

カメラとレンジセンサの疑似時刻合わせによる前景領域抽出

中島 秀真^{*1} 満上 育久^{*1} 山添 大丈^{*1}
波部 斉^{*2} 榎原 靖^{*1} 八木 康史^{*1}

Foreground Extraction by Pseudo-synchronization of Camera and Range Sensor

Hozuma Nakajima^{*1}, Ikuhisa Mitsugami^{*1}, Hirotake Yamazoe^{*1}
Hitoshi Habe^{*2}, Yasushi Makihara^{*1}, Yasushi Yagi^{*1}

Abstract – Existing background subtraction methods often fail to extract a foreground region whose color is similar to that of the background. When we use a co-located camera and range sensor, by which we can obtain both a color image and depth map simultaneously, it is expected to get a better foreground region by integrating the two kind of images. However, it is not straightforward when a moving object is observed because the camera and range sensor do not capture the scene synchronously. In this paper, we propose a novel method that pseudo-synchronize the camera and range sensor and integrate the background subtraction of the color and depth images to realize a good foreground extraction. Experimental results of a walking human show its effectiveness.

Keywords : Kinect, pseudo synchronization, earth mover's morphing, graph cut

1 はじめに

Microsoft社のXbox Kinect(以下Kinect)は、同社製ゲーム機器における直感的インタフェースのためのセンサとして販売されたが、同社[1]およびOpenNI[2]などのグループによってPCから制御できるSDKが提供されたことで、最近ではロボットビジョン・ヒューマンインタフェースなど幅広い分野で様々な用途で利用されている。Kinectによって得られるのは、これまでの一般的なカメラで得られていた二次元カラー画像と、そのシーンのメトリックな奥行き画像である。多くの利用事例において、この奥行き画像が簡単に得られるという点が重視されており、Shottonらはこの奥行き画像からシーン中の人物の三次元姿勢推定を行っている[3]。この技術は、MicrosoftのSDKの姿勢推定機能としても実装されており、容易に利用できる。また、IzadiらはKinectを環境中で動かしながら撮影することで環境の三次元形状復元を行う手法を提案している[4]。いずれの手法でも、推定や統合の対象は奥行き画像であり、同時に撮影されているカラー画像はそれらの処理結果に対して奥行き画像へのテクスチャマッピング用として付加的な形で利用されることが多い。

一方で、Kinectのように安価なレンジセンサが登場

する以前から、通常のカメラで撮影される二次元のカラー画像によって人物の姿勢推定[5, 6]や、Structure from MotionやMulti-view Stereoに代表されるような環境の三次元形状復元を行う手法[7, 8]などが数多く行われていることを考えると、Kinectに搭載されているカメラによって得られるカラー画像も積極的に利用した上で、レンジセンサと組み合わせるのが有用であると考えられる。観測値の属性や観測原理が異なれば苦手とするシーンの特徴は異なるため、同一シーンのカラー画像と奥行き画像を相補的に利用することができれば、処理全体のパフォーマンス向上が期待できる。

本論文では、前景領域抽出というタスクにおいて、この考え方を適用する。固定カメラによって人物などの前景を抽出する際、背景差分法は有力な手段である。しかし、カラー画像における背景差分は、前景色と背景色が類似している場合や背景に変動が起こった場合に弱い。これは古くから知られた問題であり、Gaussian Mixture ModelやParzen推定などさまざまな確率モデルによって背景自体の変動と前景の出現を区別する試みが行われている[9, 10]が、前景色と背景色が偶然一致している場合など、その原理上抽出が不可能なケースが存在する。一方、奥行き画像に対する背景差分は、色の類似性や背景変動には強いが、人の接地部分は背景との奥行き差分値が小さいため抽出が不安定となり、また、その計測原理上投影光の反射を正しく取得できない頭髪部分やオクルーディングエッジ近くの領域は正しく計測を行うことはできない。そこで本

^{*1}大阪大学 産業科学研究所

^{*2}近畿大学 理工学部

^{*1}Institute of Scientific and Industrial Research, Osaka University

^{*2}Faculty of Science and Technology, Kinki University

論文では、Kinect に代表されるカメラとレンジセンサの統合センサで観測されるカラー画像と奥行き画像の背景差分結果を統合することにより高精度な前景領域抽出処理を行うことを目的とする。

これを実現する上で、大きな足かせとなるのが、カメラとレンジセンサが同期して撮影を行うことができないという点である。市販のカメラとレンジセンサを適当に組み合わせるだけでは同期が取れないのは当然として、上述の Kinect においても、内蔵のカメラとレンジセンサは同期が取れておらず、この非同期性は、Kinect を含むすべてのカメラ・レンジセンサの統合センサに共通する問題である。本論文では、この問題を克服するため、モーフィング技術を利用して疑似的な同期画像系列を生成し、それをグラフカットによってカラー画像と奥行き画像の背景差分を統合することで、より正確で安定な人物領域抽出を実現する手法を提案した。歩容解析 [11] のための人物シルエット取得に Kinect を用いる状況を想定した実環境実験を行い、このような状況下での提案手法の有効性を確認した。

2 カメラとレンジセンサの統合

レンジセンサは、その奥行き計測原理には Time-of-Flight や構造化光投影法などのバリエーションがあるものの、いずれも中心射影的に、センサの光学中心から画角内の各方向への奥行きを計測し、その奥行きを画素値とする 2 次元画像情報として出力する。このレンジセンサとカメラを、その視野が重なり合うように設置することで、シーンのカラー情報と奥行き情報が同時に取得できる。さらに、このレンジセンサとカメラの距離が計測対象の距離に対して十分に小さければ、これらの光学中心は一致しているとみなすことができ、事前のキャリブレーションによって、共通の画像平面上にカラー情報と奥行き情報をマッピングできる。Microsoft 社の Kinect も、このようにカラー情報と奥行き情報を同時計測できる統合センサであり、図 1 のようにシーンのカラー画像と奥行きを同時に取得できる。カラー画像は内蔵されたカメラで撮影され、奥行き画像は同じく内蔵の赤外カメラと赤外光ドットパターン投影によって取得される。カラー画像用のカメラと奥行き画像用の赤外カメラは物理的に異なる位置にあり、その内部パラメータも異なるが、SDK により工場出荷時のキャリブレーション情報を利用することでカラー画像と奥行き画像の位置合わせが実現されている。本論文では、このように内部キャリブレーションが完了し、カラー画像と奥行き画像の位置合わせが済んだものとして議論を進める。

このような統合センサを用いることで、カラー情報と奥行き情報という 2 つの異なる観測情報で背景差分



図 1 Microsoft Xbox Kinect
Fig.1 Microsoft Xbox Kinect.

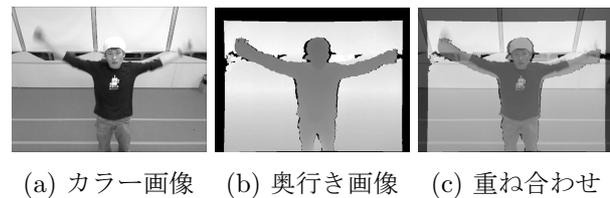


図 2 Kinect の撮影非同期問題
Fig.2 Asynchronization problem of Kinect.

を行うことができる。これらの 2 つの情報は異なった計測原理で取得されており、背景差分に失敗する条件も異なるため、それぞれに対する背景差分結果を統合することにより、より安定な前景抽出を行うことができると考えられる。しかし、このような処理を実現する際に問題となるのは、カメラとレンジセンサの非同期動作である。カメラとレンジセンサは、それぞれに内蔵された別のクロックによって撮影を行なっている。たとえ仕様上同じフレームレートを指定しても、それらは一般には同期しておらず、さらにはクロックの精度に依存するが、2 つのクロックは一般に完全には一致しないため、撮影時刻のズレ量が時間の経過とともに変化してしまう。実際、上述した Kinect でもこの問題が存在する。Kinect はカラー画像と奥行き画像が同時に得られる得られるデバイスとして知られているが、実はカメラとレンジセンサは同期していない。SDK によって同期撮影を指定することができる¹ が、これは実はソフトウェアレベルでの同期に過ぎない。すなわち非同期で撮影されたカラー画像と奥行き画像に対して、時間的に最も近接している画像ペアを PC に返すのみであり、例えば 30fps で撮影を行った場合、撮影時刻のズレが $(1000/30)/2 = 16.6\text{ms}$ 以下になることが保障されるだけである。図 2 に、ソフトウェア同期を行った際の画像ペアの一例を示す。図 2(a)(b) はそれぞれ Kinect から取得されたカラー画像と奥行き画像であり、図 2(c) はそれらを重ね合わせたものである。16.6ms という撮影時刻ズレは数値的には小さく思われるが、観測対象として人物を考えた場合、こ

¹OpenNI では、撮影時に WaitAndUpdateAll 関数を用いてカラー画像と奥行き画像の両方のノードで新しいデータが取得可能になるのを待ってから二つのノードを更新し、さらに FrameSyncWith 関数を用いてフレームの同期機能を有効にする。

の図のように腕を回した場合や歩行時などに腕や足の位置ズレの要因となる．本論文で目的とする，カラー画像・奥行き画像の背景差分の統合を行うにあたり，この時刻ズレによる位置ずれは大きな問題となる．

3 疑似時刻合わせによる前景領域抽出

本論文では，2章で述べたカメラとレンジセンサの非同期性を考慮して，カラー画像と奥行き画像の背景差分の統合による前景抽出法を提案する．この手法では，まず異なるクロックで動作するカメラとレンジセンサによる撮影画像系列を取得し，両系列の時刻合わせを行う．本論文においては前景は人物とし，その人物の歩行や運動によるシルエット形状変化に着目して，その一致度を評価して時刻合わせを行う．その際，カメラとレンジセンサが非同期で撮影されるという問題を克服するために，一方の系列に対してモーフィング処理を施して時間超解像化させ，他方の系列との時刻一致度の評価を可能としている．さらに，その結果得られる疑似的に同期のとれたカラー画像と奥行き画像から得られる前景抽出結果を用いて，グラフカットによって正確な前景領域抽出を行う．

なお，前段の時刻合わせ処理については，Kinect などのようにデバイス間で共通のタイムスタンプが発行されるデバイスでは不要である．しかし，Kinect によって取得したデータでも，そこにタイムスタンプを記録していない場合もあり，そのようなケースに対して適用可能である．

3.1 カラー・奥行き画像系列の疑似時刻合わせ

まず，カラー画像・奥行き画像の時系列それぞれに対して背景差分法を施し前景のシルエット画像列を得る．この2つのシルエット画像列は非同期で撮影されているため，図3に示すように，あるカラー画像に対して厳密に同時刻に撮影された奥行き画像は一般には存在しない．そこで，一方のシルエット画像系列を疑似的に時間超解像化し，他方の系列の各フレームと撮影時刻の一致するフレームを探索し対応付けていくことで，両系列間の対応関係を算出する．シルエット画像系列の時間超解像化においては，本論文では，画像系列中で時間的に隣接した2画像に対して楨原らの Earth Mover's Morphing (EMM) [12] によりモーフィングを行う．EMM では分割数 M を任意に設定することが可能で，これにより隣接2画像の間に遷移率 $1/M, 2/M, \dots, (M-1)/M$ の画像を生成し，系列のフレームレートを疑似的に M 倍にすることができる(図4)．

$$D(T, S) = \sum_t d_t(T, S) \quad (1)$$

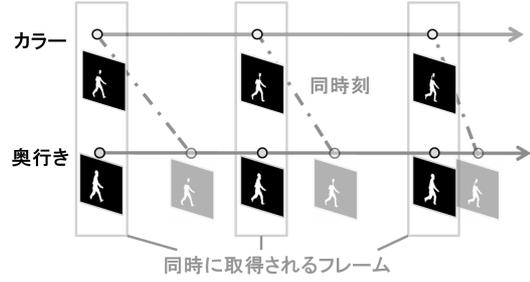


図3 カラー画像と奥行き画像のシルエット画像系列
Fig. 3 Silhouette image series of color images and depth maps.

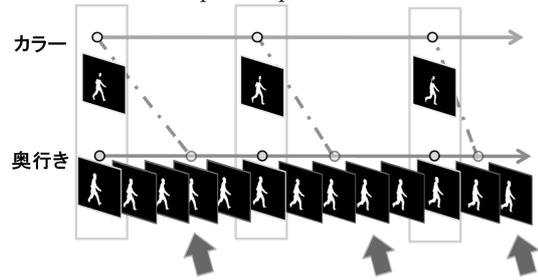


図4 モーフィングによる高フレームレート画像系列
Fig. 4 Generating high frame rate images.

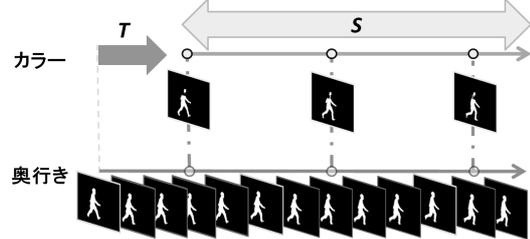


図5 カラー画像と奥行き画像の系列全体の時刻合わせ
Fig. 5 Time adjustment of color and depth image series.

$$T_o, S_o = \arg \max_{T, S} D(T, S) \quad (2)$$

2系列の時間対応は，それぞれの系列が一定のフレームレートで撮影されていれば，2者間の時間方向の平行移動量 T と時間伸縮量 S という2つのパラメータによって定義される¹ (図5)．一方の系列の f_1 フレーム目に対応する他方の系列のフレーム番号を f_2 とするとき，これらの関係は次式で表現される．

$$f_2 = S f_1 + T \quad (3)$$

2系列が時間的に一致する最適な $(T, S) = (T_o, S_o)$ を推定するには，理論上は異なる2時刻での対応するフレームが得られればよいが，背景差分によって抽出されたシルエットにはノイズが乗ることや，一方の画像系列はモーフィングによって生成された疑似的なシ

¹ T の単位は「フレーム」である．

シルエット画像であることを考慮すると、時刻の選択によっては対応関係の推定を大きく誤る可能性がある。そこで、2時刻のみに限定せず、すべての時刻 t においてその時刻に対応するシルエット画像ペアの時刻一致度 d_t を評価し、その総和を両シルエット画像系列全体の一致度 D とし、それが最大となる (T, S) として (T_o, S_o) を求める方法をとる。

各時刻の時刻一致度 d_t は、図 6(a-1)(b-1) のようなシルエット画像ペアに対して定義する。一致度の評価尺度として考えられる最も単純なものは、シルエットの共通部分の面積の大きさである。図 6 の (c-1) は (a-1) と (b-1) を重ね合わせたもので、白色の領域が両者の共通部分、灰色の領域が非共通部分になっている。時刻の差が大きいほど図 6 の (d-1) のようにシルエット全体のズレが大きくなり共通部分の面積は減少するため、この評価尺度は時刻一致度と一応の相関が期待される。しかし、図 6(e-1) のようにシルエットの一部が背景差分の失敗により大きく欠落してしまう場合も、(d-1) と同様に評価値が下がってしまう。各画像からのシルエット抽出処理においてこのような欠落が頻繁に見られることを考えると、時刻の一致度を正しく評価するには、そのような欠落の影響を受けにくく、時刻ズレの影響には敏感な評価尺度が望ましい。そこで、本論文では、シルエットの輪郭線(図 6(a-2)(b-2))と中心線(図 6(a-3)(b-3))に着目し、シルエット画像のペアに対してこれらの線の共通部の面積を評価尺度とする。図 6(c-2)(d-2)(e-2)(c-3)(d-3)(e-3) と図 7 に、これらの評価尺度での、時刻ズレ発生時とシルエット欠落時の評価値の減少割合の違いを示す。ここでは、2画像の時刻が一致している場合におけるシルエットの共通部分の面積を基準とし、時刻ズレの発生とシルエットの欠落によって、それぞれ共通部分の面積が 95% となるように設定している。結果から、輪郭線や中心線の重なりを時刻一致度の評価尺度に利用すると、シルエット自体の重なりを利用したときと比較して、時刻ズレの影響による評価値の減少が大きく、シルエットの欠落による影響が相対的に小さくなることがわかる。なお、実際には、輪郭線や中心線そのままでは共通部分の画素数が少なすぎて不安定なため、それぞれ膨張処理を加えたものの共通部分の大きさにより評価している。

系列間の時刻一致度 D が最大となる (T_o, S_o) は、 (T, S) 探索範囲内から全探索的に求める。図 8 に T, S の変化に対する D の分布を示す。この図のように、通常、分布は巨視的に見て単峰性を示すことが分かる。しかしデータによっては、輪郭線や中心線画像の離散化誤差により、図 8(b) のように微視的に見ると増減が確認される場合がある。従って、式 (2) の通りにこ

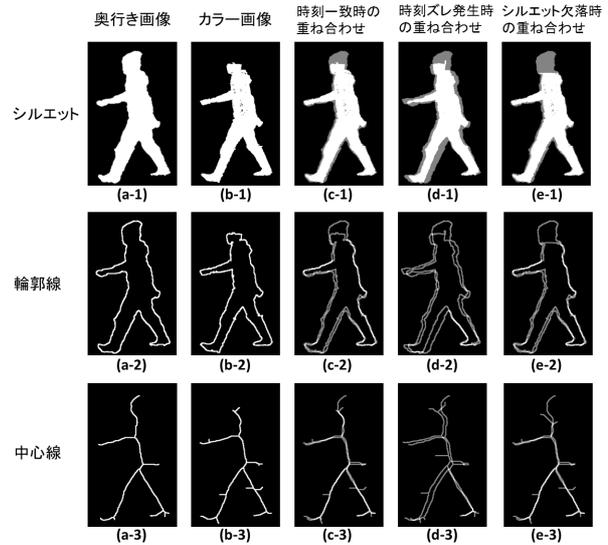


図 6 時刻一致度の評価尺度の比較
Fig.6 Comparison of matching measurements.

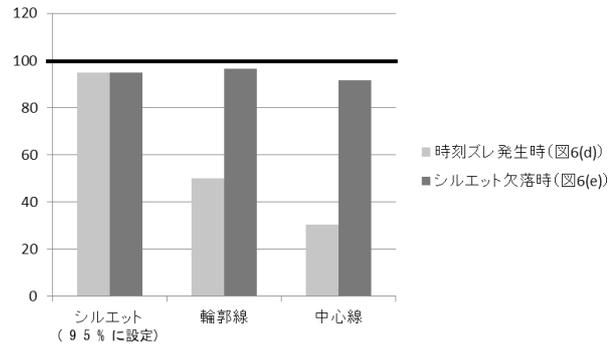
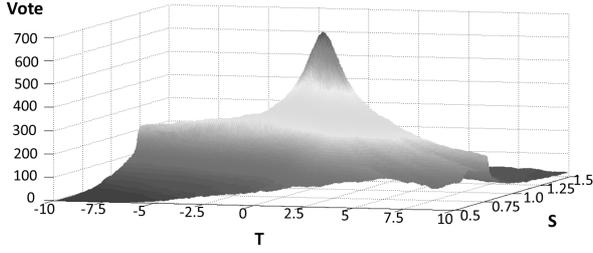


図 7 時刻ズレ発生時とシルエット欠落時の評価値の減少割合の違い
Fig.7 Decrease of evaluation value caused by time lag and partial crack.

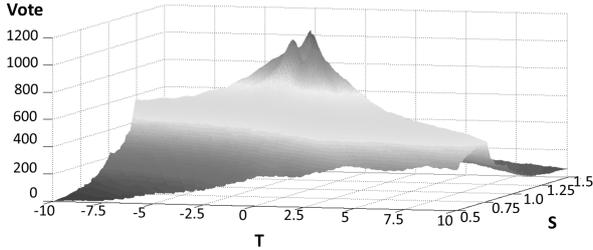
の分布に対して単純に最大値をとるのではなく、巨視的な分布のピークを選択するのが望ましいと考えられる。そこで本論文では、この分布 D に以下の二次元ガウス分布をフィッティングして、その中心 (μ_T, μ_S) をこの分布のピークとし、 (T_o, S_o) として採用する。

$$F(S, T) = \alpha \frac{1}{2\pi\sigma_S\sigma_T\sqrt{1-\rho_{ST}}} e^{-\frac{1}{2(1-\rho_{ST}^2)} \left\{ \frac{(S-\mu_S)^2}{\sigma_S^2} + \frac{(T-\mu_T)^2}{\sigma_T^2} - \frac{2\rho_{ST}(S-\mu_S)(T-\mu_T)}{\sigma_S\sigma_T} \right\}} \quad (4)$$

このフィッティングには、レーベンバーグ・マーカート法による非線形最適化を利用する。最適化のための初期値は以下の方法によって得る。まず、分布のピー



(a)



(b)

図 8 系列間の時刻一致度 $D(T, S)$
Fig. 8 Matching score of two series $D(T, S)$.

クから離れた裾野部分はピークの推定には無関係なため、あるオフセット値を定め、それ以下の評価値の領域は削除する。次に、各 (T, S) における評価値を、その (T, S) の点の個数とすることで、この 3次元曲面を (T, S) の二次元平面上の点群とみなし、その点群に対する主成分分析を行う。これによって、この分布の中心と軸が算出され、式 (4) の各パラメータの初期値が決定される。

3.2 時刻合わせ済みの画像系列を用いた前景抽出

前節で得られた (T_o, S_o) によって、一方の系列 F_1 の各フレーム f_1 に対応する他方の系列 F_2 のフレーム f_2 を前節同様 EMM を用いて生成する。これは、 F_2 の $[Sf_1 + T]$ フレーム目と $[Sf_1 + T] + 1$ フレーム目を遷移率 $Sf_1 + T - [Sf_1 + T]$ でモーフィングして得られる。Kinect など、カメラとレンジセンサで共通のタイムスタンプが得られる統合センサの場合は、得られるタイムスタンプから同様の比例計算によって遷移率を設定しシルエットのモーフィングを行うことも可能である。

以上の処理によって得られた疑似的に同期した 2 つのシルエット系列に対してグラフカット [13] を適用し、最終的な前景領域を取得する。画像 V の各画素を v としたとき、ラベルを $\mathbf{X} = \{X_1, X_2, \dots, X_v, \dots, X_V\}$ とし、各 X_v には 0 (背景) か 1 (前景) の値を与える。以下の式で定義されるエネルギー関数 $E(\mathbf{X})$ を最小化するような \mathbf{X} とし前景領域を決定する (図 9)。

$$E(\mathbf{X}) = \lambda \sum_{v \in V} g_v(X_v) + \kappa \sum_{(u,v) \in E} h_{uv}(X_u, X_v) \quad (5)$$

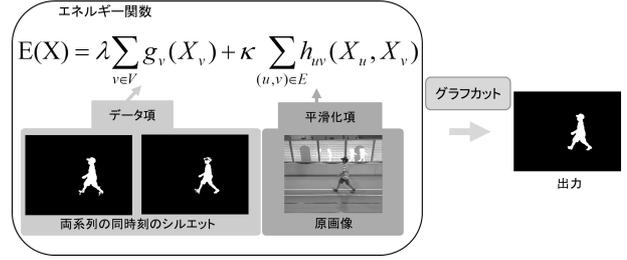


図 9 グラフカットによるシルエット統合
Fig. 9 Integration of two silhouette based on graph cut.

$$g_v(1) = \exp(-(\alpha_c c_v + \alpha_d d_v)) \quad (6)$$

$$g_v(0) = 1 - g_v(1) \quad (7)$$

$$h_{u,v} = \|I_u - I_v\| \quad (8)$$

式 (5) の第一項はデータ項、第二項は平滑化項である。データ項の c_v, d_v はそれぞれカラー画像、奥行き画像から抽出したシルエットの位置 v の画素の値、平滑化項の I_u, I_v は原画像における位置 u, v の画素の輝度値である。また、 α_c, α_d はバランス調整のためのハイパーパラメータである。

4 実験

4.1 実験環境

本手法の応用用途を歩容解析のためのシルエット抽出と想定し、直線歩行路を設けた実環境で人物が歩行するシーンを、カメラとレンジセンサの統合センサで側面から観測し、提案手法の有効性を確認した。統合センサとして、Microsoft 社の Kinect を使用した。ただし、一般的な統合センサと条件を揃えるため、タイムスタンプは記録せず撮影したデータを対象とした。計測範囲は Kinect の奥行き推定の有効距離や人間の全身を画角に収める必要性から 1.5m~3.5m とした。カラー画像・奥行き画像における背景差分は、各画素を単一ガウス分布でモデル化する方法を用いた。ただし、シーンに若干の照明変動があったため、カラー画像の背景差分においては、直近のある時間幅で観測された背景色情報でこのモデルを動的に更新する。一方、奥行き画像の背景画像については事前に取得した固定の背景モデルを用いた。さらに、ノイズ除去のために、背景差分結果の 2 値化画像をセグメンテーションし、ある一定の大きさ以下の領域を削除した。

また、提案手法の時刻合わせ処理における使用フレーム数は 50、EMM の分割数 M は 50、探索する S, T の範囲は S が $[0.5, 1.5]$ 、 T が $[-10, 10]$ 、刻み幅は S が 0.01、 T が 0.02 とし、時刻一致度の評価に用いる輪郭線と中心線の膨脹回数は 1 回とした。

4.2 実験結果

まず、時刻合わせによって対応付けられた画像を重ね合わせた結果を図 10 に示す。図中の白色の領域が共通部分の領域、灰色の領域がカラー画像のシルエットもしくは奥行き画像のシルエットのみの領域である。背景差分の失敗により、奥行き画像では床面に接地している領域で、カラー画像では背景色と前景色が類似している領域で欠落が発生し、灰色になっている。(a) は Kinect のソフトウェア同期によって対応付けられたフレームの画像の重ね合わせ、(b)~(e) は提案手法で時刻合わせを行い対応付けた疑似的な同時刻画像の重ね合わせである。(b),(c) は奥行き画像の系列でモーフィングを行い各カラー画像と同時刻のものを対応付けている。一方、(d),(e) は逆にモーフィングしたカラー画像の系列を各奥行き画像に対応付けている。また、(b),(d) ではシルエットの輪郭線を、(c),(e) では中心線をそれぞれ時刻一致度の計算に用いた。

それぞれ時刻合わせによって、Kinect のソフトウェア同期で対応付けられたフレームよりも同時刻らしいシルエット対応付けが行われているのが確認できる。(d)(e) の最下段では人物シルエットの首部分にノイズのようなものが観測されているが、これはカラー画像の連続フレーム間で図 11 のようにシルエットの欠落の仕方が異なるため、このフレーム間でモーフィングした場合、大きくノイズが乗ったようになってしまうためである。カラー画像・奥行き画像のいずれが背景差分法に失敗しやすいかはシーンに依存するものの、一般に色の観測は非常に不安定なため、カラー画像による背景差分の方が失敗を生じやすい。そこで、以降の実験では、カラー画像系列を基準とし、奥行き画像系列を EMM によって時間超解像化することとした。

また、図 10 の (b) と (c) を比較すると、この例ではほぼ差が生じておらず、輪郭線・中心線いずれを用いても性能は変わらないことが確認できた。ただし、これはカメラ・レンジセンサの計測特性に依存しており、モーションブレンダーが起こった際に観測画像がどのような挙動を示すかに応じて適切に選択すべきである。例えば、Kinect の場合、腕や足など高速で運動する部位は、その運動方向に沿って奥行き画像が太く観測されることがある。このような場合は、時刻一致度の評価には輪郭線よりも中心線を用いたほうがよいと考えられる。

次に、図 10 の時刻合わせの結果の後段のグラフカットによる前景抽出を行った結果を図 12 に示す。(b)~(e) の各図は、得られた前景領域をマスクした原画像である。(b) の画像はカラー画像のみの背景差分により得られた前景領域、(c) は Kinect のソフトウェア同期で対応付けられたフレームにグラフカットを適用す

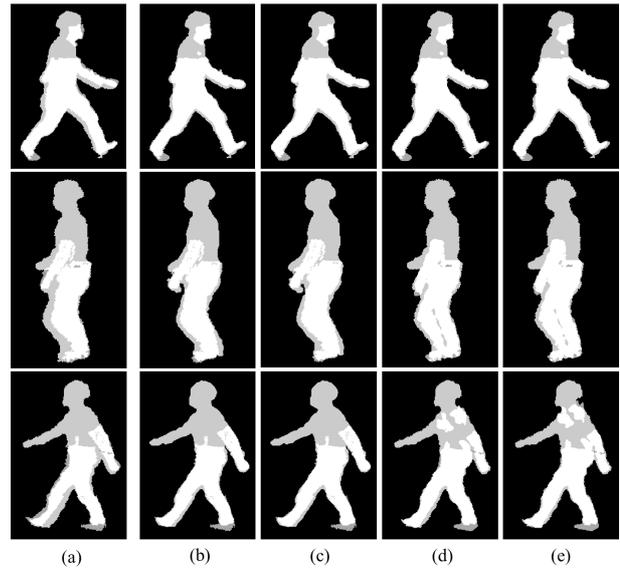


図 10 両系列のシルエットの重ね合わせ
Fig. 10 Two silhouettes aligned by the proposed method.



図 11 欠落の仕方が違う隣接フレームのモーフィング結果
Fig. 11 Morphing of differently cracked two silhouettes.

ることにより得られた前景領域、(d),(e) は提案手法によって得られた前景領域である。(d) はシルエットの輪郭線、(e) は中心線を時刻一致度の計算に利用した結果である。

(b) では、背景差分の失敗により上半身が大きく欠けてしまっている。(c) は (b) と比較すると、上半身の大きな欠落はないものの、時刻ズレのある背景差分結果を統合しているため背景領域を多く含んでる。これらと比較して提案手法 (d)(e) では、モーションブレンダーなどの影響で多少ノイズは残るものの、(c) で見られたような大きなノイズの削減には成功している。

4.3 ガウス分布のフィッティングによる (T, S) 最適化の効果

3.1 節で述べた通り、系列間の時刻一致度 D は、輪郭線や中心線画像の離散化誤差等の原因で微視的には単峰性を示さないこともあるので、巨視的な分布のピークを求めるために、単純な最大値探索でなくガウス分布のフィッティングを行っている。この節では、そのフィッティングの効果を確認する。時刻一致度の分布 D について、そのピーク付近を拡大したものと、それ

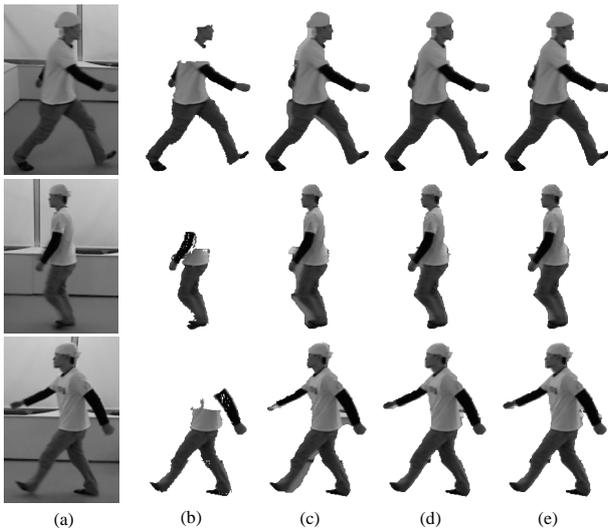


図 12 提案手法による前景領域抽出

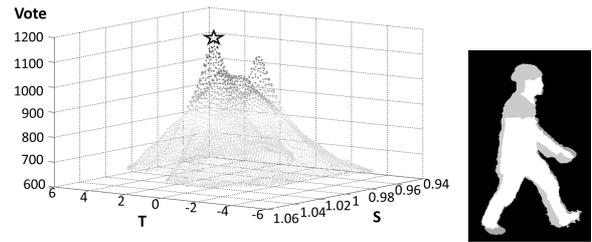
Fig. 12 Foreground extraction by proposed method.

に対し単純な最大値探索により求めたピーク, ガウス分布のフィッティングにより求めたピークと, 各ピークの (T_0, S_0) を基に対応付けられた画像を重ね合わせた結果を図 13 に示す. 結果から, ガウス分布のフィッティングによって分布の巨視的なピークが正しく得られていることが分かる.

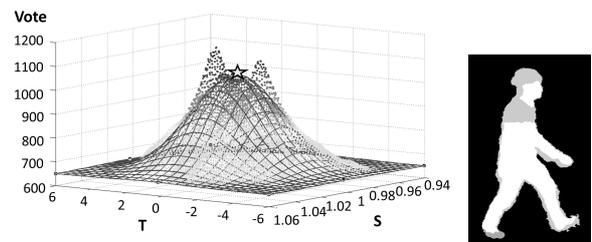
4.4 フレームレートと抽出性能の関係

提案手法の有効性は, EMM によるシルエットモーフィングの性能に強く依存するが, 統合センサのフレームレートが低下するに従い, モーフィングで推定しなければならない時間幅が増加するため, 生成されるシルエットの信頼性が低下する. このことを踏まえ, 歩行者を対象とした際に, 提案手法がどの程度のフレームレートまで有効に働くかを調査するための実験を行った. 実験では, 前節と同様に統合センサとして Kinect を使用し, そのタイムスタンプを使用しない. カラー画像・奥行き画像はそれぞれ 30fps で撮影しているが, 各画像系列を 1/2, 1/4, 1/8 に間引くことで, 仮想的にそれぞれ 15fps, 7.5fps, 3.75fps のフレームレートの画像系列を作成し, 提案手法を適用した.

実験結果を図 14 に示す. いずれの結果も提案手法の結果であり, (a),(b) はシルエットの輪郭線, (c),(d) は中心線をそれぞれ時刻一致度の評価に用いた結果である. なお, (a),(c) は図 10 と同様に, 両画像系列の時刻合わせを行い対応するシルエットを重ねあわせたもの, (b),(d) は図 10 と同様に, 時刻合わせ結果を元に前景抽出を行ったものである. (b) の結果を見ると, 7.5fps および 3.75fps では, 前景領域に背景部分が多く観測されていることから, 前景抽出性能が低下していることが分かる. これより, 提案手法によって信頼



(a) 最大値探索で求めたピーク



(b) フィッティングで求めたピーク

図 13 ガウス分布のフィッティングによる (T, S) の最適化

Fig. 13 Gauss fitting to obtain global peak (T, S) .

できる前景抽出結果を得るためには, 15fps 程度のフレームレートが必要であることが分かる.

5 おわりに

本論文では, 非同期で動作するカメラとレンジセンサを用いて, それらによって取得されるカラー画像と奥行き画像の背景差分結果を統合することで, 前景領域抽出を高精度に行う手法を提案した. これを実現するために, 提案手法では, カラー画像と奥行き画像から抽出したシルエット系列に対し, EMM により中間時刻の画像列を作成することで系列を時間超解像化し, 両画像系列の疑似的な時刻合わせを行った. そして, 時刻合わせが施されたカラー画像と奥行き画像それぞれについて背景差分を行い, グラフカットにより統合して, 前景領域抽出を行った. 実環境で歩行者を対象に抽出を行い, カラー画像のみを用いたときの抽出結果や, Kinect から同時に取得されたフレームの奥行き画像と対応付け統合した場合の抽出結果と比較することで, その有効性を確認した.

今後の課題としては, 最終的な前景領域抽出処理における時系列情報の利用が挙げられる. 現在は, 系列間の時間合わせが済んだ後は, 同時刻のカラー画像・奥行き画像の統合は各時刻ごとに独立に行なっているが, 時間方向の連続性を利用することで, さらに抽出精度を向上させられる可能性がある. また, 本論文では, 視点が一致しているとみなすことのできる 1 対のカメラとレンジセンサを対象にその疑似的な同期を実現したが, さらに台数を増やし, 視点の異なるカメラ

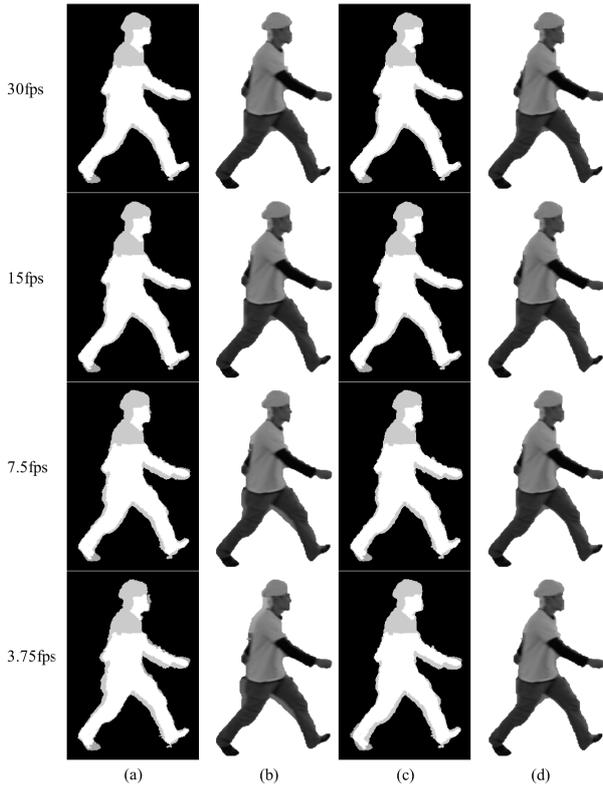
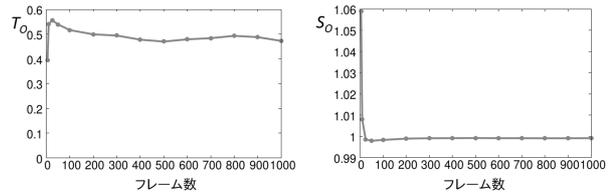


図 14 フレームレートの低下による前景領域抽出精度の低下

Fig. 14 Results of low-frame-rate images.

やレンジセンサとも疑似的な同期を行い、高精度な人物三次元形状復元を行うことも今後の課題である。さらなる課題として、計算コストの改善が挙げられる。一つの方法として、試行毎に (T_o, S_o) の推定を行うのではなく、予め (T_o, S_o) を推定しておき、以降にその数値と各センサ内のクロックを基に時刻合わせを行うことで、実行回数を減らすことができる。ただしその場合、推定した (T_o, S_o) の値が正確でなければ、時間の経過に伴って時刻合わせに大きな誤差を生じてしまうため、ある程度以上の枚数のフレームを用いて高精度に推定を行う必要がある。図 15 に示したのは、ある試行における、使用フレーム数に対する (T_o, S_o) の変化の様子である。200 フレーム前後で収束していることが確認できる。この結果から、正確な (T_o, S_o) の推定を行うには 200 枚程度のフレーム数、時間にして 7 秒分のデータがあれば良いことが分かる。ただし、この場合も初めの (T_o, S_o) の推定には時間を要する。例えば、現状の実装では、200 枚を用いた (T_o, S_o) の推定に、時刻一致度の評価尺度を中心線、分割数 M は 50、探索範囲に関しては T が $[-5, 5]$ の範囲で刻み幅 0.05、 S が $[0.9, 1.1]$ の範囲で刻み幅 0.001 と設定したとき、約 8 時間を要することを実験的に確認した。また、改善のもう一つの方法は、この 1 回の (T_o, S_o)



(a) T_o の値の変化 (b) S_o の値の変化

図 15 推定に使用するフレーム数による (T_o, S_o) の変化

Fig. 15 Variation of (T_o, S_o) by the number of frames used for the estimate.

の推定にかかる時間自体を短縮することである。例えば現状の実装では、分布の巨視的なピークを求める必要性から実際には取り得ない範囲の (T, S) も探索しており、かつ (T, S) の探索範囲内全域で一定の細かい刻み幅を用いているが、まず大きい刻み幅で探索することで分布の大まかなピークを把握し、そのピークの付近で細かい刻み幅で探索することにより、推定の精度を落とさずに計算量を削減できると考えられる。

謝辞

本研究は、科学技術振興機構 (JST) 戦略的創造研究推進事業 (CREST) の支援のもとに推進された。

参考文献

- [1] <http://kinectforwindows.org/>
- [2] <http://www.openni.org/>
- [3] J. Shotton, A. Fitzgibbon, M. Cook, A. Blake, "Real-time Human Pose Recognition in Parts from Single Depth Images," Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2011.
- [4] S. Izadi, R. Newcombe, D. Kim, O. Hilliges, D. Molyneaux, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, A. Fitzgibbon, "KinectFusion: Real-Time Dynamic 3D Surface Reconstruction and Interaction, Proc. SIGGRAPH, 2011.
- [5] Y. Yang, D. Ramanan, "Articulated Pose estimation with Flexible Mixtures-of-parts," Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2011.
- [6] Y. Wang, D. Tran, Z. Liao, "Learning Hierarchical Poselets for Human Parsing," Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2011.
- [7] N. Snavely, S. M. Seitz, R. Szeliski, "Modeling the World from Internet Photo Collections," International Journal of Computer Vision, Vol.80, No.2, pp.189–210, 2007.
- [8] Y. Furukawa, J. Ponce, "Accurate, Dense, and Robust Multi-View Stereopsis," Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2007.
- [9] I. Pavlidis, V. Morellas, P. Tsiamyrtzis, and S. Harp, "Urban surveillance systems: from the laboratory to the commercial world," Proc. the IEEE,

- Vol.89, No.10, pp.1478-1497, 2001.
- [10] T. Tanaka, A. Shimada, D. Arita, R. Taniguchi, "A Fast Algorithm for Adaptive Background Model Construction Using Parzen Density Estimation," Proc. IEEE International Conference on Advanced Video and Signal based Surveillance, 2007.
- [11] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi, "Adaptation to Walking Direction Changes for Gait Identification", Proc. of the 18th Int. Conf. on Pattern Recognition, Vol. 2, pp. 96-99, Hong Kong, China, Aug. 2006.
- [12] Y. Makihara and Y. Yagi, "Earth Mover's Morphing: Topology-Free Shape Morphing Using Cluster-Based EMD Flows," Proc. the 10th Asian. Conf. on Computer Vision, pp. 2302-2315, Queenstown, New Zealand, Nov. 2010.
- [13] 石川博: "グラフカット (チュートリアル)", 情報処理学会研究報告. CVIM, pp. 193-204, 2007.

(2012年2月27日受付)

[著者紹介]

中島秀真



大阪大学大学院情報科学研究科コンピュータサイエンス専攻博士前期課程在学中. Kinect による動物体三次元形状復元に関する研究に従事.

満上育久 (正会員)



2001 京大・工・電気電子工学卒. 2007 奈良先端科学技術大学院大学情報科学研究科博士後期課程了. 同年京都大学学術情報メディアセンター研究員. 2010 より大阪大学産業科学研究所助教. ジオメトリを中心としたコンピュータビジョン, 画像理解, 対象検出・追跡等に関する研究に従事. IEEE, 電子情報通信学会, 情報処理学会, 日本バーチャルリアリティ学会各会員. 博士 (工学).

山添大丈



2000 大阪大学基礎工学部システム工学科卒. 2002 同大学院基礎工学研究科修士課程了. 2005 同大学院基礎工学研究科博士課程了. 同年 ATR メディア情報科学研究科専任研究員. 2006 年 ATR 知能ロボティクス研究所専任研究員. 2011 年 4 月より, 大阪大学産業科学研究所特任助教. 博士 (工学).

波部 齊



1999 京都大学大学院工学研究科電子通信工学専攻修士課程了. 三菱電機 (株), 京都大学助手, 奈良先端科学技術大学院大学助手, 同助教を経て, 2011 大阪大学産業科学研究所特任講師 (常勤). 2012 近畿大学理工学部情報学科講師. コンピュータビジョンに関する研究に従事. IEEE, ACM, 電子情報通信学会, 情報処理学会各会員. 博士 (情報学).

榎原 靖



2001 年大阪大学工学部応用理工学科卒業. 2002, 2005 年にそれぞれ同大学院工学研究科博士前期・後期課程を了. 同年大阪大学産業科学研究所特任助手, 2006 年同研究所助手, 2007 年同研究所助教となり現在に至る. 歩容認証をはじめとするコンピュータビジョン・パターン認識に関する研究に従事. 日本機械学会第 79 期ロボティクス・メカトロニクス部門ベストプレゼンテーション表彰受賞. 日本ロボット学会第 18 回研究奨励賞受賞. 情報処理学会平成 18 年度山下記念研究賞, 第 15 回画像の認識・理解シンポジウム優秀論文賞受賞. 情報処理学会, 日本ロボット学会, 日本機械学会各会員. 博士 (工学).

八木康史



1983 阪大・基礎工・制御卒. 1985 同大学院修士課程了. 同年三菱電機 (株) 入社. 同社産業システム研究所にてロボットビジョンの研究に従事. 1990 大阪大学基礎工学部情報工学科助手. 同学部システム工学科講師, 同大学院助教授を経て, 2003 より同大学産業科学研究所教授. 2012 より同大学同研究所所長. コンピュータビジョンに関する研究に従事. 情報処理学会フェロー, IEEE, 情報処理学会, 日本ロボット学会各会員. 博士 (工学).