Epipolar Geometry Estimation and Its Application to Image Coding

Koichi Hata Minoru Etoh

Tokyo Communication Systems Research Laboratory,
Matsushita Electric Industrial Co., Ltd.
Tokyo, Japan 140-8632

Abstract

This paper presents image-based rendering related techniques; (1)a computer vision technique to estimate accurate epipolar geometry and (2) an image coding technique to compress multiple-viewpoint images, which are stored for reconstruction of any arbitrary view of 3-D object, with the epipolar geometry. Unlike other epipolar geometry estimation techniques based on the correct sets of feature point correspondence, which is hardly given in practice, the proposed method minimizes absolute luminance difference between two images with respect to seven epipolar geometry parameters. In the framework of block matching, the absolute luminance difference between the two images can be represented by the seven parameters with which a nonlinear robust regression follows. We apply the epipolar geometry estimation technique to multiple-viewpoint image coding. That practically contributes to improving the coding efficiency, where more than 1000 viewpoint images are coded by a type of predictive coding based on epipolar geometry. The experimental results show that the compression ratio attains to 1/100 and more.

1 Introduction

Image-based rendering is a promising approach in computer graphics(CG) to reconstruct any arbitrary view of 3-D real object without having its 3-D data[1]. Unlike 3-D object rendering in traditional CG, image-based rendering requires a huge set of images. Fig. 1 shows our image capturing system that takes 1500 images (i.e., 350M Bytes images) to reconstruct any arbitrary view. We take an approach,"compression of a huge set of multiple-viewpoint images." Suppose more than 1000 images are captured by a moving camera depicted in Fig. 2, the 3-D relative motion between the camera and the object result in temporal change[1] in the 2-D image plane. In multi-viewpoint images(in short, MV images), the 2-D temporal change is small and there is similarity between two adjacent viewpoints. Thus, we can apply predictive image coding scheme taking one image or more as reference, like moving picture coding standards such as MPEG[2].

[1]To be exact it is not "temporal" but "inter-viewpoint."

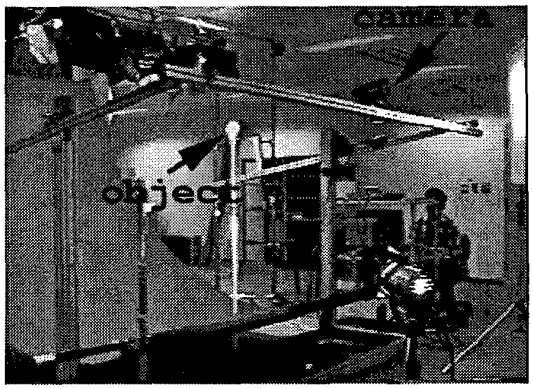


Figure 1: Multi-Viewpoint Image Capturing System.

To compress MV images, actually we use motion-compensated DC (MC-DCT) scheme but extend it with a sophisticated motion compensation method. To reduce inter-frame redundancy, a traditional MC-DCT scheme takes 16×16 pixels square block with 2D motion vector as depicted in Fig. 3. Here, note that the 2-D temporal change can be modeled by a *rigid motion projection model* in MV images. That is the epipolar geometry, the only generally-known geometric constraint that exists between two projections of a single rigid object from different viewpoints.

Tsai[3] and B. Girod[4] suggested the idea of *motion compensation along a linear trace defined by Epipolar geometry(epipolar motion compensation)* in predictive image coding. With the epipolar constraints, motion compensation in coding MV images can take a block with 1-D motion vector as depicted in Fig. 4. They, however, have not shown the real usage of Fig. 4, since neither a geometry estimation technique nor a compression application with rigid body motion has not been practically available before. On the contrary, we have a well-motivated application, compression of MV images, which have only rigid body motion in the 2-D image plane.

The remaining issues are a geometry estimation

0-7803-5467-2/99/ $ 10.00 © 1999 IEEE

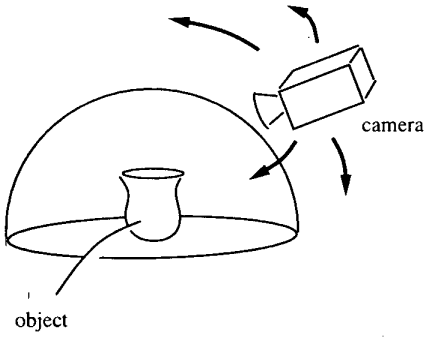


Figure 2: Relative motion between object and camera.

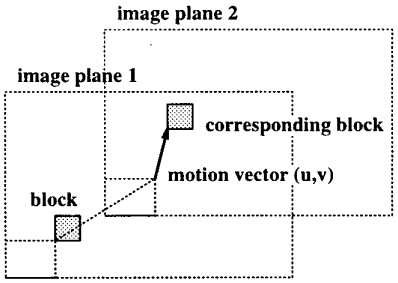


Figure 3: Traditional motion compensation scheme.

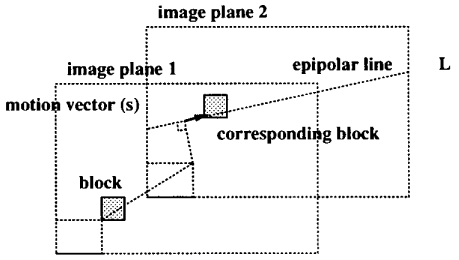


Figure 4: Epipolar motion compensation.

technique and to show how it works for the compression of MV images. We will describe them in the following sections.

2 Epipolar Geometry

2.1 Basic idea

Longuet-Higgins[5] has proposed an epipolar constraint estimation algorithm under the assumption that the correct sets of feature point correspondence are given. It is known as eight-point algorithm. He has shown that if at least eight sets of correspondence are obtained, the epipolar constraint is obtained by solving eight simultaneous linear equations. One common approach is based on repetitive hypothesis-and-test procedures. Olsen[6] has proposed an epipolar estimation algorithm that uses a linear system with a robust method. In his method, he solved the linear equations by a least squares approach with an M-estimator[7]. One drawback of his method is that the algorithm's performance depends heavily on heuristic feature-point matching. Zhang, Deriche and Xu[8][9][10] have proposed a method using homography in which they parameterized the epipolar constraint in a different way. Homography is the relation between the epipolar lines in the first image and those in the second image. The nonlinear system has only one more parameter than the exact number of the epipolar constraint. However, the algorithm's performance still depends on a matching heuristics.

In place of the feature-point matching, we propose *to minimize the absolute luminance difference between two images with respect to seven epipolar geometry parameters.* For this minimization, we approximate the block matching results of the two images by quadratic scalar functions. Each quadratic function represents a local absolute luminance difference associated with local motion vectors, which are further associated with the epipolar parameters. We call the quadratic function value a parameterized block correlation[11]. Thus, the absolute luminance difference is represented by the sum of parameterized block correlations. We estimate the epipolar geometry by minimizing the sum of parameterized block correlations with a nonlinear robust regression technique. Zhang et al.[8] pointed out that block matching might be irrelevant at occlusion boundaries and within featureless regions. It must be noted that the localization of a featureless region has two-dimensional uncertainty, and that of a boundary region has one-dimensional uncertainty along the boundaries (i.e., well-known *aperture problem*). For that reason, the feature point matching approach by necessity discards unreliable matches even if some of these matches are reliable in a certain direction. The use of parameterized block correlation enables us to utilize not only two-dimensional reliability (i.e., corner point matching) but also directional reliability. Moreover, since parameterized block correlation provides continuous measure with respect to the epipolar parameters, we can easily incorporate robust regression techniques (e.g., M-estimator) into our estimation process.

2.2 Formulation

Fig. 5 depicts epipolar geometry that exists between two projections of a single rigid object from different viewpoints. In Fig. 5, the $\mathbf{C}_1$ and $\mathbf{C}_2$ indicate the optical centers of cameras. Let $\mathbf{M}$, $\mathbf{m}_1 = (x_1, y_1)^T$ and $\mathbf{m}_2 = (x_2, y_2)^T$ be a single point in three-dimensional space and its projected points on the different image planes. The epipolar geometry between $\mathbf{m}_1$ and $\mathbf{m}_2$ is represented by a 3x3 fundamental matrix[10] $\mathbf{F}$ as follows.

$$\tilde{\mathbf{m}_2}^T \mathbf{F} \tilde{\mathbf{m}_1} = 0 \ , \qquad (1)$$

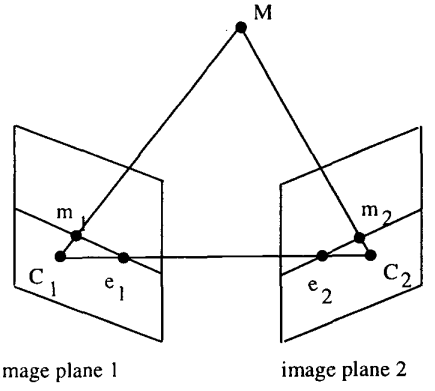


Figure 5: Epipolar geometry

where $\tilde{\mathbf{m}}_1$ and $\tilde{\mathbf{m}}_2$ denote homogeneous coordinate representation of $\mathbf{m}_1$ and $\mathbf{m}_2$ such that $\tilde{\mathbf{m}}_1 = (x_1, y_1, 1)^T$ and $\mathbf{m}_2 = (x_2, y_2, 1)^T$. The fundamental matrix $\mathbf{F}$ is defined by intrinsic parameters (which represent the origin of the image plane, the image coordinate axes, and sampling rate of the imaging device), the rotation and translation of the cameras.

Let us suppose that the $\mathbf{F}$ is known but $\mathbf{m}_2$ and $\mathbf{M}$ are unknown. We can find the line $\mathbf{l}_2$ on which $\mathbf{m}_2$ exists as follows.

$$\mathbf{l}_2 : \{\mathbf{x} | \tilde{\mathbf{x}}^T \mathbf{F} \tilde{\mathbf{m}}_1 = 0\} \tag{2}$$

where $\mathbf{x} = (x, y)^T$. We call the line $\mathbf{l}_2$ the epipolar line corresponding to the point $\mathbf{m}_1$. Thus, given the $\mathbf{F}$, we can obtain the epipolar line for every point on the image plane.

Next, let us introduce epipolar transformation. Epipolar transformation is defined by the transformation of the epipoles and the epipolar lines between the two different images. The epipoles are intersections of all epipolar lines. Let $\mathbf{e}_1$ and $\mathbf{e}_2$ be the epipoles in image 1 and image 2, respectively, and let $\mathbf{H}$ be the 2x2 matrix that maps the epipolar lines in images 1 to the epipolar lines in image 2. This is called the homography matrix.

The epipolar line corresponding to point $\mathbf{m}_1$ in image 1 is represented as follows by using a scale parameter s.

$$\mathbf{l}_2 : \mathbf{e}_2 + s\mathbf{H}(\mathbf{m}_1 - \mathbf{e}_1) \tag{3}$$

Thus the epipolar transformation is defined by the following seven parameters (x_{e_1}, y_{e_1}, x_{e_2}, y_{e_2}, a, b and c).

$$\mathbf{e}_1 = (x_{e_1}, y_{e_1})^T, \mathbf{e}_2 = (x_{e_2}, y_{e_2})^T, \mathbf{H} = \begin{pmatrix} c & 1 \\ a & b \end{pmatrix} . \tag{4}$$

We can rewrite the fundamental matrix by using these seven parameters as equation (5).

$$\mathbf{F}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{H}) = \begin{pmatrix} b & a & -(ay_{e_2} + bx_{e_2}) \\ -1 & -c & cy_{e_2} + x_{e_2} \\ (y_{e_1} - bx_{e_1}) & (cy_{e_1} - ax_{e_1}) & -(cy_{e_2} + x_{e_2})y_{e_1} + (ay_{e_2} + bx_{e_2})x_{e_1} \end{pmatrix} . \tag{5}$$

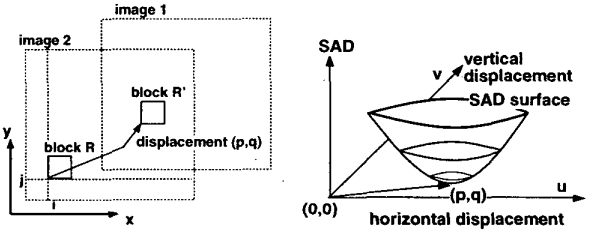
The epipolar geometry estimation now becomes the problem of estimating these seven parameters[2]. We will show the algorithm to solve the seven-parameter problem in the next section.

3 Estimation Algorithm

The epipolar constraint estimation scheme is based on the sum of absolute difference (SAD) block matching criterion, which uses a 16×16 pixels square block. In SAD calculation, the best match displacement (p, q) and its surrounding 3×3 positions' SAD measures are registered. Using these registered data, we apply Beaudet's differential operator to the surrounding SAD values(see [11] for detail). Thus, we can approximate the SAD measures at the (i, j)th block as expressed by the quadratic function:

$$E_{(i,j)}(u, v) = \bar{E}(p, q) + \frac{\partial E}{\partial u}(u - p) + \frac{\partial E}{\partial v}(v - q) + \frac{\partial^2 E}{\partial u^2}\frac{(u-p)^2}{2} + \frac{\partial^2 E}{\partial v^2}\frac{(v-q)^2}{2} + \frac{\partial^2 E}{\partial u \partial v}(u - p)(v - q) , \tag{6}$$

where (u, v) is the parameter of the function and denotes the motion vector of block correlation.



(a) Block search. (b) SAD surface.

Figure 6: Building SAD surface

The quadratic function is the second order Taylor series expansion of the SAD measure, where we assume it has a parabolic surface as depicted in Fig. 6. We call the quadratic function value a parameterized block correlation[11]. Since the absolute luminance difference is represented by the sum of parameterized block correlations, we can estimate the epipolar geometry by minimizing the sum of parameterized block correlations.

Letting $\mathbf{b}_{i,j}$ be the left-upper point of the (i, j)th block, the local motion vector (u, v) of the corresponding point, which must be on the epipolar line, is constrained by the following equation:

$$\tilde{\mathbf{v}}^T \mathbf{F}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{H}) \tilde{\mathbf{b}_{i,j}} = 0 , \tag{7}$$

[2]Further discussions on the epipolar geometry can be found in [10].

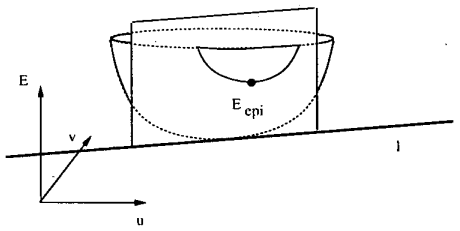


Figure 7: Energy on the epipolar line.

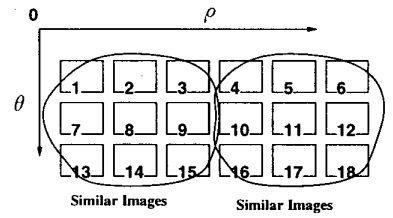
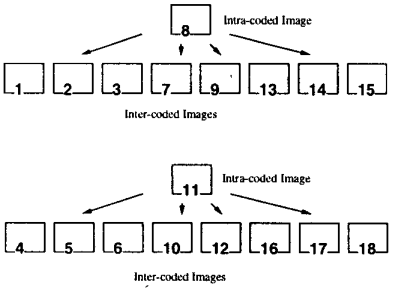


Figure 8: Images in two-dimensional space



where $\tilde{\mathbf{v}} = (u, v, 1)^T$. Therefore, the quadratic function $E_{(i,j)}(u, v)$ can be re-parameterized as $E^{epi}_{(i,j)}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{H})$ by the seven epipolar geometry parameters. Unlike the (u, v)-parameterization, we must consider the minimum $E_{(i,j)}(u, v)$ along the epipolar line. Fig. 7 depicts the relationship between $E_{(i,j)}(u, v)$ and $E^{epi}_{(i,j)}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{H})$.

We can obtain the seven epipolar parameters by minimizing the following equation:

$$E(\mathbf{e}_1, \mathbf{e}_2, \mathbf{H}) = \sum_{i,j} E^{epi}_{(i,j)}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{H}) \ , \qquad (8)$$

which is equivalent to the absolute luminance difference between two images, with respect to seven epipolar geometry parameters. Since $E^{epi}_{(i,j)}()$ is a nonlinear function, we apply a nonlinear robust regression technique to the seven-parameter estimation. For the nonlinear minimization we adopt a trust region method[12], which is a type of second derivative method.

In applying the robust regression, we discard irrelevant block matching. The irrelevancy is due to the two-dimensional correspondence problem. We can discriminate outliers in terms of $E^{epi}_{(i,j)}()$, which indicates degree of agreement to the estimated epipolar geometry. Please recall that we minimize the sum of $E^{epi}_{(i,j)}()$, accordingly, we can statistically treat the outlier rejection in the framework of M-estimation[7], in which we re-formalize Eq. (8) as

$$E'(\mathbf{e}_1, \mathbf{e}_2, \mathbf{H}) = \sum_{i,j} w_{ij} E^{epi}_{(i,j)}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{H}) \qquad (9)$$

where w_{ij} is a weight coefficient depending on the value of $E^{epi}_{(i,j)}()$.

4 Multi-viewpoint Image Coding

As mentioned in the introduction, we compress the multi-viewpoint images (MV images) with the epipolar motion compensation shown in Fig. 4. As illustrated in Fig. 2, a camera moves on a sphere and always looks at the center of the sphere, where an object is placed. We represent a viewpoint as (ρ, θ), where ρ and θ indicate the camera's horizontal rotation and vertical rotation. We control the viewpoint roughly in our system(Fig. 1); in other words, the system is not designed to accurately control the viewpoints. Thus, we don't know the exact viewpoint (ρ, θ) of each image. Furthermore, we don't know the exact intrinsic parameters of the camera. Therefore, we must estimate the epipolar constraint between two images from the image themselves.

Since we roughly know the viewpoints (ρ, θ) of images in our system, we associate the captured images to the two dimensional $\rho - \theta$ parameter space as shown in Fig. 8. In the $\rho - \theta$ parameter space, we group the adjacent images into clusters. Each cluster consists of one intra-frame coded image, which is used as the reference for prediction, and inter-frame coded images with reference to the intra-frame coded image (Fig. 4) The intra-frame coded image is encoded by a block-based DCT coding scheme, i.e., MPEG-4 [3], The inter-frame coded images, on the other hand, are coded by modified MPEG-4; the modification is only motion estimation and compensation part(i.e., changes from Fig 3 to Fig 4).

For the inter-frame coding, we first estimate the epipolar geometry between the intra- and inter- coded images. Next, we search the best block match along the estimated epipolar line as shown in Fig 4. Finally, the displacement along the epipolar line and the difference of luminance are encoded. Thus, we utilize epipolar geometry as a more advanced prediction model to achieve higher coding efficiency. Specifically, the epipolar constraint eliminates the necessity for the two-dimensional motion vectors needed by the conven-

[3]This is because the authors are in MPEG-4 standardization activity.

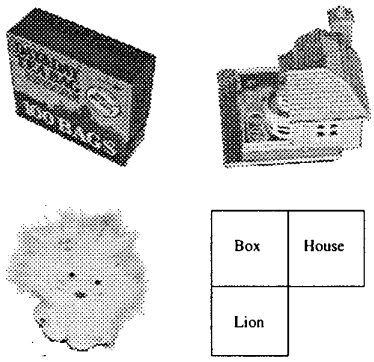


Figure 9: Images used in the experiments.

tional predictive coding scheme.

5 Experimental Results

We applied the proposed epipolar geometry estimation method to the compression of MV images using the coding scheme described in Section 4. Fig. 9 shows three test objects. Three experiments are conducted under the following conditions.

Experiment 1 : Inter-viewpoint coding

Experiment 2 : Inter-viewpoint coding + optimization of intra-coded frame selection

Experiment 3 : Inter-viewpoint coding + optimization of intra-coded frame selection + epipolar motion compensation

Roughly speaking, 100 viewpoint images were selected as intra-coded frame from 1500 viewpoint images, and the reset(i.e., adjacent viewpoint images) were coded as inter-coded frame. Since coding efficiency is affected by that prediction structure, which is selection of intra-coded frame, we optimized the selection by an annealing method in Experiment 2. Then we applied epipolar motion compensation in Experiment 3.

Table 1: Experimental results (average Kbit/frame).

	Box	House	Lion
Experiment 1	22.8	25.6	15.1
Experiment 2	18.2	21.7	11.4
Experiment 3	17.8	21.6	11.2

Tab.1 shows the experimental results, where we can summarize as follows. In Experiment 2, MV images were coded at 20 Kbit/frame in average. That corresponds to 1:80 compression ratio. In Experiment 2, the optimization worked well and compression ratio attained to 1:110. In Experiment 3, the epipolar motion compensation improved the coding efficiency about 2%, where compression ratio attained to 1:112.

6 Conclusion

We have proposed a robust and accurate epipolar geometry estimation for compact representation of multi-viewpoint images. The use of parameterized block correlation, which represent local absolute luminance difference, allows us to evaluate matching reliability in sub-pixel precision and reject outliers without any heuristics. *The point is that we do not need the correct sets of feature point correspondence.* We demonstrated the advantages of our estimation method by applying it to multi-viewpoint image coding. According to the results, more than 1000 images(i.e., 300-500 Mbytes images) are encoded into 3-5 Mbytes.

References

[1] S. E. Chen, "QuickTime VR – An Image-Based Approach to Virtual Environment Navigation," in Proc. SIGGRAPH'90, pp. 29–38, 1990.

[2] D. LeGall, "MPEG: a video compression standard for multimedia applications," Communications, ACM, 1991.

[3] R. Y. Tsai and T. S. Huang, "Estimating Three-Dimentional Motion Parameters of a Rigid Planar Patch," IEEE Trans. on ASSP, Vol. 29, No. 6, Dec, 1981.

[4] B. Girod and P. Wagner, "Displacement estimation with a rigid body motion constraint," in Proc. Picture Coding Symp. (PCS-90), paper 8.7, 1990.

[5] H.C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," Nature, Vol. 293, pp.133–135 1981.

[6] S.I. Olsen, "Epipolar Line Estimation," in Proc. ECCV, pp. 307–311, 1992.

[7] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, Inc., 1987.

[8] Z. Zhang, R. Deriche and O. Faugeras, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," Artificial Intelligence, Vol. 78, pp. 87–119, 1995.

[9] R. Deriche, Z. Zhang Q.-T. Luong and O. Faugeras, "Robust Recovery of the Epipolar Geometry for an Uncalibrated Stereo Rig," in Proc. ECCV, pp. 567–576, 1994.

[10] G. Xu and Z. Zhang, Epipolar Geometry in Stereo, Motion and Object Recognition – A Unified Approach, Kluwer Academic Publishers, 1996.

[11] M. Etoh, "Promotion of Block Matching: Parameteric Representation for Motion Estimation", in Proc. ICPR'98, pp. 282–285, 1998.

[12] P.E. Gill, W. Murray and M.H. Wright, Practical Optimization, Academic Press, 1981.