コンピュータビジョンと
イメージメディア 119−16
(1999. 11. 18)

多数のカメラによるダイナミックイベントの仮想化

斎藤 英雄†* 金出 武雄*

† 慶應義塾大学理工学部情報工学科

*カーネギーメロン大学ロボティクス研究所

本稿では，多数のカメラにより撮影される動画像シーケンスから，動きのあるイベントの３次元再構築を行い，それを仮想空間に入力し，さらに仮想空間で表示するVirtualized Reality を目的とした最近の我々の研究について述べる．この目的のために，我々は５０個程度のCCDカメラを部屋の壁面と天井に取り付け，各カメラからの画像をリアルタイムでデジタル化してPC に取り込むことのできるシステム “3DRoom” を構築した．この 3DRoom により撮影された多数のカメラからのデジタル動画像列を用いて，動きのあるイベントの３次元形状を再構築し，再構築された３次元情報を利用することによって任意視点からの画像を自由に合成することが可能となる．本稿では，3DRoomの構成，そしてそれにより得られた画像からの３次元再構築法，任意視点画像の合成法について述べ，それにより得られた任意視点画像を示す．

Virtualizing Dynamic Real Events from Muliple Cameras

Hideo Saito †* Takeo Kanade*

† Department of Information and Computer Science, Keio University

*Robotics Institute, Carnegie Mellon University

In this paper, we present recent research of “Virtualized Reality”, that aims to create virtual model of dynamic real events from multiple view image sequences. We develop “3D Room”, for digitizing all images of multiple cameras mounted on the ceiling and walls of the room. This article presents the way to create the virtual model of the input dynamic events. The mesh representation of the object is reconstructed via multiple camera stereo based method. We introduce the virtual appearance view generation method is based on simple interpolation between two selected views. The correspondence between the views are automatically generated from the multiple images by use of the volumetric model shape reconstruction framework.

1 はじめに

シーンの３次元モデリング，レンダリングに関する研究は，近年のコンピュータのデータ処理容量及び能力の増大により，非常に盛んになってきた．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盛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にした研究がコンピュータビジョン及びコンピュータグラフィックスの分野で非常に盛んになってきている．

CMUにおけるVirtualized Reality では，多視点のCCDカメラシステムからの画像入力による３次元モデル獲得とレンダリングによって，動きのあるイベントの仮想化を行うことにより任意視点画像を生成する研究に取り組んでいる。本稿では，このCMUにおいて開発された，多視点カメラシステムである “3DRoom” システムを紹介する。そして，3D Room により撮影された多視点画像列から３次元モデルを

獲得し，任意視点画像を生成する方法について述べ，この手法に基づいて生成された任意視点画像を示す。

2 関連研究

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(モデルベースドレンダリング)，そして，入力画像から直接任意視点画像を合成するもの(イメージベースドレンダリング)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 et. al. [6], Curless and Levoy[3], Masuda and Yokoya [12], そして Wheeler et. al. [21] 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，大抵の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出来なかった．

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 et. al. 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]. Levoy and Hanrahan [10] and Gortler et al. [5] 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．

我々の研究している Virtualized Reality では，動きのあるイベントを対象としているために，上記のように３次元スキャナの利用や，大量の視点からの入力画像を用いることができないので，現実的に利用可能な数１０個程度のCCDカメラからを用いて，そこから入力される画像列を用いている．そして，対象の３次元構造を複数のカメラの画像からのステレオマッチングに基づいて構築するアプローチを取っている．

このように多くのカメラを用いたシステムとしては，Davis et. al. [4] が開発した人間のモーションキャプチャシステムが挙げられる．ここでは，シーンの３次元形状を再構成する代りに，人体の関節角を多視点の画像から推定している．

さて，入力された視点の異る画像から任意視点画像を合成するための手法として，View interpolation [2, 20] と呼ばれる任意視点画像合成法がある．この手法は，複数の入力画像間の対応関係からその中間の画像を内挿により合成するものである．View morphing [16] は，image morphing [1] の拡張であり，ここでは，カメラ間の３次元的構造を正確に扱うことができるようなアルゴリズムが提案されている．いずれも，各画像間の対応関係に基いた中間画像生成法である．本稿で紹介する任意視点画像生成法も，この画像の対応関係に基づくものであるが，対応関係を入力された多視点画像から再構築した３次元モデルから与えることにより画像生成の自動化を実現している．

3 3D Room

“3D room” [8] は，動きのあるイベントの３次元情報をコンピュータ内にデジタル化して取り込む－つまり４次元デジタル化－するためのシステムとして開発したものである．これは，これを目的として最初に構築された “3D Dome” [7, 13, 19] では複数のCCDカメラからの画像を一旦ビデオテープに記録したのちにデジタル画像化するという手順をとるものであったところを，画像撮影とリアルタイムでデジタル化するという点で大きな改善がなされている。.

さて，CMUで開発した 3D Room [8] では，図1及び図2に示すように，約６メートル四方，高さ約2.7メートルの部屋の４つの壁面に各10台，天井に９台のカメラが現時点で取り付けられている．これらの４９台のカメラは共通の外部同期信号で同期が取られ，さらにカメラからの映像信号には共通のタイムコード情報が，白黒のビットパターンの画像情報として，画像上部数本の走査線(実際にはNTSC

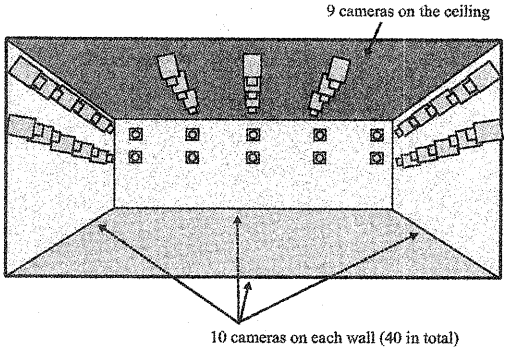


図 1: 3D Room におけるカメラ配置

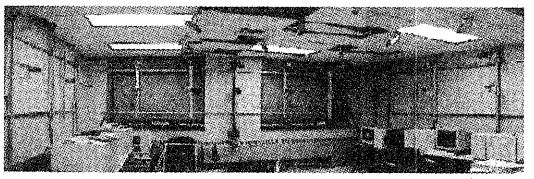


図 2: 3D Room の概観

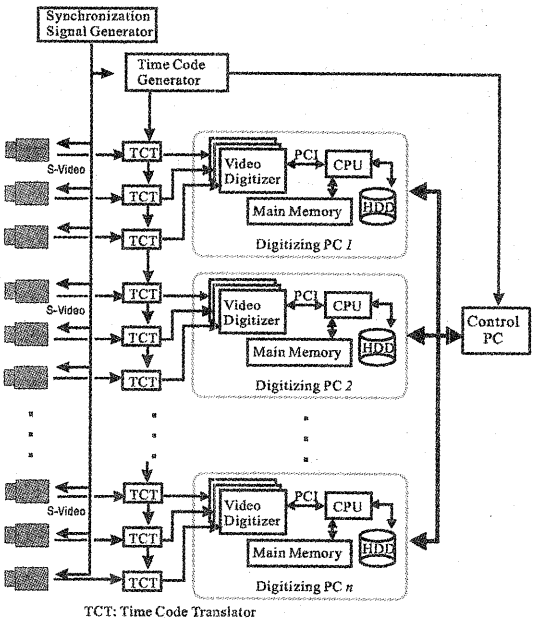


図 3: 3D Room におけるディジタル画像取得システム.

画像領域の外側になる)に書き込まれている．タイムコードが書き込まれた映像信号は，画像デジタイザ(米Imagenation社PXC200)により，640×480画素のカラー画像(1画素あたり2バイト :輝度，色差にそれぞれ1バイト／画素)として30フレーム／秒でフレーム落ち無しで直接PCにデジタル化して取り込まれる．デジタイザはPC 1台あたり3枚同時にフレーム落ち無しでデジタル化が可能となるため，合計17台のPC クラスタを構成し，デジタル画像を取得する仕組みになっている．図 3 に，このデジタル化システムの構成を示す.

4 3次元モデルの取得

上記の 3D Room により撮影された多視点動画像列から，対象の3次元構造を再構築する．この3次元再構築には，マルチベースラインステレオ法(MBS)[11]を用いて各カメラ毎に求めた距離画像を体積空間で融合し，3次元モデルを構築する手法を用いる.

具体的には，各カメラ毎に近傍のカメラを2～4個組合せて，基準となるカメラに対して，近傍の各カメラのエピポーラ線上でN×N画素(Nは5～9画素程度) の範囲のパターンのマッチングを行い，各カメラに対するマッチング評価値の合計から対応点を探索し，距離を推定する．この処理を基準カメラ画像の各画素について行い，距離画像を得る.

こうして各カメラ毎に得られた距離画像は，体積空間で融合される．この融合では，各距離画像が物体表面の推定位置を与えると考え，各ボクセル毎に物体表面までの推定距離を計算する．そして，この推定距離＝0となる点の集合が表面となると考え，この表面を与える距離情報をマーチング ・キューブアルゴリズムにより抽出する．こうして抽出された表面は3角メッシュの集合で表現されるが，このメッシュの数を削減し，3次元モデルとする．図5に，この処理によって復元された3次元モデルの例を示す.

さて，上記のような手法で物体の3次元形状を復元するためには，各カメラが完全にキャリブレーションされている必要がある．我々は，Tsaiにより提案されたカメラモデルとキャリブレーションの手法[18]を適用した．この手法により，6自由度のカメラ外部パラメータと，焦点距離，画素のアスペクト比，光軸の画像平面上の2次元座標，そしてレンズのラディ

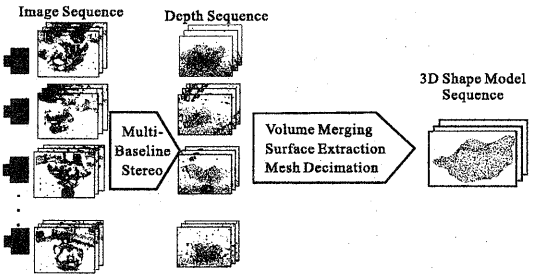


図 4: 多視点画像から３次元形状モデルを得るまでの流れ

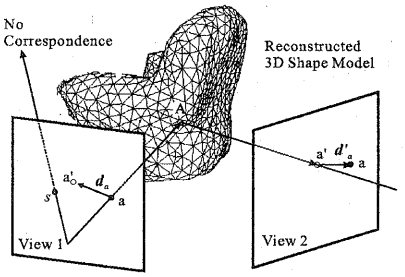


図 6: ３次元形状モデルから求められる対応点

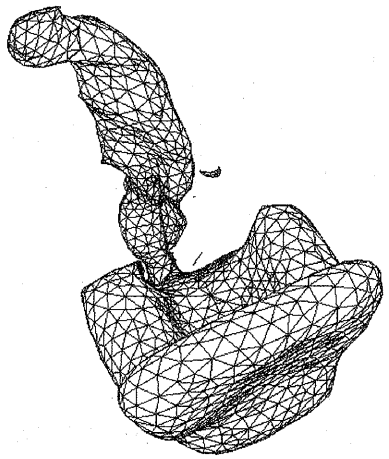


図 5: 再構成された３次元形状モデルの例(三角メッシュ数10,000).

アル歪の第一次パラメータの合計５つのカメラ内部パラメータを推定する．この推定には，あらかじめ対象空間内に３次元位置既知の点を複数与え，これがカメラに投影される位置を検出する必要があるが，このために300mm毎に８×８個のLEDが配置されたプレートを作成し，上下方向に位置を５種類の位置で変化させた画像を撮影した．

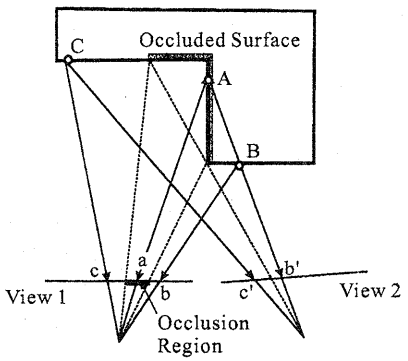
5 任意視点画像の生成

上記のようにして構築された３次元モデルを用いて，任意視点の画像をレンダリングする手法について述べる．我々は，２つのアプローチを試みた．一つは，再構築された３次元モデルの各三角メッシュ上に画像からテクスチャを割り当て，そして，任意視点からテクスチャマッピングによりレンダリングする手法[19]．である．そしてもう一つは，合成しようとする視点付近の２枚，もしくは３枚の入力画像を選び，これらの入力画像間の対応関係を復元された３次元モデルから求め，この対応関係に元づいて見かけ上の中間画像を合成する手法[14]である．

前者では，三角メッシュを各画像面に投影し，投影された三角領域内のテクスチャを三角メッシュのテクスチャとして割り当てる．このとき，同じ三角メッシュを見るカメラは複数あるため，複数のカメラ間でテクスチャを平均する．そして，任意視点からの画像を合成する際に割り当てられたテクスチャを各三角メッシュにレンダリングするものである．

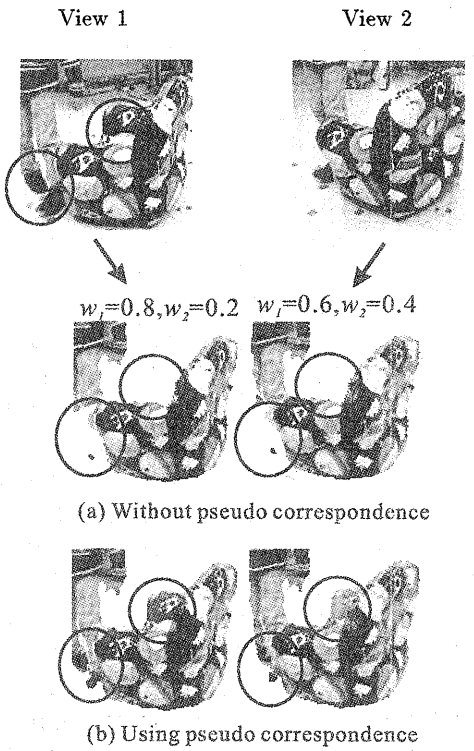
この手法の場合，各三角メッシュに割り当てられるテクスチャは複数のカメラの入力画像の平均となるため，推定した３次元モデルの誤差にがあると，複数のカメラに投影された三角領域のテクスチャに位置ずれを生じる．そして，位置のずれを起こした複数のテクスチャを平均するために，結果として割り当てられるテクスチャにゴーストのようなぼけを生じてしまう．このため，もし，入力画像を撮影したカメラと同じ位置で任意視点の合成を行ったとしても，このテクスチャの劣化のために，その位置で実際に入力された画像に比べると画質が劣るという問題がある．

一方，後者の手法では，まず再構築した３次元モデルから画像間の対応点を図6のようにして求める．View1の点aは，表面上の点Aを貫く．そして，これは，View2のa'に投影される．この場合，View2におけるa'がView1におけるaに対する対応点となる．もし，物体上に貫く点がなければ，View1にお



A a→b' :pseudo correspondence
B b↔ b':consistent correspondence
C c↔ c':consistent correspondence

図 7: 無矛盾対応と擬似対応.



(a) Without pseudo correspondence

(b) Using pseudo correspondence

図 8: 中間画像生成の際の擬似対応利用の効果. もし2枚の画像しか与えれていなければ, ○で示されるような隠れた領域の合成は不可能であるが, 再構成されている3次元モデルから得られる擬似対応を用いることによって, 隠れた領域に対する中間画像の生成が可能になっている.

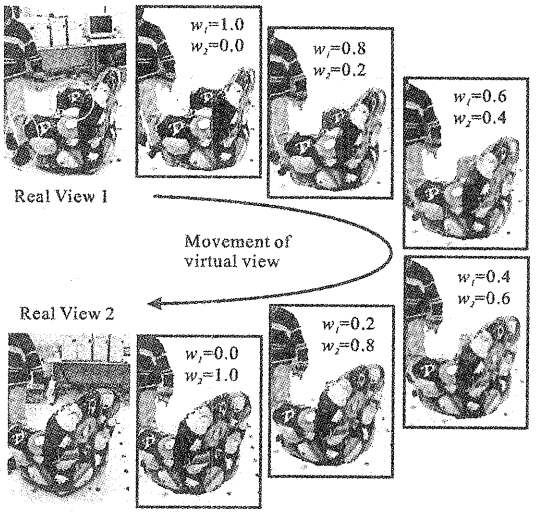


図 9: 2つの画像から生成された中間画像群. 中間画像の仮想視点は, それぞれの重みを変化させることにより動かすことができている.

ける s のように, 対応点を持たないことになる. 対応点を持つ各点について, 視差ベクトルが, 例えば点 a については, a から a' へのベクトル $\mathbf{d}_a$ のように定義できる. この視差ベクトルは, 図6における点 A のように, 片方の画像には投影されるがもう一方には投影されないといった場合については, View1 の a に対しては擬似的に点 a から点 b' へのベクトルとして定義することにする. このような対応関係を, 擬似対応 (pseudo correspondence) と呼ぶ. 一方, 図6 における点 A や, 図7における点 B や点 C に関する対応関係については, 両者の画像両者に投影され, 両者で視差ベクトルが定義でき, これを無矛盾対応 (consistent occlusion) と呼ぶ.

このような視差ベクトルを用いれば, 視差ベクトル上の中間の位置に画素値をマッピングすることにより, 画像間の中間の視点の画像 (部分中間画像) を合成できる. そして, 各画像について合成された画像をさらに合成して, 中間画像を生成する. この中間画像を合成する際に, 上記のように擬似対応の情報も用いることにより, 隠れの影響の少ない画像が合成できる. また, この際には, 任意視点は選択したカメラ間の合成比率 (重み) として与えるため, 入力画像と同じ視点の画像を合成しようとする場合は, 同じ視点の入力画像の重みが1となり, それ以外は

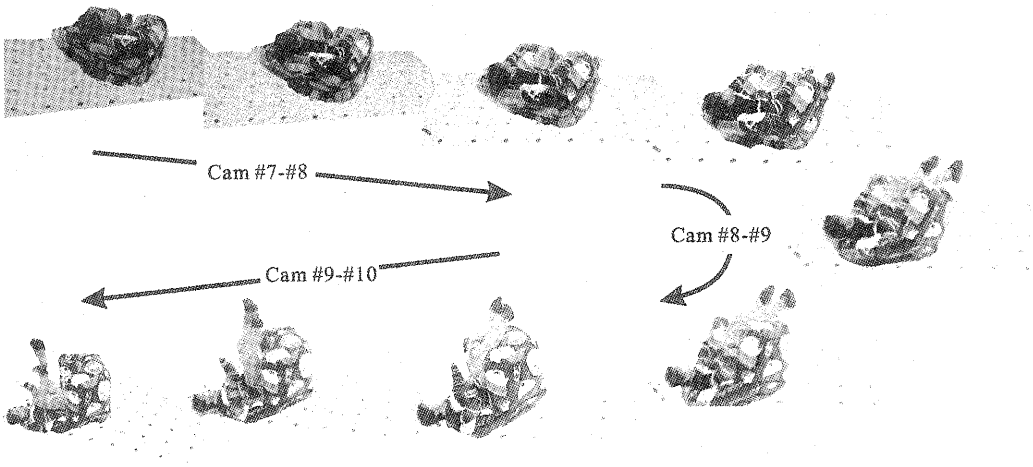


図 10: ４つのカメラに撮影された動画像から生成された任意視点画像列の例

０となることから，３次元モデル復元に誤差が含まれていても，合成される画像は全く入力画像と同じ画質のものが得られる．この性質のため，合成される任意視点動画像列の画質は，３次元モデルの各三角メッシュに，複数の視点から見える画像の部分テクスチャを割り当てる手法に比べ，良好なものとなることが期待できる。

図8に，擬似対応の利用の効果が示されている．もし，この図に示した２枚の入力画像しか与えられていない場合は，View1において○で囲まれたエリアの隠れ領域については対応関係が得られないので，結果として合成される中間画像では完全にこの領域の画像情報が消えてしまう(図8(a))．一方，多視点画像から再構築して得られた３次元モデルを利用して擬似対応を求め，これを用いて中間画像を合成した場合は，隠れ領域についても中間画像で再現されていることがわかる (図8(b))．

図9に，同一の時刻において２つのカメラにより撮影された入力画像より生成した中間画像の例を示す。２つのカメラの合成比を変化させることにより，見かけ上の視点を変化させることができる．また，擬似対応の利用により，隠れの発生するような領域についても良好な画像生成を行うことができている．また，図10に，複数のカメラの画像列から，仮想的に視点移動したときに生成した画像列を示す．

6 多視点画像における射影幾何学の利用

この3D Room のように多数のカメラを利用して３次元復元を行おうとする場合，各カメラのキャリブレーションが非常に重要な要素となる．

ここで示した実験結果を得るために行ったカメラキャリブレーションでは，各カメラについて同時に撮影されたマーカ点からキャリブレーションを行ってはいるものの，各カメラ独立にカメラパラメータの推定を行っているために，各カメラ間で推定されたカメラパラメータ間に矛盾が推定誤差により含まれていることが予想される．具体的には，各カメラで独立に推定されたカメラパラメータから計算したカメラ間のエピポーラ幾何と，各カメラに共通に見えるマーカ点から検出したカメラ間の対応関係から推定したエピポーラ幾何とに微妙な誤差を発生してしまうことがわかった．この誤差は，特に離れたカメラ間で大きく，時には，エピポーラ線の誤差が約数画素程度のものになってしまうこともあった．

このような多数カメラのシステムにおいては，基本的には複数のカメラ間で共通に見える点を検出して，三角測量の原理により３次元構造を復元するものであるため，カメラ間の相対的幾何学関係を出来るだけ正確に扱うことが重要となる．また一方で，多

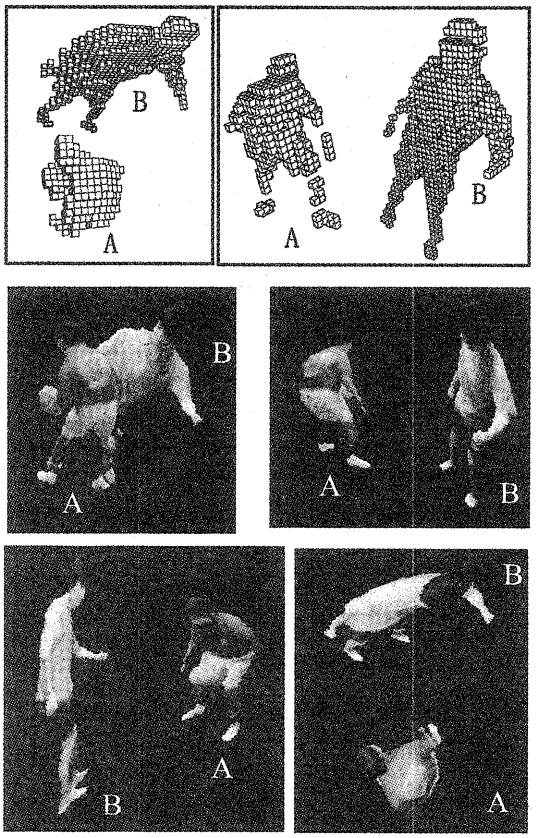


図 11: 射影グリッド空間における形状再構成の例.

数のカメラに対して同時に3次元位置の既知なマーカを撮影させてキャリブレーションするということの手間も大きな問題である．そこで，カメラ間の相対的幾何学関係を明示的に表現し，しかも，3次元点の未知な幾つかのマーカの対応関係情報のみから推定できるエピポーラ幾何に基づいて多数カメラから3次元復元を行うことの重要性は高い．

そこで，この多数のカメラからの射影幾何学に基づいた3次元復元を行うために，カメラ間のエピポーラ幾何から“射影グリッド空間”を構成し，この射影グリッド空間において3次元復元する枠組を提案した[15].

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，通常は，各カメラ毎に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毎に推定する必要がある．これに対し，我々の提案した“射影グリット空間”では，このグリッド空間と画像上の点との関係を，カメラ間のエピポーラ幾何を表すFundamental行列のみを用いて記述することができるため，カメラ毎に関する射影行列を復元することなしに，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．

その構成法は，多数のカメラの中から2つの基底カメラを選び，これらの基底カメラ間でFundamental行列を求め，このFundamental行列により，3次元のグリッド位置を定義する．このグリッド位置と基底カメラ以外のカメラの画像位置との関係は，基底カメラとのFundamental行列により記述される．そこで，各カメラと基底カメラとのFundamental行列を推定しておくだけで，基底カメラによって定義される3次元グリッド空間の任意の点と各カメラの2次元位置の関係がわかることになるので，多数のカメラからの3次元再構成が可能になる．この枠組では，全カメラ数をNとすると，2つの基底カメラ間のFundamental行列，それぞれの基底カメラと，それ以外の各カメラ間のFundamental行列の合計$1+(N-2)\times 2$組のFundamental行列のみから3次元復元を行うことが可能になる．Fundamental行列の推定には，位置のわかっているマーカ点を設定する必要が無く，幾つかの3次元位置の未知な点に関してカメラ間の対応関係を検出するだけで良いので，多数のカメラからの3次元復元を行うために必要となるカメラキャリブレーションの手間を削減できる．

図11に，3D Roomの前身である3D Domeの51個のカメラから撮影された画像から，提案する射影グリッド空間の枠組を用いて再構成した射影3次元形状の例を示す．ここでは，各カメラ毎に射影行列やカメラ・パラメータを推定することなく，カメラ間のFundamental行列のみを用いることにより，3次元形状が再構成されている．上の図において，表示されているボクセルのサイズは，実際に体積再構成のために定義しらボクセルサイズの4倍(体積比64倍)となっている．実際のボクセルサイズは，射影グリッド空間を定義する基底カメラの投影中心からの距離に比例して大きくなる．この例では，射影グリッド空間を定義する基底カメラが人物Bの後ろに位置しているため，人物B付近の実際のボクセルサイズが人物A付近よりも小さくなり，この結果人

物Aが小さく表現されている.

なお, この再構成には, Seitzらにより提案されているVoxel Coloring [17] の手法を用いている. 彼らの元の手法では, 各カメラのカメラ・パラメータを推定する必要があったが, 我々の提案する射影グリッド空間の枠組を適用することにより, 各カメラのカメラ・パラメータの推定無しに3次元再構成を行うことができている.

7 おわりに

動きのあるためのイベントの仮想化のためにCMUで開発した, 多数のカメラからの動画像列を収集するためのシステム"3D Room"を紹介した. そして, この3D Roomを用いて多数のカメラにより撮影された動画像列から, 動いている対象の形状を3次元復元する手法を紹介した. また, 仮想視点における画像を合成するために画像の内挿に基づく方法について述べ, この手法により合成した任意視点画像の例を示した. また, 射影幾何学に基づいたFundamental行列による多視点画像からの形状再構成法を紹介した.

参考文献

[1] T. Beier, S.Neely, "Feature-Based Image Metamorphosis", *Proc. of SIGGRAPH'92*, pp.35-42, 1992.

[2] S. Chen, and L.Williams, "View Interpolation for Image Synthesis", *Proc. of SIGGRAPH'93*, pp.279-288, 1993.

[3] B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images", *Proc.of SIGGRAPH '96*, 1996.

[4] D.M.Gavrila and L.S.Davis, "3-D Model Based Tracking of Humans in Action : Multi-View Approach",*Proc. Computer Vision and Pattern Recognition 96*, pp. 73-80, 1996.

[5] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen, "The Lumigraph", *Proc. of SIGGRAPH'96*, 1996.

[6] A. Hilton, J. Stoddart, J. Illingworth, and T. Windeatt, "Reliable Surface Reconstruction From Multiple Range Images", *Proc. of ECCV'96* pp.117-126, 1996.

[7] T. Kanade, P. W. Rander, and P. J. Narayanan, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes", *IEEE MultiMedia*, Vol.4, No.1, 1997.

[8] Takeo Kanade, Hideo Saito, and Sundar Vedula, "The 3D Room: Digitizing Time-Varying 3D Events by Synchronized Multiple Video Streams", *CMU-RI-TR-98-34*, 1998.

[9] A. Katayama, K. Tanaka, T. Oshino, and H. Tamura, "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] M. Levoy and P. Hanrahan, "Light Field Rendering", *Proc. of SIGGRAPH'96*, 1996.

[11] M. Okutomi and T.Kanade, "A Multiple-Baseline Stereo", *IEEE Trans. on PAMI*, Vol.15, No.4, pp.353-363, 1993.

[12] T.Masuda, N.Yokoya, "A Robust MEthod for Registration and Segmentation of Multiple Range Images", Computer Vision and Image Understanding, Vol.61, No.3, pp.295-307, 1995.

[13] P. J. Narayanan, P.W.Rander, and T.Kanade, "Constructing Virtual Worlds using Dense Stereo", Proc. ICCV '98, 1998.

[14] H. Saito, S. Baba, M. Kimura, S. Vedula, T. Kanade, "Appearance-Based Virtu al View Generation of Temporally-Varying Events from Multi-Camera Images in the 3D Room", Proc. International Conference on 3D Imaging and Modeling (3DIM99), pp.516-525, Ottawa, Oct., 1999.

[15] H. Saito, T. Kanade, "Shape Reconstruction in Projective Grid Space from Large Number of Images", Proc. Computer Vision and Pattern Recognition (CVPR'99), Vol.2, pp.49-54, Fort Collins, CO, June 1999.

[16] S.M.Seitz and C.R.Dyer, "View Morphing", *Proc. of SIGGRAPH'96*, pp.21-30, 1996.

[17] S.Seitz, C.Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring", Proc. Computer Vision and Pattern Recognition (CVPR97),pp.1067-1073, 1997.

[18] R. Tsai. "A Versatile Camera Calibration Technique for High- Accuracy 3D Machine Vision Metrology Using Off-the-Shelf Tv Cameras and Lenses",*IEEE Journal of Robotics and Auto mation* RA-3, 4, pp.323-344, 1987.

[19] S.Vedula, P.W.Rander, H.Saito, and T.Kanade, "Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences", *Proc. 4th Conf. Virtual Systems and MultiMedia*, Vol.1, pp.326-332, 1998.

[20] T. Werner, R. D. Hersch, and V. Hlavac, "Rendering Real-World Objects Using View Interpolation", In IEEE Int'l Conference on Computer Vision:ICCV95, pp.957-962, 1995.

[21] M.D. Wheeler, Y. Sato, and K. Ikeuchi, "Consensus surfaces for modeling 3D objects from multiple range images", *DARPA Image Understanding Workshop*, 1997.