

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 6月 5日現在

機関番号：11201

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21500090

研究課題名（和文） 内外方向撮影を利用した複雑物体のための高品質ビデオベースCG技術の開発

研究課題名（英文） Development of high-quality video-based computer graphics methods for complicated objects using inward and outward camera capturing

研究代表者

藤本 忠博（FUJIMOTO TADAHIRO）

岩手大学・工学部・准教授

研究者番号：00312512

研究成果の概要（和文）：本研究では、内向きカメラ（前景カメラ）と外向きカメラ（背景カメラ）による内外方向撮影を利用した複雑物体のための高品質ビデオベース・コンピュータグラフィックス技術として、前景カメラと背景カメラを用いた効率的な前景物体の抽出、複雑で微細な半透明物体のための α 値を用いた形状表現と映像表示、多視点映像による自由視点映像の効率的な生成、テクスチャ合成を利用した多視点画像からの高品質な自由視点画像の生成を行う技術を開発した。

研究成果の概要（英文）：In this research, we developed high-quality video-based computer graphics methods for complicated objects using inward and outward camera capturing by inward cameras (foreground cameras) and outward cameras (background cameras): an efficient method for extracting foreground objects using foreground and background cameras, modeling and rendering methods for complicated semitransparent objects using alpha values, an efficient rendering method for free-viewpoint images from multi-viewpoint images, a method for generating high-quality free-viewpoint images from multi-viewpoint images using texture synthesis.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	3,000,000	900,000	3,900,000
2010年度	200,000	60,000	260,000
2011年度	500,000	150,000	650,000
年度			
年度			
総計	3,700,000	1,110,000	4,810,000

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：コンピュータグラフィックス、イメージベースCG、ビデオベースCG、コンピュータビジョン、多視点映像

1. 研究開始当初の背景

異なる視点位置からカメラ撮影した複数の参照画像（多視点画像）を用いて新たな自由な視点位置から見た画像（自由視点画像）を合成する技術を一般にイメージベース・コンピュータグラフィックス（イメージベース

CG）と呼ぶ。その研究萌芽期は1990年代前半であり、当初は静止画像が対象であったが、2000年代に入りビデオ（動画）映像を扱う研究が主流となり、ビデオベースCGとも呼ばれるようになった。ビデオベースCGでは、映像品質とともに、2つのリアルタイム

(実時間処理)性が重視される。一つは、ユーザによる自由視点の移動操作にリアルタイムで追従して映像を合成する視点移動リアルタイム性である。もう一つは、カメラ撮影にリアルタイムで同期し、ビデオフレームの獲得速度に合わせて映像を合成する撮影同期リアルタイム性である。現在まで、この両者を実現して十分に高品質な映像を合成する技術は提案されていない。例えば、W. Matusik (MIT) らの手法 (SIGGRAPH 2000, Eurographics Workshop 2001) は、両リアルタイム性を満たしてビジュアルハルと呼ぶ対象物の近似形状を構築するが、その近似精度と映像品質は低い。C. L. Zitnick (Microsoft Research) らの手法 (SIGGRAPH 2004) は、高度なコンピュータビジョン (CV) 技術を用いて視点移動リアルタイム性を満たす高品質映像を生成するが、元のビデオ映像の前処理に時間がかかり、撮影同期リアルタイム性はない。国内でも、京都大学松山研究室におけるボクセル視体積交差法に関する一連の研究や、東京大学相澤研究室におけるデータ圧縮等を考慮した3次元ビデオの研究など、様々な研究が行なわれている。しかし、国内外の研究事例を通して、本研究代表者が知る範囲では、映像の高品質化と両リアルタイム性をともに十分に満たす技術はないのが現状である。

一方、一般に、限られた画素 (ピクセル) 数のカメラとディスプレイによって複雑かつ微細な対象物を高品質映像化するには、各画素の不透明度 (α 値) の利用が有効である。 α 値とは、1画素内に対象物と背景が映り込んだ際、対象物が占める割合を表す値 ($0 \leq \alpha \leq 1$) である。これを利用する技術として、画像上で対象物が映る画素領域を背景領域から α 値付きで分離抽出し、対象物だけの RGB α 画像を求める α マッピングがある。複雑な自然背景を持つ静止画像に対する Y. Chuang (Washington 大学) らによるベジアンマッピング (CVPR 2001) や、これをビデオ映像 (単視点) に拡張したベジアンビデオマッピング (SIGGRAPH 2002) など、高度な技術が提案されている。また、N. Joshi (California 大学) らによって多視点ビデオ映像から自動的に対象物の RGB α 映像を得る手法 (SIGGRAPH 2006) が提案された。これは、焦平面 (フォーカルプレーン) を利用した効率的な対象物の位置推定による撮影同期リアルタイム性はあるものの、自由視点移動ができない。自由視点移動が可能な手法には、W. Matusik (MIT) らによるビジュアルハルに α 値と反射モデルを導入したオパシティハル (SIGGRAPH 2002)、A. Reche (REVES/INRIA) らによるボクセルに α 値を導入した樹木モデル (SIGGRAPH 2004) などがあるが、多視点静止画像を用いている。

現状、ビデオベースCGにおける既存技術の多くは、人物や建物など、表面や輪郭が鮮明な単純形状のみを扱う。例えば、燃え上がる炎、液体の流れ、風で揺らぐ樹木のような、複雑で微細な形状と動きを伴う対象物の高品質な自由視点映像化を目的として、多視点ビデオ映像に対して視点移動かつ撮影同期リアルタイムで自由視点から見た RGB α 映像を得る方法は提案されていない。

2. 研究の目的

本研究では、複雑で微細な形状と動きを伴う対象物に対して、視点移動と撮影同期の両リアルタイム性を満たしながら、多視点ビデオ映像から高品質な自由視点 RGB α 映像を生成するビデオベースCG技術を開発する。これを実現するには、各ビデオ映像上での対象物領域の分離抽出を高速かつ高精度に行う必要がある。そこで、本技術では、対象物を2系統のカメラ群で撮影する。一方は、対象物を周囲から撮影する複数の内向きカメラ (対象物カメラ、前景カメラ) である。他方は、逆に、対象物付近から周囲を見渡す全周背景を撮影する外向きカメラ (背景カメラ) である。これらの内外方向撮影で獲得される映像情報を相互に利用し、対象物近傍を中心とした対象物と背景を含むシーン全体に関する詳細な3次元映像空間情報を構築し、対象物領域抽出の高速化と高精度化の実現を目指す。さらに、映像中の各画素の不透明度 (α 値) を利用することで、微細な形状と動きを高品質に自由視点映像化する。

3. 研究の方法

前節までに述べた研究の背景と目的のもとで本研究を開始した。しかし、研究を進めていく過程で生じた様々な問題や新たに検討が必要と思われた課題の発見などにより、適宜、研究の方向性を変更する必要が生じた。その結果、当初の最終目標であった、複雑で微細な対象物の高品質な自由視点 RGB α 映像を視点移動かつ撮影同期リアルタイムで生成する技術の完成までには至らず、その実現のための幾つかの基礎技術の開発で研究期間を終えることとなった。そこで、本報告書では、実際に開発を行った内容について整理し、報告を行う。

本研究で開発を行った研究項目について、その内容と方法を以下に述べる。

(1) 前景カメラと背景カメラを用いた効率的な前景物体の抽出

ビデオ映像中で背景から対象物 (前景物体) のみを抽出する方法に背景差分がある。最も簡単な方法は、背景のみが映る1枚の背景画像を用い、前景物体を含めて撮影したビデオ映像のフレーム画像ごとに背景画像との差分を求めることで前景物体の画素領域

のみを抽出する。しかし、この方法は、背景が不変であることが前提で、背景に時刻による環境変化（明るさの変化など）や動的な物体がある場合などは、それらの画素領域も抽出されてしまう。そのような問題を解決するために多くの手法が提案されてきたが、ほとんどは背景の微小な緩やかな変化に対応するものであり、前景物体と同等の動きを伴う物体が背景にあるような場合には前景物体のみの抽出は難しい。一方、物体を周囲から撮影した複数のビデオ映像から3次元形状を復元するイメージベースCGでは、各映像上で物体を抽出する必要がある場合が多い。

そこで、本研究では、前景物体を周囲から撮影した複数の映像から前景物体の画素領域（前景領域）を抽出することを最終目標とし、背景に大きな変化がある場合にも適切に前景領域のみを抽出する手法を提案する。本研究では、前景カメラと背景カメラを用いた方法を採用する。図1のように、前景物体を取り囲むように前景カメラと背景カメラを配置し、前景カメラは内側の物体に向け、背景カメラは外側の背景に向ける。そして、背景カメラによる背景映像（背景のみ）と前景カメラによる前景映像（前景物体と背景）の間で適切な座標変換に基づく背景差分を行い、前景映像中の前景領域のみを抽出する。本研究では、上記の最終目標を実現するための基礎技術として、図1のように、背景カメラ3台と前景カメラ1台を用い、次の2つの方法を提案する。

【方法1】 複数の背景映像をホモグラフィ行列により一枚の背景合成映像に合成し、背景合成映像と前景映像との背景差分を行う。

【方法2】 エピポーラ幾何を用いて背景映像ごとに前景映像との背景差分を行う。

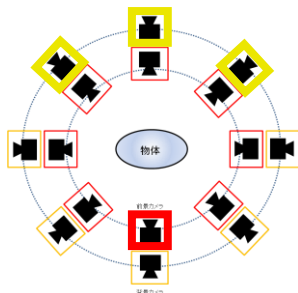


図1: 前景カメラ（赤枠）と背景カメラ（黄枠）の配置。本研究では、太枠のカメラのみを使用した基礎技術を提案する。

(2) 複雑で微細な半透明物体のための α 値を用いた形状表現と映像表示

(2-1) 3次元点群の α 値の推定

本研究では、複数の α マット画像に対して、各ピクセルを通過するレイに依存した空間サンプリング（レイベースサンプリング）を行うことで、カメラ配置に応じて適切に分布する3次元点群により半透明物体の形状を復元する手法を提案する。まず、半透明物体を異なるカメラ位置から撮影して生成した

複数の α マット画像を用い、各カメラ位置から各画像のピクセルにレイを飛ばし、レイ同士の交点をサンプリング点として求める。次に、各 α マット画像上の物体輪郭に基づく形状復元を行う。そして、各サンプリング点の α 値を推定することで半透明要素を伴う形状復元を行う。

(2-2) α 値を伴う3次元点群の表示

本研究では、物体形状を3次元点群として効率的に管理するLDI (Layered Depth Image) を利用した半透明物体の効率的な表示法について提案する。本手法では、 α 値を持つ3次元点群で表現された半透明物体に対して、異なる基準視点による複数のLDIを用いた点群の分割登録を行うことにより、単一のLDIの場合に生じる3次元空間上の位置によるサンプリング密度の疎密の問題を改善する。また、複数のLDIに対してマクミランの順序付けアルゴリズムを導入することで、高品質な自由視点画像を効率的に生成する。

(3) 多視点映像による自由視点映像の効率的な生成

(3-1) テクスチャ投影によるビジュアルハルの効率的な表示

本研究では、複数のカメラによる多視点カメラ映像から、テクスチャの投影マッピングを利用した視体積交差法により、リアルタイムで対象物体のビジュアルハルを復元して描画する手法を提案する。本手法では、複数の平板を平行かつ等間隔に並べた平行平板群を3軸方向に組み合わせた平行平板モデルによりボクセルモデルを疑似表現し、ボクセルベースの視体積交差法を模擬する。本手法は、通常のボクセルベース視体積交差法に比べて非常に高速であり、動きを伴う対象物体に対してもリアルタイムで自由視点映像を生成することが可能である。また、メモリ使用量も少なく、メモリ不足により高解像度のボクセルモデルを扱えないという従来の問題も解決される。本手法では、まず、カメラごとに、フレーム画像に背景差分を適用し、ピクセルごとに α 値演算に必要な α 値と視点依存型カラーマッピングに必要なRGB値を設定して背景差分画像を生成し、投影用テクスチャとする。そして、全てのカメラの投影用テクスチャを平行平板モデルに対して投影マッピングし、各投影用テクスチャの α 値による積演算によってビジュアルハルを求め、視点依存型カラーマッピングによる自由視点からのレンダリングにより自由視点フレーム画像を得る。

(3-2) ビルボードによる自由視点映像の効率的な表示

多視点映像から自由視点映像を生成する代表的な手法として、以下の方法で3次元形

状を復元して表示するものがある：1) 映像上の対象物体の輪郭や色の情報からボクセルモデルにより復元する。2) ステレオマッチングによる異なる映像間のピクセルの対応付けにより対象物体までのデプス値を求めて復元する。しかし、前者はボクセル空間の解像度により復元精度が限定される問題があり、後者はピクセルの対応付けに時間を要する問題がある。そこで、本研究では、リアルタイムでの自由視点映像の生成を目標とし、ビルボードを用いて効率的に自由視点映像を生成する手法を提案する。本手法は、カメラ映像のフレーム時刻ごとに物体表面上の点群をリアルタイムで推定し、その各点上にビルボードを配置することで自由視点映像を生成する。フレーム時刻ごとに、まず、各映像上で SURF 特徴点を抽出し、異なる映像間で共通の特徴点を対応付ける。次に、カメラ校正によるパラメータを用いて各特徴点の3次元位置を推定する。そして、その3次元位置に適切な大きさのビルボードを自由視点に向けて配置する。最後に、ビルボード上に適切なカメラ映像の投影を行うことで自由視点映像をリアルタイムで生成する。

(4) テクスチャ合成を利用した多視点画像からの高品質な自由視点画像の生成

多視点画像から自由視点画像を生成する代表的なアプローチの一つとして、物体の3次元形状をボクセルモデルで復元して自由視点からレンダリングする方法がある。しかし、ボクセル空間の解像度の制限から、復元精度が低く、結果として自由視点画像の品質が低下する問題がある。また、一般に、参照画像の枚数が少ない場合、参照画像を撮影したカメラ位置から自由視点位置が離れるほど自由視点画像の品質は低下する。そこで、本研究では、復元したボクセルモデルを自由視点に対してレンダリングした画像を“仮の”自由視点画像とし、その画像に対してテクスチャ合成を行なうことで品質を改善する手法を提案する。本手法では、まず、対象物体を異なるカメラ位置から撮影した複数の参照画像に対して、視体積交差法とボクセルカラーリング法を適用することで、ボクセルモデルで3次元形状を復元する。そして、そのモデルを自由視点からレンダリングして仮の自由視点画像を得る。最後に、その画像に対して PatchMatch 法とユーザ制御型テクスチャ合成法を適用することで高品質な自由視点画像を生成する。

4. 研究成果

前節の研究項目(1)～(4)の成果を述べる。

(1) 前景カメラと背景カメラを用いた効率的な前景物体の抽出

実験の結果、方法1では、背景の多くの部

分が前景物体とみなされ、正確な前景物体領域を得ることができなかった。一方、方法2では、ある程度の精度で前景物体領域を得ることができた。背景に動的な物体が置かれた場合も背景とみなし、前景カメラと背景カメラの間に位置する前景物体のみの抽出が可能となった。方法2による実験結果のフレーム画像を図2、処理時間を表1に示す。各映像の解像度は640×480であり、表中の参照画素数はエビポーラ線上で探索を行う画素数である。

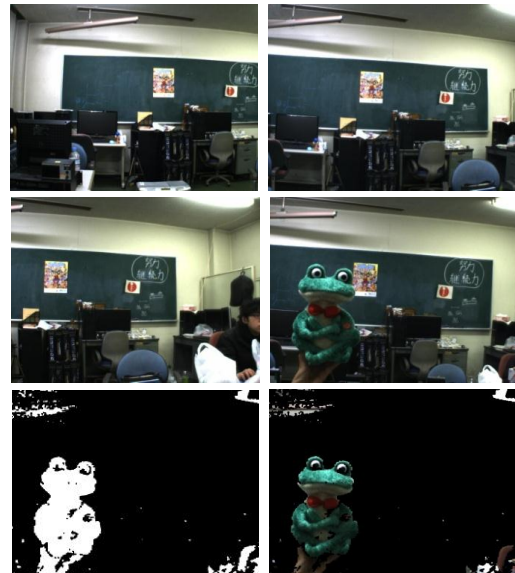


図2：実験結果。上段左右，中段左：背景映像。中段右：前景映像。下段左：各背景映像に対する背景差分の結果を論理積で統合したマスク映像(白：前景画素，黒：背景画素)。下段右：前景映像にマスク映像を重ねて生成した前景物体の抽出結果。

表1：処理時間

参照画素数	処理時間(s)	フレームレート(fps)
5	0.094	10.64
11	0.110	9.09
21	0.150	6.67
31	0.170	5.88
41	0.220	4.55

(2) 複雑で微細な半透明物体のための α 値を用いた形状表現と映像表示

(2-1) 3次元点群の α 値の推定

4枚の α マップ画像に対し、従来のボクセルベースサンプリングと本手法のレイベースサンプリングにより3次元点群の α 値を推定した結果を図3、処理時間を表2に示す。ほぼ同等のサンプリング点数に対し、本手法は、カメラ配置に依存した適切な位置でのサンプリング点の生成と妥当な精度による効率的な α 値の推定がなされている。

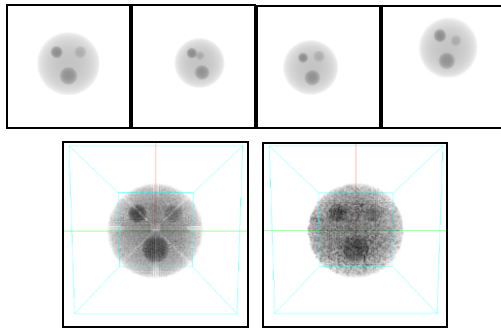


図 3：実験結果. 上段： α マット画像. 下段左：ボクセルベースサンプリングによる結果. 下段右：レイベースサンプリングによる結果.

表 2：処理時間

	ボクセルベース	レイベース
サンプリング点数	125,000 (= 50^3)	157,732
α 値推定時間(秒)	9.46918	2.25122

(2-2) α 値を伴う 3 次元点群の表示

具体的な 3 次元点群として図 4 のボクセルモデルを用い、LDI を 1 個だけ用いた通常の方法と 6 個を用いた本手法による自由視点映像の描画結果を比較した。前者では、自由視点に対して重なり合う複数のボクセルの前後関係が視点移動に伴い急激に入れ替わる部分が多数あり、画面のちらつきが見られた。後者では、その問題が軽減されており、本手法の有効性が確認された。

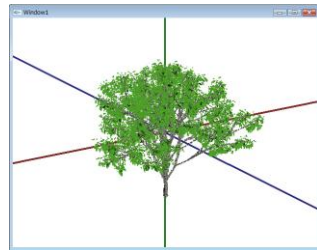


図 4：樹木のボクセルモデル

(3) 多視点映像による自由視点映像の効率的な生成

(3-1) テクスチャ投影によるビジュアルハルの効率的な表示

4 台のカメラを用いた実験結果を図 5 に示す。自由視点の位置に応じた適切な自由視点映像が生成されている。また、処理時間を図 6 に示す。比較手法は従来のボクセルを用いた視体積交差法であり、復元空間はボクセル空間を意味する。提案手法は、比較手法ではメモリ不足で扱えない高解像度のボクセル空間を処理でき、解像度が高まるほど優位性が増している。なお、提案手法では、あるボクセル解像度から処理時間が急激に増加しているが、ビデオカードの性能等に起因するものと考えられる。提案手法は通常必要と考えられるボクセル解像度の範囲ではリアルタイムで自由視点映像の生成ができています。



図 5：実験結果. 上段：カメラ撮影したフレーム画像. 下段：自由視点フレーム画像.

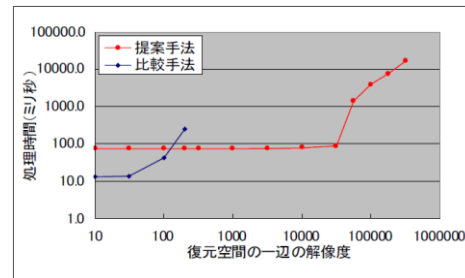


図 6：処理時間

(3-2) ビルボードによる自由視点映像の効率的な表示

実験結果のフレーム画像を図 7 に示す。左図と中図は 2 つのカメラにより撮影した映像中のフレーム画像であり、画像中の青と緑の点は抽出・対応付け・3 次元位置推定を行った特徴点である。右図は、それらのカメラ位置のほぼ中間に位置する自由視点に対して本手法によりビルボードを配置して生成したフレーム画像である。おおまかには自由視点位置に応じた画像が生成されているが、特徴点の分布が疎な部分についてはビルボードの大きさが目立ち、不自然な結果となっている。表 3 は処理時間を示す。リアルタイム表示のためには不十分であり、特に大きな割合を占める SURF 特徴点の抽出と対応付けの効率化が今後の課題である。



図 7：実験結果

表 3：処理時間 (秒/フレーム)

解像度	320×240	640×480
SURF 特徴点抽出	0.034	0.146
特徴点对応付け	0.065	0.270
3 次元位置計算	0.004	0.004
ビルボード生成	0.014	0.057
描画時間	0.002	0.003
合計	0.119	0.484

(4) テクスチャ合成を利用した多視点画像からの高品質な自由視点画像の生成

対象物体（植木鉢）を周囲の等間隔の8方向からカメラで撮影して得られた参照画像を用いて、本手法により自由視点画像を生成する実験を行なった。一つのカメラ（参照カメラ）位置から右隣のカメラ位置に向けて物体を中心に右回りに一定の角度ずつ離れる位置に自由視点を設定して自由視点画像を生成した結果を図8に示す。画像解像度400×400に対して低解像度（256³）のボクセルモデルをレンダリングしただけの低品質な画像に比べ、本手法によるテクスチャ合成を適用した画像では品質の改善がなされた。しかし、特に参照カメラとの角度差が大きいほど、品質は改善された一方で、自由視点から見えるべき実際の画像とは異なる、物体の見え方の妥当性の低い画像が得られることが多かった。この改善が今後の課題である。



図8：参照カメラと自由視点を異なる角度差（左：0°，中央：10°，右：20°）とした実験結果。それぞれ、上図はボクセルモデルをレンダリングした仮の自由視点画像，下図はテクスチャ合成を適用後の画像。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計2件）

- ① 竹中史雄，藤本忠博，原美オサマ，千葉則茂，投影マッピングを利用した多視点カメラ映像からのリアルタイムな仮想視点映像生成法，芸術科学会論文誌，Vol. 10，No. 4，pp. 263-275，2011，査読有
- ② 吉田勝久，藤本忠博，原美オサマ，千葉則茂，複数LDIを用いた半透明点群の効率的な描画法，芸術科学会論文誌，Vol. 2，No. 9，pp. 73-84，2010，査読有

〔学会発表〕（計16件）

- ① 盛内翔太，藤本忠博，背景映像を利用したビデオ映像からの効率的な前景物体抽

出法，グラフィクスとCAD研究会第146回研究発表会，2012.2.8，東京大学（東京都）

- ② 金野龍人，藤本忠博，ビルボードを用いた多視点映像からの効率的な仮想視点映像生成法，グラフィクスとCAD研究会第146回研究発表会，2012.2.8，東京大学（東京都）
- ③ 沢田奈保子，高橋実，藤本忠博，テクスチャ合成を利用した多視点画像からの高品質な自由視点画像生成法，グラフィクスとCAD研究会第146回研究発表会，2012.2.8，東京大学（東京都）
- ④ 樋口拓馬，藤本忠博，可視性マップの逐次更新を利用したブレンスワイプ法，グラフィクスとCAD研究会第146回研究発表会，2012.2.8，東京大学（東京都）
- ⑤ 竹中史雄，藤本忠博，原美オサマ，千葉則茂，テクスチャ投影を利用した多視点カメラ映像からのリアルタイムな形状復元法，第26回NICOGGRAPH論文コンテスト，2010.9.24，アイーナ・いわて県民情報交流センター（岩手県）（審査員特別賞受賞，芸術科学会）
- ⑥ 吉田勝久，藤本忠博，原美オサマ，千葉則茂，複数LDIを利用した半透明点群の効率的な表示法，第25回NICOGGRAPH論文コンテスト，2009.10.24，東京工科大学（東京都）
- ⑦ 高橋実，藤本忠博，原美オサマ，千葉則茂，テクスチャ合成を用いた複数カメラ画像からの高品質ボクセルモデルの構築法，第25回NICOGGRAPH論文コンテスト，2009.10.23，東京工科大学（東京都）
- ⑧ 三國大介，藤本忠博，原美オサマ，千葉則茂，レイベースサンプリングによる半透明物体の形状復元，第25回NICOGGRAPH論文コンテスト，2009.10.23，東京工科大学（東京都）

6. 研究組織

(1) 研究代表者

藤本 忠博 (FUJIMOTO TADAHIRO)

岩手大学・工学部・准教授

研究者番号：00312512