

動画像を用いた人物の位置検出手法

宮地 正大

1 はじめに

近年、情報通信技術の発展によりユビキタスコンピューティングが生活の中に浸透してきており、人々が特に意識していない部分でも多様なセンサを用いたサービスが数多く存在している。このようなセンサ情報を利用したサービスとして自動ドアや照明の自動点灯などが挙げられる。これらはセンサにより特定の位置に利用者が存在しているかを判断することでサービスを提供している。しかし、このような単純な情報に加えて、利用者の位置・方向情報の取得が可能になれば、より高度なサービスの提供や利用形態の分析などが実現できると考えられる。

例えば、室内における利用者の位置・方向情報があれば、その空間内での利用者の関係性が推測できる。具体的には、一カ所に利用者が集合し、かつそれらが向き合っているならば、その集団は話し合いをしている可能性が高いと考えられる。そこで本研究では、利用者の状況の把握、行動予測とそれに基づいたサービスの提供を目指して、人物の位置及び方向を検出する手法について検討する。本発表では、これらのうち、室内での人物の位置検出手法について動画像を用いて検討した。

2 人物の位置検出手法

本研究では、公共空間内における利用状況の分析などを目的としているため、人物の位置及び方向の検出においては以下の条件を満たす必要がある。

- 公共空間での検出の為、個人は特定しない
- 個人にセンサを携帯させない

主な室内における人物の位置検出手法 Table 1 に示した。

Table1 主な位置検出手法評価

手法	方向情報の取得	ユーザのセンサ携帯
動画像処理	可	不要
圧力センサ	不可	不要
赤外線センサ	不可	不要
RFID の電波強度	可	必要
GPS・地磁気センサ	可	必要

前述の条件を考慮すると、不特定多数の人物が出入りする公共空間にという環境を想定しているため、個人に RFID タグや GPS・地磁気センサなど特別なセンサを持たせることが出来ない。また、Table 1 より、現状では圧力センサや赤外線センサでは方向情報の取得が難しく、導入にも手間がかかる。これより本研究では人物の位置及び方

向の検出に動画像処理を用いるものとした。

3 動画像による人物の位置検出

本研究では、Fig. 1 のように 4m 四方の空間に設置した 2 つの Web カメラの動画像から位置検出を行う。具

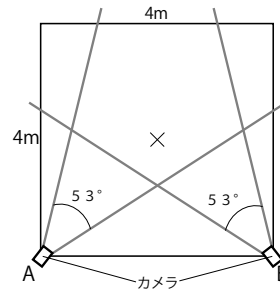


Fig.1 実験環境俯瞰図

体的には下記のように人物の顔位置を画像処理で取得し、2 点のカメラ画像から三角測量で人物の三次元位置を特定する。顔位置の取得には OpenCV のオブジェクト検出を使用して、顔正面の特徴量から判断する。そのため、カメラが人物の顔を正面から捉えることが必要条件となる。以下に本研究での位置検出の手順を示す。

- 顔位置候補の限定
- 特徴量の評価
- 顔正面画像の判断
- 人物追跡
- 三角測量で人物の三次元位置特定

使用したカメラは Buffalo 社の USB カメラである BWC-35H01 を使用した。基本性能を Table2 に示す。このカメラを選んだ理由としては、軽量で様々な場所に

Table2 使用カメラ基本性能

解像度	フレームレート	画角	映像素子
320*280	15fps	53°	1/4 インチ CMOS

設置可能なこと、安価で複数台入手し易いことが挙げられる。

4 顔位置の特定

4.1 雑音除去・顔位置候補の限定

取得した動画像のすべての領域に対してオブジェクト検出を行うと計算量としても膨大になるうえ、誤検出の可能性が高くなる。そのため、本研究では、背景差分、肌色抽出、膨張・収縮によって、明らかに人物でない部分

(以下、雑音)を予め除去する。

4.1.1 輝度情報を用いた動的背景差分法

雑音除去の手法としてカメラを設置した空間の画像を背景画像の初期値として保存しておき、それ以降取得した入力画像との差分を計算することで前景画像となる物体を検出する。実際の環境では単純に画素のRGB情報を比較する手法では環境の変化に弱く、設置したカメラが僅かに動くだけで誤検出をしてしまうため、ある程度揺らぎを持たせる必要がある。そのため、画像全体の輝度パターンの変化を検出する手法を用いて背景を特定した。また、環境光の変化に対応する為、ある程度動的に背景情報を更新する必要がある。Fig. 2 に用いた背景差分手法の処理内容を示した。カメラを設置した際に、背

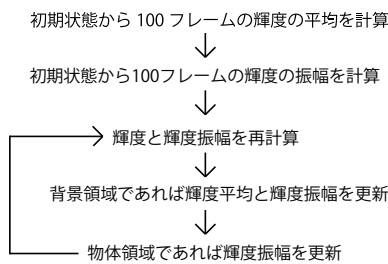


Fig.2 背景差分の処理手順

景画像の初期値として連続した100フレームの画像を取得する。画素ごとに輝度の平均 \bar{I} と輝度振幅 I' を計算する。

$$I' = \sqrt{2 \times (I - \bar{I})} \quad (1)$$

計算した I' が、 $\bar{I} - I' \leq I \leq \bar{I} + I'$ の場合に、その画素は背景領域中に存在する画素であると判断する。背景と判定された領域では輝度平均値 \bar{I} と振幅 I' を以下の式 (2) 式 (3) を用いて更新する。

$$\bar{I}' = \frac{(n-1)}{n} \times \bar{I} + \frac{1}{n} \times I \quad (2)$$

$$I' = \frac{(n-1)}{n} \times I' + \frac{1}{n} \times \sqrt{2 \times (I - \bar{I}^2)} \quad (3)$$

ここで、 n は更新速度パラメータであり、本研究では $n=1.0 / 500$ とした。

一方、物体領域と判定された領域では、輝度平均値は元の値を保持し、振幅 I' のみを以下の式 (4) 式 (5) を用いて更新する。

$$\bar{I}' = \bar{I} \quad (4)$$

$$I' = \frac{(m-1)}{m} \times I' + \frac{1}{m} \times \sqrt{2 \times (I - \bar{I}^2)} \quad (5)$$

ここで、 m は物体領域更新速度パラメータであり、本研究では $m=1.0 / 2000$ とした。これらの処理を毎フレーム行うことで背景画像と前景画像の分離を行っている。Fig. 3 に実際に入力画像から前景画像を抽出した例を示す。

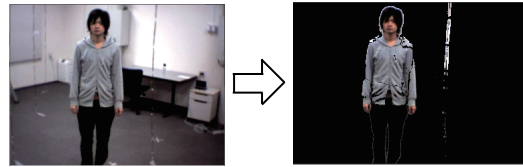


Fig.3 背景差分処理

4.1.2 膨張収縮

背景差分法で検出された前景画像にはごま塩雑音と呼ばれる小さな粒の雑音が残る。これらの雑音を除去するために膨張・収縮処理を適応した。これらは二値画像の雑音除去に対して用いられる手法であり、入力画像に対して膨張・収縮処理を同じ回数繰り返すことで任意の大きさの雑音を除去することが可能である。

● 膨張

膨張は物体領域を大きくする処理である。ある画素の近傍 (4近傍, 8近傍) に一つでも1 (白)があれば、その画素を1にする処理である。Fig. 4 に二値画像に膨張処理を加えた例を示す。

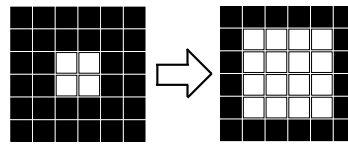


Fig.4 膨張処理

● 収縮

収縮は物体領域を一回り小さくする処理である。ある画素の近傍に一つでも0があれば、その画素を0にする処理である。Fig. 5 に二値画像に収縮処理を加えた例を示す。

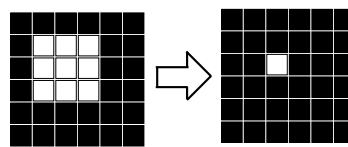


Fig.5 収縮処理

実際に Fig. 6 にごま塩雑音の存在する入力画像から膨張・収縮処理により、雑音除去した例を示す。

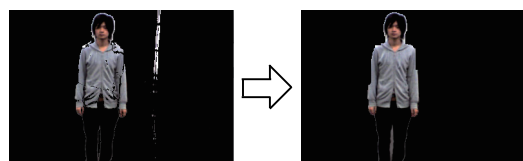


Fig.6 膨張・収縮処理

4.2 肌色抽出

人物の顔を検出の精度を上げるために、人の顔となり得る箇所の候補を特定する必要がある。本研究は動画像中に現れる肌色領域付近を顔の候補としている。肌色領域の定義として式 (6) を満たす画素とした。

$$(0, 50, 0) \leq (H, S, V) \leq (40, 255, 255) \quad (6)$$

また肌色領域に関して、200 ピクセル以上の面積を持つ集合のみを有効な領域とした。実際に Fig. 7 に入力画像から有効な肌色領域のみを抽出した例を示す。

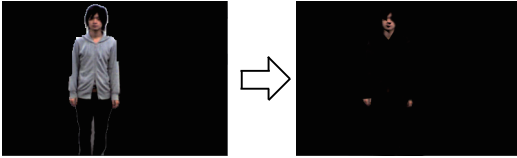


Fig.7 肌色抽出処理

4.3 OpenCV による顔認識

人物の顔位置特定には OpenCV のオブジェクト検出を用いて顔正面の特徴量から判断した。OpenCV は Intel の研究所が提供している動画像処理ライブラリである。提供される言語として C/C++・Python でのライブラリが用意されている。OpenCV のオブジェクト検出手法では、画像特徴量として Haar-Like 特徴量、学習アルゴリズムとして Adaboost と呼ばれるアルゴリズムを用いて作成された識別器を用いて行われる。

4.3.1 AdaBoost による強識別器の作成

Adaboost は、単純な識別関数を複数組み合わせることによって高い精度の識別器を得る学習、判断アルゴリズムである。学習データに対するリサンプリングの際に重みを逐次的に更新していき、作成した識別器を最後に学習機械に対する重みをつけて足し合わせて統合学習機械を作成する。

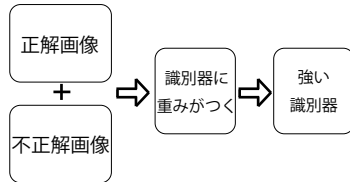


Fig.8 adaboost 学習アルゴリズム

4.3.2 Haar-Like 特徴量を用いた弱識別器

Haar-Like 特徴量は、矩形領域の平均明度の差分値として求められるスカラー量であり、その値は明度勾配の強度を表す。Fig. 9 にその特徴の種類を示す。入力画像に対してこの特徴を用いて画像から顔を検出する。Fig. 9 の矩形特徴の黒色領域に相当する部分と、グレースケールへと変換した入力画像の明度情報に値する黒色領域の特徴との一致度によって、それぞれの識別器の重みを判

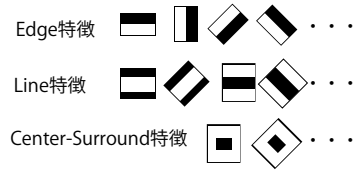


Fig.9 Haar-Like 特徴

断する。この矩形特徴の位置、種類の組み合わせで入力画像の特徴量を判断することで、顔の特徴量を持つ部分を検出する。

5 特定した人物の追跡

顔認識は毎フレーム (10fps) において行われるため、各フレームにおいて検出された顔位置はそれぞれ独立である。そのため、途中に顔が検出されないフレームがあった場合、途中の顔位置を補完することが出来ない。これを解決するため、パーティクルフィルタによる物体追跡を用いて、検出された顔の位置を予測した。

パーティクルフィルタとはパーティクル (粒子) と呼ばれるそれぞれが重みを持った点を空間に散布し、その空間内全体の確率分布を近似する手法で、その確率をもとにパーティクルの次の遷移場所を決める。パーティクル配置場所に関しての尤度 L を肌色色相へのユークリッド距離 d に対して、平均 0、分散 σ^2 の正規分布として定義しており、式 (7) のように表される。

$$L(d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{d^2}{2\sigma^2}\right) \quad (7)$$

つまり、パーティクルは次の遷移場所を空間全体のパーティクルの尤度を基に決定するが、式 (7) によって求められた尤度が高い、より肌色に近い画素を持つパーティクル付近へと遷移する。本研究ではパラメータを色相 40、分散 σ^2 を 10、パーティクル数を 100 としている。このパーティクルを顔認識で得られた肌色領域を含む最小矩形にここでは撒き直すことで顔の肌色領域のみを追跡する。追跡で得られたパーティクルの集合の重心が顔の中心であるとした。

6 人物の三次元位置特定

複数の動画像から得られた顔位置それぞれの座標を空間全体の座標系に変換する必要がある。実験環境としてカメラの位置と設置角度は固定しているので、Fig. 10 のように方向情報とカメラ間の相対的な位置関係と方向を用いて、三角測量で今後特定する予定である。

7 まとめ

本稿では、公共空間内における人物の行動分析や行動予測を行うために、複数台のカメラの動画像から人物の位置を検出する手法を提案した。OpenCV の Haar-Like 特徴量による検出手法だけでは適切に顔を認識できないため、前処理として雑音除去及び顔位置候補となる領域の抽出を背景差分、肌色抽出、膨張・収縮で行い、それら

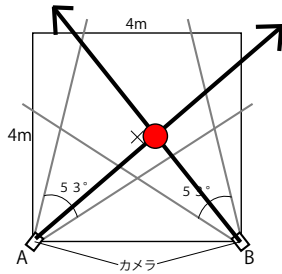


Fig.10 三角測量による位置特定

の処理で得られた顔位置をパーティクルフィルタによって追跡、予測することで安定した顔位置検出が行える。今後はカメラを4台に増設することで、死角を減らし、視点を増やしてより人物の顔を正面から捉えやすくする。また現状では検出対象となる人物を1人であることから、ラベリング処理によって検出された人物にIDを振り分けることによって複数人に対応させる予定である。

参考文献

- 1) Document OpenCV-1.1pre
<http://opencv.jp/opencv-1.1.0/document/>
- 2) Open Source Computer Vision Library, Intel
<http://www.intel.com/technology/computing/opencv>
- 3) 出水祐樹, 佐野睦夫, 西口敏司, "AAMとパーティクルフィルタを用いた視線追跡", 電子情報通信学会研究報告, PRMU, パターン認識・メディア理解, Vol.108, No.198, 2008-73, p.163-168, Sep. 2008
- 4) Hongliang Bai, Jianping Wu, Changpin Liu, "Motion and Haar-like Features Based Vehicle Detection", IEEE Multimedia Modelling Conference, 2006