Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 53691, Pages 1–6
DOI 10.1155/ASP/2006/53691

Global Motion Model for Stereovision-Based Motion Analysis

Jia Wang,[1] Zhencheng Hu,[2] Keiichi Uchimura,[2] and Hanqing Lu[1]

[1] *National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun East Road, Beijing 100080, China*
[2] *Department of Computer Science, Faculty of Engineering, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan*

Received 26 October 2005; Revised 11 January 2006; Accepted 21 January 2006

Recommended for Publication by Dimitrios Tzovaras

An advantage of stereovision-based motion analysis is that the depth information is available, thus motion can be estimated more precisely in 2.5D stereo coordinate system (SCS) constructed by the depth and the image coordinates. In this paper, *stereo global motion* in SCS, which is induced by 3D camera motion in real-world coordinate system (WCS), is parameterized by a five-parameter global motion model (GMM). Based on such model, global motion can be estimated and identified directly in SCS without knowing the physical parameters about camera motion and camera setup in WCS. The reconstructed global motion field accords with the spatial structure of the scene much better. Experiments on both synthetic data and real-world images illustrate its promising performance.

1. INTRODUCTION

The advantage of stereovision-based motion analysis is that the depth/disparity information can be computed. Considering the depth information together with the image coordinates, motion can be analyzed more precisely in a 2.5D space rather than the traditional 2D image plane. This paper, by expressing the 2.5D space as stereo coordinate system (SCS), addresses the problem of global motion modeling in SCS.

Global motion model (GMM) is commonly used to describe the effect of camera motion (global motion) acting on video image. By GMM, global motion can be distinguished from image motion induced by moving objects (local motion), thus moving objects can be extracted from the image.

In the literature, single-camera-based GMM approaches [1–3], which analyze the camera motion based on 2D image-space shifts [1], cannot describe the global motion accurately when the depth of field is great. By using stereovision, global motion can be estimated more precisely from 2.5D stereo-motion analysis using the depth and image coordinates. The reconstructed global motion field will accord with the spatial structure of the scene much better, which makes moving object's detection much easier.

In this paper, a five-parameter stereo GMM is proposed to parameterize global motion in SCS based on the analysis of 3D camera motion. Different from the previous works aiming to recover the physical parameters of camera motion in real-world coordinate system (WCS) [4–8], the presented model pays more attention to the fast distinguishing of global motion and local motion directly from stereo data. Thus instead of estimating the real camera motion in WCS, global motion is estimated and identified directly in SCS without knowing the physical camera parameters. The proposed model is provided as a tool for stereo-motion analysis where disparity can be fast calculated. It is very useful for many stereovision-based real-time applications, such as surveillance, robot vision, especially of our research on stereovision-based adaptive cruise control (ACC) systems [9].

2. STEREO GLOBAL MOTION MODEL

A typical stereovision system consists of two coplanar cameras with the same intrinsic parameters [9]. By projecting a point in WCS (x, y, z) to the stereo left/right ICSs (u, v), the following equation is held:

$$u_{l,r} = \frac{f(x \pm b/2)}{z} \qquad v_{l,r} = \frac{fy}{z}, \tag{1}$$

where f is camera focal and b is baseline distance between the cameras. Then, disparity Δ can be achieved by

$$\Delta = u_l - u_r = \frac{fb}{z}. \tag{2}$$

Only considering the left ICS, the relationship between WCS and SCS (u, v, Δ) can be described by

$$u = \frac{f(x+b/2)}{z}, \quad v = \frac{fy}{z}, \quad \Delta = \frac{fb}{z}, \iff \begin{cases} x = \dfrac{ub}{\Delta} - \dfrac{b}{2}, \\ y = \dfrac{vb}{\Delta}, \\ z = \dfrac{fb}{\Delta}. \end{cases} \tag{3}$$

Note that in stereovision, the intrinsic parameters f and b always remain unchanged. Global motion with respect to WCS can be generally referred, as a composition of rotations about x-, y-, z-axes followed by translations along them [10]. To express the stereo GMM in SCS, this paper analyzed the rotation and translation separately based on the relationship between WCS and SCS.

In WCS, "rotation about x-axis" with angle α can be described by

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{4}$$

Mapping to SCS based on (3), there is

$$u' = \frac{u}{\cos\alpha - \sin\alpha \cdot (v/f)}, \quad v' = \frac{\cos\alpha \cdot v + f\sin\alpha}{\cos\alpha - \sin\alpha \cdot (v/f)}, \quad \Delta' = \frac{\Delta}{\cos\alpha - \sin\alpha \cdot (v/f)}. \tag{5}$$

In order to deduce a simple expression, we assume that the rotation angle α is small. Since v/f is generally smaller than 1, when α is small, we approximate that

$$\cos\alpha \approx 1, \qquad \frac{v\sin\alpha}{f} \approx 0. \tag{6}$$

Then (5) can be simplified as

$$u' \approx u, \qquad v' \approx v + f\sin\alpha, \qquad \Delta' \approx \Delta, \tag{7}$$

which results in a Δ-independent global displacement of v. Note that such simplification is also permitted by the variety of depths that are being reconstructed.

Similarly, "rotation about y-axis" with angle β is described by

$$u' = \frac{\cos\beta \cdot u - f\sin\beta}{\cos\beta + \sin\beta \cdot (u-\Delta/2)/f} \overset{\beta\to 0}{\approx} u - f\sin\beta, \quad v' = \frac{v}{\cos\beta + \sin\beta \cdot (u-\Delta/2)/f} \overset{\beta\to 0}{\approx} v, \quad \Delta' = \frac{\Delta}{\cos\beta + \sin\beta \cdot (u-\Delta/2)/f} \overset{\beta\to 0}{\approx} \Delta, \tag{8}$$

which results in a Δ-independent global displacement of u.

"Rotation about z-axis" with angle γ can be described by

$$u' = \left(u - \frac{\Delta}{2}\right)\cos\gamma + v\sin\gamma + \frac{\Delta}{2}, \quad v' = -\left(u - \frac{\Delta}{2}\right)\sin\gamma + v\cos\gamma, \qquad \Delta' = \Delta. \tag{9}$$

Because "rotation about z-axis" occurs less frequently than the other motions [2], it will not be considered by the model presented in this paper.

Translations within WCS and their mappings in SCS can be described by

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \Longrightarrow \begin{cases} u' = \dfrac{u+\Delta t_x/b}{1+\Delta t_z/fb}, \\ v' = \dfrac{v+\Delta t_y/b}{1+\Delta t_z/fb}, \\ \Delta' = \dfrac{\Delta}{1+\Delta t_z/fb}, \end{cases} \tag{10}$$

which demonstrate that translation along x/y-axis will bring a Δ-dependent displacement to u/v, respectively, yet translation along z-axis will change the scale of u, v, and Δ simultaneously.

Based on the above analysis, camera's rotation about x-, y-axes and translation along x-, y-, z-axes are parameterized into such a stereo GMM as

$$u' = \frac{u + R_Y + T_X\Delta}{1 + T_Z\Delta}, \quad v' = \frac{v + R_X + T_Y\Delta}{1 + T_Z\Delta}, \qquad \Delta' = \frac{\Delta}{1 + T_Z\Delta}, \tag{11}$$

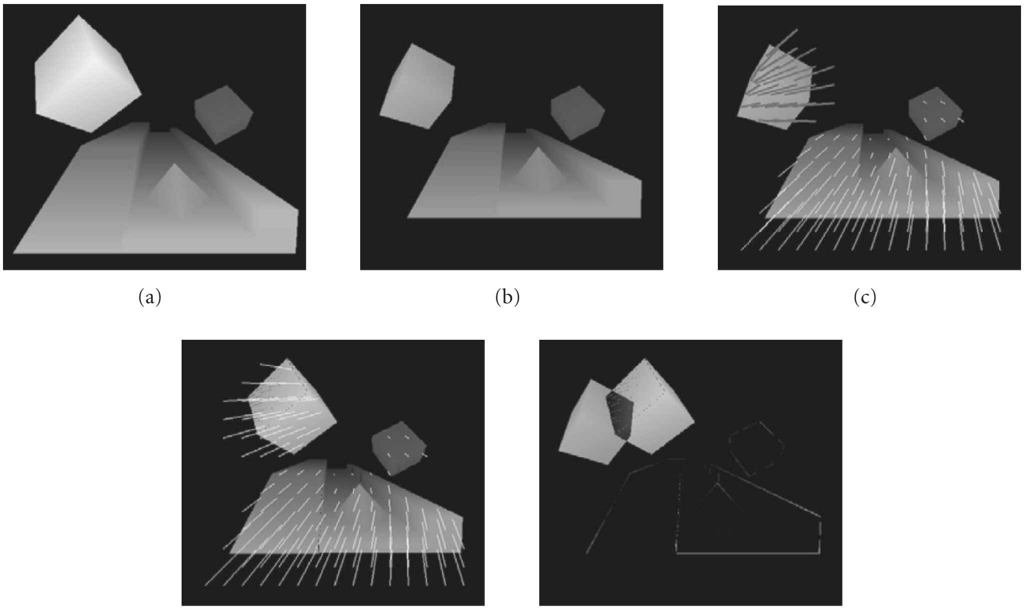
where $(R_X, R_Y, T_X, T_Y, T_Z)$ are introduced to describe the rotations and translations by letting $R_X = f\sin\alpha$, $R_Y = -f\sin\beta$, $T_X = t_x/b$, $T_Y = t_y/b$, $T_Z = t_z/fb$.

3. PARAMETER ESTIMATION

Corresponding pixel pairs (measured by corner matching or block matching) between successive frames are used to estimate the five parameters. Assuming there are N pairs, each pair consists of a pixel $k = 1, \dots, N$ with SCS coordinate (u_k, v_k, Δ_k) in the first frame, and its counterpoint (u'_k, v'_k, Δ'_k) in the next frame. The five parameters $(R_X, R_Y, T_X, T_Y, T_Z)$ are estimated by a least-square method following two steps.

Step 1. Estimating T_Z. Based on the third subformula of (11), T_Z is first estimated by minimizing the following least-square criterion:

$$\begin{aligned} T_Z &= \arg\min_{T_Z} \sum_{k=1}^{N} \left[\Delta'_k - \frac{1}{1 + T_Z\Delta_k}\Delta_k\right]^2 \\ &= \arg\min_{T_Z} \sum_{k=1}^{N} \left[\Delta'_k + T_Z\Delta'_k\Delta_k - \Delta_k\right]^2. \end{aligned} \tag{12}$$


(a) (b) (c)

(d) (e)

Figure 1: Synthetic data involving global motion and local motion. (a) and (b) are the succesive frames where the image intensity denotes disparity. (c) gives the 2D motion field between (a) and (b). (d) is the camera motion compensated image of (a) based on the estimated parameters $(R_X, R_Y, T_X, T_Y, T_Z) = (6.76, -6.83, 31.17, -28.62, 0.248103)$. (e) shows the difference between (d) and the actual image (b).

Differentiating (12) with respect to T_Z and setting the derivative to zero, T_Z is achieved by

$$T_Z = \frac{\left\{ \sum_{k=1}^{N} \left[\Delta'_k (\Delta_k)^2 \right] - \sum_{k=1}^{N} \left[\Delta_k (\Delta'_k)^2 \right] \right\}}{\sum_{k=1}^{N} \left[(\Delta'_k \Delta_k)^2 \right]}. \tag{13}$$

Three sums are needed in (13), which can be computed by summing Δ of all the N pixel pairs. In addition, in order to avoid the influence of local motion, the above estimating procedure is performed iteratively, which is similar to the method in [2]. In each iteration, every pixel pair is evaluated based on the computed T_Z by comparing the original Δ' with the computed Δ' using (11). If the difference exceeds a predefined threshold, corresponding pixel pair will be referred to as local motion pairs and will be discarded. (In our experiments, the threshold is an experiential value which is selected as 0.1.) Then the remaining pairs are used to reestimate T_Z. Using such method, the influence of pixel pairs that do not follow global motion will be removed gradually and the convergence of T_Z will occur after a very few iterations. (Generally, the convergence will occur with 4 iterations.)

Step 2. Estimating (T_X, T_Y, R_X, R_Y). Based on T_Z, we introduced an auxiliary variable $Z_k = 1 + \Delta_k T_z$. Then similar to the solving of T_Z, the following criteria are minimized:

$$\begin{aligned} (T_X, R_Y) &= \arg\min_{T_X, R_Y} \sum_{k=1}^{N} \left[Z_k u'_k - u_k - R_Y - T_X \Delta_K \right]^2, \\ (T_Y, R_X) &= \arg\min_{T_Y, R_X} \sum_{k=1}^{N} \left[Z_k v'_k - v_k - R_X - T_Y \Delta_k \right]^2, \end{aligned} \tag{14}$$

and (T_X, T_Y, R_X, R_Y) are achieved by

$$\begin{aligned} T_X &= \frac{N(\sum Zu'\Delta - \sum u\Delta) + (\sum u - \sum Zu') \sum \Delta}{N \sum (\Delta)^2 - (\sum \Delta)^2}, \\ R_Y &= \frac{(\sum Zu' - \sum u) \sum (\Delta)^2 + (\sum u\Delta - \sum Zu'\Delta) \sum \Delta}{N \sum (\Delta)^2 - (\sum \Delta)^2}, \\ T_Y &= \frac{N(\sum Zv'\Delta - \sum v\Delta) + (\sum v - \sum Zv') \sum \Delta}{N \sum (\Delta)^2 - (\sum \Delta)^2}, \\ R_X &= \frac{(\sum Zv' - \sum v) \sum (\Delta)^2 + (\sum v\Delta - \sum Zv'\Delta) \sum \Delta}{N \sum (\Delta)^2 - (\sum \Delta)^2}, \end{aligned} \tag{15}$$

where the subscript k is omitted for simplification. Note that the estimation of (T_X, T_Y, R_X, R_Y) also follows an iterative scheme, aiming to eliminate the influence of local motion.

4. SIMULATION RESULTS

The proposed GMM has been tested on both synthetic data, in which we know the camera motion parameters and corresponding pixel pairs exactly, and a variety of real-world images with unknown camera motion parameters. For the real-world images, we use corner pairs, which is a good feature to be tracked in image sequences [11], to estimate the global motion parameters.

4.1. *Simulation*

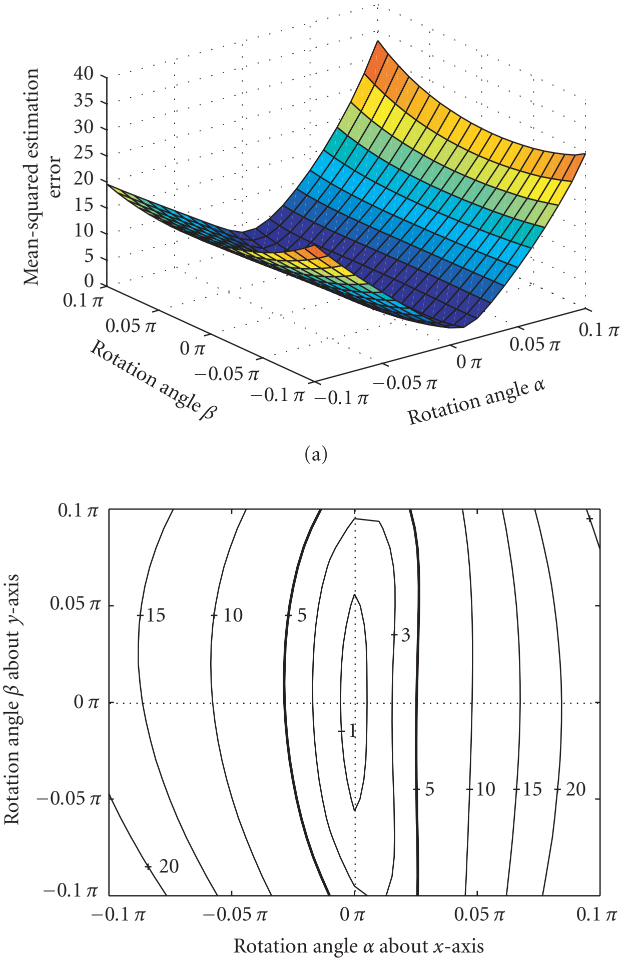
Figure 1 shows the synthetic data where image intensity denotes disparity. Between Figures 1(a) and 1(b), the camera ($f = 200, b = 100$) undergoes a rotation $(\alpha, \beta) = (0.01\pi, 0.01\pi)$ and a translation $(t_x, t_y, t_z) = (3000, -3000, 5000)$, while the closest cube undergoes an isolated motion. Figure 1(c) illustrates their correspondence relationship in format of 2D motion vectors, where the local motion is marked by gray vectors. Based on the physical parameters of camera motion, the reference GMM parameters are $(R_X, R_Y, T_X, T_Y, T_Z) = (6.28, -6.28, 30, -30, 0.250000)$. Using the two-step iterative estimator, the estimated parameters come to be $(6.76, -6.83, 31.17, -28.62, 0.228103)$. In order to measure the accuracy of the model and the parameters, Figure 1(d) shows the predicted image of Figure 1(a) after camera motion compensation based on the estimated parameters. A 2D global motion field is also constructed by recomputing the 2D motion vectors using such parameters. The predicted image is compared with the actual image (see Figure 1(b)), and their difference is shown in Figure 1(e).

Accuracy analysis

To analyze the accuracy of the stereo GMM quantitatively, we defined a mean-squared estimation error (MSEE) as follows:

$$\text{MSEE} = \frac{1}{N}\sum_{k=1}^{N}\left[\left(u_k^* - u_k'\right)^2 + \left(v_k^* - v_k'\right)^2 + \left(\Delta_k^* - \Delta_k'\right)^2\right], \tag{16}$$

where N is the number of the corresponding pairs, $(u_k^*, v_k^*, \Delta_k^*)$ is the actual SCS coordinate after camera motion, and (u_k', v_k', Δ_k') is the estimated SCS coordinate. Based on the synthetic data in Figure 1, Figure 1 shows the calculated MSEE versus physical parameters of camera rotation (α, β). It can be seen that when the rotation angles are small, the MSEE is small (MSEE ≤ 5 corresponding to regions surrounded by thick boundaries in Figure 2(b)) and the stereo GMM can work well. Comparing the two kinds of rotations, MSEE is more sensitive to rotation angle α about the x-axis. This comes from the fact that many points with large v/f ($v/f \approx 1$) are existing at the bottom of the image (as shown in Figure 1(a), where the image size is 400 × 400, and $f = 200$), thus the approximation of $v \sin \alpha / f \approx 0$ in formula (6) becomes false when α increases. For translations, the presented GMM can cope with large translations along x-, y-, and z-axes. We have tested the GMM by $-10000 \leq t_X, t_Y, t_Z \leq 10000$, and the MSEE remains within 10^{-6}.



(a)

(b)

Figure 2: MSEE versus rotation angle about x-, y-axis (α, β): (a) 3D map; (b) contour map.

4.2. *Real-world images*

Experiment 1

The testing image is taken from *flower garden* sequence, which involves an apparent camera motion of translation along x-axis. Corresponding disparity image is taken from [12] as shown in Figure 3(b). Note that disparity Δ can be computed by various methods with different resolutions, it is normalized into $0 \leq \Delta \leq 1$ before estimating $(T_X, T_Y, T_Z, R_X, R_Y)$. Figure 3(a) shows the detected corners and their 2D motion vectors pointing to the counterpoints in the next frame. Based on such corner pairs, the iterative scheme computes the parameters as $T_x = -12.434820$, $T_y = 1.723545$, $T_z = 0.004598$, $R_x = -0.117074$, $R_y = -0.099431$, where the most apparent global motion is parameterized by T_x. Figure 3(b) shows the reconstructed 2D global motion field by recomputing the 2D UV motion vectors of the pixels using the above parameters. Comparing with such field, corners which match with the field are marked by white color in

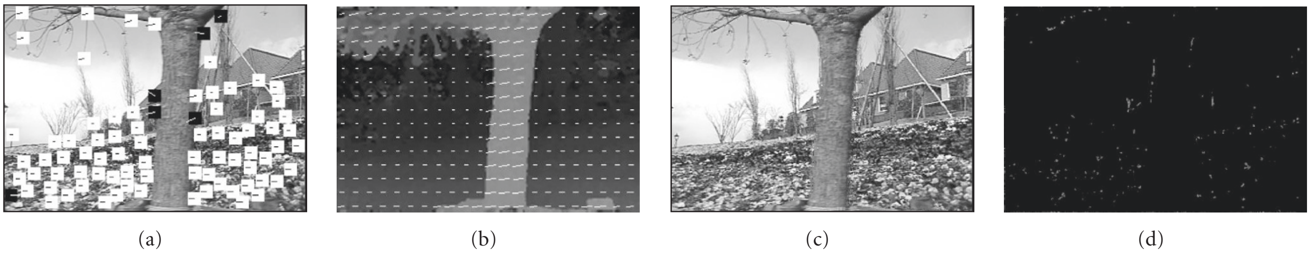


FIGURE 3: Experimental results on *flower garden* sequence: (a) original image with the detected corners and their motion vectors; (b) disparity image and reconstructed global motion field using the estimated parameters ($T_x = -12.434820$, $T_y = 1.723545$, $T_z = 0.004598$, $R_x = -0.117074$, $R_y = -0.099431$); (c) camera motion compensated image; (d) difference image between the compensated image and the original image.

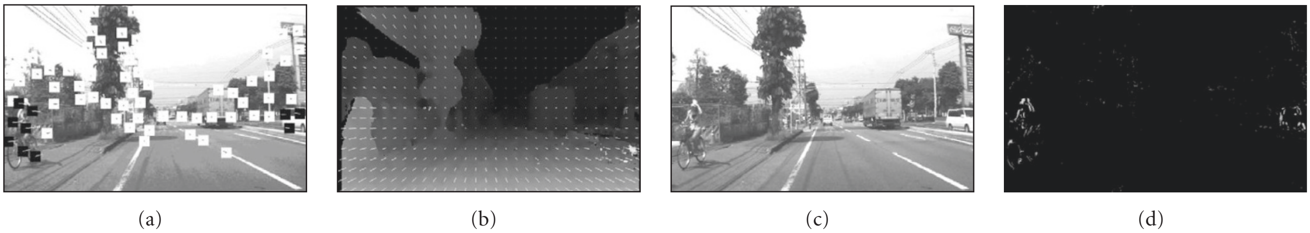


FIGURE 4: Experimental results on a traffic scene: (a) original image with the detected corners and their motion vectors; (b) disparity image and reconstructed global motion field using the estimated parameters $(R_X, R_Y, T_X, T_Y, T_Z) = (0.05, 0.03, 0.05, 0.71, -0.067826)$; (c) camera motion compensated image; (d) difference image between the compensated image and the original image.

Figure 3(a), those which do not match are marked by black color. To our experience, corners which do not follow the global motion are either belonging to the moving objects or the overlapped regions. Figure 3(c) shows the predicted image of Figure 3(a) by camera motion compensation. To reduce edge distortions, bilinear interpolation has been applied for image compensation by using the image intensities of the four nearest neighboring pixels. The difference image between Figure 3(c) and the actual image Figure 3(a) is shown in Figure 3(d).

Experiment 2

The stereo GMM is also applied to our own stereo sequence of traffic scene (Figure 4), which is obtained by a binocular system [9] mounted on a moving vehicle. Disparity image corresponding to Figure 4(a) is shown in Figure 4(b). For estimating the global motion parameters, Figure 4(a) shows the detected corners and their 2D motion vectors pointing to the counterpoints in the previous frame. Based on such corner pairs, the two step estimator computes the parameters as $(R_X, R_Y, T_X, T_Y, T_Z) = (0.05, 0.03, 0.05, 0.71, -0.067826)$. Then the reconstructed 2D global motion field is given in Figure 4(b). Comparing with such field, corners which match with the field are marked by white color in Figure 4(a), those that do not match are marked by black color. From Figure 4(a), it can be seen that corners belonging to the major moving objects are successfully detected, while those belonging to the slightly moving objects are confused with the background noise. Figure 4(c) shows the predicted image of Figure 4(a) by camera motion compensation. To reduce edge distortions, bilinear interpolation has been applied for image compensation by using the image intensities of the four nearest neighboring pixels. The difference image between Figure 4(c) and the actual image Figure 4(a) is shown in Figure 4(d).

5. CONCLUSION

Experimental results demonstrate that the proposed stereo GMM works well for stereovision-based camera motion analysis and the motion parameters can be efficiently estimated by the two-step iterative estimator. Based on the presented model, global motion can be estimated more precisely according with the spatial structure of the scene, which makes further motion analysis, such as real moving object's detection, much easier. In addition, the computational simplicity of the presented method also makes it suitable for real-time applications such as ACC systems.

ACKNOWLEDGMENT

This research is sponsored by the National Natural Science Foundation of China (Grants no. 60135020, 60475010, and 60121302).

REFERENCES

[1] T. S. Huang and R. Y. Tsai, "Three-dimensional motion estimation from image-space shifts," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '81)*, vol. 6, pp. 1136–1139, Atlanta, Ga, USA, March-April 1981.

[2] G. B. Rath and A. Makur, "Iterative least squares and compression based estimations for a four-parameter linear global motion model and global motion compensation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 7, pp. 1075–1099, 1999.

[3] G. Qian, R. Chellappa, and Q. Zheng, "Robust Bayesian cameras motion estimation using random sampling," in *International Conference on Image Processing (ICIP '04)*, vol. 2, pp. 1361–1364, Singapore, Republic of Singapore, October 2004.

[4] T. S. Huang and S. D. Blostein, "Robust algorithm for motion estimation based on two sequential stereo image pairs," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '85)*, pp. 518–523, San Francisco, Calif, USA, June 1985.

[5] G.-S. J. Young and R. Chellappa, "3-D motion estimation using a sequence of noisy stereo images: models, estimation, and uniqueness results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 8, pp. 735–759, 1990.

[6] J. Shieh, H. Zhuang, and R. Sudhakar, "Motion esimtation from a sequence of stereo images: a direct method," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 24, no. 7, pp. 1044–1053, 1994.

[7] H. Hirschmuller, P. R. Innocent, and J. M. Garibaldi, "Fast, unconstrained camera motion estimation from stereo without tracking and robust statistics," in *The 7th International Conference on Control, Automation, Robotics and Vision*, vol. 2, pp. 1099–1104, Singapore, Republic of Singapore, December 2002.

[8] Z. Xiang and Y. Genc, "Bootstrapped real-time ego motion estimation and scene modeling," in *The 5th International Conference on 3-D Digital Imaging and Modeling (3DIM '05)*, pp. 514–521, Ottawa, Ontario, Canada, June 2005.

[9] Z. Hu and K. Uchimura, "U-V-disparity: an efficient algorithm for stereovision based scene analysis," in *IEEE Intelligent Vehicle Symposium (IV '05)*, Las Vegas, Nev, USA, June 2005.

[10] A. M. Tekalp, *Digital Video Processing*, Prentice Hall, Beijing, China, 1998.

[11] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pp. 593–600, Seattle, Wash, USA, June 1994.

[12] S. B. Kang, R. Szeliski, and J. Chai, "Handling occlusions in dense multi-view stereo," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 103–110, Kauai, Hawaii, USA, December 2001.

Jia Wang received his B.E. degree and M.S. degree from Huazhong University of Science and Technology, China, in 1999 and 2002, respectively. He is currently a Ph.D. Candidate in National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China. His research interests include video motion analysis, image segmentation, and machine vision applications in ITS.

Zhencheng Hu received his B.E. degree from Shanghai Jiao Tong University, China in 1992, and his M.E. degree from Kumamoto University, Japan, in 1998. He received his Ph.D. degree in system science from Kumamoto University, Japan, in 2001. He is currently an Associate Professor with the Department of Computer Science, Kumamoto University, Japan. His research interests include camera motion analysis, augmented reality, and Machine vision applications in industry and ITS. He is a Member of IEEE, and the Institute of Electronics and Information Communication Engineers of Japan (IEICE).

Keiichi Uchimura received the B.E. and M.E. degrees from Kumamoto University, Kumamoto, Japan, in 1975 and 1977, respectively, and the Ph.D. degree from Tohoku University, Miyagi, Japan, in 1987. He is currently a Professor with the Department of Computer Science, Kumamoto University. He is engaged in research on intelligent transportation systems, and computer vision. He is a Member of the Institute of Electronics and Information Communication Engineers of Japan.

Hanqing Lu was born in 1961. He received his B.E. degree and M.S. degree both from Harbin Institute of Technology in 1982 and 1985, respectively. He received his Ph.D degree from Huazhong University of Science and Technology in 1992. Since 1992, he has been with the Institute of Automation of Chinese Academy of Sciences, where he is now a Professor. His research interests include image processing, content-based image and video retrieval, object tracking, and recognition.