{j-1})) = \prod_j f(\Delta\gamma(\boldsymbol{p}_j)), \tag{4}$$

which is interpreted as the smoothness constraints for disparity values[5]. To implement a block-wise *MAP* disparity estimation, statistical models for the *likelihood* term $f(\bar{s}_{l,B} \mid \gamma, \bar{s}_{r,B})$ and the *a-priori* term $f(\gamma)$ have to be determined first.

3. STATISTICAL MODELS

To select a suitable model for block-wise *MAP* disparity estimation, the performance of the disparity estimation with three commonly used statistical models, namely *Cauchy* model with a parameter a

$$f_C(x) = \frac{a}{\pi} \frac{1}{a^2 + x^2}, \tag{5}$$

Gaussian model with a variance σ^2

$$f_G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{x^2}{2\sigma^2}\right\}, \tag{6}$$

and *Laplacian* model with a variance σ^2

$$f_L(x) = \frac{1}{\sqrt{2}\sigma} \exp\left\{\frac{|x|}{\sigma/\sqrt{2}}\right\}, \tag{7}$$

were compared.

3.1. Statistical model for the *likelihood* term

For modeling the *likelihood* term, we have found in aprevious study that a *Gaussian* model did not work as well as a *Laplacian* model for block-wise *maximum likelihood* (*ML*) estimation[10]. Here, we compare whether a *Laplacian* model is superior to a *Cauchy* model or not. To this end, we assumed that the *a-priori* term $f_{\Delta d}(\Delta\gamma(\boldsymbol{p}_j))$ in (4) is a uniform model. With this condition, the *MAP* estimator becomes a *ML* estimator and can be represented as[10]

$$d(\boldsymbol{p}_j) = \min_{\gamma} \left\{\delta(\boldsymbol{p}_j) \cdot c_o + (1 - \delta(\boldsymbol{p}_j)) \cdot c_m\right\} \quad , \tag{8}$$

where c_o is an occlusion cost, c_m is a matching cost and $\delta(\boldsymbol{p}_j)$ is an indicator variable that defined as

$$\delta(\boldsymbol{p}) = \begin{cases} 1, & if \ \ \boldsymbol{p} \in occlusion \\ 0, & if \ \ \boldsymbol{p} \notin occlusion \end{cases} .$$

The formulation of c_o and c_m depends on the statistical model used. Let $B(\boldsymbol{p}_j)$ be the block centered at the pixel position $\boldsymbol{p}_j$ in the image. N_B is the total pixel number within the block $B(\boldsymbol{p}_j)$. f_o is a constant describing the probability of pixel position $\boldsymbol{p}_j$ in the occluded area. Then, in the case of the *Laplacian* model,

$$c_o = \ln\left(1/(f_o \cdot \sqrt{2} \cdot \sigma_w)\right) \tag{9a}$$

and

$$c_m = \frac{\frac{1}{N_B} \cdot \sum_{\boldsymbol{p}_m \in B(\boldsymbol{p}_j)} \left| s_l(\boldsymbol{p}_m) - s_r(\boldsymbol{p}_m - d(\boldsymbol{p}_j)) \right|}{\sigma_w / \sqrt{2}} . \tag{9b}$$

In the case of the *Cauchy* model,

$$c_o = \ln \left(a / (f_o \pi) \right) \tag{10a}$$

and

$$c_m = \frac{\sum_{\boldsymbol{p}_m \in B(\boldsymbol{p}_j)} \ln\left(\left(s_l(\boldsymbol{p}_m) - s_r(\boldsymbol{p}_m - d(\boldsymbol{p}_j)) \right)^2 + a^2 \right)}{N_B} . \tag{10b}$$

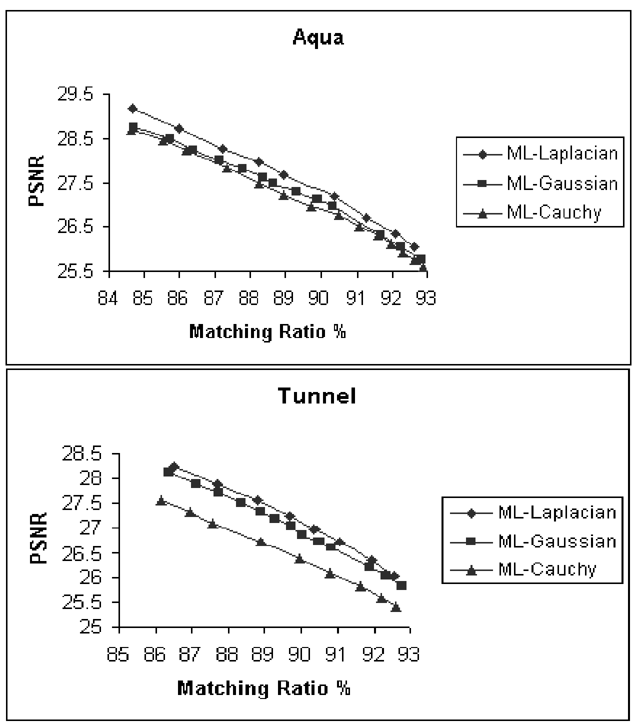

Fig. 1. Comparison of different *likelihood* terms, with either a *Gaussian*, *Laplacain* or *Cauchy* model using stereo image pair.

As criteria for comparison, the *PSNR* of disparity compensation errors and the number of corresponding matches are taken into consideration[10]. Fig. 1 shows the comparison results with the natural stereo image pairs *Auqa* and *Tunnel*. For comparison, the result with a *Gaussian* model is also shown in the figure. It can be seen from Fig. 1 that the *ML* disparity estimation based on the *Laplacian* model (*ML-Laplacian*) outperforms the *ML* estimation based on the

Gaussian (*ML-Gaussian*) or the *Cauchy* model (*ML-Cauchy*) at the same matching ratio. The *ML-Laplacian* has a higher *PSNR* value compared with the *ML-Cauchy* and the *ML-Gaussian*.

This result is different from the conclusion made by Sebe[9]. However, the findings were for pixel-wise disparity estimation that showed that the *ML-Cauchy* provides the best results for disparity estimation. Our results can be explained as follows. For the block-wise *ML-Cauchy* estimator, the matching cost c_m is defined as the sum of the logarithms of intensity differences (see (10b)). The logarithmic function reduces the relative contribution of large intensity differences to the sum. Large intensity differences are usually an indicator of a mismatch between two blocks. Using (10b), the ability to distinguish two different blocks matched to the reference block is reduced compared with using (9b). Therefore, the block-wise *ML-Cauchy* estimator is less effective than the block-wise *ML-Laplacian* estimator.

3.2. Statistical model for the *a-priori* term

To examine statistical characteristics of disparity differences $\Delta d(\boldsymbol{p}_j)$, the disparity maps obtained with the *ML-Laplacian* approaches are exploited. At first, difference Δd of the disparity value at each pixel position $\boldsymbol{p}_j$ is calculated as follows:

$$\Delta d(\boldsymbol{p}_j)= d(\boldsymbol{p}_j)-d(\boldsymbol{p}_{j-1}). \tag{11}$$

Different theoretical models are then used to model the measured statistical characteristics of the disparity difference values. The model error determines which model is the best for representing the data.

Fig. 2 shows the measured data and the approximated theoretical models of disparity differences Δd. The measured data is obtained from the disparity maps estimated by the *ML-Laplacian* approach using natural stereo images *Aqua* and *Tunnel*. The theoretical models, a *Cauchy*, a *Laplacian* and a *Gaussian* model, are then used to approximate this measured data, respectively. It can be seen that the *Cauchy* model represents the measured data more accurately than the *Laplacian* or the *Gaussian* model.

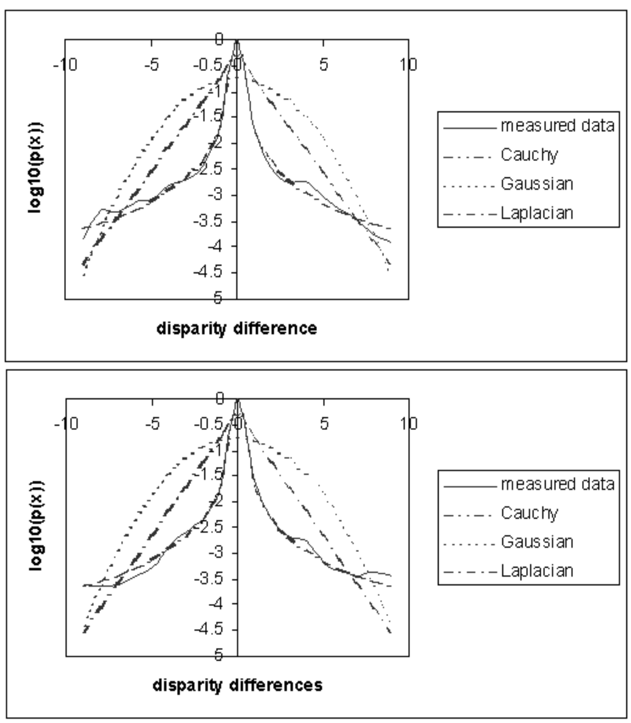

Fig.2. Model of disparity differences. The top shows the results for the stereo image pair is *Aqua*, and the bottom for *Tunnel*.

4. IMPLEMENTATION

According to the investigation in the previous section, the block-wise *MAP* disparity estimation is implemented using a *Laplacian* model for the *likelihood* term and a *Cauchy* model for the *a-priori* term. After some mathematical deduction, we have the *MAP* estimator

$$d = arg \min_{\gamma} \{c_o \delta(s_l(\boldsymbol{p})) + c_m (1 - \delta(s_l(\boldsymbol{p}))) + c_{\Delta \boldsymbol{d}} (\Delta d)\}. \tag{12}$$

where

$$c_{\Delta \boldsymbol{d}} (\Delta d) = \ln \left(1 + \left(\frac{|\Delta d|}{a} \right)^2 \right),$$

c_o and c_m are defined in (9) .

To minimize the cost presented in (12), the dynamic programming technique is used, where only the costs of three possible nearest neighboring candidate pairs with a maximum disparity difference of 1 pixel are taken into account[11]. In this case, $c_{\Delta \boldsymbol{d}}(0) = 0$ and $c_{\Delta \boldsymbol{d}}(1) = \ln\left(1 + 1/a^2\right)$. In this implementation model values of disparity difference larger than one pixel are not taken into consideration.

5. EXPERIMENTAL RESULTS

The proposed algorithm for *MAP* disparity estimation was evaluated based on the accuracy of the disparity estimation and its performance in intermediate view reconstruction. To evaluate the effectiveness of statistical models used for block-wise *MAP* disparity estimation, natural stereo image pairs *Aqua* and *Tunnel* are used and the peak signal-to-noise ratio (*PSNR*) of disparity-compensation errors within matching regions are measured. The matching ratio r is defined as a ratio of number of corresponding matches over total image pixel number. Fig.3 shows the performance of *MAP* disparity estimation with different values of $c_{\Delta d}(1)$ compared to block-wise *maximum likelihood* disparity estimation based on the *Laplacian* model. The value of $c_{\Delta \boldsymbol{d}}(1)$ is equal to 1.1 for *MAP_Cauchy_1*, 1.3 for *MAP_Cauchy_2* and 0.8 for *MAP_Cauchy_3*. From Fig. 3, it can be seen that the proposed *MAP* disparity estimation, on average, increases the *PSNR* by around 0.2 dB compared with *ML_Laplacian* for the same number of corresponding matches. This means that the smoothness constraints can really improve the performance of the disparity estimation. The optimal value of $c_{\Delta \boldsymbol{d}}(1)$ for these two image pairs is in the interval of 0.8 and 1.3.

The proposed *MAP* disparity estimation was further incorporated into an algorithm for intermediate view reconstruction from stereoscopic images. Based on the framework of the reconstruction algorithm[12], we compared the image quality of the reconstructed intermediate views with different disparity algorithms. As a reference[12], *maximum likelihood* disparity estimation with a *Laplacian* model was exploited and the reconstruction algorithm was referred to as *ML-REC*. The reconstruction algorithm with the proposed *MAP* disparity estimation was referred to as *MAP-REC*. For the *ML-REC* and *MAP-REC* algorithms, they were only different in how they modeled the *a-priori* term.

For objective comparison, the test image sequence “*Flower Garden*” was used. “*Flower Garden*” can be considered a "stereoscopic" image sequence due to multiple views of a static scene resulting from camera translation. This provides “ground truth” intermediate views for numerical performance evaluation using a *PSNR* metric. For testing, we chose the frames with the series numbers of 1, 4, 7, …, 148 from the original sequence as the right-eye images, the frames with the series numbers of 2, 5, 8, …, 149 as the center images, and the frames with the series numbers of 3, 6, 9, …, 150 as the left-eye images. Fig. 4 shows the first frame of the left-eye, the center and the right-eye image sequence. The intermediate views at position α=0.5 were reconstructed from this stereoscopic image sequence and compared to the “ground truth” centre images. Fig. 5 shows the comparison results, in terms of *PSNR* of the error signals between the reconstructed view and the “ground truth” image at position α=0.5 for each frames. Table 1 shows the average, the minimum and the maximum *PSNR* values of the total 50 reconstructed intermediate views. It can be seen that the *MAP-REC* algorithm, on average, achieved a *PSNR* gain of 0.1 dB compared with the *ML-REC* algorithm.

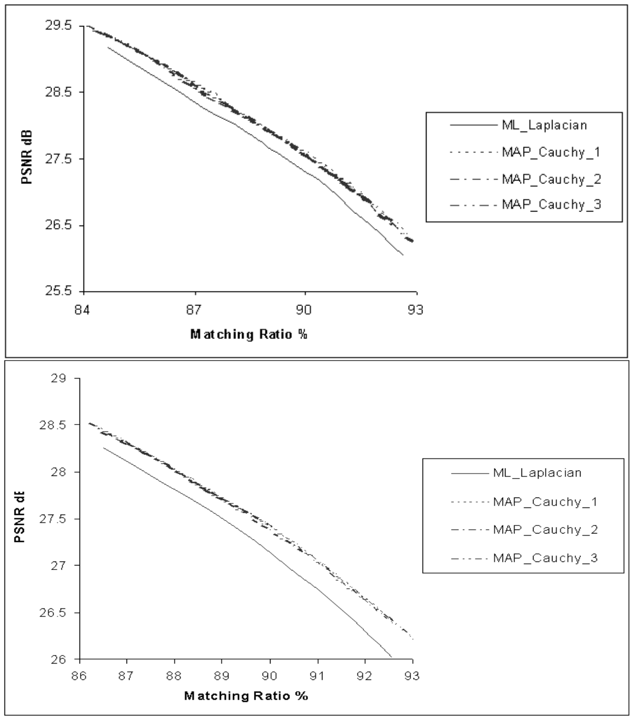

Fig. 3. Performances of *MAP* disparity estimation. The top shows the results for the stereo image pair *Aqua*, and the bottom for *Tunnel*.

Table 1. *PSNR* results for a total of 50 intermediate views reconstructed at position α=0.5.

	ML-REC (dB)	*MAP-REC* (dB)	*PSNR gain*
average *PSNR*	30.4782	30.5828	+0.1046
min. *PSNR*	24.8239	25.1025	+0.2786
max. *PSNR*	32.799	32.826	+0.025

6. CONCLUSIONS

In this paper, statistical characteristics of data related to the *likelihood* term and the *a-priori* term for block-wise *MAP* disparity estimation was studied using natural stereo images. The study, which included an experimental comparison, showed that a *Laplacian* model was the best model compared with the *Cauchy* and *Gaussian* models for modeling the *likelihood* terms. For modeling the *a-priori* term, a *Cauchy* model was the best model compared to a *Laplacian* and a *Gaussian* model.

The proposed MAP disparity estimation was also incorporated into an algorithm for intermediate view reconstruction. The experimental results obtained with the natural video "*Flower Garden*" showed that the *MAP* disparity estimation using the determined models improved the image quality of reconstructed intermediate views compared to a reference reconstruction algorithm *REC-ML*.

Fig. 4. Example of stereoscopic images and the center image made from the image sequence *Flower Garden*. From the top to the bottom are the left-eye image, the center image and the right-eye image.

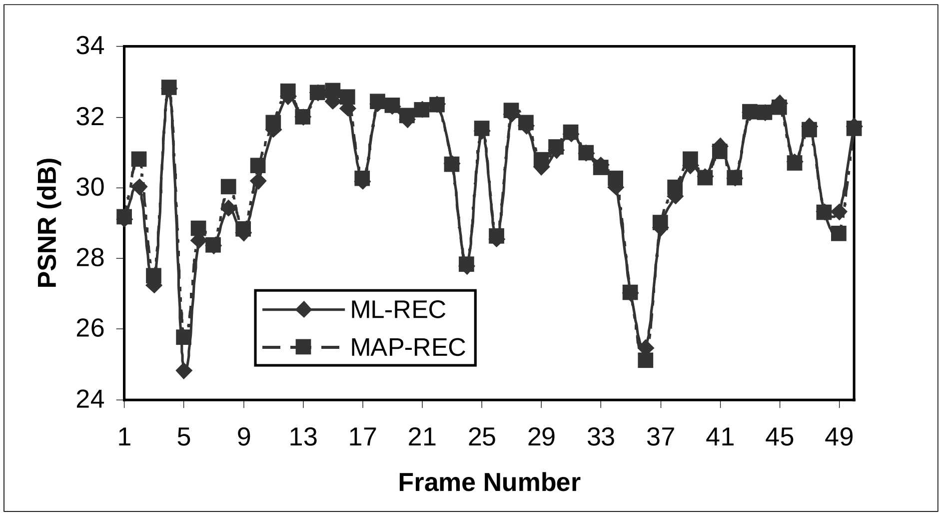

Fig. 5. *PSNR* comparison of the reconstructed intermediate views using the *ML-REC* and *MAP-REC* algorithms *versus* frame number.

REFERENCES

1. ISO/IEC JTC1/SC29/WG11 N4982, "Applications and requirements for 3DAV", Klagenfurt, Austria, July 2002.
2. W.J. Tam and L. Stelmach, "Stereo depth perception in a sample of young television viewers", Int. Workshop on Stereoscopic and 3D Imaging, Santorini, Greece, Sept. 1995, pp. 149-156.
3. J. Konrad, "Enhancement of viewer comfort in stereoscopic viewing: parallax adjustment", *SPIE Stereoscopic Displays Virtual Reality Systems*, Vol. 3639, January 1999, pp. 179-190.
4. A. Redert, E. Hendriks, J. Biemond, "Correspondence estimation in image pairs", *IEEE Signal Processing Magazine*, May 1999, pp. 29-46.
5. P. N. Belhumeur, "A bayesian approach to binocular stereopsis", *International Journal of Computer Vision*, Vol. 19, No. 3, 1996, pp. 237-262.
6. I. Cox, S. Hingoraini, S. Rao, "A maximum likelihood stereo algorithm", Computer Vision and Image Understanding, Vol. 63, No. 3, May 1996, pp. 542-567.
7. L. Falkenhagen, "Blockbasierte Disparitaetsschaetzung unter Beruecksichtigung statistischer Abhaengigkeiten der Disparitaeten", Dissertation, University of Hanniover, VDI Verlag, ISBN 3-18-365710-4, 2001.
8. A. Mansouri and J. Konrad, "Bayesian winner-take-all reconstruction of intermediate Views from stereoscopic images", IEEE Transactions on Image Processing, Vol. 9, No. 10, October 2000, pp.1710-1722.
9. N. Sebe, M. S. Lew and D. P. Huijsmans, "Toward improved Ranking Metrics", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 10, October, 2000, pp. 1132-1143.
10. L. Zhang, "Statistical model for intensity differences of corresponding points between stereo image pairs", *IEEE International Conference on Multimedia and Expo* (ICME) Baltimore, MD, July 6-9, 2003, Vol. I, pp. 365-368.
11. L. Zhang, "Hierarchical block-based disparity estimation using mean absolute difference and dynamic programming", *International Workshop on VLBV*, Athens, Greece, 11-12 October 2001.
12. L. Zhang, D. Wang, A. Vincent, "Reliability measure of disparity estimates for intermediate view reconstruction", *IEEE International Conference on Image Processing*, Vol. III, Rochester, NY, Sept. 22-25, 2003, pp.837-840.