

研究テーマ 生活環境内対象物の分析と位置認識に関する研究

研究者 斎藤文彦
伊藤昭

岐阜大学
岐阜大学

共同研究員
共同研究員

フェーズⅡ

1 研究の概要

本研究は、フェーズⅠで実施した「人物検出に関する研究(4-1)」におけるパーセプトルームのような生活環境内から、より安定した人物の検出を可能とすることを目的とし、人周辺環境下における静止背景画像生成手法に関する研究、2次元エッジベクトル差分による微動人物領域の検出に関する研究、3次元ポリゴンモデルを用いたステレオ人物頭部追跡に関する研究を実施した。

人周辺環境下における静止背景画像生成手法に関する研究では、常に移動物の存在する動画から背景画像を安定して生成する手法を提案した。

2次元エッジベクトル差分による微動人物領域の検出に関する研究では、人間が静止している場合、完全に静止している訳ではなく、微動していることを利用し、人間の微動を捉えることにより、長時間留まっている人物の検出を可能にした。ここでは、時系列2次元エッジを用いることで、画素ごとの空間的・時間的変化量を求め、さらに精度よく変化を捉えるために、エッジ方向を組み合わせた手法を提案した。

3次元ポリゴンモデルを用いたステレオ人物頭部追跡に関する研究では、ステレオカメラを用いて人物の頭部を追跡する方法を開発した。ここでは、ステレオカメラから得られる時系列距離画像の中で、人物頭部の位置と姿勢を推定するために、パーティクルフィルタを用いた。頭部の位置・姿勢の事後確率分布は3次元ポリゴンモデルの重み付き集合によって表現され、人物の頭部は肌色や頭髪の色などを色情報として持つポリゴンモデルとしてあらわされる。必要なパーティクルの個数は対象の次元に対して指数関数で増加するため、頭部の位置追跡と位置推定は別の処理として行った。

2 研究の目標

より安定した人物検出を可能とするために、生活環境を撮影した画像を分析し、人物検出のための画像処理手法の開発を目指す。人周辺環境下における静止背景画像生成手法に関する研究では、濃度情報以外の背景情報として、空間エッジと時間エッジからなる時系列2次元エッジ差分を用いることによって、背景らしいフレームのみを選択してヒストグラムに累積し、ノイズの少ない背景画像を生成する。2次元エッジベクトル差分による微動人物領域の検出に関する研究では、画像の空間情報である空間エッジと、その勾配の向きを表すエッジ方向を成分にもつエッジ方向ベクトルを用い、この変化を検出することで濃度だけでは検出できないようなわずかな変化を捉える。3次元ポリゴンモデルを用いたステレオ人物頭部追跡に関する研究では、ノイズのある環境において、ロバストな頭部追跡を実現する。

3 実施内容と結果

本研究で実施した、人周辺環境下における静止背景画像生成手法に関する研究、2次元エッジベクトル差分による微動人物領域の検出に関する研究、3次元ポリゴンモデルを用いたステレオ人物頭部追跡に関する研究について、それぞれの内容を以下に述べる。

3. 1 人周辺環境下における静止背景画像生成手法に関する研究

3. 1. 1 はじめに

近年、コンピュータの処理速度の向上や記憶媒体の容量の増大に伴い、動画認識技術に対する期待が高まっている。動画認識分野における重要な研究テーマとして、移動物体抽出がある[1][2][3]。移動物体抽出の代表的な手法に背景差分法がある[4]。この手法は、移動物が含まれていない画像を背景として、背景画像と入力画像との差を画素ごとに計算することで、差の大きい部分を移動物体として抽出する。この手法では、事前に背景を採取しておく必要がある。しかし実際の場面では、道路や人通りが多い場所など移動物体が含まれない背景画像を取得することが難しい状況が多い。またこのような場所を風景として撮影したい場合において、車や歩行者は邪魔となる。このような場合には、移動物体を含んだ動画から背景を推定し生成する必要がある。

本研究では、常に移動物体の存在する動画から背景画像を安定して生成する手法を提案する。提案手法によって生成された背景画像に対して定量評価実験を行うことで本手法の有効性を示す。

3. 1. 2 背景画像生成法

3. 1. 2. 1 従来手法

背景画像を生成する従来手法の一つに背景積分法がある[1][2]。この手法は、次に示すような背景生成式によって背景生成が行われる。

$$b(t) = \alpha f(t) + (1 - \alpha)b(t-1) \quad (1)$$

ここで、 $b(t)$ は時刻 t におけるある画素の推定背景濃度であり、 $f(t)$ は時刻 t におけるフレームのある画素での濃度値である。背景画像 $b(t)$ を1つ前のフレームで生成された背景画像 $b(t-1)$ と現在のフレームの入力画像 $f(t)$ との加重平均によって生成する。 α は背景積分パラメータであり、 $0 \leq \alpha \leq 1$ の値をとる。 $(1 - \alpha)$ は前フレームまでの背景画像における情報の減衰量と考えることができる。 α の値を大きく設定すると背景自体の変化に強く、小さく設定すると背景の誤生成が生じにくくなる。よって、双方を満たす α の設定値の決定が問題となる。この手法では、入力画像中で移動物体の存在する時間的な割合が大きくないことが前提となっており、移動物体領域が大きい場合には安定した背景生成が行えないという問題がある。

3. 1. 2. 2 単純濃度ヒストグラム法

3. 1. 2. 1の手法は背景画像を逐次生成し更新する方法であるが、ここではある一定の動画から一つの背景画像を生成する手法について検討する。そのため、まず動画中における背景画素がどのような変動をするかについての検討を行う。実際に撮影した動画からある同位置画素の濃度変化を計測した。入力となる動画は、学内の庭を撮影したものである。記録する画素位置は、次の2つの位置を選択した。この2つの位置画素における濃度変化の様子を図1に示す。

(a) 移動物体が出現することのない常に背景である画素

(b) 移動物体が頻繁に出現する画素

(a)において濃度値は小さく振動しているが、これは画像取得時の雑音であり、動画中における背景とは濃度変動が十分小さいところと考えることが出来る。(b)では、常に濃度値が変動しているが、このように変動の大きな部分が移動物体部分であると推測される。

このように濃度変動する動画から背景を推定する手法として、ヒストグラムを利用する方法が考えられる[5][6][7]。画素位置毎の時系列データに対して濃度値をそれぞれ累積していき、時系列濃度ヒストグラムを作成する。そこから最頻度値を検出しその値を背景画素濃度とする[8][9]。(a)、(b)における時系列濃度ヒストグラムを図2に示す。

以上に述べた方法では、次のような問題点がある。

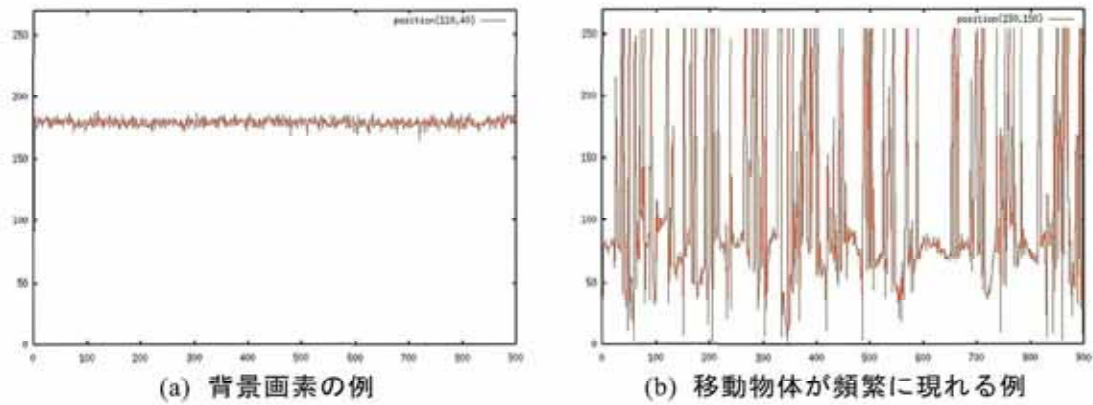


図1 動画像の同画素における濃度変化

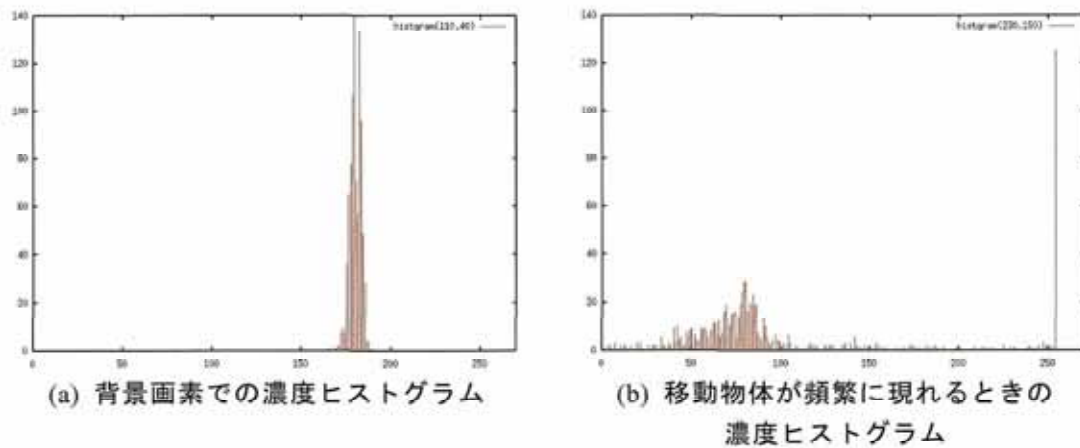


図2 時系列濃度ヒストグラム



図3 単純濃度ヒストグラム法による背景生成

- ・雑音の影響でヒストグラムに鋭いピークは現れない。
- ・ダイナミックレンジの影響によって濃度値 0、255 ではヒストグラムに累積値が集中しやすい。

これによって図 2(b)のヒストグラムでは濃度値 255 が最頻度値として検出され、誤認識を発生する。解決策として、次のことを行った。まず(a)における濃度値の変動の幅を調べたところ、濃度平均値から ± 5 の中にはほぼ全ての値が累積されている。そこで、画像取得時の雑音による影響は ± 5 以内でランダムであると仮定した。ヒストグラムでは ± 5 の幅で最も多く累積している部分が背景部分だと考えられる。このことから ± 5 の幅をもってヒストグラムを走査し、平均化したピークを求めることで雑音の影響に対して頑強に背景推定が行える。これを行った結果が図 3 である。

この結果から、ほぼ正しい背景が推定できていると考えられるが、図 3(c)において、白く囲

った移動物の存在する時間的割合が大きい部分では、誤生成を行っている。この手法では、動画像における背景の情報のなかで濃度情報のみを用いており、背景生成を行うには限界があると考えられる。そこで、時系列2次元エッジ差分に着目することにより、背景である信頼性の高い画素のみを選択し時系列濃度ヒストグラムに累積することで、安定した背景生成を行う手法を提案する。

3. 1. 3 時系列2次元エッジ差分を用いた背景画像生成

3. 1. 3. 1 空間エッジと時間エッジ

単純濃度ヒストグラム法では、ある画素位置の時系列データに対して、全ての濃度値を時系列濃度ヒストグラムに加算していったが、この方法では移動物の情報も加算しているため、移動物の存在する時間的割合が大きい部分では背景の誤生成が生じた。濃度値をヒストグラムに加算する際に、背景が出現しているフレームだけを選択して加算すれば、背景生成の精度が向上すると考えられる。ここでは、背景フレームを選択する方法として、時系列2次元エッジ差分を利用する方法を提案する。本手法では、動画像中における背景の特徴として、注目画素と周辺画素との関係を用いる。これには2つの関係が考えられる。すなわち、同フレームにおける空間的な情報と前後フレームにおける時間的な情報である。

これらの情報を用いる理由を以下に述べる。動画像中においてある画素が背景であるとき、その画素と周辺画素との空間的な関係では一定の相関関係が成り立つが、移動物体が背景を覆った場合では相関関係が崩れる。すなわち、その画素が背景である間は周辺画素との濃度差は一定であるが、移動物体が覆ったときは周辺画素との濃度差は変化する。その画素の前後フレームにおける時間的な関係では、背景であるときは、撮像系の雑音による緩やかな濃度変動をしており前フレームと現フレームとの濃度値には相関が高いが、移動物体が出現した場合での急激な濃度変動では相関はない。よって、相関が高いと考えられるときの情報のみを用いれば背景が推定できる。

本研究では2つの関係を空間エッジ、時間エッジとして扱う。これら2つのエッジの定義を行う。空間エッジとは隣接画素との差分であり、エッジ検出には、ノイズの影響を受けにくい Sobel オペレータを用いた。時間エッジとは同位置画素での前フレームとの差分である。移動物体が出現したとき、背景は数フレームにわたって覆われるので、前1フレームとの比較では、背景推定の有効性が低くなってしまふ。そこで、実験的に前5フレームの平均値との差分を用いることにした。時間エッジは、濃度値に対する値と、空間エッジに対する値の2つが考えられるが、ここでは空間エッジに対する値を用いる。動画像中で空間エッジ、時間エッジの双方を求めることで、時系列2次元エッジ差分が得られる。移動物体を含む動画像で背景が出現している部分では、空間エッジの値は一定であり、時間エッジの値は0となる。逆に移動物体が覆っているときは空間エッジの値は違う値をとり、時間エッジは0ではなくなる。この情報を用いることで背景であるフレームを選択する。

3. 1. 3. 2 提案手法の流れ

本手法の処理の流れを図4に示す。これは、ある1画素についての処理の流れである。まず空間エッジ、時間エッジの双方の値を求める。時間エッジが0であるときは、前フレームとの変化がおきていないことを表しており、そのフレームは背景である可能性が高い。よって、このときの空間エッジの値を時系列2次元エッジ差分ヒストグラムへ累積する。しかし画像取得時の雑音の影響から、背景が出現しているフレーム中でも緩やかな濃度変動がある。このため、時間エッジの値が厳密に0をとる場合は移動物体が存在しない場合でも少なくなってしまう。これに対応するために、雑音の影響による空間エッジ変動の振幅である ± 3 を実験的に求めた。そして時間エッジが ± 3 の値をとる場合に空間エッジの値を累積する。

以上の処理を全フレームについて行った結果得られる時系列2次元エッジ差分ヒストグラムから、最頻度値を検出する。この値はフレーム中の背景部分での空間エッジの値と定できるので、これを背景空間エッジとする。次に、全フレームにおいて背景空間エッジの値をもつとき

のフレームを検出する。検出されたフレームでの濃度値を、時系列濃度ヒストグラムに累積していく。そして最頻度値を検出することで、最終的な背景濃度が得られる。以上の処理を全画素についてそれぞれ行うことで背景画像を生成する。

以上が本手法による背景生成の基本的な流れになるが、今回、時系列2次元エッジ差分ヒストグラムの作成には、更に工夫を加えている。先に述べた時間エッジ情報は、1フレームずつ単独で求めているが、さらに有効に利用するため、時間エッジの継続情報を用いる。時系列2次元エッジ差分ヒストグラムに値を記録する際に、次の式によって求められる値を累積する。

$$h(\phi(x, y)) \leftarrow h(\phi(x, y)) + t; \tag{2}$$

ある画素 $\phi(x, y)$ において時間エッジの値が0であるフレームが t 回連続出現した場合、その部分は背景である信頼性がより大きくなる。そこでヒストグラム $h(\phi(x, y))$ には時間エッジが0であるフレームが連続出現した数 t を累積する。

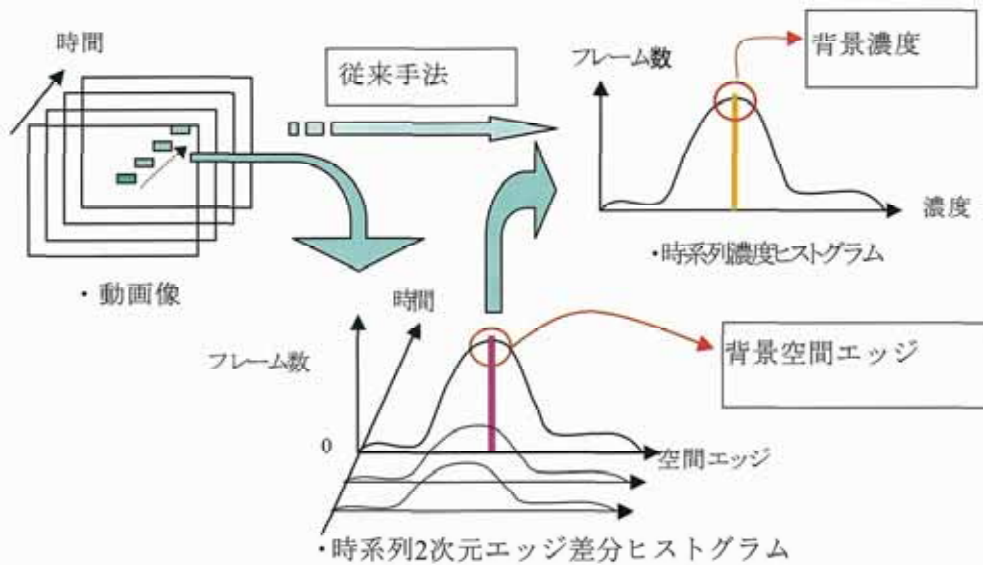


図4 提案手法

3. 1. 4 評価実験

3. 1. 4. 1 実験条件

ここでは、本手法を定量的に評価するために、あらかじめ背景画像を取得しておいた実験用動画を用意し、背景生成を行った。撮影環境を表1に示す。

表1 実験用動画の撮影環境

	(a) 研究室	(b) 廊下
フォーマット	256×256 (グレイスケール 256階調)	
フレームレート (frame/sec)	10	
フレーム数	1000	600
カメラ	固定 (Sony CCD-TRV80)	

3. 1. 4. 2 真の背景との比較評価

あらかじめ撮影しておいた真の背景画像と、背景生成によって作成された生成画像とで比較をおこなった。雑音による影響を回避するため、各画素に対して背景用の動画の50フレームから平均値を求め、その値を真の背景とした[10]。比較用手法としては、従来手法である単純濃度ヒストグラム法を用いた。背景生成を行った結果画像を図5に示す。従来手法において(a)、

(b)いずれの画像でも、ほぼ正しい背景が生成できているが、白丸で囲った部分において、誤生成が発生した。これに対して提案手法では、同じ部分についても正しい背景が得られている。

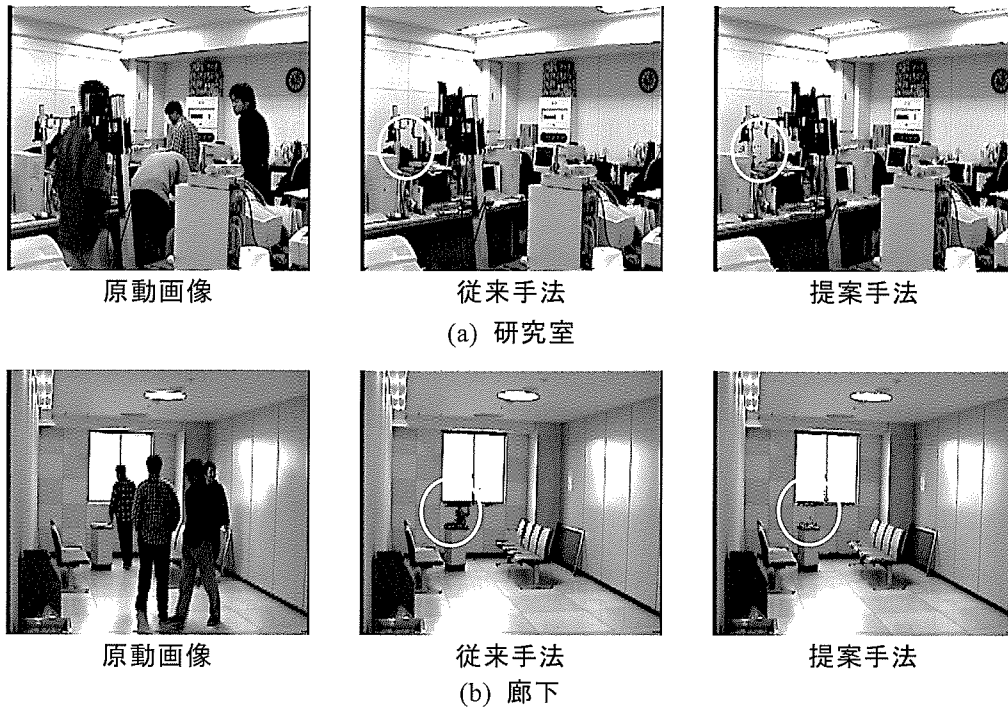


図5 背景画像生成結果

3. 1. 4. 3 定量評価

本手法を定量的に評価するために、次式によって、背景濃度誤差の平均値 T を求めた。

$$T = \frac{1}{MN} \sum_x^M \sum_y^N (\phi(x, y) - \varphi(x, y)) \quad (3)$$

この評価値 T は、真の背景画像 $\phi(x, y)$ と背景生成を行った結果画像 $\varphi(x, y)$ との各画素での濃度差を累積し、画像サイズ $M \times N$ によって正規化することで求めたもので、1画素あたりの背景との濃度差の平均値を表している。値が小さいほど真の背景との絶対的な濃度差が小さいということになり、より背景画像に近いと考えることができる。

結果を表2に示す。いずれの画像についても提案手法の方が小さな値となった。よって、定量的にも提案手法によって生成された画像のほうがより真の背景に近いと判断できる。(a)の画像において評価値が高めになっているが、これはあらかじめ撮影しておいた背景用の映像と、背景生成用の画像とで微妙な位置ずれが生じたためである。実際の評価値の差となって現れたのは、白く囲った部分の差によるものであると考えられる。

表2 従来手法との評価値の比較

画像種類	従来手法	提案手法
(a) 研究室	12.82	12.47
(b) 廊下	5.21	4.21

3. 1. 4. 4 他の動画像での実験と結果

3. 1. 4. 1～3. 1. 4. 3では、定量評価のためにあらかじめ背景が取得可能な映像を用意したが、ここでは他の屋内、屋外で実際に自然な環境で撮影した動画像に対して、背景生成を行った。撮影した動画像は、(c)駅前の交差点、(d)デパート展示会、(e)学内の庭とした。(c)、(d)では休日の昼時、(e)では平日の昼の休憩時間といった、人通りの激しい時間帯において

撮影した。背景生成した結果画像を図6に示す。従来手法においてそれぞれ誤生成を起こした部分が白く囲った部分において確認できる。しかし提案手法における同じ部分では、ノイズになっていたところが正しい背景として抜き出されていることが、視覚的に判断できる。(c)の動画画像では、交差点で信号待ちのために、移動物体であるべき歩行者が一時的に留まっていた。このために従来手法において誤生成が発生しているが、提案手法においては正しい背景が得られている。このことから、本手法では移動物体がある程度留まっている場合においても、正しい背景生成が行えることがわかる。

ここで扱った動画画像では、移動物体を含まない背景画像の撮影ができないため、視覚的な判断により、被験者によって主観的な評価をしてもらった。被験者は20歳代の学生の8人である。まず動画画像の様子を見てもらい、その後、従来手法と提案手法それぞれによって背景生成された画像について、どちらの移動物が適切に除去できているかを判断する。評価対象画像は3.1.4.2で用いた動画画像2種類と3.1.4.4で用いた動画画像3種類である。この結果を表3に示す。これらの評価結果を総合的にみると、提案手法の方がより背景画像に近いと判断できる。しかし画像によっては票が割れている。その画像は、従来手法においても十分に背景生成が行われているためと考えられる。

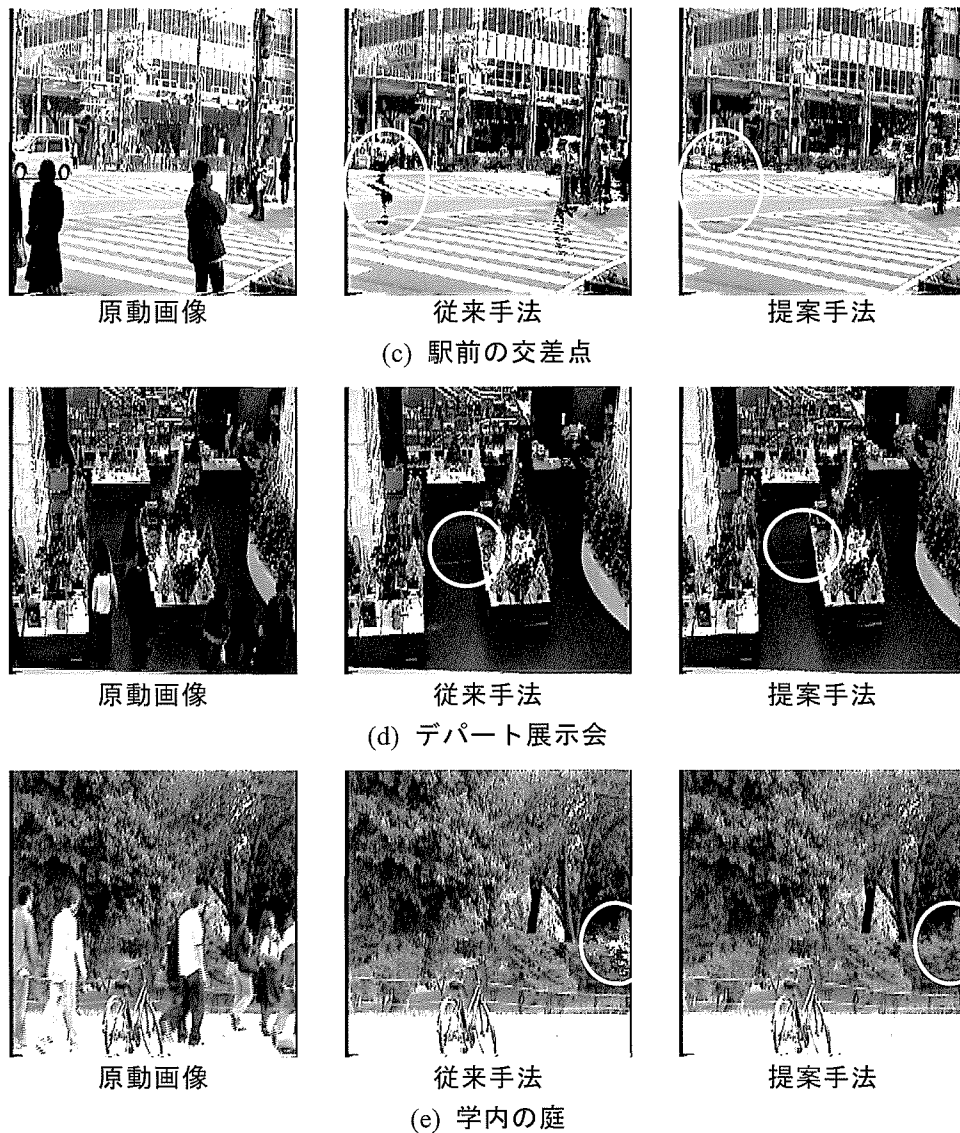


図6 他の動画画像における背景生成結果

表 3 被験者による相対評価

画像	(a)	(b)	(c)	(d)	(e)
従来手法	0	0	1	2	0
提案手法	8	8	7	6	8

3. 1. 5 まとめ

常に移動物体の存在する動画像から、安定して背景画像を生成する手法を提案した。提案手法は、濃度情報以外の背景情報として、空間エッジと時間エッジからなる時系列2次元エッジ差分を用いることによって、背景らしいフレームのみを選択してヒストグラムに累積していく事によって、ノイズの少ない背景画像を生成できることを示した。従来手法において、撮像系による雑音の影響を考慮することで、ある程度の精度で背景生成を行うことが出来たが、移動物体の存在する時間的割合の大きいところでは背景の誤生成が発生した。しかし本手法では、そのような部分に対しても正しい背景を抜き出すことが出来た。

今後の課題として、木々の葉等の自然物での背景自体の変動に対応する方法を検討したい。また、現段階では処理はオフラインにて行っているが、処理方法の改善、変更をすることで、オンラインにて背景生成を行うことも検討したい。

参考文献

- [1] 川端敦, 谷籐真也, 諸岡康男, “移動物体像の抽出技術,” 情処論, Vol.28, No.4, pp.395-402, 1987.
- [2] 佐久間喜郎, 伊東潔, 増田功, “フレーム間差分を用いた侵入物体検出法,” テレビジョン学会技術報告, IPCV '90-27, AIPS '90-50, pp.1-6, 1990.
- [3] 長屋茂喜, 宮武孝文, 藤田武洋, 伊藤渡, 上田博唯, “時間相関型背景判定法による移動物体検出,” 信学論, J79-D-II, pp.251-259, 2001.
- [4] 篠原利章, 都築健吾, 西澤真人, 林敏彦, 井上弘昭, “背景差分法における移動物体の検出精度の向上とカメラの揺れによる誤検出の低減手法の検討,” 第2回画像センシングシンポジウム講演論文集, pp.249-254, 1996.
- [5] P. Meer, D. Mints and A. Rosenfeld, “Robust regression methods for computer vision: A review,” International Journal of Computer Vision, Vol.6, No.1, pp.59-70, 1991.
- [6] P. J. Huber, “Robust Statistics,” John Wiley & Sons, 1981.
- [7] R. J. Rousseeuw and A. M. Leroy, “Robust Regression and Outlier Detection,” John Wiley & Sons, 1986.
- [8] 早坂光晴, 富永英義, “動画像からの背景画像生成を用いた移動物体抽出方法に関する一検討,” 情処研報, AVM29-1, 2000.
- [9] T. Kurita, K. Hotta and T. Mishima, “Scale and Rotation Invariant Recognition Method using Higher-Order Local Autocorrelation Features of Log-Polar Image,” In Proc. of 3rd Asian Conference on Computer Vision, Vol. II, pp.89-96, 1998.
- [10] M. Yamada, K. Ebihara and J. Ohya, “A new robust real-time method for extracting human silhouettes from color images,” In Proc. of IEEE Inter. Conf. on Automatic Face and Gesture Recognition, pp.528-533, 1998.

3. 2 2次元エッジベクトル差分による微動人物領域の検出に関する研究

3. 2. 1 はじめに

近年、安全に対する要求や治安の悪化などにより、映像監視システムの導入が進んでいる。また、コンピュータの処理速度の向上、記憶媒体の容量の増大に伴い、動画像処理が比較的容易に行えるようになった。そこで、動画像処理を用いて映像監視システムを自動化させる高度

監視システムの研究が盛んに行われている。監視システムの自動化の役割として、映像中から人間を見つける人物検出技術がある[11][12]。人物検出の基本的な手法として、背景差分法とフレーム間差分法の二つがよく用いられる。

背景差分法では、あらかじめ背景画像を作成しておき、入力画像と背景画像との差分をとる。差分の大きい部分が人物部分として検出される。背景の環境が安定している場合では、効率よく人物を検出できる。しかし実際には、屋内においては照明状態の変化や、机や椅子などの移動によって背景自体が変化すること、屋外においては日照状態の変動などにより、背景画像は安定していない。このように動的に変化する環境において背景差分法を行うために、照明状態の変化や新しい背景物体の出現などを感知し、適応的に背景を更新する技術が提案されている[13][14][15]。しかし、背景の変化を感知するまでは誤認識が発生しやすいなどの問題があり、幅広い環境での実利用にまでは至っていない。

一方、フレーム間差分法では、現在の入力画像と N フレーム前の画像との差分をとり、差分の大きい部分を人物部分とする。通常は $N=1$ であり、つまり一つ前のフレームと現在のフレームを比較する。この手法では、常に背景更新を行っていることになり、背景自体の変化は無視できると考えてよい。すなわち、背景シーンの環境の変化に強い手法であるといえる[16][17]。しかしながら、検出対象となる人物が、映像内で静止している場合は、人物部分に差が現れず、検出が不安定になるという欠点がある。

さて、図書館やオフィスなどにおいて、人物検出を行うことで、空調や照明などを効率的に利用しようという試みがある。このようなシーンでは、机や椅子の移動、照明の変化により背景は常に変動しており、かつ検出対象である人物は、読書やデスクワークのため、一箇所に長時間留まっていると考えられる。このようなシーンで背景差分法を用いる場合、人物部分の背景が出現していることがないために、適切な背景画像が得られない。よって正しく人物検出を行うことは難しい。また、フレーム間差分法を用いる場合、人物が同じ場所に留まっているときは検出されない。

本研究では、人間が静止している場合、完全に静止している訳ではなく、微動していることを利用する。人間の微動を捉えることにより、長時間留まっている人物を検出可能にすることが、本研究の目的である。本研究では、時系列2次元エッジを用いることで、画素ごとの空間的・時間的変化量を求める。さらに精度よく変化を捉えるため、エッジ方向を組み合わせた手法を提案する。提案手法により、人物の微動状態を捉えることが可能であることを、性能評価実験によって確認する。さらに、実際の監視状況を想定した図書館などの映像によって、本手法の応用例などを示す。

3. 2. 2 人物検出手法

ここでは、従来の人物検出手法であるフレーム間差分法と、それを発展させた提案手法について述べる。

3. 2. 2. 1 フレーム間差分

図7にフレーム間差分の概要を示す。フレーム間差分とは、現在のフレームと、 N フレーム前の画像とで差分をとり、差の大きい部分を人物領域として検出する手法である。通常は、 $N=1$ を用いる。次式で定義される。

$$\delta(t) = f(t) - f(t-1) \quad (4)$$

ここで、 $f(t)$ は現フレームの画素値であり $f(t-1)$ は前フレームの画素値である。背景が出現しているときは、その画素の濃度値はほとんど変動しない。移動物体が出現した場合は、濃度値は前フレームに比べ大きく変動する。従って、現在のシーンに移動物体が存在している条件は、次式で示される。

$$|\delta(t)| \geq Th_{scene} \quad (5)$$

判定しきい値 Th_{scene} は、撮影した映像における濃度の量子化誤差から定める。映像中で、検出対象である人物が常に移動している場合には、ほぼ確実に人物が検出できる手法である。しかし、人物が静止している場合は、前後フレーム間で差が得られないため、人物検出することは難しい。

人物が微動状態であった場合を考える。前後フレーム間で人物が1画素移動していたとする。フレーム間差分では、画素ごとの濃度差を計算しているため、人物の輪郭部分が検出される。しかし、人物の内部は大きな濃度差は得られず、検出されない。この手法では、映像中の変化部分を求めるのに、フレーム間の濃度情報のみを用いるため、小さな変化は捉えることができない。そこで本手法では、より詳しく変化を捉えるために、空間エッジ情報とエッジ方向の二つの特徴量を用いる。

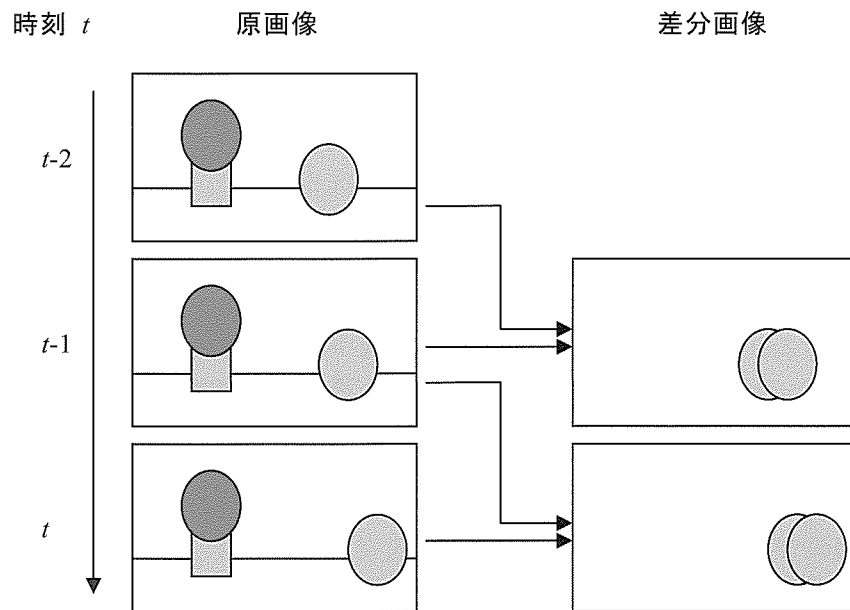


図7 フレーム間差分法

3. 2. 2. 2 空間エッジとエッジ方向

空間エッジとは、画像平面上における注目画素とその周辺画素との濃度差を表している。注目画素が背景部分である場合、この画素における局所的な空間差分値は、濃度値と同様にほぼ一定の値をとる。しかし、移動物体として人物が横切る場合、衣服の皺や影によって、局所領域内で僅かな濃度差が生じる。小さな濃度差では注目画素一点だけを比較した場合では捉えられないが、空間エッジでは、局所領域内で空間的に変化をみているので、濃度差よりもより明確に変化部分を捉えることができると考えられる[18]。

エッジ方向とは、空間差分によって求められたエッジ部分のエッジ勾配を表している[19]。エッジ勾配はエッジ点 (x, y) の水平からの角度 $\theta(x, y)$ で表す。角度 θ は濃度値を $f(x, y)$ とすると、式(6)を用いて求められる。

$$\theta(x, y) = \tan^{-1} \left(\frac{\partial f(x, y)}{\partial y} / \frac{\partial f(x, y)}{\partial x} \right) \quad (6)$$

注目画素が空間エッジ値のある背景部分である場合、エッジ方向は安定して同じ方向を向いている。しかし、注目画素が人物部分であった場合、人物がわずかに動いただけでも、衣服の皺では、エッジ方向が大きく変動すると考えられる。よって、空間エッジ値では大きな差を生じなかった部分でも、エッジ方向を見た場合には人物部分では大きく変動していることがあり、

その部分の変化をエッジ方向では捉えることができる。提案手法では、この二つの特徴量を組み合わせ、ベクトルとして考える。図8に示すように、このベクトルはエッジ強度を絶対値にもち、向きがエッジ方向である。これをエッジ方向ベクトルと呼ぶことにする。

提案したベクトルをフレーム間で画素ごとに比較する。その差が大きい部分が人物部分であると推定される。この二つの特徴量を用いることで、フレーム間差分よりも詳しく変化を捉えることが期待される。

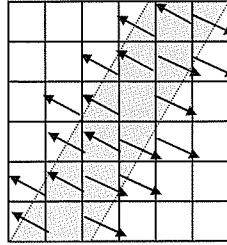


図8 エッジ方向ベクトル

3. 2. 2. 3 提案手法

図9に提案手法の概要を示す。各フレームにおいて、入力である濃淡画像から、エッジ検出を行い、エッジ画像を求める。同時に、式(6)により各画素のエッジ方向も求める。エッジ検出には次に示すような5x5のフィルタを用いた。

$$h_x = \begin{bmatrix} -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \end{bmatrix}, h_y = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (7)$$

これにより、注目画素一点に対して、周囲24画素の空間的な変化を捉えることが出来る。本研究では、人物は同一場所に長時間留まっている場合を想定している。そのため、1フレーム前の画像と現在の画像との2枚の画像で比較するよりも、複数のフレームとで比較するほうが、変化をより捉えることができると考えられる。本手法の利用目的を考えれば、映像内を高速に通過するような移動物体は検出対象外と考えられるので、前後複数フレームで比較を行っても問題はないと思われる。次式により、画素ごとの差分値を求める。

$$\phi(t) = g(t) - \gamma \quad (8)$$

$$\gamma = \frac{1}{N} \sum_{i=1}^N g(t-i)$$

今回は実験的に $N=5$ と定めた。時系列での差分値は映像に依存するところが大きい。定義式の改良次第で、更なる改善の余地があると考えられる。上式において、時系列エッジ方向ベクトル差分 $\phi(t)$ は、現フレームのエッジ方向ベクトル $g(t)$ とその前5フレームの平均ベクトルとの差分を意味する。この値が次式(9)を満たす場合、その画素は人物領域であると推定される。

$$|\phi(t)| \leq Th_{edge} \quad (9)$$

判定しきい値 Th_{edge} は、量子化誤差や画像キャプチャ時のノイズによる影響で考えられる最大誤差とする。予備実験により移動物体の映っていない照明状態の安定した環境で撮影した映像を用いて、エッジ方向ベクトルの変動幅を計測し、その最大値を判定しきい値 Th_{edge} に設定した。

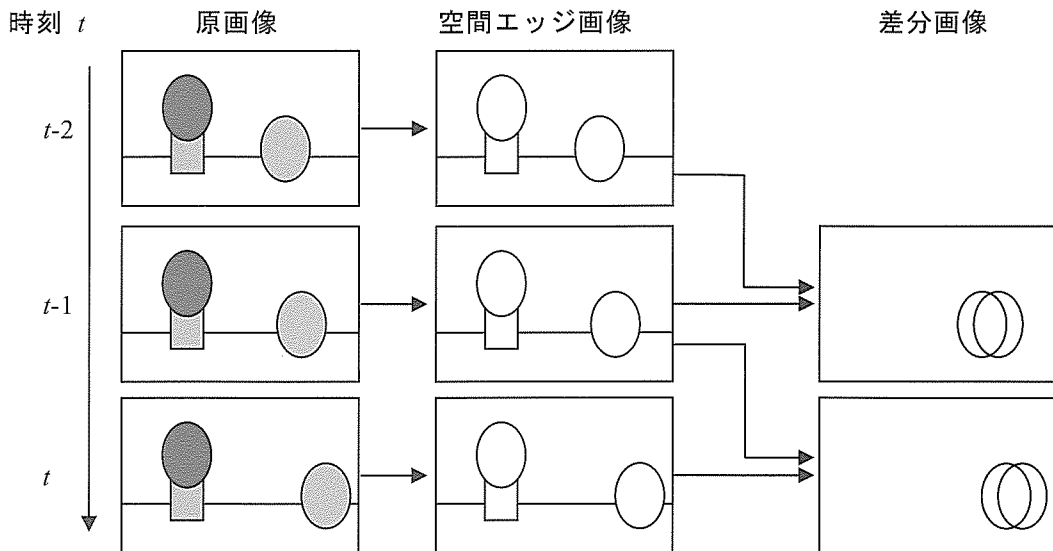


図9 エッジ方向ベクトル差分法

3. 2. 2. 4 累積処理

以上の処理によって、人物の微動状態を捉えることが期待される。本手法では、空間エッジを用いているため、照明変化に対して、フレーム間差分よりも頑健性を持っていると考えられるが、背景部分において局所的な濃度変化が起こった場合、その部分の変化を捉えてしまう。こういった誤検出部分と、人物部分とを区別するために、検出部分の累積処理を行う。本手法によっても映像内で全く変動のないタイミングでは人物検出は困難であるが、長期的に映像を見れば、人物は微動をしているので、検出が可能になると考えられる。そこで、ある程度のフレーム数から検出部分を累積し、閾値以上の検出数があった領域を人物領域とする。必要なフレーム数は、対象となる映像によって差があると考えられる。次節の性能評価実験において、必要なフレーム数の検討を行う。

3. 2. 3 性能評価実験

本手法では、空間エッジを基にしているので、人物検出の性能は、その人物の衣服の種類によって差が生じる。また人物の留まっている状態によっても性能差が生じる。そこで、実際にどの程度人物抽出が可能であるのかを確認するため、次のような実験を行った。人物の姿勢によってシーンを分類し、それぞれのシーンで、フレーム数ごとに人物検出状態を評価する。それぞれ比較手法として、フレーム間差分法による人物検出も行った。撮影場所は実験室内で、安定した照明状態である。カメラは Sony-CCD-TRV80 を用いた。画像サイズは 320x240、階調数は 256 階調のグレースケールである。また実験条件を表 4 に示す。

表 4 実験環境

画像サイズ	320×240
階調数	256label [gray-scale]
フレームレート	10 [frame/second]
カメラ	固定 [Sony CCD-TRV80]

図 10 に、実験映像の例を示す。実験室内にて、人物が椅子に座っている映像である。左が原画像、右がエッジ画像となっている。被験者には、意識的に静止するように指示をしてある。このシーンから、視覚的に変動の見られない連続したフレームに対して、人物検出を行った。図 11 に検出結果を示す。上から順にフレームを累積させていった結果画像である。使用フレーム数は、10、50、100 となっている。フレーム間差分法では、100 フレーム累積した結果画像でかろうじて目の辺りが検出できているが、その他のノイズ部分と検出した大きさが変わらない。

提案手法では、フレームを累積するごとに上半身部分が検出されていることが分かる。足の部分は、椅子と床とで固定されており非常に安定した状態である。そのため全く変動が見られず、本手法によっても検出は困難であった。しかしながら、上半身の特に頭部では、完全に静止させておくことは出来ず、微動している状態であったと考えられる。そのため、本手法によって検出することが出来ている。このことから、人物が映像内で静止している場合でも、本手法によれば50~100フレーム程度の累積処理を行えば検出が可能であると考えられる。

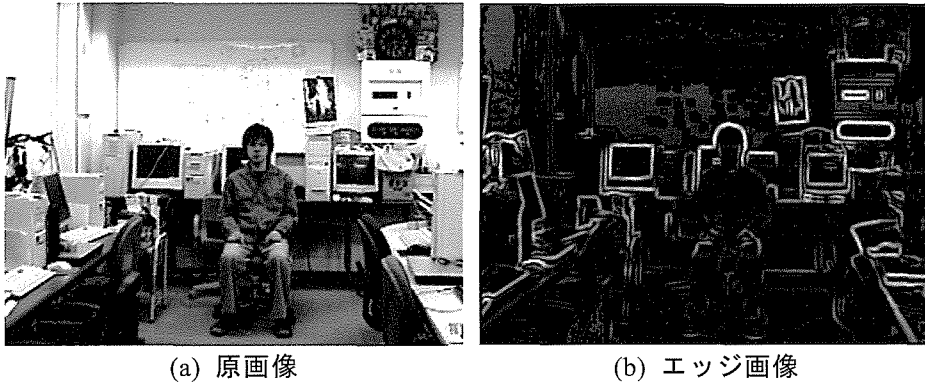


図10 実験映像

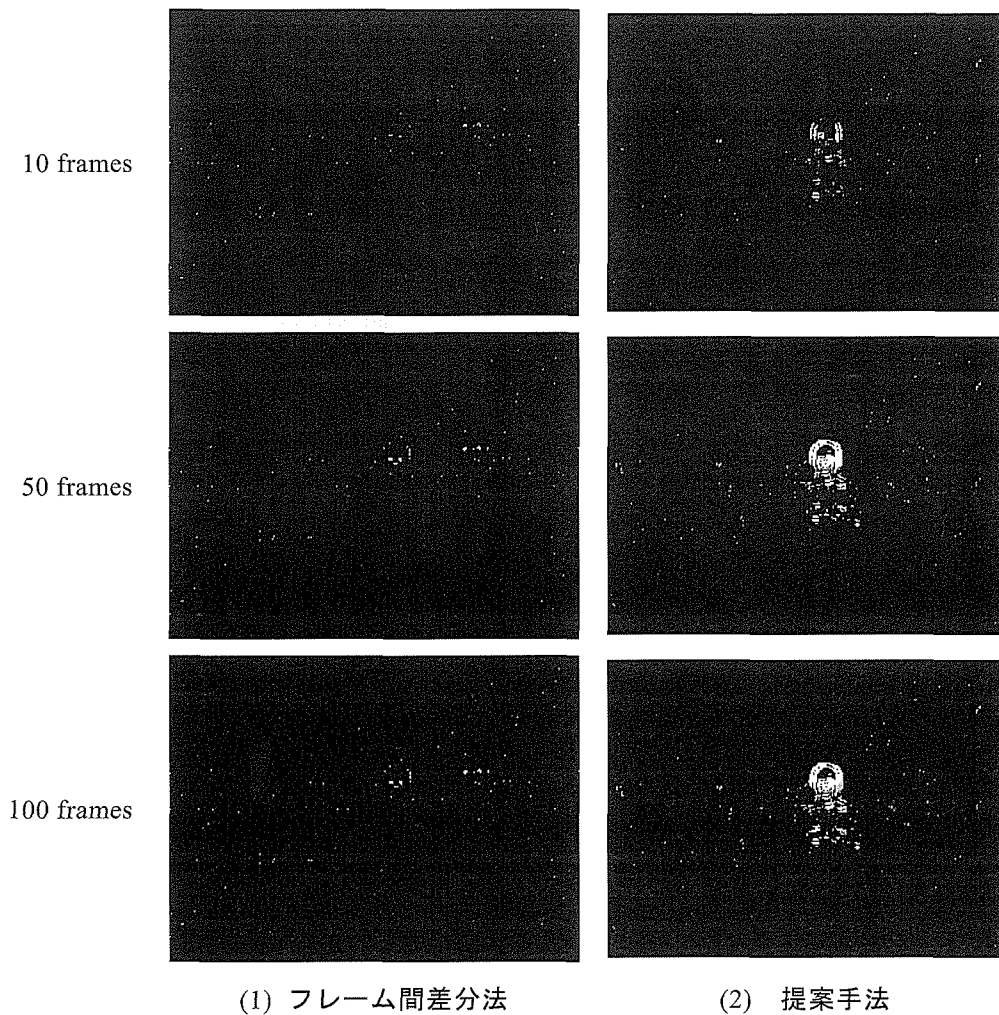


図11 人物検出結果

また、同じシーンに対して視覚的に変動がみられたフレームでの人物検出結果を図 12 に示す。フレーム間差分では、人物の輪郭が検出されている。しかし本手法では、空間エッジによって衣服の皺などの変化を捉えているため、人物の内部まで検出されている。次のシーンは、人物が直立している状態である(図 13)。これも先程と同じく、視覚的には変化が見られない映像を用いて実験を行った。実験結果を図 14 に示す。フレーム間差分法では、人物の輪郭部分が徐々に検出されているが、わずかな部分しか検出されておらず、そこから全体を把握するのは困難である。一方提案手法では、10 フレームで、人物全体を認識するのに十分な領域が検出されている。人物が立っている状態は完全な静止状態ではなく、絶えず微動していることが結果から推測される。

以上の結果より、次のことが言える。人物が椅子に座っており安定した姿勢で静止している場合でも、上半身の特に頭部では微動を起こしている。提案手法では、50~100 フレーム程度の累積を行うことで、検出が可能である。人物が立っている状態では、全体的に微動を起こしており、提案手法を用いると 10 フレーム程度で検出することが出来る。

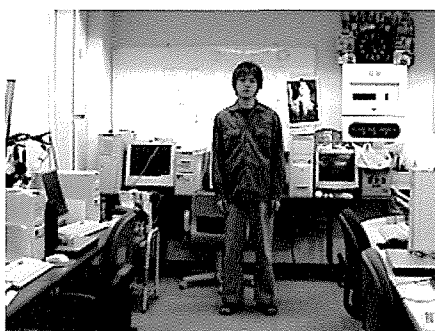


(1) フレーム間差分法



(2) 提案手法

図 12 変動時の人物検出結果



(a) 原画像



(b) エッジ画像

図 13 実験映像

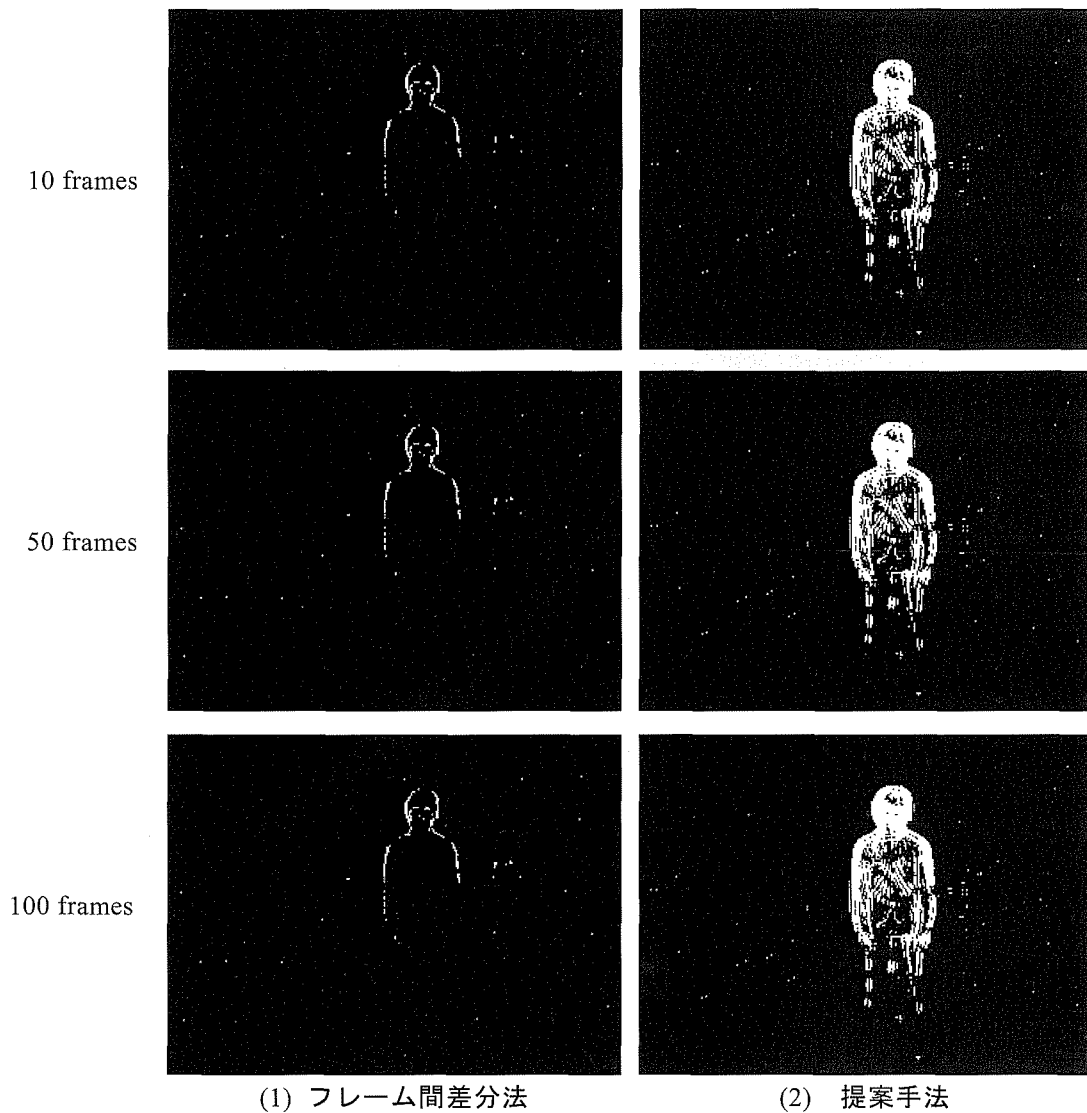


図 14 人物検出結果

3. 2. 4 実映像による実験

実験室外で撮影した映像に対して人物検出を行った例を示す。図 15(a)は、図書館にて学生が自習している様子である。このような映像に対して人物検出を行った結果を図 15(b)に示す。検出結果画像は 100 フレームの累積処理を行った画像である。学生は机の前に座っている。結果画像から人物部分を検出できていることが分かる。しかしながら、画面左側の円で囲った部分にいる学生は、撮影中に居眠りをしており、全く動くことがなかった。そのため本手法によっても検出することは出来なかった。窓の部分では、日照の急な変化により、時折、誤検出することが見られたが、本手法では累積処理を行ったことで、結果画像ではノイズとなる部分を除去することが出来ている。

図 16(a)は、教室の様子を撮影した映像である。このような映像に対して本手法を適用した結果を図 16(b)に示す。結果画像は、先程と同じく 100 フレームの累積処理を行った画像である。この映像では、全ての人物が検出できている。しかしながら、人物の足部分は検出できていないところがある。これは、椅子に座っている状態では、足は床と椅子でしっかりと固定されており、動きがなかったためと考えられる。また今回の映像では全員が検出できているが、仮に学生が居眠りをしていた場合では、動きがなくなるために、やはり検出は難しいと考えられる。

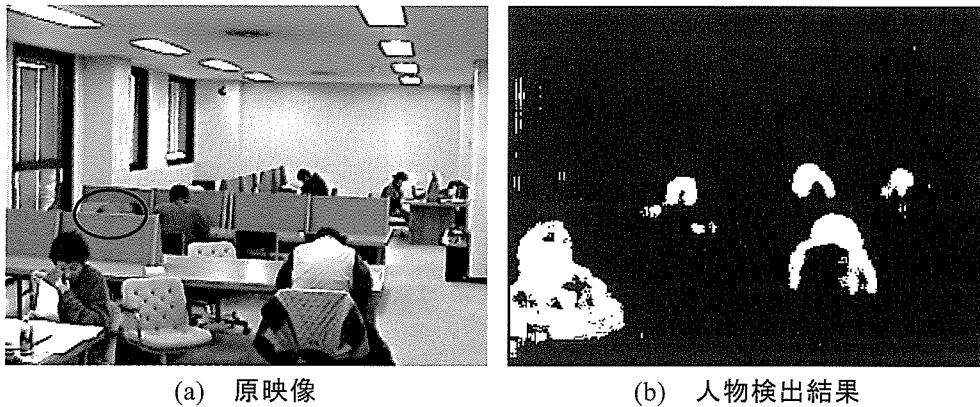


図 15 図書館の映像例と検出結果

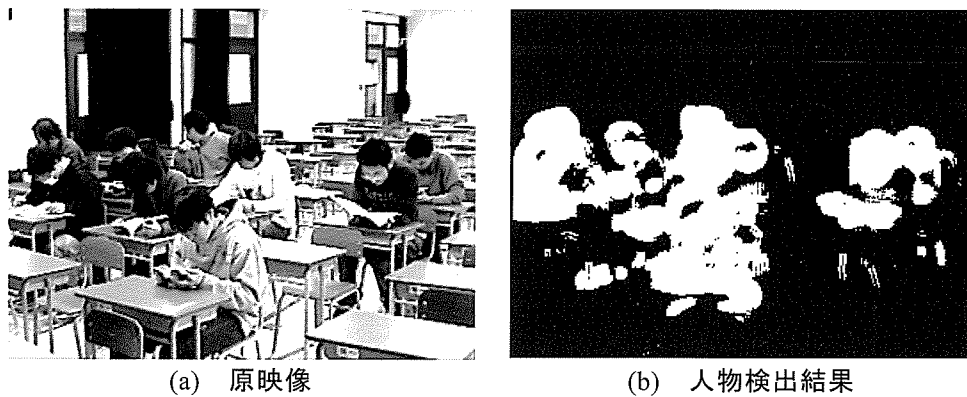


図 16 教室の映像例と検出結果

3. 2. 5 まとめ

人物が長時間留まっている映像から、人物検出を可能にする手法を提案した。本稿では、人物は完全に静止することは出来ず、必ず微動しているという性質に着目し、わずかな変化を捉えることで、人物検出を可能にした。画像の空間情報である空間エッジと、その勾配の向きを表すエッジ方向を成分にもつエッジ方向ベクトルを用い、この変化を検出することで濃度だけでは検出できないようなわずかな変化を捉えることが出来た。図書館や教室の映像を用いて実験を行った結果、人物部分を検出できていることが確認された。

今後の課題としては、完全に静止している人物への対応、人数計測や人物追跡への応用などが挙げられる。

参考文献

- [11] 川端敦, 谷藤真也, 諸岡康男, “移動物体像の抽出技術,” 情処論, Vol.28, No.4, pp.395-402, 1987.
- [12] 早坂光晴, 富永英義, “動画像からの背景画像生成を用いた移動物体抽出方法に関する一検討,” 情処研報, vol.AVM29-1, 2000.
- [13] 長屋茂喜, 宮武孝文, 藤田武洋, 伊藤渡, 上田博唯, “時間相関型背景判定法による移動物体検出,” 信学論, vol.J79-D-II, No.4, pp.251-259, 2001.
- [14] 高藤政雄, 北村忠明, 小林芳樹, “空間微分及び差分処理を用いた車両抽出法,” 信学論, Vol.J80-D-II, No.11, pp.2976-2985, 1997.
- [15] 島井博行, 三島健稔, 栗田多喜夫, 梅山伸二, “移動物体検出のためのロバスト統計を用いた適応的な背景推定法,” 画像の認識・理解シンポジウム (MIRU2000) 講演論文集 I, pp.391-396, 2000.
- [16] 佐久間喜郎, 伊東潔, 増田功, “フレーム間差分を用いた侵入物体検出法,” テレビジョン学会技術報告, IPCV'90-27, AIPS '90-50, pp.1-6, 1990.

- [17] 馬場功淳, 大橋健, 乃万司, 松尾英明, 江島俊朗, “Head Finder: フレーム間差分をベースにした人物追跡,” 画像センシングシンポジウム 2000, pp329-334, 2000.
- [18] 大西正輝, 泉正夫, 福永邦雄, “講義映像における板書領域のブロック分割とその応用,” 信学論, Vol.J-83-D-I, No-11, pp.1187-1195, 2000.
- [19] 井上誠喜, 八木伸行, 林正樹, 中須英輔, 三谷公二, 奥井誠人, “C 言語で学ぶ実践画像処理,” オーム社, 1999.

3. 3 3 次元ポリゴンモデルを用いたステレオ人物頭部追跡に関する研究

3. 3. 1 概要

視覚を用いて人間の動作を知覚することは、機械と人間が円滑な相互作用を行うために必要不可欠である。我々は、ステレオカメラを用いて人物の頭部を追跡する方法を開発した。ノイズを含んだ環境における対象追跡問題はベイズ的枠組で定式化される。ベイズフィルタでは、視覚入力を得られた条件のもとでの、対象の位置の事後確率分布を推定する。ノイズがガウス分布としてモデル化される場合、この枠組はカルマンフィルタによって解かれるが、多くの場合、ガウス分布は良い近似とはならない。これに対して、非ガウスでモデル化するために、事後確率分布を参照点と重みによって表現するパーティクルフィルタ (CONDENSATION) が提案されている。本研究では、ステレオカメラから得られる時系列距離画像の中で、人物頭部の位置と姿勢を推定するために、パーティクルフィルタを用いた。頭部の位置・姿勢の事後確率分布は3次元ポリゴンモデルの重み付き集合によって表現される。人物の頭部は肌色や頭髮の色などを色情報として持つポリゴンモデルとしてあらわされる。必要なパーティクルの個数は対象の次元に対して指数関数で増加するため、頭部の位置追跡と位置推定は別の処理として行った。実験の結果、ノイズのある環境において、頑健な頭部追跡が可能であることが示された。

3. 3. 2 はじめに

視覚を用いた人間の動作解析は、ロボットと人間の知的な相互作用を実現するために、必要不可欠な要素である。動作解析における重要な課題として、頭部追跡と頭部の姿勢推定がある。この課題に対して、多くの手法が提案されている。

視覚による動作追跡[31]において、モデルを用いた方法は一般的であるが、この方法は2次元モデルを用いる方法と3次元モデルを用いる方法の2つのカテゴリーに大別される。2次元モデルを用いる方法としては、エッジなどの画像特徴を用いる方法[26]、固有顔を用いる方法、そして類似度に基づく方法[36]などがある。Isardらは2次元画像中で人間の頭部を追跡するためにB-Spline曲線をモデルとして用いた[26]。この手法では、移動による頭部の姿勢変化は画像中における輪郭のアフィン変換として現される。しかしながら、2次元画像上で3次元の動作を正確にモデル化することは困難である。この問題はモデルとして3次元モデルを使用することによって解決される。この手法では、観測された画像特徴と2次元上に投影された3次元モデルを比較することによって頭部の状態を推定する。この手法は合成による解析 (analysis-by-synthesis) として知られる。円筒[22][37]、楕円体[20]、ポリゴンメッシュ[21][28][33][40][39][23]など様々な種類の3次元モデルを用いた手法が提案されている。Casciaらはテクスチャマッピングを行った円筒をモデルとして用いて頭部追跡を行った[22]。また、Scholdらは同様なテクスチャマッピングの手法をポリゴンモデルに対して適用した[35]。

視覚的な追跡問題における不確かさを扱う上で、ベイズ的枠組みがしばしば持ち出される。視覚的不確かさは照明やオクルージョン、背景の乱雑さなど様々な要因によって発生する。ベイズ的枠組みでは、このような視覚的不確かさをモデル化するだけでなく、人間の歩行時には頭部が一定の高さを保持するというような事前知識も同時に統計モデルとして用いる。ベイズフィルタは視覚入力を得られたときの対象の姿勢の事後確率分布を推定する。もしノイズがガウス分布で近似され、ダイナミクスが線形であれば、事後確率分布は解析的に求めることができる。この方法はカルマンフィルタとして知られる。頭部追跡問題にカルマンフィルタを用いる手法が提案されている[32]。しかし、追跡問題においてノイズをガウス分布によって近似す

るよりも、非ガウスモデルで近似するほうが現実的である。パーティクルフィルタ (CONDENSATIONアルゴリズムとしても知られる) はモデルが非ガウスである場合に用いられる手法である。多くの研究者が人物追跡問題に対してパーティクルフィルタを用いている [36][37][26][25][27][24]。パーティクルフィルタの特徴は非ガウスな状態空間を重み付けされたパーティクルの集合によって近似的に表現可能な自由度を持つことである。

追跡問題においては単眼視覚がよく用いられ、ステレオ視覚が用いられることは少ない [32][27]。しかしながら、ステレオ視覚を用いて得られる奥行き画像は、乱雑な背景や照明の変化に対して頑健なため、利用価値は高い。RussakoとHermanは奥行き画像を用いて頭部追跡を行った。

我々は奥行き画像と確率的に生成された3次元のポリゴンモデルを比較することによって頭部の3次元追跡を行う方法を提案する。

3. 3. 3 提案手法の概要

我々はステレオ画像から得られた奥行き画像系列から頭部追跡を行うためにパーティクルフィルタを適用する。提案手法の特徴は以下の通りである。

- ・ 頭部の姿勢の事後確率分布は重み付けされたポリゴンモデル仮説集合によって表される。図 17 はポリゴンモデル仮説集合の概念図である。重み付けを持つそれぞれのポリゴンがその姿勢での頭部の存在確率に相当する。
- ・ 頭部は肌色や頭髪に関する色情報を含んだポリゴンメッシュによって表される。幾何学的な情報と色情報はそれぞれ、位置追跡と方向推定に用いられる。
- ・ 位置追跡と方向推定は別に行われる。このためパーティクルの数を減らすことができる。一般に状態空間が増えると必要なパーティクルの数は指数関数的に増えることが知られている [38]。我々の手法では頭部の位置推定がまず行われ、次にその位置に基づいた方向推定が行われる。

また、我々の提案手法は下記の仮定に基づいている。

- ・ ステレオ視覚のパラメータと光学系のモデルは事前に分かっている。これらのパラメータは奥行き画像を生成するときと観測モデルを生成するときに必要なものになる。
- ・ 人間の頭部の幾何学的モデルと色のモデルは事前に分かっている。

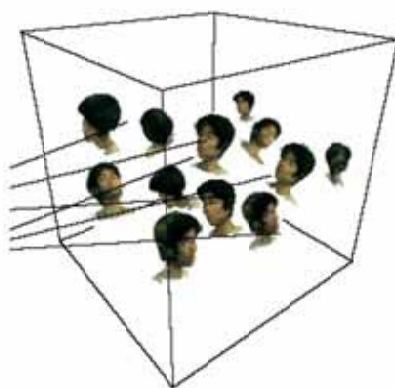


図 17 Hypothesized polygon based human head tracking

【パーティクルフィルタ】

3次元顔追跡は、時刻 t までに得られる奥行き画像系列 $\mathbf{O}_t = (o_1, \dots, o_t)$ を用いて、人間の頭部の姿勢 $\mathbf{X} = [x, y, z, \theta_x, \theta_y, \theta_z]^T$ の事後確率分布を求めることである。もし観測対象がマルコフ性、すなわち $p(\mathbf{X}_t | \mathbf{X}_{t-1}, \dots, \mathbf{X}_0) = p(\mathbf{X}_t | \mathbf{X}_{t-1})$ を満たすなら、事前確率分布 $p(\mathbf{X}_t | \mathbf{O}_{t-1})$ は次のように表される。

$$p(\mathbf{X}_t | \mathbf{O}_{t-1}) = \int p(\mathbf{X}_t | \mathbf{X}_{t-1}) p(\mathbf{X}_{t-1} | \mathbf{O}_{t-1}) d\mathbf{X}_{t-1}$$

ここで $p(\mathbf{X}_t | \mathbf{X}_{t-1})$ は状態遷移確率である。そして事後確率分布はベイズのルールを用いて、

$$p(\mathbf{X}_t | \mathbf{O}_t) = \kappa p(\mathbf{O}_t | \mathbf{X}_t) p(\mathbf{X}_t | \mathbf{O}_{t-1})$$

のように表すことができる。ここで κ は \mathbf{X}_t に独立な正規化定数である。 $p(\mathbf{O}_t | \mathbf{X}_t)$ は観測モデルである。事後確率分布はモデルが非ガウス分布の場合には解析的に計算できないので、確率分布を近似するためにパーティクルフィルタが用いられる。パーティクルフィルタの基本的な考え方は、確率分布 $p(\mathbf{X}_t | \mathbf{O}_t)$ を重み付けされたサンプル集合 $\{s_t^{(n)}, \pi_t^{(n)}; n=1, \dots, N\}$ によって表すというものである。このサンプル集合を再帰的に更新することで事後確率分布を推定する。このプロセスは以下の通りである。

1. $p(\mathbf{X}_{t-1} | \mathbf{O}_{t-1})$ の分布に従ってサンプリングを行う。

$$\{s_{t-1}^{(n)}; n=1, \dots, N\} \sim p(\mathbf{X}_{t-1} | \mathbf{O}_{t-1})$$

2. このサンプル集合に対して状態遷移を計算する。

$$p(\mathbf{X}_t | \mathbf{X}_{t-1} = s_{t-1}^{(n)})$$

その結果 $s_t^{(n)}$ が得られる。

3. 新たな重みを次のように計算する。

$$\pi_t^{(n)} = p(\mathbf{O}_t | \mathbf{X}_t = s_t^{(n)})$$

3. 3. 4 観測モデル

尤度関数 $p(\mathbf{O} | \mathbf{X})$ は、頭部が姿勢 \mathbf{X} にあるときに観測される奥行き画像 \mathbf{O} の確率分布を表す。この尤度関数を得るために、ステレオカメラの構成、頭部の形状、ステレオマッチングの不確かさをモデル化する必要がある。上述のように、提案手法では頭部追跡と方向推定は別に行われるので、この尤度関数は次のように表すことができる。

$$p(\mathbf{X} | \mathbf{O}) = p(\mathbf{X}_p | \mathbf{O}_d) p(\mathbf{X}_o | \mathbf{O}_c)$$

ここで、 $\mathbf{X}_p = [x, y, z]^T$ 、 $\mathbf{X}_o = [\theta_x, \theta_y, \theta_z]^T$ である。 \mathbf{O}_d はステレオカメラから得られる奥行き情報、 \mathbf{O}_c は奥行き画像の色情報である。

頭部追跡のための尤度関数は、奥行き画像を評価することによって得られる。

$$p(\mathbf{O}_d | \mathbf{X}_p)$$

奥行きの値はカメラのノイズ、キャリブレーションの誤差など様々なノイズによって影響を受ける。奥行き画像に関するノイズをモデル化する方法がいくつか提案されている[30][29]が、我々は次のようにモデル化する。

$$\Sigma = \begin{pmatrix} \sigma_x & 0 & 0 \\ 0 & \sigma_y & 0 \\ 0 & 0 & \sigma_z \end{pmatrix}$$

頭部が N 個のポリゴンによって構成されているとする。

$$\mathbf{M} = \{m_j; j = 1 \dots N\}$$

ポリゴン片 m_j に対応するピクセル \mathbf{o}_p (3次元ベクトル) が奥行き画像から検出される確率は、3次元正規分布を用いて次のように表すことができる。

$$p(\mathbf{O}_d | \mathbf{X}_p) = \prod_{j=1}^N \mu_j p_j$$

ここで

$$p_j = \frac{\exp(-1/2(\mathbf{o}_d - \mathbf{x}_{pj})^T \Sigma^{-1}(\mathbf{o}_d - \mathbf{x}_{pj}))}{\sqrt{(2\pi)^3 |\Sigma|}}$$

である。 \mathbf{o}_d は奥行き画像中に観測されたピクセルの3次元ベクトルである。 \mathbf{x}_{pj} は m_j の3次元位置ベクトルである。 $|\Sigma|$ は共分散行列の行列式である。 μ_j はカメラから見える面のポリゴン片だけを抽出するための係数であり、次のように求める。

$$\mu_j = \begin{cases} 1 & \mathbf{n}_o \cdot \mathbf{n}_p^j < 0 \\ 0 & \text{otherwise} \end{cases}$$

ここで \mathbf{n}_o は光軸である。 \mathbf{n}_p^j は j 番目のポリゴン片の法線ベクトルである。 \mathbf{o}_d は次の式によって選ぶ。

$$\mathbf{o}_d = \arg \min_j \|\mathbf{o}_{di} - \mathbf{x}_{pj}\|$$

方向推定のための尤度関数は、奥行き画像の色を評価することによって得られる。尤度関数 $p(\mathbf{O}_c | \mathbf{X}_o)$ は、頭部が \mathbf{X}_o にあるときに \mathbf{O}_c が観測される確率分布である。これは、仮説として生成されたポリゴンと実際に得られた色分布を比較することによって計算される。 i 番目のポリゴンが平均 \mathbf{u}_j 分散 Σ_{cj} の色情報を持っていた場合、尤度関数は次のようになる。

$$p(\mathbf{O}_c | \mathbf{X}_o) = \frac{\mu_i \lambda_j}{\Sigma \mu_j}$$

ここで、 λ_j は一致度を表す係数であり、次の式で計算される。

$$\lambda_j = \begin{cases} 1 & d_j < \tau_j \\ 0 & \text{otherwise} \end{cases}$$

τ_j は閾値である。 d_j は次のようなマハラノビス距離である。

$$d_j = \sqrt{(\mathbf{o}_{cj} - \mathbf{u}_j)^T \Sigma_{cj}^{-1} (\mathbf{o}_{cj} - \mathbf{u}_j)}$$

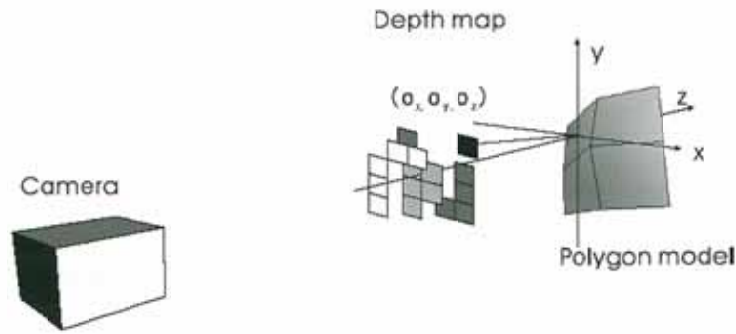


図18 Likelihood computation

3.3.5 実験結果

提案した手法を確認するために実験を行った。使用したコンピュータはAthlon XP2400のCPUを有するものである。図19は実験環境と使用したカメラである。使用したカメラはSONY FCB-IX47である。カメラの間隔は12cmである。図20左に示すように入力画像の大きさは320x240ピクセルである。また奥行き画像の大きさは図20右に示すように100x80ピクセルである。ポリゴンモデルは、図21に示すような簡単なものを用いた。ポリゴン数は108である。

図22に推定された軌跡のX-Z平面の図を示す。黒い線は実際に被験者が移動した軌跡を表す。赤い線は追跡結果を示す。図23は追跡時の画像系列である。緑色の円が推定された頭部の位置を示す。1フレームあたりの処理時間は約1秒である。



図19 Experimental environment and stereo camera configuration

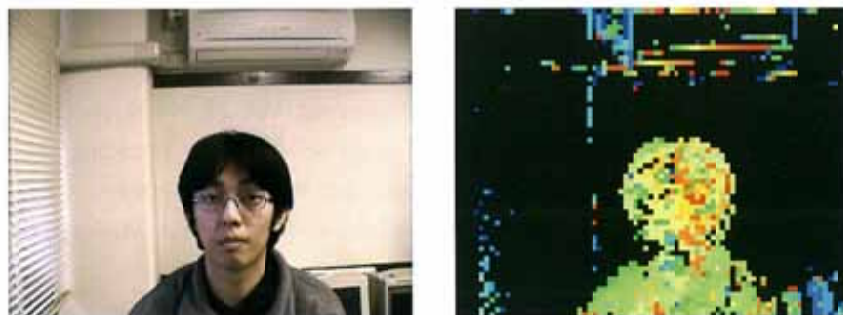


図20 Input image and depth map

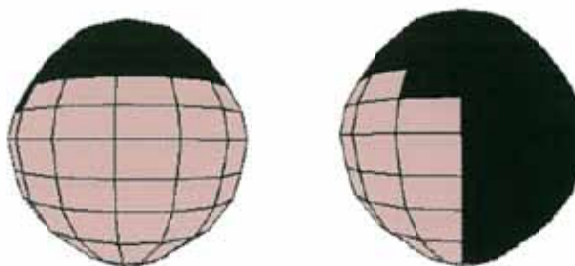


図21 Polygon model used in our experiment

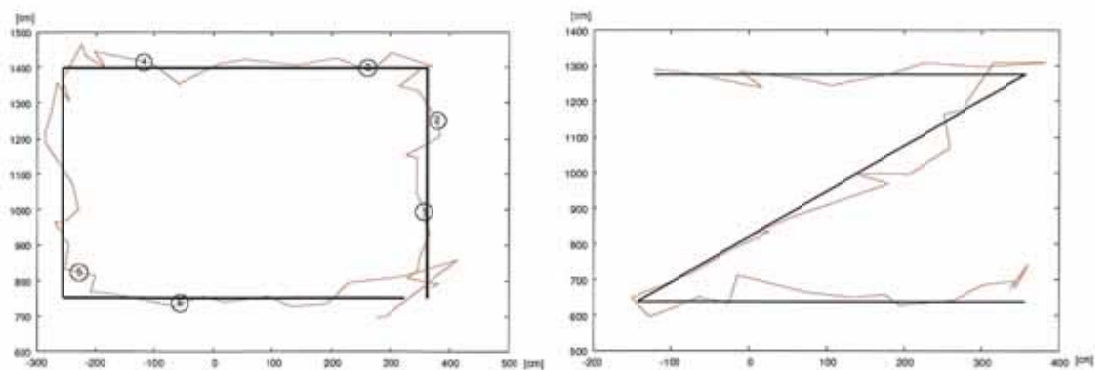


图 22 Tracking result

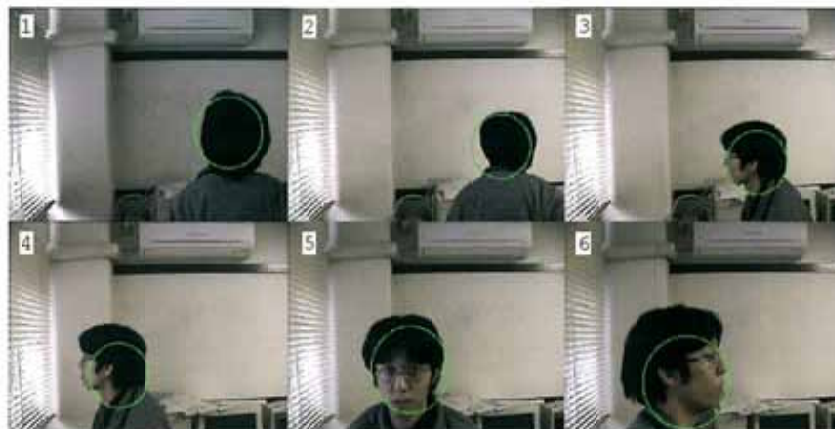


图 23 An example of tracking sequence

参考文献

- [20] Sumit Basu, Irfan Issa, and Alex Pentland, "Motion regularization for model-based head tracking," In Proceedings of the International Conference on Pattern Recognition, 1996.
- [21] Xavier L.C. Broly, Constantinos Stratelos and Jeffrey B. Mulligan, "Model-based head pose estimation for air-traffic controllers," In Proceedings of the International Conference on Image Processing, 2003.
- [22] Marco La Cascia, Stan Sclaroff and Vassilis Athitsos, "Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 4, pp. 322-336, 2000.
- [23] F. Dornaika and J. Ahlberg, "Model-based Head and Facial Motion Tracking," In Proceedings of the 12th International Conference on Image Analysis and Processing, 2003.
- [24] Arnaud Doucet, Nando De Freitas and Neil Gordon, "Sequential Monte Carlo Methods in Practice," Springer Verlag, 2001.
- [25] Michael Acheson Isard, "Visual Motion Analysis by Probabilistic Propagation of Conditional Density," D.Phil. Thesis, Oxford University, 1998.
- [26] Michael Isard and Andrew Blake, "CONDENSATION -- conditional density propagation for visual tracking," International Journal of Computer Vision, Vol. 29, No. 1, pp. 5-28, 1998.
- [27] S. Malasiotis and M.G. Strintzis, "Real-Time Head Tracking and 3D Pose Estimation From Range Data," In Proc. of the International Conference on Image Processing, 2003.
- [28] Marius Malciu and Françoise Prêteux, "A Robust Model-Based Approach for 3D Head Tracking in Video Sequences," In Proc. of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, Vol. 1, pp. 169-174, 2000.
- [29] G. Kamberova, R. Mandelbaum, and M. Mintz, "Stereo Depth Estimation: A Confidence Interval Approach," In Proceedings of the International Conference on Computer Vision, pp. 503-509,

1998.

- [30] Larry Matthies and Steven A. Shafer, "Error Modeling in Stereo Navigation," IEEE Journal of Robotics and Automation, Vol. RA-3, No. 3, pp. 239-250, 1987.
- [31] Thomas B. Moeslund and Erik Granum, "A Survey of Computer Vision-Based Human Motion Capture," Computer Vision and Image Understanding, Vol. 81, No. 3, pp. 231-268, 2001.
- [32] Francesc Moreno, Adrià Tarrida, Juan Andrade-Cetto, and Alberto Sanfeliu, "3D Real-Time Head Tracking Fusing Color Histograms and Stereovision," In Proc. 16th IAPR International Conference on Pattern Recognition, Vol. I, pp. 368-371, 2002.
- [33] Frederic H. Pighin, Richard Szeliski, and David H. Salesin, "Resynthesizing Facial Animation through 3D Model-based Tracking," In Proc. of the International Conference on Computer Vision, pp. 143-150, 1999.
- [34] Daniel B. Russakoff and Martin Herman, "Head tracking using stereo," In Proc. of the IAPR Workshop on Machine Vision and Applications, Vol. 13, No. 3, pp. 164-173, 2002.
- [35] Arno Schödl, Antonio Haro, and Irfan A. Essa, "Head Tracking Using a Textured Polygonal Model," In Proceedings of Workshop on Perceptual User Interfaces, 1998.
- [36] Jamie Sherrah and Shaogang Gong, "Fusion of perceptual cues for robust tracking of head pose and position," Pattern Recognition, Vol. 34, No. 8, pp. 1565-1572, 2001.
- [37] Hedvig Sidenbladh, Michael J. Black, and David J. Fleet, "Stochastic Tracking of 3D Human Figures Using 2D Image Motion," In Proc. of the 6th European Conference on Computer Vision, pp. 702-718, 2000.
- [38] B. Stenger, A. Thayananthan, P.H.S. Torr and R. Cipolla, "Filtering Using a Tree-Based Estimator," In Proc. of the 9th International Conference on Computer Vision, Vol. II, pp. 1063-1070, 2003.
- [39] Jacob Ström, "Model-Based Real-Time Head Tracking," EURASIP Journal on Applied Signal Processing, Vol. 2002, No. 10, pp. 1039-1052, 2002.
- [40] Ruigang Yang and Zhengyou Zhang, "Model-based head pose tracking with stereovision," In Proc. of the 5th IEEE International Conference on Automatic Face and Gesture Recognition, 2002.

フェーズⅢ

今後の取り組み

これまでの研究において、ノイズの少ない背景画像の生成技術、微動人物領域の検出技術、安定した人物の頭部追跡技術の有効性が示された。今後の取り組みとしては、これらの研究で得られた成果を応用し、知的空間における人物の検出や追跡等への展開が考えられる。そのためには、実際の生活環境内での動作を考慮した課題等を検討し、それらに対応した技術の確立が必要である。