

高精細3次元形状復元のための多視点カメラワーク最適化

西山 学* 牧 淳人 松山 隆司

京都大学大学院 情報学研究科

概要 本論文では、視体積交差法において高解像度かつ精度良く3次元形状復元を行うためのカメラワークについて考察する。これまでも視体積交差法におけるカメラ位置についての研究は行われているが[2]、本研究での対象は人体であり、形状が複雑で自己遮蔽が多く発生することが問題を複雑にしており、単純にカメラ間の位置関係を考慮するだけでは不十分である。そこで対象の形状モデルを導入し、誤差分布表現によるカメラワークの評価および分枝限定法と誤差分布表現を利用した効率的な最適化手法を提案する。

1 はじめに

3次元ビデオ映像という新たなメディアについての研究が進められている。3次元の映像といえば、これまでもビデオゲーム等で人物や物体の表現手法として用いられてきたが、基本的に手作業で作成されるものであった。それに対し3次元ビデオの実用化に当たっては、撮影装置さえあればビデオカメラで映像を撮るように3次元データを容易に獲得できるようになることが目指される。

3次元ビデオ映像は現在のところ多数のカメラで対象を撮影し、形状に関する情報および表面のテクスチャ情報を取得して組み合わせることで作られる。このうち、テクスチャ情報は撮影像から手に入る。一方、複数の2次元画像から3次元形状情報を獲得する手法については、数多くの研究が行われている。

そういった中でも、3次元ビデオ用の形状情報を得る上で有望な手法として、視体積交差法と呼ばれる手法がある[4][5]。これは図1に示すように、(1)複数台のカメラで対象物体の全周囲画像を撮影、(2)撮影画像から対象のシルエット部分を切り出し、(3)各カメラのシルエット像から visual-cone を構成(図1(b))、(4)全 visual-cone の共通領域を対象の形状として出力(図1(c))することで撮影対象の3次元形状情報を得ようとする手法である。

*現在 (株) 東芝 研究開発センターに勤務。

Optimizing Camera Control for High-Resolution Volume Intersection

Manabu Nishiyama, Atsuto Maki and Takashi Matsuyama
Graduate School of Informatics, Kyoto University

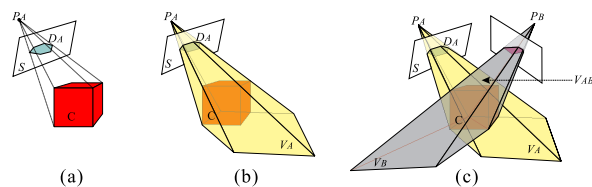


図 1: 視体積交差法による3次元形状復元

この手法による利点には、奥行き画像だけでなく対象の全周囲形状が得られる、カメラ間の距離が離れている環境下でも使える、カメラ台数を増やすほど正確さが増すなどがある。しかし一方で、原則として全カメラで対象全体をとらえる必要があるため、ズームをコントロールできるカメラでも制限が生じ、結果として得られる3次元形状の空間解像度は低く留まる傾向にある、という欠点が存在する。

このような問題点を克服するため、本研究では視体積交差法の枠組を前提に、実画像からの3次元形状復元の高空間解像度化手法を提案する。高い空間解像度を得られるようになれば人体の指先などの細かな形状までも獲得できるだけでなく、同時にテクスチャ情報が高精細化されることをも意味する。テクスチャの細かさが3次元ビデオの見栄えに果たす役割は大きく、それゆえ高解像度化により3次元ビデオ映像の品質を大きく改善できる可能性がある。

1.1 想定する環境

まず本研究で考える形状復元の実行環境について述べる。形状復元に用いられるカメラ群は図2のように実験室内に分散配置されている。各カメラは(a)視点

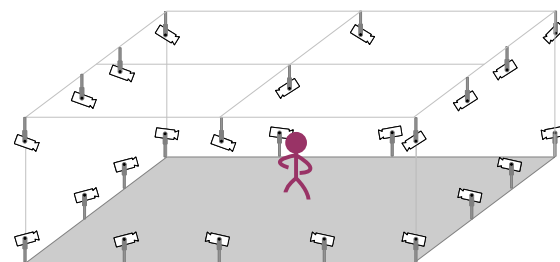


図 2: 想定する環境

を中心とした回転機構を持っており、パンおよびチルト角の指定で視線方向の制御が可能、(b)ズームを制御可能、といった特徴を持っているとする。

一方、撮影対象は人体の全身とする。以降ではある時点(フレーム)で撮影された画像群を入力として形状復元の高精細化の問題を考えていく。

1.2 提案手法の概要

冒頭で述べた視体積交差法を適用するには、用いるカメラ群全てが対象全体を視野内に収めておく、という前提がある。すなわち図3の例のように全身を画像内にとらえておく必要がある。これに対して、本研究ではまず、対象の一部分にズームアップして撮影する(図4)ことで得られる対象の部分撮影画像から対象の3次元形状を推定し、高解像度の3次元形状の獲得を目指す。つまり次のような考え方で行う。

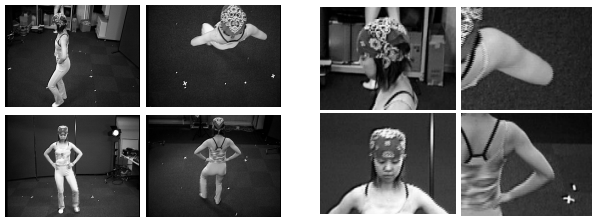


図3: 従来の形状復元で用いている画像の例

1. 対象を複数パーツから構成される物体であると見なす。人体においては、関節によって分けられる頭・胴体・上腕・下腕などを1つのパーツと考える。
2. カメラはそれぞれ1つ(または複数)のパーツをクローズアップして撮影する。このとき、カメラ群全体として全てのパーツをカバーできるようにカメラを割り振る必要がある。
3. 空間解像度の高い対象の部分撮影画像群を用いて3次元形状を推定する。

つまり、各カメラは従来の視体積交差法に存在したズームの制限を破って対象の一部分だけをクローズアップする。これは上述した前提に反するが、その代わりに撮影する部位をカメラごとにうまく割り振ることで全身の高解像度画像を獲得し、それを用いて全身形状の獲得を行う。したがって、“各カメラがどこにズームアップするか”を決定することが提案の枠組みにおける主要な問題となる。従来の視体積交差法では原理上、部分撮影画像からの全身形状復元は不可能であるが、本研究ではアルゴリズムをアレンジし、視野外にも対象存在の可能性があるとしてそれを可能にしている[6]。

1.3 人体形状モデルの導入

上記したような手法を実現するにはまず、人体をどのようにパーツの集合として認識するかという問題を解決する必要がある。また、“どこにズームアップするか”を決定するため、各カメラがズームアップしたとき形状復元に与える影響を評価しなければならない。

本研究ではこれらの目的を達成するため、図5のような対象(人体)の多関節形状モデルを導入する。形状モデル中でパーツは一つの円柱で表現する。このモデルで実際の人体の姿勢を再現すれば[3]、任意の位置からの対象の見え方を再現できる。したがって、(1)従来手法で荒い形状を復元してモデルを当てはめ、パーツ位置を検出、(2)モデルを使って見え方を再現、カメラコントロールを決定しクローズアップ撮影、の2段階の処理を行う。この中でクローズアップ撮影時にカメラの視線方向およびズームを決めることを特にカメラワークの決定問題と呼ぶことにする。本論文ではこのカメラワークの決定プロセスに焦点を当て、形状モデルを使ったカメラワークの評価手法および、それを利用したカメラワークの決定アルゴリズムを提案する。

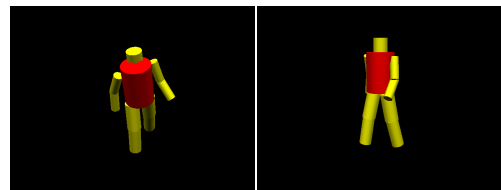


図5: 人体形状モデル

2 人体形状モデルに基づくカメラワークの評価手法

3次元形状の品質の評価尺度には形状の細かさ(空間解像度)と形状の正確さ(形状復元誤差)の2つがあると考えられるが、提案手法により期待される形状の品質を従来手法と比較すると(a)形状の細かさについては当然向上する、(b)正確さに関してはカメラの台数・位置関係により誤差は概して大きくなる、ということがいえる。提案手法では形状復元の入力画像に対象の一部分へのクローズアップ画像を用いるため、各カメラが対象のどの部分にクローズアップするのか、カメラワークの選択の余地がある。そのような選択肢のうちどれを選べばよいのか、評価、比較の上で最適に決定する必要がある。本節では3次元形状復元における撮影時のカメラワークの評価手法について考察する。

2.1 カメラワークの評価基準

カメラワークを評価するためにその基準を決める。本研究では形状復元の高解像度化が最大の目的であるため、達成する空間解像度をあらかじめ決めておき、

これを実現するようにズームをかける。この制約の範囲内でカメラワークを決めるとすると、前述の通り形状の正確さがカメラワークにより異なってくる。よって多数あるカメラワーク候補を比較する場合、復元される形状がより正確な方が“良い”カメラワークといえる。したがって形状復元誤差をカメラワークの評価基準として用いる。

2.2 形状復元誤差の表現手法

前述した形状モデルの人体への当てはめ結果から、復元したい人体のおおよその形状を推定できる。また、モデルを使って任意視点からの対象の見え方を再現できることから、

1. 全てのカメラからの対象の見え方 (部分撮影像) を形状モデル (この体積を $\mathcal{V}$ とする) で再現し、
2. 再現画像を用いて形状復元を行い (つまりモデルの形状を復元している)、これを $\hat{\mathcal{V}}$ とする。
3. 形状モデル $\mathcal{V}$ と復元形状 $\hat{\mathcal{V}}$ とを比較することで形状復元誤差を測定する。

この $\mathcal{V}$ と $\hat{\mathcal{V}}$ との誤差評価方法であるが、これはカメラワークを決定するために用いられるものである。したがって、カメラ群中で形状復元に貢献している/貢献していないといった情報が評価結果から分析できるよう手法があれば、より正確な形状を復元できるカメラワークを見つける手がかりになる。

そこで本研究では、形状復元誤差の表現手法にモデル $\mathcal{V}$ の表面から復元形状 $\hat{\mathcal{V}}$ の表面までの距離を用いることにする (図 6)。この距離を形状復元エラー値と称する。そしてモデル全体でのエラー値の分布を形状復元エラー分布と呼び、これを使ってカメラワークの形状復元の正確さを評価することを提案する。復元形状が元の形状に近づくほど、このエラー分布は全体的に小さくなることは明らかなので、正確さを評価する尺度として妥当である。しかもこのエラー分布を採用することにより、形状復元にとって最適なカメラワークを見つけるのに有効な手がかりを得ることが可能になる。

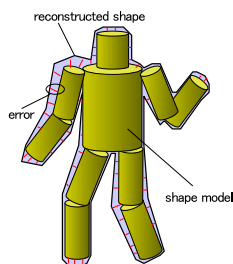


図 6: エラー分布による形状復元誤差の表現

2.3 形状復元エラー分布の計算方法

提案する形状復元エラー分布がどのように得られるのかについて述べる。まずカメラが 1 台の場合、視体積交差法で得られるのは撮影像からの visual-cone そのものである。よってこのときの形状復元エラー値はモデル表面から visual-cone 表面までの距離となる (図 7)。この値はモデルを用いた再現画像上で容易に求められる。

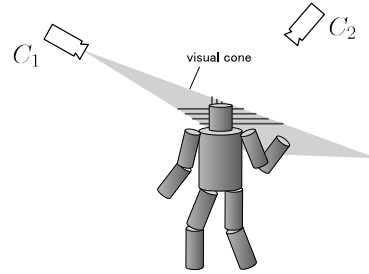


図 7: カメラ単体での形状復元エラー分布

次にカメラを複数台にしたときのエラー分布を求める。視体積交差法では各カメラからの visual-cone の共通領域を対象の形状とする。よってこのときの形状復元エラー値は各カメラ単体でのエラーのうち最小のものになる (図 8)。したがって形状復元エラー分布は複雑な計算無しに求めることが可能である。

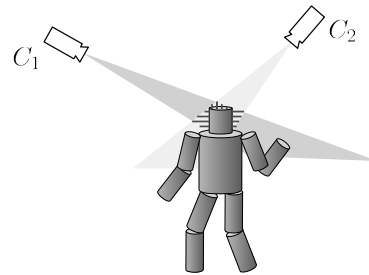


図 8: カメラ複数台での形状復元エラー分布

3 カメラワークの決定アルゴリズム

前節であるカメラワークに対する形状復元誤差を分布で表現することを提案した。ここでいう“カメラワーク”とは具体的には、視線方向を決めるパン角 p 、チルト角 t 、およびズーム値 z というコントロールパラメータのことである。これまでの議論により、 n 台のカメラ $C_1, C_2, \dots, C_n$ それぞれについてカメラワーク、すなわちパラメータ (p_i, t_i, z_i) を決定すると形状復元エラー分布が得られるようになっている。ここではエラー分布を利用してカメラワークを最適化する具体的手法について述べる。

3.1 カメラワークの変化によるエラー分布の変化

まずカメラワーク最適化の戦略を決めるために、コントロールパラメータを微小変動させたときにエラー分布がどのように変化するかについて考察する。対象をパーツ1つにし、パン、チルト角の変化で対象の見え方が図9のように変化したとき円筒の円周方向のエラー分布は図10のように変化する。この図で横軸は軸周りの角度、縦軸はエラー値を表す。

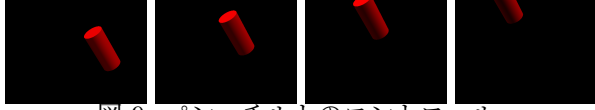


図9: パン・チルトのコントロール

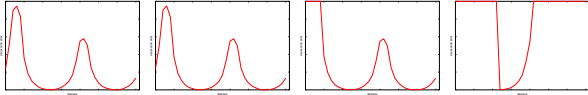


図10: パン・チルトの変化に対するエラー分布の変化

この図から対象がカメラの視野内に収まっている間はエラー分布は不変であり、一部が画像からはみ出して初めてエラー値が増加することが分かる。その理由は、カメラの回転機構が視点固定だとしていることに起因する。この場合、視線方向を変化させても視野領域が移動するだけで、対象の見え方自体は変わらない。ズームの変化についても視野の広さが変化するだけなので、同様のことがいえる。

したがってカメラワークを決定するにあたり、各カメラがどのパーツをズームアップして撮影するかを選択して大まかにカメラワークを変化させて探索していった方が効率的である。一方、カメラのズームは視野の広さと復元形状の空間解像度の両方に影響し、しかも2つはトレードオフの関係にある。本研究では高空間解像度3次元形状を得ることが主な目的であるから、復元される形状に要求する空間解像度をパーツごとにあらかじめ設定し、パーツ撮影時のズームに対し制約を設ける。

したがってカメラワークの決定は、(1) あらかじめ各パーツについて要求する空間解像度を決めておき、(2) 各カメラについて、どのパーツを視野内に収めるかで視線方向を決め、ズームは要求空間解像度を満たすように決めることにする。これはそれぞれのカメラについて、いくつかあるカメラワークの候補のうち1つを選択する問題であると見なせる。どの候補が良いかは前章で述べた形状復元誤差分布に基づき判断する。

3.2 分枝限定法による最適化

3.2.1 候補カメラワーク

まず各カメラについてカメラワークの選択肢を与える必要がある。前節の考察より、あるパーツ P_j のエ

ラー分布は、パーツ P_j が画像領域内に収まっている場合とそうでない場合のほぼ二者択一といえる。そこで各カメラでのカメラワークの候補としては、(1) パーツ P_1 を視野に入れる (2) パーツ P_2 を視野に入れる ... (m) パーツ P_m を視野に入れる、の m 個を与えることにする。各候補について具体的なカメラのコントロールパラメータは、(1) パーツの重心が画像中心に来るようにパン、チルトを決定し、(2) 空間解像度の要求値を満たし、なおかつ視野が最も広くなるようズームを設定する。

カメラ C_i の候補カメラワークの集合を W_i とする。 W_i の要素であるカメラワーク $w_{ij} = (p_{ij}, t_{ij}, z_{ij})$ それぞれに対し、カメラ1台による形状復元エラー分布が求まる。以降の議論ではこれをベクトルの要素として並べたものを $e_{ij} = (e_{ij1}, e_{ij2}, \dots, e_{ijl})$ と表記する。 l はエラー分布中のエラー値の個数である。

3.2.2 問題の定式化

それぞれのカメラ C_i には $|W_i|$ 個のカメラワークの候補が存在する。ここで、カメラワーク w_{ij} を選択する場合は $x_{ij} = 1$ 、選択しない場合 $x_{ij} = 0$ とする変数 x を導入する。カメラワークは同時に1つしか選べないので、

$$\sum_{j=1}^{|W_i|} x_{ij} = 1 \quad (1)$$

という制約条件が存在する。また、このときの C_i 1台に対する形状復元エラー分布は、変数 x_{ij} を利用して

$$\sum_{j=1}^{|W_i|} x_{ij} e_{ij} \quad (2)$$

とすることで求められる。式(2)のようなエラー分布は各カメラで独立に得られる。 $C_1, C_2, \dots, C_n$ の n 台のカメラからのエラー分布を統合すると、その結果 e は

$$e = (e_1, e_2, \dots, e_l) = \bigotimes_{i=1}^n \sum_{j=1}^{|W_i|} x_{ij} e_{ij} \quad (3)$$

となる。記号 $\otimes$ は $e_a = (e_{a1}, e_{a2}, \dots, e_{al})$ 、 $e_b = (e_{b1}, e_{b2}, \dots, e_{bl})$ のとき

$$e_a \otimes e_b = (\min\{e_{a1}, e_{b1}\}, \min\{e_{a2}, e_{b2}\}, \dots, \min\{e_{al}, e_{bl}\})$$

となるような演算子とする。式(3)はカメラ単体でのエラー分布からカメラ群全体での形状復元エラー分布を求める計算式である。

誤差の小さな形状を得るには e の要素が全体的に小さくなることが望ましい。そこで目標関数をベクトル e の要素中の最大値

$$\max_{1 \leq k \leq l} e_k \quad (4)$$

と設定する。この値を最小にする x_{ij} の組み合わせを求めることでカメラワークを決定する。

3.2.3 分枝限定法の適用

ここまでの議論により、高空間解像度を実現するためのカメラワークの決定を組み合わせ最適化問題として定式化できた。しかし実際の問題の規模を考えると、カメラが n 台、パーツが m 個の場合、組み合わせの総数は m^n 通りと非常に大きく、全てを列挙して解くのは事実上不可能である。そこで、この問題に対し分枝限定法を適用して効率的に探索を行っている。

4 シミュレーションによる検証

本論文で提案した形状復元の高解像度化手法をシミュレーションにより検証する。シミュレーションは既にモデル当てはめが終了していると仮定して対象の形状モデルのポーズ情報を入力として与える。

4.1 シミュレーション環境

4.1.1 対象の形状モデル

対象の多関節モデルは図5のものをを用いることにする。ポーズも同じである。このモデルでは全てのパーツが円筒で構成されている。モデルは10個のパーツで構成されており、それぞれに空間解像度を設定できる。今回は全てのパーツが一辺1mmのボクセルの集合として形状復元できるように対象にズームアップする。このことを以下では空間解像度が1mm/voxelであると表記する。

4.1.2 撮影環境

次に、撮影環境は図2のような空間中にカメラが分散して配置されているものとする。空間の大きさは縦7.5m、横9m、高さ2.5m、カメラは合計で24台ある。このうち11台が天井に、残りの13台が床面から約70cmの高さに設置されている。対象はこのスペースの中心付近に立ってポーズをとっている。

設置されたアクティブカメラはあらゆる方向を向くことができる。また、撮影時の画像サイズは640×480ピクセルとする。このサイズと要求空間解像度でズームアップ撮影時のズーム値が決まる。

4.2 提案手法により決定されたカメラワーク

上記の条件で実際にカメラワークの決定を行った。その結果得られるカメラワークで対象を撮影すると一部のカメラでは図11のようになった。各カメラがズームアップするポイントが対象の全身に分散されていることが分かる。なお、このときの実行結果は以下の通りであった。

- 空間解像度: 1mm/voxel
- 評価値 (エラー分布中の最大値): 56mm

- 実行時間: 2.18 秒
- 探索したカメラワークの組み合わせ: 9460 通り (総組み合わせ数 10^{24} 通り中)

実行時間、および全探索空間に対する探索した組み合わせの数を見ると非常に高速にカメラワークの決定が行われている。

次に、得られたカメラワークで対象を撮影し、形状復元するときのエラー分布を度数分布表示したのが図12である。

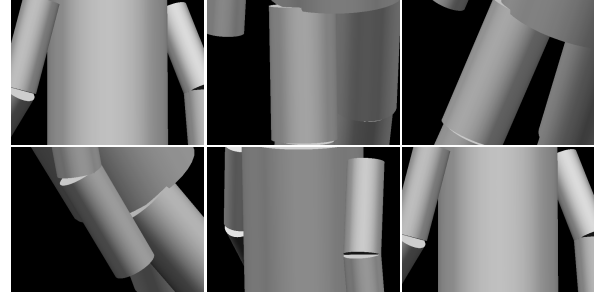


図 11: 提案手法により決定されたカメラワークでの撮影画像

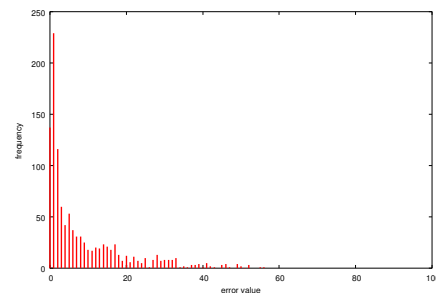


図 12: 全身エラー分布の度数分布 (提案手法)

4.3 最適化の初期段階でのカメラワーク

提案手法で得られるのがどのくらい良いカメラワークであるかを検証するため、決定手法の初期段階で得られるカメラワークとの比較を行う。その結果は

- 空間解像度: 1mm/voxel
- エラー分布中の最大値: 366mm

であった。エラー値を0mmから100mmまで切り出したときの度数分布は図13のようになった。提案手法の場合と比較すると、度数分布上ではエラーの小さな部分での度数分布は大きくは変わらない。しかしエラーの大きな部位を調べてみると、いずれもパーツ同士の距離が接近しているという共通点があった。カメラワークの最適化の効果はこういったエラーが大きくなりやすい箇所に顕著に現れているといえる。

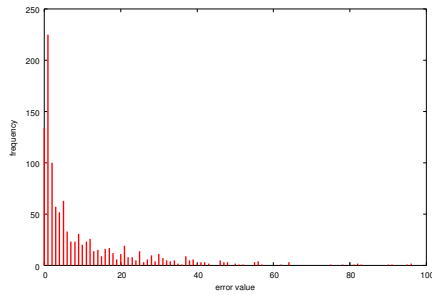


図 13: 初期段階での度数分布 (100mm までを抽出)

4.4 従来の視体積交差法でのカメラワーク

続いて、従来の視体積交差法で形状復元するために対象の全身が収まるように撮影すべくカメラの視線方向、ズームをコントロールした場合の結果は次のようになった。

- 空間解像度: 5mm/voxel
- 最大エラー値: 56mm

このときの度数分布を図 14 に示す。この結果を提案手法での結果と比較すると、エラーの最大値は提案手法と同じ値である一方、その他のエラーの度数分布を見ると、やはり提案手法でのカメラワークの形状復元誤差は全体として単純に視体積交差法で形状復元した場合よりも大きい。ただし、空間解像度は 5mm/voxel であり、提案手法での 1mm/voxel よりもはるかに低い計測精度である。そのため、従来手法では正しい形状を得られるポテンシャルを持っていても結局粗い形状しか得られないのに加え、テクスチャの取得精度も悪い。一方、3次元ビデオのためのデータ獲得方法として見た場合、誤差を低く抑えた形状と高解像度のテクスチャ情報が同時に得られる提案手法は、従来手法に基づく3次元ビデオ撮影手法に比べて、かなりの品質向上が望めると考えられる。

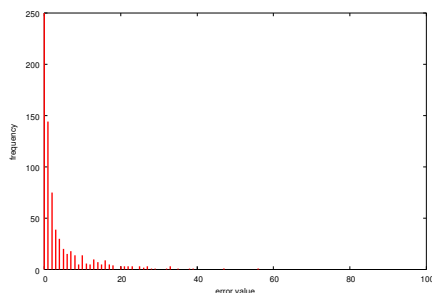


図 14: 全身エラー分布の度数分布 (従来手法)

5 結論と今後の課題

3次元の情報を持った新たなメディアである3次元ビデオ映像が一般に普及していくためには、従来の2次元映像に匹敵する品質を実現する必要がある。本研究ではこの高品質化という目的に対し、カメラをコントロールし、対象にズームアップ撮影して対象の形状およびテクスチャを高解像度化する、というアプローチをとった。そこで発生したのが、カメラは一体どこに注目すれば良いのかという問題である。

本研究ではまず対象の多関節形状モデルを導入し、対象の見え方を再現可能にすることで、カメラワークを評価可能にした。そしてカメラワークの評価に形状復元誤差の分布を用いることを提案した。さらにこの分布表現の特徴を活用してカメラワーク決定手法を提案した。以上の提案をシミュレーション環境で試し、単純にズームを掛ける場合と比較して形状復元誤差が大きく削減されることを確認した。この結果を踏まえて今後の課題としては以下のようなことがある。

動的環境への対応 本研究では1フレームのみの高解像度化に焦点を当ててきたが、3次元ビデオの撮影でこれを行うには対象の動きを予測したカメラコントロールや動対象に対するカメラワーク評価の枠組が必要になってくる。

カメラワーク決定の高速化 本研究で提案したカメラワークの決定手法は動的環境で用いる場合、決定までの所要時間の保証がないという問題がある。カメラ群をいくつかに分けて問題の規模を小さくしたり、最適化のプロセスを一定時間で打ち切るなどの工夫が必要となる。

本研究は基盤研究 (A)(2) 13308017、特定領域研究 (2) 13224051 の補助を受けて行った。

参考文献

- [1] Steven Abrams, Peter K. Allen, and Konstantinos A. Tarabanis. Dynamic sensor planning. *IEEE Proc. International Conference on Robotics and Automation*, pp. 605–610, 1993.
- [2] W. Niem, M. Steinmetz. Camera Viewpoint Control for the Automatic Reconstruction of 3D Objects. *ICIP 96, Lausanne*, pp. 16–19, September 1996.
- [3] Ivana Mikic, Mohan Trivedi, Edward Hunter and Pamela Cosman. Human Body Model Acquisition and Tracking using Voxel Data. *International Journal of Computer Vision*, pp. 199–223, July 2003.
- [4] A. Laurentini. How far 3d shapes can be understood from 2d silhouettes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17(2), pp. 188–195, 1995.
- [5] ウ小軍, 和田俊和, 東海彰吾, 松山隆司. 平面透視投影を用いた並列視体積交差法. 情報処理学会論文誌, Vol. 42, No. SIG6(CVIM2), pp. 33–43, 2000.
- [6] 西山 学. 視体積交差法を用いた高解像度形状復元のためのカメラワーク最適化. 京都大学大学院 修士論文, 2004.