

多眼ステレオを用いた3次元画像表示における画像生成法

菅谷 保之 北原 格 佐藤 清秀 大田 友一

筑波大学 電子・情報工学系

〒305 つくば市天王台 1-1-1

sugaya@image.is.tsukuba.ac.jp

あらし 実情景を対象とした運動視差の再現可能な3次元画像表示を実現するために、多眼ステレオによって得られた情景の3次元情報を用いて、任意視点における表示画像を生成し表示する方法について述べる。撮影画像間の内挿画像を生成することにより、見え方の変化に応じた物体の輝度値の微妙な変化を再現する手法と、奥行き毎に分割したプレーン表現を生成することにより、実時間での画像生成、表示を可能にする手法、および、これらの手法を利用して構築した3次元画像表示システムを紹介する。

キーワード 画像生成, カメラマトリクスステレオ, 3次元画像表示, 内挿処理, プレーン表現

Image synthesis for 3D display by using camera matrix stereo

Yasuyuki SUGAYA Itaru KITAHARA Kiyohide SATOH Yuichi OHTA

Institute of Information Science and Electronics, University of Tsukuba

Tennoudai 1-1-1, Tsukuba 305, Japan

sugaya@image.is.tsukuba.ac.jp

Abstract In this paper, we propose two algorithms for synthesizing images from arbitrary viewing positions based on 3D representation of a real scene generated by using camera matrix stereo. One algorithm is based on view interpolation, which has an advantage of synthesizing realistic images. The another is based on plane representation, in which a set of plane images is generated by dividing the 3D scene representation into slices. This algorithm has an advantage of synthesizing the images from arbitrary viewing positions in real time. We also introduce a 3D image display system using these algorithms.

key words image synthesis, camera matrix stereo, 3D image display, view interpolation, plane representation

1 はじめに

人間は、主に両眼視差と運動視差の2つの要素によって物体の奥行きを認識している[1][2]。従ってこれらの要素を再現した画像表示を行うことで、観察者にあたかもそこに本物の物体が存在するかのような錯覚を起こさせることが可能となる。

我々はこの2つの要素のうち運動視差に注目し、運動視差を再現する3次元画像表示を考えている。運動視差を再現する3次元画像表示方式の例としては、レンヂキュラススクリーンなどを用いた多視点画像投影やホログラフィによる受動的な方式[3][4]と、観察者の動きを検知し、適切な運動視差を再現した画像を生成または選択して2次元ディスプレイに表示する能動的な方式の2つに大別できる。能動的な方式の実現にはディスプレイに強力なデジタル画像処理機能を持たせることが必須であるが、それは今日のマルチメディア技術の素直な延長線上にあると考えられる。

能動的な方式としては、3次元幾何形状を用いる視点追従への柔軟性に優れた記述方式(geometry based rendering)と、実画像を用いる実世界表現力に勝る記述方式(image based rendering)が考案されている。3次元幾何形状を用いた記述方式では、表示すべき対象の3次元モデルがシステム中に保持されていることから、観察者の動きに応じた適切な観測画像の生成・表示が可能になる。しかし、実際のシーンの3次元モデルを自動的に作成するのは困難であり、VRシステムのほとんどは計算機内部で合成された3次元モデルのみを用いているのが現状である。一方、実画像を用いた記述方式は、異なる視点から観測される画像を予めシステムに保持し、観察者の動きに応じて適切な観測画像を選択表示するもので、実際に撮影した画像をそのまま用いるため実世界表現力に優れている。この方式を用いた画像表示によって得られる映像の臨場感は、システムを持つ画像の撮影時の視点間隔と画像全体でカバーするシーンの範囲に依存する。そのため観察者に高い臨場感を与えるためには膨大な量のデータが必要になる。これら双方の欠点を補い、かつ利点を合わせ持つ方式として、カメラにより実際のシーンを撮影した多視点画像から、シーンの3次元形状情報を自動獲得し、任意視点位置からの観測画像を生成し表示する方式が注目されている[5]-[8]。

2 運動視差の再現可能な3次元画像表示

我々は撮影された多視点画像からステレオ法によりシーンの3次元情報を獲得し、得られた情報をもとに任意視点における観測画像を生成し表示する3次元画像表示システムの実現を目指す[9][10]。多視点画像からシーンの3次元形状情報を自動獲得する手法として、佐藤らはマトリクス状に配置された画像を用いた多眼ステレオ法を提案した[11]-[14]。この手法はマトリクス状に配置された画像の幾何学的特性と各画像から得られる情報の冗長性を利用することで一般的なステレオ法で起こる隠れによる画像間での誤対応を低減することができる。この手法を用いることで、任意視点における観測画像生成に必要な物体輪郭が鮮明で

密な奥行き情報を得ることができる。

本稿では、ステレオ法により得られたシーンの奥行き情報をもとに任意視点における観測画像を生成する手法について述べる。以下の節では撮影された多視点画像の奥行き情報は既に得られているものとする。画像生成手法として画素の内挿処理による手法と、シーンを同じ奥行き情報ごとに分割したプレーン表現による手法を提案し、それぞれの手法の得失について説明する。またこれらの手法を利用して構築した3次元画像表示システムを紹介する。

3 入力画像の内挿による画像生成法

本節では各視点から観測されるシーンの奥行き情報から、画素の内挿処理により任意視点における観測画像を生成する手法について述べる。

3.1 カメラマトリクス平面内の任意視点

観察者の視点カメラマトリクス平面内の時、その視点から観測される画像は、カメラマトリクスで撮影された画像の内挿処理によって生成される。カメラマトリクスによって撮影された二枚の画像 I^m, I^{m+1} を用いて、視点 V^{m+k} , $0 < k < 1$, で観測される内挿画像 I^{m+k} の生成を考える。ここで、画像 I^{m+k} で観測されるシーン中の点は、 I^m, I^{m+1} の少なくとも一方の画像で観測されており、また、画像 I^m, I^{m+1} の各点における正確な視差画像 D^m, D^{m+1} は得られているものとする。

$I^{m+k}(x)$ の真の視差が \bar{d} のとき、その I^m 上での対応点は $I^m(x+k \times \bar{d})$ となる¹。しかし \bar{d} は未知であるため、様々な視差 d を仮定した時の対応点の視差情報 $D^m(x+k \times d)$ から、その仮定の真偽を判定する。

$$D^m(x+k \times d) = d \rightarrow \text{矛盾しない}$$

$$D^m(x+k \times d) \neq d \rightarrow \text{矛盾する}$$

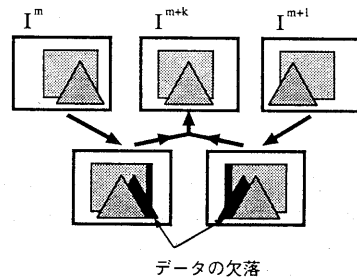


図1: 内挿処理による画像生成

¹記述を簡単化するため、まず1次元画像として説明する。

$D^{m+1}(x-(1-k) \times d)$ からも、同様に情報を獲得できる。いずれかの画像 I^m, I^{m+1} で対応点の視差が d であれば、その仮定は正しいと判断される。このような d のうち最大の d を \bar{d} として選択する。

$I^{m+k}(x)$ の輝度値は、対応点が両方の画像で観測されている場合はそれぞれの観測値の視点の距離に応じた重みつき平均を、いずれかの画像でのみ観測されている場合はその観測値を与える。このように、近い視点で得られる輝度情報を優先的に利用することで、ハイライトのように視点位置に依存して同一点の観測値が異なる場合でも正しいシーンの再現が可能になる。

以上のアルゴリズムを2次元に拡張することで、 $Z=0$ の格子平面内の任意の視点位置 $V(X, Y, 0)$ ($|X| \leq (M-1) \cdot b; |Y| \leq (N-1) \cdot b$) から観測した画像の生成が可能となる。

3.2 カメラマトリクス平面外の任意視点

片山らは、カメラマトリクス平面内の任意の視点位置からの観測画像が利用可能であれば、この平面外の任意の視点位置からの観測画像が生成可能であることを示した[8]。前節で示した方法により $Z=0$ 平面内の任意の視点位置からの観測画像を取得できるため、この方法によって、任意の視点位置 $V(X, Y, Z)$ で観測される画像の生成が可能となる。

4 プレーン表現を用いた画像生成法

シーン中の物体の3次元位置は観察者の視点位置に依存しない情報である。従って全ての画像で同一点における奥行き情報を持つことは冗長である。またシーンの3次元情報はステレオ法により量子化された視差として得られる。本節では、多視点画像から得られる情報を奥行き毎に分割し統合するプレーン表現を提案し、これを用いた任意視点における観測画像の生成手法について説明する。

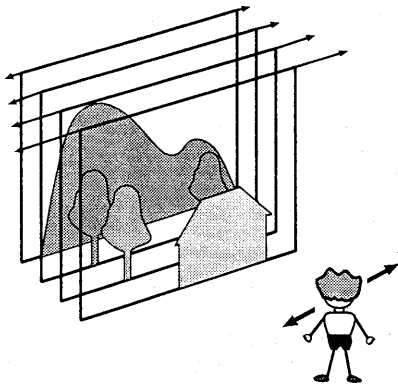


図 2: プレーン表現による画像生成

4.1 プレーン画像生成の基本原則

視点 (X, Y) で観測される画像 $I^{X, Y}$ 上の点 (x, y) の推定視差が d であるとき、この点の中心画像 $I^{0,0}$ での対応点は式(1)に示される座標で観測される。

$$I^{m,n}(x, y) \Leftrightarrow I^{0,0}(x+md, y+n) \quad (1)$$

式(1)の関係をもとに各視点から得られる画像上の点の中心画像での座標を求め、各視点ごとに分割した画像として統合することでプレーン画像を生成する。ここでは、プレーン画像をテクスチャ情報 T_d と存在情報 E_d のセットと定める。以下にプレーン画像生成の具体的なアルゴリズムを示す。

1. 画像 $I^{m,n}$ 上の点 (x, y) の中心画像での座標 (x', y') を式(2)から求める。

$$\begin{aligned} x' &\Leftarrow x + m \cdot D^{m,n}(x, y) \\ y' &\Leftarrow y + n \cdot D^{m,n}(x, y) \end{aligned} \quad (2)$$

ここで $D^{m,n}(x, y)$ は画像 $I^{m,n}$ 上の点 (x, y) の視差を表す。

2. 1で求めた座標 (x', y') にテクスチャ情報、存在情報を与える。

$$T_d(x', y') \Leftarrow I^{m,n}(x, y) \quad (3)$$

$$E_d(x', y') \Leftarrow 1 \quad (4)$$

$I^{m,n}(x, y)$ は画像 $I^{m,n}$ 上の点 (x, y) での輝度情報であり $E_d(x, y)$ はその座標における物体の有無を1または0で表す。

このようにシーンの情報をプレーン画像として表現することで、各画像ごとに保持していた冗長なデータを削減する。データを削減することで先に述べた内挿処理による画像生成手法のように、視点の移動に応じた物体の輝度変化は再現できないが、ハードウェアにより実時間で画像生成が行えるという利点がある。一方、奥行き情報の精度は誤った視差の蓄積によってデータ統合前の精度よりも低くなる。これは、各視点の画像上に分散している誤推定された視差を持つ点を1つの画像に統合してしまうことが原因である。プレーン画像から観測画像を生成する際にこの影響が顕著に現れると考えられるので、プレーン画像における視差精度の低下は解決しなければならない問題である。

4.2 視差の信頼度を考慮したプレーン画像の生成法

プレーン画像生成時に生じる誤推定された視差の影響を減らしプレーン画像での視差の精度を向上させるために、推定された視差の信頼度を定義し推定視差の信頼度の高い点だけをプレーン画像生成に採用する手法を提案する。

ある点において観測される対応点の視差が全て正しければ、その点は統合された画像上の1点に投影される。従ってプレーン画像上に投影された点の多い視差がその座標における正しい視差である可能性が高い。

実際には隠れの影響によりそれぞれの点ごとに観測可能な対応点の個数が異なるため、プレーン画像上に投影された点の個数 $N(x, y, d)$ と観測可能な個数 $\bar{N}(x, y, d)$ の比 $N(x, y, d)/\bar{N}(x, y, d)$ を視差の信頼度と定義する。

$$N(x, y, d) = \sum_{m,n,x',y'} f^{m,n}(x', y') \quad (5)$$

$$f^{m,n}(x', y') = \begin{cases} 1 & D^{m,n}(x', y') = d \\ 0 & D^{m,n}(x', y') \neq d \end{cases} \quad (6)$$

$$\begin{aligned} x &= x' + D^{m,n}(x', y') \cdot X/b \\ y &= y' + D^{m,n}(x', y') \cdot Y/b \end{aligned} \quad (7)$$

$$\bar{N}(x, y, d) = \frac{M}{N(x, y, d)} \sum_{m,n,x',y'} n^{m,n}(x', y') \cdot f^{m,n}(x', y') \quad (8)$$

M は多視点画像から得られる視差画像の総数である。 $n^{m,n}(x', y')$ は画像 $I^{m,n}$ 上の点 (x', y') の対応点が画像 $I^{m,n}$ を中心とする 3×3 の画像マトリクス中で観測される割合を表す。この値は多眼ステレオ法 SEA[11]-[14] を用いてシーンの奥行き情報を獲得するときに得られる隠れの情報を利用して求めることができる。

ここで定義した視差の信頼度を利用して、前節で述べた基本原理に従って生成したプレーン画像から信頼度が閾値に満たない点を削除することでプレーン画像の視差精度が向上する。

4.3 プレーン表現による任意視点の観測画像生成

観察者の視点 (X, Y, Z) に従って、プレーン画像に対してその視差によって決まる平行移動、拡大縮小処理を施し、奥行き方向に垂直に配置した平面へ各プレーンをテクスチャマッピングすることで、任意視点における観測画像を生成する。物体の存在しない点へは、その点を透明なテクスチャとして扱うことで対応する。

視点 (X, Y, Z) と視差によって決定する各プレーン画像の平行移動成分は式 (1) の対応関係から $(-d \cdot X/b, -d \cdot Y/b)$ となる。また、視点の前後移動に伴う画素のスケール係数は視差の Z 座標とプレーンの奥行き a により、 $a/(a-Z)$ となる。ここで a は、視差 d [pixel] と焦点距離 f [mm]、隣合ったカメラ間の基線長 b [mm]、撮像面での一画素の大きさ p [mm/pixel] から次式のように求められる。

$$a = \frac{fb}{dp} \quad (9)$$

4.2 実験結果

以上に述べたアルゴリズムを用いてプレーン画像を生成する実験を行った。使用したデータは 7×7 の視点から撮影した画像と、ステレオ法により得られる 5×5 枚の視差画像である。

図 6(a) はプレーン画像生成に用いた中心視点の視差画像であり、図 6(b)、図 6(c) はそれぞれ基本原理、提案したアルゴリズムにより生成したプレーン画像から再構成した中

心視点の視差画像である。これらの画像は視差を画素の明るさとして表現したものであり、明るいところ程視差が大きいことを表している。図 6(b) により、誤推定された視差の蓄積によるプレーン画像での視差精度の低下が確認できる。図 6(c) では、図 6(b) に見られる誤推定された視差の影響が取り除かれていることがわかる。また、生成したプレーン画像から再構成した視差画像と手入力で作成した正しい視差を持つ視差画像と比較することで、プレーン画像の視差精度を調べた。表 1 は正しい視差との比較により求めた各視差画像における視差の正答率である。この結果から、視差の信頼度を考慮して生成したプレーン画像から再構成した視差画像は、入力に用いた視差画像よりも視差精度が高いことがわかる。このことから提案した手法の有効性が確認できた。

表 1: 復元された視差画像での視差の正答率

画像	画像全体	隠れなし	隠れ境界
入力視差画像	80.75 %	83.38 %	72.86 %
再構成 (基本原理)	63.26 %	68.09 %	48.79 %
再構成 (提案手法)	82.76 %	84.33 %	78.06 %

5 3次元画像表示システムの構築

前節まで、撮影されたシーンの奥行き情報から任意視点における観測画像を生成する手法について述べた。運動視差を再現する 3次元画像表示を実現するために、これらの手法によって生成された画像をスクリーン上に表示する。しかし、利用する表示系によって、表示画像に対して適切な幾何学的補正を加えなければ、各視点における最適な観測画像を提示したことはない。具体的には、視点の移動に追従して移動するスクリーンか、視点の移動に関わらず定位置に固定されたスクリーンかのどちらを用いるかによって画像の補正法は異なる。

本章では、観察者とスクリーンの 3次元空間における相対的な位置関係にもとづいて、正しい視差を再現するための画像提示法について述べる。また、実際に固定式スクリーンを用いた 3次元画像表示システムを構築し、前節で述べた二つの手法により生成された画像を提示する実験を行った。

ここでは説明の簡略化のために Y 軸を除いた $X-Z$ 軸の 2次元座標系での関係について説明する。この関係は Y 軸を含む 3次元座標系に容易に拡張できる。

5.1 移動式スクリーンにおける運動視差の再現

HMD のように観察者と共に移動するスクリーンを用いる場合、正しい運動視差の再現は単純に実現できる。図 3 に観察者と移動式スクリーンとの配置を示す。このシステムでは、観察者の視点位置が $V'(X, Z)$ の場合、スクリーンを Z 軸と垂直に、その中心が常に $(X, Z+a)$ となるように配置し、カメラ位置 $V(X, Z)$ からの観測画像を提示する。

正しい運動視差を再現するためには、シーン中の点 P を実際に視点 V から観測した時の視線と、視点 V' の時にス

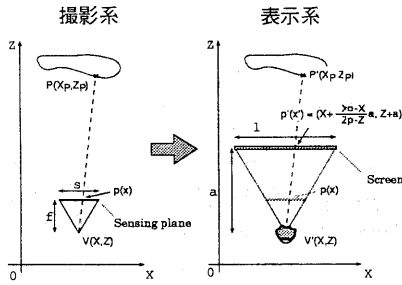


図3: 移動式スクリーン

クリーンに提示される点 P' の像に対する視線 $V'P'$ が平行となるようにすれば良い。

これらの幾何学的な関係を導き出すために、点 $P(X_P, Z_P)$ がカメラ位置 $V(X, Z)$ で撮影され、視点位置 $V'(X, Z)$ の観察者に対して提示された場合について考える。撮像系の焦点距離を f 、撮像面の大きさを s 、スクリーンに提示される画像の大きさを l とすると、 P は画像上の座標 $p(x) = p((X_P - X)f / (Z_P - Z))$ で観測され、スクリーン上の座標 $p'(l/s)x$ に表示される。このとき、 p' のシーン座標は、次式によって与えられる。

$$(X_{p'}, Z_{p'}) = (X + \frac{X_P - X}{Z_P - Z} \cdot \frac{l}{s} f, Z + a) \quad (10)$$

従って、次式の関係を満たすことにより、視線 $V'P'$ が視線 VP と常に平行になることは明らかである。

$$a = \frac{l}{s} f \quad (11)$$

これは、撮像系の画角と、提示された画像への観察者の視角を等しくすることを意味している。

5.2 固定式スクリーンにおける運動視差の再現

図4に観察者の視点位置によらず中心が常に $(0, a)$ となるような固定式スクリーンの配置を示す。表示面までの距離 a は、観察者の視点が原点にある時に、式(11)によって与えられる。

移動式のスクリーンと同様に補正を加えなかった場合、この点は、スクリーン上の点 $p'(l/s)x = p'((X_P - X)a / (Z_P - Z))$ に投影される。しかし、 p' のシーン座標は式(10)とは異なり、次式によって与えられる。

$$(X_{p'}, Z_{p'}) = (\frac{X_P - X}{Z_P - Z} a, a) \quad (12)$$

一方、 VP と平行な視線 $V'P'$ がスクリーンと交差する点 p'' を考えると、そのシーン座標は、次式ようになる。

$$(X_{p''}, Z_{p''}) = (\frac{X_P - X}{Z_P - Z} (a - Z) + X, a) \quad (13)$$

式(12)(13)より、正しい運動視差を再現するためには、観察者の視点位置から観測される画像を提示するだけでは不十分であることがわかる。しかし同時に、これらの式に基

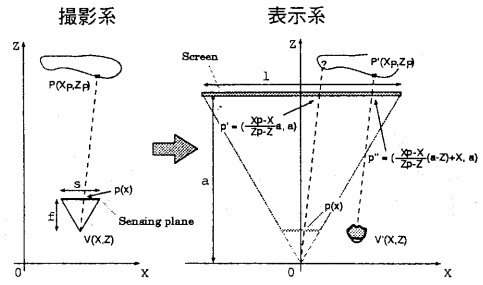


図4: 固定式スクリーン

づいて画像の適正な補正法を以下のように求めることができる。すなわち、点 p' と p'' との関係が

$$X_{p''} = \frac{a - Z}{a} X_{p'} + X \quad (14)$$

として導き出されるため、提示画像を $(a - Z)/a$ だけスケールリングした後に、 X だけ平行移動させることで、視線が VP と平行になるように補正できる。

5.3 固定式スクリーンを用いたシステムの構築

固定式スクリーンを用いた3次元画像表示システムとして、観察者の視点位置を磁気センサを用いて検出し、視点位置に適した画像を生成してビデオプロジェクタによってスクリーンに投影するシステムを構築した。図5にシステムの概要を示す。

観察者の視点位置検出を行うセンサは、磁場発生装置から最大半径約 $300[\text{cm}]$ の位置を約 $0.01[\text{mm}]$ の精度で、毎秒120回検出可能である。表示デバイスには、表示画素数 480×640 画素、最大表示サイズ100インチのビデオプロジェクタを使用した。内挿による画像生成手法では画像生成に時間がかかるため、観察者の視点移動に応じた画像の生成・表示を実時間で行うことができない。そこで、密な視点間隔で生成した画像を予め生成し、JPEGで圧縮・蓄積しておくことで観察者の視点に適した画像を実時間に表示する蓄積再生型のシステムを実現した。視点検出から画像表示までにかかる時間は約0.04秒である。従って、このシステムでは25枚/秒の画像表示が可能である。一方プレーン表現を用いた画像生成では、視点の変化に応じた物体の輝度変化は再現できないものの、視点に応じた画像を実時間に生成できる。従って視点移動に伴った滑らかな画像表示が可能である。実際の画像生成はINDIGO²によって生成した。

実験で使用したスクリーンおよびCCDの撮像面の大きさは $115 \times 80[\text{cm}]$ 、 $6.46 \times 4.83[\text{mm}]$ である。撮影に用いたカメラの焦点距離は $15[\text{mm}]$ であった。従って式(11)より観察者からスクリーンまでの適切な距離 a は $267[\text{cm}]$ となる。

5.2節で述べた補正を施した画像を表示することで、実験で撮影したシーン中の、観察者からスクリーンまでの距離と等しい奥行きにある石膏像(約 $260[\text{cm}]$)が、常にスクリーン上に静止して一定の大きさで表示されていることが確認できた。

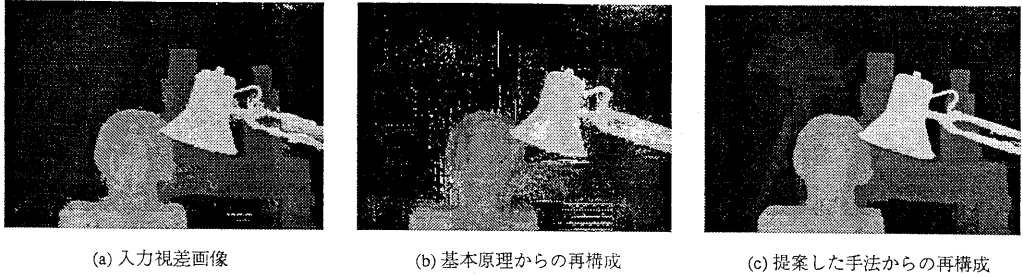


図 6: プレーン画像から復元された視差画像

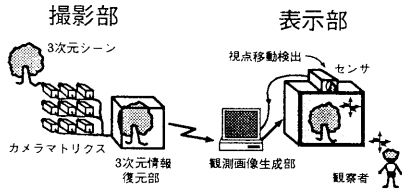


図 5: 運動視差を再現する 3 次元画像システム

6 まとめ

本論文では運動視差を再現する 3 次元画像表示のための画像生成手法を提案した。内挿処理による手法では視点の移動に応じた物体の輝度変化を再現した画像生成が可能となった。プレーン表現による手法では視差の信頼度を考慮することで誤推定された視差を持つ点の蓄積を軽減し、データの統合による視差精度の低下を解決した。この手法を用いることで実時間での画像生成が可能である。また、実際に 3 次元画像表示システムを構築することで、それぞれの画像生成手法の有効性及びスクリーンに合わせた幾何学補正の正当性を確かめた。

参考文献

- [1] Borner, R., "Autostereoscopic 3D-imaging by front and rear projection and on flat panel displays," *Displays*, Vol.14, No.1, pp.39-46, 1993.
- [2] Pastoor, S., "Human factors of 3D displays in advanced image communications," *Displays*, Vol.14, No.3, pp.150-157, 1993.
- [3] 大越孝敬: 「3次元画像工学」, 朝倉書店 (1991)
- [4] 濱崎襄二, "三次元表示技術の現状と動向", *ディスプレイアンドイメージング*, Vol.3, No.3, pp.255-267, 1995
- [5] 高橋、本田、山口、大山、岩田, "少ない視差画像からのホログラフィック・ステレオグラム of 原画作成", 3

次元画像コンファレンス'93 講演論文集, pp.5-10, 1993.

- [6] Skerjanc, R., and Liu, J., "Computation of Intermediate Views for 3DTV," *Proc. 5th Workshop 1992 on Theoretical Foundations of Computer Vision*, pp.190-201, 1992.
- [7] Fujii, T., and Harashima, H., "Data Compression and Interpolation of Multi-View Image Set," *IEICE Trans. on Information and Systems*, Vol.E77-D, No.9, pp.987-995, 1994.
- [8] 片山、田中、押野、田村, "多視点画像データの補間処理による視点追従型立体画像表示", 3次元画像コンファレンス'94 講演論文集, pp.7-12, 1994.
- [9] 佐藤、北原、大田, "多眼ステレオ法を用いた運動視差の再現可能な 3次元画像表示", 3次元画像コンファレンス'95 講演論文集, pp.178-183, 1995.
- [10] 北原、佐藤、大田, "多眼ステレオ法を用いた運動視差の再現可能な 3次元画像表示", *テレビジョン学会誌*, Vol.50, No.9, pp.1268-1276, 1996
- [11] Satoh, K., and Ohta, Y., "Occlusion Detectable Stereo Using A Camera Matrix," *Proc. Second Asian Conf. on Computer Vision (ACCV'95)*, pp.II-331-335, 1995
- [12] Nakamura, Y., Matsuura, M., Satoh, K., Ohta, Y., "Occlusion Detectable Stereo -Occlusion Patterns in Camera Matrix-", *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'96)*, 1996
- [13] 佐藤、大田, "Occlusion Detectable Stereo-Systematic Comparison of Detection Algorithms", *ICPR'96*, Vol.I, Track A, pp.280-286, 1996
- [14] 佐藤、大田, "カメラマトリクスを用いた高精細ステレオ SEA における隠れ検出法の検討", *信学技報, PRU95-245*, 1996