

前背景同時撮影による自然背景に適したビデオベースCG技術に関する研究

代表研究者 藤本 忠博 岩手大学 工学部 准教授

1 研究の背景と目的

コンピュータグラフィックス (CG) 技術は、映画やテレビ、インターネットなどの映像コンテンツ制作において不可欠であり、大容量の高精細な映像情報を高速に配信する高度な技術開発を目指す電気通信分野への関わりは非常に大きい。CGの分野では、近年、人物などの撮影対象を取り囲むように複数のビデオカメラを固定配置して撮影した動画映像 (多視点映像) を利用し、その撮影対象を実際にはカメラが無い自由な視点位置から見た映像 (自由視点映像) を生成するビデオベースCGと呼ばれる技術が活発に研究され、映画産業などでは応用も進んでいる。しかし、テレビやインターネットなどによるリアルタイムの映像配信では、現状、カメラで撮影された動画映像をそのままユーザの画面に提示するだけである。映像配信へのビデオベースCG技術の融合は、ユーザにまるで撮影場所にいるかのような大きな臨場感を与えうる。また、限られた台数のカメラからでも自由視点映像の生成が可能であり、映像配信の効率が大幅に向上する可能性を持つ技術と言える。

ビデオベースCG技術に関しては、その基本をなす技術として、静止画像を対象として多視点画像から自由視点画像を生成するイメージベースCG技術が研究されており、S. E. Chen らの画像補間による手法 (1993年 [1]) を草分けとし、M. Levoy らのライトフィールドレンダリング (1996年 [2]) と S. J. Gortler らのルミグラフィ (1996年 [3]) が以降の研究を活発化させた。2000年代に入ると動画映像を対象としたビデオベースCG技術へと展開され、代表例として、ビジュアルハルの表示 (Matusik 2000 [4], Matusik 2001 [5])、高度なコンピュータビジョン技術を用いた高品質映像生成 (Zitnick 2004 [6])、カメラアレイの開発 (Wilburn 2005 [7])、半透明物体の表示 (Joshi 2006 [8])、焦点平面スイープにより奥行きを推定する手法 (Liu 2006 [9])、デプスハルの表示 (Bogomjakov 2008 [10])、物体輪郭を考慮した高品質映像生成 (Hornung 2009 [11]) などがある。また、多視点画像から3次元モデルを構築する技術では、A. Laurentini が物体輪郭の逆投影によるビジュアルハルの概念を提案 (1994年 [12]) して以降、S. M. Seitz らが色情報によりビジュアルハルを高精度化するフォトハルを構築するボクセルカラーリング法を提案 (1997年 [13]) し、その後、高速高精度化を目指す様々な拡張法が提案された。例えば、M. Li らによるビジュアルハルとフォトハルの構築をGPU上で高速実行する手法 (2003年 [14], 2004年 [15]) を代表として、様々なGPUによる高速化手法が提案されている。また、国内の研究事例としては、名古屋大学谷本研究室の自由視点テレビに関する研究 [16]、京都大学松山研究室のボクセル視体積交差法に関する研究 [17] などが有名である。

ビデオベースCG技術では、撮影対象の自由視点映像を生成するため、それぞれのビデオカメラによる動画映像上で撮影対象 (前景物体) と背景の画素領域を分離して前景物体のみを抽出する必要があり、高精度な抽出手法の実現が重要な課題となっている。前景物体の画素領域を抽出する手法として「背景差分」と呼ばれる手法がよく用いられる。背景差分法は、映像コンテンツ制作や監視システムなど、様々な目的に対して応用が期待される基礎技術として非常に有益であり、これまでに様々な研究が進められてきている。背景差分の基本手法では、背景には動く物体 (背景物体) や輝度変化などが無く、前景物体のみが動きや輝度変化を持つという、背景を不変とした固定背景を前提とする。そして、動画映像の撮影時と同じカメラで同じ位置から背景のみを撮影した背景の静止画像 (背景参照画像) を用いて、動画映像上の前景物体と背景を分離する。具体的には、動画映像中の各フレーム時刻におけるフレーム画像について、各画素が持つ色値 (画素値) を背景参照画像のものと比較することで、画素値が異なる (両者の差分が閾値を超えた) 画素領域が前景物体を映していると判定する。しかし、背景が全く変化しない固定背景シーンは非常に限定的であり、背景中に動く物体や輝度変化が生じる通常的な自然背景シーンに対して上記の基本手法を適用した場合には背景の変化が前景物体として判定されてしまい、前景物体の画素領域だけを良好に抽出することが難しい。そこで、そのような欠点を解決するため、これまでに、動画映像上で背景要素が映る画素に時間の経過に伴う微小な変化 (例えば、揺れる樹木により葉、枝、空が交互に映る、日差しの変化で輝度が変わる等) が生じるような自然背景シーンに対しては、その変化を確率統計的に扱う様々な背景モデルが提案されている。これらのモデルでは、画素に背景が映る確率を確率密度関数で表し、その関数の表現方法によってパラメトリ

ック手法とノンパラメトリック手法に分類される。パラメトリック手法は、確率密度関数を特定の形式の関数で表現し、そのパラメータをサンプルデータから特定する手法であり、ガウス分布モデル (Wren 1997 [18]) や混合ガウス分布モデル (Friedman 1997 [19], Stauffer 1999 [20], KaewTraKulPong 2001 [21], Power 2002 [22], Hayman 2003 [23], Lee 2005 [24], Zivkovic 2006 [25]) などが代表的である。一方、ノンパラメトリック手法は、サンプルデータを利用してカーネル関数の加重和として確率密度関数を表現する手法であり、カーネル関数にガウス分布関数を用いた方法 (Elgammal 2000 [26], Elgammal 2002 [27])、カーネルを可変サイズ化する方法 (Mittal 2004 [28], Zivkovic 2006 [25]) などが提案されている。しかし、これらの確率統計的手法は、背景の微小な変化や緩やかな変化に対応するものであり、例えば、背景内に動く物体が存在するなど、前景物体と同等の大きな変化が背景にあるような場合には、良好な結果が得られない。また、一方で、コンピュータやビデオカメラの性能の向上に伴い、複数のビデオカメラによって動画映像を撮影しながらリアルタイムで自由視点映像を生成するビデオベースCG技術も提案されてきており、その場合、全てのビデオカメラの映像上で同時にリアルタイムで前景物体の画素領域を抽出することを可能とする効率的な背景差分法が要求される。

従来の背景差分法における上記の問題を考慮し、本代表研究者らの研究グループでは、これまでに、自然背景に適したビデオベースCG技術の開発を目的として、前景物体を取り囲むように異なる方向から撮影した複数のカメラ映像上で同時にリアルタイムで前景物体の画素領域を抽出する手法の開発を行ってきた[29]。具体的には、前景物体と背景を含むシーン全体を撮影する従来のカメラ (前景カメラ) に加えて、前景カメラの映像上に映る背景のみを撮影するための別のカメラ (背景カメラ) を前景カメラの前方に設置し、それぞれのカメラで同時に撮影 (前背景同時撮影) を行うことで、前景物体と同等の大きな変化を含む自然背景に対しても前景カメラ映像上で前景物体のみを適切に抽出する背景差分法を開発してきた。この前背景同時撮影による背景差分法の実施環境を図1に示す。図1の実施環境では、1台の前景カメラと2台の背景カメラが設置されており、3次元空間上で前景カメラ映像だけに映る (前景物体が存在する) 領域を「前景空間領域」、前景カメラと背景カメラにともに映る領域を「背景空間領域」と呼ぶこととする。背景カメラの導入により、背景空間領域上に大きな変化 (動く物体の存在など) が生じる場合であっても、前景カメラと背景カメラの映像を比較することで、前景カメラだけに映る前景空間領域上の前景物体を前景カメラ映像上でリアルタイムで抽出することができる。なお、1台の前景カメラに対して複数台の背景カメラを利用することにより、前景カメラ映像上に映る全ての背景部分が少なくともいずれか一つの背景カメラ映像上に映るような状況を実現することで、前景物体を抽出する精度を高めることが可能となる。しかし、実際の様々な実施環境下において、特に、広い空間において、上記の先行研究によって開発済みの手法を用いて実験を行った結果、カメラから遠方にある多様な異なる奥行きを持つ複数の背景物体の影響や、複雑な光環境による明暗の変化などにより、多くの場合に良好な結果が得られないことが分かった。また、リアルタイム処理としては処理速度が不十分であることも分かった。そこで、本研究では、より高い精度でロバストに前景物体を抽出し、かつ、処理速度の向上も図るため、先行研究による手法に対して以下の改善を行うこととした。

- 従来の背景差分法の導入
- カメラパラメータを用いた基礎行列の算出
- シェーダ言語によるプログラム実装

以下、提案手法の詳細について述べる。

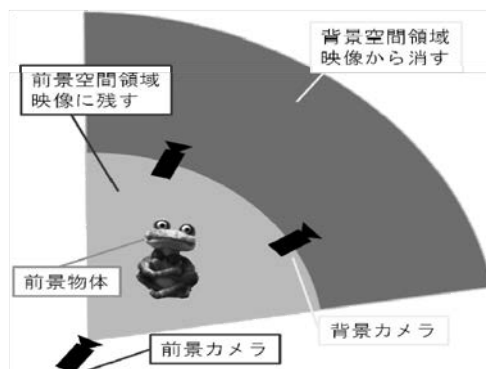


図1 前背景同時撮影による背景差分法の実施環境

2 本研究による提案手法

2-1 提案手法の概要

これ以降、前景カメラ映像中のフレーム時刻ごとのフレーム画像を「前景フレーム画像」、背景カメラ映像中のフレーム時刻ごとのフレーム画像を「背景フレーム画像」と呼ぶこととする。また、前景物体が映る画素を「前景画素」、背景が映る画素を「背景画素」と呼ぶこととする。本研究による提案手法では、前景フレーム画像上で前景画素だけをリアルタイムで抽出するために、前景フレーム画像の画素ごとに以下の処理を行う。

- ① 従来の背景差分法による背景画素の決定 (2-2 節)
- ② 背景フレーム画像上の対応背景画素の探索 (2-3 節)
- ③ 2 値画像の積演算による前景画素と背景画素の決定 (2-4 節)

以下、それぞれの処理の詳細を説明する。

2-2 従来の背景差分法による背景画素の決定

本研究における前背景同時撮影による背景差分法では、次節で述べるように、前景フレーム画像の画素ごとにエピポラ幾何を用いた探索によって背景フレーム画像上の対応背景画素の決定を行う。そして、この探索が膨大な処理時間を要することが先行研究で問題とされていた。そこで、提案手法では、前景画素を抽出するリアルタイム処理の前に、前景カメラによって前景物体が存在せずに背景だけが映る背景参照画像を撮影しておく。そして、フレーム時刻ごとに取得される前景フレーム画像上で前景画素を抽出するリアルタイム処理において、最初の処理として、背景参照画像を用いた従来の背景差分法を適用することで、明らかに背景であると判定される変化のない背景部分が映る背景画素を決定する。一方、それ以外の画素には、前景物体が映る前景画素だけではなく、変化を伴う背景が映る背景画素も含まれる。そこで、この処理で背景画素と判定されなかった画素に対してのみ、次節で述べる背景フレーム画像上での対応背景画素の探索を行う。変化のない背景部分が大部分を占めるようなシーンでは、この処理の時点で多くの画素が背景画素と判定されるため、背景フレーム画像上で探索を行う画素数を大幅に削減することができ、全体的な処理の高速化が実現される。なお、この従来の背景差分法では、前景フレーム画像と背景参照画像がともに前景カメラ、すなわち、同じカメラで同じ位置から撮影されるため、単純に2つの画像上の同じ位置の画素どうしを比較すればよく、対応する2つの画素の画素値の差分を事前に決めた閾値と比較することで背景画素であるか否かを決定する。このとき、時間の経過による照明の変化等に起因する画素値の明るさの違いの影響を緩和するため、画素値としてHSV表色系を用い、色相 (Hue) と彩度 (Saturation) に対して明度 (Value) の重みを小さくして差分を計算することとした。

2-3 背景フレーム画像上の対応背景画素の探索

前背景同時撮影による背景差分法では、前景物体と背景がともに映る前景フレーム画像と背景だけが映る背景フレーム画像の間で画素値の比較を行い、前景フレーム画像上の各画素が前景画素か背景画素かの判定を行う。このとき、図1に示すように前景カメラと背景カメラの設置される位置が異なることから、背景内の同じ部分が映っている画素の位置が前景フレーム画像と背景フレーム画像の間で互いに異なってしまうという、従来の背景差分法では考慮する必要の無い問題が生じる。そこで、前景フレーム画像上の各画素に対して、背景内の同じ部分が映っている背景フレーム画像上の背景画素 (対応背景画素) を求めることが必要となる。この対応背景画素の決定を効率的に行うためにエピポラ幾何を利用する。ここで、1台の前景カメラに対して N 台の背景カメラがあるとする。このとき、前景フレーム画像 I_f の各画素 $P_{f,m}$, $m = 0, \dots, M-1$, に対して背景フレーム画像 $I_{b,n}$, $n = 0, \dots, N-1$, 上のエピポラ線 $L_{b,n,m}$ を求め、画素 $P_{f,m}$ に対応する背景画素 $P_{b,n,m}$ をエピポラ線 $L_{b,n,m}$ 上で探索する。具体的には、画素 $P_{f,m}$ とエピポラ線 $L_{b,n,m}$ 上の全ての画素を比較し、画素値が類似する画素があれば、その画素を対応背景画素 $P_{b,n,m}$ とし、画素 $P_{f,m}$ を背景が映っている背景画素であると判定する。逆に、類似する画素が無い場合には、画素 $P_{f,m}$ を前景物体が映っている前景画素であると判定する。なお、画素値の類似度の計算にはHSV表色系による前節と同様の方法を用いた。

3次元空間上の同じ点を映している前景フレーム画像 I_f 上の画素 P_f の座標 \mathbf{p}_f と背景フレーム画像 I_b 上の画素 P_b の座標 \mathbf{p}_b は、それらの同次座標を用いて、基礎行列 \mathbf{F} による次式の関係を満たす。

$$\tilde{\mathbf{p}}_f = \begin{bmatrix} x_f \\ y_f \\ 1 \end{bmatrix}, \quad \tilde{\mathbf{p}}_b = \begin{bmatrix} x_b \\ y_b \\ 1 \end{bmatrix} \quad (1)$$

$$\mathbf{F} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}, \quad \tilde{\mathbf{p}}_f^T \mathbf{F} \tilde{\mathbf{p}}_b = 0 \quad (2)$$

基礎行列 \mathbf{F} は、前景カメラと背景カメラのカメラパラメータ（内部パラメータと外部パラメータ）によって決定される。基礎行列 \mathbf{F} を用いることで、前景フレーム画像上の座標 $\mathbf{p}_{f,m}$ の画素 $P_{f,m}$ に対応する背景フレーム画像上のエピポーラ線 $L_{b,m}$ は、背景フレーム画像上の画素 P_b の座標 \mathbf{p}_b に関して次式で表される。

$$L_{b,m} : \mathbf{n}_m^T \tilde{\mathbf{p}}_b = a_m x_b + b_m y_b + c_m = 0$$

$$\tilde{\mathbf{p}}_{f,m} = \begin{bmatrix} x_{f,m} \\ y_{f,m} \\ 1 \end{bmatrix}, \quad \mathbf{n}_m = (\tilde{\mathbf{p}}_{f,m}^T \mathbf{F})^T = \mathbf{F}^T \tilde{\mathbf{p}}_{f,m} = \begin{bmatrix} a_m \\ b_m \\ c_m \end{bmatrix} \quad (3)$$

また、特に、背景が平面であると仮定した場合、前景フレーム画像と背景フレーム画像で背景内の同じ点、言い換えれば、背景平面上に位置する同じ点を映している画素の座標 \mathbf{p}_f と \mathbf{p}_b は、ホモグラフィ行列 \mathbf{H} によって次式で対応付けられる。

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}, \quad \tilde{\mathbf{p}}_b = \mathbf{H} \tilde{\mathbf{p}}_f \quad (4)$$

ホモグラフィ行列 \mathbf{H} は、前景カメラと背景カメラのカメラパラメータ、ならびに、前景カメラと背景カメラに対する背景平面の位置と向きによって決定される。理論上、式(4)で対応付けられる座標 \mathbf{p}_f と \mathbf{p}_b は式(2)の関係も満たす。すなわち、前景フレーム画像上の座標 $\mathbf{p}_f = \mathbf{p}_{f,m}$ に対して式(4)から得られる背景フレーム画像上の座標 \mathbf{p}_b は式(3)を満たす。また、上述した背景フレーム画像のエピポーラ線上の対応背景画素の探索は、前景フレーム画像の画素 $P_{f,m}$ に対応する背景画素が背景フレーム画像上にあるかどうかを調べることが目的である。よって、実際に背景を十分に近似するような平面を仮定することができる場合、その平面に関するホモグラフィ行列を用いて前景フレーム画像上の画素 $P_{f,m}$ に対応する背景フレーム画像上の画素 $P_{b,H,m}$ を求め、エピポーラ線 $L_{b,m}$ 上で画素 $P_{b,H,m}$ の付近の数画素だけを探索することで、画素 $P_{f,m}$ が背景画素であるかどうかを効率的に調べることができる。なお、実際には背景は完全な平面ではないため、画素 $P_{b,H,m}$ がエピポーラ線 $L_{b,m}$ 上に正確に位置する保証はない。そこで、画素 $P_{b,H,m}$ に最も近いエピポーラ線 $L_{b,m}$ 上の画素を起点として、その付近を探索することとする。

先行研究では、前景画素を抽出するリアルタイム処理の前に、前景カメラと背景カメラのフレーム画像上で SURF 特徴量[30]を用いて対応付けた特徴点群から基礎行列 \mathbf{F} を求めた。この方法では、まず、前景カメラ映像と背景カメラ映像のそれぞれについて、同じフレーム時刻の適当なフレーム画像上で SURF 特徴量による特徴点の抽出を行う。次に、互いのフレーム画像上の特徴点どうしの SURF 特徴量を比較することで、共通する特徴点を対応付ける。そして、対応特徴点の組を制約条件とすることで最適な基礎行列を推定する。なお、上記の適当なフレーム画像としては、任意に選択した幾つかのフレーム時刻におけるフレーム画像を用いて基礎行列の推定を行い、最終的に最も精度の高いものを採用した。また、SURF 特徴量による特徴点の抽出、ならびに、対応特徴点群を制約条件とした基礎行列の推定を行うプログラムの実装には、OpenCV ライブラリを利用した。しかし、この方法では、基礎行列の推定精度が特徴点の対応付けの正確さに依存してしまい、特に、照明条件が良好ではない環境下において抽出される特徴点の個数が少ない場合には、安定した基礎行列が得られないという問題が生じた。そこで、提案手法では、安定した基礎行列を得るために、前景カメラと背景カメラに対するキャリブレーション[31]を行い、得られたカメラパラメータから基礎行列の算出を行うこととした。前景フレーム画像と背景フレーム画像の座標 \mathbf{p}_f と \mathbf{p}_b に対して、カメラパラメータを用いた基礎行列 \mathbf{F} の算出は次式のようになる。

$$\mathbf{F} = \mathbf{K}_b^{-T} \mathbf{E} \mathbf{K}_f^{-1}, \quad \mathbf{E} = [\mathbf{t}]_{\times} \mathbf{R}, \quad \tilde{\mathbf{p}}_f = \mathbf{R} \tilde{\mathbf{p}}_b + \mathbf{t} \quad (5)$$

式(5)において、 3×3 行列 \mathbf{K}_f と \mathbf{K}_b は前景カメラと背景カメラの内部パラメータ行列、 3×3 行列 \mathbf{R} と 3次元ベクトル \mathbf{t} は前景カメラと背景カメラの外部パラメータから求められるカメラ間の回転行列と並進ベクトルである。また、 3×3 行列 \mathbf{E} は基本行列と呼ばれ、並進ベクトル \mathbf{t} の外積を表す 3×3 行列 $[\mathbf{t}]_{\times}$ と回転行列 \mathbf{R} の積で定義され、以下のように外積で表される。

$$E \tilde{p}_f = ([t], R) \tilde{p}_f = t \times (R \tilde{p}_f) \quad (6)$$

式(5)によって算出した基礎行列 F を用いることで、照明条件等に依存せず、安定して精度の高いエピポーラ線を得ることが可能となった。なお、カメラキャリブレーションのプログラム実装にも OpenCV ライブラリを利用した。

また、背景フレーム画像のエピポーラ線上で探索の起点とする画素を得るための画素 $P_{b,h,m}$ を求めるホモグラフィ行列を推定する方法には、先行研究と同様に、SURF 特徴量によって求めた前景フレーム画像と背景フレーム画像の対応特徴点群を用いる方法を採用した。このプログラム実装にも OpenCV ライブラリを利用した。

2-4 2 値画像の積演算による前景画素と背景画素の決定

前節で述べた方法により、 N 個の背景フレーム画像 $I_{b,n}$, $n=0, \dots, N-1$, のそれぞれに対して、前景フレーム画像 I_f の各画素 $P_{f,m}$ が前景画素であるか背景画素であるかが決定される。ここで、背景に大きな変化を含む自然背景シーンでは、実際には背景が映っている背景画素を前景画素として誤判定してしまう傾向が大きいと考えられる。実際、実験によって、その傾向が確認された。そこで、 N 個の背景フレーム画像の全てについて前景画素であると判定された画素 $P_{f,m}$ だけを最終的に前景画素と判定することとした。これを行うため、各背景フレーム画像 $I_{b,n}$ に対して、画素 $P_{f,m}$ の判定結果が前景画素であれば 1、背景画素であれば 0 という 2 値を与えた 2 値画像 B_n を生成し、 N 個の 2 値画像 B_n の画素ごとの積演算を行うことで、最終的に画素 $P_{f,m}$ が前景画素か背景画素かの決定を行うこととした。

2-5 高速化のためのプログラム実装

高速な処理を実現するため、GPU (Graphics Processing Unit) 上で高速に動作するプログラムの開発を目指し、シェーダ言語 GLSL による実装を行った。具体的には、フラグメントシェーダを利用することで、背景フレーム画像におけるエピポーラ線上の画素探索を高速に実行するように実装を行った。

3 実験

提案手法の有効性を検証するため、図 1 の実施環境に示すような 1 台の前景カメラと 2 台の背景カメラを用いた実験を行った。その結果について、以下に述べる。

3-1 基礎行列の精度に関する実験

基礎行列の求め方として、対応特徴点を制約条件とした推定によって求める先行研究の方法、ならびに、キャリブレーションにより得られたカメラパラメータを用いて算出する提案手法の方法を採用し、それぞれの方法によって求められた基礎行列の精度を比較する実験を行った。この実験では、前景カメラと背景カメラに対して図 1 の背景空間領域内に存在する背景物体としてぬいぐるみを提示した。そして、前景フレーム画像上でぬいぐるみが映っている一つの画素 $P_{f,m}$ に対応する背景フレーム画像上のエピポーラ線をそれぞれの方法で求めた基礎行列から計算し、そのエピポーラ線上に画素 $P_{f,m}$ と同じ部分が映っているかを調べることでエピポーラ線の精度の比較を行った。図 2, 3 に実験結果のフレーム画像を示す。それぞれの図において、上段はぬいぐるみがカメラに近い場合、下段は遠い場合の結果である。そして、それぞれの場合において、左図は前景フレーム画像、右図は背景フレーム画像であり、左図中の緑色の点が画素 $P_{f,m}$ 、右図中の赤色の線が対応するエピポーラ線を示す。また、右図中の緑色の点は、エピポーラ線上の探索の起点を得るためにホモグラフィ行列によって求められた画素 $P_{b,h,m}$ である。なお、背景フレーム画像については、実際に用いた 2 台の背景カメラのうちの一つのものを示している。先行研究の方法による図 2 では、ぬいぐるみがカメラから遠い場合 (下段) では画素 $P_{f,m}$ に映るぬいぐるみの顔部分の付近をエピポーラ線が通過しているのに対して、ぬいぐるみがカメラに近い場合 (上段) では離れたところを通過しており、エピポーラ線の精度が低いことが分かる。一方、提案手法の方法による図 3 では、いずれの場合でも画素 $P_{f,m}$ に映る顔部分の付近をエピポーラ線が通過している。この結果から、先行研究の方法に比べて、提案手法の方法によって求めたエピポーラ線の精度が高く、良好な基礎行列が得られていることが分かった。

3-2 前景画素の抽出精度に関する実験

前景フレーム画像上の前景画素の抽出精度を比較するため、図 1 の前景空間領域に前景物体としてぬいぐるみを配置して実験を行った。先行研究の手法を用いて前景画素を抽出した結果を図 4、提案手法を用いた結果を図 5 に示す。それぞれの図において、左上図は前景カメラによる前景フレーム画像、左下図と右上図は 2 台の背景カメラによる背景フレーム画像、右下図が前景フレーム画像上で前景画素を抽出した結果である。図 4 では、ほとんどの前景画素が抽出されてはいるが、背景空間領域内に位置する青色のボードと人物

の緑色の洋服が映る多くの背景画素が前景画素として誤抽出されてしまっている。一方、図5では、図4に比べて、前景画素として誤抽出される背景画素が大幅に減少しており、また、前景画素の抽出精度も向上している。これは、提案手法における従来の背景差分法の導入と基礎行列の精度の向上が原因であると考えられる。この結果から、先行研究の手法に比べて、提案手法によって高い精度で前景画素の抽出を行うことが可能となったことが確認された。

3-3 高速化のためのプログラム実装

フラグメントシェーダを利用したプログラムの実装により、前景フレーム画像上の画素ごとに背景フレーム画像のエピポーラ線上で画素探索を行うという膨大な処理をGPU上で効率的に実行することが可能となった。しかし、現状では1秒あたり数フレームを処理する程度であり、リアルタイム処理には不十分な処理速度であることが分かった。

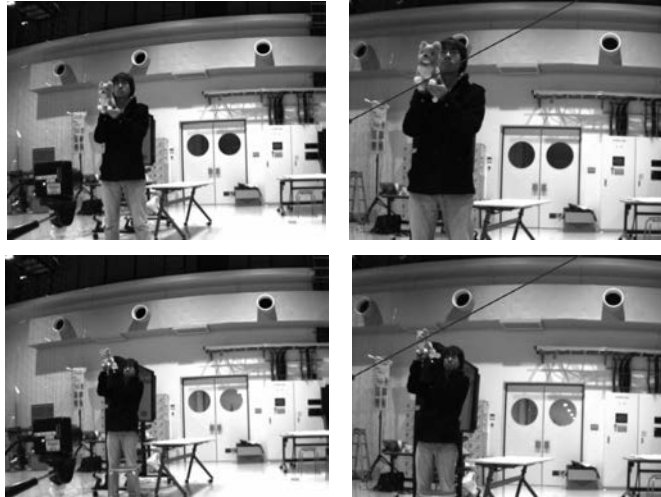


図2 先行研究の方法で求めたエピポーラ線

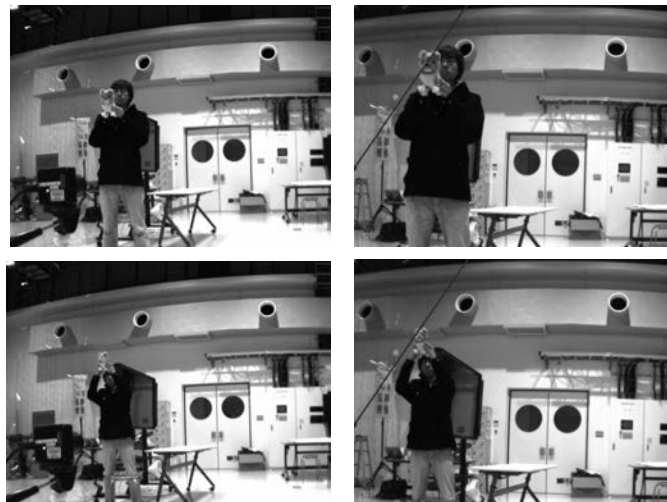


図3 提案手法の方法で求めたエピポーラ線



図4 先行研究の手法による前景画素の抽出



図5 提案手法による前景画素の抽出

4 結論と今後の課題

本研究では、前景カメラと背景カメラを用いた前背景同時撮影によって、背景に大きな変化を含む自然背景シーンに対してロバストに前景物体の画素領域を抽出する手法を提案した。提案手法では、先行研究による手法に対して、従来の背景差分法の導入、ならびに、キャリブレーションで得られたカメラパラメータを用いた基礎行列の算出による改善を行うことで、より高精度な前景画素の抽出を実現した。一方、シェーダ言語を用いたプログラム実装については、処理速度の向上は見られたものの、現状ではリアルタイム処理としては不十分な結果であった。

今後の課題としては、まず、シェーダ言語によるプログラム実装を改善し、より高速な GPU 上での処理を実現することが挙げられる。また、前景画素の抽出精度をさらに向上させるため、画素値の類似度の評価方法を検討することも挙げられる。これに関しては、現状では明度の違いを考慮できる HSV 表色系で画素値を比較しているが、他の表色系の利用も含めた方法の検討が必要である。また、その他の様々な観点からも改善策を検討し、前景画素の抽出精度と処理速度の向上を実現していく予定である。

【参考文献】

- [1] S. E. Chen, L. Williams, View Interpolation for Image Synthesis, Proc. of SIGGRAPH 1993, pp.279–288, 1993.
- [2] M. Levoy, P. Hanrahan, Light Field Rendering, Proc. of SIGGRAPH 1996, pp.31–42, 1996.

- [3] S. J. Gortler, R. Grzeszczuk, R. Szeliski, M. F. Cohen, The Lumigraph, Proc. of SIGGRAPH 1996, pp.43–54, 1996.
- [4] W. Matusik, C. Buehler, R. Raskar, S. Gortler, L. McMillan, Image Based Visual Hulls, Proc. of SIGGRAPH 2000, pp.369–374, 2000.
- [5] W. Matusik, C. Buehler, L. McMillan, Polyhedral Visual Hulls for Real-Time Rendering, Proc. of Eurographics Workshop on Rendering 2001, pp.115–125, 2001.
- [6] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, High-quality Video View Interpolation using a Layered Representation, Proc. of SIGGRAPH 2004, pp.600–608, 2004.
- [7] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, M. Levoy, High Performance Imaging Using Large Camera Arrays, Proc. of SIGGRAPH 2005, pp.765–776, 2005.
- [8] N. Joshi, W. Matusik, S. Avidan, Natural Video Matting using Camera Arrays, Proc. of SIGGRAPH 2006, pp.779–786, 2006.
- [9] Y. Liu, G. Chen, N. Max, C. Hofsetz, P. McGuinness, Undersampled Light Field Rendering by a Plane Sweep, Computer Graphics Forum, Vol.25, No.2, pp.225–236, 2006.
- [10] A. Bogomjakov, C. Gotsman, Reduced Depth and Visual Hulls of Complex 3D Scenes, Computer Graphics Forum, Vol.27, No.2, pp.175–182, 2008.
- [11] A. Hornung, L. Kobbelt, Interactive Pixel-Accurate Free Viewpoint Rendering from Images with Silhouette Aware Sampling, Computer Graphics Forum, Vol.28, No.8, pp.2090–2103, 2009.
- [12] A. Laurentini, The Visual Hull Concept for Silhouette-based Image Understanding, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.16, No.2, pp.150–162, 1994.
- [13] S. M. Seitz, C. R. Dyer, Photorealistic Scene Reconstruction by Voxel Coloring, Computer Vision and Pattern Recognition Conf., pp.1067–1073, 1997.
- [14] M. Li, M. Magnor, H.-P. Seidel, Hardware-Accelerated Visual Hull Reconstruction and Rendering, Proc. of Graphics Interface 2003, pp.65–71, 2003.
- [15] M. Li, M. Magnor, H.-P. Seidel, Hardware-Accelerated Rendering of Photo Hulls, Computer Graphics Forum, Vol.23, No.3, pp.635–642, 2004.
- [16] 谷本正幸, 藤井俊彰, 自由視点映像技術, 映像情報メディア学会誌, Vol. 60, No. 1, pp. 29–34, 2006.
- [17] T. Matsuyama, X. Wu, T. Takai, 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.
- [18] C. R. Wren, A. Azarbayejani, T. Darrell, A. P. Pentland, Pfinder: Real-Time Tracking of the Human Body, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.19, No.7, pp.780–785, 1997.
- [19] N. Friedman, S. Russell, Image Segmentation in Video Sequences: A Probabilistic Approach, Proc. of Conf. on Uncertainty in Artificial Intelligence, pp.175–181, 1997.
- [20] C. Stauffer, W. E. L. Grimson, Adaptive Background Mixture Models for Real-Time Tracking, Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition, Vol.2, pp.246–252, 1999.
- [21] P. KaewTraKulPong, R. Bowden, An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection, Proc. of European Workshop Advanced Video Based Surveillance Systems, 2001.
- [22] P. W. Power, J. A. Schoonees, Understanding Background Mixture Models for Foreground Segmentation, Proc. of the Image and Vision Computing, 2002.
- [23] E. Hayman, J.-O. Eklundh, Statistical Background Subtraction for a Mobile Observer, Proc. of the 9th IEEE International Conf. on Computer Vision, pp.67–74, 2003.
- [24] D.-S. Lee, Effective Gaussian Mixture Learning for Video Background Subtraction, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.27, No.5, pp.827–832, 2005.
- [25] Z. Zivkovic, F. v. d. Heijden, Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction, Journal of Pattern Recognition, Vol.27, No.7, pp.773–780, 2006.
- [26] A. Elgammal, D. Harwood, L. S. Davis, Non-parametric Model for Background Subtraction, Proc. of the European Conf. on Computer Vision 2000, pp.751–767, 2000.

- [27] A. Elgammal, R. Duraiswami, D. Harwood, L. S. Davis, Background and Foreground Modeling using Nonparametric Kernel Density Estimation for Visual Surveillance, Proc. of the IEEE, Vol.90, No.7, pp.1151–1163, 2002.
- [28] A. Mittal, N. Paragios, Motion-Based Background Subtraction using Adaptive Kernel Density Estimation, Proc. of the Conf. on Computer Vision and Pattern Recognition 2004, pp.302–309, 2004.
- [29] 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, 第146回グラフィクスとCAD研究会, 2012.
- [30] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, SURF: Speeded Up Robust Features, Computer Vision and Image Understanding, Vol.110, No.3, pp.346–359, 2008.
- [31] Z. Zhang, A Flexible New Technique for Camera Calibration, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.22, No.11, pp.1330–1334, 2000.

〈 発 表 資 料 〉

題 名	掲載誌・学会名等	発表年月
背景変化を伴うカメラ映像からの前景物体の抽出 (目黒拓樹, 藤本忠博)	平成25年度第3回芸術科学会東北支部研究会	2013年12月7日
複数カメラを用いた背景変化にロバストな前景物体の抽出法 (目黒拓樹, 藤本忠博)	映像情報メディア学会メディア工学研究会	2014年2月22日