

雑踏における複数歩行者の追跡*

内田 恭輔**・三浦 純**・白井 良明**

Tracking Multiple Pedestrians in Crowd*

Kyouzuke UCHIDA**, Jun MIURA** and Yoshiaki SHIRAI**

This paper proposes a robust method for tracking many pedestrians viewed from an upper oblique angle. Persons are extracted while they are isolated, and they are tracked forward and backward. In tracking, if the region of a tracked pedestrian is ambiguous because of overlapping with other pedestrians, plural candidates for a pedestrian path are generated, and false candidates are eliminated in other frames. We make an experiment using real image sequences to show the effectiveness of the method.

1. はじめに

近年、ビジョンを用いた移動物体の検出と追跡の研究が数多く行われている¹⁾。その応用例には、不審者や侵入者を検知するセキュリティシステム、動画像データベースの検索、スポーツシーンの解析、道路環境における交通監視システムなどがある²⁾⁻⁴⁾。一方、都市デザインの分野では、快適な都市づくりのために歩行者の群の振るまい(生成・移動・消滅など)の解析^{5),6)}が行われている。解析者は都市の雑踏が数時間に渡って撮影されたビデオテープを調べる。しかしテープには常に30人以上の歩行者が行き交っていて作業は難航する。このため歩行者追跡ツールの開発の必要性が高まっている。そこで本稿では都市の雑踏における複数歩行者の追跡手法を提案する。

都市の雑踏では歩行者が他の歩行者や電柱などの陰に入って隠蔽されることがあり、歩行者同士の隠蔽では三人以上が一組に重なることも珍しくない。こうした隠蔽がFig.1では上部歩道だけでも10個近く見られ、歩行者の検出と追跡を困難にしている。

検出で問題になるのは動物体が重なっている場合に一



Fig. 1 Urban crowd viewed from an upper oblique angle

つの物体と見なされてしまうことである。これは従来手法が物体の現れたフレームで検出を行っていたからである⁷⁾。

本手法は全画像を通して単独に分離している歩行者のみを検出する。ここから未来および過去へ追跡することで、歩行者が重なることがあっても追跡を続けることができる。

追跡で問題になるのは隠蔽前後における対応付けである。従来手法は隠蔽の前後で見かけの類似性から対応付けを決めていた⁸⁾。しかし、見かけの似た歩行者同士が重なる場合には対応付けのあいまいさにより誤対応を生じてしまうことがある。このため隠蔽の前後で手動で軌跡を修正する必要がたびたびあった。

* 原稿受付 2000年6月12日

** 大阪大学 大学院 工学研究科 Graduate School of Engineering, Osaka University; 2-1 Yamada-oka, Suita city, Osaka 565-0871, JAPAN

Key Words: vision tracking, template matching, SAD(Sum of Absolute Difference), occlusion, motion prediction.

本手法は対応付けにあいまいさがある場合は対応の可能性のある候補を保持して追跡を続け、その後の追跡の結果を利用して誤った候補を排除する．このようにして隠蔽前後のあいまいさを克服して信頼性の高い追跡を実現している．

本論文は、まず2.で歩行者の検出方法を述べる．次に3.で追跡方法を述べたのち最後に実画像を用いた実験を示す．

2. 歩行者の検出

動物体を背景差分法で抽出する．画像中には常に歩行者が行き交い背景を撮影することが出来ないで背景を生成する必要がある．一連の動画から最も高い頻度で現れる画を背景と見なし、各ピクセルのR,G,B成分(各8ビット)それぞれについてヒストグラムを作りモード(最頻値)を選ぶ．こうして5秒間の動画(75枚)から得られた背景をFig.2に示す．



Fig. 2 Background created from image sequence

この背景を用いて背景差分を行う．カラーの差分にはRGB成分を用いたユークリッド距離やマンハッタン距離などがある．本手法はカラーの差分を(1)式で求める．

$$D(c_1, c_2) = \max\{|r_1 - r_2|, |b_1 - b_2|, |g_1 - g_2|\} \quad (1)$$

ここで c_1, c_2 は差分を行う二つのピクセル、 r_i, g_i, b_i はピクセル c_i のRGB成分の値を示す．

Fig.3にFig.1の背景差分像を示すが、原画像で発見できない様々な雑音を含んでいるのが分かる．雑音を除去して歩行者を抽出するために二値化を行う．二値化の場合閾値の決定に局所的な値のヒストグラムを利用する方法がある．Fig.3の長方形で囲まれた領域のヒストグラムをFig.4に示すがこれから分かるように歩行者と雑音の区別がつきにくく、閾値の決定が困難である．

そこで閾値 τ_i を複数設定する．ここでは0~255の値を取る差分値に対して、閾値 τ_i を15から60まで5きざみに10個設定した(15未満や60以上の閾値では歩行者をほとんど検出できない)． $\tau_i = 15$ から始めて検出し



Fig. 3 Subtracted image of Fig.1

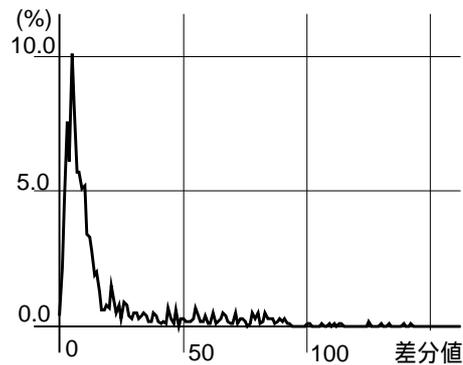


Fig. 4 Local histogram

た歩行者を登録していく．歩行者の検出がまだされていない領域について $\tau_i + 5$ として再度検出を行う．

それぞれの閾値で差分領域を求めるのであるが、歩行者の色と背景の色がしばしば似ているため、歩行者は一般にちぎれちぎれに検出される．歩行者の領域をまとめるため、二つの領域の重心間の距離がある閾値以下ならば領域を併合する．

孤立している歩行者を検出するために、ある一定の大きさを満たす領域を歩行者として選ぶ．今回のカメラ設定では歩行者の大きさは 5×7 から 8×20 の範囲である．

Fig.1に対して、閾値 $\tau_i = 15$ で歩行者を検出すると、9人が検出される．残りの閾値で検出を繰り返し、最終結果では21人が検出される(Fig.5)．図中で長方形に囲まれた白く塗りつぶされていない領域が追跡のためにテンプレートとして使われている．

こうして検出された歩行者の位置と既に求められている軌跡の位置が重複すれば、その歩行者については追跡済みと見なし、一方、そうでない歩行者については追跡を行う．

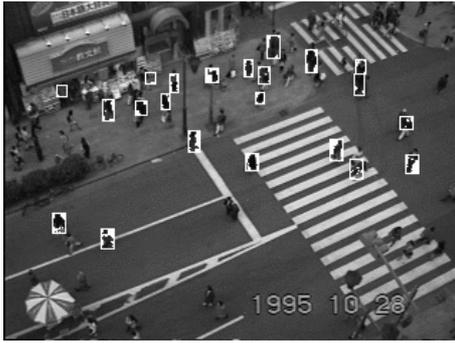


Fig. 5 Detected pedestrians by the background subtraction

3. 歩行者の追跡

3.1 フレーム間の対応付け

対応付けを行うとき、歩行者の動きを予測して探索範囲を決める。微小時間内で歩行者が等速直線運動していると仮定し、過去の連続する六つの位置 (1/15 秒間隔) を用い、(2) 式により現在の速度を求める。ここでは、過去 1 ~ 3 フレームの歩行者の平均位置と 4 ~ 6 フレームの平均位置の差をとって速度を求めている。

$$v(t) = \left[\frac{\sum_{i=1}^3 x(t-i) / 3 - \sum_{i=4}^6 x(t-i) / 3}{3} \right] / 3 \quad (2)$$

次のフレームでは予測される位置を中心に、予測位置から歩行者がずれる分だけを考慮して、そのまわりの領域を探索する。画像データを調べた結果、予測される位置の ± 1 ピクセルの範囲に歩行者が入っていたので、予測位置まわりの 3×3 ウィンドウについてテンプレートをずらしながら探索を行う。

テンプレートのある位置に置いたとき対応の評価には計算の容易な SAD (Sum of Absolute Difference) を用いる。マッチングは多少の姿勢変化に対応するよう、各ピクセルの AD (Absolute Difference) は (4) 式により対応するピクセルおよびその近傍 3×3 のピクセルとの差分のうち最小値を取る。この AD を (3) 式によりテンプレート内で総和を取って SAD を求める。

$$SAD = \sum_{i=1}^h \sum_{j=1}^w AD(i, j) \quad (3)$$

$$AD(i, j) = \max_{-1 \leq m, n \leq 1} D(T(i, j), I(i+m, j+n)) \quad (4)$$

ここで (i, j) はテンプレート内の座標、 (w, h) はテンプレートの幅、高さ、 $T(i, j)$ と $I(i, j)$ はテンプレートと画像の対応するピクセルを示す。

ところで、歩行者の一部が隠蔽される場合や他の歩行

者と近接する場合には、対応点で SAD が常に最小値を取るとは限らない。そこで、SAD が極小値かつ閾値 τ_t 以下の点を対応点とする。

閾値 τ_t は対応を評価するのに重要な役割を果たす。たとえば τ_t が高過ぎると対応点が過剰に得られてしまい、一方低過ぎると対応点が得られなくなってしまい、いずれも追跡失敗の要因となる。あらかじめ τ_t を与えて何人かの歩行者について実験を行うと、追跡に成功した場合の値は、1 ピクセルあたりおよそ 8 ~ 20 と歩行者により大きく異なった (明るい歩行者ほど高い値であった) が、どの歩行者についても毎フレームの SAD の最小値に近いことが頻繁に確認された。そこで、この事実に基づき τ_t を次のように与えた。まず、歩行者の明度平均に比例した値を初期値に与える。その後毎フレームにおいて、 3×3 ウィンドウでそれぞれ計算した九つの SAD のうち最小の値 SAD_{min} を用いて (5) 式により更新する。軌跡候補が二つある場合などは、二つの 3×3 ウィンドウで計算した 18 の SAD から最小の SAD_{min} を選ぶ。 $\tau_t > SAD_{min}$ の場合、閾値を SAD_{min} に近づけるため重み平均をとる。ここでは、 $w_1 = 0.8$, $w_2 = 0.2$ としている。 $\tau_t \leq SAD_{min}$ の場合には対応点が全く得られない。この場合、追跡対象が隠蔽されている可能性があり SAD 値を信用できないため、重み付け平均をとらずに一定数だけ上げるようにした。ここでは、 $c = 1 \times$ (歩行者のピクセル数) としている。

$$\tau_{t+1} = \begin{cases} w_1 \cdot \tau_t + w_2 \cdot SAD_{min} & (\tau_t > SAD_{min}) \\ \tau_t + c & (\tau_t \leq SAD_{min}) \end{cases} \quad (5)$$

Fig. 6 にピクセル数 67 の歩行者における τ_t と SAD_{min} の各フレームにおける値を示す。 τ_t が SAD_{min} 以下の場合にも、次のフレームでは一定値上昇して SAD_{min} を上回る様子や、全体を通して SAD_{min} に近い値が得られているのが分かる。

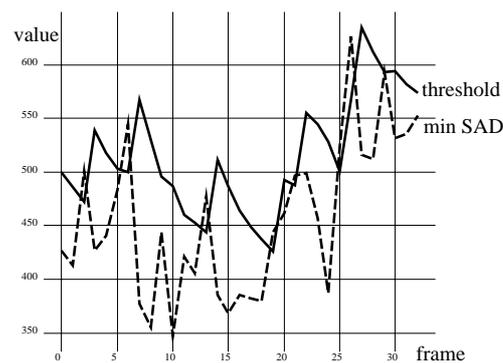


Fig. 6 Minimum SAD and threshold

対応点を得られないときは、等速運動させた位置を中心に探索を続ける。探索範囲は、歩行者の位置がしいに予測からずれることを考慮して、隠蔽が生じてから5フレーム間は 5×5 、6～10フレーム間は 7×7 に設定した。ここで、対応点を得られない原因としては次の三つが考えられる。

- (1) τ_t が小さすぎる場合
- (2) 歩行者の一部分が他の歩行者や街灯に隠蔽されて対応が十分に取れない場合
- (3) 歩行者を完全に見失っている場合

それぞれの原因において再び対応が取れるのは、

- (1) τ_t が十分上昇した場合
- (2) 歩行者の姿が再び現れた場合

である。一定時間(ここでは10フレーム)探索し続けても対応がとれないときは、歩行者を見失った((3)の場合)と見なして追跡を打ち切っている。

3.2 軌跡候補が複数の場合

複数の軌跡候補が生成される可能性がある。たとえば、Fig.7の(a)～(c)にかけて追跡中の歩行者と色の似ている歩行者が重なるので、複数の対応点候補が存在し、(b)において三つに軌跡が分岐し、(b)の下二つの軌跡が(c)ではさらにそれぞれ二つに分岐して合計五つの軌跡候補が生成する。

ここで、二つの軌跡候補の絞られ方が異なる。一つは、分岐した軌跡が合流する場合である。この場合、歩行者が最短距離で進みだる性質を考慮して分岐していた軌跡のうち道のりの短いものを選ぶ。これによりFig.7の(c)で五つある軌跡候補のうちの下三つの軌跡が合流した結果、軌跡候補が(d)で三つに絞られている。

もう一つは、分岐した軌跡が打ち切られる場合である。一般に、誤った軌跡では閾値 τ_t 以下のSADを得るのが難しく、一定フレーム以上対応が見つからずに追跡が打ち切られる。Fig.7の(d)から(f)にかけて誤りの軌跡候補が除外されて最後に正しい軌跡が残っている様子がわかる。

歩行者がフレームの外へ出た場合や動画像が途切れた場合、追跡を終了する。このとき、軌跡候補が複数残っている場合は全軌跡候補の共通部分だけを抽出し、その追跡時間が十分(ここでは10フレーム以上)であれば軌跡を登録する。

3.3 誤りのある軌跡の検出

検出と追跡を行った後に、誤りのある軌跡を削除するために求められた軌跡を検証する。

Fig.8は二人の歩行者AとBが交錯する例である。図中の歩行者(楕円)内の数字はフレーム番号、長方形はテンプレートを検出した領域を示す。まずFig.8(a)の歩行



(a) Initial tracking

(b) Split-up into three paths



(c) Furthermore, split-up

(d) Merger of paths



(e) Elimination of false paths

(f) Last single path

Fig. 7 Plural candidates for the pedestrian path

者Aの追跡は成功した例である。歩行者Aをフレーム0で検出して正しく追跡した結果が示されている。次にFig.8(b)の歩行者Bの追跡は誤った例である。歩行者Bをフレーム1で検出したものの、フレーム2以降で誤って追跡した結果が示されている。この場合Fig.8(c)に見られるように追跡が後に再び行われる。歩行者Bをフレーム4で検出して正しく追跡した結果が示されている。しかし、二つの軌跡がFig.8(c)右上の長方形を通過することがテンプレートを検出した領域を単独の歩行者領域と見なしている仮定に反する。このため、この二つの軌跡の一方を誤りとして除外する必要性が生じる。そこで、目的地へ最短距離で向かおうとする歩行者の性質に基づき、二つの軌跡のうちより直線的な方が正しく他方が誤りであると見なし、誤りを除外する。こうしてFig.8(d)に見られるように(b)における誤った追跡が削除され、

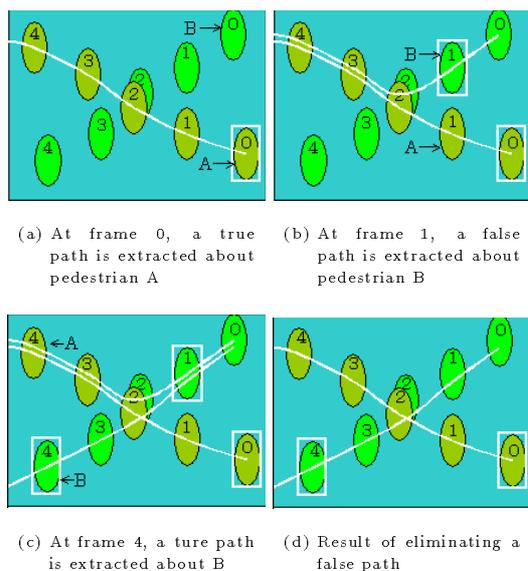


Fig. 8 Detecting an error locus

正しい軌跡だけが残る。

ここで、軌跡の直線性は(6)式により軌跡の道のり l と軌跡の始点から終点までの距離 d の比 r で求め、 r の小さい方が直線的であるとする。

$$r = \frac{l}{d} = \frac{\sum_{t=1}^{t_e} |x(t) - x(t-1)|}{|x(t_e) - x(0)|} \quad (6)$$

3.4 実験結果

Fig. 1 を含む 5 秒間 (75 フレーム) の動画像に対して本手法を適用した。2/3 秒 (10 フレーム) おきの各歩行者の頭部の軌跡を Fig. 9 に示す。

歩行者 A が画像 (c) から (f) にかけて、右前方から来る歩行者 B および C を左に迂回して避けている様子や、この間、歩行者 C が歩行者 A にじゃまされてあまり前へ進めず、歩行者 B に追い越された様子が良く分かる。

人手による計測と比較した結果、本手法の追跡状況を Table 1 および Table 2 に示す。ここで、動画像を通して歩行者を最初から最後まで追跡しているものを追跡の成功あるいは正しい軌跡としている。

Table 1 Result of tracking 1

全歩行者数	追跡成功数	未追跡数
86 人	66 (77%)	20 (23%)

Table 1 から分かるように全歩行者の約 77% (66/86) について軌跡の抽出に成功している。追跡されないのは背景と色が良く似ている歩行者で、この場合背景と区別し難く歩行者を見失うので追跡が困難である。また、毎回検出を行ったが、最初から最後まで重なっている歩行者は、最後まで一人の歩行者と見なされる。

Table 2 Result of tracking 2

抽出軌跡数	正しい軌跡数	誤った軌跡数
68	66 (97%)	2 (3%)

Table 2 から分かるように得られた軌跡の正解率は 97% (66/68) であった。誤りが生じやすいのは見かけの似ている歩行者が近接する場合で、このような場合も軌跡候補の生成・検証により安定して追跡できた。また、求められた軌跡間に背理を見つけることが誤りの軌跡を除外するのに役立った。

4. おわりに

雑踏において複数歩行者を追跡する手法を述べた。雑踏における追跡の問題点として重なる頻度の多さが挙げられるが、本手法では複数の軌跡候補を生成・検証することにより、十分この問題に対処できた。

最後に本手法をより多くの状況において実験を行うことと、本手法によって得られた軌跡データを適用して都市における群衆の解析などの応用研究を行うことが今後の課題である。

謝 辞

本研究を進めるにあたり、鍛佳代子氏 (東京工芸大学) が撮影されました都市の雑踏のビデオを利用させていただきました。

参考文献

- 谷内田ほか: コンピュータービジョン, 丸善, pp. 167-189 (1990)
- 筒井, 白井, 三浦: オプティカルフローを用いた複数動物体の実時間追跡; 第 41 回システム制御情報学会研究発表講演会講演論文集, pp. 693-694 (1997)
- 和田, 浮田, 松山: 視点固定型パン・チルト・ズームカメラとその応用; 信学論, D-II, Vol.J 81-D-II, No. 6, pp. 1182-1193 (Jun.1998)
- 大野, 三浦, 白井: サッカーゲームにおける選手とボールの追跡; 情報処理学会研究報告, 99-CVIM-114, pp. 49-56 (1999)
- 竹内, 長田: 画像処理による群衆行動性状評価に関する研究; 日本建築学会大会学術講演梗概集 E5518 (1994)
- K.Kitai: Space Activity - Quantitative analysis and evaluation with experiments on the street; *Proc.*

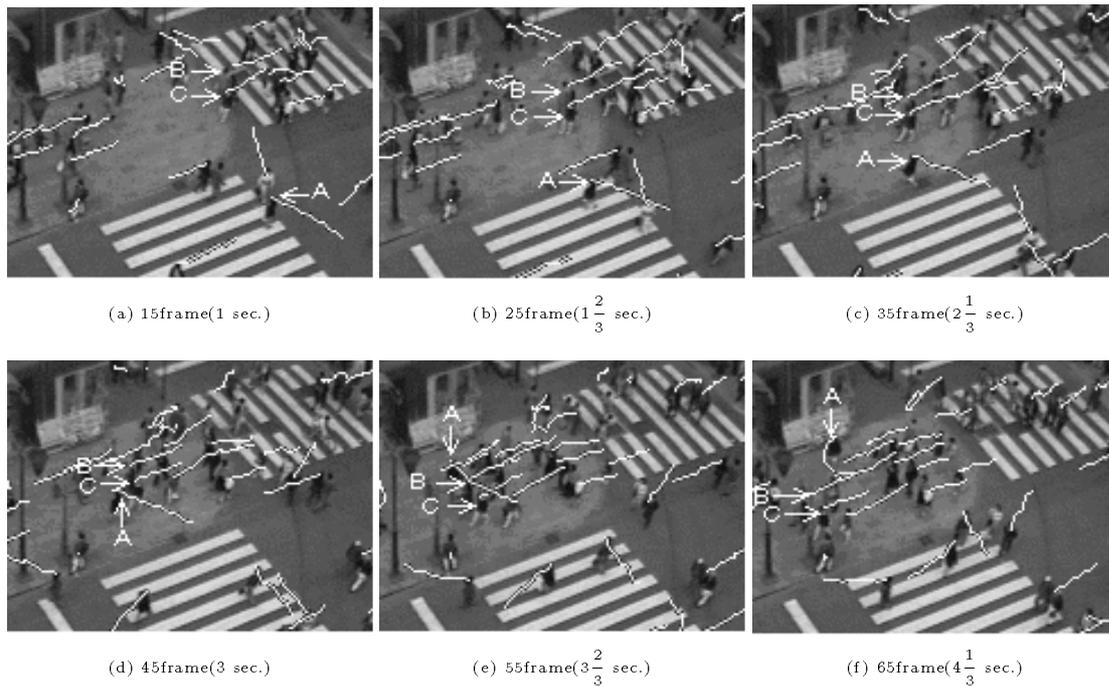


Fig. 9 Experimental result of pedestrian tracking

S&T/SPIE Conference on Videometrics VI, pp. 268-275 (1999)

7) M.Yachida, M.Asada and S.Tsuji : Automatic Analysis of Moving Images ; *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.PAMI- 3, No. 1, pp. 12-20 (1981)

8) S.S. Intille, J.W. Davis and A.F. Bobick : Real-Time Closed-World Tracking ; *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society Press, pp. 697-703 (June 1997)

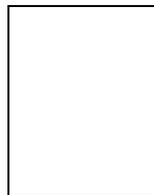
著者略歴

うちだ きょうすけ
内田 恭 輔



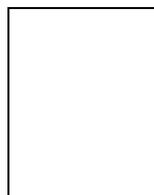
1976年12月13日生。1999年3月大阪大学工学部機械物理工学専攻卒業。同年4月大阪大学大学院工学研究科電子制御機械専攻入学。現在に至る。コンピュータービジョンの研究に従事。

みうら じゅん
三浦 純 (正会員)



1989年3月東京大学大学院工学系研究科情報工学専攻博士課程修了。同年4月大阪大学工学部電子制御機械工学科助手。現在、同助教授。知能ロボット、人工知能、コンピュータービジョンの研究に従事。工学博士。1994年～1995年カーネギーメロン大学計算機科学科客員研究員。1997年日本ロボット学会論文賞受賞。ロボット学会、人工知能学会、日本機械学会、電子情報通信学会、情報処理学会、IEEE、AAAIなどの会員。

しら い よしあき
白井 良 明 (正会員)



1964年3月東京大学大学院工学系研究科機械工学専攻博士課程修了。同年、電子技術総合研究所入所。1971～1972年 MIT AIラボの客員研究員。1979年視覚情報研究室長。1985年 制御部部長。1988年大阪大学工学部電子制御機械工学科教授となり現在に至る。コンピュータービジョン、知能ロボット、人工知能の研究に従事。工学博士。情報処理学会、電子情報通信学会、日本機械学会、計測自動制御学会、ロボット学会、画像電子学会、エレクトロニクス実装学会、人工知能学会、マルチメディアコンテンツ振興協会、IEEE、AAAIなどの会員。