

## 多視点カメラを用いた記録と提示

齋藤 英雄  
慶應義塾大学  
理工学部情報工学科  
saito@ozawa.ics.keio.ac.jp

北原 格  
ATR  
知能ロボティクス研究所  
kitahara@atr.jp

木村 誠 \*\*, \*\*  
\*産業技術総合研究所デジタル  
ヒューマン研究センター  
\*\*CREST, 科学技術振興機構  
makoto.kimura@aist.go.jp

多視点カメラを用いた記録と提示に関連した研究について、1) 多視点カメラのキャリブレーション、2) 多視点画像からの3次元復元、3) 多視点画像からの自由視点画像生成と提示、4) 多視点カメラを利用した監視・サーベイランス、という4つの観点毎に最近の動向を紹介する。

## Capturing and Displaying from Multiple Cameras

Hideo Saito  
Dept. of Information and  
Computer Science,  
Keio University  
saito@ozawa.ics.keio.ac.jp

Itaru Kitahara  
ATR  
Intelligent Robotics and  
Communication Laboratories  
kitahara@atr.jp

Makoto Kimura \*\*, \*\*  
\* Digital Human Research Center,  
National Institute of Advanced  
Industrial Science and Technology  
\*\* CREST, JST  
makoto.kimura@aist.go.jp

In this survey article, we introduce recent researches related to capturing and displaying from multiple cameras. We categorized the related researches into 4 groups: 1) Calibration of multiple cameras, 2) 3D shape reconstruction from multiple-viewpoint images, 3) Synthesize and display of free-viewpoint images from multiple-viewpoint images, and 4) Surveillance via multiple cameras.

### 1 はじめに

コンピュータビジョンの研究は、カメラで撮影された画像データをコンピュータに入力し、それを処理することにより行われる。80年代から90年代前半までは、上記の枠組みを実現するためのハードウェアの制約から、処理対象が静止画像に限られていたり、動画像を処理対象にするような場合は、一旦テープ等の媒体に保存してから処理のためにコンピュータに入力する、ということが必要とされていた。しかし、1990年代後半から、コンピュータへの画像データ入力及びデータ蓄積容量の増大によって、複数のカメラに

撮影される動画像をコンピュータに入力し、蓄積することが容易に行えるようになってきた。このような、コンピュータビジョン研究に利用可能なハードウェアの進歩に伴って、近年、複数のカメラを用いて同一シーンを多視点から記録し、記録した多視点画像から対象物体形状を求めたり、対象シーンの解析を行ったり、また、それらの情報を新しい形の画像として提示したりする、といった研究が盛んに行われるようになってきた。

このサーベイでは、このような、複数のカメラにより撮影された多視点画像を入力としたコンピュータビジョン関連の研究について紹介する。それぞれの研究

事例の目的や応用は様々であるが、本稿では、

- 多視点カメラのキャリブレーション (2章)
- 多視点画像からの3次元復元 (3章)
- 多視点画像からの自由視点画像生成と提示 (4章)
- 多視点カメラを利用した監視・サーベイランス (5章)

という4つの処理に注目し、紹介する。

## 2 多視点カメラのキャリブレーション

一般的にカメラ・キャリブレーションと呼ばれる処理には、カメラの位置や姿勢などを求める処理 [Tsai86, Lenz88, Wei94] を指す場合と、カメラの色補正などを校正する処理を指す場合 [Porikli03] があるが、本稿では前者を扱うものとする。

コンピュータ・ビジョンの処理の多くで、カメラ・キャリブレーションは既知であることが前提となっており、特に多視点カメラを用いている場合には、その傾向が顕著である。それらの処理でキャリブレーションを行う目的は、対象となるシーン内の3次元座標と撮影される画像中の2次元座標の間の射影関係を求めることや、複数のカメラ間の相対的な関係を求めることにある。そのため、キャリブレーションと言っても実際にはカメラの個々のパラメータ (焦点位置、姿勢、焦点距離など) は必要とされていないケース [Wei94] もある。例えばステレオ・ビジョンがその代表的な例として挙げられる。ステレオ・ビジョンでは、3次元空間と画像の間の射影関係さえ得られれば、個々のパラメータは必要とされない場合が多い。

しかしながら、従来提案されている手法の多くは、実際には何らかのカメラ・モデルを想定し、そのモデルのパラメータを推定する形でキャリブレーションを行っている。その際、用いられるモデルとしてはピンホール・カメラが一般的である。また、理想的なピンホール・カメラのモデルと実在のカメラの間の違いとして、最も大きなものとしては、レンズによる歪み (radial distortion) が挙げられる。この歪みについても別途モデル化して対処した手法は存在する。

### 2.1 固定カメラのキャリブレーション

まず、1台の固定カメラに対するキャリブレーションについて述べる。Tsai [Tsai86] は、3次元座標とその点が投影される画像中の2次元座標の組合せを大量に計測し、それらのデータ群からキャリブレーションを行った。さらに、Tsaiら [Lenz88] はモデル化されたレンズによる歪みとカメラの各パラメータについて非線形推定を行った。この他にも、形状既知の物体をシーン内に設置して撮影するようなキャリブレーション手法が数多く [Hartley00] 存在する。これらの手法は、以降に述べる手法に比べ一般的に安定性や精度などの面で長けており、複数カメラを用いるような場合であっても、全カメラで共通の3次元座標を基準に固定カメラのキャリブレーション手法を適用することは多い。初期のCMUでのVirtualized Realityプロジェクト [Narayanan98, Vedula98, Saito03] では約50台のカメラに対してTsai手法が用いられていた。

しかしながら、このような手法で必要とされるデータを正確に得るためには、シーン内に精密なオブジェクトを正確に配置する [北原02] ような作業が必要であり、その作業が繁雑であるという欠点がある。したがって、例えば屋外環境のように撮影環境が不自由な場合や、システムの制約でキャリブレーションに時間がかけられない場合など、このような手法が適用できない場合も考えられる。

### 2.2 セルフ・キャリブレーション

カメラで撮影した画像中から、3次元位置が未知の点を用いてキャリブレーションするような、セルフ・キャリブレーションと呼ばれる手法が存在する。

初期のセルフ・キャリブレーション手法の多くは、全ての画像においてカメラの内部パラメータが不変であるという制約を設けていたため、移動ロボットのカメラや、複合現実感 (Augmented Reality: AR) のように、主にカメラが移動するシステムで多用されてきた。

しかしながら、近年では内部パラメータが未知な条件でも適用可能なキャリブレーション手法が提案 [Pollefeys99, Matsunaga01] されている。これらの手法では、ズームやフォーカスを変化させて撮影した画像列から、トラッキングされた特徴点や平面を元にセルフ・キャリブレーションを行うことが可能である。Vi-

Room プロジェクト [Svoboda02] では、そのような手法を利用して、3 台以上のカメラが存在する場合に適用可能なキャリブレーションを用いている。

### 2.3 平面を用いたキャリブレーション

Zhang[Zhang98] は、セルフ・キャリブレーションの手法を応用し、簡易な入力データによるカメラ・キャリブレーション手法を提案した。Zhang 手法では 1 台の固定カメラの撮影シーン内で平面を動かして撮影する。撮影された平面を固定された X-Y 平面と想定し、相対的にカメラが移動したという問題に置き換えることにより、カメラ・キャリブレーションを行う。これは、撮影対象と画像がどちらも 2 次元の平面であることと、カメラの内部パラメータが固定であることをふまえて、後述の Homography を中間に利用することによってキャリブレーションを行うものである。Zhang 手法は、Zhang 自身による実装の他にも MatLab や OpenCV に実装されており、広く用いられている。例えば 128 台のカメラを擁する The Stanford Multi-Camera Array システム [Wilburn02] や、最近の CMU での Virtualized Reality Project [Cheung03] において Zhang 手法が用いられている。

Zhang 手法を用いて複数カメラを共通の座標系でキャリブレーションする場合、対象となる平面を複数カメラで同時に撮影する必要がある。植芝 [植芝 02] らは、そのように複数カメラを用いることを前提とした場合について、Zhang 手法における無駄な中間処理で安定性や精度が下がる可能性があることを指摘し、複数カメラに最適化された改良手法を提案した。また、数十台のカメラに対して適用する際の改良手法 [Baker00] や、複数カメラを入力とした Image-Based-Rendering のための改良手法 [Sebe02] など提案されている。

### 2.4 弱キャリブレーション

複数のカメラを用いたシステムでカメラ間の相対的な関係のみを記述する、いわゆる射影幾何のキャリブレーションが存在する。射影幾何のキャリブレーションは、前述のキャリブレーション手法に比べて簡易なデータで得られるが、3 次元座標と画像中の画素の間の射影関係は未知のままである。そのため、前述のよ

うなキャリブレーション手法との対比で、3 次元座標とカメラの間のキャリブレーションは強キャリブレーション (Strong Calibration)、射影幾何のキャリブレーションは弱キャリブレーション (Weak Calibration) と呼ばれる。

Homography は平面の射影を記述する  $3 \times 3$  行列である。ある共通の平面を複数のカメラで撮影し、平面上の対応点を 4 点用いることにより、Homography を線形解法で求めることが可能である。一度 Homography が求まれば、その平面上の任意の点をカメラ間で相互に射影することが可能である。前述の平面を用いた手法や、セルフ・キャリブレーションの手法で利用されることが多いが、それ以外にもコンピュータ・ビジョン全般で多用されている手法である。

Fundamental Matrix [Zhang96, Hartley95] は 2 つのカメラ間のエピポラ幾何を記述する  $3 \times 3$  の行列であり、画像間の任意の 7 つ (非線形解法) あるいは 8 つ (線形解法) の対応点から求めることが可能である。Fundamental Matrix は内部パラメータと外部パラメータを全て内包している。内部パラメータが既知である場合には、Fundamental Matrix から、外部パラメータを内包する  $3 \times 3$  の行列 (Essential Matrix) を得ることができる。従って、内部パラメータが既知であれば、Fundamental Matrix から強キャリブレーションを行うことが可能 [Hartley97, Ihrke04] である。Fundamental Matrix や Homography による弱キャリブレーションのみを用いて複数画像から Image-Based-Rendering を行う手法 [Seitz96, 木村誠 02, Avidan98, 稲本 02] も研究されている。また、弱キャリブレーションしかできていない多視点カメラのうち 2 台のカメラを利用して定義される「射影グリッド空間」を利用して、3 次元形状復元を行う手法も提案されている [Saito99, Iso03]。

より多数のカメラを用いる場合には、3 枚の画像間のエピポラ幾何を記述する Trifocal Tensor や、4 枚の画像間のエピポラ幾何を記述する Quadrifocal Tensor といった手法が存在 [Hartley00] する。また、あるシーンを取り囲むように多数のカメラが存在する際に、各カメラ間の相互投影を利用してエピポラ幾何をより安定的に求める手法 [杉村 03] も提案されている。

### 3 多視点画像からの3次元復元

画像という2次元空間に投影された3次元空間情報を、画像から求めることを目的にした研究はコンピュータビジョンの主要な分野を形成している。その代表である、2台のカメラを用いたステレオ法に関する研究は、長い間多くの研究が行われてきた。この章では、カメラの台数を増やすことにより撮影された多視点画像からの3次元復元について、最近の研究事例について紹介する。なお、1台のカメラを移動する等して視点を変化させた画像列からの3次元復元法も多視点画像からの3次元復元と位置づけることも可能であり、ここでは、そのような多視点画像からの3次元復元も取り上げる。

#### 3.1 形状復元法

**マルチベースラインステレオ** 上記のような2台のカメラを用いたステレオ法を2台を超える数のカメラに適用した手法が、奥富と金出により提案されている [Okutomi93]。このマルチベースラインステレオ法 (MBS) は、1つのカメラにおける奥行き分布を求めるために、各画素の周辺領域の濃度パターンに一致するパターンを、そのカメラの周囲の数台のカメラから同時に探索することによって、誤対応を減らすものである。Narayananらは、このMBS法を50台ものカメラにより撮影された多視点画像に適用し、撮影対象の3次元形状を復元した [Narayanan98]。しかし、対象を取り囲むように取り付けられた多視点カメラでは、たとえば50台とはいえ、カメラ間の基線 (Baseline) 長が長くなり、その結果オクルージョンが多発してしまうという問題があった。このような問題を解決するための手法として、オクルージョンを検出し、影響を軽減する手法 [松浦 97, KangSB01] や、視点位置に依存しないアフィン不変量を画像の角特徴に対して求め、RANSACを利用したマッチングによって広い基線長でも安定に角特徴の対応関係を安定に求める手法 [Pritchett98] や、奥行き分布の局所的な滑らかさを制約条件として利用した最適化を利用することによって、長い基線長でも良好な結果が得られる手法 [Strecha03] が提案されている。また、物体表面の鏡面反射が存在しても良好な深さ分布が得られる手法も提案されている [Li02]。

**形状体積データ復元** 上記のように、ステレオ法に基づいた形状復元法は、対象を取り囲むように取り付けられた多視点カメラでは基線長が長くなるため、適用が困難になる。これに対して、多視点画像に撮影された対象のシルエットとカメラの射影中心が形成する「シルエットコーン」を多視点画像間で交差させ、交差する部分の体積を求めることにより対象形状を復元する「視体積交差法」が提案されている [Laurentini94]。この手法を利用した形状復元の基本的アルゴリズムは、3次元空間をボクセルで表現し、各ボクセルを各カメラの画像に投影したとき、全ての画像についてシルエット画像内部に含まれているボクセルの集合を復元形状とするというものである。この計算を高速化するための手法として、ウらは、各ボクセルについてカメラに投影するのではなく、空間を平面スライスとして表現し、各平面毎にカメラに投影する、平面間投影に基づく視体積交差法の高速化と、その並列化手法を提案した。そして、この手法をPCクラスタ上に実装することにより、画像撮影から形状復元までを高速に実行可能なシステムを構築した [ウ 01]。同様にPCを複数利用して視体積交差法の高速化を実現したシステムとしては、[Cheung00, 濱崎 01] 等が報告されている。

上記のような視体積交差法では、各視点のシルエットコーンの内側に凹んでいる部分については形状復元が行われぬ。このような部分を、異なる視点間で対応画素の色の同一性をチェックすることにより削り取ることによって、より詳細な形状を復元可能にした手法が、Seitzらの提案したVoxel Coloring法/Space Carving法 [Seitz97, Kutulakos00] である。この手法では、対象3次元ボクセル空間をスライスする平面を走査しながら、その平面より外側の空間に存在するカメラを可視カメラとして、この可視カメラ間で各ボクセルに対応する画素の色が一致した場合に物体表面ボクセルと判定し、そうでない場合は物体外ボクセルと判定して削り取りながら形状復元を行っていく。このため、画素の色の一致判定の正確さが重要となる。そこで、このような問題を改善するために、各ボクセルに不透明度という連続的な値を導入することによって、(半)透明な物体が存在するような場合でも形状復元を可能にしようとした手法 [Bonet99] が Bonetらによって提案された。同様の改善法として、画素の色の一致判定に確率を用いて誤判定を減らす手法 [Broadhurst01] や、各ボクセルが物体表面である可能性を連続量として表現して形

状復元する手法 [Agrawal01] が提案されている。また、Slabaugh らは、Space Carving 法において可視カメラの決定と、画素値の同一判定法に色々な改善を施すことによって良好な形状復元を可能とした [Slabaugh04]。これらの Voxel Coloring 法に基づいた手法では、各ボクセルにおいて、そのボクセルに対応した画素値の均一性を評価して物体表面ボクセルを決定するが、同じボクセルでも視線の違いによって画素値が変化してしまうことが誤差の原因となる。これに対して、[Kim04] らは、ボクセルと各画像との投影関係を考慮して、ボクセルではなく画像の画素上でボクセルに対する色の均一性を考慮して表面ボクセルを決定する手法を提案し、復元形状の精度が向上することを示した。また、物体表面が鏡面反射する場合でも Space Carving を可能にする手法も提案されている [Bonfort03]。

表面メッシュの復元 上記の手法は、対象空間を有限な個数のボクセル空間にサンプリングし、物体内部ボクセルの集合として対象形状を復元するものであった。しかし、実際の応用では、物体表面をポリゴンメッシュ表現等により必要とする場合がある。このとき、ボクセル集合から表面を抽出する代表的な手法として知られるのが、CT や MRI のボクセルデータをメッシュとして可視化することを目的に考案された Marching Cubes 法 [Lorenzen87] である。このような手法により抽出された物体表面形状には、ボクセルサイズの誤差が含まれてしまうことになる。この誤差を回避するために、Slabaugh [Slabaugh02] らは、流体解析の分野で 2 つの流体間の界面を数値解法により求めるために考案された「レベルセット法」を Space Carving 法に利用して、高精度な物体表面のポリゴンメッシュ表現を得る手法を提案した。延原らは、視体積交差法で復元した形状のメッシュモデルの高精度化をはかるために、ポリゴンメッシュを複数の視点で撮影された画像上に再投影し、このエリア内における画像テクスチャが一致するようにメッシュ頂点の 3 次元位置を修正していく手法 [延原 02] を提案した。

一方、上記のように対象形状のボリュームを復元するのではなく、目的である物体表面の多角形メッシュ表現そのものを、多視点画像から直接復元する手法が考案されている。Lazebnik [Lazebnik01] らは、多視点からのシルエットコーンが交差する形状表面を、曲線からなる一般化多角形により表現したメッシュ推定す

るための手法を提案している。しかし、ここで得られるは、物体表面形状の完全な多角形メッシュ表現ではなく、曲線分からなる「多角形」による表面形状表現である。Boyer ら [Boyer03] は、ある視点の画像 1 のエッジ点から、別の画像 2 に対してエッジ線を引き、画像 2 のシルエットと交差する線分を Viewing Edge として検出していき、このような Viewing Edge を色々な 2 視点を組み合わせて求め、それらを 3 次元空間中で連結することによりポリゴンメッシュ直接復元する手法を提案している。この手法では、Viewing Edge から最初に復元されるポリゴンメッシュは粗く、また完全に Viewing Edge が連結されない部分もあるため、局所的に形状を精密なものにする処理を 2 段階目の処理として行うことによって、ポリゴンメッシュの高精度化を図っている。山口ら [山口 03] は、未校正多視点カメラにより撮影されたシルエットを多角形近似し、この頂点から 2 台のカメラで定義される射影グリッド空間内に引いた直線によって、ポリゴンメッシュで表現されている対象形状モデル形状を逐次更新していく手法を提案した。

### 3.2 人体動作計測のための多視点画像の利用

多視点カメラを人体動作計測に利用する、いわゆる「モーションキャプチャ」を行おうとする研究も、近年盛んに行われている。

Gavrila らは、人体を多関節モデルと考え、多視点カメラで撮影された人体の多関節モデルの各部位の姿勢を 3 次元的に推定するための手法を提案した [Gavrila96]。ここでは、対象人物に腕、胸、といった各部位毎に色の異なるスーツを着せ、多視点画像において各部位をセグメンテーションしておく。そして、各部位に分割された人体の 3 次元モデルの各部位の姿勢を、体の中心に近い部位から階層的に推定している。

Ong らは、多視点カメラで撮影された人物の 2 次元シルエット形状と、それに対応する 3 次元スケルトンモデルとを結びつける関係を、あらかじめ学習によって求めておき、入力多視点画像のシルエット形状から 3 次元スケルトンモデルを推定する手法を提案した [Ong99]。

米元らは、多視点カメラシステムを高速ネットワークで接続することによって、各カメラに撮影された画像データをビデオレートで転送・処理することの可能

なシステムを構築した。そして、このシステムを用いて実時間で対象人物のモーションキャプチャを行い、仮想世界とのインタラクションを可能にした [米元 00]。

Cheung らは、入力多視点画像のシルエットを多関節モデルの各部位にセグメンテーションし、そして多関節モデル毎の形状と姿勢を求め、さらにその結果を利用して、入力多視点画像のシルエットを多関節モデルの各部位にセグメンテーションしていく、という繰り返しを行うアルゴリズム [Cheung03] を提案した。

Chu らは、あらかじめ用意しておく形状モデルやマーカーを利用しない多視点画像からのモーションキャプチャ法を提案 [Chu03] した。この手法では、人間が色々な姿勢を取っている時の 3 次元形状は、部分的な姿勢が異なるだけで基本的には同じトポロジーになっていることに着目し、入力多視点画像のシルエットから復元される 3 次元ボクセル形状モデルから人体の姿勢に依存しない成分を抽出するために Isomap [Tenenbaum00] を適用する。そして、抽出された姿勢不変 3 次元形状から人体のスケルトンを抽出して関節部位毎に分割する。そして、この分割を元の 3 次元ボクセル形状モデルに適用することによって部位毎の姿勢を推定するものである。

## 4 多視点画像からの自由視点画像生成と提示

多視点カメラにより撮影された多視点画像を利用して、実際に撮影したカメラ位置とは異なる新たな画像を生成し提示するという研究が近年盛んになっている。この手法は、ユーザの好みに対して「自由な視点位置」の画像を提示するものであり、自由視点画像生成と呼ばれる。

**3 次元モデル復元に基づく手法** このような自由視点画像は、対象シーンの 3 次元形状と、その表面の光学特性が既知であれば容易に行うことができる。そこで、前章に述べたような 3 次元復元を多視点画像から行えば、原理的には自由視点画像を生成可能である。

多視点画像からの 3 次元形状復元を基本にして自由視点映像を生成しようとする手法としては、まず、2 台のステレオ画像を入力として、それらとは異なる視点の画像を生成する [Scharstein99] ものが挙げられる。

これは、例えばテレビ会議システムなどにおける視線一致を実現しようとするものであり、多くの研究が行われている [Ott93, Yang02]。

カメラの台数を増やして 3 次元モデル形状全体をより高精度に復元し、自由視点映像を生成しようとする研究は、金出らが Virtualized Reality のコンセプトを提案 [Kanade95] して以来、様々な研究が行われてきた。金出らのグループでは、マルチベースラインステレオと視体積交差法の考え方を組み合わせて 50 台のカメラにより復元した形状から得られた表面のメッシュモデルの各メッシュに、入力画像のテクスチャを割り当てることによって、自由視点画像を生成した [Vedula98]。同様に Moezzi らは、17 台のカメラのシルエット画像から視体積交差法により Visual Hull を復元し、この表面を表すポリゴンメッシュに入力画像を割り当てることによって自由視点画像を生成するシステムを構築した [Moezzi97]。これらの手法では、形状復元の際の誤差が、メッシュ割り当てられたテクスチャ画像を劣化させてしまう。そこで、この画質劣化を軽減するために、仮想視点に隣接する 2 カメラの画像の視点内挿によって自由視点画像を生成した [Saito03]。さらに Vedula らは、視点だけではなく、時間軸方向への形状の内挿を行い、任意の時刻における自由視点画像を生成した [Vedula02]。Yaguchi らは、多視点カメラのキャリブレーションの手間を避けるために、数台の未校正カメラにより撮影された多視点画像から、射影グリッド空間内で 3 次元モデルを復元し、[Saito03] と同様の考え方により仮想視点に隣接する 3 カメラ画像から視点内挿を行うことによって自由視点画像を生成する手法を示した [Yaguchi04]。Kitahara らは、仮想視点が対象物体に近い場合でも遠い場合でも、十分な解像度の自由視点画像を合成可能にするために、対象までの距離に応じて異なるサイズのメッシュで 3 次元モデルを表現する手法を示した [Kitahara03]。

Carranza らは、対象を人と限定し、あらかじめシステムが所持している人の 3 次元モデル形状利用し、入力多視点画像のシルエットから人体の各部位の姿勢を推定し、さらにモデル形状を入力画像のテクスチャを用いて修正することによって、精度の高い 3 次元モデルを復元し、モデルを利用して自由視点画像を生成する手法を示した [Carranza03]。

Grau らは、ここで紹介しているような技術を応用して、登場人物に仮想環境で演技させ、それを放送用映

像の制作に利用するようなスタジオシステム [Grau04] を構築した。このシステムでは、映像確認用の低解像度映像はリアルタイムで合成可能であり、実際に方法に利用可能な画質は後処理で達成することが可能である。また、このような高速な自由視点映像生成を実現したシステムとして、Goldlucke らの構築した自由視点ビデオシステム [Goldlucke03] がある。このシステムでは、 $64 \times 64 \times 64$  という低解像度のボクセルで対象形状を高速に復元し、自由視点画像のレンダリングの際に、Yamazaki らが提案した Microfacet Billboarding [Yamazaki02] という、自由視点の視線方向に平行な法線を持つ小さな面素の集合で対象形状が表されているものと考えて自由視点画像を合成するものである。この Microfacet Billboarding の利用により、低解像度のボクセル形状でも、良好な画質の自由視点画像の生成が行える。

ETH Zurich の “Blu-c project” [Gross03] では、多視点カメラシステムと、没入型ディスプレイ CAVE を利用した仮想環境提示を組み合わせたシステムを開発した。このシステムでは、3面をスクリーンで囲まれた環境に存在する人物を多視点カメラで撮影して得られる人物のシルエットからボクセルモデル構築を行い、それに基づいてリアルタイム自由視点レンダリングを行いながら、スクリーンに仮想環境画像を提示することができる。スクリーンへの画像提示と多視点カメラ画像からの人物シルエットの安定な抽出を両立するために、スクリーンは液晶パネルと組み合わせられている。そして、対象人物への提示用画像とカメラによるシルエット検出のための均一な背景色画像を液晶パネルにより高速に切り替えながら表示する仕組みになっている。

3次元モデル復元を明示的に行わない手法 上記で紹介した研究例において利用されている自由視点画像生成手法は、多視点カメラ画像から3次元モデルを復元して自由視点画像を生成するものであった。これに対して、このような3次元形状復元を明示的に行わず、仮想視点画像を生成するために必要となる情報のみを多視点画像から求めることによる自由視点画像生成法として、Matusik らは、多視点画像に撮影された物体のシルエットから得られるシルエットコーンの積集合である Visual Hull を3次元空間で復元する代わりに、自由に設定した仮想視点における各画素に対する入力カメラ画像上のエピポーラ線上で、Visual Hull の表面上

の点を検出することによって、自由視点画像を生成する手法を提案した [Matusik00]。この手法では、Visual Hull 形状をエピポーラ線上の Visual Hull の表面上の点の座標により間接的に表現しており、この表現を Image Based Visual Hull (IBVH) と名づけた。IBVH は、Visual Hull を3次元空間で復元して自由視点画像を合成する場合に比べて計算量を削減することが可能であるため、4台の IEEE1394 カメラを利用したシステムによりビデオレートの自由視点画像生成に成功している。

上記は、室内に設置された多視点カメラからの自由視点画像生成である。これらに対して、大規模空間であるサッカースタジアムに設置した多視点カメラシステムからの自由視点画像生成として、Koyama らは、サッカー選手を抽出し、各選手を仮想視点からの視線に平行な法線を持つ一平面素と仮定して、この平面素を利用して実時間で自由視点映像を生成することが可能なシステムを提案した [Koyama03]。稲本らは、View Interpolation [Chen93] をサッカースタジアムで撮影した多視点画像に適用した自由視点画像生成法を提案した [稲本02]。木村らは、同様の手法をテニスの多視点画像に適用し、プレイヤー視点での画像生成を行った [木村健二03]。

## 5 多視点カメラを利用した監視・サーベイランス

90年代後半に米国 DARPA 主導のもと、米国内の研究組織によって動画像理解技術を使用した監視システムに関する研究プロジェクト VSAM (Video Surveillance and Monitoring) が行われた。特に CMU では、照明条件の変化にロバストな移動物体検出・追跡・識別手法、人物の行動を解析し異常行動を検出する手法、様々なタイプの映像センサで獲得した情報のキャリブレーション技術、少数のオペレータによって効率的に広範囲の監視が可能になるユーザインタフェースの開発など、多視点カメラを用いた屋外監視システムを実際に構築することを目的とした多数の基礎技術の開発が行われた [Collins00]。本節では、VSAM プロジェクトにより活性化された映像監視システムに関する研究領域のその後の展開について、映像センサ技術、アルゴリズム、アプリケーションに分類して紹介する。

## 5.1 映像センサ技術

多くの映像監視システムでは、映像センサの数や機能の制限により監視可能領域が限定される。しかし、監視すべきイベントは必ずしも設定した領域内で発生するわけではない。映像センサとして、パン・チルト制御が可能なアクティブカメラを導入することにより、様々な監視目的に適応可能で、かつ撮影対象空間の動的な変化に対応可能な映像監視システムの実現を目的とした研究が行われている。Kanade は、アメリカンフットボールの競技場を 30 台以上のアクティブカメラで取り囲み、それらを同時に遠隔制御しながら撮影を行うことで、競技場のあらゆる地点において任意のタイミングで発生するイベントを記録することが可能な Eye Vision システムを開発した [Kanade01]。浮田らは、能動視覚エージェントと名付けたアクティブカメラを制御する論理モデルを提案し、自律的に動作する多視点アクティブカメラを用いて複数の撮影対象を実時間で追跡・撮影するシステムを実現している [浮田 02]。

全方位カメラを用いて撮影を行うことによっても監視領域の拡大が可能である。さらに全方位カメラを分散的に配置することで、カメラ同士が観測しあう状態を作ることができ、カメラの相対的な位置決めが可能であるという特長が生まれる [十河 01]。これらの特長を有効活用することにより、長時間・広範囲に渡る人間の大局的な行動監視が可能になる [Morita03]。山本らは、3 眼ステレオユニットを正 20 面体の各面上に配置し、全方位の 3 次元情報が獲得可能な全方位ステレオシステム SOS を開発した。このシステムを用いて撮影を行うことにより、照明条件が変化するような悪条件下においても奥行き情報を用いることで安定した人物検出が可能になる [山本 01]。

## 5.2 検出・追跡アルゴリズム

複雑対象空間が複雑または広大なため、隣接する多視点映像上で重なり合って撮影される領域が存在しない場合、各カメラ間で観測の切れ目が発生する。このような場合には、ベイジアンネットを用いて多視点画像間の物体移動をモデル化することにより、切れ目における追跡処理を行う手法が提案されている [Kettner99]。

多視点カメラによって同一領域が撮影されている場合、その冗長性を利用することで物体同士の相互干渉

の影響を取り除き、安定した物体検出・追跡が可能となる。中島らは、多視点画像に Shape from Silhouette 法を適用した結果求められる物体形状が非ガウスの振る舞いをすることに着目し、CONDENSATION を用いた追跡手法を提案している [中島 02]。しかし、さらに物体の個数が増加する、もしくはより密集して存在する場合、全ての多視点画像上において何らかの相互干渉が発生し、検出・追跡処理が困難になるという問題が生じる。M2Tracker では、領域ベースステレオ法を用いて人物領域の位置検出を行うのと同時に、その見え方モデルを順次更新することにより、物体同士の隠れの問題に対応している [Mittal03]。撮影空間の背景領域の 3 次元形状が既知である場合には、多視点画像をその 3 次元モデル上に投影し、その上での整合性を検証することにより、安定した位置検出・追跡が可能である。Kang らは、サッカーの試合を撮影対象とし、グラウンド面と多視点画像面間のホモグラフィ射影関係を求め、グラウンド上に多視点画像を投影した結果を用いて物体の対応関係を求めることで、サッカーのように同じような見え方の物体が重なり合う悪条件下においても安定して動作する手法を提案している [KangJ04]。

多視点画像を用いて追跡処理を行う場合、全てのカメラがあらかじめキャリブレーション済であることが前提とされることが多いが、追跡対象空間の拡大に伴うカメラ台数の増加により、キャリブレーション処理の負荷もまた増加すると考えられる。Lee らは、平面上を移動する対象物体を基準カメラとその他のカメラで同時に観測することで、それらの間の相対的な位置・姿勢の推定を行っている [Lee00]。山添らは、多視点カメラからなる人物追跡システムにおいて、人物の 3 次元運動情報を用いた多視点カメラの位置・姿勢推定手法を提案している [山添 02]。これらの手法によれば、追跡処理と同時にカメラのキャリブレーションを行うことができるため、カメラの台数増加時のキャリブレーション作業の負荷が軽減されるだけでなく、運用中の障害によるカメラの位置・姿勢の変化に動的に対応可能であるというメリットがある。

## 5.3 新たなアプリケーション

これまでは、対象物体の物体検出や追跡などの処理の自動化を目的とした研究について紹介してきたが、計算機性能・ネットワーク速度の向上に伴い、検出・



追跡処理の実時間実行が可能となるにつれ、人間との協調的監視を目的とした研究が行われるようになってきた。Grimsonらは、あらゆる検出物体をプロトタイプとして生成した見え方モデル（共起関係行列）を用いて物体の検出を行い、それら組み合わせの統計的変動により、通常／異常の判定を行う枠組みについて述べている [Stauffer00]。IBMのSmart Surveillanceシステムは、大量にある映像情報から監視に必要となる情報を選択した後、注意すべき箇所のみを監視者に提示することで、監視作業の負荷を軽減する手段を提案している [Hampapur03]。

映像監視システムによって獲得した情報を人間対人間または人間対コンピュータのコミュニケーションに活用しようという、知的環境に対する取り組みが始まっている [Pentland00, Trivedi00]。依田らは、ユビキタスステレオビジョンシステムによって獲得した空間の3次元情報を用いてユーザごとに特化したインタラクションが可能な空間の実現を目指した研究を行っている [依田 03]。遠隔教育への応用も盛んである。京都大学の遠隔講義システムでは、多視点画像を用いて講師の位置や受講者の活性度など講義状況を検出し、その時々に応じた撮影方針により受講者に必要とされる情報を獲得するシステムを構築している [椋木 02]。

実際に公の空間で映像監視システムを利用するためには、被写体のプライバシーに対する配慮が必要となる。被写体の3次元モデルを実時間で生成し、そのモデルの投影領域として映像上でプライバシーを保護すべき領域を算出。画像処理を施すことにより被写体のプライバシーを保護するStealth Visionシステム [北原 04] など、プライバシー保護と映像情報獲得の両立を目的とした研究に注目が集まりはじめている。

## 6 まとめ

多視カメラを用いた記録と提示に関連した研究について、1) 多視点カメラのキャリブレーション、2) 多視点画像からの3次元復元、3) 多視点画像からの自由視点画像生成と提示、4) 多視点カメラを利用した監視・サーベイランス、という4つの面から、最近の動向を紹介した。本稿では、各研究事例については簡単な解説のみにとどめ、できるだけ多くの研究を紹介できるように心がけたつもりではある。しかしなが

ら、この分野を網羅的に紹介したというよりは、ある「視点」からの解説に留まっていると思う。本稿で不足している部分については、是非、著者らまでご指摘頂ければ幸いである。

## 参考文献

- [Agrawal01] M.Agrawal, L.S. Davis, "A Probabilistic Framework for Surface Reconstruction from Multiple Images," Proc. CVPR01, Vol.II, pp.470-476, 2001.
- [Avidan98] S. Avidan and A. Shashua, "Novel View Synthesis by Cascading Trilinear Tensors," IEEE Trans. on Visualization and Computer Graphics, Vol.4, No.4, pp.293-306, 1998.
- [Baker00] P. Baker, Y. Aloimonos, "Complete Calibration of a Multi-camera Network", Proc. of IEEE Workshop on Omnidirectional Vision (OMNIVIS'00) pp. 134-141, 2000.
- [Bonfort03] T.Bonfort, P.Sturm, "Voxel Carving for Specular Surfaces," Proc.ICCV03, pp.591-596, 2003.
- [Bonet99] Jeremy S. de Bonet, Paul Viola, "Roxels: Responsibility Weighted 3D Volume Reconstruction," Proc. ICCV99, Vol.1, pp.418-425, 1999.
- [Boyer03] E. Boyer, J.-S.Franco, "A hybrid approach for computing visual hulls of complex objects," Proc.CVPR03, Vol.I, pp.695-701, 2003.
- [Broadhurst01] A. Broadhurst, T. Drummond, R. Cipolla, "A probabilistic framework for the Space Carving," Proc. ICCV01, Vol.I, pp.388-393, 2001.
- [Carranza03] J. Carranza, C. Theobalt, M. Magnor, H.-P. Seidel, "Free-Viewpoint Video of Human Actors," ACM Trans. on Computer Graphics (SIGGRAPH'03), vol. 22, no. 3, pp. 569-577, July 2003 .
- [Chen93] S. Chen and L. Williams, "View Interpolation for Image Synthesis," Proc. of SIGGRAPH '93, pp. 279-288, 1993.
- [Cheung00] G. K. M. Cheung, T. Kanade, J. Y. Bouguet, M. Holler, " A real time system for robust 3D voxel reconstruction of human motions ", Proc. CVPR2000, Vol. 2,pp. 714-29, 2000.
- [Cheung03] G.K.M.Cheung, S.Baker, T.Kanade, "Shape-From-Silhouette of Articulated Objects and its Use for Human Kinematics Estimation and Motion Capture," Proc. CVPR03, 2003.
- [Chu03] Chi-Wei Chu, Odest Chadwicke Jenkins, Maja J Mataric, "Markerless Kinematic Model and Motion Capture from Volume Sequences," Proc. CVPR 2003, 2003.
- [Collins00] R.T.Collins, A.J.Lipton, T.Kanade, H.Fujiyoshi,D.Duggins, Y.Tsin,D.Tolliver, N.Enomoto, and O. Hasegawa, Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000.

- [Franco03] J.-S. Franco and E. Boyer, "Exact polyhedral visual hulls," Proc. the British Machine Vision Conference 2003, 2003.
- [Gavrila96] D.M.Gavrila, and L.S.Davis, "3-D model-based tracking of humans in action: a multi-view approach," Proc. CVPR96 pp.73–80,1996
- [Goldlucke03] B. Goldlucke, M. Magnor, "Real-Time Microfacet Billboarding for Free-Viewpoint Video Rendering," Proc. IEEE International Conference on Image Processing (ICIP'03), Barcelona, Spain, vol. 3, pp. 713-716, September 2003
- [Grau04] O.Grau, T.Pullen, G.A.Thomas, "A Combined Studio Production System for 3D Capturing of Live Action and Immersive Actor Feedback," IEEE Trans. Circuits and Systems for Video Technology, March 2004.
- [Gross03] M. Gross, S.Wurmlin, M.Naef, E. Lamboray, C.Spagno, A.Kunz, E. Koller-Meier, T. Svoboda Luc Van Gool, S. Lang, K. Strehlke, A.Vande Moere, O. Staadt, "blue-c: A Spatially Immersive Display and 3D Video Portal for Telepresence," Proceedings of ACM SIGGRAPH 2003, pp. 819-827, 2003.
- [Hampapur03] A.Hampapur, L.Brown, J.Connell, S.Pankanti, A.Senior, Y.Tian, "Smart Surveillance: Applications, Technologies and Implications," Proc. of IEEE and ACM Pacific-Rim Conf. on Multimedia, CD-ROM 2C1-2, 2003.
- [Hartley95] R. I. Hartley, "In Defense of the 8-Point Algorithm", Proc. ICCV95, pp. 1065-1070, 1995.
- [Hartley97] R. I. Hartley, "Kruppa's Equations Derived from the Fundamental Matrix", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.19, no.2, pp. 133-135, 1997.
- [Hartley00] R. Hartley and A. Zisserman, "Multiple View Geometry," ISBN 0521540518, Cambridge University Press, 2000.
- [Ihrke04] I. Ihrke, L. Ahrenberg, and M. Magnor, "External camera calibration for synchronized multi-video systems", Journal of WSCG, Vol.12, No.1-3, pp. 537-544, Jan 2004.
- [Iso03] D.Iso, H.Saito, "Modeling and rendering in 3D coordinate defined by two cameras for shared virtual space communication," Proc. Computer Vision/Computer Graphics Collaboration for Model-based Imaging, Rendering, image Analysis and Graphical special Effects (MIRAGE2003), pp.28-37, 2003.
- [Kanade95] T. Kanade, P. W. Rander, and P. J. Narayanan, "Concepts and early results," IEEE Workshop on Representation of Visual Scenes, pp.69076,
- [Kanade01] T. Kanade, "Carnegie Mellon Goes to the Super Bowl," <http://www.ri.cmu.edu/events/sb35/tksuperbowl.html>, 2001.
- [KangSB01] S.B. Kang, R. Szeliski, J. Chai, "Handling Occlusions in Dense Multi-view Stereo," Proc. CVPR01, Vol.I,pp.103–110, 2001.
- [KangJ04] J. Kang, I. Cohen, G.Medioni, "Tracking People in Crowded Scenes Across Multiple Cameras," Proc.ACCV04, pp.390-395, 2004.
- [Kettner99] V.Kettner, R.Zabih, "Bayesian multi-camera surveillance," Proc. CVPR99, pp. 2253-2259,1999.
- [Kim04] H.W. Kim, I.S.Kweon, "Optimal Photo Hull Recovery for the Image-Based Modeling," Proc.ACCV04, 2004.
- [Kitahara03] I. Kitahara and Y. Ohta, "Scalable 3D Representation for 3D Video Display in a Large-scale Space," Proc. of the IEEE Virtual Reality 2003,pp.45-52,2003.
- [Koyama03] T. Koyama, I. Kitahara, and Y. Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium," The Second International Symposium on Mixed and Augmented Reality (ISMAR03), pp.178-187, 2003.
- [Kutulakos00] K.N. Kutulakos, S.M. Seitz, "A Theory of Shape by Space Carving," IJCV(38), No. 3, pp. 199-218, 2000.
- [Laurentini94] A. Laurentini, "The Visual Hull Concept for Silhouette Based Image Understanding," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, no.2, pp.150-162, 1994.
- [Lazebnik01] S.Lazebnik, E.Boyer, J.Ponce, "On Computing Exact Visual Hulls of Solids Bounded by Smooth Surfaces," Proc.CVPR01, Vol.I, pp.156–161, 2001.
- [Lee00] L. Lee, R.Romano, G.Stein, "Monitoring Activities from Multiple Video Streams," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.22, No.8, pp.758-767, 2000.
- [Lenz88] R. Lenz and R. Y. Tsai, "Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3-D Machine Vision Metrology," IEEE Trans. Pattern Anal. Machine Intell., Vol. 10, No. 5, pp. 713-720, May 1988.
- [Li02] Y. Li, S. Lin, H. Lu, S.B. Kang, H.Y. Shum, "Multibase-line stereo in the presence of specular reflections," Proc. ICPR02, Vol.III,pp.573–576, 2002.
- [Lorensen87] W. E. Lorensen and H. E. Cline, "Marching Cubes: A High Resolution 3D Surface Construction Algorithm," Computer Graphics (Proceedings of SIGGRAPH '87), Vol. 21, No. 4, pp. 163-169, 1987.
- [Matusik00] W. Matusik, C. Buehler, R. Raskar, S. Gortler and L. McMillan, "Image-Based Visual Hulls," Proc. of SIGGRAPH 2000, pp.369-374, 2000.
- [Matsunaga01] C. Matsunaga and K. Kanatani, "Calibration of a moving camera using a planar pattern: Optimal computation, reliability evaluation and stabilization by the geometric AIC," Electronics and Communications in Japan, Part 3: Fundamental Electronic Science, Vol. 84, No. 7 pp. 12-21, 2001.
- [Mittal03] A.Mittal, L.S.Davis, "M2Tracker:A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene," Int. Journal of Computer Vision, Vol. 51, No.3, pp.189-203, 2003.

- [Moezzi97] S. Moezzi, L.C.Tai, P.Gerard, "Virtual View Generation for 3D Digital Video," IEEE Multimedia, Vol 4, Issue 1 pp.18–26, 1997.
- [Morita03] S. Morita, K. Yamazawa, and N. Yokoya, "Networked video surveillance using multiple omnidirectional cameras," Proc. 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp. 1245-1250, 2003.
- [Narayanan98] P. J. Narayanan, P.W.Rander, and T.Kanade, "Constructing Virtual Worlds using Dense Stereo", Proc. ICCV 98, 1998.
- [Okutomi93] M. Okutomi and T.Kanade, "A Multiple-Baseline Stereo," IEEE Trans. on PAMI, Vol.15, No.4, pp.353-363, 1993.
- [Ong99] E. J. Ong and S. Gong, "Tracking hybrid 2D-3D human models from multiple views," in International Workshop on Modeling People at ICCV99, Corfu, Greece, September 1999.
- [Ott93] M. Ott, J. Lewis, and I. Cox, "Teleconferencing Eye Contact Using a Virtual Camera," In INTERCHI93, pp.119–110, 1993.
- [Pentland00] A.Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.10, No.22, pp.107–118,2000.
- [Pollefeys99] M. Pollefeys, R. Koch, and L. V. Gool, "Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters," International Journal of Computer Vision, 32(1), pp. 7-25, 1999.
- [Porikli03] F. M. Porikli, A. Divakaran, "Multi-Camera Calibration, Object Tracking and Query Generation," IEEE International Conference on Multimedia and Expo (ICME), Vol. 1, pp. 653-656, July 2003.
- [Pritchett98] P. Pritchett and A. Zisserman, "Wide baseline stereo matching," Proc. ICCV '98, pp. 754-760, January, 1998.
- [Saito99] H. Saito, T. Kanade, "Shape Reconstruction in Projective Grid Space from Large Number of Images," Proc. CVPR'99, Vol.2, pp.49–54, 1999.
- [Saito03] H.Saito, S.Baba, T.Kanade, "Appearance-Based Virtual View Generation From Multicamera Videos Captured in the 3-D Room," IEEE Trans. on Multimedia, vol.5, no.3, pp.303-316,2003.
- [Scharstein99] D.Scharstein, View Synthesis Using Stereo Vision, Lecture Notes in Computer Science 1583, Springer, 1999
- [Sebe02] I. O. Sebe and G. Q. Chen, "Multi Camera Calibration", STMicroelectronics Technical Report, San Diego, July 2002.
- [Seitz96] S. M. Seitz and C. R. Dyer, "View morphing", Proceedings of SIGGRAPH '96, pp. 21-30, 1996.
- [Seitz97] S. M. Seitz and C. R. Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring," Proc.CVPR97, pp. 1067-1073, 1997.
- [Slabaugh02] G.G.Slabaugh, R.W.Schafer, M.C. Hans, "Multi-resolution space carving using level set methods," Proc. ICIP02, Vol.II, pp.545–548, 2002.
- [Slabaugh04] G.G.Slabaugh, W. B. Culbertson, T. Malzbender, M. R. Stevens, R. W. Schafer, "Methods for Volumetric Reconstruction of Visual Scenes," The International Journal of Computer Vision, 57(3), pp. 179-199, 2004.
- [Stauffer00] C.Stauffer, E.W.Gimson, "Learning Patterns of Activity Using Real-Time Tracking," IEEE Trans on Pattern Analysis and Machine Intelligence, Vol.22, No.8, pp.747-757, 2000.
- [Strecha03] C.Strecha, T.Tuytelaars, L.V.Gool, "Dense Matching of Multiple Wide-baseline Views," ICCV03, pp.1194–1201,2003.
- [Svoboda02] T. Svoboda, H. Hug, and L. V. Gool, "ViRoom - low cost synchronized multicamera system and its self-calibration" In Pattern Recognition, 24th DAGM Symposium, number 2449 in LNCS, pp. 515-522, Springer, September 2002.
- [Tenenbaum00] J. B. Tenenbaum, V. de Silva and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science 290 (5500): 2319-2323, 22 December 2000.
- [Trivedi00] M.M. Trivedi, I. Mikic, S. K. Bhonsle, "Active Camera Networks and Semantic Event Database for Intelligent Environments," IEEE Workshop on Human Modeling, Analysis and Synthesis (in conjunction with CVPR), June, 2000.
- [Tsai86] R. Y. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision", Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 364-374, Miami Beach, FL, 1986,
- [Vedula98] 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. 326-322, 1998.
- [Vedula02] S.Vedula, S.Baker, T.Kanade, "Spatio-temporal view interpolation," Proceedings of the 13th Eurographics workshop on Rendering, pp.65 – 76, 2002.
- [Wei94] G. Q. Wei, S. D. Ma "Implicit and Explicit Camera Calibration: Theory and Experiments" IEEE PAMI, Vol. 16, No. 5, pp. 469-480, May 1994.
- [Wilburn02] B. Wilburn, M. Smulski, H. K. Lee, and M. Horowitz, "The Light Field Video Camera", Proc. Media Processors 2002, SPIE Electronic Imaging 2002.
- [Yaguchi04] S.Yaguchi , H.Saito, "Arbitrary Viewpoint Video Synthesis from Multiple Uncalibrated Cameras," IEEE Trans. on Systems, Man and Cybernetics, PartB, vol. 34, no.1, pp.430-439, 2004.

- [Yamazaki02] S. Yamazaki, R. Sagawa, H. Kawasaki, K. Ikeuchi, M. Sakauchi, "Microfacet Billboarding," 13th Eurographics Workshop on Rendering, pp.169-179,327, 2002.
- [Yang02] R. Yang and Z. Zhang, "Eye Gaze Correction with Stereovision for Video Tele-Conferencing". In Proc. 7th European Conference on Computer Vision (ECCV2002), Volume II, pages 479-494, Copenhagen, Denmark, May 28-31, 2002
- [Zhang96] Z. Zhang, "Determining the epipolar geometry and its uncertainty: a review", Research Report 2927, INRIA Sophia-Antipolis, France, July 1996.
- [Zhang98] Z. Zhang, "A Flexible New Technique for Camera Calibration", Technical Report MSR-TR-98-71, Microsoft Research, 1998.
- [稲本 02] 稲本奈穂, 斎藤英雄, "サッカーシーンにおける自由視点映像生成のための視点内挿法," 日本バーチャルリアリティー学会論文誌 TVRSJ, Vol.7, No.4, pp.513-520, 2002.
- [ウ 01] ウ小軍, 和田俊和, 東海彰吾, 松山隆司, "平面間透視投影を用いた並列視体積交差法," 情報処理学会 CVIM 研究会論文誌, Vol.42, No.SIG 6(CVIM 2), pp.33-43, 2001.6
- [植芝 02] 植芝俊夫, 富田文明, "平面パターンを用いた複数カメラシステムのキャリブレーション," 情処研報, No.2002-CVIM-135, pp. 47-54, 2002.
- [浮田 02] 浮田宗伯, 松山隆司, "能動視覚エージェント群による複数対象の実時間協調追跡," 情報処理学会 CVIM 研究会論文誌, Vol.43, No. SIG11, pp.64-79, 2002.
- [北原 02] 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", 映像メディア学会誌, Vol.56, No.8, pp.120-125, 2002.
- [北原 04] 北原格, 小暮潔, 萩田紀博, "Stealth Vision:被写体のプライバシーを保護する映像獲得方式," 信学技報, PRMU2003-299, pp.89-94, 2004.
- [木村誠 02] 木村誠, 斎藤英雄 "3 カメラ間の射影幾何を用いた View Morphing" 情報処理学会 CVIM 研究会, No.135-16,2002.
- [木村健二 03] 木村健二, 斎藤英雄, "テニスの多視点画像からのプレイヤー視点画像の生成法," 電子情報通信学会技術報告, PRMU2003-173, 2003.
- [杉村 03] 杉村良彦, 佐藤淳, "未校正マルチカメラの相互投影を用いた弱校正と物体形状の復元," 情報処理学会研究報告コンピュータビジョンとイメージメディア, No.140-003, 2003.
- [十河 01] 十河卓司, 石黒浩, "分散全方位視覚システムの研究," 情報処理学会 CVIM 研究会論文誌, Vol.42, No.SIG13, pp.33-40, 2001.
- [中島 02] 中島平, 浜崎浩二, 岡谷貴之, 出口光一郎, "CONDENSATION を用いた多視点画像の融合による複数人物の追跡," 画像の認識・理解シンポジウム (MIRU2002) 講演論文集 II, pp.317-319, 2002.
- [延原 02] 延原章平, 和田俊和, 松山隆司, "弾性メッシュモデルを用いた多視点画像からの高精度 3 次元形状復元," 情報処理学会 CVIM 研究会論文誌, Vol.43, SIG11(CVIM5), pp.53-63, 2002.
- [濱崎 01] 濱崎 省吾, 吉田 裕之, 重永 信一, "多視点シルエット画像からの高速な 3 次元形状復元手法", 第 7 回画像センシングシンポジウム講演論文集, pp.59-64, June. 2001.
- [松浦 97] 松浦友彦, 佐藤清秀, 中村裕一, 大田友一, "隠れ検出の可能な多眼ステレオ法," 電子情報通信学会論文誌 D-II, Vol.J80-DII, pp.1432-1440, 1997
- [椋木 02] 椋木雅之, 西口敏司, 美濃導彦, "複数台カメラによる講義の観測と撮影," 情処研報, CVIM131-11, pp.77-84, 2002.
- [山口 03] 山口義隆, 斎藤英雄, "未校正多視点画像からの多面体表面モデルの直接復元法," 情報処理学会 CVIM 研究会, 138-6, 2003.
- [山添 02] 山添大丈, 内海章, 鉄谷信二, 谷内田正彦, "分散型人物追跡システムにおける多数カメラの位置・姿勢推定," 画像の認識・理解シンポジウム (MIRU2002) 講演論文集 II, pp.311-316, 2002.
- [山本 01] 山本和彦, 棚橋英樹, 桑島茂純, 丹羽義典, "実環境センシングのための全方向ステレオシステム (SOS)," 電気学会論文誌 C (電子・情報・システム部門誌) Trans. IEE of Japan, Vol. 121-C, No. 5, pp. 876-881, 2001.
- [依田 03] 依田育士, 坂上勝彦, "多地点ステレオ距離情報による人の認識と表現手法の提案," 信学技報, PRMU2003-150, pp.59-64, 2003.
- [米元 00] 米元 聡, 有田 大作, 谷口 倫一郎, "多視点動画像処理による実時間全身モーションキャプチャシステム - 視覚に基づく仮想世界とのインタラクション -," 映像情報メディア学会誌, vol.53, no.3, pp.409-416, 2000.