

未校正多視点カメラシステムを用いた任意視点画像生成

矢口 悟 志[†] 木村 誠[†]
斎藤 英雄^{†,††} 金出 武雄^{†††}

本手法では、未校正の多視点カメラシステムから任意視点画像を作成する手法を提案する。多視点カメラのうち、お互いのエピポラ幾何学が既知な2台の基底カメラ画像により射影グリッド空間を構成する。この射影グリッド空間内の各点は、基底カメラに対するエピポラ幾何学が既知なカメラであれば、それを表す基礎行列(F行列)を用いて画像面に投影することができる。そこで、各カメラのシルエットに射影グリッド空間内の各点を投影することによって、Shape-From-Silhouette(SS法)の原理に基づいて3次元モデルを射影グリッド空間内に再構成する。この射影グリッド空間内に再構成された3次元モデルから、各視点の画像間の対応点関係が密に得られるため、この対応点関係を用いて2枚の入力視点画像からその中間にあたる任意視点画像を得る。さらに、連続画像に適用することにより任意の視点からの映像を作成する。

Arbitrary View Generation from Uncalibrated Multiple View Camera System

SATOSHI YAGUCHI,[†] MAKOTO KIMURA,[†] HIDEO SAITO^{†,††}
and TAKEO KANADE^{†††}

In this paper, we propose a method for arbitrary view generation from multiple view images taken with uncalibrated camera system. In Projective Grid Space (PGS), that is the three dimensional space which is defined by epipolar geometry between the two basis cameras in the multiple cameras, we reconstruct three dimensional shape model from the silhouette images of the multiple cameras. For the shape reconstruction in the PGS, the multiple cameras do not have to be fully calibrated, but the fundamental matrices of each camera to the two basis cameras must be collected. By using the three dimensional model reconstructed in the PGS, we can obtain the point correspondence between arbitrary pair of images which can generate the image of arbitrary view between the pair of images.

1. はじめに

近年、コンピュータ処理能力および容量の増大にともない、複数台のカメラによって撮影された画像からの3次元モデリング、レンダリングに関する研究は、非常にさかんになってきた。複数の視点の画像から、3次元形状を再構成したり形状をモデリングしたりする研究は、ロボット視覚システムやマシンビジョンに古くからさかんに応用されてきたが、近年、新しい画像を合成し、ユーザに提示する映像生成を目的とした研究がコンピュータビジョンおよびコンピュータグラ

フィックスの分野で非常にさかんになっている¹⁵⁾。

このような新しい視点の画像合成に関する研究は、大きく2つのカテゴリに分割することができる：何らかの手法でコンピュータに取り込んだ対象の3次元構造モデルから新しい視点の画像を生成するもの(モデルベースレンダリング)、そして、入力画像から直接任意視点画像を合成するもの(イメージベースレンダリング)である。前者のカテゴリでは、3次元構造モデルを入力することが重要になるが、Hiltonら⁸⁾、Curlessら⁴⁾、Masudaら¹⁴⁾、そしてWheelerら²⁴⁾等は、複数の距離画像を融合して、3次元構造を体積空間で再構築する手法を提案した。これらの手法で用いられる距離画像は、主に、様々な原理による3次元スキャナを用いて収集されることが多い。しかし、たいていの3次元スキャナは、3次元データを得るために数秒程度の時間を要することが多く、動きのあるイベントに対しては適用することができなかった。

[†] 慶應義塾大学理工学部

Faculty of Science and Technology, Keio University

^{††} 科学技術振興事業団、さきがけ研究21

PRESTO, Japan Science and Technology Corporation

^{†††} カーネギーメロン大学ロボティクス研究所

Robotics Institute, Carnegie Mellon University

明示的な3次元復元を行わずに画像に基づき任意視点画像を合成しようという、イメージベースドレンダリングもまた近年非常に発展してきた。Katayamaらは、視点を密に変化させて得られる画像列から、任意の視点の画像が合成できることを示した⁹⁾。Levyら¹³⁾やGortlerら⁷⁾は、このコンセプトを拡張し、3次元空間における任意の光線を表す直線が4つのパラメータで表されることに着目し、視点の異なる大量の入力画像から4次元の光線空間を構築し、この光線空間において任意視点の各画素に対応する光線の色を推定するという新しい枠組みを提案した。このような手法の本質的な問題点は、視点の異なる画像を非常に多く必要とすることであるため、動きのあるイベントに適用することは、非常に多くのカメラを必要としてしまうため困難となることである。

我々は、動きのあるイベントを仮想化するという目的で、現実的に利用可能な数台から十数台程度のCCDカメラを用いて、そこから入力される画像列から任意視点画像を生成する研究を行っている¹⁹⁾。その応用例として、サッカーや、ボクシング、相撲などスポーツのTV中継などに利用し、ユーザが自由に自分の見たい視点からの映像を表示させることのできるシステムなどが考えられ、将来性の高いシステムとなることが予想される。

このような多数のカメラを利用したシステムで3次元復元を行う場合、従来は各カメラを正確に校正しておくことが必要であった^{6),10)~12),23),25),26)}。カメラを校正するには、空間上の数点に対し、ワールド座標系における3次元座標と画像座標系における2次元座標を正確に測定し計算する方法²²⁾が一般的である。このように、空間中に正確に数点の校正用ポイントを設置して測定する必要があるため、特に多数のカメラの校正は非常に手間のかかるものであった。しかも、屋外の場合のように広い撮影空間を対象とする場合には、校正用ポイントを広範囲にわたって正確に設置することは非常に困難であった。そこで、カメラ校正を必要とせず、取得した画像からの情報のみで対象シーンの3次元復元を行うことが可能であれば非常に有効であり、より制約の少ないシステムが可能になると期待される。

本論文では、未校正の多視点カメラシステムから得られた画像から対象物体の3次元モデルの再構成を行い、3次元モデルをもとに任意視点画像を合成する手法を提案する。提案する手法では、画像から得られる基礎行列(F行列)の表すカメラ間のエピポーラ幾何を用いて「射影グリッド空間」¹⁷⁾を構築する。この

射影グリッド空間は、従来困難であった未校正の多視点カメラの画像からの3次元復元の枠組みを提供するものである。そしてこの射影グリッド空間においてShape from Silhouette(SS法)を用いることにより、3次元モデルを再構成し、得られた3次元モデルから得られる入力視点画像間の対応点関係を用いて、2枚の入力視点画像からその中間にあたる任意視点画像をイメージベースドレンダリングによって生成する。また本論文では、入力画像における背景領域から抽出した特徴点を射影グリッド空間に投影することにより、背景の床平面を任意視点画像に合成する手法を示す。そして9台の未校正多視点カメラにより撮影された連続実画像に本手法を適用して生成した任意視点における連続画像を示し、本手法の有効性についての検証を行う。

本論文では2章において射影グリッド空間の定義および射影グリッド空間における3次元再構成の手法について述べる。また3章で作成された3次元モデルから任意視点画像を合成する手法および、対象物体の背景にあたる床平面の合成方法について述べ、4章で実際に未校正の多視点カメラシステムから得られた画像に対して本手法を用いて行った実験について、また5章においてそれに対する考察を述べる。

2. 3次元モデルの作成

一般に、多数のカメラからの3次元復元のためには、対象を再構成しようとする3次元空間の各点と、多数のカメラ各々についての画像上に投影される位置との関係が必要となる。そこで、従来は、各カメラごとに3次元空間とカメラの画像座標とを関連づけるための射影行列をカメラごとに推定していた^{10),23)}。この射影行列の推定がカメラの校正に相当する。

本論文で提案する手法では、射影行列を用いることなしに3次元空間とカメラ座標とを関連づけるために、カメラ間のエピポーラ幾何を表す基礎行列(F行列)⁷⁾のみを用いて構成される3次元空間である「射影グリッド空間」¹⁷⁾によって、実空間とカメラ座標の関係を記述する。この射影グリッド空間の利用により、カメラごとに射影行列を復元せずに多数のカメラからの3次元復元を行うことが可能になる。

3次元復元の手法としては、Voxel Coloring Method²¹⁾による手法とShape from Silhouette(SS法)による手法が考えられる。前者は、室内のように照明条件が場所によらず安定な環境で、1つのカメラを用いて静止物体を多視点から撮影したような場合のように、各画像間で対応する点どうしの色調差が小さ

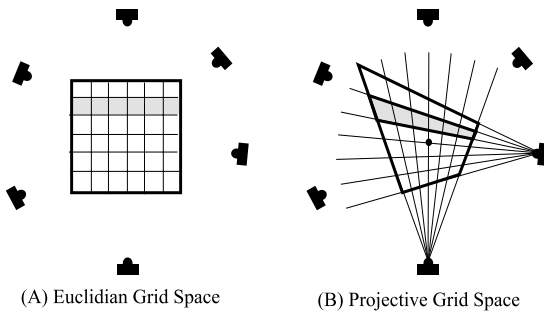


図1 ユークリッド空間と射影グリッド空間

Fig. 1 Definition of Euclidian Grid Space and Projective Grid Space.

い場合、後者より誤差の少ない復元結果が得られるが、屋外や大規模空間でカメラ間の距離が大きく、各カメラの感度などが大きく異なるような場合は、対応点どうしても異なる視点間で色調差が大きくなるために、Voxel Coloring Methodによる復元結果の誤差は非常に大きくなる。そこで本手法では大規模空間への適用も考慮し、射影グリッド空間における3次元モデルの復元に、Shape from Silhouette (SS法)を利用することにした。SS法において射影グリッド空間の各点から各カメラ座標へ投影される位置は、カメラ間のF行列のみによって決定され、この位置が各カメラに撮影されたシルエットの内部か外部かを判定することにより、物体領域を決定し、3次元モデルを復元する。

なお、本手法では、カメラレンズ歪みが生じていないカメラを想定している。このため、実験では、レンズ歪みの影響を無視することができるように、レンズ歪みの少ないカメラを用いた。

下記に、本手法における3次元モデル復元法の詳細を述べる。

2.1 射影グリッド空間

射影グリッド空間¹⁷⁾は次のように定義される(図1)。数台あるカメラシステムのうちの任意の2台を基底カメラ1, 2とする。この2台のカメラそれぞれの視点から中心射影によって3次元空間を定義する。ここで空間を定義する3軸は基底カメラ1から得られる画像のX軸, Y軸, 基底カメラ2から得られる画像のX軸になる(以下各P, Q, R軸とする)。また1グリッドは各画像の画素によって定められる。射影グリッド空間はユークリッド空間で見るとそのボクセルの大きさはそれぞれ異なり、視点に近いほど小さいという性質を持つ。

2.2 3次元モデル復元

本手法では3次元形状を復元するためにShape from Silhouette (SS法)を用いる。従来のSS法では、カメ

ラパラメータによって計算される各カメラごとの射影行列を用いて、一定の空間中のボクセルを各シルエット画像上に投影し存在を決定することで3次元モデルを構築していた^{2),16)}。

本手法では、カメラの校正を行っていないため射影行列を用いることができないが、射影グリッド空間のボクセルを画像間のF行列を用いて各カメラのシルエット画像へ投影することによりSS法を実行する。

F行列は2画像間のエピソード幾何の関係を表す 3×3 の行列である。F行列は2画像間の8点以上の点対応によって決定することができ、ワールド座標系における基準点の3次元座標を用いずに得ることができる。F行列を用いると一方の画像上の点からもう一方の画像中のそのエピソード線である直線を求められる。本手法ではこのF行列を用いて射影グリッド空間の一定の空間中のボクセルを各シルエット画像上に投影し存在を決定することで3次元モデルを構築する。

次に実際に射影グリッド空間上のボクセルを各カメラの画像に投影する方法について述べる。投影するボクセルを $A(p, q, r)$ 、基底カメラ1, 2から得られる各画像を画像1, 画像2、基底カメラ以外のカメラから得られる画像を画像*i*とする。またそれぞれの画像上のAの投影点を a_1, a_2, a_i 、画像*h*の画像*k*に対するF行列を F_{hk} と表すものとする。

ボクセル $A(p, q, r)$ は射影グリッド空間の定義より、画像1では $a_1(p, q)$ に投影されることが分かる(図2)。また画像2の画像1に対するF行列を用いて a_1 を画像2に直線*l*として投影すると直線*l*は式(1)のように定義される。

$$l = F_{21} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \quad (1)$$

射影グリッド空間の定義よりボクセル $A(p, q, r)$ の画像2における投影点 a_2 の*x*座標は*r*であるから、 a_2 はこの直線*l*上の*x*座標が*r*の点として定めることができる。

最後に画像*i*の投影点 a_i の決定法を述べる。先ほどと同様に F_{i1} を用いて a_1 を画像*i*に直線 l_1 として投影する。また a_2 を F_{i2} で画像*i*上に直線 l_2 として投影する。この2直線 l_1 と l_2 の交点が画像*i*の投影点 a_i となる(図3)。

実際の3次元モデルの作成の流れは次のようになる。まず射影グリッド空間にある一定の空間を定義し、その空間に含まれるすべてのボクセルを各シルエット画像上に投影する。各シルエット画像における投影点が

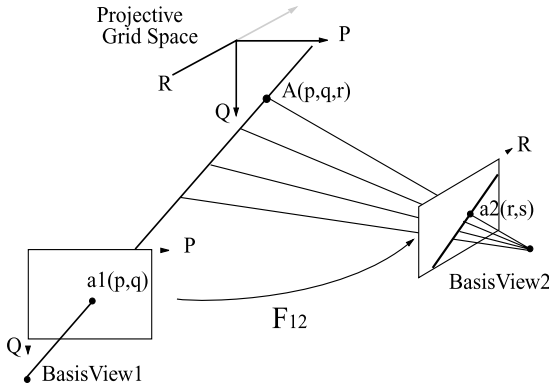


図2 基底カメラ画像への投影法

Fig. 2 Projection on to the two basis view images.

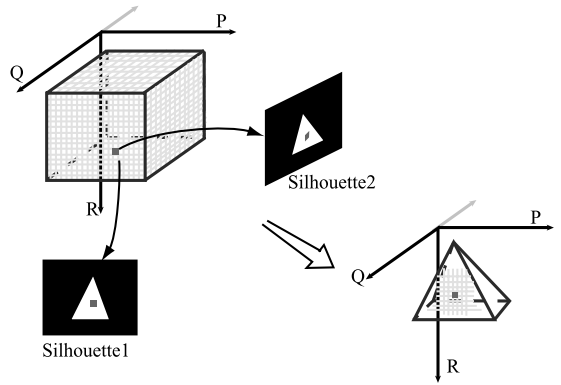


図4 3次元モデル再構築

Fig. 4 Reconstruction of 3D shape model.

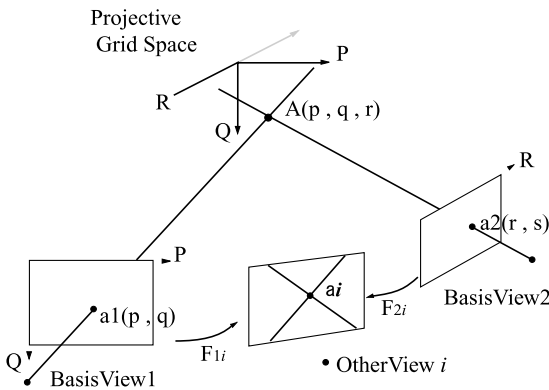


図3 その他のカメラ画像への投影法

Fig. 3 Projection on to non-basis images.

すべて対象物体のシルエット上にあるならばそのボクセルは存在，1つでもシルエット上にないならば存在しないものと判定し，存在するボクセルの集合として3次元ボクセルモデルとし(図4)，この3次元モデルからモデル表面のボクセル集合である表面形状モデルを求める．

その際，計算時間はボクセルの画像への投影時間によって決定されるが，本手法におけるボクセル投影の性質を用いることでより効率良く計算ができる．

図2のように基底カメラ1の画像には，各ボクセルのP,Q座標がそのままX,Y座標として投影されるため計算の必要はなく，また基底カメラ2の画像には基底カメラ1とのエピポーラ線を求めるだけで投影することができる．図3のように基底カメラ以外のカメラの画像では両基底カメラとの間のエピポーラ線計算をする必要があり，基底カメラの場合よりも計算量は多くなる．そこでシルエットへの投影順を基底カメラ1,2その他のカメラの順にし，各シルエットに投

影するたびに不要なボクセルを削除するようにする．このように処理することで3次元再構成をより効率良く行うことができる．

3. 任意視点画像の合成

次に作成された3次元モデルをもとに任意視点画像を生成する．3次元モデルを用いて任意の視点からのレンダリングをするには，3次元モデルの各3角メッシュに入力画像のテクスチャ情報を割り当て，テクスチャマッピングによって行う手法²³⁾と，合成しようとする付近の視点入力画像を複数枚選び，画像間の対応関係をモデルから求め，この対応関係に基づいて見かけ上の中間画像を合成する手法^{1),3),18),20)}が考えられる．

前者は入力視点のカメラ画像のテクスチャをいったん3次元モデルに逆投影し，そして任意視点に再投影することにより画像を生成するため任意視点の位置によらず3次元モデル復元誤差が合成画像に現れてしまうが，仮想視点をどこにでも設定することができる．

後者は任意視点の位置が入力カメラ視点に近くなればなるほど3次元モデル復元の誤差による画質劣化が少なくなり，入力カメラ視点と同一の視点では3次元モデル復元誤差の影響が現れないため，前者よりも画質劣化が少なく画像を合成できるが，仮想視点は現実のカメラ視点の中間にしか設定することはできない．

通常カメラはイベントに対し良い方向から撮影できる位置に設置されることが多く，またその場合いくつかの視点の中間に仮想視点が設定できれば，必要とされる視点をほぼ合成することができる．そのため本手法ではこのような場合比較的画質劣化の少ない後者を用いることとした．

3.1 入力視点におけるレンダリング

任意視点の画像は合成しようとする視点付近の入力画像を 2 枚用い、その中間に位置する視点の画像として作成する。ここでは、入力画像の視点からモデルを見たとき、モデルのどの面が手前にあるか、つまりどの面が画像に反映されるかを決定する。

処理は入力視点において Z-Buffer 法によるレンダリングを行う。前章で作成された表面形状モデルのボクセルは十分に密であるとする。この各ボクセルを、シルエット画像に投影するときと同じ要領 (2.2 節) で入力画像に投影する。入力画像に Z-Buffer を用意し、各画素の Z-Buffer には、その画素に投影されたボクセルの 3 次元位置とその画像の視点の 3 次元位置との距離を格納する。同画素に投影されるボクセルが複数ある場合、視点が一番近いボクセルの距離を格納する。また投影されるボクセルがない画素の画素値は初期化しておく。このようにして Z-Buffer に距離情報を持った入力画像を作成する。

ここでの距離は、射影グリッド空間において下記のように定義される。

$$D = \sqrt{(p_1 - p_2)^2 + (q_1 - q_2)^2 + (r_1 - r_2)^2} \quad (2)$$

ただし、 (p_1, q_1, r_1) 、 (p_2, q_2, r_2) は射影グリッド空間における任意の 2 点とする。この距離の定義では射影グリッド空間のすべての場所における距離の大小を比較することはできない。しかしここで必要なのは同じ画素に投影されるボクセルどうしの視点との距離の大小であり、これらのボクセルは視点とを結ぶ同一直線上にある。同一直線上であれば、射影グリッド空間中であっても、ここで示した定義により距離の大小関係の比較は可能である。

また、ここでボクセルと視点との距離を求めるために、射影グリッド空間における各視点の 3 次元座標が必要になるが、それらはエピポーラ幾何を用いて次のように決定できる。

基底カメラ 1, 2 の各視点を C_1, C_2 としその他カメラの視点を C_i とする (図 5)。射影グリッド空間の定義より基底カメラ 1 から得られる画像の X 軸, Y 軸によってカメラ 1 の視点 C_1 から中心射影的に P 軸, Q 軸が定められているから、 C_1 の P 座標, Q 座標は一意には定まらずあらゆる値をとりうるが、ここでは処理の汎用性を考えて基底カメラ 1 の画像中央の X 座標 X_{1c} , Y 座標 Y_{1c} をそれぞれ C_1 の P 座標, Q 座標とする。射影グリッド空間の R 軸は画像 2 の X 軸として定義されることから、画像 2 に対するエピポールとなる C_1 の R 座標は e_{21} の X 座標 (e_{21x}) と

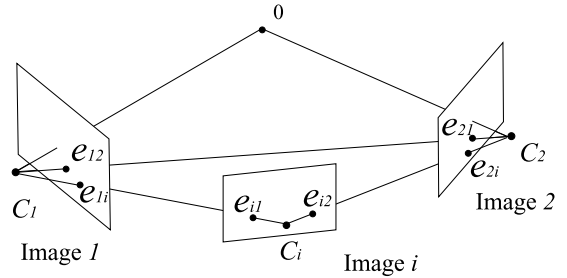


図 5 射影グリッド空間における各カメラの視点位置
Fig. 5 Camera position in the PGS.

なる。つまり、 C_1 の座標は $C_1 (X_{1c}, X_{2c}, e_{21x})$ になる。同様に、 C_2 の射影グリッド空間における R 座標も一意には定まらないが、基底カメラ 2 の画像中央の X 座標 X_{2c} を用いて定義されるものとする。 C_2 は画像 1 に対するエピポールとなるので、その P 座標, Q 座標はそれぞれ e_{12x}, e_{12y} となる。つまり、 C_2 の座標は $(e_{12x}, e_{12y}, X_{2c})$ と表すことができる。また、基底カメラ 1, 2 以外のカメラの視点 C_i の座標は、基底カメラ 1, 2 のエピポールの座標を用いて、 $C_i (e_{1ix}, e_{1iy}, e_{2ix})$ であることが分かる。

3.2 任意視点画像合成

次に先ほど作成された距離情報を持った入力画像を 2 枚用いて任意視点の画像を作成する。ここでは物体上の各点と視点との距離の違いによって生じる画像上での視差を利用して、その 2 視点の中間における各点の画像上の座標を決定する。

いま投影すべき 3 次元モデル上のボクセルを V 、その距離情報を持った 2 枚の画像上での投影点を $v_1 (x_1, y_1)$ 、 $v_2 (x_2, y_2)$ とし、画像間の距離を調節するウェイトを w とすると任意視点の画像における V の投影点 v_3 の座標は

$$v_3 = w \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + (1 - w) \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \quad (3)$$

と表される。画素値も v_1, v_2 の画素値を用いて同様に決定される。

3 次元モデル上の V を画像上に投影したとき、 V とその画像の視点との距離と、 v_1 の Z-Buffer に格納されている距離が異なる場合、その画像では V がオクルージョンになっていることになる。この場合、その画像の座標は任意視点画像における投影点の座標計算に用いるが画素値は用いない。画素値はもう一方の画像の v_2 の画素値をそのまま用いる。 v_1, v_2 両方がオクルージョンの場合、この視点では見えないものと

判断し画素値は求めない．これを 3 次元表面モデル上のすべてのボクセルについて行うことで任意視点画像を得る．

3.3 床平面の合成

床はシルエット作成時に背景として取り除かれるため，先に述べたモデル作成法では床の三次元モデルは作成されない．そこで，前景として 3 次元モデルを生成した後に，別の手法を用いて床平面を前景の 3 次元モデルと合成することにする．その方法であるが，床平面をいくつかの 3 角メッシュに区切り，頂点を F 行列を用いて投影し，その 3 角メッシュ内部は Affine 変換を用いて投影することで行う．まずその際用いる射影グリッド空間の性質について述べる．

射影グリッド空間の座標軸は基底カメラの設定によって定義されているため，射影グリッド空間内部ではユークリッド空間における直線や平面の方程式は成り立たない．そのため射影グリッド空間中に直線や平面の方程式を用いて仮想平面を作成することはできない．しかし画像中における物体を一度，射影グリッド空間の座標系に投影し，それをもとに本手法のように任意視点画像を作成した場合，任意視点画像におけるそれらの物体の前後関係や，平面や直線などの幾何関係は保存される．

本手法ではこれらの性質をふまえ，背景差分に使用した各カメラの背景画像を用い実際に画像に映り込んでいる床の画像からその床平面上に存在する点を 3 点以上取り出し，それらを射影グリッド空間の座標系に投影し，それらをもとに任意視点画像の床平面を合成する手法をとる．

まず基底カメラ間で床面上に対応点を数点取得する．ここでは F 行列を求める際床平面から取得した対応点を用いてもかまわない．射影グリッド空間の定義から，取得した対応点 A の座標が基底カメラ 1 の画像で $A_1(x_1, y_1)$ ，基底カメラ 2 の画像で $A_2(x_2, y_2)$ の場合この点 A の射影グリッド空間における 3 次元座標は $A(x_1, y_1, x_2)$ になる (図 6)．射影グリッド空間における 3 次元座標が決定できればこの点は今までの処理と同様にカメラ間の F 行列を用いてどの入力画像へも投影することができる．

床平面の任意視点画像を作成する場合 3 角メッシュを最小単位とするため，取得した基底カメラ間の対応点を基底カメラ 1 の画像中で 2 次元的にドロネー網を作成する．ここではあくまで取得した点群を 3 角メッシュに分割するだけでよいので 3 次元的な情報は用いる必要はない．こうして作成されたドロネー網の 3 角メッシュの頂点を人物モデルの場合と同様に F 行列を

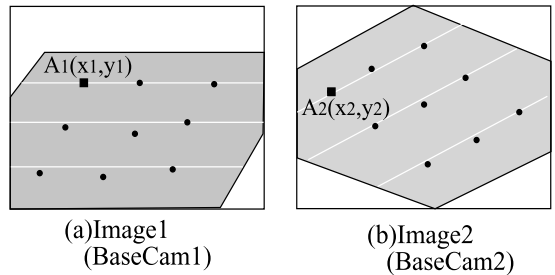


図 6 対応点の取得

Fig. 6 Pointing the correspondence points between two basis images.

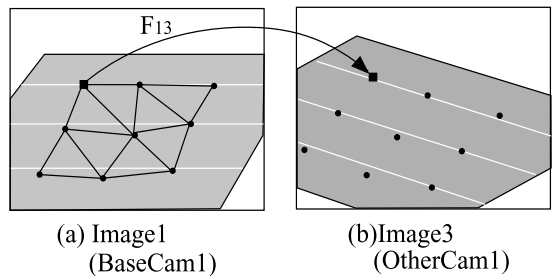


図 7 ドロネー網の作成と他画像への投影

Fig. 7 Making the Delaunay meshes, and projecting it on to the non-basis images.

用いて各背景画像に投影する (図 7)．

次に先ほどドロネー網によって作成された 3 角メッシュを用い任意視点画像における床平面を合成する．各頂点の任意視点画像での座標は 3 次元モデルのときと同様の手法 (3.2 節) で決定する．

まず合成したい視点付近の背景画像を 2 枚選択する．そして各 3 角メッシュの頂点の 2 枚の画像における対応点の座標に，画像間の距離を表すウェイトで重みづけした値を任意画像におけるその頂点の座標に決定する (図 8)．そして 1 つの 3 角メッシュを形成する 3 つの頂点の任意視点画像における座標が決定したとき，その座標と各入力画像における座標とをもとに Affine 行列を計算する．そして最後に任意視点画像におけるその 3 角メッシュ内部を逆変換に基づき Affine 変換によって内挿する．この際各画素の画素値は両画像のその点の対応点に頂点のときと同様のウェイトで重みづけされたものになる．この操作をすべての 3 角メッシュで行い，任意視点画像の床平面を合成する．

最後に，以上のように作成された床平面の任意視点画像と，人物の任意視点画像を合成する．それぞれ手法が違うため両者同時にレンダリングすることはできないが，床はあくまで人物の背景であるという明示的な関係から，はじめに床平面をレンダリングし，その画像上に人物を合成するという順序で行うことで人物

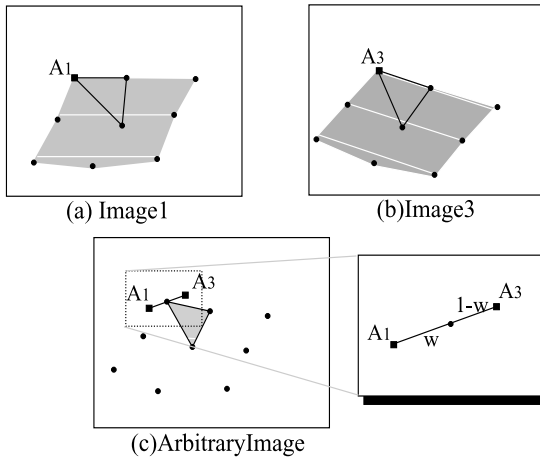


図 8 任意視点画像の合成
Fig. 8 Synthesizing the arbitrary view images.

と床の任意視点画像を作成する。

3.4 任意テクスチャ画像の床平面への合成

前節に述べた手法では、各カメラにより実際に撮影された背景画像から床のテクスチャを床平面として合成することは可能であるが、これに対して、任意のテクスチャを使用した画像を合成したい場合も考えられる。この場合、任意のテクスチャ画像を撮影された背景画像のいずれかに射影できれば、任意テクスチャが重畳された背景画像から、前節に述べた方法と同じ手法により床が合成できるが、任意のテクスチャ画像を背景画像に射影する射影変換を決定することは一般には不可能である。そこで、天井カメラは床とほぼ垂直の関係にあり、また任意テクスチャ画像は、テクスチャを持った平面に垂直に設置したカメラから撮影された画像になっていることが多いことに着目し、天井カメラ画像に任意テクスチャ画像を適当なスケールで貼り付け、それを背景画像に重畳し、前節で述べたような床面合成を行うものとした。なお、鏡面反射などの光学的要素は考えないものとする。

具体的には、床の領域を基底カメラの背景画像間の対応点を用いて、前節と同様に取得し、基底カメラ 1 の背景画像にドローネー網を作成する(図 9(a))。作成したドローネー網の各三角メッシュの頂点を床面と垂直に設置された天井カメラの背景画像に F 行列を用いて投影する(図 9(b))。次に投影された点と同座標の点を床に使用する任意テクスチャ画像上にとる(図 9(c))。そして最後に各三角メッシュの基底カメラ画像と天井カメラ画像間の Affine 行列を計算し、Affine 変換を用いて基底カメラにおける任意テクスチャの床平面を合成する(図 9(d))。このようにして、任意テクスチャ

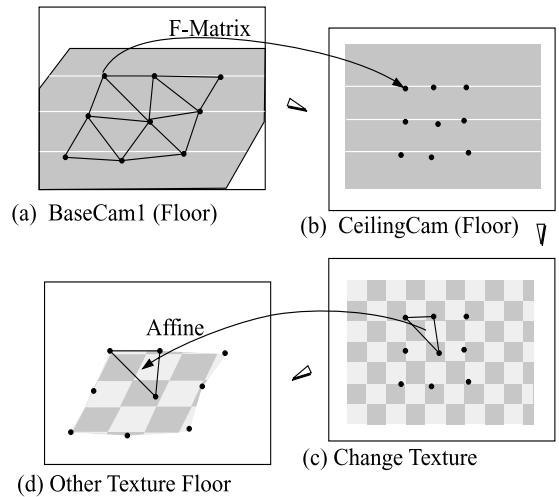


図 9 任意背景との合成
Fig. 9 Synthesizing foreground object and background image.

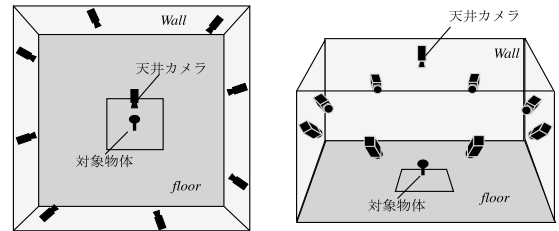


図 10 カメラ配置
Fig. 10 Camera location.

の床平面を合成することができる。

4. 実験

実験に使用した画像は、大分県別府市のピーコンプラザ大ホールの壁面および天井に設置された、壁面 8 台、天井 1 台の計 9 台のカメラで撮影したものである。カメラの配置状況を図 10 に示す。壁面の 8 台のカメラは、床面からの高さがほぼ同じ位置に対象を取り囲むように設置されている。これらのカメラから得られた画像列のうちの 20 フレームに対し本手法を適用し、各フレームごとに 3 次元モデル、壁面の 8 台の各カメラ間での任意視点画像の合成を行った。1 枚の画像サイズは 640×480 画素の RGB 画像である。入力画像例を図 11 に示す。

入力画像に対し背景画像を用いて背景差分、ノイズ除去などを行ってシルエット画像を作成し 3 次元モデルを作成した。射影グリッド空間に作成された 3 次元モデルを、便宜的に P, Q, R 軸を X, Y, Z 軸に対応させることによりユークリッド空間で表し、任意の視点から合成した画像を図 12 (図 11 と同一フレー

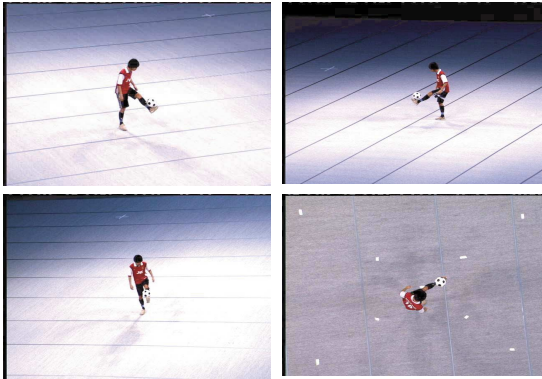


図 11 入力画像例 (frame 1)
Fig. 11 Example of input images.



図 12 3D モデル (frame 1)
Fig. 12 Reconstructed 3D shape model (frame 1).

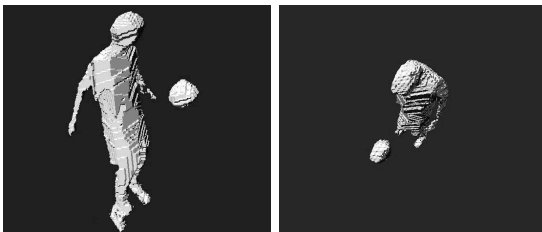
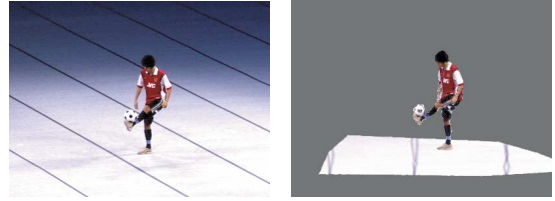


図 13 3D モデル (frame 10)
Fig. 13 Reconstructed 3D shape model (frame 10).

ム), 図 13 (図 11 と異なるフレーム) に示す .

次に作成した 3 次元モデルを用いて合成した任意視点画像を図 14 に示す . (a) は入力画像として用いた画像であり , (b) が (a) のカメラと隣接するカメラ間でウェイトを 5:5 にして合成した画像である . これらと比較すると , 隣接する 2 視点のちょうど中間の視点でも大きな画質劣化なしに , 良好な合成画像が得られていることが分かる . また同一フレームで 4 視点間においてウェイトを変化させ合成した画像を図 15 に示す .

上下の画像が入力画像である . この各 2 枚の画像間の中間の画像をウェイトを変化させて合成したものが任意視点の画像にあたる . 2 枚の入力画像の中間視点から見たボールと人物の位置関係 , 人物の配色などの観点から , 結果画像が良好に作成されていることが分



(a) 入力画像 (b) 任意視点画像

図 14 入力画像と任意視点画像の比較
Fig. 14 The comparison of image quality between input image and synthesized image.

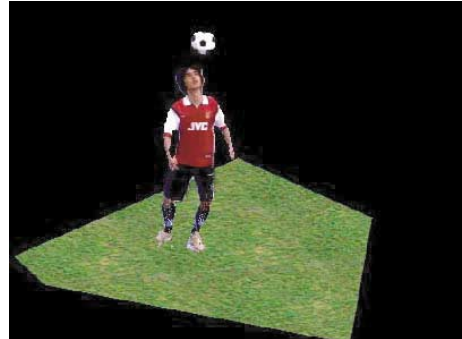


図 17 任意のテクスチャとの合成
Fig. 17 The object synthesized with arbitrary texture.

かる . また各フレームに対し画像 2 , 3 間でウェイトを 3 : 7 にし合成した任意視点画像を図 16 に示す . 人物の輪郭などにノイズが生じているが , これはシルエットの誤差などによって , 3 次元モデルを入力画像に投影した際に , 入力画像の人物領域よりも外側の画素も任意視点画像合成に用いてしまったためと考えられる . このノイズの影響を軽減するため , 2 枚の画像を合成する際のウェイトを , エッジ強度に応じて変化させる⁵⁾などの対策が考えられる . 最後に任意のテクスチャ (芝) を床面に合成した画像を図 17 に示す . 天井のカメラを利用し , 視点方向の関係で生じる違和感をなくすことで , 人物と任意テクスチャの背景が自然に合成されていることが分かる .

5. 考察・検討

以上の結果から未校正のカメラから得られた画像列から入力視点にない画像を合成し良好な結果が得られた . 実験を通して明らかになった課題などについて述べる .

本手法を適用するにあたり , 各カメラの配置に関する検討が必要である . 本手法は 2.2 節で述べたように SS 法による基底カメラ以外のカメラにおける投影点決定の際に , 各基底カメラとの間のエビポーラ線の交点を利用する . このとき F 行列の推定誤差の影響によ

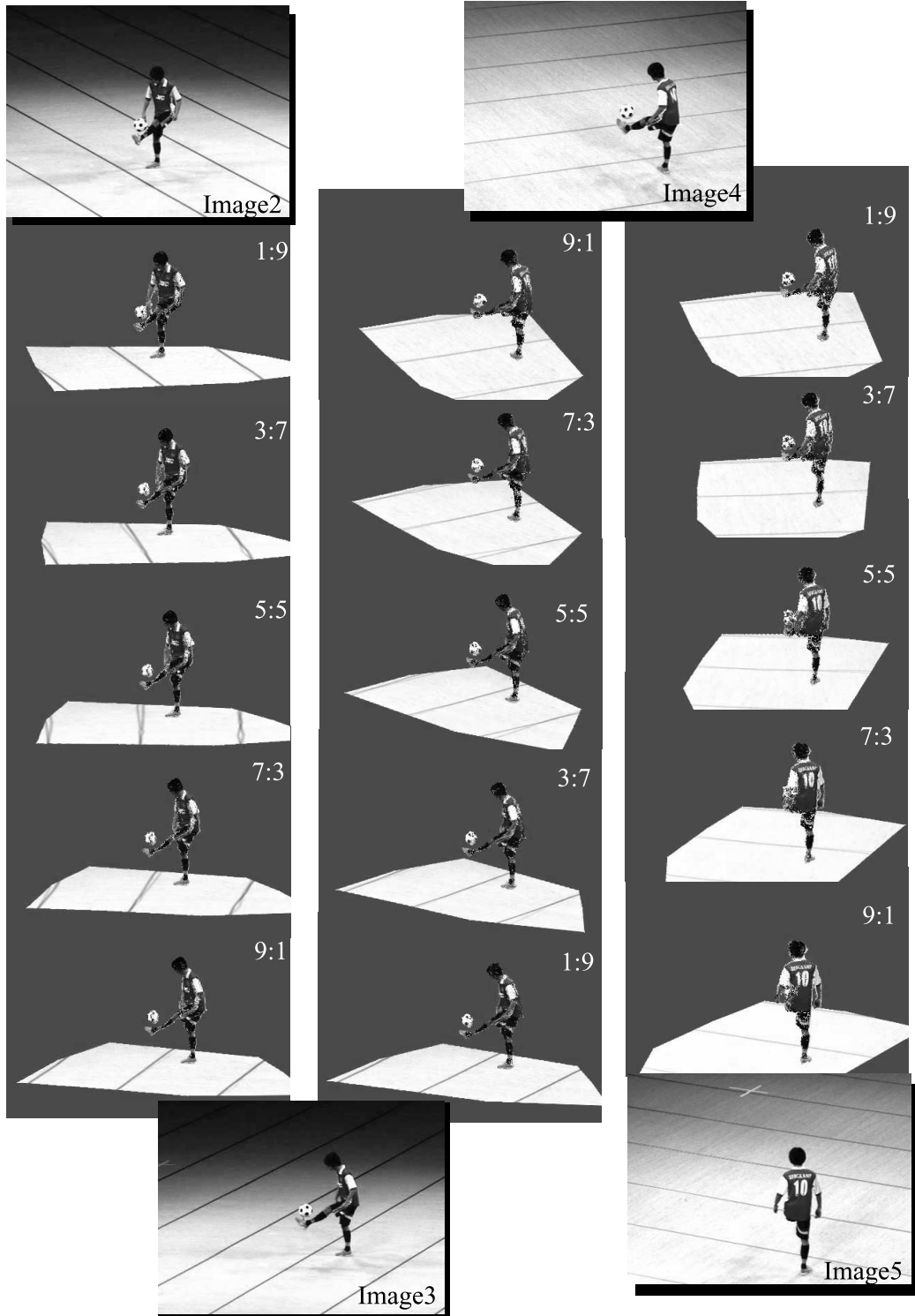


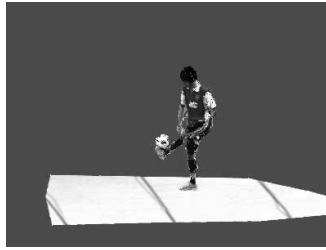
図 15 weight を変えて作成した任意視点画像

Fig. 15 Arbitrary view images changing weighting factor.

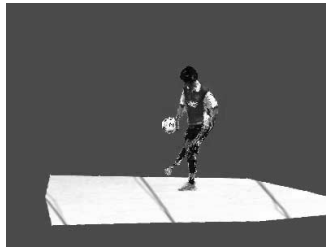
Input Image (Camera2)

Image of Arbitrary View
(weight = 0.7)

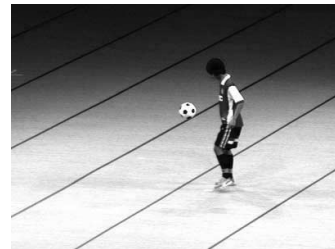
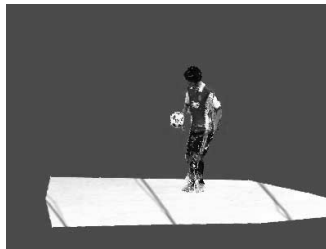
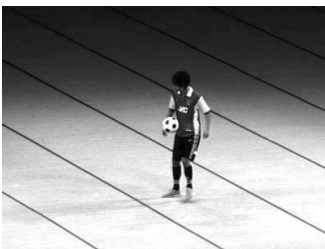
Input Image (Camera3)



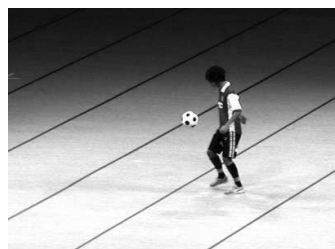
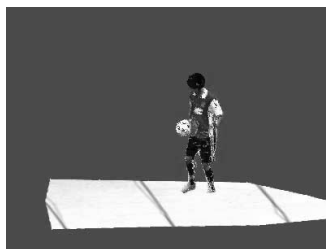
(a)Frame 0



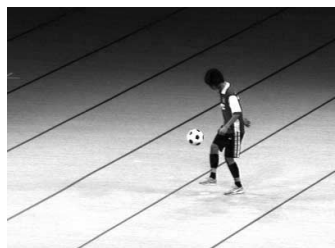
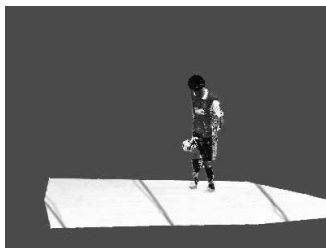
(b)Frame 4



(c)Frame 8



(d)Frame 12



(e)Frame 16

図 16 カメラ 2-3 間, weight=0.7 の場合の連続画像

Fig. 16 Image sequence. Camera 2-Camera 3, weight=0.7.

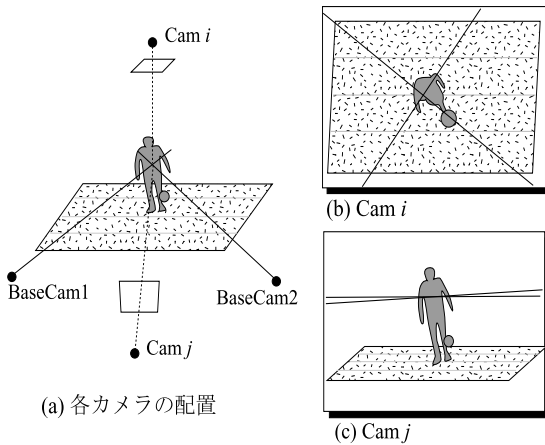


図 18 エピポーラ線の交差

Fig. 18 Cross of the two epipolar line.

る問題がある。

図 18 の (a) において 2 台の基底カメラと空間中の注目点がなすエピポーラ平面を考える。カメラ j はその平面上に近いところ、カメラ i はその平面の垂直上方に近いところからそれぞれ注目点を見ているものとする。この場合、各カメラからの各基底カメラとの 2 本のエピポーラ線の見え方はそれぞれ (b), (c) のようになる。図を見て分かる通り、カメラ j から 2 本のエピポーラ線を見ると交差角が非常に小さく平行に近い。一方カメラ i から見た場合 2 本のエピポーラ線は直交に近い角度で交差する。このとき前者は F 行列の推定誤差が及ぼす影響が大きく、たとえ推定誤差が非常に小さくても交点を実際交差すべき場所から大きくずれてしまう場合がある。

特にカメラと対象物との距離が十分大きい場合、対象物体上のあらゆる点におけるエピポーラ線どうしが平行に近くなるため、前者のような誤差は対象物体全域で生じることになる。本手法で適用した SS 法のアルゴリズムは、注目ボクセルが各シルエット画像のうち 1 枚でもシルエットの外側にある場合、そのボクセルを取り除くというアルゴリズムであるため、カメラ j のような画像を用いると作成される 3 次元モデルが必要以上に削られてしまう場合がある。

これらの対策として、事前にエピポーラ線の交差角が小さくなるカメラの画像は用いないで 3 次元モデルを作成する、すべてのカメラの配置を基底カメラのエピポーラ線の交差角が大きくなるように工夫するなどが考えられるが、すべてのカメラにこのような条件をあてはめるのは難しく、また任意視点画像は各 2 画像間の中間画像として得られるため、省かれたカメラとの間の任意視点画像が合成できなくなるという問題が

ある。

今回用いたカメラシステムは人物を取り囲む周囲の壁のほぼ同じ高さには 8 台、天井に 1 台の計 9 台のカメラからなるため、任意に基底カメラセットを選択しても、そのうちの 1, 2 台が上記のような条件を満たさない場合があった。そのため条件を満たさないカメラ画像とその両隣のカメラ画像間の中間画像を合成する際は、条件を満たさなかったカメラとその両隣のカメラに対しこの条件を満たすような基底カメラのセットを選択し合成することによって行った。

しかし処理の汎用性を考えると基底カメラ 1 セットですべてのカメラ間の中間画像ができることが望ましい。そこで上記のような条件を満たさない場合、一度たとえば天井に設置されたカメラで交点を求め、その交点のエピポーラ線を目的の画像に投影し、片方の基底カメラのエピポーラ線との交点を求めることで、エピポーラ線の交差角が小さくならないようにすることが考えられる。今回このような手法は試みていないが、今後検討すべき課題である。

また、今回合成した任意視点画像は、射影グリッド空間で再構成された 3 次元モデルから、隣接する入力カメラ画像間の対応点を密に求め、それらの対応点から任意の比率で画素値を単純に補間することにより合成した。しかし、この方法では、隣接するカメラの光軸が平行であり、また 2 台のカメラの光軸周りの回転角が等しい場合でない限り、隣接するカメラ間の幾何学的関係を正確に反映したものにはならない。対象物体とカメラとの距離が大きい場合は、ある程度光軸が平行でなくても、合成画像に及ぼす影響は小さいが、カメラの回転角が違う場合は、単純な補間に基づく方法では、合成画像が大きく歪む場合がある。今回の実験では対象とカメラ間の距離が遠く、また各カメラが水平面に対してほぼ回転のない状態に設置したため、任意視点画像はほとんど歪の影響なく合成することができた。しかし今後、手法の汎用性を考え、Seitz らの提案した隣接するカメラ間の幾何学的関係を正確に反映させた View Morphing²⁰⁾などの方法を適用して、より幾何学的に正確な任意視点合成を行う予定である。

6. む す び

未校正の数台のカメラ画像からの情報のみをもとに、3 次元形状再構築、任意視点の画像合成を行う手法を提案し、良好な結果を得た。本手法の特徴は中心射影によって定義される仮想 3 次元空間である射影グリッド空間上に、3 次元モデルを作成することである。射影グリッド空間に画像間の関係を表す F 行列を用いる

ことで、カメラ校正を必要とせずに3次元再構成を可能にしている。

今回の実験では3次元モデル作成に1分程度、また任意視点画像合成に数秒程度の処理時間がかかっているが、今後の課題として連続画像に適用した際、実時間処理が可能のように3次元形状再構築にかかる計算量をさらに削減する必要がある。またユーザがマウスなど簡単な操作だけで視点を自由に变化させられるインタフェースの作成、任意の背景との合成を行う予定である。

謝辞 本実験に用いた多視点画像は、通信・放送機構委託研究(地域提案型研究開発制度)「仮想化現実技術による自由視点三次元映像スタジオ通信の実証研究」の一環として撮影されたものである。本研究プロジェクトの関係各位に感謝する。

参 考 文 献

- 1) Beier, T. and Neely, S.: Feature-Based Image Metamorphosis, *Proc. SIGGRAPH '92*, pp.35-42 (1992).
- 2) Chein, C.H. and Aggarwal, J.K.: Identification of 3D Objects from Multiple Silhouettes using Quadrees/Octrees, *Computer Vision, Graphics and Image Processing*, Vol.36, pp.100-113 (1986).
- 3) Chen, S. and Williams, L.: View Interpolation for Image Synthesis, *Proc. SIGGRAPH '93*, pp.279-288 (1993).
- 4) Curless, B. and Levoy, M.: A Volumetric Method for Building Complex Models from Range Images, *Proc. SIGGRAPH '96* (1996).
- 5) Debevec, P.E., Taylor, C.J. and Malik, J.: Modeling and Rendering Architecture from Photographs: A hybrid geometry- and image-based approach, *Proc. SIGGRAPH '97* (1997).
- 6) Gavrilu, D.M. and Davis, L.S.: 3-D Models Based Tracking of Humans in Action: Multi-View Approach, *Proc. Computer Vision and Pattern Recognition 96*, pp.73-80 (1996).
- 7) Gortler, S.J., Grzeszczuk, R., Szeliski, R. and Cohen, M.F.: The Lumigraph, *Proc. SIGGRAPH'96* (1996).
- 8) Hilton, A., Stoddart, J., Illingworth, J. and Winder, T.: Reliable Surface Reconstruction from Multiple Range Images, *Proc. ECCV'96*, pp.117-126 (1996).
- 9) Katayama, A., Tanaka, K., Oshino, T. and Tamura, H.: A view point dependent stereoscopic display using interpolation of multi-viewpoint images, *SPIE Proc.*, Vol.2409, *Stereoscopic Displays and Virtual Reality Systems II*, pp.11-20 (1995).
- 10) Kanade, T., Rander, P.W., Vedula, S. and Saito, H.: Virtualized Reality: Digitizing a 3D Time-Varying Event As Is and in Real Time, *International Symposium on Mixed Reality (ISMR99)*, Yokohama, Japan, pp.41-57 (Mar. 1999).
- 11) 北原 格, 大田友一, 斎藤英雄, 秋道慎志, 尾野 徹, 金出武雄: 大規模空間における多視点映像の撮影と自由視点映像生成, 3次元画像工学コンファレンス (Jul. 2000).
- 12) 北原 格, 大田友一, 金出武雄: 多視点映像の融合によるスポーツシーンの自由視点映像生成, 画像の認識・理解シンポジウム (MIRU'2000) 論文集, Vol.I, pp.13-18 (Jul. 2000).
- 13) Levoy, M. and Hanrahan, P.: Light Field Rendering, *Proc. SIGGRAPH'96* (1996).
- 14) Masuda, T. and Yokoya, N.: A Robust Method for Registration and Segmentation of Multiple Range Images, *Computer Vision and Image Understanding*, Vol.61, No.3, pp.295-307 (1995).
- 15) 大田友一: 複合現実感のための実環境入力技術の展望, 情報処理学会コンピュータビジョンとイメージメディア研究会, 119-13, Vol.99, No.93, pp.95-99 (Nov. 1999).
- 16) Potmesil, M.: Generating Octree Models of 3D Objects from Their Silhouettes in a Sequence of Images, *Computer Vision, Graphics and Image Processing*, Vol.40, pp.277-283 (1987).
- 17) Saito, H. and Kanade, T.: Shape Reconstruction in Projective Grid Space from Large Number of Images, *IEEE Proc. Computer Vision and Pattern Recognition*, Vol.2, pp.49-54 (1999).
- 18) Saito, H., Baba, S., Kimura, M., Vedula, S. and Kanade, T.: Appearance-Based Virtual View Generation of Temporally-Varying Events from Multi-Camera Images in 3D Room, Computer Science Technical Report, CMU-CS-99-127 (Apr. 1999).
- 19) 斎藤英雄, 金出武雄: 多数のカメラによるダイナミックイベントの仮想化, 情報処理学会コンピュータビジョンとイメージメディア研究会, 119-13, Vol.99, No.93, pp.117-124 (1999).
- 20) Seitz, S.M. and Dyer, C.R.: View Morphing, *Proc. SIGGRAPH '96*, pp.21-30 (1996).
- 21) Seitz, S.M. and Dyer, C.R.: Photorealistic Scene Reconstruction by Voxel Coloring, *Proc. Computer Vision and Pattern Recognition (CVPR)*, pp.1067-1073 (1997).
- 22) Tsai, R.: A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cam-

eras and Lenses, *IEEE Journal of Robotics and Automation* RA-3, 4, pp.323-344 (1987).

- 23) Vedula, S., Rander, P.W., Saito, H. and Kanade, T.: Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences, *Proc. 4th Conf. Virtual Systems and Multimedia*, Vol.1, pp.326-322 (1998).
- 24) Wheeler, M.D., Sato, Y. and Ikeuchi, K.: Consensus surfaces for modeling 3D objects from multiple range images, *DARPA Image Understanding Workshop* (1997).
- 25) 山本正信, 川田 聡, 近藤拓也, 川越和忠: ロボットモデルに基づく人間動作の3次元動画追跡, 信学論(D-II), Vol.J79-D-II, No.1, pp.71-83 (1996).
- 26) 米元 聡, 布巻 崇, 鶴田直之, 谷口倫一郎: 多視点動画における3次元多関節物体の追跡, 画像の認識・理解シンポジウム(MIRU'98)論文集, Vol.II, pp.423-428 (Jul. 1998).
- 27) Zhang, Z.: Determining the Epipolar Geometry and its Uncertainty: A Review, INRIA research report, 2927 (1996).

(平成12年9月15日受付)

(平成13年1月22日採録)

(担当編集委員 内海 章)



矢口 悟志

平成12年慶應義塾大学理工学部情報工学科卒業。現在同大学大学院修士課程在学中。コンピュータビジョン, 画像処理に関する研究に従事。電子情報通信学会学生会員。



木村 誠

平成7年慶應義塾大学理工学部電気工学科卒業。平成10年同大学大学院博士課程入学, 現在在学中。コンピュータビジョン, および画像処理の分野にて研究に従事。電子情報通信学会学生会員。



齋藤 英雄(正会員)

昭和62年慶應義塾大学理工学部電気工学科卒業。平成4年同大学大学院博士課程修了。同年同大学助手, 専任講師を経て平成13年助教授。工学博士。コンピュータビジョン, 画像処理等に関する研究に従事。平成9年から平成11年まで, 日本学術振興会海外特別研究員として米国カーネギーメロン大学に留学。平成12年より, 科学技術振興事業団さきがけ研究21(「情報と知」領域)研究員兼任。電子情報通信学会, 計測自動制御学会, IEEE各会員。



金出 武雄

昭和48年京都大学大学院博士課程修了, 同年同大学情報工学科助手。昭和51年同助教授。昭和55年米国カーネギーメロン大学計算機科学科高等研究員。昭和60年同教授。現在 U.A. and Helen Whitaker 全学教授, 同大学ロボティクス研究所所長。工学博士。計算機視覚, ロボットの腕, 自律走行車, VLSI センサに関する研究に従事。米国アカデミー外国特別会員。IEEE Fellow, AAAI Fellow, ACM Fellow。