

動画像からの歩行者検出法

塩 昭 夫

NTT ヒューマンインタフェース研究所

ジャック ス克蘭スキー

カリフォルニア大学アーバイン校

あらまし

単眼視による動画像のシーケンスから動いている歩行者を自動検出する方法について述べる。この方法は、画像強度、動き情報、および物体モデル(動いている歩行者の画像のモデル)に基づいている。画像中の人物は、類似色(強度)の服を着ていることが多く、しかも他の人物に遮蔽されることも多い。また、個々の人物は髪の毛、皮膚、服など色(強度)の異なった複数の部分に分割される。このため、画像から個々の人物を検出するのは容易ではない。しかし、同一人物に属する各部は協調して動くことを利用すれば、動き情報に基づいて領域の分割/統合が可能となる。この領域統合処理は、人間の運動視に於ける知覚的統合(perceptual grouping)に対応する。複雑な実画像のシーケンスを用いた実験の結果、提案する歩行者検出法の妥当性を裏付ける良好な結果を得た。

Segmentation of People in Motion

Akio Shio

NTT Human Interface Laboratories
Nippon Telegraph and Telephone Corp.
1-2356 Take, Yokosuka, 238-03 Japan
co@nttcvg.NTT.jp

Jack Sklansky

Department of Electrical and Computer Engineering
University of California, Irvine
Irvine, CA 92717, U.S.A.

Abstract

A method for segmenting monocular images of people in motion from a cinematic sequence of frames is described. This method is based on image intensities, motion, and an object model — i.e., a model of the image of a person in motion. People in an image may be occluded by other people, and usually it is not easy to detect their boundaries. These boundaries can be detected with motion information if they move in different directions, even if there are almost no apparent differences among object intensities or colors. Each image of a person in a scene usually can be divided into several parts, each with distinct intensities or colors. The parts of a person can be merged into a single group by an iterative merging algorithm based on the object model and the motion information because the parts move coherently. This merging is analogous to the property of perceptual grouping in human visual perception of motion. Experiments based on a sequence of complex real scenes produced results that are supportive of our approach to the segmentation of people in motion.

1 はじめに

動画像のシーケンスから物体を検出する技術は、歩行者や車の交通流の計測、監視、更には車の運転の補助などの実際の応用に非常に有用である。本検討では、多数の移動物体を含む動画像シーケンスからの移動物体の自動検出とその追跡を目指す。対象物体には、車などの剛体のみでなく、人体の様に幾つかの剛体 (rigid object) パーツの結合からなる非剛体 (nonrigid object) が含まれる。

従来、景観画像から物体を抽出する場合には、画像の強度ヒストグラムにもとづく再帰的領域分割 [6][15] や色空間でのクラスタリング [12] が行われてきた。しかし、これらの方法は、物体背景に複雑な強度変化を含む場合や、画像中の物体とその背景とのコントラストが十分でない場合には良好な結果が得られなかった。これらの困難を克服し、領域分割の精度を向上させるには、動画像のシーケンスを利用することが考えられる。すなわち、特定一枚の画像上で2つの物体の境界が得られなくても、シーケンス中の他の画像では、容易にその境界が得られる可能性がある。動画像シーケンスから得られる動きの情報もまた物体境界の検出、同一物体に属する領域の併合、遮蔽 (occlusion) の問題を解く上で有効な手掛かりとなる。

Motion の研究では、動画像シーケンスから物体の3次元構造を再生する多くの試みがなされている [2]。しかし、これらは、一部の例外 [1][18][13] を除き単一の剛体 (rigid object) を扱ったものであり、複数の非剛体 (nonrigid object) を対象にした検討は見られない [10]。物体の3次元構造の再構成は、理論的には興味深いが多くの実際の応用 (例えば、歩行者や車両交通の計測など) では必ずしも必要でない。この様な応用では、各画像には多数 (例えば 50) の移動物体が含まれる。この様な多物体を扱う場合には、限られた分解能の動画像シーケンスのみから3次元構造を再生するのは非常に困難である。更に、実際の応用を考える場合には、3次元構造よりも画像ノイズや撮像条件の変化に対するアルゴリズムの安定性の方が重要である。従って、ここでは、2次元の並進運動を視覚情報から得る問題に着目し、3次元構造再生の問題は扱わない。

人体の動きの解析など非剛体を扱った興味深い研究としては、初期の motion 研究や認知心理学の研究に見られる。Johansson は、人体の複数箇所につけた小さな光源等の動き (MLD : moving light displays) から非剛体の3次元の動きや形が知覚できることを示した [11]。Braunstein らは、人間が剛体と非剛体を区別するために必要な最小の点

の数と観測回数に関する研究で類似の方法を採用している [3]。人間の運動解析に関する研究では、O'Rourke と Badler の詳細な人体モデルと拘束伝播 (constraint propagation) による研究が興味深い [16]。しかし、実際のシーンに現われる様な多数の移動物体が存在する場合には、この様な詳細な運動解析は現実的ではない。更に、この様な解析は、物体抽出よりもむしろ解決の難しい問題である。従って、本報告では手足など人体の各部の運動には着目せず、歩く方向など時間的にグローバルな動きの解析を試みる。人体の各部は、瞬間的には異なった方向に動くが、各部分の動きを例えば数秒間平均すれば、グローバルな動きの平均に収束する。一人の人間の頭、手足などの各部は協調して動くため、動き情報からそれらが連結したものであることが判る。また、ある領域内の部分が異なった方向に移動する場合には、その領域は異なる複数の物体に対応すると考えられる。この様な領域の統合や分割は、丁度人間が協調して運動する多数の移動点を同一グループと知覚する *perceptual grouping* と似ている。ただし、複数の物体が協調して動く場合には、動き情報から個々の物体を分離するのは困難であり、この場合には物体モデルなどの付加的な情報が必要である。

本報告では、一連の単眼の動画像シーケンスから歩行者を自動検出する新しい方法を提案する。まず、アルゴリズムの概要を次章で述べる。アルゴリズムは、動き情報の抽出と歩行者抽出の2つのフェーズからなり、第3章で動き情報抽出フェーズ、第4章で歩行者抽出フェーズについて述べる。続いて第5章では、複雑な実画像のシーケンスに対する実験結果を示し、第6章で考察と結論を述べる。

2 アルゴリズムの概要

開発すべきアルゴリズムは、歩行者など多数の非剛体を扱えること、実画像に於ける雑音に対してロバストであることが必要である。ここでは、課題を限定するため、カメラパラメータが既知の静止カメラからの観測を前提とした2次元並進運動のみを扱う。

ここで提案する歩行者検出アルゴリズムは、画像強度、動き情報、および移動物体モデルに基づいている。このアルゴリズムは、(1) 動き情報 (motion) の抽出、および歩行者抽出の2つのフェーズからなる。第1のフェーズでは、単眼の動画像シーケンス中の連続した2枚のフレームの局所領域マッチングによって動き情報 (初期 flow) を求める。この初期 flow では、画像の強度変化の明確な物体

輪郭付近での動きが検出され、強度の様な物体内部では動きは検出されない。これを改善するため、画像強度の情報から得られる領域画像を利用する。この領域画像は、入力画像から背景画像(移動物体を含まない画像)を差し引いて、しきい値処理することにより得られる。

歩行者は、非剛体であることから、一人の歩行者に対応する領域内の flow は一様にならない。この様な歩行者の非剛性(nonrigidity)を吸収するため、flow を数秒に渡って時間的に平滑化する。この時間的平滑化を行うと、一人の歩行者の各部の運動は、体全体の平均の運動ベクトルに収束する。このフェーズの詳細は、第3章で述べる。

第2のフェーズ：歩行者抽出では、第1のフェーズで得られた時間平均 flow を用いて、個々の物体の抽出を行う。まず、flow ベクトルの方向の粗量子化によって、flow 検出された領域をサブ領域に再分割する(簡単のため flow ベクトルの絶対値は無視)。第2のステップでは、これらの領域を任意の2つの領域の統合仮説に基づいて繰り返し併合する。これら仮説は、歩行者画像の確率的モデルによって与えられる適合度によって評価される。このモデルは、物体の大きさと運動方向に関する情報を含む。各繰り返しのステップ毎に最大適合度を与える領域の組合せに関する仮説が取り入れられ、領域の統合が行われる。このフェーズに関する詳細は、第4章で述べる。

3 flow の抽出

動き情報の抽出法としては、(1) 勾配法と(2) 特徴マッチング法の2つが提案されている[9][2]。勾配法は、画像強度の空間的および時間的勾配に基づいて密な flow を求める方法である[8][7][14]。この方法は、例えば滑らか拘束の様な付加的な拘束条件が必要であり[8]、滑らかな物体表面に対しては良好な結果を与えるが、物体輪郭付近での flow 抽出精度に問題がある。

一方、特徴マッチング法は、特徴点、エッジ、屈曲点、領域などの特徴のフレーム間の対応付けに基づく方法であり、安定な特徴抽出が前提となる。この対応付け問題は、多くの物体が画像に含まれる場合や、特徴抽出に曖昧さが残っている場合には困難になる。両方の手法ともノイズに対して敏感であり、実際の応用に現われる様な屋外実画像に対しては十分な結果が得られないことが多い。

ここでは、ノイズに対して比較的安定な特徴マッチング法を選び、特別な特徴抽出を必要としないことから局所領域の画像強度を特徴に選んだ(局所領域マッチング)。こ

れと類似の手法は、屋外実画像に対する flow の算出に Burt ら[4]も採用している。ここで提案する手法は、次の4つのステップからなる。

1. 画像強度に基づく領域分割,
2. 初期 flow の算出,
3. flow の伝播,
4. flow の時間的平滑化.

第1のステップでは、類似した強度を持つ様な領域を抽出する。この領域画像は、次のステップで抽出される初期 flow の改善に使われる(第3ステップ)。第2ステップでは、2枚の連続フレーム間の局所領域マッチングによって初期 flow を求める。得られる flow は、エッジ付近でのみ検出され、それも有限の大きさ局所マスクを使うためエッジの周辺にボケを生じる。そこで、第3ステップでは、上記の領域画像を参照しながらこの初期 flow を物体領域の内側に伝播させながら平滑化する。このステップによって物体領域内部の flow が修復される。得られた flow は、対象が剛体の場合には各物体ごとに一樣になるが、非剛体の場合には各部が異なった方向に運動するため一樣にはならない。第4ステップでは、この flow を時間的に繰り返し平滑化し、物体内部の flow が個々の物体で一樣な値に収束する様にする。

3.1 領域抽出

画像強度は、物体抽出に非常に有効な情報であるが、背景の複雑な強度変化があると抽出が困難になる。もし、動画像シーケンスから移動物体を含まない背景のみの画像が抽出できれば、背景の強度変化を入力画像から取り除くことができる。背景画像は、各画素 (i, j) 毎に時間方向のモード(画像強度の最頻値)フィルタリングによって求めることができる。この背景画像は、屋外照明の変化を追従するため一定時間毎に更新する必要がある。

入力画像 $F_{i,j}^t$ と背景画像 $B_{i,j}^t$ との差分画像 $D_{i,j}^t$:

$$D_{i,j}^t = F_{i,j}^t - B_{i,j}^t \quad (1)$$

を考えると、動かない背景の情報は消えて、動きのある部分のみが浮かび上がる。

エッジ画像を用いると異なった物体に属する類似した強度を持つ領域の輪郭がより明確に抽出することができる。エッジ画像 $E_{i,j}^t$ は、例えば Canny のエッジオペレータ[5]により得ることが出来る。ここで、エッジ上の点を $E_{i,j}^t = 1$ 、その他の点を $E_{i,j}^t = 0$ としておく。次式により差分

画像 $D_{i,j}^t$ とエッジ画像 $E_{i,j}^t$ から 3 値の領域画像 $R_{i,j}^t$ を求める。

$$R_{i,j}^t = \begin{cases} 1 & \text{if } T_d \leq D_{i,j}^t \text{ かつ } E_{i,j}^t = 0 \\ 0 & \text{if } -T_d < D_{i,j}^t < T_d \text{ または } E_{i,j}^t = 1 \\ -1 & \text{if } D_{i,j}^t \leq -T_d \text{ かつ } E_{i,j}^t = 0 \end{cases} \quad (2)$$

ここで、 $-T_d$ と T_d は実験的に得られる差分画像に対するしきい値である。この領域画像に於て、物体輪郭と背景では 0 の値、物体領域では -1 または 1 の値をとる。

3.2 初期 flow の抽出

初期 flow ベクトル $\vec{\delta} = (\delta_x, \delta_y)$ は、2 つの連続フレーム F^t, F^{t+1} の局所領域マッチングによって求める。ここで、 δ_x, δ_y は、空間座標 (i, j) と時間 (t) の離散関数である。

$$\vec{\delta} = (\delta_x, \delta_y) = \{(m, n) \mid \min_{m,n \in \{0, \dots, \pm 1\}} C_{m,n}\}, \quad (3)$$

ここで

$$C_{m,n} = \sum_{i=-1}^1 \sum_{j=-1}^1 |F_{i,j}^{t+1} - F_{i+m,j+n}^t|. \quad (4)$$

式 (3) と (4) で観測される flow は、 $-l$ から l の間の整数値をとり、局所領域マスクの大きさは計算量の低減のため 3×3 としている。

この初期 flow には、物体抽出するうえで 4 つの欠点がある。まず、(1) エッジ付近でのみ flow が検出され、物体内部の一樣強度の部分では flow は検出されない。次に、(2) 得られる flow はエッジ部分にボケて現われる。更に、(3) 有限の局所領域マスクを使うことにより窓枠問題 (aperture problem) を生じる。また、(4) 対象が非剛体の場合には個々の物体に対する flow が一様にならないため、領域境界を得にくい。以下の節では、これらに対処する方法について述べる。

3.3 Flow の伝播

初期 flow を空間的に平滑化することによって、上で述べた (1) から (3) の問題が解決可能である。ただし、背景が静止している場合にはもともと物体輪郭の外側では動いているものがないから、平滑化処理は物体輪郭の内側のみに限定する必要がある。そこで、領域境界を保存し、かつ領域境界 (エッジ) 付近の flow のボケを取り除く様に第 3.1 節で述べた領域画像を利用しながら、初期 flow を空間的に伝播させる方法を探った。この flow 伝播処理は、重み付けの平均処理の繰り返しで実現される。まず、このステップでは、領域境界の外側の点 ($R_{i,j}^t = 0$) における

flow を消去し、境界上の点 ($E_{i,j}^t = 0$) を除く全ての点に対して 3×3 のマスク処理

$$v_x^t = \frac{1}{k} \sum_{i=-1}^1 \sum_{j=-1}^1 \delta_x(i, j), \quad (5)$$

$$v_y^t = \frac{1}{k} \sum_{i=-1}^1 \sum_{j=-1}^1 \delta_y(i, j), \quad (6)$$

を繰り返すことによって flow ベクトル $\vec{v}^t = (v_x^t, v_y^t)$ を求める。ここで、

$$k = \sum_{i=-1}^1 \sum_{j=-1}^1 a_{i,j}, \quad (7)$$

$$a_{i,j} = \begin{cases} 1 & \text{if } \delta_x \neq 0 \text{ or } \delta_y \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

である。

3.4 Flow の時間的平滑化

物体が非剛体の場合、時刻 t に於ける flow ベクトル \vec{v}^t は同一物体領域内でも場所によって異なった値になる (欠点 (4))。しかし、この flow を適当な時間だけ平均すれば、その値は物体全体の平均移動方向に収束すると考えられる。ここでは、以下の式に従って、瞬間の flow ベクトル \vec{v}^t を時間的に平滑化し、時間平均 flow ベクトル \vec{V}^t を求める。

$$\vec{V}^t = (V_x^t, V_y^t) = \frac{1}{1 + w^{t-1}} \{ \vec{v}^t(i, j) + w^{t-1} \vec{V}^{t-1}(i^*, j^*) \}, \quad (8)$$

ここで、

$$i^* = i - [v_x^{t-1}(i, j)], \quad (9)$$

$$j^* = j - [v_y^{t-1}(i, j)], \quad (10)$$

w^{t-1} は一時点前 ($t-1$) の flow に対する重み、式 (9)、(10) に於ける括弧 $[\]$ は四捨五入を意味する。これらの式に於て、瞬間の flow ベクトル \vec{v}^{t-1} は、一時点前の flow ベクトル \vec{V}^{t-1} の位置ずれの補正にも用いられている。

4 歩行者の抽出

ここで動画シーケンス中の歩行者を抽出する方法について述べる。前章で述べた処理で得られた flow から個々の物体 (歩行者) を抽出するためには以下の 3 つの解決すべき問題がある。

- 統合された物体領域: 連続した 1 つの領域が複数の物体に対応する場合。
- 分割された物体領域: 物体が複数の領域に対応する場合。

- 物体の遮蔽: 物体が他の物体の陰になる場合.

そこで我々は、以下の2つのステップからなる領域分割/統合法を提案する。

1. 動き情報に基づく領域分割.
2. 物体モデルに基づく領域統合.

第1のステップでは、運動方向の量子化により、各領域をサブ領域に分割する。第2のステップでは、歩いている人間の画像のモデルを利用してこれら領域の統合を行う。この統合処理では、任意の2つの領域を統合する仮説を設け、最も妥当な仮説に基づいて領域統合を繰り返し行う。このステップは、人間の運動知覚に於ける知覚的統合 (perceptual grouping) に類似している。

4.1 Motion に基づく領域分割

第1のフェーズで flow の抽出された領域には、隣接した2つ以上の物体が含まれることが少なくない。もし、それらが異なった方向に移動しているならば、動き情報から物体の輪郭を検出できる。物体輪郭の検出には、速度場のクラスタリングによる方法が考えられるが、含まれるクラスタ数が未知であり必ずしも容易でない。そこで、ここでは flow の方向の粗量子化によって領域をサブ領域に分割する方法を採った。画像中の各点 (i, j) に於ける運動方向 $d_{i,j}$ ($0 \leq d_{i,j} < 2\pi$) は、flow ベクトル

$$\vec{V}^t = (V_x^t(i, j), V_y^t(i, j)),$$

から計算され、これを量子化することによって複数のサブ領域に分割し物体の遮蔽輪郭を検出する。

4.2 物体モデルに基づく領域統合

個々の歩行者は、以上の処理で得られる画像上では複数の領域に分割されている。ここでは、これらの領域を統合して、個々の歩行者を抽出する処理について述べる。

(1) 物体モデル

前ステップで分離された隣接する領域は、物体の確率的モデルを手がかりに併合される。このモデルは、単純であり、確率密度関数 $p(w, h, \Delta\theta)$ で表わされる。ここで、 w, h , および $\Delta\theta$ は、それぞれ物体の幅、高さ、および運動方向の差である。もし、これらが互いに独立であるなら、

$$p(w, h, \Delta\theta) = p(w)p(h)p(\Delta\theta). \quad (11)$$

と書ける。また、これらがそれぞれ適当な平均値 ($\mu_w, \mu_h, 0$) の周りに分散 ($\sigma_w^2, \sigma_h^2, \sigma_{\Delta\theta}^2$) で正規分布するとすれば、

これらの密度関数 $p(w), p(h), p(\Delta\theta)$ は、以下の式で与えられる。

$$p(w) = \frac{1}{\sigma_w \sqrt{2\pi}} e^{-(w-\mu_w)^2/2\sigma_w^2}, \quad (12)$$

$$p(h) = \frac{1}{\sigma_h \sqrt{2\pi}} e^{-(h-\mu_h)^2/2\sigma_h^2}, \quad (13)$$

$$p(\Delta\theta) = \frac{1}{\sigma_{\Delta\theta} \sqrt{2\pi}} e^{-\Delta\theta^2/2\sigma_{\Delta\theta}^2}. \quad (14)$$

なお、これらに含まれるパラメータ $\mu_w, \mu_h, \sigma_w, \sigma_h, \sigma_{\Delta\theta}$ は、実験的に求められる。

(2) 統合仮説

図1に示す2つの領域 i と j の統合処理を考える。各領域の x 座標の最小および最大値をそれぞれ x^s, x^e ($x^s \leq x^e$)、 y 座標の最小および最大値をそれぞれ y^s, y^e ($y^s \leq y^e$) とするとき、両者を統合して得られる新しい領域 k の幅 w_k と高さ h_k は、それぞれ

$$w_k = \max(x_i^e, x_j^e) - \min(x_i^s, x_j^s), \quad (15)$$

$$h_k = \max(y_i^e, y_j^e) - \min(y_i^s, y_j^s). \quad (16)$$

で与えられる。また、統合後の運動方向の差は、

$$\Delta\theta_k = \theta_j - \theta_i. \quad (17)$$

となる。

もし、 w_k と h_k が共に平均値 μ_w and μ_h に比べて十分小さく、 $|\Delta\theta_k|$ もほぼ0であれば、領域 j を i に統合する仮説は受け入れられる。一方、 w_k または h_k が平均値 (それぞれ μ_w, μ_h) よりも十分大きいか、または $|\Delta\theta_k|$ が大きい時には、その仮説は棄却されるべきである。この様な振舞いをする統合仮説の適合度として次式で示す確率 P を選んだ。

$$P = P(w_k < w)P(h_k < h)P(|\Delta\theta_k| < |\Delta\theta|), \quad (18)$$

ここで、右辺の各項は互いに独立であり、以下の式で表わされるものとする。

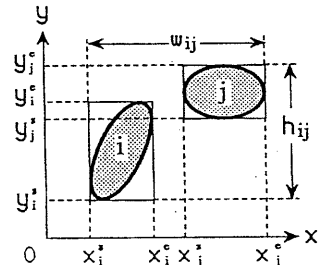


図1: 領域統合

$$P(w_k < w) = 1 - \int_{-\infty}^{w_k} p(w)dw \quad (19)$$

$$P(h_k < h) = 1 - \int_{-\infty}^{h_k} p(h)dh \quad (20)$$

$$P(|\Delta\theta_k| < |\Delta\theta|) = 1 - 2 \int_0^{\Delta\theta_k} p(\Delta\theta)d(\Delta\theta) \quad (21)$$

(3) 領域統合

領域統合処理は、統合順序によって大きく結果が変わるため、以下に示す手順により徐々に進める必要がある。

1. 2つの領域 i と j ($i \neq j$) の全ての組合せについて統合仮説を生成する。
2. 領域 i に関する仮説のうち式 (18) の適合度を最大にする仮説を選択する。
3. 選ばれた仮説に対する適合度 P がしきい値 P_{th} よりも大きければ、領域 i と j を統合して新しい領域 k を作成する (領域 i と j は削除)。
4. 新たな統合処理が起こらなくなるまで、ステップ1から3を繰り返す。

5 実験

アルゴリズムの評価には、実際の屋外景観の動画シーケンスを用いた。このシーケンスは、横断歩道付近を斜め上から固定カメラにより連続的に撮影したもので、各画像フレームには30～50人程度の歩行者が含まれる。この為、頻繁に物体の遮蔽 (occlusion) が起こっている。図2に入力画像フレームの例を示す。撮影に使用したCCDカメラでは、インタレイス走査のため奇数フィールドと偶数フィールドの間でブレを生ずる。そこで実験には、片フィールドのみを使用し、縦横の分解能を合わせるため処理画像の画素数を 256×240 とした。また、計算量を減少させるためフレームレートを半分の15枚/秒とした。



図2: 入力画像の例



図3: 背景画像



図4: 差分画像

図3は、240フレーム (8秒の長さのシーケンスに相当) から等間隔にサンプルした20枚の画像を入力として、時間方向のモードフィルタリングによって得た背景画像を示す。同図では、歩行者の過密な画像右上と左下角付近で若干のノイズが見られるが、ほとんど完全に運動物体が消去されている。より長い動画シーケンスが使用できれば、より善い結果が期待できる。

図4(a)(b)は、それぞれ図3に示した背景画像を用いて得られた(a)フレーム #064と(b)フレーム #128に対する

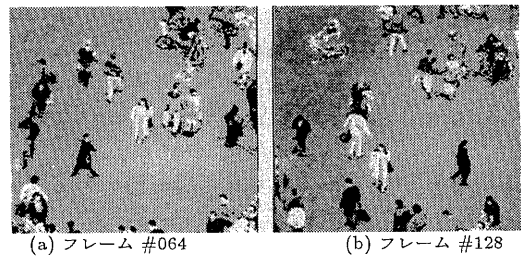


図5: 領域画像

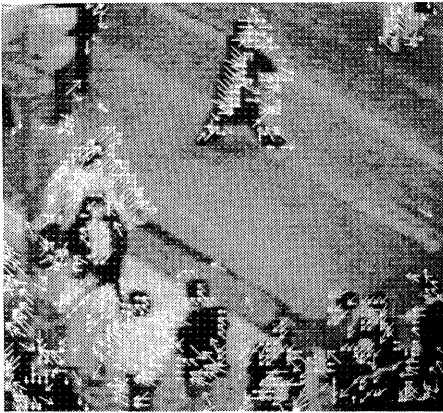


図 6: 初期 flow

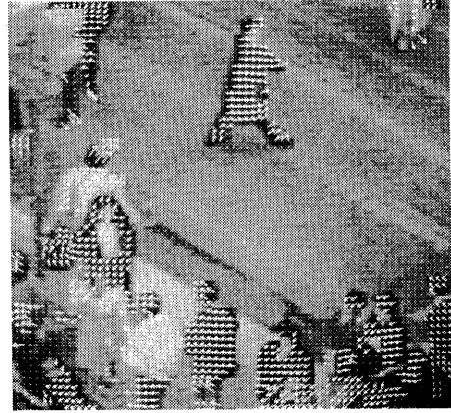


図 8: 時間的に平滑化された flow



図 7: 空間的平滑化 flow

る差分画像の例である。図 5(a)(b) は、これらの画像をしきい値処理して得られた領域画像の例である。

図 6 は、2 つの連続フレーム #64, #66 から得られた初期 flow の例 (部分) であり、物体輪郭付近に集中して flow が検出されているのが判る。Flow の空間的伝播処理では、図 7 に示すように初期 flow が物体の輪郭付近から内部に向かって伝播され大幅に改善されている。

図 8 は、図 7 に示した画像に対する時間的平滑化処理の結果を示している。個々の人物に対する flow の一様性は、10 回程度の平滑化の繰り返し (約 2/3 秒に相当) ではほぼ収束した。

図 9 は、シーケンス中の 2 枚の画像に対する歩行者検出結果を示す。図中の矩形は検出された歩行者の領域である。同図では、若干のエラーも見られるが、頻繁に遮蔽が起こっ

ているにも拘らず殆ど正しい結果が得られている。

6 むすび

歩行者を動画シーケンスから自動的に検出する方法について述べた。この方法は、(1)flow 抽出、および (2) 歩行者抽出の 2 つのフェーズからなる。第 1 のフェーズでは、初期 flow の精度向上のため、移動物体のみを含む領域画像を利用した。この領域画像は、動画シーケンスの時間軸方向のモードフィルタリングによって得た背景画像と入力画像との差のしきい値処理により得ている。実験で得られた背景画像では、歩行者など全ての運動物体が消去され、背景のみが残った。Flow を時間方向に平滑化することによって歩行者の非剛性を取り除くことができた。実験では、約 10 回程度の平滑化により、各歩行者に対する flow がほぼ一様な値に収束した。

歩行者抽出フェーズの第 1 ステップでは、flow の得られている各領域を flow 方向の粗量子化によってサブ領域に分割した。次のステップでは、これらの領域を歩行者モデルに基づいて統合し、個々の歩行者領域を抽出した。実際の屋外景観の動画シーケンスに対する実験結果により、提案手法が歩行者の抽出に有効であることが確認出来た。

歩行者の検出精度は、アルゴリズムの時間方向への拡張によって改善されると考えられ、これが今後の課題である。

謝辞

本研究の機会を与えて戴くとともに研究の進め方についてご討論戴いたヒューマンインタフェース研究所の小林 幸雄 部長、末永 康仁 グループリーダ、石井 健一郎 グループリーダに感謝致します。また、研究内容についてご討論戴いた赤松 茂 主幹

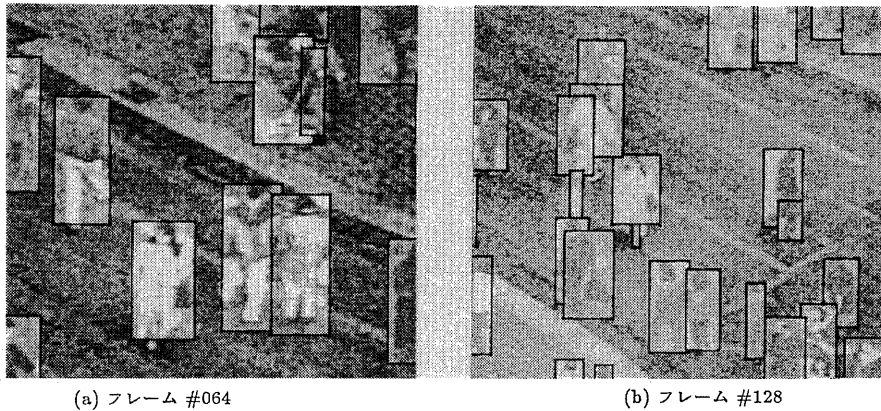


図 9: 歩行者検出結果

研究員、Dan E. Gutfinger と Rohit Agarwal (共に UCI : カリフォルニア大学アーバイン校)、動画像データを提供して戴いた高橋 裕子 氏、橋本 昌明 氏、草稿にコメントを戴いた倉掛 正治 氏 (当時、USC : 南カリフォルニア大学) に感謝します。本研究は、NTT のサポートにより 89 年冬から 91 年春にかけて UCI で行ったものである。

参考文献

- [1] Adiv, G., "Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, No. 4, pp. 384-401, July 1985.
- [2] Aggarwal, J. K. and N. Nandhakumar, "On Computation of Motion from Sequences of Images — A Review," *Proc. of IEEE*, Vol. 76, No. 8, pp. 917-935, 1988.
- [3] Braunstein, M. L., D. D. Hoffman, and F. E. Pollick, "Discriminating Rigid from Nonrigid Motion: Minimum Points and Views," *Perception & Psychophysics*, 43(3), pp. 205-214, 1990.
- [4] Burt, P. J., J. R. Bergen, et al., "Object Tracking with a Moving Camera: An Application of Dynamic Motion Analysis," *Proc. of IEEE Workshop on Visual Motion*, pp. 2-12, Irvine, CA, March 20-22, 1989.
- [5] Canny, J., "A Computational Approach to Edge Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 6, pp. 679-698, November 1986.
- [6] Chow, C. K. and T. Kaneko, "Automatic Boundary Detection of the Left Ventricle from Cineangiograms," *Computers in Biomedical Research*, Vol. 5, pp. 388-410, 1972.
- [7] Heeger, D. J., "Optical Flow Using Spatiotemporal Filters," *International Journal of Computer Vision*, Vol. 1, No. 1, pp. 279-302, 1988.
- [8] Horn, B. K. P. and B. G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, Vol. 17, pp. 185-203, 1981.
- [9] Horn, B. K. P., "Robot Vision," *The MIT Press and McGraw-Hill Book Company*, 1986.
- [10] Huang, T. S., "Modeling, Analysis, and Visualization of Nonrigid Object Motion," *Proc. of International Conf. on Pattern Recognition*, Vol. 1, pp. 361-364, Atlantic City, NJ, June 1990.
- [11] Johansson, G., "Perception of Motion and Changing Form," *Scandinavian J. Psychology*, Vol. 5, pp. 181-208, 1964.
- [12] Klinker, G. J., S. A. Shafer, and T. Kanade, "The Measurement of Highlights in Color Images," *International Journal of Computer Vision*, Vol. 2, pp. 7-32, 1988.
- [13] Murray, D. W. and B. F. Buxton, "Scene Segmentation from Visual Motion Using Global Optimization," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 2, pp. 220-228, 1987.
- [14] Nagel, H. and W. Enkelmann, "An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 5, pp. 565-578, 1986.
- [15] Ohlander, R., K. Price, and D. R. Reddy, "Picture Segmentation Using a Recursive Region Splitting Method," *Computer Graphics and Image Processing*, Vol. 8, pp. 313-333, 1978.
- [16] O'Rourke, J. and N. L. Badler, "Model-Based Image Analysis of Human Motion Using Constraint Propagation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, No. 6, pp. 522-536, 1980.
- [17] Palmer, S. E., "The Psychology of Perceptual Organization: A Transformational Approach," in J. Beck, B. Hope and A. Rosenfeld (Ed.), "Human and Machine Vision," pp. 269-339, *Academic Press, Inc.*, 1983.
- [18] Peleg, S. and H. Rom, "Motion Based Segmentation," *Proc. of International Conf. on Pattern Recognition*, Atlantic City, NJ, June 16-21, 1990.