

3次元ビデオ映像

3D Video

松山隆司

Takashi Matsuyama

京都大学大学院 情報学研究科

Graduate School of Informatics, Kyoto University

e-mail: tm@i.kyoto-u.ac.jp

ABSTRACT

3D video is NOT an artificial CG animation but a real 3D movie recording the full 3D shape, motion, and precise surface color & texture of real world objects. It enables us to observe real object behaviors from any viewpoints as well as to see pop-up 3D object images and to analyze accurate 3D behavior characteristics. We believe the exploitation of 3D video technologies will open up a new world of versatile cultural and social activities: visual communication, entertainment, training & learning, archiving, and so on. This paper gives an overview of our research attainments so far obtained: (1) a PC cluster system with distributed active cameras for real-time 3D shape reconstruction, (2) a dynamic 3D mesh deformation method for obtaining accurate 3D object shape, (3) a texture mapping algorithm for high fidelity visualization, and (4) user friendly 3D video editing system.

Keywords: 3次元ビデオ, PC クラスタ, 実時間3次元形状復元, 弾性メッシュモデル, 自由視点映像生成

1 3次元ビデオ映像とは

世界史を紐解くと、人類の文化・文明が記録・伝達メディアの開拓によって大きく進展したことが分かる。古くは石板、パピルス、紙、15世紀のグーテンベルグの印刷装置による書籍、新聞、19世紀の写真の発明へとメディア世界が広がり、20世紀には電子技術によるラジオ・テレビ放送、テープレコーダ・ビデオによる記録、そして最近ではデジタル・ネットワーク技術によるWWWへと発展し、今ではだれもがマルチメディアを使った国際的情報発信が可能となった。

一方、記録・伝達される内容（コンテンツ）としては、静止した文字や図形から、時間的に変化する1次元の音声・音楽、2次元の画像・映像を経て最近ではCGによる3次元アニメーションへと発展してきた。

では、21世紀におけるメディアにはどういった発展が期待され、それが文化・文明をどのように変えるのであろうか。その答えの1つとして考えられるのが3次元ビデオ映像である。

3次元ビデオ映像は、CGによる仮想的・人工的な3次元アニメーションではなく、ダンスやスポーツをする人間、自然界の動物などの生の姿・形・色の時間的変化を3次元的にそのまま記録した実写立体ビデオ映像で、実世界における対象の振る舞い・動作を余すところなく記録した究極の映像メディアである。

3次元ビデオ映像は、ラジオやテレビが人類のあらゆる活動に与えた影響と同じあるいはそれ以上のもの

をもたらすのではないかと考えている。具体的には、3次元ビデオ映像を利用することによって、以下のような応用システムが実現できる。

1. 人間国宝やオリンピック選手の動作をそのまま3次元的に記録・分析できる身体技能・芸能デジタルアーカイブ（今しか撮れない動作を完全に記録し、後世に伝える。）
2. 動物のありのままの生態を多角的に観察できる3次元ビデオ映像DVD図鑑
3. 自分の動作と手本とを3次元的に比較しながらトレーニングできるリハビリ、スポーツ練習システム（タイガー・ウッズと自分のゴルフスウィングを比較する。）
4. デジタルテレビ放送や広帯域インターネットをインフラとして使った3次元テレビ放送
5. 遠隔地の雰囲気をもそのまま伝えることができる高臨場感遠隔会議・講義・実験システム
6. 人間の動作や仕種を見守り、それに反応するインタラクティブな情報システム（現在のコンピュータシステムのように、キーボードやマウスを使ってコマンドを入力しなくてもよい新たな情報システム）

これらの例や3次元ビデオ映像という名前からイメージされるのは映像が飛び出して見える立体映像であり、

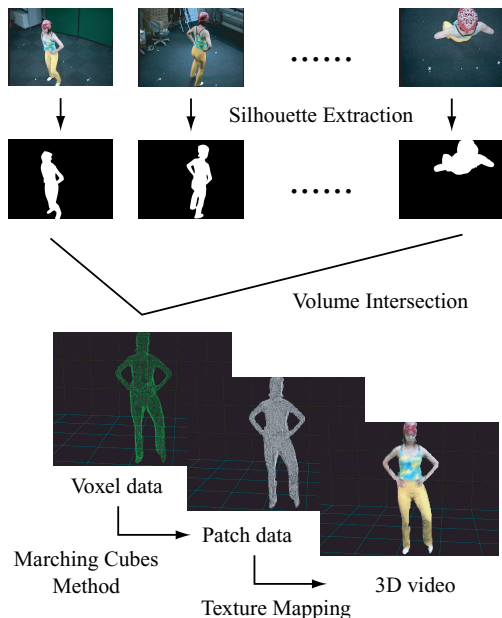


図 1: 3次元ビデオ映像の生成プロセス

3次元ビデオ映像は立体ディスプレイ用の特殊なコンテンツだと思われるかもしれないが、通常の2次元のディスプレイを使った場合でも、3次元ビデオ映像は映像の楽しみ方に革新的変化をもたらす。具体的には、3次元ビデオ映像を使えば、視聴者が映像を鑑賞する位置や方向、視野をその場でインタラクティブに変えることができる。つまり、大相撲の放送が3次元ビデオ映像を使って行われているとすると、チャンネルはすべて同じであるにも関わらず、ある家ではかぶりつきからある力士をズームアップした映像が、別の家庭では土俵全体の様子がテレビに映し出されているといったことができる。

このように、3次元ビデオ映像の持つ大きな特徴は、視聴者がその場でインタラクティブに映像と楽しめることにあり、映像の楽しみ方に大きな変革をもたらすものと考えられる。インタラクティブ映像の効果は、テレビゲームによってテレビ番組を見る時間が減ったことから明らかである。

一方、インタラクティブな映像鑑賞を実現する方法としては、自由視点テレビ [1] に代表される Image-Based Rendering があるが、3次元ビデオ映像では対象の3次元形状モデルを持つため、上記の1, 3, 6のような応用が可能となり、単なる任意視点映像生成を越えた機能が実現できるという大きな利点がある。

2 3次元ビデオ映像撮影の基本方式

図1は、現在我々が用いている3次元ビデオ映像撮影・生成過程を示したもので、以下のプロセスから構

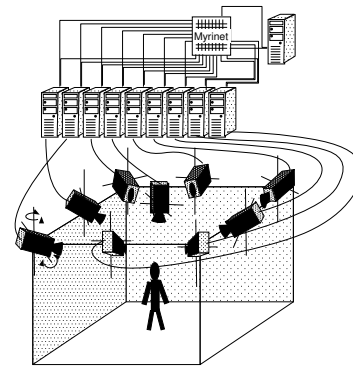


図 2: 能動的実時間3次元形状復元のためのPCクラスタシステム

成されている。

1. **多視点同期ビデオ撮影:** 異なる視点に取り付けられたアクティブカメラが実時間で対象を追跡し、対象の多視点ビデオ映像を同期撮影する (図2, 図1最上段)。
2. **対象シルエットの抽出:** 背景差分を用いて各ビデオフレームから対象のシルエットを抽出する (図1上から2段目)。
3. **視体積交差法による3次元ボクセルデータの獲得:** 異なる視点から得られた対象のシルエットを共通の3次元空間に逆投影し、得られた視体積の共通部分を求め、対象の3次元形状をボクセル表現として得る (図1下から3段目)。
4. **表面形状の計算:** 離散 Marching cubes 法 [2] を用いて3次元形状のボクセル表現を表面パッチ表現に変換し、それを初期値として弾性メッシュモデルを適用し対象の詳細な表面形状を求める (図1下から2段目)。
5. **テクスチャマッピング:** 対象の3次元形状を表す各表面パッチに対するテクスチャを、多視点ビデオ映像から求めマッピング処理を行う (図1最下段)。
6. 以上の処理で、1フレームの3次元ビデオ映像が生成される。動画を作るには上記の処理を繰り返せばよい。

3 研究成果の概要

我々の研究室では、5年ほど前から3次元ビデオ映像の生成、表示、編集、符号化について研究を展開してきており、これまでに得られた成果をまとめると以

下ようになる。なお、技術的詳細、性能評価などについては参考文献を参照して頂きたい。

(1) 能動的実時間3次元ビデオ映像撮影システムの開発

IEEE1394 インターフェイスを持つデジタルビデオカメラを搭載した視点固定型パン・チルト・ズームカメラを新たに製作するとともに、カメラ25台、PC30台からなるPCクラスタシステムを開発した。この新システムでは、図2のように天井カメラに加えて、床カメラが設置されており、約100平米の広さのスタジオで移動する対象を高解像度で撮影ができるように設計されている[3]。

(2) 準視点固定型パン・チルト・ズームカメラのキャリブレーション法の開発

準視点固定型パン・チルト・ズームカメラという首振りカメラに対する高精度なカメラモデルを考案し、そのキャリブレーション法を開発した[4]。これによって、広範囲を移動する対象を追跡しながら多視点映像を撮影することが可能となった。

(3) 実時間3次元動作計測システムの開発

多視点ビデオ映像から対象の3次元動作を実時間で復元するための方法として、まず平面ベースの視体積交差法を開発し、その並列パイプライン処理化をPCクラスタを用いて実現した[3]。

この方法では、

1. 対象が存在する3次元空間を平行な平面群として表し、
2. 各撮影画像から抽出されたシルエットを平行平面群中の基準平面に投影する（図3左図の1の投影計算）。
3. 次に基準平面上のシルエットを他の平行平面群に投影し（図3左図の2の投影計算）、
4. それぞれの平面上で各撮影画像から得られた2次元シルエットの共通部分を求め、それを対象の断面形状とする（図3右図）。
5. 得られた断面形状が平行平面間の間隔の厚みを持つものと考え、それらを3次元的に積み重ねて対象の3次元ボクセル形状を求める。

図4は、PCクラスタを用いた並列化法を示したもので、図の各列は1台のPCでの処理過程を示している。

1. カメラを備えたPCに撮影用同期信号を出し、多視点画像の同期撮影を行う。
2. 各PCが背景差分により対象のシルエットを抽出し、それを基準平面（実験では床面）に投影して基準シルエットを求める。

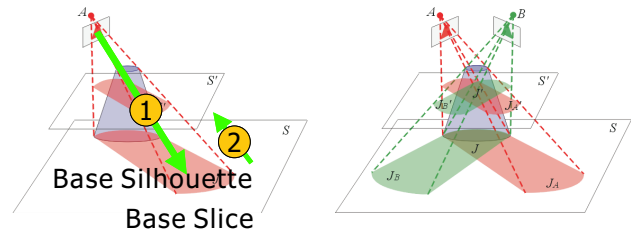


図3: 平行平面間透視投影に基づく視体積交差法

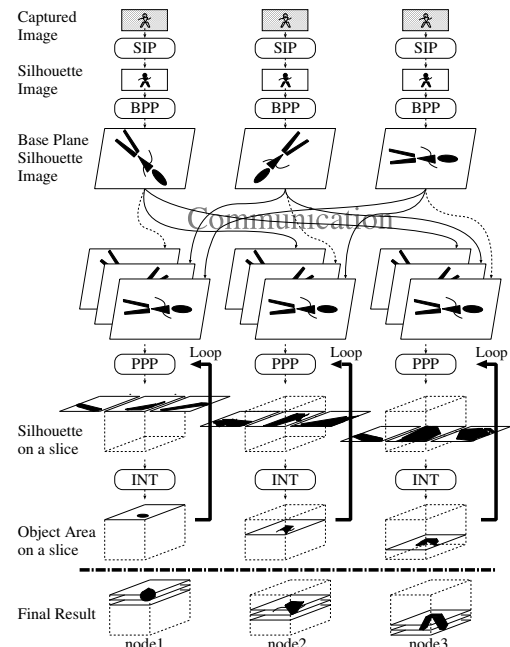


図4: 並列パイプライン処理の流れ

3. カメラを持たないPCも含め、全てのPCが全ての基準シルエットのコピーを持つように相互通信を行う。
4. 平行平面群をグループに分け、各グループをそれぞれ1台のPCに割り当てる。各PCは割り当てられた平行平面群を対象に基準シルエット群の平行平面への投影とその上でのシルエット交差計算を行い、対象の断面形状を求める。

試作したシステムでは、さらなる高速化を図るため、各PCで実行される処理を画像キャプチャ、シルエット画像生成、基準平面への逆投影、基準シルエットの相互コピー、基準シルエットの平行平面群への投影と交差計算の5ステージに分け、パイプライン処理を行わせるようにした。

このシステムでは、2cmの空間解像度で毎秒10フレームを越える速度で人間の3次元形状が復



図 5: 人体部位のズームアップ撮影の例

元できることが実証されている。さらに、現在は、3平面視体積交差法という新たな3次元動作計測法を考案し、対象が移動したり、カメラが首を振ったりした場合でも、安定かつ高速に3次元動作が復元できることを示し、その性能評価を行っている。

(4) 人物の姿勢に基づく多視点カメラワークの最適化法の開発

視体積交差法によって得られた3次元ボクセルデータに対して人体の多関節モデルをあてはめ、得られた姿勢に応じてカメラ群の視線方向およびズーム倍率を制御して各人体部位をズームアップ撮影することによって（図5）、詳細な3次元形状を復元するアルゴリズムを開発した[5]。これによって、人物の姿勢に応じた高解像度の多視点映像の撮影、3次元形状の復元が可能となった。

(5) 高精度3次元形状復元アルゴリズムの考案

視体積交差法で得られた対象の3次元ボクセルデータの表面に三角形パッチを張ったものを初期形状として、画像に写された(a)対象表面のテクスチャおよび(b)対象の輪郭形状、(c)3次元空間中での対象表面の滑らかさを評価基準としてパッチ形状を変形させることにより、対象の3次元形状を高精度に復元する弾性メッシュモデルを考案し、実験によってその有効性を示した[6]。さらに、考案した弾性メッシュモデルを拡張し、動的に変化する3次元ボクセルデータから対象の3次元動作を高精度に復元する動的弾性メッシュモデルを考案し、実験によってその有効性を示した。図6は、提案した弾性メッシュモデルの有効性を示したもので、滑らかな対象表面形状の復元とともに、視体積交差法では復元できなかった凹部の形状もうまく復元できている。

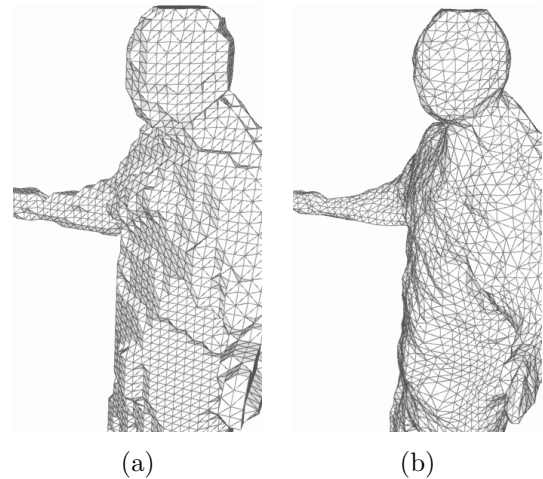
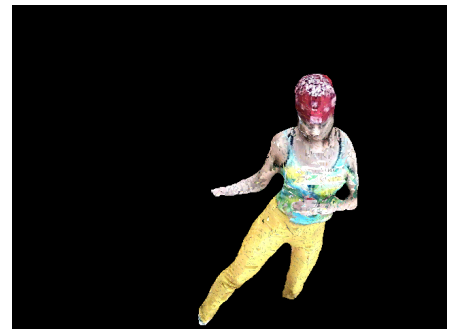


図 6: (a) ボクセルデータから離散マーチングキューブ法で得られた3次元メッシュ(b) 弾性メッシュモデルによる変形を行って得られた3次元メッシュ

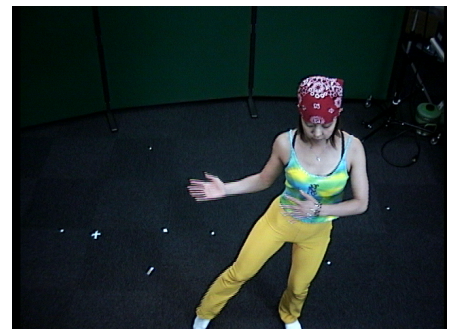
VDVBM



VIPBM



Original
sequence



frame # 126

図 7: 生成された画像と元画像



図 8: 3次元ビデオと全方位パノラマ画像を使ったコンテンツ

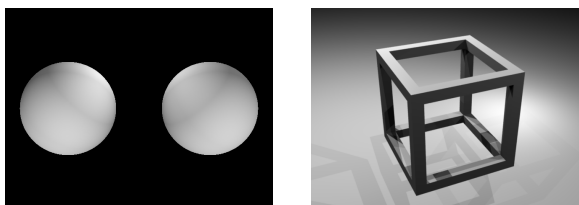


図 9: 複数球を使った光環境センシング

(6) 高品質テクスチャマッピングアルゴリズムの考案

撮影された多視点ビデオ映像に基づいて、対象表面を構成する各三角形パッチに対するテクスチャを求めるアルゴリズムとして、視点依存型頂点ベース法 (VDVBM) を新たに考案し、従来の視点独立型面ベース法 (VIPBM) に比べて大幅に画質が向上することを実験的に明らかにした [7]. 図 7 は、両者によって生成された画像を比較したもので、提案手法 (VDVBM) ではジッタの少ない滑らかな映像が得られることが分かる.

また、前述の (5) の弾性メッシュモデルによって復元された 3 次元形状を使うことによって、生成される画像の品質が向上することも実証でき、この意味でも弾性メッシュモデルの有効性が確認された.

(7) 3次元ビデオ映像と全方位パノラマビデオを使ったコンテンツ編集

生成された 3 次元ビデオと全方位パノラマ画像とを素材として、多様な映像コンテンツをインタラクティブに編集することができるシステムを開発した [7]. 図 8 に 3 次元コピーを使った編集後の映像例を示す.

(8) 参照球を用いた光環境センシング法の考案

図 9 のように複数の球を実世界に配置し、その表面上での陰影パターンを解析することによって光源の種類 (無限遠光, 近接光, 環境光), 3 次元位置, 方位を推定するアルゴリズムを考案し、シミュレーションによってその有効性を示した [8].

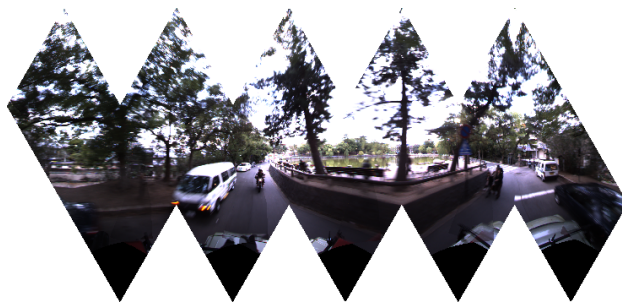


図 11: 正 20 面体展開図を用いた全方位ビデオ映像の符号化

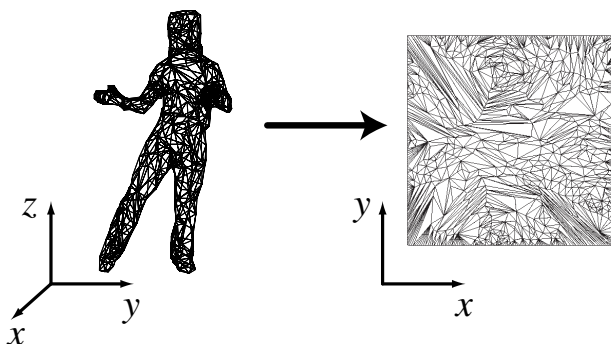


図 12: 人物の 3 次元メッシュを展開した 2 次元矩形メッシュ

これによって、3 次元ビデオ映像に対して、光源の付加や削除といった新たな編集が可能となる.

(9) セルフシャドウを用いた光環境センシング法の考案

図 10 のようなスケルトン・キューブ (枠のみからなる立方体) が多数の照明によって照らされてできるセルフシャドウおよびシェーディングから、光源種類, 位置, 方位を推定するアルゴリズムを考案し、シミュレーションによってその有効性を示した.

(10) 全方位ビデオ映像の符号化法の開発

高解像度全方位ビデオ映像を符号化する方法として、正多面体展開図法を考案し、いずれの正多面体を用いるのが最も効率的であるのかを実験的に評価し、MPEG 会議にてその結果を報告した [9]. 図 11 は、球面スクリーン上に写された全方位ビデオ映像を正 20 面体表面に投影し、その展開図によって 2 次元平面上の映像として表現したもので、これを MPEG 方式で圧縮・伝送する.

(11) 3次元ビデオ映像の符号法の開発

3 次元ビデオ映像の符号化法として、メッシュデータとして表された対象の 3 次元表面を平面矩形領

域に展開・マッピングする Geometry Image 法 [10] を適用し、それに MPEG 符号化を施すことによって高効率の 3 次元ビデオ映像符号化が可能であることを示した。図 12 は、人物の 3 次元メッシュを平面展開して得られた 2 次元矩形メッシュで、矩形中の各画素に 3 次元メッシュにおける (x, y, z) の 3 次元座標値を記録することによって、3 次元形状が 2 次元画像として表現できる。

4 おわりに

以上述べたように、これまでの研究によって、3 次元ビデオ映像の能動的実時間撮影・圧縮・編集・表示法の開発に関する基本的な手法が揃ったが、3 次元ビデオ映像の実用化のためには以下の課題を解決する必要がある、現在それに向けた研究を進めている。

ビデオレート処理の実現：現在の PC クラスタでは、カメラのハードウェア特性および PC 台数の制限から、3 次元動作復元処理のスピードは最大 15 フレーム/秒となっている。これをビデオレートに上げるには、カメラ、PC クラスタといったハードウェア機能の向上および処理アルゴリズムの効率化が必要である。特に、本研究で考案した弾性メッシュモデルによる高精度 3 次元形状復元アルゴリズムは、実時間処理を考慮しておらず、処理の並列化や効率化をどのようにするかは今後の課題となっている。

広域、多人数の撮影：現在のシステムは、撮影対象を 1 人としているが、実用化のためには、広い空間で行われる複数の人物の動作を撮影する必要がある、そのためのカメラワークプランニング法を考案する必要がある。

照明の影響の削除と編集：撮影された際の照明環境を推定し、その影響を映像から取り除き、新たな照明環境の下での映像を生成することができれば、3 次元ビデオ映像の編集にさらなる多様性を持たせることができる。

魅力的なコンテンツの開発：3 次元ビデオ映像を広く普及させるには、技術の開発に加え、魅力的なコンテンツ製作が不可欠である。我々のグループでは、デジタルアーカイブ開発として意義がある人間国宝の方々による無形文化財を 3 次元ビデオ映像として記録することを計画している。

国際標準化：現在 MPEG で行われている 3 次元ビデオ映像の符号化に関する標準化活動では、我々が提案しているモデルベース符号化とイメージベース符号化法が並列的に議論されており、国際標準化のためには、新たな符号化法の提案に加え、符

号化効率の比較などを行い、継続的に活動を展開する必要がある。

本研究は文部科学省科学研究費基盤研究 (A) (No. 13308017) の補助を受けて行った。研究と一緒に進めてくれている研究室のメンバに感謝します。

参考文献

- [1] T. Fujii, T. Kimoto, and M. Tanimoto: A New Flexible Acquisition System of Ray-Space Data for Arbitrary Objects, IEEE Trans. on Circuits and Systems for Video Technology, Vol.10, No.2, pp.218-224, 2000
- [2] 剣持 雪子, 小谷 一孔, 井宮 淳: 点の連結性を考慮したマーチング・キューブ法, 電子情報通信学会技術報告, PRMU98-218, pp.197-204, 1999.
- [3] T. Matsuyama, X. Wu, T. Takai, and T. Wada: Real-Time Dynamic 3D Object Shape Reconstruction and High-Fidelity Texture Mapping for 3D Video, IEEE Trans. on Circuits and Systems for Video Technology, Vol.CSVT-14, No.3, pp.357-369, 2004
- [4] 近藤純司, ウ 小軍, 松山隆司: 準視点固定型パン・チルト・ズームカメラのキャリブレーション, 情報処理学会研究会資料, CVIM137-19, 2003
- [5] 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, 京都大学大学院情報科学研究科修士論文, 2004
- [6] 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, No.SIG-11(CVIM 5), pp.53-63, 2002.12
- [7] 松山隆司, 高井勇志, ウ 小軍, 延原章平: 3 次元ビデオ映像の撮影・編集・表示, 日本バーチャリアリティ学会論文誌, Vol.7, N0.4, pp.521-532, 2002
- [8] T. Takai, K. Niinuma, A. Maki, and T. Matsuyama: Difference Sphere: An Approach to Near Light Source Estimation, CVPR, 2004 (to appear)
- [9] 曾良洋介, 波部斉, 山澤一誠, 松山隆司: 多面体モデルを用いた全方位ビデオの圧縮, 情報処理学会研究会資料, CVIM138-9, 2003.5
- [10] X. Gu, S.J. Gortler, and H. Hoppe: Geometry Images, Proc. of SIGGRAPH 2002, pp.355-361, 2002