

[公開]

TR-M-0045

3D Image/Video Processing for Image Expression

朴 鍾一

Jong-Il PARK

1999.8.16

ATR 知能映像通信研究所

3D Image/Video Processing for Image Expression

Jong-Il Park

Dept. 3

ATR Media Integration and Communications
Research Laboratories

Table of Contents

- Final Report
 - Introduction
 - Explored Technical Topics
 - Application
 - Concluding Remarks
- Publications
 - List of Publications
 - Reprints
- Selected Presentations
 - List of Presentations
 - Reprints
- Appendix - Procedure of depth map extraction

Part 1.

Final Report

Final Report

Jong-Il Park
ATR MIC Dept.3

Introduction

For the 3 years (from March 1996 to March 1999) of my research life in ATR MIC, I have enjoyed working on 3D image and video processing for image expression. The technical topics covered include multi-view stereo matching, image-based rendering, video motion analysis, and multi-image alignment. A number of systems toward flexible and realistic image expression have been tried. Some of the systems such as 3D video compositing system, real-image-based virtual studio, vision-based real-time interaction system, and MPEG2 compressed-domain motion analyzer have been successfully demonstrated.

Explored Technical Topics

(1) Multi-View Stereo Matching

Toward getting a dense and sharp depth map, multi-view stereo matching algorithms are developed. We use 5 synchronized cameras (center, right, left, upper, and lower). Exploiting the symmetry of camera location, a simple matching algorithm is devised and the harmful influence of occlusions in stereo matching is considerably suppressed [6][7]. Moreover, two kinds of hierarchical implementation of the algorithm are implemented on a resolution pyramid [8][9]. Throughout a number of experiments with real images, the developed algorithms have shown impressive results [2][4][10][12][22][24]. Furthermore, a totally different approach has been tried to get a dense and sharp depth map. It is a diffusion-based stereo matching algorithm. A Bayesian formulation is first established and then solved by a diffusion which iteratively maximizes the a posteriori probability of the disparity field over the image [5][15][26][28].

They are applied to 3D video compositing demonstration in Image Expression Room, arbitrary view generation system, and multimedia performance at '98 ATR Lab. Exposition.

All the source codes are stored in miris53:/usr/people/pji/msm. A brief guide for experiments is attached in the appendix.

(2) Arbitrary View Synthesis

A framework of generating arbitrary views from multi-view images is developed [3][11][14][25]. We attempt to synthesize a new view with the best

attainable quality given limited information. It is based on estimation of a depth map and successive image mapping: transformation of the depth map, guessing the depth map for the uncovered area through a simple directional interpolation, and back-mapping using the filled depth map with a reasonable check. The uncovered area due to a view change can be handled on the basis of the observable viewpoint. An acceptable image quality could be generated for an arbitrary viewpoint within a scope [25]. Furthermore, in order to speed up the frame rate of synthesis, we developed an efficient algorithm based on an investigation into the epipolar geometry of multi-view images [29]. At present, the frame rate of the arbitrary view generation is about 4 to 10 frames/second for 320x240 images on a WindowsNT machine with Alpha 533 MHz CPU [20].

The view synthesis algorithm is a key technology in the Real-Image-Based Virtual Studio [27].

All the source codes are stored in miris53:/usr/people/pji/view_synth.

(3) Video Annotation

Automatic annotation of video plays a key role in constructing a *useful*, in the sense that easy and fast search and easy and fast retrieval, multimedia database. Camera parameters in video sequence are very important feature that characterize a scene. A direct extraction of camera parameters from encoded video is developed [19][30]. Rather than using non-encoded or completely decoded images, we use only motion vectors to estimate camera parameters. Two types of methods are implemented. One is based on the extended Kalman filter and the other is based on the Levenberg-Marquardt method. The EKF approach shows better performance in accuracy while the LM method is superior in processing speed.

As an application, a compressed-domain motion analyzer for MPEG2 bit stream has been developed on Windows machine. To cope with interlace images in MPEG2 sequence, a novel estimation strategy is devised [19].

All the source codes are stored in mpc18 c:\usr\pji\camview.

(4) Synchronized Video Compositing and Mosaicking

In video compositing systems, viewpoints and image planes of multiple images to be composited have to be properly integrated for acquiring natural-looking composited images, especially when some of the images undergo camera operations. A robust method of estimating camera parameters from image sequence is developed [1]. It is based on an outlier rejection scheme and nonlinear least-square data fitting. The method can be applied to synchronized video compositing in virtual studio where all the virtual cameras involved are synchronized in camera work. The developed method can replace the conventional measurement system such as rotary encoder, zoom sensor, etc.

The estimated camera parameters can also be utilized in the acquisition of high-resolution images from video. Each image is aligned on a virtual image plane by using the camera parameters. Temporal nonlinear filtering is executed to eliminate moving objects. To enhance the quality of mosaic, some direct method of image alignment may be incorporated at the final alignment stage. The generated large image can be used as a virtual environment in various VR applications.

All the source codes are stored in miris53:/usr/people/pji/mosaic.

Application

▪ 3D video compositing system in Image Expression Room

An important part of the COMI&CS(Computer Organized Media Integration & Communication System) demonstration is the 3D depth-map-based real-time compositing system [13][15][23]. An actor in blue-back studio is extracted by a chroma-key technique and then inserted in a 3D virtual world constructed by using the depth estimation method [4]. In this way, we can construct a virtual environment for image expression in the IE Room.

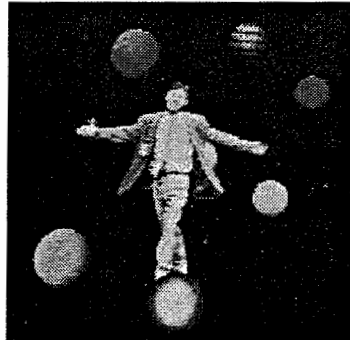


Fig. 1. An example of 3D compositing in the IE Room.

▪ Real-image-based virtual studio

Real-image-based virtual studio is an integration of all the developed technologies: 3D video compositing [23], depth map extraction [4], arbitrary view synthesis [25][29], and vision-based interaction. Unlike usual virtual studio systems, both of fore- and back-ground images can come from real images. According to the camera work in the studio, the view synthesis method can generate an arbitrary virtual scene from multi-view real images and thus we get natural-looking composited images [17][18][27].



Fig. 2. A result image shot in the Real-Image-Based Virtual Studio.

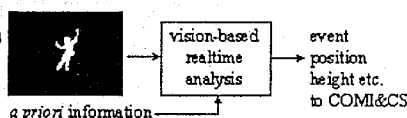
▪ **Vision-based real-time interaction for multimedia performance**

A vision-based real-time interaction system is developed for the multimedia performance during the ATR Lab. Exposition in Nov. 1998. It is implemented on an SGI INDIGO2 (miris61). Various events are detected as expected and the system is successfully applied to the multimedia performance.

□ **Interaction In Image Expression Room**

☆ **Realtime interaction**

- realtime extraction of 3D information
- realtime event handling
- realtime 3D video compositing



Application Example

Multimedia Performance

- live dance
- video
- CG
- music



Fig. 3. Illustration of vision-based real-time interaction.

▪ **Compressed-domain video annotation system**

A compressed-domain video motion analysis system is developed. The key technology in the system is camera parameter estimation. Both extended Kalman filter and Levenberg-Marquardt method are implemented [30]. An interesting feature of the system lies in the capability of dealing with the complicated motion compensation schemes in MPEG2 format [19]. Figure 4 shows the GUI of the system.

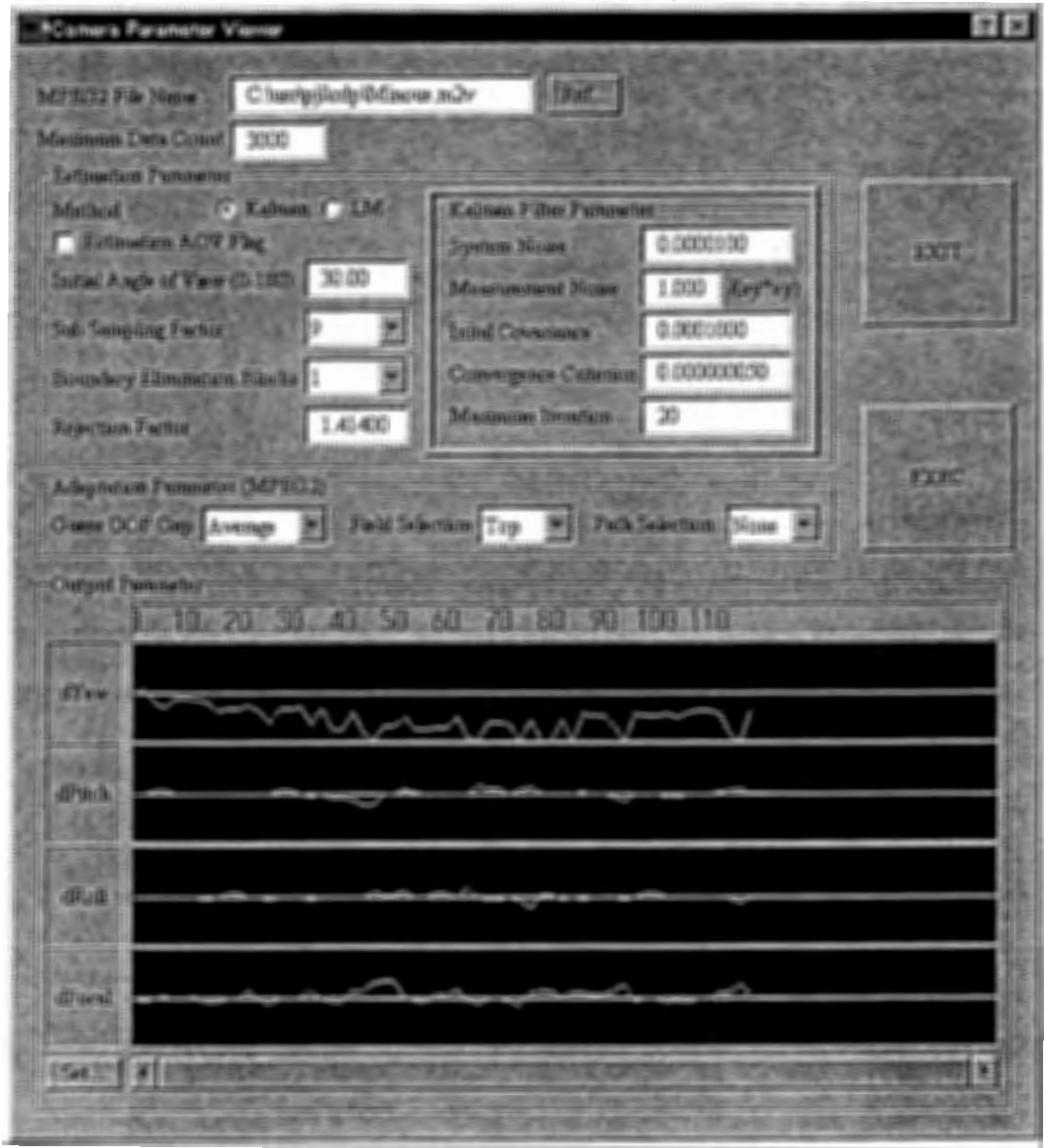


Fig. 4. The GUI of the compressed-domain video motion analysis system.

Concluding Remarks

Three-dimensional image and video processing for image expression has been my research area during my stay in ATR MIC for 3 years. I have explored multi-view stereo matching, image-based rendering, video motion analysis, and multi-image alignment.

The most prominent trend in this field is the convergence of computer vision and computer graphics. The recent rapid advance in computer vision technology and computing power enables us to dream of dealing with real images interactively, i.e. giving 3D functionality to real images. Thus, computer vision people are coming closer to computer graphics field. On the other hand, computer graphics people often find their solution in the advanced results of computer vision field when it comes to bottle-neck problems such as reality and processing speed. As a researcher in computer vision and image processing, my recent interest has been in this big stream. Many interesting trials have come true although the developed technologies have much room for exploration, especially, in high-end application.

The followings would be very interesting as a future work.

- Multi-view stereo matching for high-quality images
- Fast view synthesis working in video rate
- Multi-view video coding for 3D visual communications
- Stereoscopic application in a virtual environment
- Video restoration for motion blur
- Efficient multi-view camera calibration
- Long-term gesture analysis for Kansei processing.

Acknowledgement: I am grateful to ATR MIC for providing the wonderful research environment. I would like to take this opportunity to express my sincere acknowledgement to Dr. Nakatsu, Dr. Inoue, Dr. Iwadate, and members of Dept.3 and Planning Section for their great support. I would also like to thank Mr. Kita for his help throughout my experiments.

Contact Information

Jong-Il Park, Ph.D.
Assistant Professor
Division of Electrical and Computer Engineering
Hanyang University
17, Haengdang-dong, Sungdong-ku, Seoul, Korea
Tel.+82-2-2290-0368 Fax.+82-2-2281-9912
Email: jipark@email.hanyang.ac.kr

Part 2.

Publications

Publications

□ Journal Papers

1. **Jong-II Park** and Choong Woong Lee, "Robust estimation of camera parameters from image sequence for video composition," *Signal Processing: Image Communication*, vol.9, pp.43-53, Nov.1996.
2. **Jong-II Park** and Seiki Inoue, "A hierarchical depth mapping algorithm using multiple cameras," *Journal of Korean Society of Broadcasting Engineers*, vol.2, no.1, pp.45-56, March 1997 (in Korean).
3. **Jong-II Park** and Seiki Inoue, "Image-based rendering from multi-view images," *Journal of ITE*, vol.52, no.3, pp.371-376, March 1998
4. **Jong-II Park** and Seiki Inoue, "Acquisition of sharp depth map from multiple cameras," *Signal Processing: Image Communication*, vol.14, pp.7-19, Nov. 1998
5. Sang Hwa Lee, **Jong-II Park**, Seiki Inoue, and Choong Woong Lee, "Disparity estimation based on Bayesian maximum a posteriori algorithm," *IEICE Trans. on Fundamentals*, vol.E82-A, no.7, pp.1367-1376, July 1999.

□ Domestic Conference Papers

6. **Jong-II Park**, Hiroshi Fukuda, and Seiki Inoue, "Extraction of depth information for scene description and its application," *Proc. ITE'96*, pp.112-113, Nagoya, Japan, July 1996 (in Japanese).
7. **Jong-II Park** and Seiki Inoue, "Toward occlusion-free estimation of depth from multiple stereo images," *IEICE Autumn Conference'1996*, SD-7-1, Kanazawa, Japan, Sept. 1996 (in Japanese).
8. **Jong-II Park** and Seiki Inoue, "Hierarchical disparity estimation using multiple cameras," *Proc. IMPS'96*, pp.105-106, Yokohama, Japan, Oct. 1996.
9. **Jong-II Park** and Seiki Inoue, "Efficient acquisition of depth information using multiple cameras," *Proc. ITE Winter Conference'96*, 1-4, p.54, Tokyo, Japan, Dec. 1996 (in Japanese).
10. **Jong-II Park** and Seiki Inoue, "Depth extraction using multiple cameras for video composition," *Technical Report of IEICE*, PRMU96-133, pp.33-40, Kyoto, Japan, Jan. 1997 (in Japanese).
11. **Jong-II Park** and Seiki Inoue, "New view generation from multi-view image sequence," *Technical Report of IEICE*, IE96-121, pp.91-98, Sapporo, Japan, Feb. 1997 (in Japanese).

12. **Jong-II Park** and Seiki Inoue, "Extraction of depth map from multiple cameras and its application to image expression," *Sensing via Image Information'97*, pp.247-252, Yokohama, Japan, June 1997 (in Japanese).
13. Seiki Inoue, Michitoshi Ishiwaka, Shoji Tanaka, and **Jong-II Park**, "A study on image expression environment IE Room," *Proc. ITE'97*, pp.290-291, Tokyo, Japan, July 1997 (in Japanese).
14. Akira Matsui, **Jong-II Park**, Seiki Inoue, "View synthesis from multi-view images," *Proc. ITE'97*, pp.70-71, Tokyo, Japan, July 1997 (in Japanese).
15. **Jong-II Park**, Sang Hwa Lee, and Seiki Inoue, "Depth estimation from multiple stereo images with small disparity range," *Proc. IMPS'97*, pp.53-54, Karuizawa, Japan, Oct. 1997 (in Japanese).
16. Seiki Inoue, Michitoshi Ishiwaka, **Jong-II Park**, and Shoji Tanaka, "Image expression environment IE Room," *Third Symposium on Intelligent Information Media*, pp.89-96, Dec. 1997 (in Japanese).
17. **Jong-II Park**, Kei Ohkawa, and Seiki Inoue, "3D video composition for Image Expression Room," *Technical Report of IEICE*, IE97-168, pp.55-60, Sapporo, Japan, Feb. 1998 (in Japanese).
18. Kei Ohkawa, **Jong-II Park**, and Seiki Inoue, "Arbitrary stereo-view generation from multiple cameras," *IEICE Annual Conference '98*, D-11-97, Kanagawa, Japan, Mar. 1998 (in Japanese).
19. **Jong-II Park** and Yuichi Iwadate, "Information extraction from digital video – camera motion," *Technical Report of IEICE*, IE98-155, pp.1-6, Sapporo, Japan, Feb. 1999.
20. **Jong-II Park**, Masahiro Chiba, and Yuichi Iwadate, "Efficient arbitrary view generation using multiple cameras," *Technical Report of IEICE*, IE98-156, pp.7-12, Sapporo, Japan, Feb. 1999.

□ International Conference Papers

21. **Jong-II Park** and Seiki Inoue, "Image expression based on disparity estimation from multiple cameras," *Proc. 3rd Joint Workshop on Multimedia Communications*, pp.7.1.1-8, Taegu, Korea, Oct.1996.
22. **Jong-II Park** and Seiki Inoue, "Toward occlusion-free depth estimation for video production," *Proc. International Workshop on New Video Media Technology '97*, pp.131-136, Tokyo, Japan, Jan. 1997.
23. Seiki Inoue, Michitoshi Ishiwaka, Shoji Tanaka, **Jong-II Park**, "An image expression room," *Proc. Intl` Conf. on Virtual System and Multimedia'97*, pp.178-187, Geneva, Switzerland, Sept. 1997.
24. **Jong-II Park** and Seiki Inoue, "Hierarchical depth mapping from multiple cameras," *Proc. ICIAP'97*, vol.1, pp.685-692, Florence, Italy, Sept. 1997.
25. **Jong-II Park** and Seiki Inoue, "Arbitrary view generation using multiple cameras," *Proc. IEEE ICIP'97*, vol.I, pp.149-153, Santa Barbara, USA, Oct. 1997.

26. Sang Hwa Lee, **Jong-II Park**, and Choong Woong Lee, "A new stereo matching algorithm based on Bayesian Model," *Proc. IEEE ICASSP'98*, vol.V, pp.2769-2772, Seattle, USA, May 1998.
27. **Jong-II Park** and Seiki Inoue, "Real-image-based virtual studio," *Proc. of the 1st Intl' Conf. on Virtual Worlds '98*, pp.117-122, Paris, France, July 1998.
28. Sang Hwa Lee, **Jong-II Park**, and Choong Woong Lee, "A new stereo matching algorithm based on probabilistic diffusion," *Eusipco'98*, vol.IV, pp.2201-2204, Rhode, Greece, Sept. 1998.
29. **Jong-II Park** and Seiki Inoue, "Image-based view rendering for 3D visual communications," *Proc. of Intl' Workshop on Very Low Bitrate Coding - Visual Content Analysis and Coding*, Urbana-Champaign, USA, Oct. 1998.
30. **Jong-II Park**, Seiki Inoue, and Yuichi Iwadate, "Estimating camera parameters from motion vectors of digital video," *IEEE Multimedia Signal Processing'98*, pp.105-110, Redondo Beach, USA, Dec. 1998.

Patents

31. **Jong-II Park** and Seiki Inoue, "Depth information extraction device and method," Under review for Japanese Patent(Application Number 09-085208), Open on 23 Oct., 1998(Open Number 10-283474).

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-283474

(43) 公開日 平成10年(1998)10月23日

1961171

(51) Int.Cl.⁶

識別記号

F I

G 0 6 T 7/00

G 0 6 F 15/62

4 1 5

G 0 1 B 11/00

G 0 1 B 11/00

H

審査請求 有 請求項の数 8 O L (全 14 頁)

(21) 出願番号 特願平9-85208

(22) 出願日 平成9年(1997)4月3日

特許法第30条第1項適用申請有り 1996年10月7日 電子情報通信学会 画像工学研究専門委員会発行の「映像メディア処理シンポジウム (IMPS96) 第1回シンポジウム資料」に発表

(71) 出願人 595147700

株式会社エイ・ティ・アール知能映像通信研究所

京都府相楽郡精華町大字乾谷小字三平谷5番地

(72) 発明者 朴 鍾一

京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社エイ・ティ・アール知能映像通信研究所内

(72) 発明者 井上 誠喜

京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社エイ・ティ・アール知能映像通信研究所内

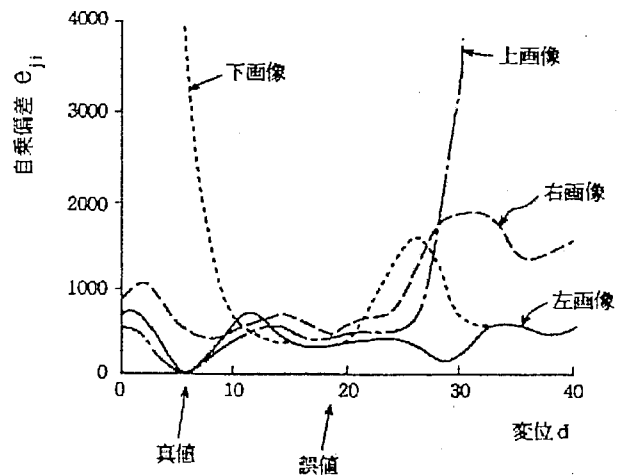
(74) 代理人 弁理士 深見 久郎 (外2名)

(54) 【発明の名称】 奥行き情報抽出装置および奥行き情報抽出方法

(57) 【要約】 (修正有)

【課題】 物体の境界線と抽出された奥行きマップのエッジの一致が良好な奥行き情報を抽出する。

【解決手段】 多眼カメラにより得られた基準画像及び参照画像に基づいて、奥行き情報の抽出を行う場合、参照画像中のマッチングウィンドウを所定画素数ずつ基準画像の方向に移動させながら、マッチングウィンドウ内の輝度値の分布について、各参照画像に対する自乗偏差を求め、自乗偏差の小さな方から所定数の値を代表値として、代表値が最小となるときの移動量を視差値として抽出する。



1

2

【特許請求の範囲】

【請求項1】 撮影したステレオ画像に基づいて、奥行き情報を抽出する奥行き情報抽出装置であって、対象画像を撮影する多眼撮像手段を備え、前記多眼撮像手段は、

基準画像を撮影し、光学像を各画素に対応する画像データに変換して出力する第1の撮像手段と、

前記第1の撮像手段の周りに光軸が互いに平行となるように配置され、それぞれが撮影した光学像を各画素に対応する画像データに変換して出力するn個(n:自然数、n≧2)の撮像手段とを含み、

前記第1の撮像手段から出力される第1の画像データと前記n個の第2の撮像手段からそれぞれ出力されるn個の第2の画像データとを記憶・保持する画像記憶手段と、

前記画像記憶手段からの出力に応じて、前記第1の画像データに属する各画素に対応する奥行き情報を抽出する奥行き情報演算手段とをさらに備え、

前記奥行き情報演算手段は、

第1の画像データの任意の1つの画素に対する奥行きデータの抽出処理を、

i) 前記第1の画像データと前記第2の画像データのうちのi番目(1≦i≦n)の画像データとの間で、前記第1の画像データ中の前記1つの画素を含む所定面積を有する第1の領域と、前記i番目の画像データ中の対応する位置の前記所定面積を有する第2の領域との画素データ値を比較して、一致の程度を表す第0番目の比較値を算出する処理を前記n個の第2の画像データのそれぞれに対して行い、算出されたn個の前記第0番目の比較値のうち小さい側からの所定数に基づいて第0番目の代表値を決定し、

ii) 前記各i番目の画像データにおいて、前記第2の領域を(移動量)=(単位移動量)×j(j:自然数)だけ第1の撮像手段の光軸に近づく方向にずらした後に、前記第1の領域と前記第2の領域との画素データ値を比較して、一致の程度を表す第j番目の比較値を算出する処理を前記n個の第2の画像データのそれぞれに対して行い、算出されたn個の前記第j番目の比較値のうち小さい側からの前記所定数に基づいて第j番目の代表値を決定するステップを、j=1からj=m(m:所定の自然数)まで繰り返し、

iii) 前記第0番目から第m番目までの代表値のうち最小のものに対応する移動量を、前記第1の画像データの任意の1つの画素に対する視差量として奥行きデータを算出することにより実行する、奥行き情報抽出装置。

【請求項2】 前記第1の画像データと前記i番目の画像データとの間の前記第j番目の比較値e_{ji}は、

前記第1の画像データの位置r=(x、y)における画素の輝度値をI₀(r)とし、前記第i番目の画像データの位置rにおける画素の輝度値をI_i(r)とし、前

記第1の領域をW₀、前記第2の領域をW₁とし、前記第i番目の画像データにおいて、位置rからd=(単位移動量)×jだけ第1の撮像手段の光軸に近づく方向にずれた位置を(r+d_{ji})とすると、

【数1】

$$e_{ji}(r, d_{ji}) = \sum_{\substack{b \in W_0 \\ b \in W_1}} [I_D(r+b) - I_i(r+b+d_{ji})]^2$$

として算出される、請求項1記載の奥行き情報抽出装置。

【請求項3】 撮影したステレオ画像に基づいて、奥行き情報を抽出する奥行き情報抽出装置であって、対象画像を撮影する多眼撮像手段を備え、前記多眼撮像手段は、

基準画像を撮影し、光学像を各画素に対応する画像データに変換して出力する第1の撮像手段と、

前記第1の撮像手段の周りに光軸が互いに平行となるように配置され、それぞれが撮影した光学像を各画素に対応する画像データに変換して出力するn個(n:自然数、n≧2)の撮像手段とを含み、

前記第1の撮像手段から出力される第1の画像データと前記n個の第2の撮像手段からそれぞれ出力されるn個の第2の画像データとを記憶・保持する画像記憶手段と、

前記画像記憶手段からの出力に応じて、前記第1の画像データに属する各画素に対応する奥行き情報を抽出する奥行き情報演算手段とをさらに備え、

前記奥行き情報演算手段は、

第1の画像データの任意の1つの画素に対する奥行きデータの抽出処理を、

i) 第0番目の探索ステップでは、探索領域を(移動量)=(単位移動量)×j(j:自然数、1≦j≦m、m:所定数)で規定される範囲として、

i-1) 前記第1の画像データと前記第2の画像データのうちのi番目(1≦i≦n)の画像データとの間で、前記第1の画像データ中の前記1つの画素を含む所定面積を有する第1の領域と、前記i番目の画像データ中の対応する位置の前記所定面積を有する第2の領域との画素データ値を比較して、一致の程度を表す第0番目の比較値を算出する処理を前記n個の第2の画像データのそれぞれに対して行い、算出されたn個の前記第0番目の比較値のうち小さい側からの所定数の比較値に基づいて第0番目の代表値を決定し、

i-2) 前記各i番目の画像データにおいて、前記第2の領域を前記移動量だけ第1の撮像手段の光軸に近づく方向にずらした後に、前記第1の領域と前記第2の領域との画素データ値を比較して、一致の程度を表す第j番目の比較値を算出する処理を前記n個の第2の画像データのそれぞれに対して行い、算出されたn個の前記第j番目の比較値のうち小さい側からの前記所定数の比較値

10

20

30

40

50

に基づいて第 j 番目の代表値を決定するステップを、j = 1 から j = m (m : 所定の自然数) まで繰り返し、
 i - 3) 前記第 0 番目から第 m 番目までの代表値のうち最小のものに対応する移動量を、前記第 1 の画像データの任意の 1 つの画素に対する第 0 番目の探索ステップにおける視差値として奥行きデータを算出し、

i i) 第 k 番目 (k ≥ 1) の探索ステップでは、前記所定面積を所定係数倍縮小し、かつ、前記縮小された所定面積を有する第 1 の領域内の画素に対して、第 (k - 1) 番目の探索ステップにおいてそれぞれ算出された p 個 (p : 自然数) の視差値の集合で規定される範囲を探索領域として、

i i - 1) 前記各 i 番目の画像データにおいて、前記第 2 の領域を前記探索領域に属する第 q 番目 (1 ≤ q ≤ p) の視差値だけ第 1 の撮像手段の光軸に近づく方向にずらした後に、前記第 1 の領域と前記第 2 の領域との画素データ値を比較して、一致の程度を表す第 q 番目のステップにおける比較値を算出する処理を前記 n 個の第 2 の画像データのそれぞれに対して行い、算出された n 個の前記第 q 番目のステップにおける比較値のうち小さい側からの前記所定数の比較値に基づいて第 q 番目の代表値を決定するステップを、q = 1 から q = p まで繰り返す、

i i - 2) 前記第 1 番目から第 p 番目までの代表値のうち最小のものに対応する移動量を、前記第 1 の画像データの任意の 1 つの画素に対する第 k 番目の探索ステップにおける視差量として奥行きデータを算出する探索ステップを k の値が所定数となるまで繰り返すことで実行する、奥行き情報抽出装置。

【請求項 4】 前記第 k 番目の探索ステップにおける、前記第 1 の画像データと前記 i 番目の画像データとの間の前記第 j 番目の比較値 e_{jik}は、

前記第 1 の画像データの位置 r = (x, y) における画素の輝度値を I₀(r) とし、前記第 i 番目の画像データの位置 r における画素の輝度値を I_i(r) とし、前記第 k 番目の探索ステップにおける前記第 1 の領域を W_{k0}、前記第 2 の領域を W_{k1} とし、前記第 i 番目の画像データにおいて、位置 r から d = (単位移動量) × j だけ第 1 の撮像手段の光軸に近づく方向にずれた位置を (r + d_{ji}) とするとき、

【数 2】

$$e_{jik}(r, d_{ji}) = \sum_{\substack{bew_0 \\ bew_1}} [I_0(r+b) - I_i(r+b+d_{ji})]^2$$

として算出される、請求項 1 記載の奥行き情報抽出装置。

【請求項 5】 平行な光軸を有する多眼カメラにより撮影した、基準画像である第 1 の画像データとそれを取り囲む n 個 (n : 自然数、n ≥ 2) の第 2 の画像データとを含むステレオ画像データに基づいて、前記第 1 の画像

データ中の任意の 1 つの画素に対する奥行き情報を抽出する奥行き情報抽出方法であって、

i) 前記第 1 の画像データと前記第 2 の画像データのうちの i 番目 (1 ≤ i ≤ n) の画像データとの間で、前記第 1 の画像データ中の前記 1 つの画素を含む所定面積を有する第 1 の領域と、前記 i 番目の画像データ中の対応する位置の前記所定面積を有する第 2 の領域との画素データ値を比較して、一致の程度を表す第 0 番目の比較値を算出する処理を前記 n 個の第 2 の画像データのそれぞれに対して行い、算出された n 個の前記第 0 番目の比較値のうち小さい側からの所定数に基づいて第 0 番目の代表値を決定するステップと、

i i) 前記各 i 番目の画像データにおいて、前記第 2 の領域を (移動量) = (単位移動量) × j (j : 自然数) だけ第 1 の撮像手段の光軸に近づく方向にずらした後に、前記第 1 の領域と前記第 2 の領域との画素データ値を比較して、一致の程度を表す第 j 番目の比較値を算出する処理を前記 n 個の第 2 の画像データのそれぞれに対して行い、算出された n 個の前記第 j 番目の比較値のうち小さい側からの前記所定数に基づいて第 j 番目の代表値を決定するサブステップを、j = 1 から j = m (m : 所定の自然数) まで繰り返すステップと、

i i i) 前記第 0 番目から第 m 番目までの代表値のうち最小のものに対応する移動量を、前記第 1 の画像データの任意の 1 つの画素に対する視差量として奥行きデータを算出するステップとを備える、奥行き情報抽出方法。

【請求項 6】 前記第 1 の画像データと前記 i 番目の画像データとの間の前記第 j 番目の比較値 e_{ji}は、

前記第 1 の画像データの位置 r = (x, y) における画素の輝度値を I₀(r) とし、前記第 i 番目の画像データの位置 r における画素の輝度値を I_i(r) とし、前記第 1 の領域を W₀、前記第 2 の領域を W₁ とし、前記第 i 番目の画像データにおいて、位置 r から d = (単位移動量) × j だけ第 1 の撮像手段の光軸に近づく方向にずれた位置を (r + d_{ji}) とするとき、

【数 3】

$$e_{ji}(r, d_{ji}) = \sum_{\substack{bew_0 \\ bew_1}} [I_0(r+b) - I_i(r+b+d_{ji})]^2$$

として算出される、請求項 2 記載の奥行き情報抽出方法。

【請求項 7】 平行な光軸を有する多眼カメラにより撮影した、基準画像である第 1 の画像データとそれを取り囲む n 個 (n : 自然数、n ≥ 2) の第 2 の画像データとを含むステレオ画像データに基づいて、前記第 1 の画像データ中の任意の 1 つの画素に対する奥行き情報を抽出する奥行き情報抽出方法であって、

第 0 番目の探索ステップと、第 1 から第 R 番目までの探索ステップとを備え、

前記第 0 番目の探索ステップは、

10

20

30

40

50

i) 探索領域を (移動量) = (単位移動量) × j (j : 自然数、1 ≤ j ≤ m, m : 所定数) で規定される範囲として、

i-1) 前記第1の画像データと前記第2の画像データのうちのi番目 (1 ≤ i ≤ n) の画像データとの間で、前記第1の画像データ中の前記1つの画素を含む所定面積を有する第1の領域と、前記i番目の画像データ中の対応する位置の前記所定面積を有する第2の領域との画素データ値を比較して、一致の程度を表す第0番目の比較値を算出する処理を前記n個の第2の画像データのそれぞれに対して行い、算出されたn個の前記第0番目の比較値のうち小さい側からの所定数の比較値に基づいて第0番目の代表値を決定するステップと、

i-2) 前記各i番目の画像データにおいて、前記第2の領域を前記移動量だけ第1の撮像手段の光軸に近づく方向にずらした後に、前記第1の領域と前記第2の領域との画素データ値を比較して、一致の程度を表す第j番目の比較値を算出する処理を前記n個の第2の画像データのそれぞれに対して行い、算出されたn個の前記第j番目の比較値のうち小さい側からの前記所定数の比較値に基づいて第j番目の代表値を決定するサブステップを、j = 1 から j = m (m : 所定の自然数) まで繰り返すステップと、

i-3) 前記第0番目から第m番目までの代表値のうち最小のものに対応する移動量を、前記第1の画像データの任意の1つの画素に対する第0番目の探索ステップにおける視差値として奥行きデータを算出するステップとを含み、

前記各第k番目 (1 ≤ k ≤ R) の探索ステップは、

ii) 前記所定面積を第(k-1)番目の探索ステップに比べて所定係数倍縮小し、かつ、前記縮小された所定面積を有する第1の領域内の画素に対して、第(k-1)番目の探索ステップにおいてそれぞれ算出されたp個 (p : 自然数) の視差値の集合で規定される範囲を探索領域として、

ii-1) 前記各i番目の画像データにおいて、前記第2の領域を前記探索領域に属する第q番目 (1 ≤ q ≤ p) の視差値だけ第1の撮像手段の光軸に近づく方向にずらした後に、前記第1の領域と前記第2の領域との画素データ値を比較して、一致の程度を表す第q番目のサブステップにおける比較値を算出する処理を前記n個の第2の画像データのそれぞれに対して行い、算出されたn個の前記第q番目のサブステップにおける比較値のうち小さい側からの前記所定数の比較値に基づいて第q番目の代表値を決定するサブステップを、q = 1 から q = p まで繰り返すステップと、

ii-2) 前記第1番目から第p番目までの代表値のうち最小のものに対応する移動量を、前記第1の画像データの任意の1つの画素に対する第k番目の探索ステップにおける視差量として奥行きデータを算出するステップ

とを含む、奥行き情報抽出方法。

【請求項8】 前記第k番目の探索ステップにおける、前記第1の画像データと前記i番目の画像データとの間の前記第j番目の比較値 e_{jik} は、

前記第1の画像データの位置 $r = (x, y)$ における画素の輝度値を $I_0(r)$ とし、前記第i番目の画像データの位置 r における画素の輝度値を $I_i(r)$ とし、前記第k番目の探索ステップにおける前記第1の領域を W_{k0} 、前記第2の領域を W_{k1} とし、前記第i番目の画像データにおいて、位置 r から $d = (\text{単位移動量}) \times j$ だけ第1の撮像手段の光軸に近づく方向にずれた位置を $(r + d_j i)$ とするとき、

【数4】

$$e_{jik}(r, d_{ji}) = \sum_{\substack{b \in W_{k0} \\ b \in W_{k1}}} [I_D(r+b) - I_i(r+b+d_{ji})]^2$$

として算出される、請求項7記載の奥行き情報抽出方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、多眼カメラにより撮影したステレオ画像からの視差情報に基づいて、奥行き情報を算出する奥行き情報抽出装置および奥行き情報抽出方法に関する。

【0002】

【従来の技術】 画像認識における特徴抽出や、柔軟かつ効率よく映像を合成、制作、加工、編集するための映像部品データベースの構築においては、カメラ等により撮影された画像情報から奥行き情報を抽出することが行われる。

【0003】 ここで、奥行き情報とは、光軸がZ軸と一致するカメラ中心座標系での各画素に対応する物体のZ座標値をいう。

【0004】 映像合成等のためには、密な、なおかつ、シャープな奥行きマップを得ることが必要となる。ここで、「密な」とは、画面全体の全ての画素ごとに奥行き値があることを意味し、「シャープな」とは、物体の境界線が明らかなことを意味する。

【0005】 ところで、奥行きマップを得るための方法としては、まず、いわゆる両眼ステレオ法が知られている。両眼ステレオ法とは、人間の両眼視を工学的に実現しようとするものである。

【0006】 図12は、標準的なステレオ撮像系の構成を示す概念図である。すなわち、視線の平行な等価な2つのカメラを設置し、このステレオ撮像系を用いて得られた2枚の画像から、画像上の各点でカメラからの距離が推定される。両眼ステレオ問題では、2枚の画像上での対応点が求めれば、幾何学的に三角測量の原理から距離を一意に定めることができる。

【0007】 図11に示した撮像系の場合、左右画像内

10

20

30

40

50

での対応点の位置は、水平走査線上(x軸)に拘束される。この水平走査線のことをエッジ線と呼ぶ。両画像中の対応点においては、明るさや明度パターンが似ていると仮定すると、対応点を含む所定面積の領域(以下、マッチングウィンドウと呼ぶ)についての明度差または明度間の相関係数を基にして、最適な対応点を探索することが可能である。このような方法を総称して、領域法と呼ぶ。

【0008】

【発明が解決しようとする課題】しかしながら、このような対応点を求める方法、すなわちステレオマッチングの方法の構成においては、以下のような問題点が存在する。

【0009】つまり、両画像間に規則的な模様が存在する場合、視点変化に伴う形変化がある場合、テキスト情報が不足している領域が存在する場合には、上述したような対応点の探索を一意に行うことができない。

【0010】さらに、一方の画像のある領域が、他方の画像からは観測されない場合があるという、いわゆるオクルージョン(occlusion)問題に対しては、両眼ステレオ法では、左右の両画像について真のマッチングをとることは不可能である。このため、根本的には対応することができず、何らかの仮定と知識に基づいて適当な補間を行うしかなかった。

【0011】上記問題点のうち、特にテキスト不足の問題、規則的なパターンの問題を改善する方法として、M. Okutomi and T. Kanade, "A Multi-baseline stereo", IEEE Trans. PAMI, vol. 15, no. 4, pp. 353-363, April 1993に開示されているように多基線ステレオマッチング法が提案されている。カメラの台数が増え、計算量が増加するという問題点はあるもののこの方法を用いることで、実時間奥行き抽出器も実現されている。

【0012】しかしながら、オクルージョン問題については、あまり注目した例が多くない。これは、かつてはロボットビジョンがステレオマッチングの主な応用であり、画面全体から見ると相対的に小さい領域であるオクルージョンには関心が払われにくかったという事情による。

【0013】ただし、後に説明するように、このオクルージョン問題の存在が、物体の境界線の明確に保存された奥行きマップの獲得を妨害する要因となる。

【0014】Y. Nakamura et al., "Occlusion detectable stereo - Occlusion patterns in camera matrix," Proc. IEEE CVPR'96, pp. 371-378, San Francisco, June 1996に開示されているように、オクルージョンのパターンを定量的に分析

し、オクルージョンが検出できるステレオマッチングも提案されているが、領域法におけるステレオマッチングの重要なポイントであるマッチングウィンドウの大きさに関する対策は検討されていない。

【0015】ここで、一般に領域法の性能はマッチングウィンドウの大きさに強く影響される。その大きさは、ある領域の特徴を表せる十分な輝度変化を含み、雑音耐性を保つような大きさとする必要がある。一方、形の変化や不連続部の影響を避けるためには、その大きさは小さくなければならない。

【0016】このようなマッチングウィンドウの問題に対する方策としては、W. E. L. Grimson, "A computer implementation of a theory of human stereo vision," Phil. Trans. Royal Soc. London, vol. B292, p. 217-253, 1981に開示されているような階層的な手法がある。

【0017】すなわち、まず低い解像度で粗い推定を行い、その結果を利用して、徐々に解像度および推定精度を高めていくという方法である。

【0018】一般には、階層的な手法では、処理しようとしている層においては、それよりも一層低い解像度の粗い推定値の付近だけを探索するように制限を加える。

【0019】これは、隣接する画素の奥行きは急には変化しないという仮定に基づいており、平坦部など奥行きの変化の滑らかな領域では、比較的正確に奥行きを求め留ることが可能である。

【0020】しかしながら、物体の境界線など奥行きの不連続な領域では、やはり十分な精度は得られず、誤差を引き起こしてしまうという問題があった。

【0021】この発明は、上記のような問題点を解決するためになされたものであって、その目的は、オクルージョンの影響を抑制して物体の境界線付近の奥行き情報の誤差を低減することが可能な奥行き情報抽出装置および奥行き情報抽出方法を提供することである。

【0022】この発明の他の目的は、物体の境界線と奥行きマップのエッジの一致を保持しつつ、雑音耐性の良好な奥行き情報抽出装置および奥行き情報抽出方法を提供することである。

【0023】

【課題を解決するための手段】請求項1記載の奥行き情報抽出装置は、撮影したステレオ画像に基づいて、奥行き情報を抽出する奥行き情報抽出装置であって、対象画像を撮影する多眼撮像手段を備え、多眼撮像手段は、基準画像を撮影し、光学像を各画素に対応する画像データに変換して出力する第1の撮像手段と、第1の撮像手段の周りに光軸が互いに平行となるように配置され、それぞれが撮影した光学像を各画素に対応する画像データに変換して出力するn個(n:自然数、n≥2)の撮像手

段とを含み、第1の撮像手段から出力される第1の画像データとn個の第2の撮像手段からそれぞれ出力されるn個の第2の画像データとを記憶・保持する画像記憶手段と、画像記憶手段からの出力に応じて、第1の画像データに属する各画素に対応する奥行き情報を抽出する奥行き情報演算手段とをさらに備え、奥行き情報演算手段は、第1の画像データの任意の1つの画素に対する奥行きデータの抽出処理を、i) 第1の画像データと第2の画像データのうちのi番目(1 ≤ i ≤ n)の画像データとの間で、第1の画像データ中の1つの画素を含む所定面積を有する第1の領域と、i番目の画像データ中の対応する位置の所定面積を有する第2の領域との画素データ値を比較して、一致の程度を表す第0番目の比較値を算出する処理をn個の第2の画像データのそれぞれに対して行い、算出されたn個の第0番目の比較値のうち小さい側からの所定数に基づいて第0番目の代表値を決定し、i) 各i番目の画像データにおいて、第2の領域を(移動量) = (単位移動量) × j (j : 自然数) だけ第1の撮像手段の光軸に近づく方向にずらした後に、第1の領域と第2の領域との画素データ値を比較して、一致の程度を表す第j番目の比較値を算出する処理をn個の第2の画像データのそれぞれに対して行い、算出されたn個の第j番目の比較値のうち小さい側からの所定数に基づいて第j番目の代表値を決定するステップを、j = 1からj = m (m : 所定の自然数) まで繰り返し、i) i) 第0番目から第m番目までの代表値のうち最小のものに対応する移動量を、第1の画像データの任意の1つの画素に対する視差量として奥行きデータを算出することにより実行する。

【0024】請求項2記載の奥行き情報抽出装置は、請求項1記載の奥行き情報抽出装置の構成において、第1の画像データとi番目の画像データとの間の第j番目の比較値 e_{ji} は、第1の画像データの位置 $r = (x, y)$ における画素の輝度値を $I_0(r)$ とし、第i番目の画像データの位置 r における画素の輝度値を $I_i(r)$ とし、第1の領域を W_0 、第2の領域を W_1 とし、第i番目の画像データにおいて、位置 r から $d = (単位移動量) \times j$ だけ第1の撮像手段の光軸に近づく方向にずれた位置を $(r + d, j, i)$ とするとき、

【0025】

【数5】

$$e_{ji}(r, d_{ji}) = \sum_{\substack{b \in W_0 \\ b \in W_1}} [I_D(r+b) - I_i(r+b+d_{ji})]^2$$

【0026】として算出される。請求項3記載の奥行き情報抽出装置は、撮影したステレオ画像に基づいて、奥行き情報を抽出する奥行き情報抽出装置であって、対象画像を撮影する多眼撮像手段を備え、多眼撮像手段は、基準画像を撮影し、光学像を各画素に対応する画像データに変換して出力する第1の撮像手段と、第1の撮像手

段の周りに光軸が互いに平行となるように配置され、それぞれが撮影した光学像を各画素に対応する画像データに変換して出力するn個 (n : 自然数、n ≥ 2) の撮像手段とを含み、第1の撮像手段から出力される第1の画像データとn個の第2の撮像手段からそれぞれ出力されるn個の第2の画像データとを記憶・保持する画像記憶手段と、画像記憶手段からの出力に応じて、第1の画像データに属する各画素に対応する奥行き情報を抽出する奥行き情報演算手段とをさらに備え、奥行き情報演算手段は、第1の画像データの任意の1つの画素に対する奥行きデータの抽出処理を、i) 第0番目の探索ステップでは、探索領域を(移動量) = (単位移動量) × j (j : 自然数、1 ≤ j ≤ m、m : 所定数) で規定される範囲として、i-1) 第1の画像データと第2の画像データのうちのi番目(1 ≤ i ≤ n)の画像データとの間で、第1の画像データ中の1つの画素を含む所定面積を有する第1の領域と、i番目の画像データ中の対応する位置の所定面積を有する第2の領域との画素データ値を比較して、一致の程度を表す第0番目の比較値を算出する処理をn個の第2の画像データのそれぞれに対して行い、算出されたn個の第0番目の比較値のうち小さい側からの所定数の比較値に基づいて第0番目の代表値を決定し、i-2) 各i番目の画像データにおいて、第2の領域を移動量だけ第1の撮像手段の光軸に近づく方向にずらした後に、第1の領域と第2の領域との画素データ値を比較して、一致の程度を表す第j番目の比較値を算出する処理をn個の第2の画像データのそれぞれに対して行い、算出されたn個の第j番目の比較値のうち小さい側からの所定数の比較値に基づいて第j番目の代表値を決定するステップを、j = 1からj = m (m : 所定の自然数) まで繰り返し、i-3) 第0番目から第m番目までの代表値のうち最小のものに対応する移動量を、第1の画像データの任意の1つの画素に対する第0番目の探索ステップにおける視差値として奥行きデータを算出し、i) i) 第k番目(k ≥ 1)の探索ステップでは、所定面積を所定係数倍縮小し、かつ、縮小された所定面積を有する第1の領域内の画素に対して、第(k-1)番目の探索ステップにおいてそれぞれ算出されたp個 (p : 自然数) の視差値の集合で規定される範囲を探索領域として、i) i-1) 各i番目の画像データにおいて、第2の領域を探索領域に属する第q番目(1 ≤ q ≤ p)の視差値だけ第1の撮像手段の光軸に近づく方向にずらした後に、第1の領域と第2の領域との画素データ値を比較して、一致の程度を表す第q番目のステップにおける比較値を算出する処理をn個の第2の画像データのそれぞれに対して行い、算出されたn個の第q番目のステップにおける比較値のうち小さい側からの所定数の比較値に基づいて第q番目の代表値を決定するステップを、q = 1からq = pまで繰り返し、i) i-2) 第1番目から第p番目までの代表値のうち最小のものに対応す

10

20

30

40

50

る移動量を、第1の画像データの任意の1つの画素に対する第k番目の探索ステップにおける視差量として奥行きデータを算出する探索ステップをkの値が所定数となるまで繰り返すことで実行する。

【0027】請求項4記載の奥行き情報抽出装置は、請求項3記載の奥行き情報抽出装置の構成において、第k番目の探索ステップにおける、第1の画像データとi番目の画像データとの間の第j番目の比較値 e_{jik} は、第1の画像データの位置 $r = (x, y)$ における画素の輝度値を $I_0(r)$ とし、第i番目の画像データの位置 r における画素の輝度値を $I_i(r)$ とし、第k番目の探索ステップにおける第1の領域を W_{k0} 、第2の領域を W_{k1} とし、第i番目の画像データにおいて、位置 r から $d = (\text{単位移動量}) \times j$ だけ第1の撮像手段の光軸に近づく方向にずれた位置を $(r + d_{ji})$ とするとき、

【0028】

【数6】

$$e_{jik}(r, d_{ji}) = \sum_{\substack{b \in W_{k0} \\ b \in W_{k1}}} [I_D(r+b) - I_i(r+b+d_{ji})]^2$$

【0029】として算出される。請求項5記載の奥行き情報抽出方法は、平行な光軸を有する多眼カメラにより撮影した、基準画像である第1の画像データとそれを取り囲む n 個(n :自然数、 $n \geq 2$)の第2の画像データとを含むステレオ画像データに基づいて、第1の画像データ中の任意の1つの画素に対する奥行き情報を抽出する奥行き情報抽出方法であって、i)第1の画像データと第2の画像データのうちのi番目($1 \leq i \leq n$)の画像データとの間で、第1の画像データ中の1つの画素を含む所定面積を有する第1の領域と、i番目の画像データ中の対応する位置の所定面積を有する第2の領域との画素データ値を比較して、一致の程度を表す第0番目の比較値を算出する処理を n 個の第2の画像データのそれぞれに対して行い、算出された n 個の第0番目の比較値のうち小さい側からの所定数に基づいて第0番目の代表値を決定するステップと、ii)各i番目の画像データにおいて、第2の領域を(移動量) = (単位移動量) $\times j$ (j :自然数)だけ第1の撮像手段の光軸に近づく方向にずらした後に、第1の領域と第2の領域との画素データ値を比較して、一致の程度を表す第j番目の比較値を算出する処理を n 個の第2の画像データのそれぞれに対して行い、算出された n 個の第j番目の比較値のうち小さい側からの所定数に基づいて第j番目の代表値を決定するサブステップを、 $j = 1$ から $j = m$ (m :所定の自然数)まで繰り返すステップと、iii)第0番目から第m番目までの代表値のうち最小のものに対応する移動量を、第1の画像データの任意の1つの画素に対する視差量として奥行きデータを算出するステップとを備える。

【0030】請求項6記載の奥行き情報抽出方法は、請求項5記載の奥行き情報抽出方法の構成において、第1の画像データとi番目の画像データとの間の第j番目の比較値 e_{ji} は、第1の画像データの位置 $r = (x, y)$ における画素の輝度値を $I_0(r)$ とし、第i番目の画像データの位置 r における画素の輝度値を $I_i(r)$ とし、第1の領域を W_0 、第2の領域を W_1 とし、第i番目の画像データにおいて、位置 r から $d = (\text{単位移動量}) \times j$ だけ第1の撮像手段の光軸に近づく方向にずれた位置を $(r + d_{ji})$ とするとき、

【0031】

【数7】

$$e_{ji}(r, d_{ji}) = \sum_{\substack{b \in W_0 \\ b \in W_1}} [I_D(r+b) - I_i(r+b+d_{ji})]^2$$

【0032】として算出される。請求項7記載の奥行き情報抽出方法は、平行な光軸を有する多眼カメラにより撮影した、基準画像である第1の画像データとそれを取り囲む n 個(n :自然数、 $n \geq 2$)の第2の画像データとを含むステレオ画像データに基づいて、第1の画像データ中の任意の1つの画素に対する奥行き情報を抽出する奥行き情報抽出方法であって、第0番目の探索ステップと、第1から第R番目までの探索ステップとを備え、第0番目の探索ステップは、i)探索領域を(移動量) = (単位移動量) $\times j$ (j :自然数、 $1 \leq j \leq m$, m :所定数)で規定される範囲として、i-1)第1の画像データと第2の画像データのうちのi番目($1 \leq i \leq n$)の画像データとの間で、第1の画像データ中の1つの画素を含む所定面積を有する第1の領域と、i番目の画像データ中の対応する位置の所定面積を有する第2の領域との画素データ値を比較して、一致の程度を表す第0番目の比較値を算出する処理を n 個の第2の画像データのそれぞれに対して行い、算出された n 個の第0番目の比較値のうち小さい側からの所定数の比較値に基づいて第0番目の代表値を決定するステップと、i-2)各i番目の画像データにおいて、第2の領域を移動量だけ第1の撮像手段の光軸に近づく方向にずらした後に、第1の領域と第2の領域との画素データ値を比較して、一致の程度を表す第j番目の比較値を算出する処理を n 個の第2の画像データのそれぞれに対して行い、算出された n 個の第j番目の比較値のうち小さい側からの所定数の比較値に基づいて第j番目の代表値を決定するサブステップを、 $j = 1$ から $j = m$ (m :所定の自然数)まで繰り返すステップと、i-3)第0番目から第m番目までの代表値のうち最小のものに対応する移動量を、第1の画像データの任意の1つの画素に対する第0番目の探索ステップにおける視差値として奥行きデータを算出するステップとを含み、各第k番目($1 \leq k \leq R$)の探索ステップは、iii)所定面積を第(k-1)番目の探索ステップに比べて所定係数倍縮小し、かつ、縮小され

た所定面積を有する第1の領域内の画素に対して、第(k-1)番目の探索ステップにおいてそれぞれ算出されたp個(p:自然数)の視差値の集合で規定される範囲を探索領域として、i(i-1)各i番目の画像データにおいて、第2の領域を探索領域に属する第q番目(1 ≤ q ≤ p)の視差値だけ第1の撮像手段の光軸に近づく方向にずらした後に、第1の領域と第2の領域との画素データ値を比較して、一致の程度を表す第q番目のサブステップにおける比較値を算出する処理をn個の第2の画像データのそれぞれに対して行い、算出されたn個の第q番目のサブステップにおける比較値のうち小さい側からの所定数の比較値に基づいて第q番目の代表値を決定するサブステップを、q=1からq=pまで繰り返すステップと、i(i-2)第1番目から第p番目までの代表値のうち最小のものに対応する移動量を、第1の画像データの任意の1つの画素に対する第k番目の探索ステップにおける視差値として奥行きデータを算出するステップとを含む。

【0033】請求項8記載の奥行き情報抽出方法は、請求項7記載の奥行き情報抽出方法の構成において、第k番目の探索ステップにおける、第1の画像データとi番目の画像データとの間の第j番目の比較値 e_{jik} は、第1の画像データの位置 $r = (x, y)$ における画素の輝度値を $I_0(r)$ とし、第i番目の画像データの位置 r における画素の輝度値を $I_i(r)$ とし、第k番目の探索ステップにおける第1の領域を W_{k0} 、第2の領域を W_{k1} とし、第i番目の画像データにおいて、位置 r から $d = (\text{単位移動量}) \times j$ だけ第1の撮像手段の光軸に近づく方向にずれた位置を $(r + d_j i)$ とするとき、

【0034】

【数8】

$$e_{jik}(r, d_{ji}) = \sum_{\substack{\text{bewk0} \\ \text{bewk1}}} [I_0(r+b) - I_i(r+b+d_{ji})]^2$$

【0035】として算出される。

【0036】

【発明の実施の形態】

【実施の形態1】

【奥行き情報抽出装置の構成】図1は、本発明の実施の形態1の奥行き抽出装置100の構成を示す概略ブロック図である。

【0037】奥行き情報抽出装置100は、対象画像を撮像する多眼カメラ10と、多眼カメラ10からの画像信号をデジタル信号に変換するA/D変換器12と、A/D変換器12からの出力を受けて、基準画像および参照画像の画像データを保持する映像記憶装置14と、映像記憶装置14に記憶された画像情報に基づいて、基準画像の各画素について奥行き情報の抽出を行い、奥行きマップを出力する奥行き計算演算ユニット16とを含む。

む。

【0038】図2は、多眼カメラ10の構成をより詳細に示す概念図である。多眼カメラ10の中央部には、基準画像を撮像する基準カメラ11aが配置され、その上下左右に、一定の距離Lにおいて参照画像を撮像する参照カメラ11b~11eが配置される。ここでは、すべてのカメラ11a~11eは、光軸が平行で、同一仕様のものであるとする。実際には、各カメラ毎に微妙に特性が異なるため、カメラの校正を行い、実効的に同一特性のカメラが光軸を平行として配置されているのと等価な構成となるように調整されているものとする。

【0039】したがって、上記のような校正が可能な範囲であれば、各カメラは必ずしも同一仕様である必要はなく、等間隔に配置される必要もなく、また、光軸も必ずしも平行である場合に限定されるものではない。さらに、カメラの台数も、上述のような5台に限定されるものではなく、以下の説明で明らかとなるように、より多くのカメラを用いる場合にも、本発明を適用することが可能である。

【0040】ただし、光軸の平行なカメラが等間隔で配置されている場合の方が、奥行き情報を抽出する手順が簡素化される。

【0041】図3は、図2のように配置された多眼カメラ10により得られる画像データの幾何学的配置を示す図である。

【0042】図3中では、基準画像にインデックス $i = 0$ を割り当て、上面像には、 $i = 1$ を、左画像には $i = 2$ を、下面像には $i = 3$ を、右画像には $i = 4$ をそれぞれ割り当てて区別するものとする。

【0043】図3を参照して、3次元空間の点 $P = (X, Y, Z)$ は、基準画像上の点 $p_0 = (x_0, y_0)$ に投影される。ここで、 $x_0 = F \times X / Z$ 、 $y_0 = F \times Y / Z$ と表される。Fはカメラの焦点距離である。

【0044】点Pはまた参照カメラ $C_i (i = 1, 2, 3, 4)$ の参照画像上の点 $p_i = (x_i, y_i)$ にも投影される。

【0045】ここで、 x_i, y_i は、以下の式で表される。

【0046】

【数9】

$$x_i = F \frac{X - D_{ix}}{Z}, \quad y_i = F \frac{Y - D_{iy}}{Z}$$

【0047】ただし、基線長ベクトル $D_i = (D_{ix}, D_{iy})$ は、それぞれ $D_1 = (0, -L)$ 、 $D_2 = (-L, 0)$ 、 $D_3 = (0, L)$ 、 $D_4 = (L, 0)$ である。

【0048】図2および図3に示したようなカメラの配置では、点Pの真の視差 d_i は、すべての参照画像に対し、

$$d_i = F \times L / Z = |p_i - p_0|$$

で表せる。したがって、視差 d_i が推定されれば、それに応じて、奥行き Z を取得することができる。

【0049】 上述したようなカメラの配置では、ある画素について1つのカメラからオクルージョンのために真のマッチングがとれない場合、対称の位置に置かれた他のカメラからは真のマッチングとれる可能性が高い。

【0050】 以下に、説明するように、本発明ではこのような特性に基づいて奥行き情報の抽出を行う。

【0051】 【奥行き情報の抽出】 続いて、図1に示した、奥行き計算演算ユニット16の動作について説明する。図4は、多眼カメラ10により撮像され、映像記憶装置14に記憶される画像データの例を示す概念図である。

【0052】 以下では、基準画像中の人物の右手の甲の上の点 r に対応する画素の奥行きデータを抽出する場合について説明する。

【0053】 まず、点 r を囲む所定面積のマッチングウィンドウ W_0 を基準画像中に設定する。たとえば、上画像 ($i=1$) について、画面中の対応する位置 (基準画面におけるマッチングウィンドウの位置と画像周辺部から同等の位置) にマッチングウィンドウ W_1 を設定すると、視差のためにマッチングウィンドウ W_0 中の画像データとマッチングウィンドウ W_1 中の画像データとは、ずれたものとなっている。

【0054】 そこで、以下では、マッチングウィンドウ W_1 を基準画像の方向に、言い換えると基準画像を撮像したカメラの光軸に近づく方向に順次ずらして行きつつ、マッチングウィンドウ W_1 中の画像データとマッチングウィンドウ W_0 中の画像データがもっとも一致する位置を探索していく。このもっとも両者が一致する位置までずらせる量が、基準画像と参照画像 ($i=1$) との視差に対応することになる。同様の手続きを他の参照画面 ($i=2, 3, 4$) に対しても行うものとする。なお、以下では参照画像中のマッチングウィンドウをすべて共通の符号 W_1 で表すものとする。

【0055】 上述した両マッチングウィンドウ内の画像データの一致の程度を表す尺度として、たとえば、以下に示すような画像データの輝度値の自乗偏差を用いることが可能である。

【0056】

【数10】

$$e_{ji}(r, d_{ji}) = \sum_{\substack{b \in W_0 \\ b \in W_1}} [I_0(r+b) - I_i(r+b+d_{ji})]^2$$

【0057】 ここで、自乗偏差 $e_{ji}(r, d_{ji})$ の式中、 $I_0(r)$ は基準画像中の位置ベクトル r で表される位置の輝度値を、 $I_i(r)$ は参照画像 ($i=1 \sim 4$) 中の位置ベクトル r で表される位置の輝度値を示し、 $b \in W_0$ とは、ベクトル b がマッチングウィンドウ

W_0 中を動くことを、 $b \in W_1$ とは、ベクトル b がマッチングウィンドウ W_1 中を動くことを、それぞれ示す。

【0058】 また、ベクトル d_{ji} とは、参照画像 i において、マッチングウィンドウを順次基準画像の方向へずらせる手続きを行う際に、 j 番目に移動した移動ベクトルを表すものとする。つまり、移動後の位置ベクトルは、 $(r + d_{ji})$ となる。したがって、一回のステップで所定の単位移動量ずつ基準画像側にマッチングウィンドウ W_1 の位置をずらしていきとすると、 j 番目のステップでは、ベクトル d_{ji} の大きさは、以下の式で表されることになる。

$$|d_{ji}| = (\text{単位移動量}) \times j$$

また、移動量の最大値としては、予めカメラ配置等を考慮して、所定の値 SR が設定され、この最大移動量までの範囲を探索領域として、視差の推定が行われるものとする。

【0060】 ここで、単位移動量としては画素数の整数倍の値でもよく、また、画素数の任意の係数倍の値でもよい。任意の係数倍の場合、移動位置の輝度値は周囲の画素の輝度値から補間により求めることができる。

【0061】 図2に示したようなカメラの配置では、原理的には視差の大きさ d は、すべての参照画像に対して一定のはずである。

【0062】 したがって、奥行きの不連続ない場所では、たとえば、単純に4つの参照画像における自乗偏差の和が最小の点における移動量が視差を表すとしても、正確な結果を得ることができる。

【0063】 しかしながら、奥行きの不連続な場所、たとえばオクルージョンの影響のある場所では、以下のような問題点が存在する。

【0064】 図5は、各参照画像について、自乗偏差 e_{ji} を求めた場合の一例を示す。図5に示した例では、真の視差の値 ($d \sim 6$) において、オクルージョンの影響のために、下画像および右画像からの自乗偏差が大きな値を持つ。このため、たとえば、単純に4つの参照画像における自乗偏差の和が最小の点が視差を表すとして、視差の推定を行うと誤った値 ($d \sim 19$) が得られてしまうことになる。

【0065】 以上のような問題点には、マッチングをとる際、推定結果に悪影響を与えるような観測データを、予め除去しておくことで対処することが可能である。

【0066】 したがって、たとえば、位置 r における視差 $d(r)$ の推定を以下の式に従って、行うことが可能である。

【0067】

【数11】

$$d(r) = \arg \min_d \sum_{i=1}^2 e'_{ji}(r, d_{ji})$$

【0068】ここで、 e'_{ji} は e_{ji} をソートしたものであり、 $i < k$ に対して $e'_{ji} > e_{jk}$ である。つまり、上記の方法では、自乗偏差 e_{ji} のうち、小さい方から2つの和により、視差 d を評価する。

【0069】もちろん、カメラの台数や観測条件等に応じて、小さい方からいくつのデータをとるかは決定されるものであって、この個数に限定されるものではない。

【0070】図6は、上述したような、奥行きデータの抽出手順を示す概念図である。まず、 $j=0$ 、つまり $d_{0i}=0$ であって、マッチングウィンドウ W_1 は初期位置にあるとして、自乗偏差 e_{0i} を各参照画像において求める。

【0071】つづいて、自乗偏差 e_{ji} をソート処理して、 e'_{0i} を求め、小さい側の2つの和($e'_{03} + e'_{04}$)を評価値(代表値)として抽出する。

【0072】同様の処理を j の値を1ずつ増やしながらか、 $j=SR$ (探索領域の最大値)となるまで繰り返す。各ステップ($j=k$)で得られた評価値のなかで最小の値を有する($e'_{k3} + e'_{k4}$)に対応する d_{ki} を視差の値とする。ここで、すべての参照画像について d_{ki} の値は同一の値であるので、これを単純に d_k で表すことにする。

【0073】図7は、図6で説明した手順を、基準画像中の任意の1つの画素に対する処理について、より詳細に示すフローチャートである。奥行き計算演算ユニット16は、処理を開始すると(ステップS100)、まず変数 j の値を初期化する(ステップS102)。

【0074】つづいて、奥行き計算演算ユニット16は、各参照画像について自乗偏差 e_{ji} を計算し(ステップS104)、各参照画像について得られた自乗偏差の値 e_{j1} 、 e_{j2} 、 e_{j3} 、 e_{j4} の値をソート処理する(ステップS106)。

【0075】ソート結果に基づいて、小さい側から2つの値の和($e'_{j3} + e'_{j4}$)が代表値として代表値配列 $E(j)$ に格納される(ステップS108)。

【0076】つづいて、奥行き計算演算ユニット16は、移動量の値が探索領域の最大値に達しているかの判断を行い(ステップS110)、最大値に達していない場合は、 j の値を1だけ増加させて(ステップS112)、ふたたび処理をステップS104に復帰させる。

【0077】一方、奥行き計算演算ユニット16は、移動量が最大値に達していると判断した場合(ステップS110)、代表値配列 $E(j)$ のなかの最小値を求める(ステップS114)。最小値に対応する移動量を視差 d_m として出力し(ステップS116)、処理を終了する(ステップS118)。

【0078】以上のような構成とすることで、奥行き情報抽出装置100は、物体の境界付近の奥行き情報の抽出においてオクルージョンの影響を抑制し、物体の境界線と奥行きマップのエッジとの一致の良好な奥行き情報

抽出動作を実現することが可能である。

【0079】また、奥行き計算演算ユニット16が行う処理に対応する奥行き情報抽出方法によっても、同様の効果を得ることができる。

【0080】[実施の形態2] 実施の形態1の奥行き情報抽出装置においては、マッチング処理において、4つの参照画像からのデータのうち、選択された2つの参照画像からのデータに基づいて視差の推定を行うという非線形な処理により、オクルージョンの影響を低減することが可能であった。

【0081】しかしながら、4つの参照画像からの情報のうち2つしか用いないということは、雑音耐性の観点からは望ましくない。

【0082】そこで、単純に雑音耐性を向上させるためだけであれば、マッチングウィンドウの大きさを大きくすることで対処可能である。しかしながら、上述したとおり、本当の物体の境界線と奥行きマップのエッジが一致するようにするためには、マッチングウィンドウの大きさは、小さい方が望ましい。

【0083】実施の形態2では、このようなトレードオフを解決することが可能な奥行き情報抽出装置および奥行き情報抽出方法を提供する。

【0084】ここで、実施の形態2の奥行き情報抽出装置の構成は、基本的に実施の形態1の奥行き情報抽出装置100の構成と同様である。

【0085】実施の形態2の奥行き情報抽出装置の構成が、奥行き情報抽出装置100の構成と異なる点は、以下に説明するとおり、奥行き計算演算ユニット16が階層的な奥行き情報抽出動作を行う点である。

【0086】以下に述べる、奥行き計算演算ユニット16の階層的な奥行き情報抽出動作は、まず大きさの大きなマッチングウィンドウで実施の形態1の奥行き情報抽出動作を行った場合、その奥行きマップ中のマッチングウィンドウ中に正しい奥行き値が存在していることを前提としている。

【0087】図8は、このような奥行き計算演算ユニット16の階層的な奥行き情報抽出動作を説明するための処理およびデータのフローを示す図である。

【0088】まず、奥行き計算演算ユニット16は、第0の階層では、マッチングウィンドウの大きさを $WS=W_{00}=W_{01}=15 \times 15$ 画素として、実施の形態1の図7において説明した手順に従って、視差情報の抽出を行い(ステップS200)、視差マップを作成する(ステップS202)。

【0089】ここで、基準画面におけるマッチングウィンドウを W_{00} で、参照画面におけるマッチングウィンドウを W_{01} で表した。

【0090】つづいて、奥行き計算演算ユニット16は、第1の階層の処理として、マッチングウィンドウの大きさを $WS=W_{10}=W_{11}=7 \times 7$ 画素として、視

差情報の抽出を行う（ステップS206）。ここで、この第1の階層での基準画面におけるマッチングウィンドウをW10で、参照画面におけるマッチングウィンドウをW11で表す。また、第1の階層においては、視差を探索する探索領域は、後に説明するように、マッチングウィンドウW10中に含まれる各画素について、第0の階層の処理で得られた視差値の集合の範囲に限定するものとする（ステップS204）。

【0091】第1の階層で得られた視差値に基づいて、視差マップが作成される（ステップS208）。

【0092】つづいて、奥行き計算演算ユニット16は、第2の階層の処理として、マッチングウィンドウの大きさをWS=W20=W21=7×7画素として、視差情報の抽出を行う（ステップS212）。ここで、この第2の階層での基準画面におけるマッチングウィンドウをW20で、参照画面におけるマッチングウィンドウをW21で表す。また、第2の階層においても、視差を探索する探索領域は、マッチングウィンドウW20中に含まれる各画素について、第1の階層の処理で得られた視差値の集合の範囲に限定するものとする（ステップS210）。

【0093】第2の階層で得られた視差値に基づいて、視差マップが作成される（ステップS214）。

【0094】図8に示した例では、処理の階層は、第2の階層までとした。したがって、ステップS214において作成された視差マップに基づいて、奥行きマップが形成される。

【0095】階層の数は、測定対象や測定条件等に応じて、変更することが可能である。上述したとおり、図8のステップS200の処理は、実施の形態1の図7で説明した処理と同様である。

【0096】図9は、図8中のステップS206またはステップS212の処理を示すフローチャートである。

【0097】まず、1つ前の階層における視差マップの抽出結果に基づいて、処理を行う階層の基準画像に対するマッチングウィンドウWk0（k=1, 2）中に存在する視差値を配列ds(j)（i=1, 2, …, jmax : jmaxは、一つ前の階層の視差マップにおいてマッチングウィンドウWk0中に存在する視差値の個数）中に読み込んでおくとともに、変数jの値を初期値”1”に、変数dの値をds(1)に初期化しておく（ステップS302）。

【0098】つぎに、奥行き計算演算ユニット16は、マッチングウィンドウWk1を移動量dで規定される量だけ、初期位置（基準画像中のマッチングウィンドウWk0の位置に対応する参照画像中の位置）からずらして、各参照画面において自乗偏差ejikを求める（ステップS304）。ここで、自乗偏差ejikは、k番目の階層の処理でのj番目のステップにおけるi番目の参照画像での自乗偏差であることを示し、以下の式で表

わされる。

【0099】

【数12】

$$e_{jik}(r, d_{ji}) = \sum_{\substack{bewk0 \\ bewk1}} [I_D(r+b) - I_i(r+b+d_{ji})]^2$$

【0100】ここで、実施の形態1と同一の符号は、同一の意味を表わす。つづいて、奥行き計算演算ユニット16は、各参照画像について得られた自乗偏差の値ej1k、ej2k、ej3k、ej4kの値をソート処理する（ステップS306）。

【0101】ソート結果に基づいて、小さい側から2つの値の和（ej3k+ej4k）が代表値として代表値配列E(j)に格納される（ステップS308）。

【0102】つづいて、奥行き計算演算ユニット16は、移動量の値が探索領域の最大値に達しているかの判断を行い（ステップS310）、最大値に達していない場合は、jの値を1だけ増加させて（ステップS312）、移動量dの値をds(j)として、ふたたび処理をステップS304に復帰させる。

【0103】一方、奥行き計算演算ユニット16は、移動量が最大値に達していると判断した場合（ステップS310）、代表値配列E(j)のなかの最小値を求める（ステップS314）。最小値に対応する移動量ds(j)を視差dmとして出力し（ステップS316）、次の画素に対する処理に移行する。

【0104】以上のような構成とすることで、実施の形態2の奥行き情報抽出装置は、物体の境界付近の奥行き情報の抽出においてオクルージョンの影響を抑制し、物体の境界線と奥行きマップのエッジとの一致が良好であるとともに、雑音耐性に優れた奥行き情報抽出動作を実現することが可能である。

【0105】図10は、上述したような階層的な奥行き情報抽出方法の流れを示す概念図であり、図11は、従来の階層的な奥行き情報抽出方法の流れを示す概念図である。

【0106】まず、図11を参照して、従来の階層的な奥行き情報抽出方法においては、まず低い解像度で粗い推定を行い、その結果を利用して、徐々に解像度および推定精度を高めていく。

【0107】つまり、処理しようとしている(n+1)層においては、それよりも一層低い第n層における解像度の粗い推定値dnの付近だけを探索するように制限を加える。同時に、マッチングウィンドウの大きさも、たとえば第n層の大きさの1/2として、視差値の推定値dn+1を求める。

【0108】このような構成とすることで、計算量を低減することは可能であるが、上述したとおり、物体の境界線など奥行きの不連続な領域では、やはり十分な精度は得られず、誤差を引き起こしてしまうという問題があ

10

20

30

40

50

った。

【0109】これに対して、図10を参照して、実施の形態2の奥行き計算演算ユニット16の処理においては、第1層の処理では、参照画像中のマッチングウィンドウの初期位置から基準画像へ向けて移動させる最大量のSRまでの範囲を探索領域として、実施の形態1で説明した方法で視差値の推定を行う。

【0110】第2層目以降では、直前の層で得られた視差マップに基づき、マッチングウィンドウ中に存在する奥行き値だけを候補として探索を行う。

【0111】したがって、マッチングウィンドウの十分に広い第1層での推定値の中から、マッチングウィンドウを小さくした場合に、推定値として最適なものを抽出することになる。

【0112】つまり、第1層での処理においては、雑音耐性の良好な処理が行われて視差値の候補が抽出され、順次マッチングウィンドウを小さくして行う第2層以降の処理により、物体の境界線と奥行きマップのエッジの一致が良好な奥行き値が、奥行き値の候補の中から選択されることになる。

【0113】以上のような構成とすることで、実施の形態2の奥行き情報抽出方法は、物体の境界付近の奥行き情報の抽出においてオクルージョンの影響を抑制し、物体の境界線と奥行きマップのエッジとの一致が良好であるとともに、雑音耐性に優れた奥行き情報抽出動作を実現することが可能である。

【0114】

【発明の効果】以上説明したとおり、本発明によれば、オクルージョンの影響を抑制して物体の境界線付近の奥行き情報の誤差を低減することが可能な奥行き情報抽出装置および奥行き情報抽出方法を提供することが可能である。

【0115】さらに、この発明によれば、物体の境界線と奥行きマップのエッジの一致を保持しつつ、雑音耐性の良好な奥行き情報抽出装置および奥行き情報抽出方法

を提供することが可能となる。

【図面の簡単な説明】

【図1】本発明の実施の形態1の奥行き情報抽出装置100の構成を示す概略ブロック図である。

【図2】多眼カメラ10の構成を示す概念図である。

【図3】多眼カメラ10により得られる画像データの幾何学的配置を示す図である。

【図4】多眼カメラ10により撮像された画像データを示す概念図である。

10 【図5】各参照画像について、自乗偏差 $e_j i$ を求めた場合を示す図である。

【図6】実施の形態1の奥行きデータの抽出手順を示す概念図である。

【図7】図6で説明した手順を、基準画像中の任意の1つの画素に対する処理について、より詳細に示すフローチャートである。

【図8】実施の形態2の階層的な奥行き情報抽出動作を説明するための処理およびデータのフローを示す図である。

20 【図9】図8に示した多眼ステレオマッチング処理の流れをより詳細に示すフローチャートである。

【図10】実施の形態2の階層的な奥行き情報抽出方法の流れを示す概念図である。

【図11】従来の階層的な奥行き情報抽出方法の流れを示す概念図である。

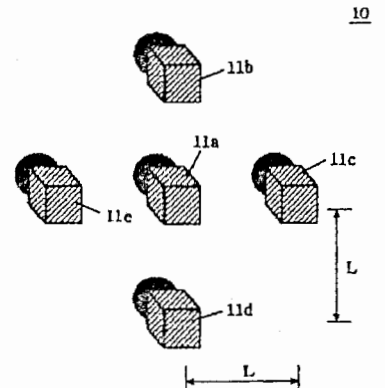
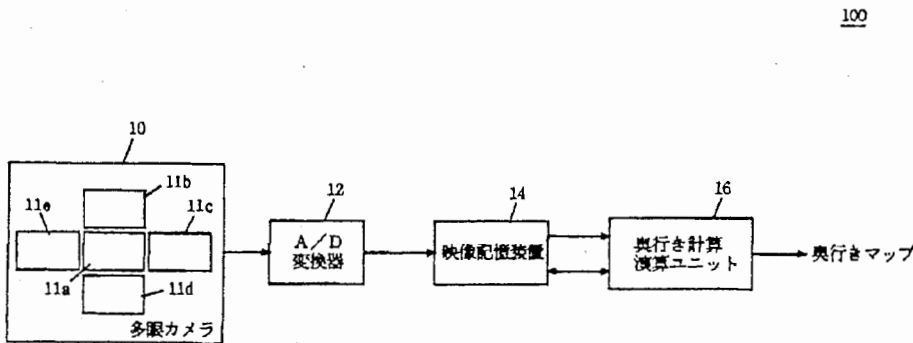
【図12】従来の両眼ステレオマッチングの幾何学的は位置を示す図である。

【符号の説明】

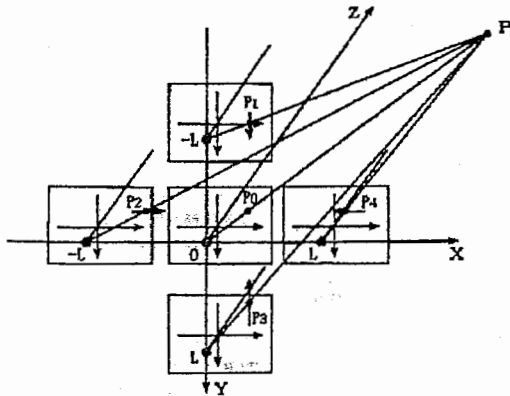
- 10 多眼カメラ
- 11 a 基準カメラ
- 11 b~e 参照カメラ
- 12 A/D変換器
- 14 映像記憶装置
- 16 奥行き計算演算ユニット
- 100 奥行き情報抽出装置

【図1】

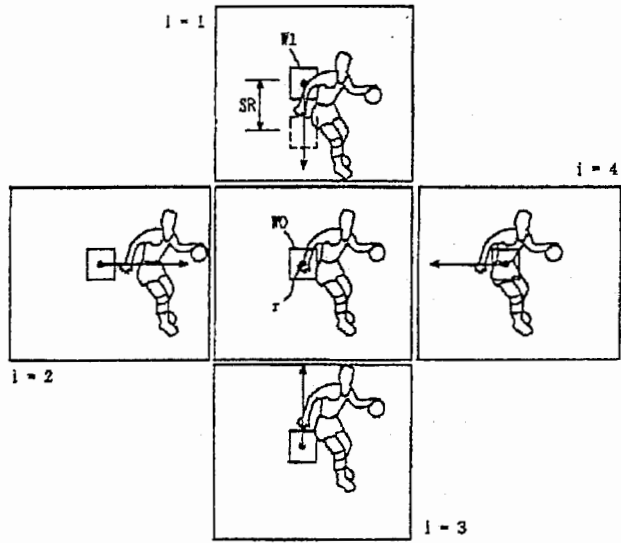
【図2】



【図3】

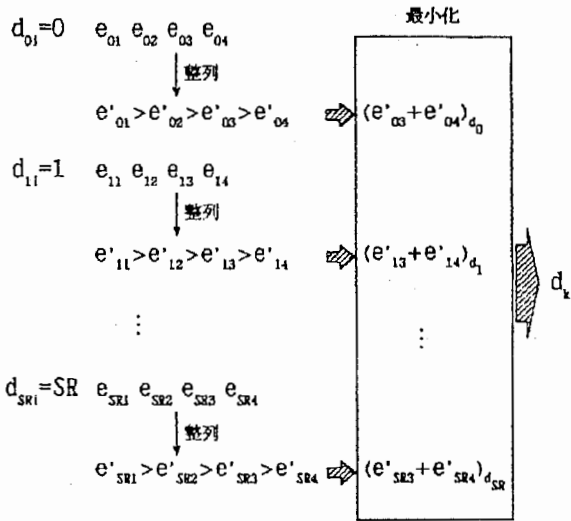
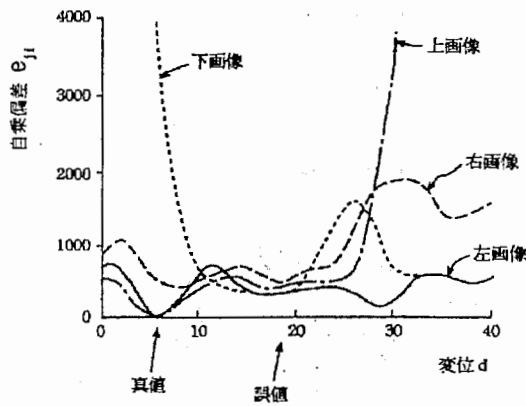


【図4】



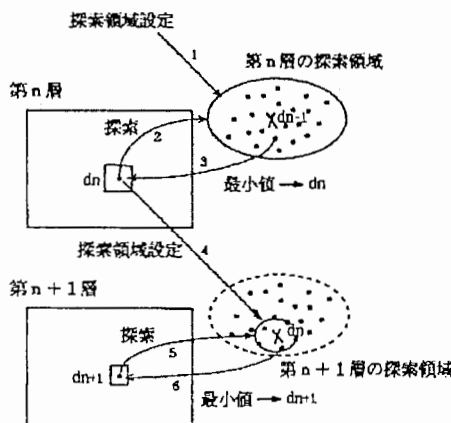
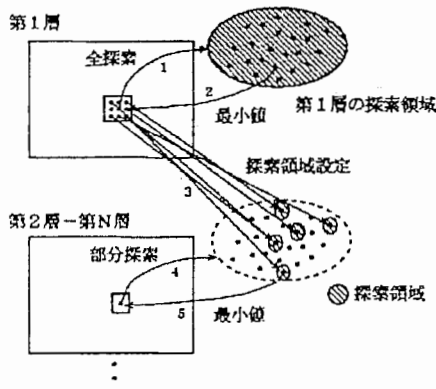
【図5】

【図6】

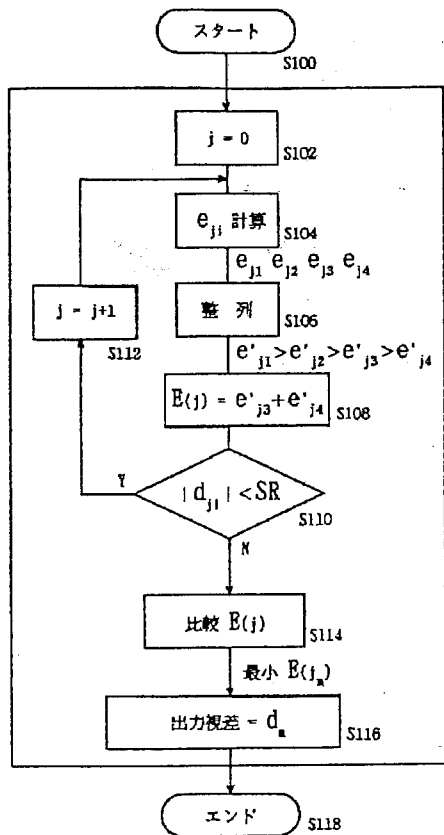


【図10】

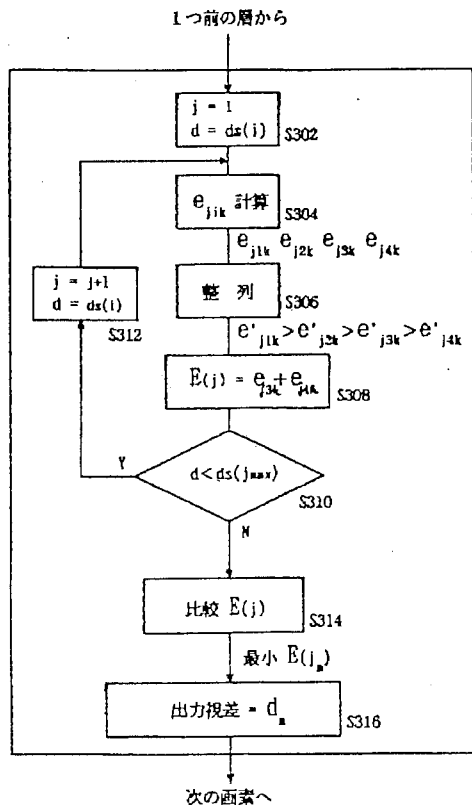
【図11】



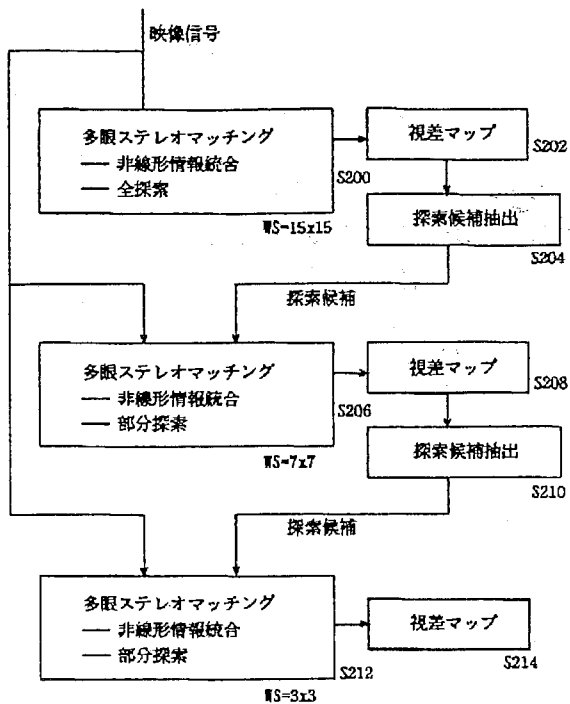
【図7】



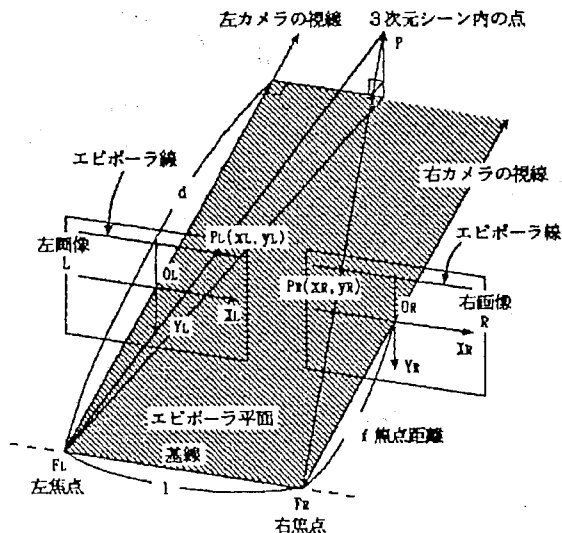
【図9】



【図8】



【図12】



Part 3.

Selected Presentations

Selected Presentations

- A. **Jong-Il Park** and Seiki Inoue, "Image expression based on disparity estimation from multiple cameras," *Proc. 3rd Joint Workshop on Multimedia Communications*, pp.7.1.1-8, Taegu, Korea, Oct.1996.
- B. **Jong-Il Park** and Seiki Inoue, "Extraction of depth map from multiple cameras and its application to image expression," *Sensing via Image Information'97*, pp.247-252, Yokohama, Japan, June 1997 (in Japanese).
- C. **Jong-Il Park** and Seiki Inoue, "Arbitrary view generation using multiple cameras," *Proc. IEEE ICIP'97*, vol.I, pp.149-153, Santa Barbara, USA, Oct. 1997.
- D. **Jong-Il Park** and Seiki Inoue, "Real-image-based virtual studio," *Proc. of the 1st Intl' Conf. on Virtual Worlds '98*, pp.117-122, Paris, France, July 1998.
- E. **Jong-Il Park** and Seiki Inoue, "Image-based view rendering for 3D visual communications," *Proc. of Intl' Workshop on Very Low Bitrate Coding - Visual Content Analysis and Coding*, Urbana-Champaign, USA, Oct. 1998.
- F. **Jong-Il Park**, Seiki Inoue, and Yuichi Iwadate, "Estimating camera parameters from motion vectors of digital video," *IEEE Multimedia Signal Processing'98*, pp.105-110, Redondo Beach, USA, Dec. 1998.

< JW-MMC'96 >

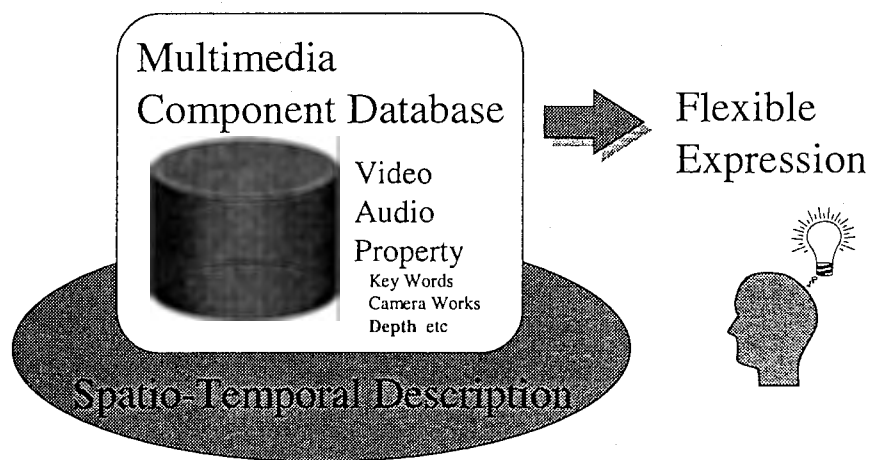
Image Expression Based on Disparity Estimation from Multiple Cameras

Jong-Il Park and Seiki Inoue

ATR Media Integration &
Communications Research Laboratories

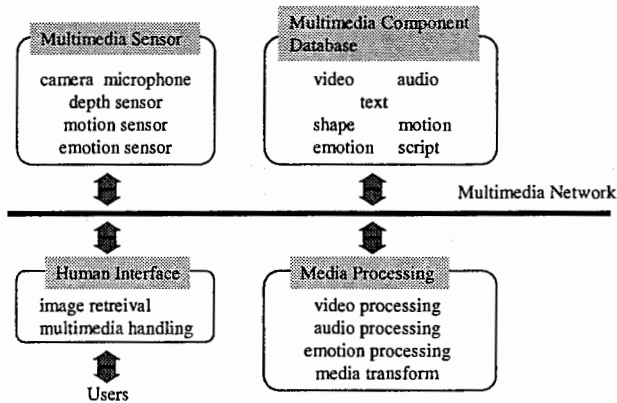
ATR

Toward Flexible Expression



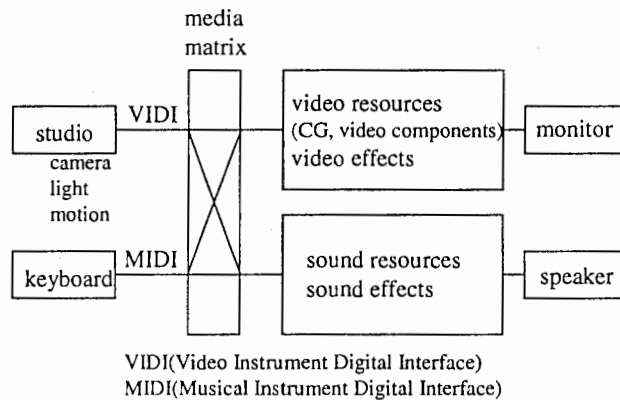
ATR

IMAGE Expression System



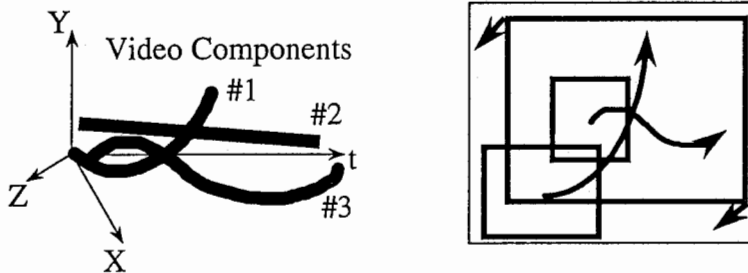
ATR

COMI&CS (Computer Organized Media Integration & Communication System)



ATR

Spatio-Temporal Description

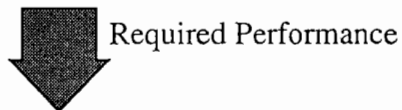


- Spatial property → Depth
- Temporal property → Motion

ATR

Use of Depth Information

- Video composition
- 3D video effects
- Automatic scene description etc.



Shape-preserving depth map

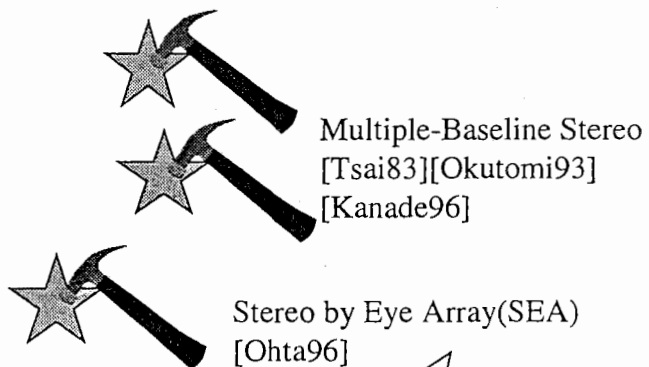
- = Occlusion-free extraction of disparity map
- = Accurate estimation near depth discontinuity

ATR

Problems in Stereo Matching

- **Lack of texture**
 - ◆ Low reliability
- **Repetitive pattern**
 - ◆ Multiple local minima
- **Occlusion**
 - ◆ Around depth discontinuity
 - ◆ Insolvable problem with 2-view stereo

ATR



Active Approach in Attitude

ATR

Still Some Questions....



■ Effect of Window Size

◆ Large Matching Window

- ✦ Reliable in almost area but noisy around depth discontinuity

◆ Small One -> vice versa

Trade-off between object boundary and other area

■ Too many cameras for occlusion-free estimation (eg. at least 9 for SEA)

ATR

Approach

■ Focus on getting

Occlusion-free depth map

- ◆ From multiple cameras
- ◆ Controlling the effect of window size

■ Considering feasibility

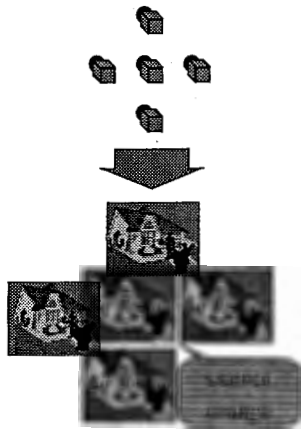
Hierarchical scheme

- ◆ Minimize the number of cameras
- ◆ Signal-level processing

5 cameras

ATR

Problem Statement

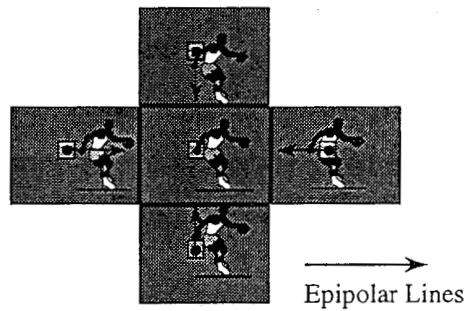


- 5 cameras
 - Equal baseline stretch
 - Parallel optical axes
 - Toward getting
 - Occlusion-free
 - Less noisy
 - Dense
- disparity map of cntr img

ATR

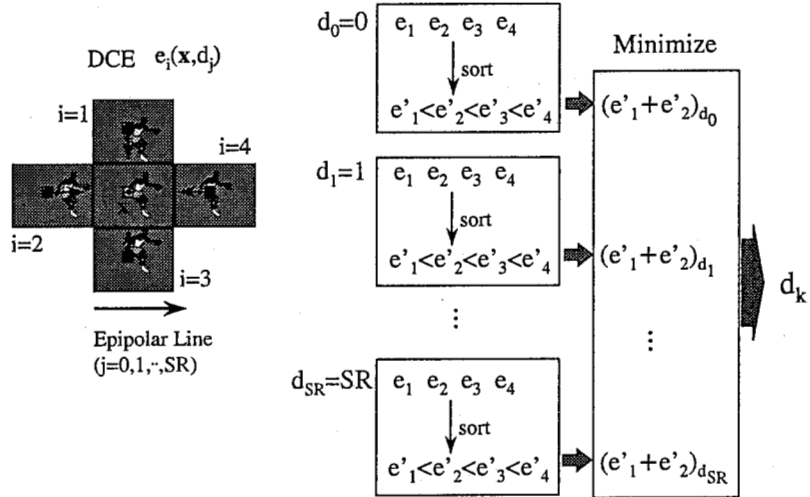
Block Matching for 4 Directions

- Pixel accuracy
- Along the epipolar lines



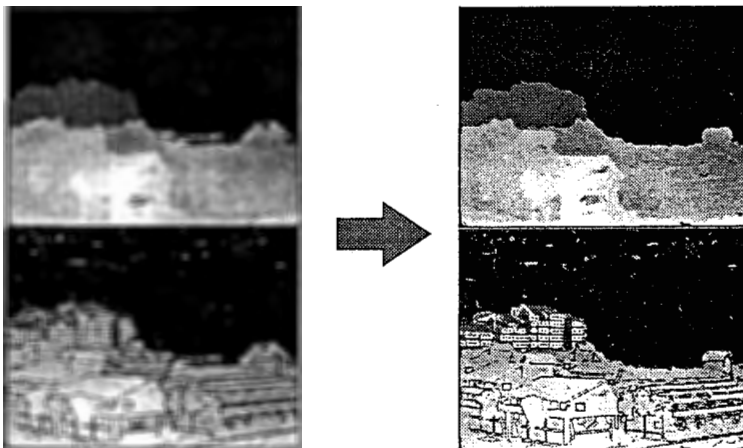
ATR

Nonlinear Collection & Summation of Displacement-Compensated-Error



ATR

Improvement by NLC&S



- block size = 7×7
- OS filter (7×7 , Yang)

ATR

Boundary Overreach

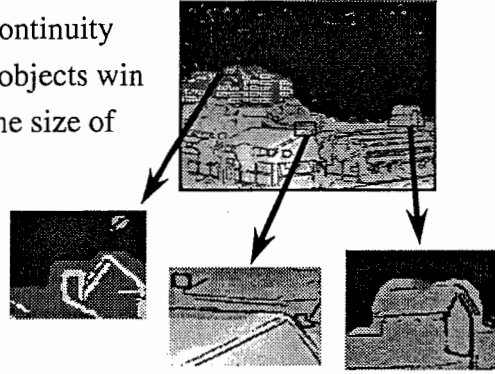


- Limitations in NLC&S

- Around disparity discontinuity
- {High-texture||Near} objects win
- Smaller than half of the size of the matching window

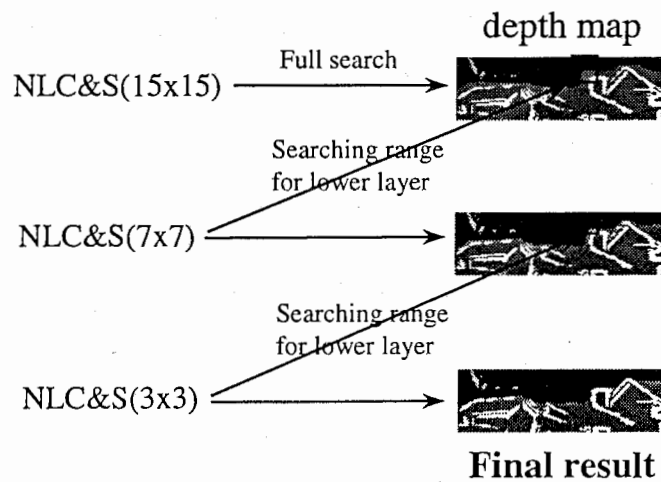


Novel Hierarchical Scheme!



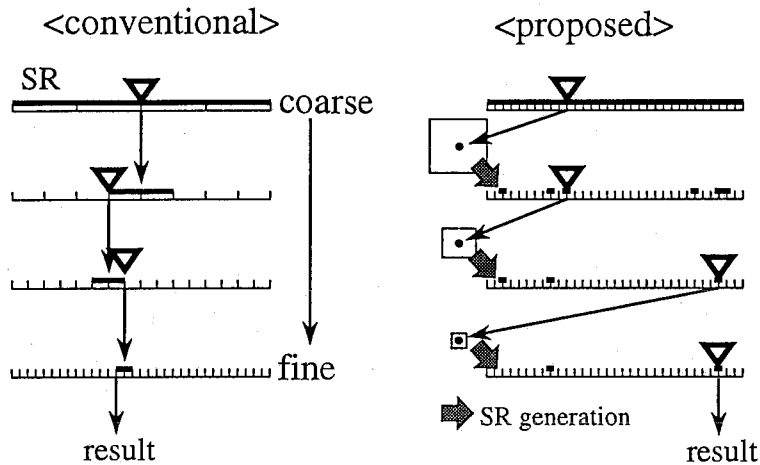
ATR

Hierarchical Scheme



ATR

Concept of Hierarchical Sceme



ATR

Some More Techniques...

■ Enhancing accuracy

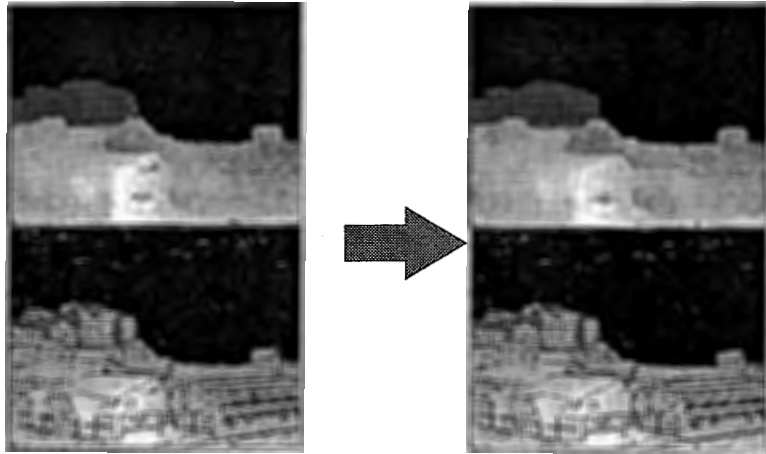
- ◆ **IF discontinuity of depth around a point**
 - ✦ Normal hierarchical processing
- ◆ **ELSE**
 - ✦ Refinement to subpixel accuracy by quadratic intpl.
 - ✦ then no more update until the end

■ Postprocessing

- ◆ Elimination of noisy points
- ◆ Adaptive order-statistics filter(7x7) [Yang89]

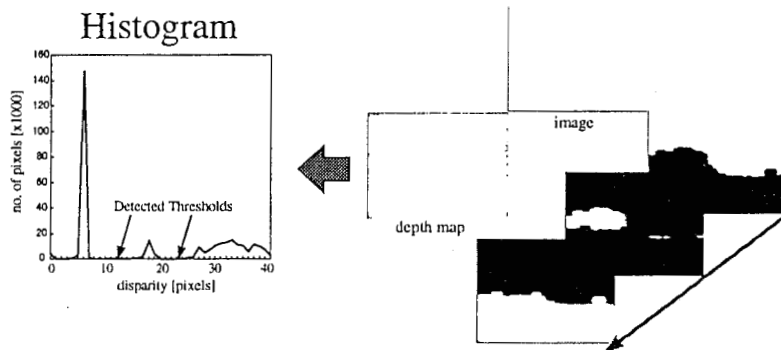
ATR

Improvement by Hierarchical Scheme



ATR

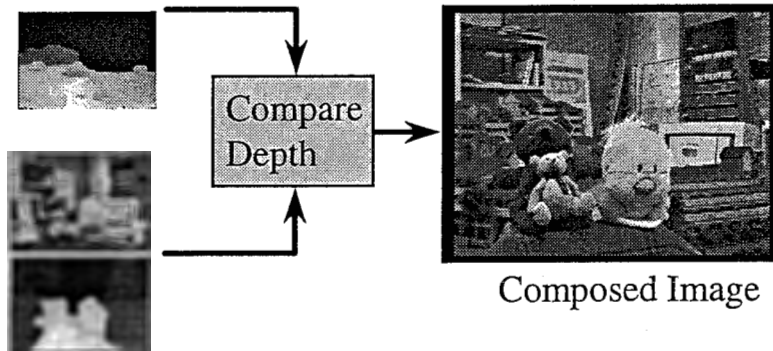
Application Layered Representation



AR

Application

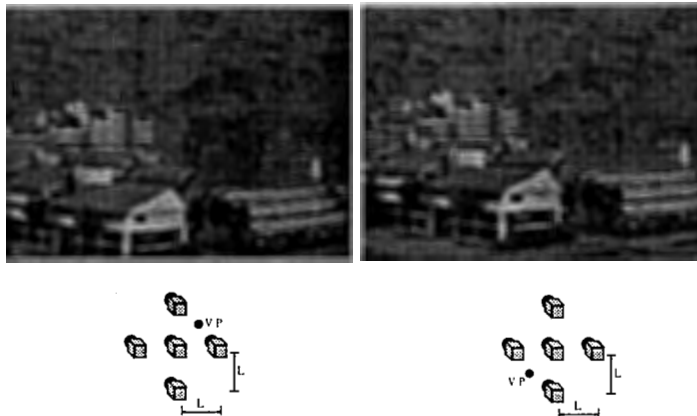
Video Composition



AR

Application

New View Generation



ATR

Application

3D Reconstruction



- Need to exploit temporal constraint with varying camera position in order to get sufficiently good quality

ATR

Concluding Remarks

- In this presentation
 - ◆ Toward hyper-realistic communications
 - ◆ Getting dense, occlusion-free depth map by **Nonlinear+Hierarchical** Scheme
 - ◆ Exploring various applications
- Future works
 - ◆ Exploit temporal redundancy in depth extraction
 - ◆ Develop user-friendly interface
 - ◆ Develop efficient computational algorithm & H/W

ATR

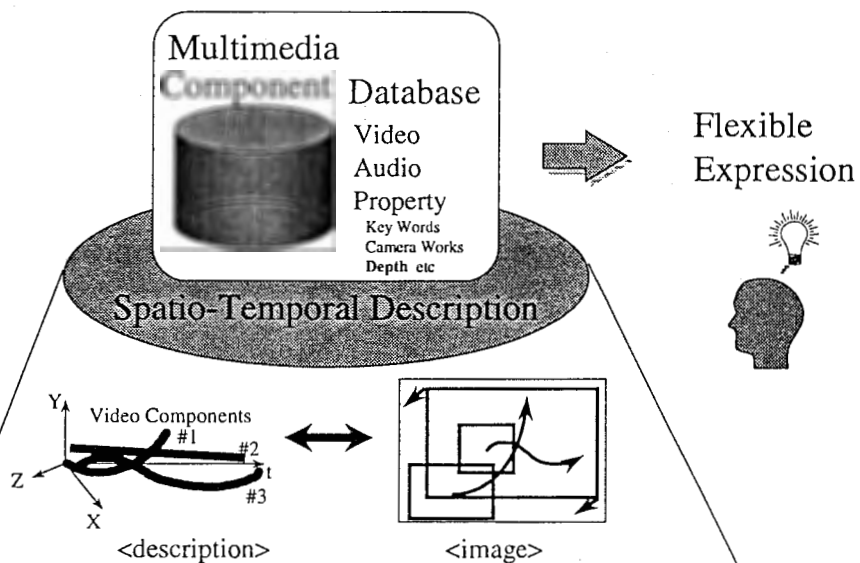
多眼カメラを用いた奥行き抽出とその映像表現への応用

ATR知能映像通信研究所

朴 鍾一、井上 誠喜

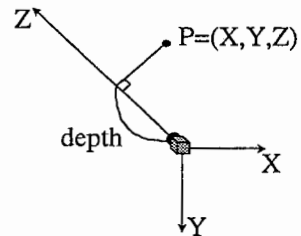
ATR

Toward Flexible Image Expression



Use of Depth Information

- Video composition
- 3D video effects
- Automatic scene description
- Arbitrary view synthesis etc.



Required Performance

Shape-preserving depth map

- = Occlusion-free extraction of disparity map
- = Accurate estimation near depth discontinuity

Approach

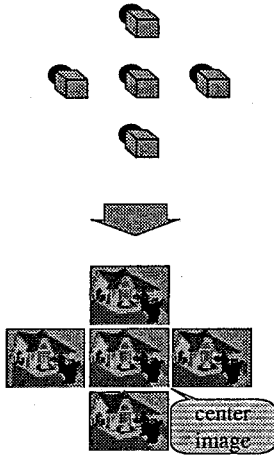
- Focus on getting

Occlusion-free depth map

- Techniques

- ◆ Using 5 cameras → occlusion overcoming
- ◆ Nonlinear summation → noise immunity
- ◆ Hierarchical scheme → computation efficiency

Problem Statement



- **Maximul Likelihood Framework**

- no *a priori* assumption on disparity field
- rectangular local support

- **Area-based method**

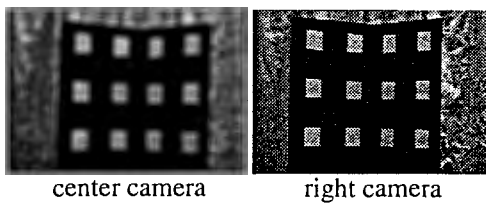
- pixelwise dense mapping

- **Strong calibration**

- almost no restriction on camera geometry

Camera Calibration (I)

- **Strong calibration using Tsai's algorithm**



<Calibration pattern>

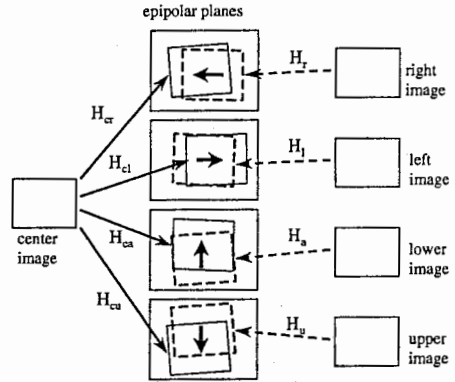
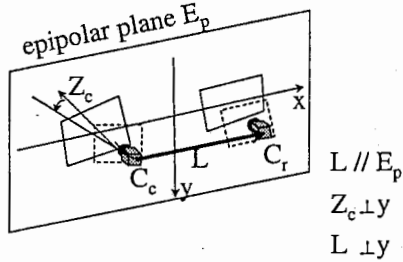


Camera
parameters

- **Epipolar plane reconstruction for each camera pair using the camera parameters (searching : 2D --> 1D)**

Camera Calibration (II)

Rectification



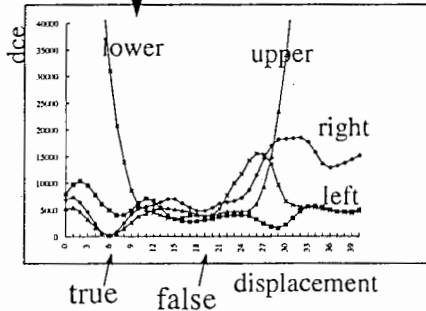
■ Using cameras and lenses of the same specification

- Center camera : internal + external parameters
- Inspection cameras : external parameters only
(assume the same internal paras. as the center camera)

Occlusion Overcoming Strategy



Point (154, 164)



- Pixelwise adaptation
- Using 3 cameras out of 5 cameras

< Possible no. of patterns = 6 >

- Used patterns



- used
- discarded

- The following patterns are not used



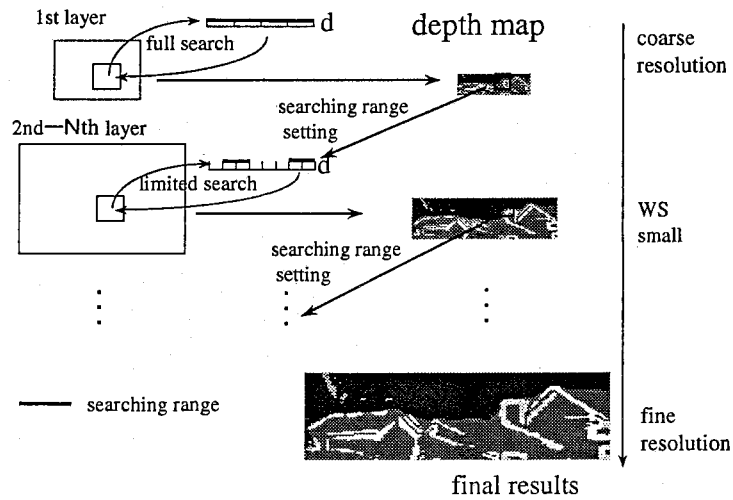
reason:



no change of SSD

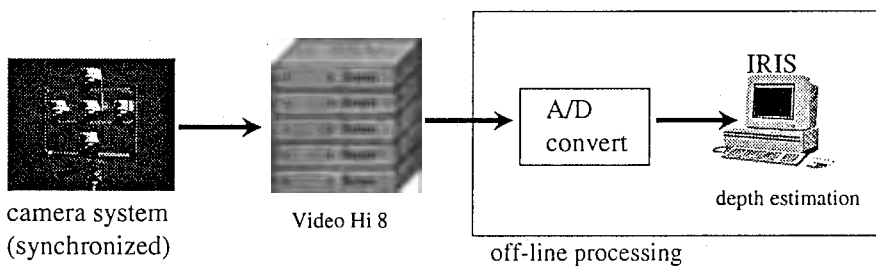
Hierarchical Scheme

- coarse-to-fine approach -



Experimental Results

<Configuration of Experiments>



- Baseline stretch $L = \text{about } 20\text{cm}$
- Camera
 - $1/2''$ CCD
 - Focal length $F = 6\text{mm}, 25\text{mm}$

ATR

Experimental Results

– Still Images –



center image



depth map

- 6mm lens
- 1/2" CCD
- image : 640x480 pixels
- depth map: 320x240 pixels
- barrel-type lens distortion
- variance matching measure
- 4-layer hierarchy
(local support 5->5->5)
- searching range : 0-70 pixels
- adaptive smoothing



depth map + edge

ATR

Experimental Results

– Moving Images –

Conditions

- 25mm lens
- 1/2" CCD
- image : 640x480 pixels
- depth map: 320x240 pixels
- baseline stretch = about 20cm
- variance matching measure
- 4-layer hierarchy
(local support 5->5->5)
- searching range : 90-150 pixels
- adaptive smoothing

See video.

ATR

Application

Video Composition

image #1

depth #1



image #2

depth #2



Depth Comparison

Z-keying



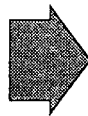
Merged
Image

<Required Performance>
High Accuracy
around object boundary

ATR

Application

Arbitrary View Synthesis



$T=(0.7L,0.3L,6.0L)$

<Required Performance>

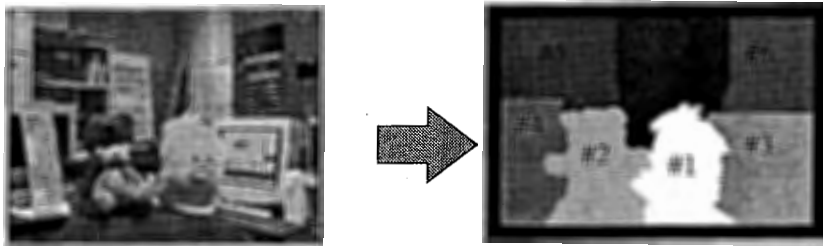
Consistency of depth map + Correctness of edge

AR

Application

Scene Segmentation & Description

- Important in the context of object-based approach
- Addition of depth information
--> more chance of meaningful segmentation



K-means algorithm
using R,G,B,X,Y,Z

AR

Concluding Remarks

- On depth estimation error
 - ◆ Cause: 1) noise, 2) motion blur, 3) calibration error, 4) ill-posedness
 - ◆ Still images :1)-3) can be avoided -> good result
 - ◆ Moving images : 3) is dominant in our experiments
- Performance evaluation of depth extraction
 - ◆ Application dependent
 - ◆ Further improvement is requested

Arbitrary View Generation from Multiple Cameras

ATR Media Integration & Comms. Res. Labs.

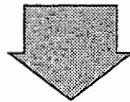
Jong-Il Park and Seiki Inoue

View Synthesis for Complex Scenes

- Complex scenes
 - ✧ Explicit handling of **occlusion** is necessary
- Related work
 - ✧ Katayama et al.(1995)
 - ✧ McMillan&Bishop(1995)
 - ✧ Kanade et al.(1995)
 - ✧ Seitz&Dyer(1996)
 - ✧ Scharstein(1996)

Acquisition of Depth Map

- ❑ Required performance
 - ✧ Sharp transition at depth discontinuities



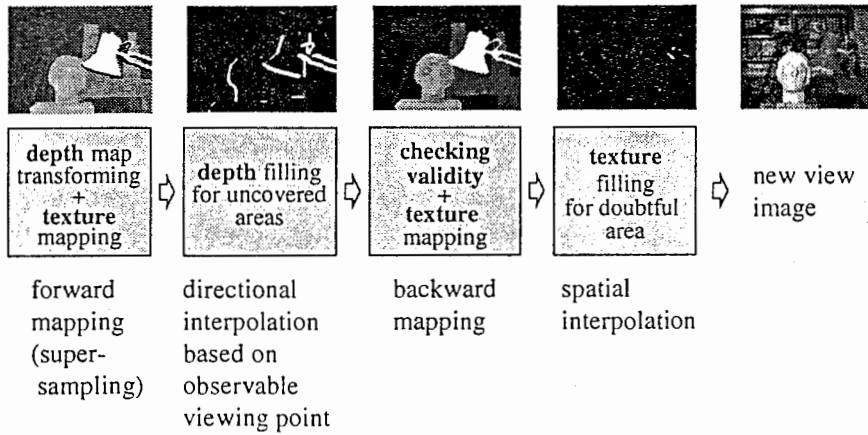
- ❑ Hierarchical method with occlusion-overcoming technique
(Park&Inoue,1997)

Overview of View Synthesis

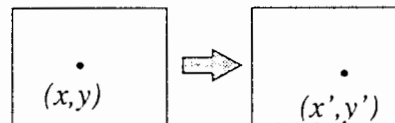
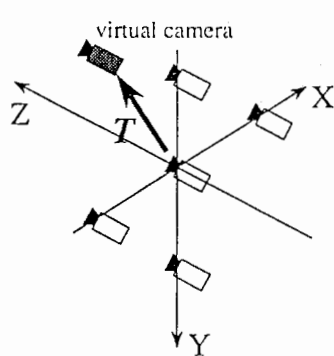
- ❑ STEP 1: Pure Translation
 - 1.1 Transformation of depth map and color
 - 1.2 Filling depth map for uncovered areas
 - 1.3 Color mapping for the uncovered areas
- ❑ STEP 2: Zoom and Pure Rotation

Flow of View Synthesis

– Pure Translation –



STEP 1: Coping with Pure Translation

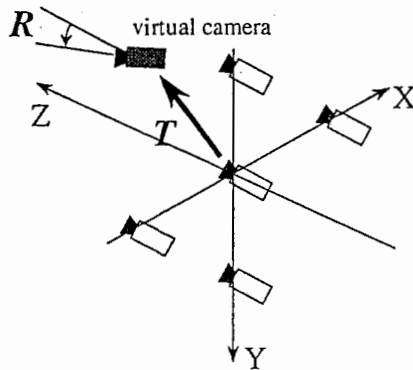


□ Transformation equation

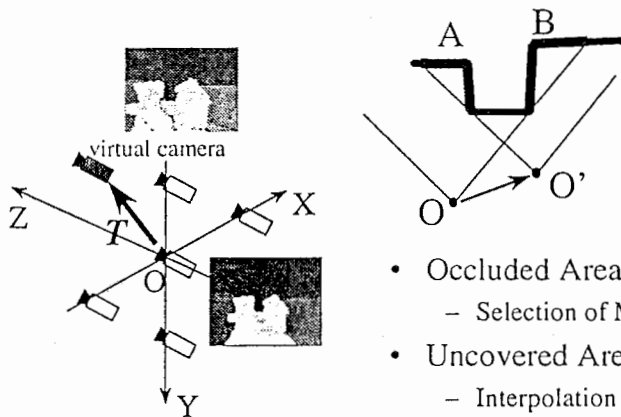
$$\begin{cases} x' = \frac{1}{1 - \Delta Z / Z(x, y)} \left(x - \frac{F \Delta X}{Z(x, y)} \right) \\ y' = \frac{1}{1 - \Delta Z / Z(x, y)} \left(y - \frac{F \Delta Y}{Z(x, y)} \right) \end{cases}$$

where $T = (\Delta X, \Delta Y, \Delta Z)$

Concept of View Synthesis



Transformation of Depth Map



- Occluded Area : A (Folds)
 - Selection of Minimum Depth
- Uncovered Area : B (Holes)
 - Interpolation

- Assume opaque objects
- Forward mapping with super-sampling

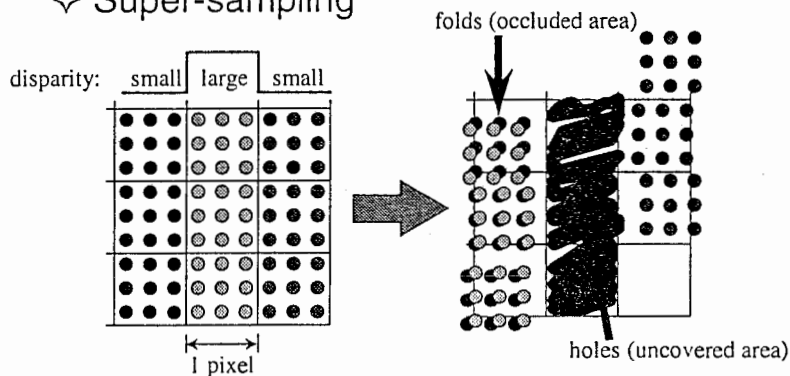
Forward Mapping

□ Why?

- ◇ Inverse mapping not known explicitly

□ How?

- ◇ Super-sampling



Filling Uncovered Areas

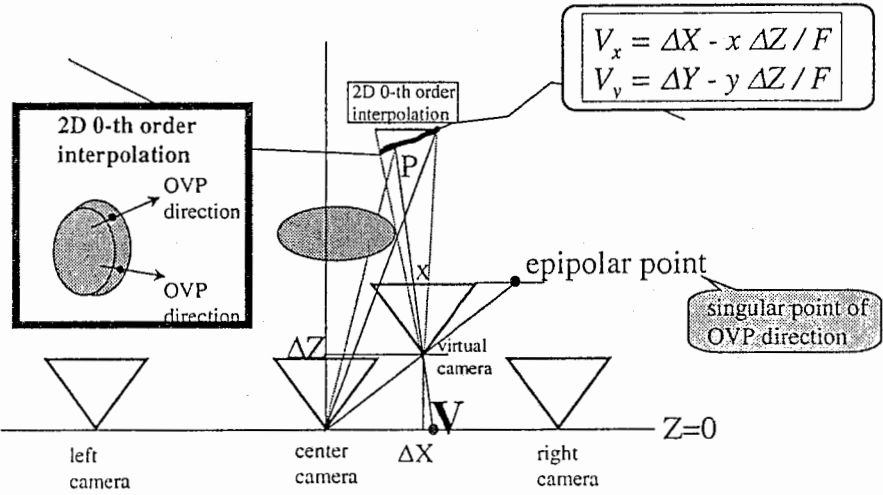
□ Interpolation of depth map

- ◇ Directional 0-th order interpolation
- ◇ Based on the Observable Viewpoint(OVP)

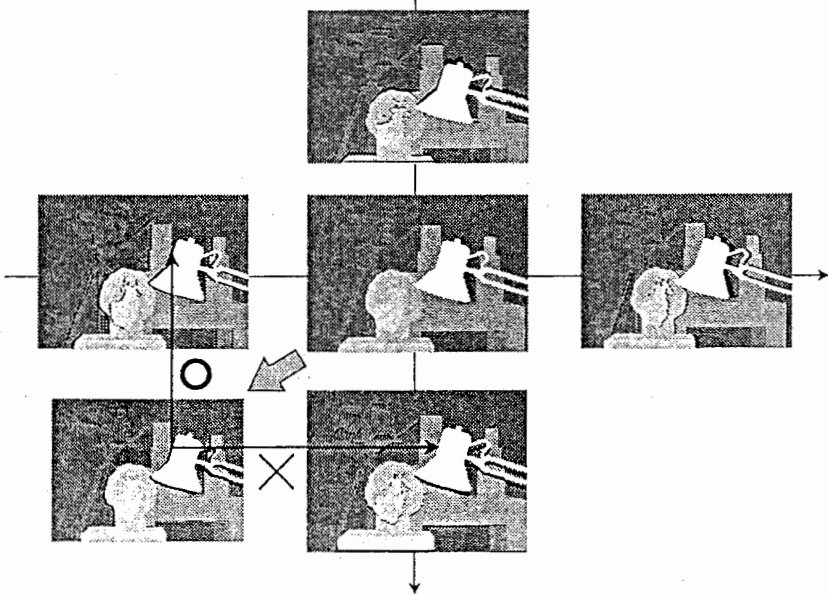
□ Interpolation of color

- ◇ Using the interpolated depths
- ◇ Inverse mapping
- ◇ Reliability check by back-mapping to the depth maps of secondary cameras

Observable Viewpoint (OVP)



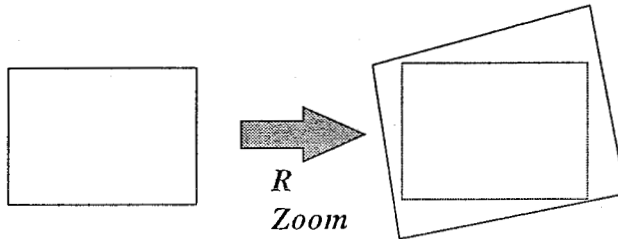
Validity Check



STEP 2: Coping with Pure Rotation and Zoom

- 2D geometric transformation
- ◇ Explicit inverse mapping exists

$$p' = f(p, R, zoom) \leftrightarrow p = f(p', R^{-1}, zoom^{-1})$$



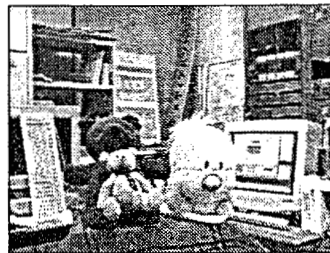
Experiments

Image #1



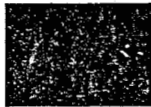
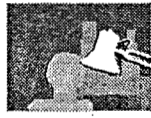
- In multi-view image database (University of Tsukuba)
- True disparity map known
- Disparity range: ~ 14 pixels
- 384x288 pixels

Image #2



- Shot by SONY 3CCD camera (DXC-930)
- Disparity estimated by occlusion-overcoming technique
- Disparity range: ~ 50 pixels
- 640x480 pixels

Experimental Results



$T=(-30,30,40)$.

$L=50, F=6, \text{CCD}=1/2''$ (assumed)

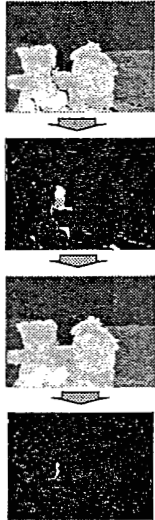
Experimental Results



$T=(0,0,-380)$.

$L=50, F=6, \text{CCD}=1/2''$ (assumed)

Experimental Results



$T=(25,15,250)$, No rotation and zoom.
 $L=50$, $F=7.5$, $CCD=1/2''$

Experimental Results



$T=(-25,-15,-250)$,
Rotated by 5 degrees and zoomed in.
 $L=50$, $F=7.5$, $CCD=1/2''$

Concluding Remarks

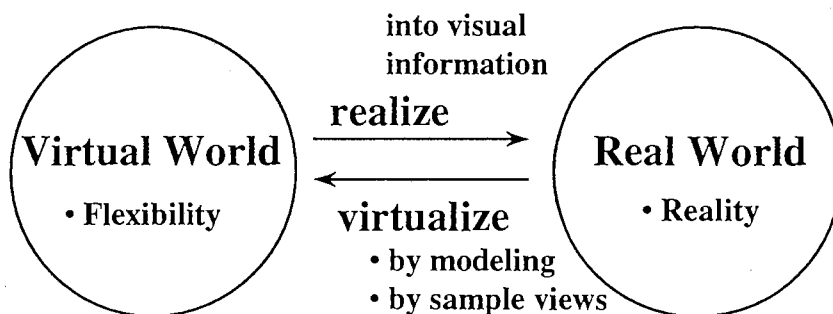
- Proposed is a view synthesis method
 - ◇ for complex scenes
 - ◇ from 5 cameras
 - ◇ based on estimation of depth map
 - ◇ explicit handling of occlusion
- Future work
 - ◇ Frame-rate rendering given a depth map
 - ◇ More accurate estimation of depth map
 - ◇ Development of application
 - Fully real-image-based virtual studio
 - View-dependent stereoscopic display

Real-Image-Based Virtual Studio

ATR Media Integration & Communications
Research Laboratories

Jong-Il Park and Seiki Inoue

Virtual Environments in Visual Communications



* CG: virtualizing by modeling and
visualizing by rendering

Video Compositing in VR

Subjective VR

- ◇ Interaction with virtual world
- ◇ Eg. Krueger's VIDEOPLACE

Objective VR

- ◇ Realistic imaging of complex reality
- ◇ Eg. Hayashi et al.'s Virtual Studio

Image
Expression
Room

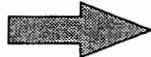


➔ **All are CG + Real video**
(CGs are used to represent virtuality)

Real-Image-Based VS

Ordinary VS

BG: CG
FG: Real images



RIB VS

BG: Real images
FG: Real images

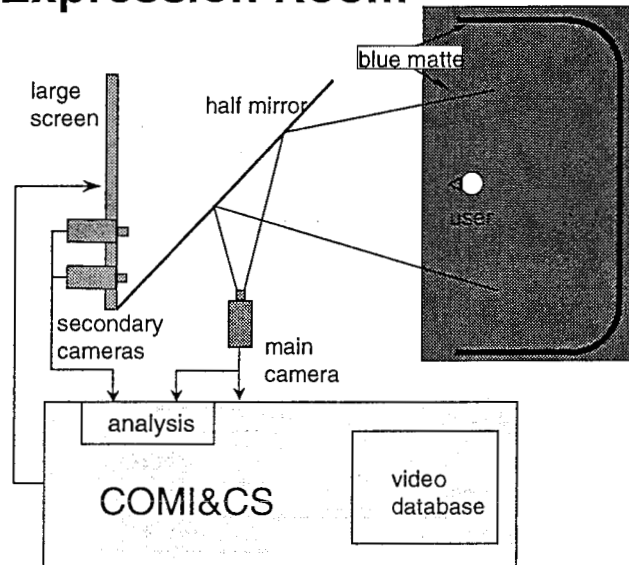
Virtualized Reality

- ◇ Virtualizing real worlds by sample views
- ◇ Visualizing virtuality by view synthesis
- ◇ Technology: **Stereo Matching + View Synthesis**

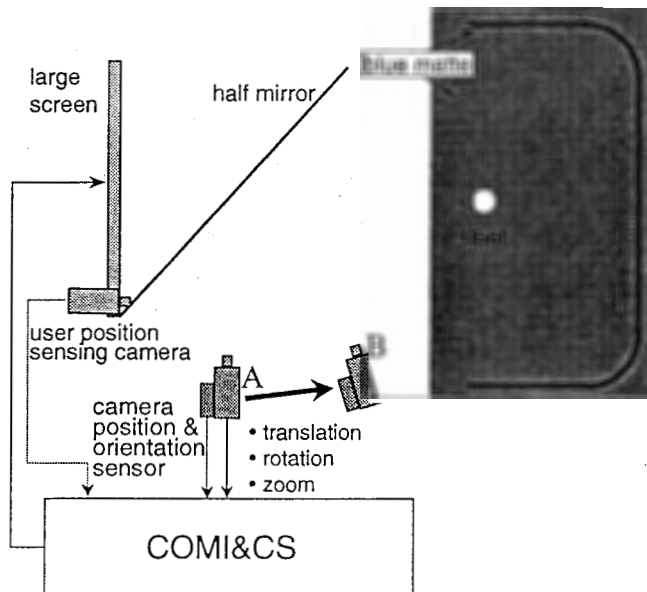
In View Synthesis, emphasis on

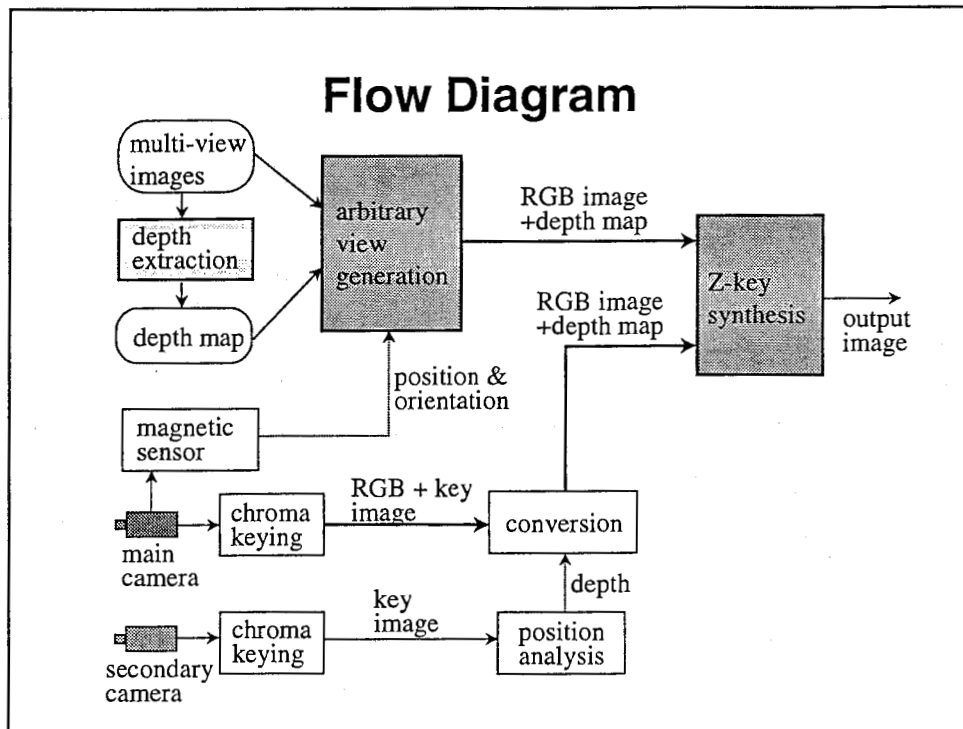
- ◇ Rendering complex backgrounds
- ◇ Simultaneous generation of depth map

Image Expression Room



IE Room with Camera Motion





Experimental Results

Composed Images



Before camera motion



After camera motion

Discussion

- Problems at the current implementation(v.0.1)
 - ◇ Low precision of the mechanical sensor
 - about 0.5 degree
 - should be enhanced to 0.01 degree
 - ◇ Not cope with moving images
 - view synthesis should be processed in video-rate
 - ◇ Limited volume and direction of acceptable views
 - fundamental limitation
 - wide-angle imaging and zooming -> low resolution
 - panoramic or omnidirectional imaging required

Concluding Remarks

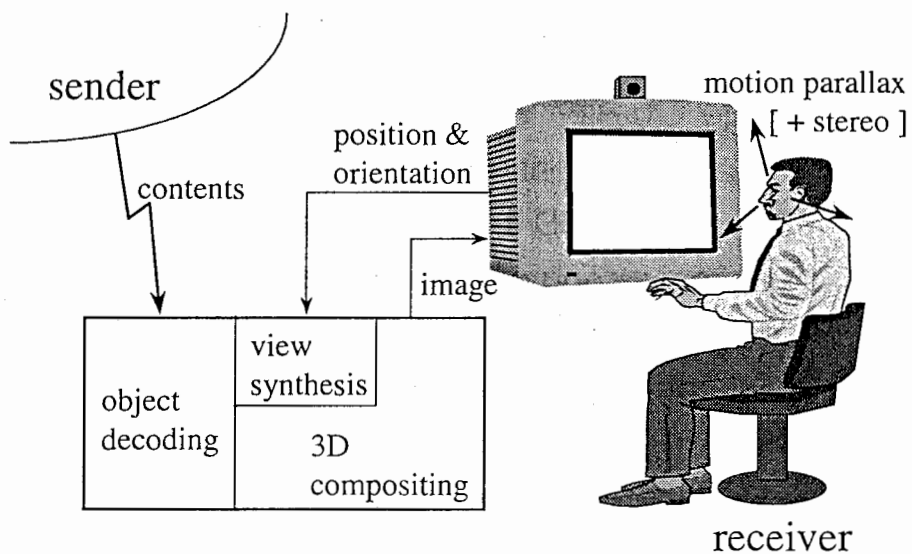
- Real-Image-Based Virtual Studio
 - ◇ Based on a View Synthesis Method
 - for complex scenes
 - from 5 cameras
 - based on estimation of depth map
 - explicit handling of occlusion
 - ◇ Prototype implementation at IERoom
- Future work
 - ◇ Enhancement of sensor precision
 - ◇ Realtime implementation of view synthesis algorithm
 - ◇ Broadening range of views

Image-Based View Rendering for 3D Visual Communications

ATR Media Integration & Comms. Res. Labs.

Jong-Il Park and Seiki Inoue

3D Visual Communications



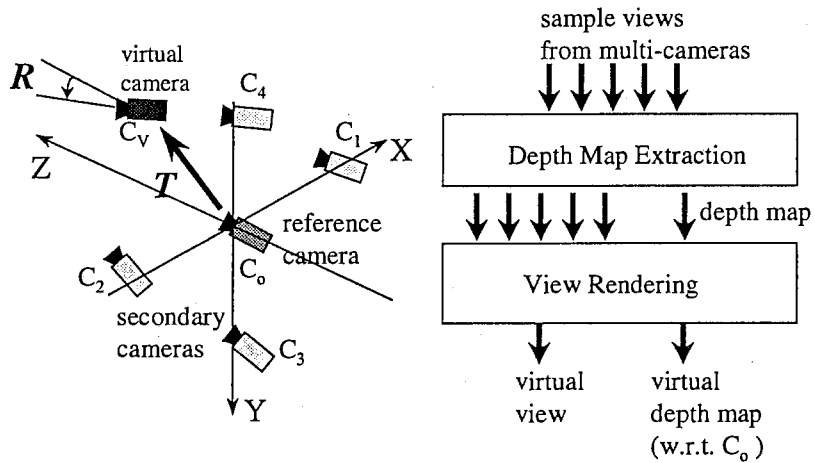
View Rendering from Multiple Sample Views

- ❑ Image-Based View Rendering
 - ✧ Without polygon models
 - ✧ Using depth map
- ❑ Approach
 - ✧ Multiple cameras (5 cameras)
 - ✧ Extracting sharp depth map
 - ✧ Properly handling uncovered areas

Improvement over Our Previous Work (ICIP'97)

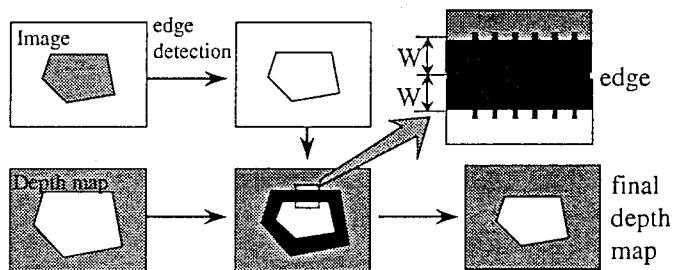
- ❑ Image Quality
 - ✧ one-step algorithm
 - by merging 3D and 2D step in [ICIP97]
 - ⇒ resampling once (cf. twice)
 - ⇒ small loss of image boundary
 - ✧ Sharper depth map by postprocessing
- ❑ Efficiency
 - ✧ only 2-pass (cf. at least 3-pass in [ICIP97])

Overview of View Rendering



Acquisition of Depth Map

- Required performance
 - ◇ Sharp transition at depth discontinuities
- Hierarchical method with occlusion-overcoming technique (Park&Inoue, 1997)
- Post-processing



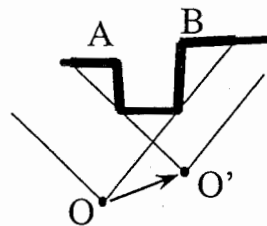
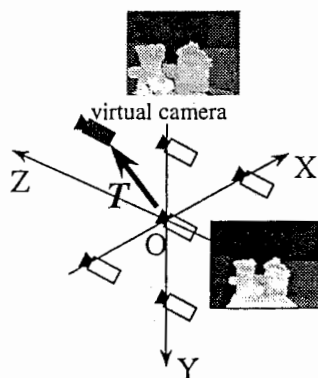
Rendering A Virtual View

Given a virtual view point,

- PASS 1: Forward Mapping
 - 1.1 Transformation of depth map and color

- PASS 2: Interpolating Uncovered Areas
 - 2.1 Filling in depth map for uncovered areas
 - 2.2 Backmapping color using the guessed depth
 - 2.3 Spatial interpolation

PASS 1: Forward Mapping

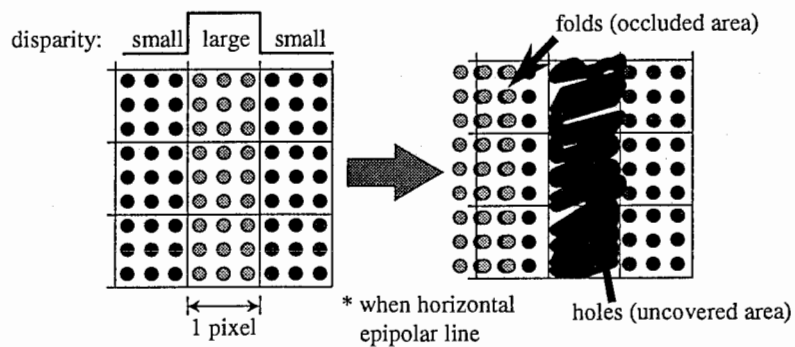


- Occlusion Area : **A (Folds)**
 - Selection of Minimum Depth
- Uncovered Area : **B (Holes)**
 - Interpolation

- Assume opaque objects
- Super-sampling

Super-sampling

- Why Forward Mapping?
 - ◇ Inverse mapping not known explicitly
- Super-sampling
 - ◇ To avoid resample crack



PASS 2: Interpolating Uncovered Area

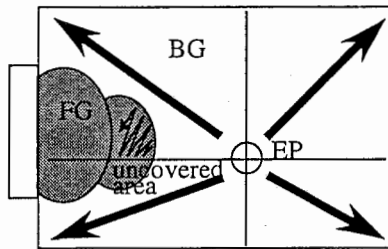
- Interpolation of depth map
 - ◇ Directional 0-th order interpolation
 - ◇ Based on the epipolar geometry
- Interpolation of color
 - ◇ Back-mapping using the interpolated depth
 - ◇ Camera selection by Observable Viewpoint
 - ◇ Validity Check

Scan Ordering

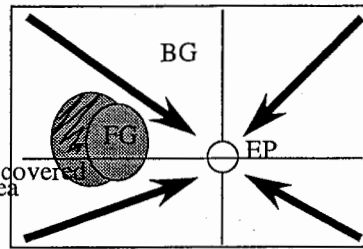
□ Based on epipolar point(EP) [Mark&Bishop97]

Diverging epipolar point
: Scan from EP

Converging epipolar point
: Scan toward EP



Camera translated forward



Camera translated backward

Illustration of PASS 2 Pointwise Operation

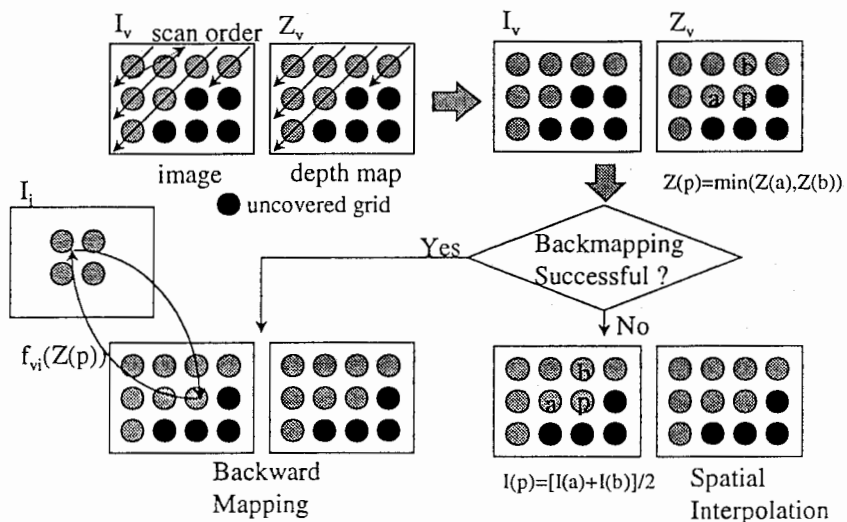
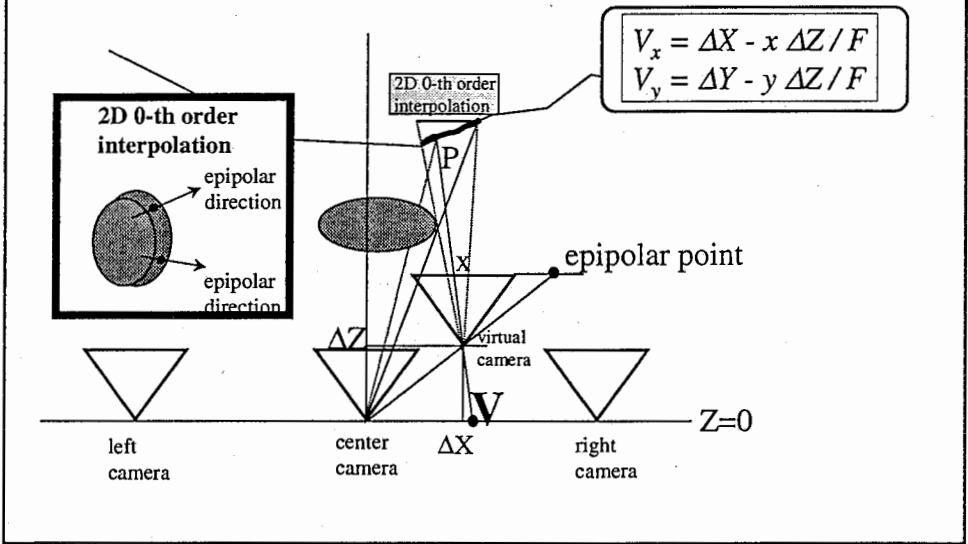
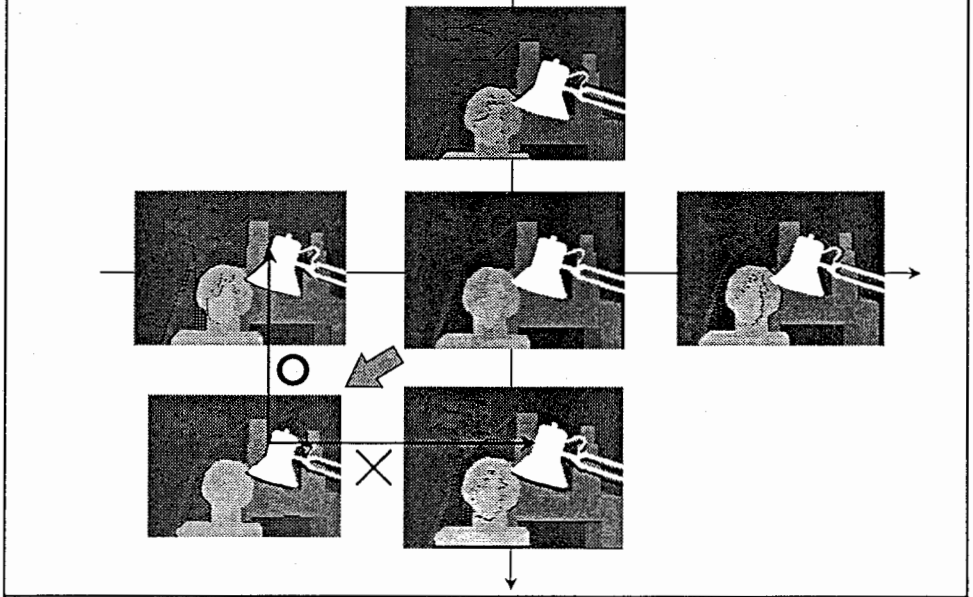


Illustration of Observable Viewpoint



Checking Mapping Validity



Experimental Results



center view I_0



disparity map D_0



D_v before interpolation

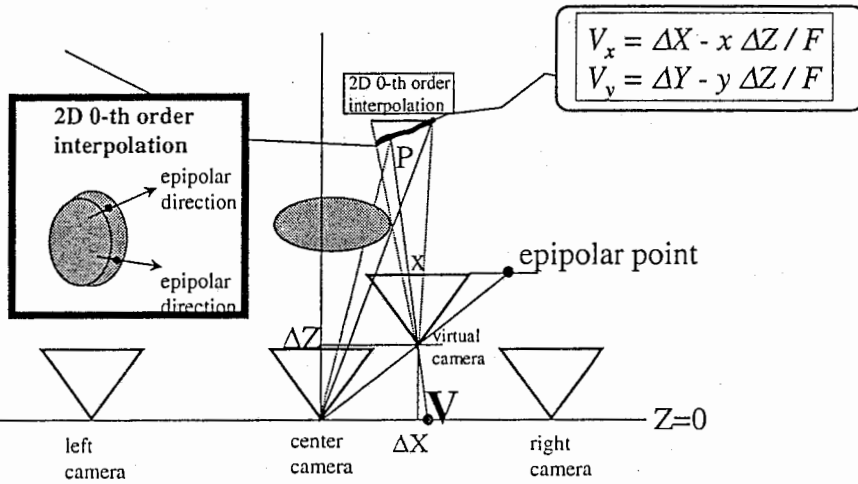


virtual view I_v

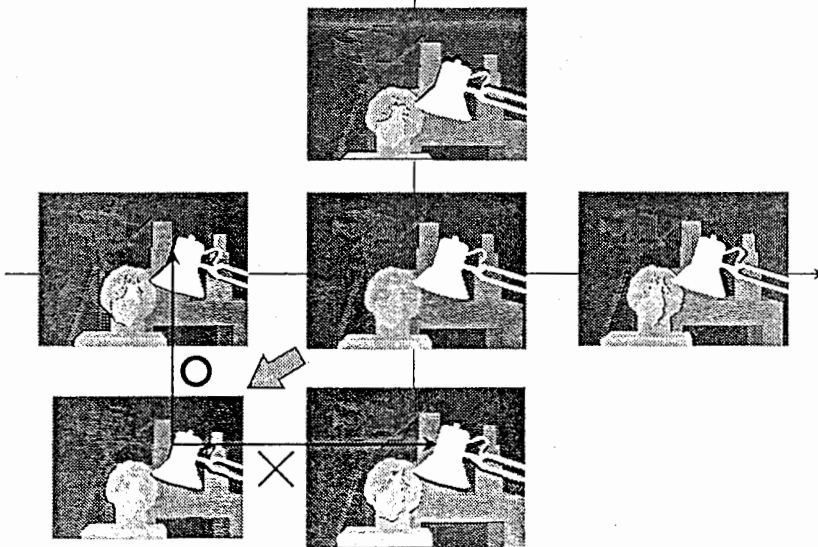
Effect of Backmapping



Illustration of Observable Viewpoint



Checking Mapping Validity



Experimental Results



center view I_0



disparity map D_0



D_v before interpolation



virtual view I_v

Effect of Backmapping

With
Backmapping



Without
Backmapping
[Mark&Bishop97]



320x240, dynamic

640x480, static

Effect of Post-Processing of Depth Map



Without post-processing



With post-processing

- Canny edge detector ($s=1.5$)
- post-processing width = 3 pixel

Discussion

- Effect of using backward mapping
 - ◇ considerable quality enhancement
- Effect of efficient implementation
 - ◇ quality is maintained
- Effect of post-processing of depth map
 - ◇ virtual view image: trade-off
 - one-pixel boundary overreach is desirable
 - ◇ virtual view depth map: quality improved

Concluding Remarks

- Image-based view rendering method
 - ◇ Using multiple cameras
 - ◇ Based on estimation of depth map
 - ◇ Explicit handling of uncovered area
 - ◇ Efficient algorithm
- Future work
 - ◇ Frame-rate rendering at full NTSC resolution
 - ◇ Coping with wider field of view(eg. panoramic view)
 - ◇ Development of application
 - Real-image-based virtual studio [VW98]
 - View-dependent display

Further Experimental Results

Available online from

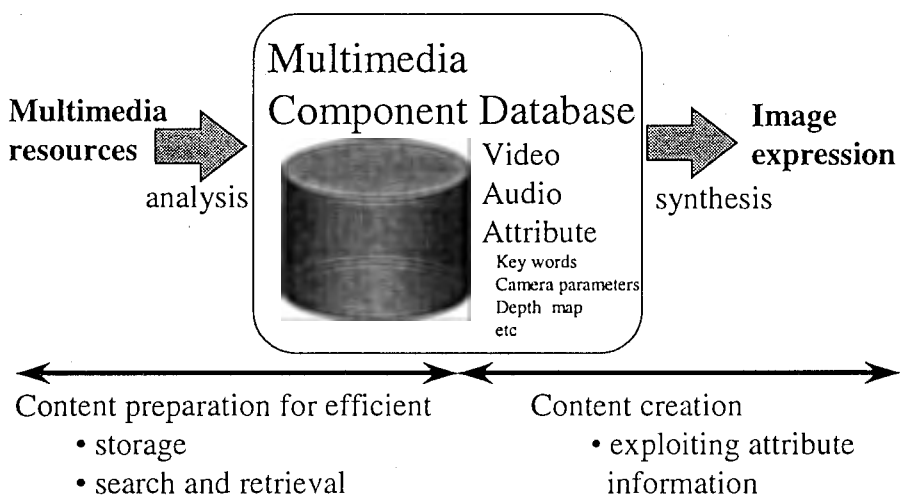
http://www.mic.atr.co.jp/~pji/view_synth.html

Estimating Camera Parameters from Motion Vectors of Digital Video

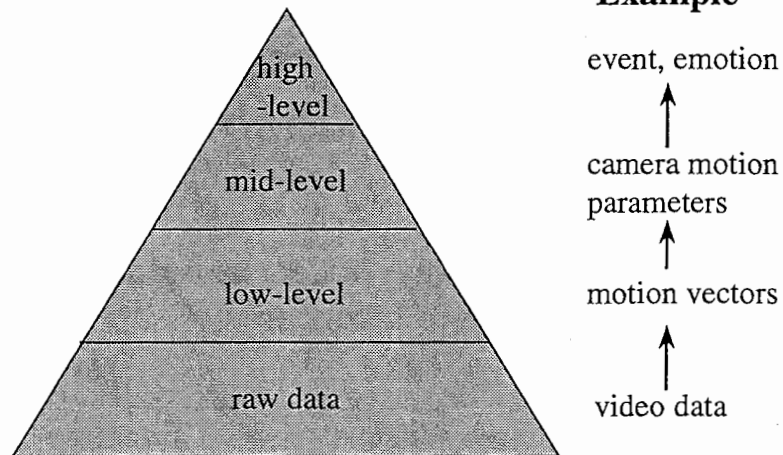
ATR Media Integration & Communications
Research Laboratories

Jong-II Park, Seiki Inoue, Yuichi Iwadate

Multimedia DB in ATR



Description Level



Why CDP?

(Compressed-Domain Processing)

- Digital resource
 - All resources are destined to be digital
- Efficiency
 - information itself in compressed-domain
 - no need of full decoding

Camera ME in CDP

- Resource: MPEG, H.263 etc. are prevalent
- Efficiency
 - Motion vectors are contained in compressed domain
 - Getting MV is major computational burden in camera ME

Relevant Work - Camera ME

- **Direct estimation** from image sequence
 - Sawhney et al.[‘95]: robust M-estimation
- **Indirect estimation** via feature matching
 - Park&Lee[‘94]: LS data fitting
 - Morimoto&Chellappa[‘97]: Ext. Kalman filter

➡ **Camera ME in CDP
is inherently based on
indirect approach**

Inherent Difficulties

in Camera ME in Compressed Domain

- No information on reliability of MVs
- Local motion estimation depends on encoder
 - RD optimization
 - accuracy of MVs

Camera Motion Estimation

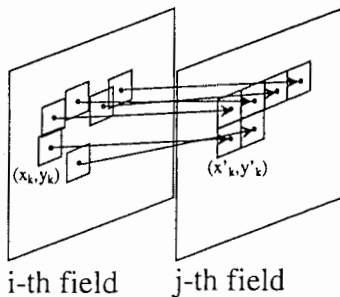
- Robust least-square estimation
 - 2-view formulation since tracking is impossible
 - 4 parameters(1 for zoom, 3 for 3D rotation)

$$x' = h_x(x, y, R, f), \quad y' = h_y(x, y, R, f)$$
 - outlier rejection for object motion/bad MVs
- Solution method
 - Method 1: Iterated Extended Kalman filter
 - Method 2: Data fitting by Levenberg-Marquardt method

Method 1:

Iterated Extended Kalman Filter

- Iterated EKF : nonlinear measurement eqn.



- measurement vector

$$X_i = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \\ y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \quad X'_j = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_N \\ y'_1 \\ y'_2 \\ \vdots \\ y'_N \end{pmatrix}$$

- For each iteration
 - outlier rejection
 - re-calculate Kalman gain
(*matrix inversion: $2N \times 2N$)
 - update state estimation(parameters)

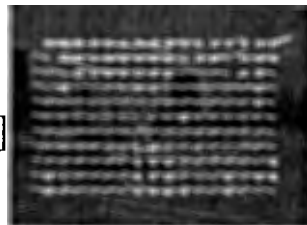
Method 2:

Levenberg-Marquardt

- Nonlinear least-square data fitting
- For each iteration
 - outlier rejection
 - solve a linear equation of dimension 4(or 5)
 - update parameter set

Outlier Rejection

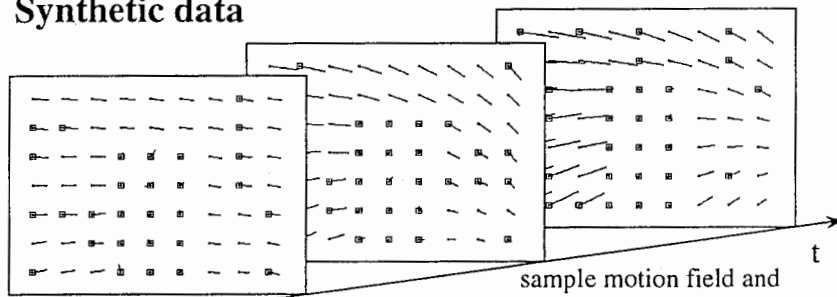
- Given current parameter set
- Calculate the residual error r_k for each observation
- Hypothesis test
 - Nonoutlier, $r_k < c \text{ med}[r]$
 - Outlier, otherwise
- Use only nonoutliers for estimation



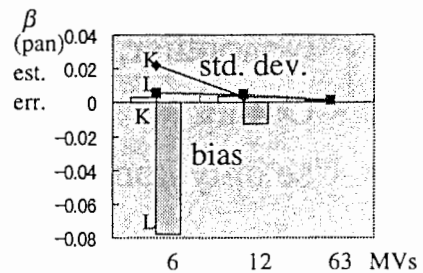
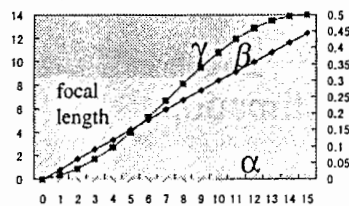
Experiments

- Synthetic data & real image sequence
- To simulate MPEG, H.263 : 16x16 ME, 320x240 image
- All calculation in normalized image coordinate (vertical image size=2.0)
- Initial parameter set: all zero
 - IEKF
 - parameter setting: measurement noise=std 0.5 pixel, system noise= 0.00001, initial cov.=0.0001

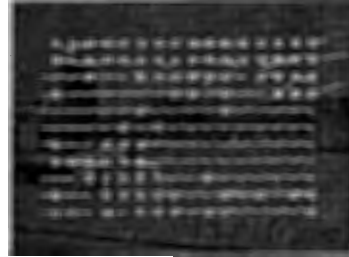
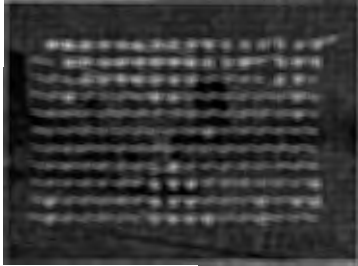
Synthetic data



- 16 frame
- 100 trial
- true parameter set known



IEKF results(almost the same as LM method)



Discussion

- IEKF
 - moderate accuracy at small # of MVs
 - computation complexity high if the # of MVs large
- LM method
 - biased estimation at small # of MVs \implies # of MVs should be large
 - computational complexity maintained at large # of MVs
- On focal length estimation
 - sensitive to noise, lens distortion
 - reliable estimation possible if large yaw/pitch [Park et al, '94]

Concluding Remarks

- Estimation of camera motion parameters from motion vectors of digital video is proposed
- Two types of implementation are presented
 - IEKF: more accurate when small # of MVs
 - Levenberg-Marquardt method: faster when large # of MVs
- Future works
 - practical implementation issues
 - MPEG2: frame/field mode, GOP gap
 - description issues: MPEG7

Appendix

Procedure of depth map extraction

多眼カメラを用いた奥行きマップ抽出実験

1998.4.30
第3研究室
朴 鍾一

実験の手順

- 1) 器材の準備
- 2) キャリブレーションパターンの撮影
- 3) 素材の撮影
- 4) 撮影映像のデータ化
- 5) カメラキャリブレーションプログラムの実行*
- 6) 奥行きマップ抽出プログラムの実行*

* All the source programs are under miris53:/usr/people/pji/msm/

1. 器材の準備

- カメラ5台
- カメラアダプター(4カメラ分のアダプター2台)
- 充分のケーブル(映像、電力)
- 8mm ビデオレコーダ 5台
- キャリブレーションパターン
- 確認用のモニタ
- 同期信号発生器(アナログ映像出力のあるもの)
- 同一フレーム確認用道具:カメラフラシまたはカット表示道具

* 注意点

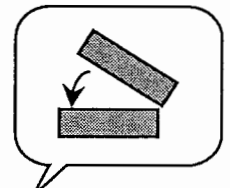
- 5台のカメラの同期が取れているか(カメラアダプターのスイッチ設定)
一> 添付資料1参照

2. キャリブレーションパターンの撮影

- 主な被写体が中心カメラの画面に入るように方向を調整
- 主な被写体が周りのカメラから観測できるかチェック
- 背景(遠距離被写体)ができれば一致するように周りのカメラの方向を調整
- カメラ露出、フォーカス、シャッタースピードのあわせ(5台を同じに)
- 5台のカメラからキャリブレーションパターンが観測できる範囲で、できるだけ近距離に配置
- パターンの撮影(充分長く、場合によって、パターンの位置を変えながら2-3回撮影する)
- 注意: これが終わってから絶対にカメラは動かさないこと。触っても絶対駄目

3. 素材の撮影

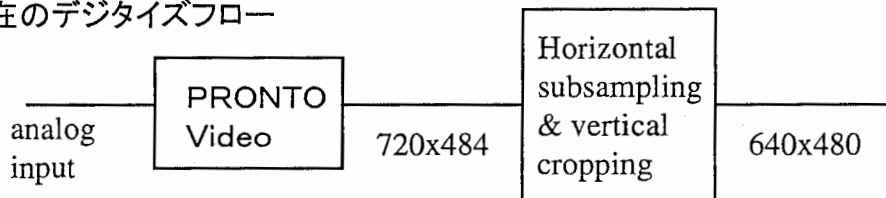
- 5台すべてのカメラから観測できる位置で、同一フレーム特定用のものを撮影
- と同時に撮影開始
- 同じものを少なくとも2、3回は繰り返し撮影した方がいい
- 注意: 5台の8mmレコーダーがすべて録画しているか確認



4. 撮影映像のデータ化

- 5つの8mmテープから良好に撮れているところを探してA/D変換
- フレームごとに ppm形式のファイルに保存
- 動画像の場合は、5つのカメラからの映像に対して、同時のフレームを正確に特定し、同じフレーム番号をつけることが重要
- 640x480画素のデジタイズが好ましい(現在奥行き抽出プログラムが720x484画素に対してはエラーを引き起こす:)
- ファイル名の付け方: {画像列ネーム}_ {カメラ位置}_ {フレーム番号}
カメラ位置=c(center), r(right), l(left), a(lower), u(upper)
フレーム番号 eg. 001, 002... (3 digit)
- 注意: キャリブレーションパターンの画像と素材画像は必ず同じデジタイズ経路であること!

* 現在のデジタイズフロー



- get reference frame
- calculate frame number

5. カメラキャリブレーションプログラムの実行

1) キャリブレーション用 画像の編集

- キャリブレーションパターン以外の領域をPhotoshopとかxvを使って削除
- キャリブレーションパターンの四角形以外はエッジが見えないようにする

2) 既知の3次元特徴点の画面座標の抽出

- “get_caldata”を実行すると“fps.dat”に特徴点の座標出力
 - 1)の結果画像を用いる
 - 入力パラメータ: エッジthreshold(150.0 - 300.0 程度)
 - 点ごとに3次元座標と画面座標、あわせて5個出力
- fps.datをrename(eg. cal_c.dat, cal_r.dat)
- 5枚の画像すべてに対して実行

3) 中心カメラのカメラパラメータ取得(Tsai's algorithm)

- “nccal_fo”の実行
 - 2)の出力ファイルを用いる
 - カメラの内部および外部パラメータの取得
 - stdoutに出力されるので“> xxx.cp”のようにredirection

4) 周りのカメラのカメラパラメータ取得(Tsai's algorithm)

- “eccal”の実行
 - 2)の出力ファイルを用いる
 - 3)の出力データを用いる(内部パラメータが同一であると仮定)
 - カメラの外部パラメータの取得
 - stdoutに出力されるので“> xxx.cp”のようにredirection

これによって、カメラの3次元位置および方向、画像へのマッピング関係が分かる。

この手順が最適とは思われない。正確にコントロールできるキャリブレーションパターンがあれば、すべてのカメラの内部パラメータを独立にかつ正確に計算でき、最も理想的であるが、いまのままでもかなりの精度は得られる。

6. 奥行きマップ抽出プログラムの実行

1) まず、必要なファイルがあるかチェック

- 入力映像ファイル
- キャリブレーションファイル
- infileの用意(右参照)

2) msmの実行

msm [-option value] infilename

infileの一例

infile_run

```
run_c_
run_r_
run_l_
run_a_
run_u_
calib/cc.cp
calib/cr.cp
calib/cl.cp
calib/ca.cp
calib/cu.cp
```

Table. Optionと valueのlist

option	value[default]	description
nf	1,2,... [1]	number of frames
sf	0,1,... [0]	starting frame number
df	1,2,... [1]	frame increment
sl	0,1,... [0]	lower limit of disparity searching range
sr	int [40]	upper “
nl	1,..,4 [4]	number of layers in resolution pyramid
ss	1,2,4 [1]	subsampling factor of output depth map
mt	1 [1]	method
pp	0 or n [1]	postprocessing(0:no n:number of iteration)
nc	0 or 1 [0]	calibration(0:no(eg. database images) 1:yes)
log	0 or 1 [0]	LOG filter(0:no 1:yes)
mg	0 or n [0]	margin to absorb calibration error

* msm(multi-view stereo matching)には2つのバージョンがある

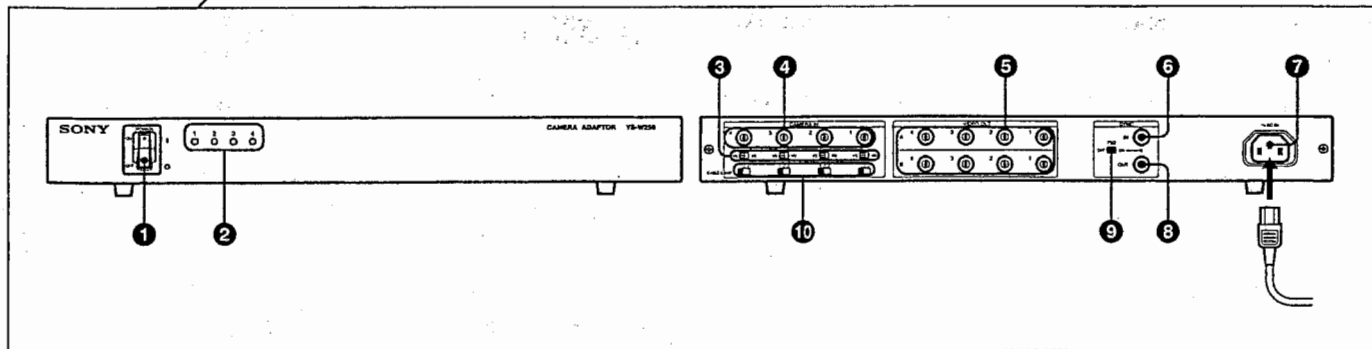
1) msm: non-interlace映像用

2) msm_fld: interlace映像用

Any questions and suggestions should be directed to
 Jong-II Park
 Hanyang Univ., Korea
 Email: jipark@email.hanyang.ac.kr

各部の名称と働き

添付資料 1



① POWER スイッチ

電源のON/OFFを切り換えます。接続したすべてのカメラにも電源が供給されます。

【ご注意】

- ・本機の電源を入れてからカメラを接続しても、画像は、映りません。接続を終えてから本機のPOWERスイッチをONにしてください。
- ・本機の電源を切ったあと再び電源を入れるときは、3～4秒たってからPOWERスイッチをONにしてください。

② LED ランプ

カメラが接続されると、ランプが点灯します。

③ VS/VD切り換えスイッチ

VD方式対応のカメラ (SSC-CX21V) を使用する場合、スイッチをVD側へ設定します。

通常は、VS (従来のVideo Sync方式) 側へ設定してください。

④ CAMERA IN (カメラ入力) 1～4端子 (BNCコネクタ)
専用カメラSSCシリーズ (重畳方式) またはDXC-107AのDC IN.VS IN/VIDEO OUT端子に接続します。

【ご注意】

- ・SSCシリーズのカメラをお使いになる場合、間違ってもMONITOR OUT端子に接続すると、カメラが壊れ、動作しなくなることがあります。必ず、VIDEO OUT端子に接続してください。

⑤ VIDEO OUT (映像出力) 1～4 A/B端子 (BNCコネクタ)
ビデオモニターの映像入力端子に接続します。CAMERA IN 1～4端子に入力された映像信号が出力されます。A/Bコネクタには同じ信号が出力されます。

⑥ SYNC IN (同期信号入力) 端子 (BNCコネクタ)

カメラに同期をかける場合に、外部同期信号 (または同期用映像信号) を入力する端子です。

⑦ 電源ソケット

付属の電源コードを差し込み、AC 100V電源へ接続します。

⑧ SYNC OUT (同期信号出力) 端子 (BNCコネクタ)

SYNC IN端子⑥に入力された同期信号がそのまま出力されます。

【ご注意】

SYNC IN端子に何も接続しないで、SYNC OUT端子から同期信号を出力させることはできません。

⑨ 75Ω (75Ω終端) スイッチ

SYNC IN端子を75Ωで終端するとき、ONに設定します。

- ・SYNC IN端子のみ使用する場合は、ONに設定します。
- ・SYNC IN端子とSYNC OUT端子の両方を使用する場合は、OFFに設定します。

【ご注意】

SYNC IN端子、SYNC OUT端子とも使用しない場合には、75ΩスイッチをONに設定してください。

⑩ CABLE COMP (ケーブル補償切り換え) スイッチ

ケーブルを伝わる途中で起きる信号の減衰を補償するため、カメラとの接続に使う同軸ケーブルの種類と長さに応じて次のように切り換えます。

SSCシリーズビデオカメラの場合

スイッチ設定	1	2	3
ケーブルの種類			
3C-2V	100mまで	200mまで	300mまで
5C-2V	160mまで	330mまで	500mまで
7C-2V	200mまで	400mまで	600mまで

DXC-107Aの場合

スイッチ設定	1	2
ケーブルの種類		
3C-2V	100mまで	
5C-2V	160mまで	300mまで

