Vol.2012-CG-146 No.20
2012/2/8

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

盛内翔太† 藤本忠博††

本研究では，複数のビデオカメラを用いることで，背景に動く物体が含まれる場合にも，ビデオ映像から前景物体のみを効率的に抽出する手法を提案する．本手法では，背景のみを撮影する 3〜4 台の背景カメラ，ならびに，背景を含めて前景物体を撮影する 1 台の前景カメラを使用する．各カメラによる背景映像と前景映像を関連付けるホモグラフィ行列と基礎行列を用いて，背景映像との背景差分により効率的に前景映像中の前景物体を抽出する．

Efficient Foreground Object Detection in Video Using Background Video

Syouta Moriuchi† and Tadahiro Fujimoto††

In this paper, we propose a method to efficiently detect foreground objects in a video. This method copes with the case in which the background has moving objects using multiple cameras. Two types of cameras are used: a few background cameras capture only the background, and one foreground camera captures foreground objects as well as the background. Using homography matrices and fundamental matrices relating background videos and a foreground video, foreground objects in the foreground video can be efficiently detected by background subtraction with the background videos.

1. はじめに

背景差分は，ビデオ映像中で背景から前景物体のみを分離して抽出する映像技術であり，映像コンテンツの制作や監視システムなど，様々な応用分野に用いられる．最も簡単な背景差分の方法としては，背景のみが映る 1 枚の背景画像をあらかじめ撮影しておき，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が映る画素領域のみを抽出する方法がよく用いられる．しかし，この方法は，背景が常に変化しないことが前提であり，背景の時刻に伴う環境変化（明るさの変化など）や背景に動的な物体がある場合などには，それらの画素領域も抽出されてしまう問題がある．そこで，そのような問題を解決するため，画素ごとに映る背景の変化を確率統計的に扱う背景モデルなど，これまでに多くの手法が提案されてきた[1-7]．しかし，その多くは，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，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，前景物体のみを抽出することは難しい．一方，物体を異なるカメラ位置から取り囲むように撮影した複数のビデオ映像から物体の 3 次元形状を復元するイメージベース CG の分野では，各ビデオ映像上で物体を抽出することが必要となる場合が多い[8-13]．

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，それぞれ，前景物体の画素領域（前景領域）を抽出した映像を得ることを最終目標とする．そして，背景に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，その方法として，複数の前景カメラと背景カメラを用いた方法を採用する．図 1 に示すように，前景物体の全周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，各背景カメラで撮影される背景のみが映る背景映像と各前景カメラで撮影される前景物体と背景がともに映る前景映像の間で，適切な画像座標の変換や背景差分等を行うことで，各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，3〜4 台ほどの背景カメラと 1 台の前景カメラを用い，前景映像から前景領域を抽出する手法を提案する．図 2 に本研究におけるカメラの配置を示す．図 2 のカメラ配置は図 1 の一部を構成しており，本研究の目的は，図 2 中の前景カメラと背景カメラの間の領域に存在する物体を前景物体として抽出することである．

† 岩手大学大学院工学研究科
Graduate School of Engineering, Iwate University
†† 岩手大学工学部
Faculty of Engineering, Iwate University

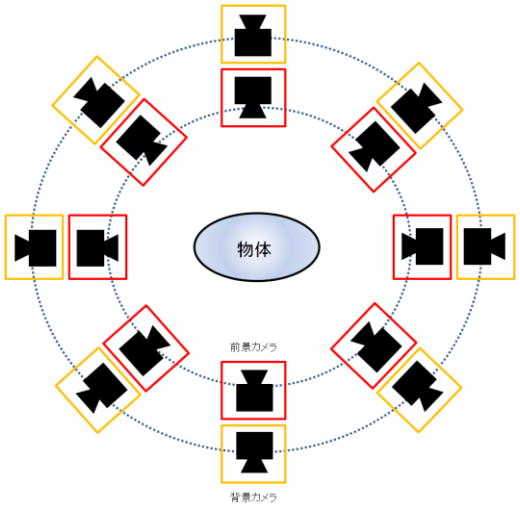

図 1　物体の全周を取り囲む前景カメラと背景カメラの配置

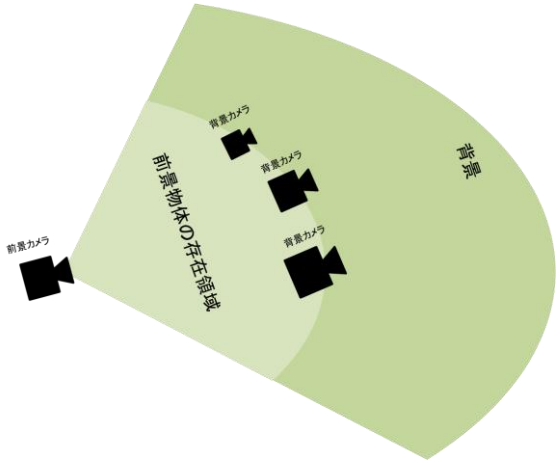

図 2　本研究における前景カメラと背景カメラの配置

以下，2 節で異なるビデオ映像の画素間の対応付けの方法，3 節で前景物体領域の抽出の方法について述べる．4 節で実験結果を示し，5 節でまとめと今後の課題について述べる．なお，これ以降の文中では，適宜，「カメラ映像」という用語を「ビデオ映像」と同義で用いる．

2.　異なるカメラ映像の画素間の対応付け

本手法では，異なるカメラ視点から撮影された前景映像と複数の背景映像の画素どうしを対応付ける必要がある．そこで，それを実現するための方法について述べる．

2.1　特徴点の抽出とマッチング

異なるカメラ映像内で 3 次元空間上の同じ部分を映す画素間の対応関係を得るために，まず，各映像内の特徴点の抽出とマッチングを行う．特徴点の抽出法には SURF（Speeded-Up Robust Features）を用いる[14]．SURF とは，画像から局所特徴量を求めて特徴点を抽出する方法であり，スケール変化，平行移動，回転，環境変化に対し頑健であり，リアルタイム性に優れる．異なる画像内で抽出された特徴点どうしのマッチングには，抽出した特徴点の局所特徴量に含まれる 128 次元からなる特徴ベクトルの比較を行う．2 枚の画像間で，各画像内で得られた特徴点どうしを特徴ベクトルのユークリッド距離で比較し，閾値により誤対応点を除外し，最も距離の近い特徴点どうしを対応点として決定する．図 3 は 2 枚の画像の対応点を黄色斜線で結んだマッチングの例である．

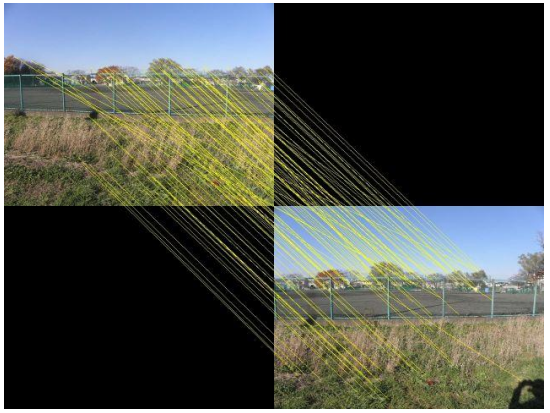

図 3　特徴点のマッチング
上段画像の特徴点数：1276
下段画像の特徴点数：1190
マッチング数：206

2.2 ホモグラフィ行列の算出

一般に，カメラ映像内に映る全ての物体が一つの平面上にあると仮定すれば，ホモグラフィ行列により，異なるカメラ映像の画素間の対応関係を容易に得ることが可能となる．図 4 のように，ある平面上の任意の点 **P** が投影される 2 つのカメラ映像上の画素座標を同次座標 $\tilde{p}_1$，$\tilde{p}_2$ とすると，式(1)の関係が成り立つ．

$$\tilde{p}_2 = H\,\tilde{p}_1 \qquad (1)$$

$$\tilde{p}_1 = [x_1, y_1, 1]^T, \quad \tilde{p}_2 = [x_2, y_2, 1]^T,$$

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

式(1)中の H をホモグラフィ行列と呼ぶ．ホモグラフィ行列は，特徴点のマッチングで得た対応点の組から算出することができ，最低で 4 組の対応点から算出可能である．

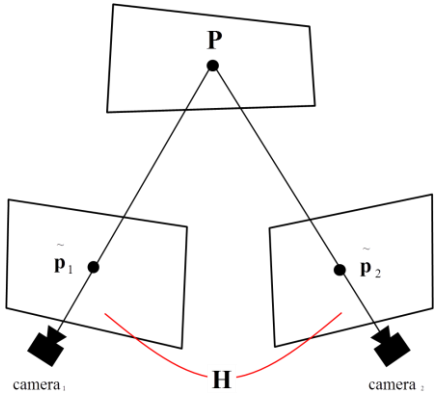

図 4　ホモグラフィ行列による異なるカメラ映像の画素間の対応関係

2.3 基礎行列の算出

ホモグラフィ行列は，画素の探索を行うことなく，映像間の画素ごとの対応関係を一度の計算で効率的に求めることができる．しかし，映像内に映る全ての物体の点が同一の平面上にあることを前提とするため，同一平面上に限定されない奥行きのある点には対応しない．その場合，一般には，基礎行列によるエピポーラ幾何を用いることが多い．これにより，一方のカメラ映像上の画素位置に対応する他方のカメラ映像上の画素位置の範囲をエピポーラ拘束によりエピポーラ線上に限定することが可能となる．図 5 にエピポーラ幾何による 2 つのカメラ映像間の画素位置の関係を示す．このとき，3 次元空間上の点 **P** が投影される各カメラ映像上の画素座標 $\tilde{p}_1$，$\tilde{p}_2$ の関係は式(2)で表せる．

$$\tilde{p}_1^{\,T} F\,\tilde{p}_2 = 0 \qquad (2)$$

$$\tilde{p}_1 = [x_1, y_1, 1]^T, \quad \tilde{p}_2 = [x_2, y_2, 1]^T,$$

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$$

式(2)中の F を基礎行列と呼ぶ．基礎行列も特徴点のマッチングで得た対応点の組から算出することができる．

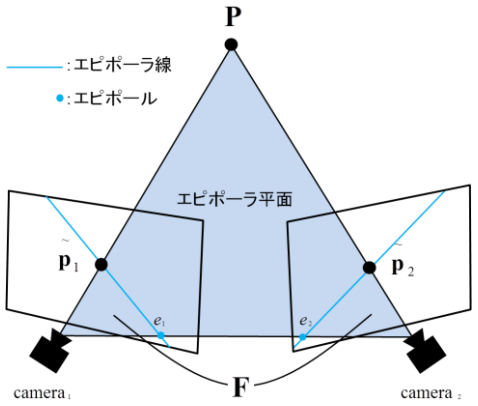

図 5　基礎行列によるエピポーラ幾何を用いた異なるカメラ映像の画素間の対応関係

カメラが固定されていれば基礎行列Fは不変であり，式(2)の $\tilde{p}_1$， $\tilde{p}_2$ の関係は映像内に映る物体の 3 次元形状に関わらずに成り立つ．よって，全ての物体の点が同一平面上にあることを前提とする式(1)の $\tilde{p}_1$， $\tilde{p}_2$ は式(2)を満たす．これは，2 つの映像間でホモグラフィ行列によって対応付けられる画素どうしは互いのエピポーラ線上にのることを意味する．

3. 前景物体領域の抽出

ホモグラフィ行列と基礎行列により，前景カメラと各背景カメラの映像間の画素の対応関係が求まる．これにより，前景映像と背景映像の背景差分により前景物体領域の抽出が可能となる．本研究では，比較のために，前景映像から前景物体領域を抽出する方法として次の 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（3.1 節）

[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方法（3.2 節）

以下，それぞれの方法について述べる．

3.1 方法 1：背景合成映像との背景差分による方法

前景物体に比べて背景が十分に遠方にあるものとすれば，背景の全ての物体が一つの平面上にあるものと仮定することができる．その場合，各背景映像間のホモグラフィ行列を求め，複数の背景映像を一枚の背景合成映像に合成することが可能となる[15]．ホモグラフィ行列は背景映像内の特徴点の抽出とマッチングの結果から求めることができる（2.2 節）．

基準とする背景映像（基準背景映像）を src_0 とし，隣接する背景映像を順に src_1，src_2，src_3，... src_{N-1} とする．また，隣接する 2 つの背景映像 src_i， src_j の間のホモグラフィ行列を H_{ij} とする．このとき，背景映像 $src_k\,(k=1,...,N-1)$ の画像座標を基準背景映像 src_0 の画像座標に変換するホモグラフィ行列 H_{0k} を式(3)により得ることができる．

$$H_{0k} = H_{01}H_{12}...H_{k-1k} \qquad (3)$$

これを用いて，基準背景映像 src_0 に他の全ての背景映像 src_k を合成することができる．

式(3)の行列 H_{0k} により背景映像 src_k を変換した映像を dst_k とする．この変換後の背景映像を基準背景映像 src_0 の画像座標系上で重ね合わせることで背景合成映像を生成する場合，背景映像どうしが重なり合う領域で輝度値の調整が必要となる．本手法では，背景映像どうしが水平方向に隣接しているものとし，例えば，基準背景映像 src_0 に背景映像 dst_1 を合成する場合，図 6 に示すように，重なり合う領域で境目からの水平方向の距離に応じて輝度値を線形に調整する．図 6 赤枠で示す重なり合う領域内の点 d における合成後の輝度値 $cmp_{01}(d)$ は，式(4)に示すように，元の 2 つの映像 src_0 と dst_1 の点 d における輝度値 $src_0(d)$ と $dst_1(d)$ に対して，図 6 中の距離 d_0 と d_1 を用いた重み付けによる α ブレンディングで求められる．

$$cmp_{01}(d) = \frac{d_1}{d_0+d_1} src_0(d) + \frac{d_0}{d_0+d_1} dst_1(d) \qquad (4)$$

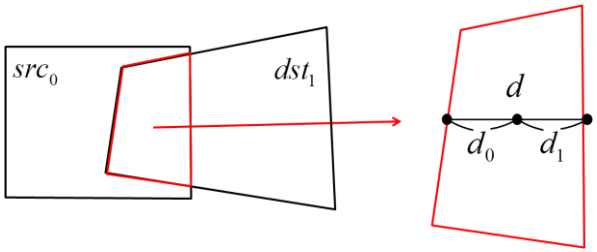

図 6　2 つの映像が重なり合う領域内の輝度値の計算

背景映像を合成後，背景合成映像を用いて背景差分により前景映像から前景物体を抽出するために，前景映像の各画素に対応する背景合成映像の画素を求める．背景の全ての物体が同一平面上にあるものと仮定すれば，背景映像の合成のときと同様，前景映像と背景合成映像の間のホモグラフィ行列を求めて画素間の対応関係を得ることができる．前景映像には前景物体が含まれるが，ホモグラフィ行列を求める際の特徴点のマッチングでは両映像中に共通に含まれる背景の特徴点のみが対応付けられるため，背景のみを考慮したホモグラフィ行列が得られる．前景映像を src_{fore}，背景合成映像を src_{back}，両映像間のホモグラフィ行列を H_{fb} とし，src_{back} を H_{fb} により座標変換した映像を dst_{back} とする．すると， src_{fore} と dst_{back} の背景差分により，3 次元空間上で前景カメラと背景カメラの間の領域に存在する前景物体を前景映像から抽出することが可能となる．背景差分における輝度値の差分には，明るさの変動の影響を考慮できる HSV 表色系を用い，対応画素間の H，S，V 値の差分の絶対値に適切な重み付けを行い，その合計値を評価値として閾値と比較することで前景画素か背景画素かの判定を行う．また，背景の全ての物体が同一平面上にあるという仮定は現実には満たされないため，実際には，ホモグラフィ行列による画素の対応は不確実である．そこで，前景映像 src_{fore}

の各画素に対応する（変換後の）背景合成映像 dst_{back} 内の画素について，周囲の 8 近傍まで参照し，合計 9 つの画素に対して上記の評価値を求め，最小の評価値となる参照画素を対応画素として扱う（図 7）．

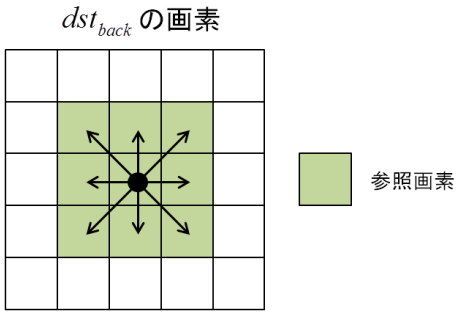

図 7　方法 1 における背景合成映像内の参照画素

3.2　方法 2：背景映像ごとの背景差分による方法

3.1 節の方法は，背景の全ての物体が同一平面上にあることを前提としているため，前景物体抽出の精度が不十分である．また，背景合成映像の生成では，元の複数の背景映像を基準背景映像の画像座標系へ座標変換する処理や，α ブレンディングによる輝度値の合成を行うため，その処理に時間を要する．そこで，これらの問題を改善するため，以下の方法を提案する．

この方法では，背景映像を合成せず，前景映像に対して各背景映像を個別に対応させて背景差分を行い，前景映像の画素ごとに 2 値化したマスク映像（前景画素：1，背景画素：0）を生成する．その後，全ての背景映像に関するマスク映像を論理積で統合した合成マスク映像を求め，前景映像との論理積を取ることで前景物体を抽出した映像を得る．

背景映像ごとの前景映像との背景差分における両映像間の画素の対応付けについては，背景の全ての物体が同一平面上にあるという制約を無くすため，前景映像と背景映像の間の基礎行列によるエピポーラ幾何を用いる（2.3 節）．基礎行列は前景映像と背景映像内の特徴点の抽出とマッチングの結果から算出する．前景映像の各画素に対応する背景映像内の画素は式(2)によるエピポーラ線上に限定されるが，さらに，前景映像と背景映像の間のホモグラフィ行列によりエピポーラ線上の探索範囲を絞り込む．すなわち，背景の全ての物体が 3 次元空間上でホモグラフィ行列による平面に近い範囲内に存在すると仮定し，式(2)によるエピポーラ線上で式(1)のホモグラフィ行列による対応画素を中心とした一定の範囲だけを探索する．このとき，2.3 節の最後に述べたように，理論上はホモグラフィ行列による対応画素はエピポーラ線上にのるはずであるが，実際には，ホモグラフィ行列や基礎行列を求める際の精度の問題などのため，そうならない場合が普通である．そこで，図 8 に示すように，ホモグラフィ行列による対応画素の位置からエピポーラ線上で最も近い画素を基準参照画素として求め，その基準参照画素から一定の画素範囲内のエピポーラ線上に沿った参照画素を対応画素の候補とする．その全ての参照画素と前景映像の着目画素との間で 3.1 節の方法と同様の H，S，V 値による評価値を求め，評価値が最小の参照画素を対応画素とする．

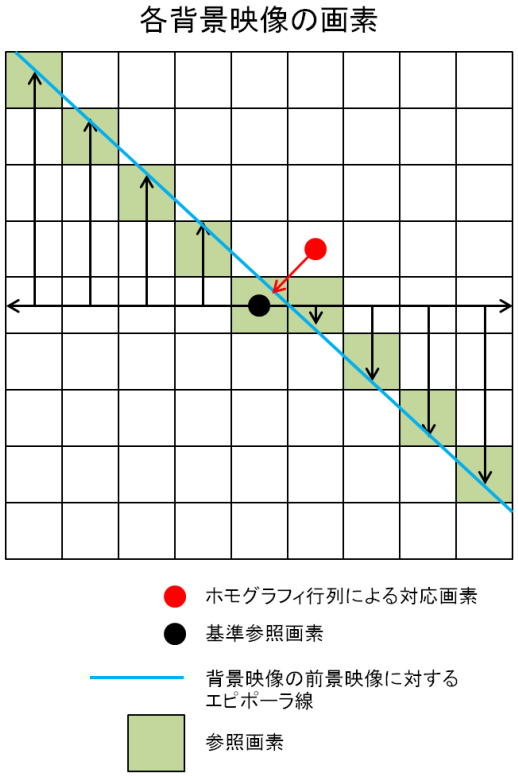

図 8　方法 2 における背景映像内の参照画素

4. 実験

提案手法の有効性を検証するために実験を行った．図 9 は実験環境を背景の側から見た画像である．背景から約 4.7 メートルほど離れた場所に 3 台の背景カメラ，さらに約 1.8 メートルほど離れた場所に 1 台の前景カメラを背景に向けて配置し，背景カメラと前景カメラの間に前景物体を置くことで，それぞれ，背景映像と前景映像を撮影した．実行環境を以下の表 1 に示す．

表 1　実行環境

CPU	AMD Phenom(tm) Ⅱ X4 965 3.40GHz
メモリ	8.00GB
グラフィックカード	ATI Radeon HD 4200
カメラ	PointGreyResearch 製 Dragonfly2
OS	Windows 7
開発環境	Visual Studio 2010, Visual C++

方法 1（3.1 節）による結果を図 10，11，12，13，14，方法 2（3.2 節）による結果を図 15，16，17，18 に示す．それぞれ，ビデオ映像上の一つのフレーム画像に関するものである．各映像の解像度は 640×480 画素とした．また，背景差分における評価値の計算において，H，S，V 値の差分の絶対値には，それぞれ，1.0，1.0，0.1 の重み付けを行い, 前景画素か背景画素かの判定のための閾値は 20 とした（0 <= H, S, V <= 255）．また，方法 2 における背景映像内の参照画素数は 21 とした．それぞれの処理時間を表 2，3 に示す．

方法 1 では，図 13 のマスク映像から分かるように，背景の全ての物体が同一平面上にあるという仮定が実際には成り立たず, 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，図 17 のマスク映像から分かるように，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，前景物体の存在領域にある物体のみの抽出が可能となった．また，HSV 表色系を用いた背景差分については，背景との色相の相違が大きい前景物体の抽出結果は良好となった．しかし，例えば，背景が白，前景物体が肌色のような，背景との色相の相違が小さい前景物体に対しては，前景物体領域も背景とみなされ，良好な結果が得られない場合が多かった．

図 9　実験環境


図 10　方法 1 の背景合成映像 src_{back}

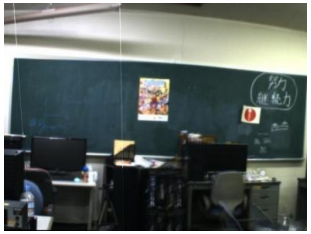
図 11　方法 1 の座標変換後の背景合成映像 dst_{back}

図 12　方法 1 の前景映像 src_{fore}

図 13　方法 1 の背景差分によるマスク映像．白：前景画素，黒：背景画素．

図 14　方法 1 による前景物体の抽出結果．図 12 の前景映像に図 13 のマスク映像を重ねて生成した．

図 15　方法 2 の各背景映像

図 16　方法 2 の前景映像

図 17　方法 2 の各背景映像に関する背景差分の結果を論理積により統合した合成マスク映像．白：前景画素，黒：背景画素．

図 18　方法 2 による前景物体の抽出結果．図 16 の前景映像に図 17 のマスク映像を重ねて生成した．

表 2　方法 1 の処理時間

参照画素領域	処理時間(s)	フレームレート(fps)
3×3 近傍	0.11	9.09
5×5 近傍	0.14	7.14

表 3　方法 2 の処理時間

参照画素数	処理時間(s)	フレームレート(fps)
5	0.094	10.64
11	0.11	9.09
21	0.15	6.67
31	0.17	5.88
41	0.22	4.55

5. まとめと今後の課題

本研究では，前景カメラと背景カメラを用いることで，背景に変化がある場合にも適切に目的とする前景物体のみをビデオ映像から抽出する手法を提案した．実験により，エピポーラ幾何とホモグラフィ行列を利用して背景映像ごとの背景差分を行う方法 2 について，良好な結果が得られた．

今後の課題としては，背景との色相の相違が小さい前景物体に対する抽出精度を上げ，また，処理速度の向上を図ることが挙げられる．さらに，本研究の手法を拡張し，物体の全周に前景カメラと背景カメラを配置し，全方向からの前景物体領域の抽出を可能にすることが今後の目標である．

謝辞

本研究を実施するにあたり，様々な貴重な御意見を頂いた岩手大学工学部千葉則茂教授に感謝致します．本研究の一部は科学研究費補助金（基盤研究(C) 21500090）の援助を受けている．

参考文献

1) A. Elgammal, D. Harwood, L. S. Davis, Non-parametric Model for Background Subtraction, Proceedings of the European Conference on Computer Vision 2000, pp.751-767, 2000.

2) T. Horprasert, D. Harwood, L. S. Davis, A Robust Background Subtraction and Shadow Detection, Proceedings of the Asian Conference on Computer Vision 2000, pp.983–988, 2000.

3) A. Elgammal, R. Duraiswami, D. Harwood, L. S. Davis, Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance, Proceedings of the IEEE, Vol.90, No.7, pp.1151-1163, 2002.

4) S.-C. S. Cheung, C. Kamath, Robust Techniques for Background Subtraction in Urban Traffic Video, Proceedings of the Conference on Visual Communications and Image Processing 2004, Vol.5308, No.1, pp.881-892, 2004.

5) A. Mittal, N. Paragios, Motion-Based Background Subtraction Using Adaptive Kernel Density Estimation, Proceedings of the Conference on Computer Vision and Pattern Recognition 2004, pp.302-309, 2004.

6) M. Heikkila, M. Pietikainen, A Texture-Based Method for Modeling the Background and Detecting Moving Objects, Transactions on Pattern Analysis and Machine Intelligence, Vol.28, No.4, pp.657-662, 2006.

7) Z. Zivkovic, F. v. d. Heijden, Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction, Pattern Recognition Letters, vol.27, no.7, pp.773-780, 2006.

8) W. Matusik, C. Buehler, R. Raskar, S. Gortler, L. McMillan, Image Based Visual Hulls, Proceedings of SIGGRAPH 2000, pp.369-374, 2000.

9) G. Slabaugh, B. Culbertson, T. Malzbender, R. Shafer, A Survey of Methods for Volumetric Scene Reconstruction from Photographs, Proceedings of the International Workshop on Volume Graphics 2001, pp.81–100, 2001.

10) C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, High-quality Video View Interpolation Using a Layered Representation, Proceedings of SIGGRAPH 2004, pp.600-608, 2004.

11) S. Corazza, L. Mundermann, A. M. Chaudhari, T. Demattio, C. Cobelli, T. P. Andriacchi, A Markerless Motion Capture System to Study Musculoskeletal Biomechanics: Visual Hull and Simulated Annealing Approach, Annals of Biomedical Engineering, Vol.34, No.6, pp.1019–1029, 2006.

12) H. Kim, R. Sakamoto, I. Kitahara, N. Orman, T. Toriyama, K. Kogure, Compensated Visual Hull for Defective Segmentation and Occlusion, Proceedings of 17th International Conference on Artificial Reality and Telexistence (ICAT2007),
pp.210-217, 2007.

13) D. Vlasic, I. Baran, W. Matusik, J. Popovic, Articulated Mesh Animation from Multi-view Silhouettes, Proceedings of SIGGRAPH 2008, pp.97:1-9, 2008.

14) H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, SURF: Speeded Up Robust Features, Computer Vision and Image Understanding, 110(3), pp.346-359, 2008.

15) 千葉直樹，蚊野浩，美濃導彦，安田昌司，画像特徴に基づくイメージモザイキング，電子情報通信学会論文誌， Vol.j82-D-Ⅱ，No.10，pp.1581-1589，1999 年.