

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4686663号
(P4686663)

(45) 発行日 平成23年5月25日(2011.5.25)

(24) 登録日 平成23年2月25日(2011.2.25)

(51) Int.Cl.		F I			
G06T	7/20	(2006.01)	G06T	7/20	300Z
H04N	7/18	(2006.01)	H04N	7/18	G
			H04N	7/18	K

請求項の数 8 (全 45 頁)

(21) 出願番号	特願2008-524722 (P2008-524722)	(73) 特許権者	504132272
(86) (22) 出願日	平成19年2月8日(2007.2.8)		国立大学法人京都大学
(86) 国際出願番号	PCT/JP2007/052271		京都府京都市左京区吉田本町36番地1
(87) 国際公開番号	W02008/007471	(74) 代理人	100084375
(87) 国際公開日	平成20年1月17日(2008.1.17)		弁理士 板谷 康夫
審査請求日	平成22年1月14日(2010.1.14)	(74) 代理人	100121692
(31) 優先権主張番号	特願2006-189849 (P2006-189849)		弁理士 田口 勝美
(32) 優先日	平成18年7月10日(2006.7.10)	(74) 代理人	100125221
(33) 優先権主張国	日本国(JP)		弁理士 水田 慎一
		(72) 発明者	官本 龍介
			京都府京都市左京区吉田本町 国立大学法 人京都大学大学院情報学研究科内
		(72) 発明者	芦田 純平
			京都府京都市左京区吉田本町 国立大学法 人京都大学大学院情報学研究科内
			最終頁に続く

(54) 【発明の名称】 歩行者トラッキング方法及び歩行者トラッキング装置

(57) 【特許請求の範囲】

【請求項1】

時間連続的に撮影した画像中に現れる特定の歩行者領域を時系列的に与えられ、その歩行者領域の動きをモンテカルロフィルタを用いて推定して時間方向に関連づけることによりその歩行者領域を追跡する歩行者トラッキング方法であって、

歩行者モデルを複数の節点とその節点間を連結する複数のスティックとを有するスケルトンモデルとして構成し、前記スケルトンモデルの状態と1対1の対応をなす点を表す多次元空間である状態空間に、その状態空間の点が示す状態の実現確率の属性をその点に付与して成る粒子を、定義すると共に、当該状態空間に初期状態として複数のスケルトンモデルの粒子を設定する工程と、

第1の時刻における前記状態空間に設定された粒子の組に対し、モンテカルロフィルタの方法に基づいて当該第1の時刻より後の第2の時刻における状態を表す粒子の組を予測する工程と、

前記第2の時刻において与えられた歩行者領域を2値画像化すると共にその2値画像の距離変換画像から成る観測データを生成する工程と、

前記予測工程によって予測された各スケルトンモデルと前記距離変換画像とを比較することにより、第2の時刻における各スケルトンモデルの状態の尤度を算出する工程と、

前記工程によって算出した各スケルトンモデルの尤度に基づいて、前記状態空間におけるスケルトンモデルの粒子の組を抽出すると共に、抽出した粒子の組を第2の時刻における前記状態空間の新たな粒子の組として設定する工程と、備え、

前記粒子の組を予測する工程から前記抽出した粒子の組を新たな粒子の組として設定する工程までの工程を、時刻を進める毎に実行することにより、スケルトンモデルの状態空間における粒子の動きを推定して時間方向に関連づけることにより歩行者領域を追跡することを特徴とする歩行者トラッキング方法。

【請求項 2】

前記スケルトンモデルは、胴体スティックと、前記胴体スティックの上端に交差して接続した肩幅スティックと、前記胴体スティックの下端に交差して接続した腰幅スティックと、前記胴体スティックの延長部としての頭長スティックと、前記腰幅スティック及び肩幅スティックの両端部にそれぞれ接続された 4 本の四肢スティックと、の 8 本のスティックからなることを特徴とする請求項 1 に記載の歩行者トラッキング方法。

10

【請求項 3】

前記スケルトンモデルは、前記肩幅及び腰幅スティックの長さをそれぞれゼロとする状態と、前記肩幅、腰幅、及び両手スティックの長さをゼロとする状態と、前記肩幅及び両手スティックの長さをゼロとする状態と、を取り得ることを特徴とする請求項 2 に記載の歩行者トラッキング方法。

【請求項 4】

前記スケルトンモデルの 1 つの状態は、そのスケルトンを構成する各スティックの位置と角度とによって決定され、前記各スティックの相互の空間配置が人体の構造上可能な空間配置となるように制約されていることを特徴とする請求項 1 乃至請求項 3 のいずれか一項に記載の歩行者トラッキング方法。

20

【請求項 5】

前記尤度を算出する工程で用いる距離変換画像は、歩行者領域を表す黒の輪郭部から輪郭内部に行くほど高輝度と成るように白黒濃度で表現した画像であり、

前記予測された各スケルトンモデルの尤度は、スケルトンモデルと距離変換画像とを重ねた状態で、スケルトンモデルを構成するスティック上の複数の点から、各点における該スティックの法線方向に存在する距離変換画像の輝度極値点までの距離を用いて算出されることを特徴とする請求項 1 乃至請求項 4 のいずれか一項に記載の歩行者トラッキング方法。

【請求項 6】

前記粒子の組を予測する工程と前記新たな粒子の組を設定する工程とにおいて乱数が用いられ、前記乱数が疑似乱数生成アルゴリズムを用いて生成されることを特徴とする請求項 1 乃至請求項 5 のいずれか一項に記載の歩行者トラッキング方法。

30

【請求項 7】

前記予測工程によって予測された各スケルトンモデルにおける、胴体スティックと頭長スティックとの節点の座標の分散を算出し、当該分散が一定の値を超える場合には追跡を終了する工程、をさらに備えることを特徴とする請求項 1 乃至請求項 6 のいずれか一項に記載の歩行者トラッキング方法。

【請求項 8】

時間連続的に撮影された画像を時系列的に受信する画像受信手段と、

前記画像受信手段によって受け取った画像から歩行者候補領域を抽出し、前記抽出した歩行者候補領域を歩行者領域と非歩行者領域とに分類すると共に、特定の歩行者領域を選択する歩行者領域選択手段と、

40

前記歩行者領域選択手段によって選択された歩行者領域を時系列的に与えられ、その歩行者領域の動きを請求項 1 乃至請求項 7 のいずれか一項に記載の歩行者トラッキング方法を用いて追跡するトラッキング手段と、

前記トラッキング手段によって追跡される歩行者領域を時系列的に表示する歩行者軌跡表示手段と、を備えることを特徴とする歩行者トラッキング装置。

【発明の詳細な説明】

【技術分野】

【0001】

50

本発明は、連続的に歩行者を撮影した画像に基づいて歩行者を追跡する歩行者トラッキング方法および歩行者トラッキング装置に関する。

【背景技術】

【0002】

従来から、画像データに基づいて移動体の動き、特に人の動きを自動的に追跡する方法や装置があり、例えば、音楽分野やスポーツ分野において、動作の解析と評価に基づく能力開発や様々な目的のために開発されている。一例を挙げると、コンピュータに入力されたダンスする被験者の映像から人物シルエット画像を抽出し、色処理によって人物シルエット画像から被験者の各部位を検出すると共に人物シルエット画像をスケルトン画像に変換し、そのスケルトン画像をハフ変換して各部位を直線近似し、各部位をカルマンフィルタにより時間追跡する身体動作解析装置が知られている（例えば、特開2005-339100号公報参照）。

10

【0003】

上述の身体動作解析装置は、時間追跡されたハフパラメータにSVD (Singular Value Decomposition; 特異値分解) を施して身体動作の動き特徴量を抽出し、その後、抽出した動き特徴量の時系列データをフーリエ変換して周波数解析し、被験者の身体動作全体のリズム要素を抽出評価する装置である。

【0004】

また、画像データに基づいて歩行者を自動的に追跡する歩行者トラッキング方法及び歩行者トラッキング装置は、事故防止や監視などの用途に応えるため、高精度化、効率化が進められている。すなわち、誤動作（誤追跡）のない、より確実で、より高速な、歩行者トラッキング方法や装置が求められている。

20

【0005】

ところで、画像を処理して歩行者の動きを推定し、時間方向に関連づける手段として、種々のフィルタ技術が用いられる。フィルタは、画像データなどの入力信号に対して望みの応答の推定を出力する方法又は装置である。歩行者トラッキングに用いられるフィルタは、雑音の加わっている現在の入力信号から雑音を除去すると共に未来の信号値を、望みの応答の推定として出力する。このようなフィルタによる未来の推定は、フィルタによる予測と呼ばれる。

【0006】

30

例えば、カルマンフィルタは、上述の特許文献（特開2005-339100号公報）で用いられているように、移動体一般の追跡を行うオブジェクトトラッキングの分野で広く用いられており、歩行者トラッキングへも適用されている。

【0007】

カルマンフィルタを用いるトラッキングの概要を説明する。カルマンフィルタは、観測ベクトル y_t から逐次的に各時刻における状態ベクトル x_t を推定する。ここで、 y_t 、 x_t 、後述の F_t などにおける下付文字の t は、ある特定の時刻を表し、その時刻よりも1ステップ前の時刻は、 $t-1$ によって表わされる。すなわち、観測（例えば、時系列画像の撮影）は、時間間隔（ステップ）毎に行われる。これらの下付き文字による時刻の表記方法は、以下においても同様に用いられる。観測ベクトル y_t は、観測可能な時系列データによって数学的に定義される観測空間のベクトルである。また、状態ベクトル x_t は、観測ベクトルを本質的に決定付けるシステムの状態を示す状態空間と呼ばれる空間のベクトルであり、ガウス-マルコフ過程 (Gauss-Markov process) に従うと仮定される（例えば、片山徹著「応用カルマンフィルタ」朝倉書店、1983年、参照）。

40

【0008】

また、カルマンフィルタは、状態ベクトルの時間方向の遷移を特徴付けるシステムモデルの式、 $x_t = F_t x_{t-1} + G_t v_t$ 、及び、状態ベクトルから観測ベクトルへの写像を特徴付ける観測モデルの式、 $y_t = H_t x_t + w_t$ の両者に線形性、及びガウス性を仮定する。ここで、 v_t と w_t は、それぞれプラント雑音、観測雑音と呼ばれるガウス白色

50

雑音であり、 F_t 、 G_t 、 H_t は、それぞれ状態遷移行列、駆動行列、観測行列と呼ばれる行列である。これらの式の形から、観測ベクトル y_t と状態ベクトル x_t が、線形性を有することがわかる。

【0009】

上述の各モデルにガウス性を仮定することは、状態空間における状態ベクトル x の確率分布、すなわち状態確率分布 $p(x)$ にガウス分布の仮定を置くことに相当する。歩行者トラッキングにおいて、状態確率分布がガウス分布から大きく外れる場合として、例えば、一時的に歩行者が陰に隠れるオクルージョンの存在、トラッキングターゲット（歩行者）の速度の急変、トラッキングターゲットに似た複数のオブジェクトの存在、などの状況がある。このような場合にカルマンフィルタを適用すると、実際の状態確率分布として、
図36(a)に示すようなガウス分布とは異なる分布が想定されるにも拘わらず、図36
(b)に示すガウス分布を用いて状態を推定することになる。従って、ガウス分布を前提
としているカルマンフィルタ適用の限界のため、このような場合に、十分な精度の状態推
定ができなくなる。

10

【0010】

そこで、ガウス性や線形性を前提としないモンテカルロフィルタを用いてトラッキングを行う手法が提案され、コンデンセーション(CONDENSATION: conditional density propagation)と呼ばれている(例えば、ミカエル・イザード(Michael Isard)、アンドリュー・ブレイク(Andrew Blake)著「コンデンセーション-コンディショナル デンシティ プロパゲーション フォー ヴィジュアル トラッキング(Conditional Density Propagation for Visual Tracking)」、インターナショナル ジャーナル オン コンピュータ ビジョン(International Journal on Computer Vision)、Vol.29, pp5-28(1989)、参照)。

20

【0011】

モンテカルロフィルタを用いる場合、カルマンフィルタを用いる場合と同様に、観測ベクトルに基づいて、各時刻における状態ベクトルが逐次的に推定される。モンテカルロフィルタを用いる推定において、状態確率分布は、状態空間の1点を示すベクトルを保持する粒子の分布に基づいて生成される。従って、モンテカルロフィルタは、上述のカルマン
フィルタにおけるシステムモデルと状態空間モデルとを一般化した非線形、非ガウス型の
モデルを扱うことができる(例えば、北川源四朗著「時系列解析入門」岩波書店、2005年、参照)。

30

【0012】

そこで、ガウス性などの仮定を置く従来の手法によると失敗するような、オクルージョンの存在や速度急変の状況においても、コンデンセーションによると、高精度な確率的状態推定、すなわち誤動作のより少ないトラッキングが実現可能と考えられる。

【0013】

(モンテカルロフィルタの概要)

ここで、モンテカルロフィルタの概要を説明する。モンテカルロフィルタにおけるシステムモデルと観測モデルとは、それぞれ次の式(1)(2)で表わされる。

40

【数1】

システムモデル

$$x_t = F(x_{t-1}, v_t) \quad (1)$$

観測モデル

$$y_t = H(x_t, w_t) \quad (2)$$

50

【 0 0 1 4 】

状態空間における状態ベクトル x_t の状態確率分布 $p(x_t)$ は、次式 (3) (4) のように N 個の粒子のセット $\{s_t^{(n)}, n=1, \dots, N\}$ によって表現される。ここで、 $s_t^{(n)}$ は n 番目の粒子が保持する状態空間 X の 1 点を示すベクトルであり、 $\delta(x)$ はデルタ関数である。

【数 2】

$$p(x_t) \cong \frac{1}{N} \sum_{n=1}^N \delta(x_t - s_t^{(n)}) \quad (3)$$

$$\delta(x) = \begin{cases} +\infty & x=0 \\ 0 & x \neq 0 \end{cases} \quad (4)$$

10

【 0 0 1 5 】

モンテカルロフィルタにおける状態確率分布は、離散的な粒子の粗密によって表現される。例えば、本来の確率分布が図 3 7 (a) に示される分布の場合に、モンテカルロフィルタにおいては、図 3 7 (b) に示すように、確率分布が離散的な粒子の粗密によって表現される。従って、粒子の数が多きほど、状態確率分布の表現がより精度良くなる。このような粒子による表現によって、どのような状態確率分布も表現可能となる。

20

【 0 0 1 6 】

(モンテカルロフィルタを用いた状態推定アルゴリズム)

次に、上述のモンテカルロフィルタを用いた状態推定のアルゴリズムを説明する。図 3 8 は、 N 個の各粒子 $s^{(1)}, s^{(2)}, \dots, s^{(N)}$ に対して時間ステップ毎に行う処理を示す。この図において、横軸は時間軸であり、縦軸は状態空間 (1 次元で表されている) を示す。また、黒丸や破線白丸で示している粒子の大きさは、その状態があり得る尤度 (もっともらしさ、実現性) を示す。モンテカルロフィルタを用いる処理は、この図に示すように、予測、計測、再抽出 (再設定) の 3 ステップの処理の繰り返しである。

【 0 0 1 7 】

上述の 3 ステップの繰り返しにより、ある時刻 t における状態確率分布 $p(x_t)$ が、その前の時刻 $t-1$ の確率分布 $p(x_{t-1})$ と観測データとを用いて逐次的に求められ、各時刻における状態確率分布が次々推定される。そして、状態確率分布は、ガウス性を前提とすることなく、各時刻の状態に応じて、柔軟に決められたものである。従って、状態確率分布が観測データによって補正され、かつ、その補正された状態確率分布を用いて、次の状態確率分布が求められるので、トラッキングの結果を表す状態空間における粒子の軌跡がより確からしいものとなる。

30

【 0 0 1 8 】

予測ステップでは、各粒子 ($n=1, \dots, N$) に対し、過程確率密度 $p(x_t | x_{t-1} = s_{t-1}^{(n)})$ に従って、次状態 $s'_t^{(n)}$ を予測する (以下、上述の北川源四朗著「時系列解析入門」参照)。

40

【 0 0 1 9 】

計測ステップでは、各粒子に対し、予測された状態の尤度 $y_t^{(n)}$ を観測確率密度 $p(y_t | x_t)$ に従って算出する。言い換えると、このステップでは、各粒子に対応するトラッキングターゲットモデルの状態と、観測データ (トラッキングターゲットの画像) との、似かより具合 (likelihood) を、適切に設定した比較方法に基づいて比較することにより求める。ここで y_t は時刻 t における観測ベクトル (観測データ) である。

【 0 0 2 0 】

再抽出ステップでは、以下の (i) (ii) (iii) の処理を、粒子の個数 N に応じ

50

てN回繰り返すことにより、時刻 t における粒子のセット $\{s_t^{(n)}, n = 1, \dots, N\}$ を抽出する。言い換えると、このステップでは、予測した状態を表す粒子の尤度に基づいて、尤度の大きな粒子のところには粒子数を多くし、尤度の小さな粒子のところには粒子数を少なくしたり、ゼロにしたりして、状態空間におけるN個の粒子の再配分（再設定）を行い、観測データによる補正を反映した、時刻 t における状態確率分布が決定される。

【数3】

(i) 一様分布に従う乱数 $u_t^{(n)} \in [0,1]$ を生成する。

(ii) 次式を満たす自然数 i を求める。

$$\frac{1}{C} \sum_{l=1}^{i-1} \pi_t^{(l)} < u_t^{(n)} \leq \frac{1}{C} \sum_{l=1}^i \pi_t^{(l)} \quad (5)$$

ただし、

$$C = \sum_{l=1}^N \pi_t^{(l)} \quad (6)$$

(iii) $s_t^{(n)} = s_t^{\prime(i)}$ とする。

【0021】

上述した予測、計測、再抽出（再設定）の3ステップの処理によって、時刻 t における粒子の状態確率分布 $p(x_t)$ が得られる。モンテカルロフィルタを用いる場合、アプリケーションに応じて、状態空間 X 、すなわちターゲットのモデルなど、をどのように構成するか、予測ステップにおいて状態遷移、特に遷移に対する制約条件など、をどのように行うか、計測ステップにおける粒子の尤度の算出法、すなわち比較方法など、をどうするか、などについて適切に設定する必要がある。

【0022】

次に、コンデンセーションを用いる例として、ターゲットの輪郭を用いてターゲットを追跡する輪郭トラッキングを説明する。輪郭トラッキングの手法では、トラッキングターゲットの輪郭をB-スプライン曲線によってモデル化し、そのスプライン曲線の制御点の座標値などからなる空間を状態空間として定義する。そして、その状態空間における点を指し示す状態ベクトルの動き（遷移）を、モンテカルロフィルタを用いて推定（予測）する。言い換えると、状態空間の1つの点が、輪郭の1つの状態に一対一対応をしており、状態空間において、時間経過とともに、ある点（状態）から、ある点（状態）に現在の状態が移動、すなわち遷移することになる。その遷移は、ある制約条件のもとで確率的に実現されると考えられる。

【0023】

状態遷移を予測する場合に、予め遷移の可能性のある状態を制限しておくことにより、つまり状態遷移に制約をかけることにより、予測の精度を上げることができる。従来行われているコンデンセーションによる輪郭トラッキングでは、主成分分析による事前学習を行うことによって、状態遷移に制約をかけることが行われている。以下、コンデンセーションによる輪郭トラッキングにおける状態空間、状態遷移、及び尤度の算出について示す。

【0024】

（状態空間）

トラッキングターゲットの輪郭をB-スプライン曲線により近似し、B-スプライン曲

10

20

30

40

50

線の制御点の位置や速度を状態空間 X として定義する。

【0025】

(状態遷移)

主成分分析を用いて教師データを事前に学習する。状態空間が M 次元の場合、第 1 主成分ベクトルから第 L 主成分ベクトル ($L < M$) の線形結合を基に状態遷移 ($s_{t+1} - s_t$) を決定して状態遷移の自由度を M から L に下げる。これにより、教師データが偏っている方向に状態が散らばるように、すなわち、教師データの特徴に従うように、状態遷移を制約する。

【0026】

(尤度の算出)

粒子の尤度は、以下の手順に基づき、 B -スプライン曲線と入力画像との比較によって算出される。まず、予測された B -スプライン曲線上の K 個の点からなる基点を設定し、そこから法線方向に長さ μ の直線を設定する。次に、この直線上において画像の特徴点であるエッジ(ターゲット部分の画像輪郭)を検出し、 k 番目の基点から検出されたエッジまでの距離を d_k とする。輪郭の状態を表す状態空間における粒子の尤度は、この d_k を用いて、次式(7)のように算出される。

$$\pi \propto \exp \left\{ - \sum_{k=1}^K \frac{1}{2rK} (\min(d_k, \mu))^2 \right\} \quad (7)$$

【0027】

上述のコンデンセーションによる輪郭トラッキングは、手のひらや葉っぱなどのトラッキングに適用され、優れた成果をあげている。また、コンデンセーションのフレームワークを用いた歩行者トラッキング手法が、提案がされている(例えば、イザード(Isard)、マッコーミック(MacCormick)著「アベイジアンマルチプル-ブロブトラッカ(A Bayesian Multiple-Blob Tracker)」、IEEE インターナショナル コンファレンス オン コンピュータ ビジョン (IEEE International Conference on Computer Vision)、pp34-41(2001)、参照)。

【0028】

しかしながら、コンデンセーションによる輪郭トラッキングを歩行者トラッキングに適用した場合、次のような問題点がある。歩行者のようなオブジェクトは、輪郭の遷移の方向や大きさが、時間とともに著しく変化する。このようなオブジェクトの輪郭に対して、状態遷移を適切に制約することは一般に困難であり、従って、コンデンセーションによる輪郭トラッキングは歩行者のトラッキングには不適である。

【0029】

歩行者の輪郭の遷移を、図39(a)(b)(c)を参照して説明する。これらの図は、ある実シーケンスにおける連続する3フレームの画像から検出した歩行者領域を示す。図39(a)(b)の2つの歩行者領域の間には、上半身の輪郭に大きな変動はないが、下半身の輪郭に大きな変動が見られる。一方、図39(b)(c)の間では、上半身の輪郭が大きく変動しているが、下半身の輪郭に大きな変動はない。このように、歩行している人の輪郭は、時間と共に変化、すなわち、状態が遷移し、その変化する部位や変化の大きさも時間的に変化している。

【0030】

また、上述したように、コンデンセーションによる輪郭トラッキングでは、主成分分析によって状態遷移を制約する。従って、状態空間における教師データが、ある特定の方向に偏っている場合に大きな制約の効果を得られる。しかしながら歩行者の輪郭の遷移は多様であり、教師データの分布がばらつくので偏りが少ない。このような状態空間に対して

10

20

30

40

50

、主成分分析に基づいて状態遷移を適切に制約することは、原理的に無理がある。状態遷移を制約するのは、状態遷移に対する推定の精度を向上させるためである。従って、制約ができなければ推定の精度の向上が望めなく、誤動作のないトラッキングができないことになる。

【0031】

そこで、安定で誤動作のない、頑健（ロバスト）な歩行者トラッキングを実現するには、輪郭を特徴量として用いる方法に替えて、状態遷移の方向や大きさが、より安定となる特徴量を採用し、その特徴量に基づいて推定の精度を向上させることが必要である。

【0032】

また、モンテカルロフィルタを用いると、カルマンフィルタを用いる場合のガウス性などの仮定は必要なくなり、より頑健なトラッキングが実現可能となるので、このフレームワークを用いた種々のトラッキング手法が提案されている。しかしながら、歩行者のトラッキングに適合した実用的な手法は、いまだ提案されていない。例えば、上述のイザード（Isard）とマッコームック（MacCormick）の文献に示されるようなコンデンセーションを用いる歩行者トラッキングは、画像中における歩行者領域が大きい場合に適用され、かつ、3次元情報などを用いた精密な歩行者モデルに基づいてトラッキングするものであり、歩行者領域が小さい画像に対応する必要がある実用的な用途には不適である。

【0033】

また、監視用途の歩行者トラッキングにおいては、一般的に、ダイナミックレンジの広いグレイスケール画像が用いられ、さらに、夜間のトラッキングの場合、赤外線カメラの使用が考えられるので、これらの状況に対応して歩行者トラッキングを可能とするためには、色情報を利用しないトラッキング手法が望まれる。

【発明の開示】

【0034】

本発明は、上記課題を解消するものであって、簡単な構成により、色情報を用いることなく画像中の歩行者の動きを推定し、頑健な歩行者トラッキングを実現できる歩行者トラッキング方法及び歩行者トラッキング装置を提供することを目的とする。

【0035】

上記目的を達成するために、本発明は、時間連続的に撮影した画像中に現れる特定の歩行者領域を時系列的に与えられ、その歩行者領域の動きをモンテカルロフィルタを用いて推定して時間方向に関連づけることによりその歩行者領域を追跡する歩行者トラッキング方法であって、歩行者モデルを複数の節点とその節点間を連結する複数のスティックとを有するスケルトンモデルとして構成し、前記スケルトンモデルの状態と1対1の対応をなす点を表す多次元空間である状態空間に、その状態空間の点が示す状態の実現確率の属性をその点に付与して成る粒子を、定義すると共に、当該状態空間に初期状態として複数のスケルトンモデルの粒子を設定する工程と、第1の時刻における前記状態空間に設定された粒子の組に対し、モンテカルロフィルタの方法に基づいて当該第1の時刻より後の第2の時刻における状態を表す粒子の組を予測する工程と、前記第2の時刻において与えられた歩行者領域を2値画像化すると共にその2値画像の距離変換画像から成る観測データを生成する工程と、前記予測工程によって予測された各スケルトンモデルと前記距離変換画像とを比較することにより、第2の時刻における各スケルトンモデルの状態の尤度を算出する工程と、前記工程によって算出した各スケルトンモデルの尤度に基づいて、前記状態空間におけるスケルトンモデルの粒子の組を抽出すると共に、抽出した粒子の組を第2の時刻における前記状態空間の新たな粒子の組として設定する工程と、備え、前記粒子の組を予測する工程から前記抽出した粒子の組を新たな粒子の組として設定する工程までの工程を、時刻を進める毎に実行することにより、スケルトンモデルの状態空間における粒子の動きを推定して時間方向に関連づけることにより歩行者領域を追跡するものである。

【0036】

このような構成によれば、モンテカルロフィルタを用いると共に、歩行者のモデルとし

10

20

30

40

50

てスケルトンモデルを用いるので、歩行者の輪郭モデルを用いる場合に比べて、画像中の歩行者領域の形状の時間変化に追従して、頑健な歩行者トラッキングを実現できる。すなわち、スケルトンモデルは、歩行者の胴体や四肢をモデル化することができ、従って、輪郭を用いる場合とは異なり、人体の胴体や四肢の構造上の制約に基づいてスケルトンモデルの形状の変化を制約できるので、状態空間におけるスケルトンモデルの状態遷移を確実に合理的に制約できる。このような制約を設けることにより、トラッキングの現実性が高められ、頑健な歩行者トラッキングが可能となる。

【0037】

また、2値画像化した歩行者領域を変換した距離変換画像を尤度算出の比較対象として用いるので、尤度をより精度良く算出することができる。すなわち、歩行者領域の距離変換画像は、歩行者領域の外形に現れた歩行者の四肢の状態を反映したものとなるので、スケルトンモデルと比較する観測データとして好適であり、精度良く尤度を算出できる。また、2値化画像を用いるので、画像の色情報を用いる必要がなく、夜間の赤外線画像などを用いる監視用の歩行者トラッキングに適用する場合においても精度良くかつ頑健な歩行者トラッキングが可能である。

10

【0038】

本発明は、上述の改良された発明において、前記スケルトンモデルは、胴体スティックと、前記胴体スティックの上端に交差して接続した肩幅スティックと、前記胴体スティックの下端に交差して接続した腰幅スティックと、前記胴体スティックの延長部としての頭長スティックと、前記腰幅スティック及び肩幅スティックの両端部にそれぞれ接続された4本の四肢スティックと、の8本のスティックからなるものである。

20

【0039】

このような構成によれば、人体の特徴的な構造を適切に反映したスケルトンモデルを、簡単に構成できる。また、スケルトンモデルの状態を表すための変数の数として、多過ぎない適切な数を選ぶことができるので、状態変数の数を減らし、遷移可能な状態数を削減でき、従って、トラッキング精度を向上させることができる。

【0040】

本発明は、上述の改良された発明において、前記スケルトンモデルは、前記肩幅及び腰幅スティックの長さをそれぞれゼロとする状態と、前記肩幅、腰幅、及び両手スティックの長さをゼロとする状態と、前記肩幅及び両手スティックの長さをゼロとする状態と、を

30

【0041】

このような構成によれば、画像中に現れる歩行者の移動方向を反映して、スケルトンモデルを選択することができる。例えば、画像中を左方または右方に移動する歩行者は、側面を見せて歩行しており、肩幅や腰幅は殆ど認識されないため、肩幅スティックや腰幅スティックの長さをゼロとしたスケルトンモデルが好適に用いられる。

【0042】

本発明は、上述の改良された発明において、前記スケルトンモデルの1つの状態は、そのスケルトンを構成する各スティックの位置と角度とによって決定され、前記各スティックの相互の空間配置が人体の構造上可能な空間配置となるように制約されているものである。

40

【0043】

このような構成によれば、互いに連結された各スティックの相互の空間配置が、人体の構造上可能な空間配置となるように、各スティックの位置と角度によって制約されているので、歩行者の外見の変化を、不自然とならないように制約して、スケルトンモデルの状態遷移を制約できる。

【0044】

本発明は、上述の改良された発明において、前記尤度を算出する工程で用いる距離変換画像は、歩行者領域を表す黒の輪郭部から輪郭内部に行くほど高輝度と成るように白黒濃度で表現した画像であり、前記予測された各スケルトンモデルの尤度は、スケルトンモデ

50

ルと距離変換画像とを重ねた状態で、スケルトンモデルを構成するスティック上の複数の点から、各点における該スティックの法線方向に存在する距離変換画像の輝度極値点までの距離を用いて算出されるものである。

【0045】

このような構成によれば、スケルトンモデルの各スティックと、距離変換画像における高輝度部分とは、共に人体の骨格を反映したものであり、これらの重なり具合を輝度極値点までの距離によって見積もることができるので、推定されたスケルトンモデルの尤度を、精度良く算出することができる。例えば、このような輝度極値点までの距離の和が小さいほど重なり度合いが大きく、尤度が大きいとされる。

【0046】

本発明は、上述の改良された発明において、前記粒子の組を予測する工程と前記新たな粒子の組を設定する工程とにおいて乱数が用いられ、前記乱数が疑似乱数生成アルゴリズムを用いて生成されるものである。

【0047】

このような構成によれば、疑似乱数生成アルゴリズムを用いて生成した乱数は、長周期、高次元均等分布の特徴を有するので、モンテカルロフィルタに好適であり、高精度の歩行者トラッキングを実現できる。

【0048】

本発明は、上述の改良された発明において、前記予測工程によって予測された各スケルトンモデルにおける、胴体スティックと頭長スティックとの節点の座標の分散を算出し、当該分散が一定の値を超える場合には追跡を終了する工程、をさらに備えるものである。

【0049】

このような構成によれば、非歩行者を誤って歩行者領域として抽出してしまった場合であっても、節点の座標の分散から、追跡対象が歩行者でないことを判別し、追跡を終了することができる。

【0050】

また、本発明は、時間連続的に撮影された画像を時系列的に受信する画像受信手段と、前記画像受信手段によって受け取った画像から歩行者候補領域を抽出し、前記抽出した歩行者候補領域を歩行者領域と非歩行者領域とに分類すると共に、特定の歩行者領域を選択する歩行者領域選択手段と、前記歩行者領域選択手段によって選択された歩行者領域を時系列的に与えられ、その歩行者領域の動きを上述した本発明の歩行者トラッキング方法のいずれかを用いて追跡するトラッキング手段と、前記トラッキング手段によって追跡される歩行者領域を時系列的に表示する歩行者軌跡表示手段と、を備えることを特徴とする歩行者トラッキング装置である。

【0051】

このような構成によれば、一時的に歩行者が陰に隠れるオクルージョンの存在、歩行者の速度の急変、注目している歩行者に似た複数の移動物体の存在、などの状況においても、簡単な構成により、色情報を用いることなく画像中の歩行者の動きを推定し、高精度で頑健な歩行者トラッキングを実現して、人の動きを自動的に追跡し、動きの意図や動作を解析したり、侵入者を監視したりすることができる。

【図面の簡単な説明】

【0052】

【図1】本発明の一実施形態に係る歩行者トラッキング装置のブロック構成図。

【図2】同上装置における歩行者トラッキング処理のフローチャート。

【図3】本発明の一実施形態に係る歩行者トラッキング方法における処理のフローチャート。

【図4】同上歩行者トラッキング方法において用いられる状態空間の概念説明図。

【図5】同上歩行者トラッキング方法に画像と歩行者モデルの関係の説明図。

【図6】同上歩行者トラッキング方法におけるモンテカルロフィルタに基づく歩行者トラッキングの概念説明図。

10

20

30

40

50

【図7】(a1)(a2)(a3)は同上歩行者トラッキング方法が処理対象とする歩行者領域の画像例を示す図、(b1)(b2)(b3)は同歩行者領域画像から得られたモルフォロジカルスケルトンの図、(c1)(c2)(c3)は同モルフォロジカルスケルトンに人の四肢の概念を反映させた図。

【図8】(a)~(d)は同上歩行者トラッキング方法において用いられるスケルトンモデルの図。

【図9】同上歩行者トラッキング方法において用いられるスケルトンモデルのスティックの座標を定義する座標系の図。

【図10】同上歩行者トラッキング方法において用いられるスケルトンモデルの形状に対する制約の説明図。

10

【図11】(a)は同上歩行者トラッキング方法の処理対象の歩行者領域の図、(b)は同歩行者領域に距離変換を施して得られた距離変換画像を示す図。

【図12】(a)は同上歩行者トラッキング方法における距離変換画像と予測されたスケルトンモデルとを重ね合わせた図、(b)は同重ね合わせた図から尤度を求める方法を説明する図。

【図13】(a)(b)は同上歩行者トラッキング方法において尤度を求める方法の他の例を説明する図。

【図14】比較対照のコンデンセーションによる歩行者輪郭トラッキング方法において用いられる座標系の説明図。

【図15】トラッキング精度の評価のための評価基準の説明図。

20

【図16】(a)~(i)は歩行者トラッキング方法の評価に用いられる画像シーケンスの一部を示す図。

【図17】(a)(b)は歩行者トラッキング方法の評価に用いられるガードレール型静的誤検出評価のための画像シーケンスの一部を示す図。

【図18】(a)(b)(c)は歩行者トラッキング方法の評価に用いられるポール型静的誤検出評価のための画像シーケンスの一部を示す図。

【図19】(a)(b)は歩行者トラッキング方法の評価に用いられる動的誤検出評価のための画像シーケンスの一部を示す図。

【図20】(a)(b)(c)は歩行者トラッキング方法の評価に用いられる実画像シーケンスの一部を示す図、(d)は(a)から得られた2値画像を示す図。

30

【図21】(a)はシーケンスNに対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図22】(a)はシーケンスRに対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図23】(a)はシーケンスG1に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図24】(a)はシーケンスG2に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図25】(a)はシーケンスP1に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

40

【図26】(a)はシーケンスP2に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図27】(a)はシーケンスP3に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図28】(a)はシーケンスC1に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図29】(a)はシーケンスC2に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図30】(a)はシーケンスC3に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

50

【図31】(a)はシーケンスC4に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図32】(a)はシーケンスUに対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図33】(a)はシーケンスS1に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

【図34】歩行者トラッキングに適応的スケルトンモデルを用いた場合の重心誤差の評価結果を示す図。

【図35】(a)はシーケンスS2に対する重心誤差の評価結果を示す図、(b)は同スケール誤差の評価結果を示す図。

10

【図36】(a)は歩行者トラッキングにおいて想定される確率密度分布の例を示す図、(b)はカルマンフィルタに用いられるガウス型の確率密度分布図。

【図37】(a)は歩行者トラッキングにおいて想定される確率密度分布の例を示す図、(b)は同確率分布のモンテカルロフィルタにおける表現例を示す分布図。

【図38】モンテカルロフィルタにおける状態予測の概念説明図。

【図39】(a)(b)(c)は歩行者トラッキングの対象となる歩行者領域の画像を示す図。

【発明を実施するための最良の形態】

【0053】

以下、本発明の実施形態に係る歩行者トラッキング装置と歩行者トラッキング方法について、図面を参照して説明する。

20

【0054】

(歩行者トラッキング装置)

図1は本発明の一実施形態に係る歩行者トラッキング装置1のブロック構成を示し、図2は歩行者トラッキング装置1における処理の流れを示す。歩行者トラッキング装置1は、撮像装置10によって時間連続的に撮影された画像を時系列的に受信する画像受信手段2と、画像受信手段2によって受け取った画像から歩行者候補領域を抽出し、抽出した歩行者候補領域を歩行者領域と非歩行者領域とに分類すると共に、特定の歩行者領域を選択する歩行者領域選択手段3と、歩行者領域選択手段3によって選択された歩行者領域を時系列的に与えられ、その歩行者領域の動きを、歩行者をモデル化したスケルトンモデルと歩行者領域から得た距離変換画像とモンテカルロフィルタとを用いて推定して時間方向に関連づけることによりその歩行者領域を追跡するトラッキング手段4と、トラッキング手段4によって追跡される歩行者領域を時系列的に表示する歩行者軌跡表示手段5と、を備えている。

30

【0055】

上述の歩行者トラッキング装置1における処理は、画像受信手段2による1フレーム画像入力(S1)、歩行者領域選択手段3による歩行者候補領域の抽出(S2)、同じく歩行者領域選択手段3による候補領域の分類(S3)、そして、トラッキング手段4による歩行者状態予測(歩行者トラッキング)(S4)、の大きな4つの処理要素から成る。

【0056】

40

上述の撮像装置10は、例えば、イメージセンサなどを用いたテレビカメラなどが用いられる。画像受信手段2は、撮像装置10から入力された画像を受け取り、その画像に対して雑音除去や輪郭強調などの適切な前処理を施し、処理後の画像を歩行者領域選択手段3に入力する。

【0057】

歩行者領域選択手段3は、入力された画像に対し、背景差分法やオプティカルフローの検出、ステレオマッチングによる3次元距離計測、肌色検出、顔輪郭検出などの処理を施し、その画像から歩行者の全身像を示す画像部分と想定される歩行者候補領域を抽出する。抽出された歩行者候補領域には、通常、誤検出された領域も多く含まれる。そこで、歩行者領域選択手段3は、抽出した候補領域を歩行者領域と非歩行者領域とに分類する。こ

50

の分類は、例えば、ニューラルネットワークやサポートベクタマシン、ブースティングなどによる機械学習を用いると共に、ハール変換やガボール変換などにより得られる画像の特微量に基づく処理によって行われる。

【0058】

トラッキング手段4は、歩行者として分類された歩行者領域の情報に基づいて、歩行者領域を時間方向に関連付けるトラッキングを行う。つまり、上述の歩行者領域選択手段3による歩行者候補領域の抽出と分類は、一般に、各フレーム毎に、すなわち1枚の画像毎に、独立に処理されるので、この処理における歩行者領域は、時間方向には何等対応付けられておらず、また対応付けすることができない。そこで、トラッキング手段4によって歩行者領域が時間方向に関連付けられる。このトラッキング手段4による処理は、後述する歩行者トラッキング方法に基づいて行われる。

10

【0059】

トラッキング手段4による歩行者トラッキングの主な目的は、歩行者の動き情報を推定し、歩行者領域を時間方向に関連付けることである。歩行者トラッキングによると、歩行者が他の物体と重なってしまい、歩行者の画像情報が得られない、いわゆるオクルージョンなどの影響のため、1枚の画像からは歩行者を検出できない場合に、歩行者の動きを推定することができる。歩行者トラッキングで得られた情報は、例えば、カメラを搭載した自動車における歩行者との衝突検出処理や、多数の歩行者の存在する状況のもとで、特定の歩行者を同定して追跡する処理などに用いることができる。

20

【0060】

(歩行者トラッキング方法の概要)

図3は本発明の一実施形態に係る歩行者トラッキング方法における処理のフローチャートを示し、図4は歩行者トラッキング方法において用いられる状態空間Xを示し、図5は歩行者トラッキング方法における画像と歩行者モデルの関係を示す。歩行者トラッキング方法は、時間連続的に撮影した画像中に現れる特定の歩行者領域を時系列的に与えられ、その歩行者領域の動きをモンテカルロフィルタを用いて推定して時間方向に関連づけることによりその歩行者領域を追跡する方法である。この方法は、上述の歩行者トラッキング装置1におけるトラッキング手段4において実行される。

【0061】

歩行者トラッキング方法は、図3に示すように、初期設定の工程(S12)、状態遷移と予測の工程(S13)、観測データ生成工程(S15)、尤度算出工程(S16)、粒子再設定工程(S17)、などから構成される。

30

【0062】

まず、工程S11において時間変数tが初期化($t = 0$)され、その後、初期設定工程(S12)が実行される。

【0063】

初期設定工程(S12)は、歩行者を複数の節点とその節点間を連結する複数のスティックとを有するスケルトンモデルとして構成し、スケルトンモデルの状態と1対1の対応をなす点を表す多次元空間である状態空間Xに、その状態空間Xの点が示す状態の実現確率の属性をその点に付与して成る粒子を定義すると共に、状態空間Xに初期状態として複数のスケルトンモデルの粒子を設定する工程である。

40

【0064】

状態遷移と予測の工程(S13)は、第1の時刻における状態空間Xに設定された粒子の組(セット)に対し、モンテカルロフィルタの方法に基づいて第1の時刻より後の第2の時刻における状態を表す粒子の組を予測する工程である。この工程の後、工程S14において、時間変数tが1つ進められる。すなわち、 $t = t + 1$ (インクリメント)とされる。

【0065】

観測データ生成工程(S15)は、第2の時刻において与えられた歩行者領域を2値画像化すると共にその2値画像を距離変換して距離変換画像から成る観測データを生成する

50

工程である。

【0066】

尤度算出工程（S16）は、予測された各スケルトンモデルと、距離変換によって形成された距離変換画像（観測データ）とを比較することにより、第2の時刻における各スケルトンモデルの状態の尤度を算出する工程である。

【0067】

粒子再設定工程（S17）は、各スケルトンモデルの尤度に基づいて、状態空間Xにおけるスケルトンモデルの粒子の組を抽出すると共に、抽出した粒子の組を第2の時刻における状態空間Xの新たな粒子の組として設定する工程である。

【0068】

歩行者トラッキングは、上述の粒子の組を予測する工程（S13）から、抽出した粒子の組を新たな粒子の組として設定する工程（S17）までの、各工程を、時刻を進める毎に実行することにより実現される。すなわち、上記各工程により、スケルトンモデルの状態空間における粒子の動きが、推定されると共に時間方向に関連づけられ、これにより、歩行者領域が追跡され、トラッキングされる。

【0069】

（確率的トラッキング）

上述の工程S13、S15、S16は、モンテカルロフィルタを用いて歩行者を確率的にトラッキングする、いわゆるコンデンセーションフレームワークに基づいている。このフレームワーク（枠組み、手法）は、状態空間Xの適切な設定、すなわち適切な歩行者モデルの設定と、適切な状態遷移モデルの設定と、状態遷移を制約するための適切な設定と、を行うことにより、背景画像や歩行者自体の画像が乱雑となる状況における歩行者トラッキングに対して、効果的となる。すなわち、対象物体つまり歩行者のオクルージョンや歩行者の速度の急激な変化などに強い頑健なトラッキングが可能となる。

【0070】

モンテカルロフィルタを用いる歩行者トラッキングは、図4に示すように、状態空間Xにおける歩行者モデルを表す点が時間tと共に移動（状態が遷移）して形成する軌跡 t_r を、確率的に求めることに対応する。状態空間Xにおける各時刻tでの歩行者モデルを表す点は、確率分布に従って分布しており、その最も存在確率の高い点を時間的に結んだ軌跡 t_r が、トラッキング結果となる。そこで、図5に示すように、実空間における歩行者領域の画像Gには、状態空間Xにおける複数の歩行者モデル M_1, M_2, \dots, M_N を表す粒子が対応している（図37(a)(b)、図38、及び背景技術におけるこれらの図の説明参照）。

【0071】

この歩行者モデルの数N、つまり状態確率分布を表現する粒子数Nを増やすことによって、計算量は増加するが、状態確率分布の近似精度を向上させることができる。すなわち、アプリケーションに応じて適切な粒子数Nを選択することによって、精度などのシステム要求を、容易に満たすことができる。

【0072】

（スケルトンの利用）

歩行者モデルとして採用する歩行者のスケルトン（骨格）は、歩行者の輪郭に比べ、時間変化が安定しており、時間方向に安定な状態遷移を示す。そこで、スケルトンに基づく状態空間Xは、歩行者トラッキングにとって適切な設定となる。すなわち、スケルトンを用いることによって、現在の状態から予測される次の状態が真値に近い、という可能性が高くなり、高精度な歩行者トラッキングが可能となる。

【0073】

（状態遷移への制約）

トラッキングする対象を歩行者に限定する場合、歩行者をスケルトンによって表現し、そのスケルトンに対し、その状態遷移を制約することにより、対象に対する事前学習を行うことなく、高精度な歩行者トラッキングを実現できる。例えば、脚を表すスティックが

10

20

30

40

50

首を表すスティックと交差するような歩行者のスケルトンの発生確率は、極めて低いと考えられる。より真値に近い状態予測は、発生確率の低い状態を状態遷移の行き先から除くことによって実現される。つまり、節点とスティックとを単純に連結した場合に可能なスケルトンモデルの状態から発生確率の低い状態を、予め状態遷移の行き先から除くように状態遷移に制約をかければよい。このように、スケルトンを用いる場合、輪郭を用いる場合と比べて、状態遷移に対するより適切、かつ、より効果的な制約が可能である。

【0074】

(距離変換画像を用いた尤度の算出)

背景技術において説明したように、コンデンセーションによる輪郭トラッキングでは、輪郭のモデル曲線(B-スプライン曲線)と画像の輪郭エッジとに基づいて、状態空間におけるモデル曲線粒子の尤度を算出している。上述の工程S15, S16では、この考えを歩行者トラッキングに応用する(輪郭にスケルトンが対応し、輪郭エッジに距離変換画像が対応)。歩行者候補領域を白とした2値画像に対して距離変換を行い、これによって生成した画像をモンテカルロフィルタに入力する観測データ y_t として用いる。

10

【0075】

(乱数発生)

上述の状態遷移と予測の工程S13、及び再抽出して新たな粒子の組を設定する工程S17に用いる乱数の生成に、擬似乱数生成アルゴリズム(Mersenne Twister: 松本眞・西村拓士らが1997年に発表)を用いる。これによって生成した乱数は、長周期、高次元均等分布の特徴を有し、モンテカルロフィルタに適した乱数であり、高精度の歩行者トラッキングを実現できる。

20

【0076】

(歩行者トラッキングの処理の流れ)

モンテカルロフィルタを用いた歩行者トラッキングの処理の流れを、数学的表現を交えて、以下にまとめる。図6は歩行者トラッキング方法におけるモンテカルロフィルタに基づく歩行者トラッキングの概念を示し、背景技術で説明した図38に対応する。

【0077】

まず、与えられた初期状態ベクトル x_0 に雑音を加え、状態確率分布 $p(x_0)$ を生成する(図3のS12)。次に、時刻 t ($t=1, 2, \dots$)において、モンテカルロフィルタを用いて、状態確率分布 $p(x_{t-1})$ から $p(x_t)$ を推定する。推定は、予測、計測、及び再抽出の3ステップを繰り返すことにより実現される。

30

【0078】

(予測)スケルトンモデルを用いると共に状態遷移に制約を加えて状態を遷移させる。各粒子 $s_{t-1}^{(n)}$ ($n=1, \dots, N$)に対し、過程確率密度 $p(x_t | x_{t-1} = s_{t-1}^{(n)})$ に従って、次状態 $s_t^{(n)}$ を予測する(図3のS13)。

【0079】

(計測)距離変換画像と各粒子との比較により、予測された各粒子の尤度を算出する。すなわち、各粒子に対し、予測された状態の尤度 $l_t^{(n)}$ を、観測確率密度 $p(y_t | x_t)$ に従って算出する(図3のS15, S16)。ここで y_t は時刻 t における観測ベクトル(観測データ、つまり距離変換画像を表す観測空間のベクトル)である。

40

【0080】

(再抽出)乱数を生成し、生成した乱数を用いて尤度を決定することにより、上記によって予測した次状態 $s_t^{(n)}$ の中から、背景技術で述べた式(5)(6)に基づいて、予測状態を抽出する。これを粒子の個数 N 回繰り返すことにより、時刻 t における粒子のセット $\{s_t^{(n)}, n=1, \dots, N\}$ を抽出し、時刻 t における粒子の組として設定する(図3のS17)。これらの処理により、最終的に各時刻 t における状態確率分布 $p(x_t)$ ($t=1, 2, \dots$)が得られる。得られた各時刻の状態確率分布に基づいて、上述の図4に示すように、状態空間 X における軌跡 t_r が決定される。これがトラッキング結果となる。

【0081】

50

上述のように、予測した状態を表す粒子の尤度に基づいて、尤度の大きな粒子のところには粒子数を多くし、尤度の小さな粒子のところには粒子数を少なくしたり、ゼロにしたりして、状態空間におけるN個の粒子を再抽出すると共にこれらの粒子の再配分（再設定）を行い、観測データによる補正を反映した、各時刻 t ($t = 1, 2, \dots$) における状態確率分布 $p(x_t)$ が決定される。

【0082】

本発明の実施形態に係る歩行者トラッキング方法によれば、モンテカルロフィルタを用いると共に、歩行者のモデルとしてスケルトンモデルを用いるので、歩行者の輪郭モデルを用いる場合に比べて、画像中の歩行者領域の形状の時間変化に追従して、頑健な歩行者トラッキングを実現できる。すなわち、スケルトンモデルは、歩行者の胴体や四肢をモデル化することができ、従って、輪郭を用いる場合とは異なり、人体の胴体や四肢の構造上の制約に基づいてスケルトンモデルの形状の変化を制約できるので、状態空間におけるスケルトンモデルの状態遷移を確実に合理的に制約できる。このような制約を設けることにより、トラッキングの確実性が高められ、頑健な歩行者トラッキングが可能となる。

10

【0083】

また、2値画像化した歩行者領域を変換した距離変換画像を尤度算出の比較対象として用いるので、尤度をより精度良く算出することができる。すなわち、歩行者領域の距離変換画像は、歩行者領域の外形に現れた歩行者の四肢の状態を反映したものとなるので、スケルトンモデルと比較する観測データとして好適であり、精度良く尤度を算出できる。また、この歩行者トラッキング方法は、2値化画像を用いるので、画像の色情報を用いる必要がなく、夜間の赤外線画像などを用いる監視用の歩行者トラッキングに適用して精度良くかつ頑健な歩行者トラッキングが可能である。

20

【0084】

以下、歩行者トラッキング方法における個々の要素技術と処理、すなわち、スケルトンモデルと状態空間 X 、状態遷移モデル、尤度 $t^{(n)}$ の算出法について詳述する。

【0085】

（歩行者のスケルトンモデルの有効性）

図7(a1)～(a3)は上歩行者トラッキング方法が処理対象とする歩行者領域の画像例を示し、図7(b1)～(b3)は前記画像から得られたモルフォロジカルスケルトンを示し、図7(c1)～(c3)は前記モルフォロジカルスケルトンに人の四肢の概念を反映させた状態を示す。ここでは、歩行者領域の画像に画像処理を施して得られる、いわゆるモルフォロジカル画像としてのモルフォロジカルスケルトンの例を示し、スティックによる歩行者のモデル化の有効性を説明する。

30

【0086】

歩行者のモルフォロジカルスケルトンは、次の処理によって得られる。 I_b を離散2値画像（図7(a1)～(a3)）とし、 P を原点 $(0, 0)$ を含む2値パターンとする。今、 P を固定して考える。 $M = \max\{i \geq 0; I_b(i, P) = 1\}$ とするとき、 $S_i = (I_b(i, P) - (I_b(i, P) \ominus P))$ で定められる各 S_i をスケルトン要素と呼ぶ。ここで、 $A \ominus B$ は、 A の B による収縮演算（erosion）であり、 $A \oplus B$ は、 A の B による切断演算（opening）を表し、 $(A \oplus B) \ominus B$ は、次式、 $(A \oplus B) \ominus B = (\dots \ominus B)$ で定義される k 回の収縮演算を表す。

40

【0087】

モルフォロジカルスケルトン $SK(I_b)$ は、次式(8)のようにすべての S_i の合併として定義される。これにより、図7(b1)～(b3)に示すような、モルフォロジカルスケルトン $SK(I_b)$ が、2値画像から一意に得られる。

【数5】

$$SK(I_b) = \bigcup_{i=0}^M S_i \quad (8)$$

【0088】

上述のモルフォロジカルスケルトン $SK(I_b)$ に対し、歩行者のスケルトンは「首の位置で4つに分岐」し、「腰の位置で3つに分岐」する、という知識を適用すると、図7(c1)～(c3)に示すように、歩行者の骨格の状態がより明確になる。逆に、人の骨格を考慮すると、画像間の歩行者の状態の変化を安定に予測することができる。

【0089】

上記により、スケルトンモデルの状態遷移は、歩行者の輪郭の状態遷移と比べて、予測がより容易であることが分かる。すなわち、スケルトンの状態遷移は、輪郭の状態遷移に比べて時間方向に安定であり、従って、予測される次の状態が、より真値に近い状態となる。そこで、歩行者のスケルトンモデルを表すパラメータから成る空間を状態空間 X とすると、より頑健な歩行者トラッキングを実現可能となる。

【0090】

(スケルトンモデルの構築)

図8(a)～(d)は歩行者トラッキング方法において用いられるスケルトンモデルを示す。歩行者の骨格を精密にモデル化すればするほど、細かな要素を粒子の尤度算出に考慮することができるので、尤度の算出精度の向上が期待できる。しかしながら、過剰な精密化によって、状態表現のためのパラメータ数が増加し、ひいては状態遷移における自由度の増加を招くので、真値に近い状態に遷移する粒子の割合が減少することになる。つまり、尤度の算出精度と、状態遷移の精度とはトレードオフの関係にある。

【0091】

そこで、本実施形態に係る歩行者トラッキング方法において、図8(c)に示すように、スケルトンモデルを、胴体スティックと、前記胴体スティックの上端に交差して接続した肩幅スティックと、前記胴体スティックの下端に交差して接続した腰幅スティックと、前記胴体スティックの延長部としての頭長スティックと、前記腰幅スティック及び肩幅スティックの両端部にそれぞれ接続された4本の四肢スティックと、の8本のスティックによって構築する。

【0092】

上述の図8(c)に示すような8本のスティックを用いた歩行者のスケルトンモデルによれば、人体の特徴的な構造を適切に反映したスケルトンモデルを、簡単に構成できる。また、スケルトンモデルの状態を表すための変数の数として、過剰でない適切な数を選ぶことができるので、状態変数の数を減らし、遷移可能な状態数を削減でき、従って、トラッキング精度を向上させることができる。

【0093】

また、スケルトンモデルの状態をさらに制限して、図8(a)に示すように、肩幅及び腰幅スティックの長さをそれぞれゼロとする状態と、図8(b)に示すように、肩幅、腰幅、及び両手スティックの長さをゼロとする状態と、図8(d)に示すように、肩幅及び両手スティックの長さをゼロとする状態と、を取り得るものとする。

【0094】

上述の制限を加えたスケルトンモデルは、状態遷移における自由度の増加を抑制するものであり、かつ、画像中に現れる歩行者の移動方向を反映するものである。例えば、画像中を左方または右方に移動する歩行者は、体側を見せて歩行しており、肩幅や腰幅は殆ど認識されないため、肩幅スティックや腰幅スティックの長さをゼロとしたスケルトンモデルが好適に用いられる。なお、スケルトンモデルを構成する複数の節点とその節点間を連結する複数のスティックについて、以下では、節点 p と節点 q を結ぶスティック $ST_{p,q}$

と表すことにする。

【0095】

上述のように、歩行者をスティックで表現する場合、例えば、図8(a)に示すように、四肢、胴体、及び頭部を表す6本のスティックによって構成したスケルトンモデルが考えられ、これをスケルトンモデル(a)とする。実際に撮影された画像から歩行者領域を検出すると、解像度が低い場合や、手が胴体に重なる場合などに、手の領域が検出されないことがある。このような場合、スケルトンモデル(a)は、尤度算出の際に ST_{bd} 、 ST_{bf} に由来する要素を適切に計算できなくなり、尤度算出精度が低くなる。

【0096】

そこで、図8(b)のように ST_{bd} 、 ST_{bf} をスケルトンモデルから削除し4本のスティックにより構成したスケルトンモデル(b)を設定する。スケルトンモデル(a)に比べて、自由度を削減でき、また、手が検出されないような状況において、尤度の算出精度が向上する。このように、手(腕)の領域が検出されるかどうかなどの状況に応じて、いずれかのスケルトンモデルを選択して用いることができる。また、歩行者が画像中で側面を見せている場合、すなわち右又は左方向に歩いている場合は、肩幅がほとんど認識されないので、スケルトンモデル(a)(b)で計算精度が十分得られる。

【0097】

ところで、歩行者が画像中で正面又は背面を見せている場合、すなわち上又は下方向、又は正面方向に歩いている場合、肩幅と腰幅とが認識されるようになる。つまり腕のスティックの節点b、及び脚のスティックの節点gが、それぞれ胴体の横幅の分だけ離れることになり、スケルトンモデル(a)(b)を用いて算出した尤度は、大きな誤差を含むことになる。そこで、スケルトンモデル(a)(b)にそれぞれ、肩幅のスティック ST_c 、腰幅のスティック ST_{hj} を追加し、スケルトン(c)(d)とする。このようなモデルを用いることにより計算精度を向上できる。

【0098】

上述のように、スケルトンモデル(a)~(d)のうち、最適なモデルは、歩行者トラッキングの対象となる画像シーケンス(以下シーケンスと略す)の特性に依存する。そこで、画面中における歩行者の向きや動きの方向に応じて、これらのスケルトンモデル(a)~(d)の中からモデルを適宜選択して、高精度な歩行者トラッキングを実現することができる。このようなモデルを、適応的スケルトンモデルと呼ぶこととし、これは、実施例において説明する。

【0099】

(状態空間Xの設定)

次に、上述のスケルトンモデルの状態を記述する状態空間Xを説明する。図9は歩行者トラッキング方法において用いられるスケルトンモデルのスティックの座標を定義する座標系を示す。

【0100】

状態空間Xは、歩行者に関する特徴点によって構成されるベクトルの指し示す多次元空間であり、歩行者をスケルトンによってモデル化したときのスケルトンモデルを表す変数(パラメータ)に基づいて定義される。状態空間Xは、精度の高い状態遷移の予測を行うために、多過ぎず、また少な過ぎない適切な自由度を持つモデルに基づくのが望ましい。そこで、上述のように、4本~8本のスティックによって構成した4パターンのスケルトンモデル(a)~(d)について状態空間Xを設定する。

【0101】

スケルトンモデルの形状を座標表示するため、歩行者トラッキングの対象となる時系列画像に固定した画像座標系を絶対座標系として用いる。スケルトンモデルを構成する各スティックを表すパラメータ(変数)の値の1組によって、歩行者の1つの状態が、状態空間Xにおける1点を占めることになる。

【0102】

そこで、各スケルトンモデル(a)~(d)をパラメータによって表現して、歩行者の

10

20

30

40

50

状態を状態空間 X における点に対応付ける。画像座標系（絶対座標系）である $x y$ 座標系を、図 9 に示すように設定し、この $x y$ 座標系における節点 p ($p = a, b, \dots$) の座標を (x_p, y_p) とする。スケルトンモデルの状態空間 X における表現方法としていくつか考えられる。例えば、次式 (9) に示す状態空間 X_{abs} のように、スケルトンの各節点の絶対座標を、そのまま用いて状態空間 X を定義することができる。ここで、大文字の X は状態空間を示し、3 本線の等号はこの式が定義式であることを示し、括弧 $\{ \}$ は括弧内の変数（パラメータ）を成分とするベクトルを示し、括弧 $\{ \}$ の右肩の T は、ベクトルや行列における転置を示す。

【数 6】

$$X_{abs} \equiv \{x_a, y_a, x_b, y_b, \dots\}^T \quad (9)$$

10

【0103】

上式 (9) の表現による状態空間は、しかしながら、頭と胴体が別々の方向へ移動する状態などのように、歩行者のスケルトンとして実現不可能な状態を簡便に排除するためには不便である。そこで、絶対座標を保持する節点を 1 点だけ定め、この点を基準となる点（基点）とし、他の節点をこの基点からの相対座標により表現する。相対座標は、次式 (10) ~ (12) に示すように、基点を原点とした直交画像座標系によって表される。ここで、 (x_{base}, y_{base}) は基点の絶対座標であり、 x'_p, y'_p はそれぞれ式 (11) (12) で与えられる節点 p の相対座標である。このような座標系で表される空間を、状態空間 X_{rel} とする。

20

【数 7】

$$X_{rel} \equiv \{x_{base}, y_{base}, x'_a, y'_a, x'_b, y'_b, \dots\}^T \quad (10)$$

$$x'_p = x_p - x_{base} \quad (11)$$

$$y'_p = y_p - y_{base} \quad (12)$$

30

【0104】

この状態空間 X_{rel} を用いることにより、状態遷移の際に、絶対座標の遷移を制御する雑音の分散と、相対座標の遷移を制御する雑音の分散と、をそれぞれ個別に調節することが容易となり、前述した頭と胴体が別々の方向へ移動するような可能性を、容易に制限することができる。

【0105】

上式 (10) の状態空間 X_{rel} をさらに改良したものとして、各スティックの長さや角度に基づく相対座標系を用いた状態空間 $X_{p.o.l}$ を定義する。これは、例えば、腰の節点 g の位置を固定したときに、脚のつま先の位置は、腰を中心とし脚の長さを半径とする円の周上にある可能性が高いという原理に基づく。この状態空間 $X_{p.o.l}$ は、状態遷移に対する制約を容易に表現することができる。状態空間 $X_{p.o.l}$ は、式 (13) のように表される。ここで、長さ l_{pq} 、角度 a_{pq} は、図 9 に示すように、それぞれ式 (14) (15) により定義する。

40

【数 8】

$$X_{pol} \equiv \{x_{base}, y_{base}, l_{ab}, l_{bg}, \dots, a_{ab}, a_{bg}, \dots\}^T \quad (13)$$

ただし、

$$l_{bd} = l_{bf} \quad [\text{スケルトンモデル(a)}]$$

$$l_{gi} = l_{gk} \quad [\text{スケルトンモデル(a)(b)}]$$

$$l_{cd} = l_{ef} \quad [\text{スケルトンモデル(c)}]$$

$$l_{hi} = l_{jk} \quad [\text{スケルトンモデル(c)(d)}]$$

$$l_{ce} = l_{hj} \quad [\text{スケルトンモデル(c)}]$$

$$a_{ce} = a_{hj} = 0 \quad [\text{スケルトンモデル(c)(d)}]$$

$$l_{pq} = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2} \quad (14)$$

$$a_{pq} = \tan^{-1} \frac{y_p - y_q}{x_p - x_q} \quad (15)$$

【0106】

上記において、3つの表現による状態空間 X_{abs} 、 X_{rel} 、 X_{pol} を説明したが、以下では、首の節点 b を基点とした状態空間 X_{pol} を用いて説明する。状態空間 X_{pol} を用いるとの前提のもとで、4つのスケルトンモデル (a) ~ (d) のそれぞれに対応する状態空間 $X_{(i)}$ 、($i = a, b, c, d$) は、次の式 (16) (17) (18) (19) のようになる。各状態空間の次元数は、それぞれ、12次元、9次元、14次元、10次元となる。この次元数は、状態遷移における自由度に相当する。

【数 9】

$$X_{(a)} \equiv \{x_b, y_b, l_{ab}, l_{bd}, l_{bg}, l_{gi}, a_{ba}, a_{bd}, a_{bf}, a_{bg}, a_{gi}, a_{gk}\}^T \quad (16)$$

$$X_{(b)} \equiv \{x_b, y_b, l_{ab}, l_{bg}, l_{gi}, a_{ba}, a_{bg}, a_{gi}, a_{gk}\}^T \quad (17)$$

$$X_{(c)} \equiv \{x_b, y_b, l_{ab}, l_{cd}, l_{hi}, l_{hj}, l_{bg}, l_{ce}, a_{ba}, a_{cd}, a_{ef}, a_{bg}, a_{hi}, a_{jk}\}^T \quad (18)$$

$$X_{(d)} \equiv \{x_b, y_b, l_{ab}, l_{hi}, l_{bg}, l_{hj}, a_{ba}, a_{bg}, a_{hi}, a_{jk}\}^T \quad (19)$$

【0107】

これらの状態空間 $X_{(i)}$ は、後述する尤度算出のため、絶対座標空間 X_{abs} に一意に変換可能であることが要求される。これは、例えば、スケルトンモデル (a) の場合、次式 (20) により達成されることが分かる。他のスケルトンモデル (b) (c) (d) についても、同様に座標変換可能である。

10

20

30

40

【数10】

$$X_{abs} = \begin{bmatrix} x_a \\ y_a \\ x_b \\ y_b \\ x_d \\ y_d \\ x_f \\ y_f \\ x_g \\ y_g \\ x_i \\ y_i \\ x_k \\ y_k \end{bmatrix} = \begin{bmatrix} x_b + l_{ab} * \cos(a_{ba}) \\ y_b + l_{ab} * \sin(a_{ba}) \\ x_b \\ y_b \\ x_b + l_{bd} * \cos(a_{bd}) \\ y_b + l_{bd} * \sin(a_{bd}) \\ x_b + l_{bd} * \cos(a_{bf}) \\ y_b + l_{bd} * \sin(a_{bf}) \\ x_b + l_{bg} * \cos(a_{bg}) \\ y_b + l_{bg} * \sin(a_{bg}) \\ x_b + l_{bg} * \cos(a_{bg}) + l_{gi} * \cos(a_{gi}) \\ x_b + l_{bg} * \sin(a_{bg}) + l_{gi} * \sin(a_{gi}) \\ x_b + l_{bg} * \cos(a_{bg}) + l_{gi} * \cos(a_{gk}) \\ x_b + l_{bg} * \sin(a_{bg}) + l_{gi} * \sin(a_{gk}) \end{bmatrix} = U(X_{(a)}) \quad (20)$$

10

20

【0108】

(状態遷移)

次に、上述のように定義された状態空間Xにおける、スケルトンモデル(a)~(d)の状態遷移のモデルについて説明する。状態遷移モデルは、歩行者の身体、つまり人体の動きとして通常あり得ない状態に遷移しないように、適切に制約を加えたモデルが望ましい。これは上述の座標系と状態空間とを用い、自己回帰モデルにおけるガウス雑音の分散を調節することによって実現する。

30

【0109】

さらに、スケルトンモデルを構成する各スティックの位置と角度とによって決定されるスケルトンモデルの状態は、各スティックの相互の空間配置が人体の構造上可能な空間配置となるように制約されているものとする。すなわち、各スティックの角度に対してスティックの可動範囲に制限を加える。また、複数のスティック相互の長さの比が、前時刻と現時刻とにおいて一定に保たれるという制限を加える。このように、取り得る状態に制限(制約)を加えるということは、スケルトンモデルを表す点が、状態空間Xにおいてある領域に局在することを意味する。また、状態遷移に制約を加えるということは、状態空間Xにおいて、互いかけ離れた点の間では、状態が遷移しないということの意味する。

40

【0110】

上述のような分散調節や状態遷移への制約を課した状態遷移モデルによると、歩行者の外見の変化を、不自然とならないように反映して、スケルトンモデルの状態遷移を制約でき、また、歩行者としては通常考えられないような状態のスケルトンに遷移する可能性を除去し、状態空間Xにおいてスケルトンの状態が存在する可能性の高い方向に、粒子を分散させることができる。

50

【 0 1 1 1 】

(基点の絶対座標の状態遷移)

上記内容を反映した状態遷移モデルの式を、各変数毎に説明する。まず、基点の絶対座標 x_b 、 y_b の状態遷移については、速度を考慮することができる 2 次の自己回帰モデル (ARモデル、上述の北川源四朗著「時系列解析入門」参照) を適用する。ここで、状態遷移後の各変数の状態の予測値を u_t とし、1 期前、2 期前 (1 期前とは、1 ステップ前のこと、以下同様) の状態をそれぞれ u_{t-1} 、 u_{t-2} とし、平均 0 のガウス雑音を w_c とすると、この状態遷移モデルは次式 (21) のように表される。

【数 1 1】

$$u_t = 2u_{t-1} - u_{t-2} + w_c \quad (21)$$

10

【 0 1 1 2 】

(スティックの長さの状態遷移)

スティックの長さ $l_{p,q}$ の状態遷移は、2 段階の処理を経て行う。まず、次式 (22) により各粒子に対し、確率的にズームパラメータ z を決定する。その後、式 (23) に示すように、 z を 1 期前の状態 u_{t-1} に乗じたのち、1 次の自己回帰モデルによって状態を遷移させる。

【数 1 2】

$$z = \gamma^{w_z} \quad (22)$$

20

$$u_t = zu_{t-1} + w_l \quad (23)$$

【 0 1 1 3 】

ここで w_z は平均 0 のガウス雑音であり、 γ は正の定数であり、 w_l は平均 0 のガウス雑音である。長さ決定に対して、ズームパラメータ z による寄与よりも、1 次自己回帰モデルによる寄与を小さくすることにより、前時刻における長さの比を維持しつつ、歩行者が接近したり遠ざかったりすることに対応するスケルトンモデルの拡大縮小に、適切に対応させることができる。このような調節は、 w_z と w_l の分散を調整することにより実現される。

30

【 0 1 1 4 】

(スティックの角度の状態遷移)

図 10 は歩行者トラッキング方法において用いられるスケルトンモデルの形状に対する制約の具体例を示す。すなわち、首のスティックの可動範囲を、上方の 90° の範囲内とし、左右の脚の可動範囲を水平位置までとしている。このような可動範囲を考慮して、スティックの角度 $a_{p,q}$ の状態遷移については、各スケルトンモデルに対し、以下に示す制限を加える。

【数 1 3】

$$-3\pi/4 < a_{ba} < -\pi/4$$

[スケルトンモデル(a) (b) (c) (d)]

$$\pi/4 < a_{bg} < 3\pi/4$$

[スケルトンモデル(a) (b) (c) (d)]

$$0 < a_{bd}, a_{bf} < \pi$$

[スケルトンモデル(a)]

$$0 < a_{gi}, a_{gk} < \pi$$

[スケルトンモデル(a) (b)]

$$0 < a_{cd}, a_{ef} < \pi$$

[スケルトンモデル(c)]

10

$$0 < a_{hi}, a_{jk} < \pi$$

[スケルトンモデル(c) (d)]

$$(a_{bd} - a_{bg})(a_{bf} - a_{bg}) < 0$$

[スケルトンモデル(a)]

$$(a_{gi} - a_{bg})(a_{gk} - a_{bg}) < 0$$

[スケルトンモデル(a) (b)]

$$(a_{cd} - a_{bg})(a_{ef} - a_{bg}) < 0$$

[スケルトンモデル(c)]

$$(a_{hi} - a_{bg})(a_{jk} - a_{bg}) < 0$$

[スケルトンモデル(c) (d)]

20

【0 1 1 5】

これらの制限内で、平均を考慮した2次自己回帰モデルを適用する。これは次式(24)のように表される。

【数 1 4】

$$u_t = \alpha \times (2u_{t-1} - u_{t-2}) + (1 - \alpha) \times \sigma + w_a \quad (24)$$

【0 1 1 6】

ここで α は平均を示し、角度 a_{ba} の遷移に対しては $\pi/2$ 、その他の角度 a_{pq} の遷移に対しては $\pi/4$ である。 σ は $0 < \sigma < 1$ を満たす定数であり、平均 σ の与える影響を制御する。特に、 $\sigma = 1$ の場合、平均は考慮されない。 w_a は平均0のガウス雑音であり、上述した制限を越えないように決定される。

30

【0 1 1 7】

以上の状態遷移モデルによると、例えば、スケルトンモデル(a)の場合、状態遷移は次式(25)のように表される。ここで、 F_1, F_2, C, W は、それぞれ式(26)~(30)で与えられる。

【数 1 5】

$$x_t = F_1 \cdot x_{t-1} + F_2 \cdot x_{t-2} + C\sigma + W \quad (25)$$

$$F_1 = \begin{bmatrix} 2I_2 & O_{2 \times 4} & O_{2 \times 6} \\ O_{4 \times 2} & \gamma^{w_z} I_4 & O_{4 \times 6} \\ O_{6 \times 2} & O_{6 \times 4} & 2\alpha I_6 \end{bmatrix} \quad (26)$$

$$F_2 = \begin{bmatrix} -I_2 & O_{2 \times 4} & O_{2 \times 6} \\ O_{4 \times 2} & O_{4 \times 4} & O_{4 \times 6} \\ O_{6 \times 2} & O_{6 \times 4} & -\alpha I_6 \end{bmatrix} \quad (27)$$

$$C = \begin{bmatrix} O_{6 \times 6} & O_{6 \times 6} \\ O_{6 \times 6} & (1-\alpha)I_6 \end{bmatrix} \quad (28)$$

$$\sigma = [O_{1 \times 6} \quad -\pi/2 \quad \pi/2 \quad \pi/2 \quad \pi/2 \quad \pi/2 \quad \pi/2 \quad \pi/2]^T \quad (29)$$

$$W = [w_c \quad w_c \quad w_l \quad w_l \quad w_l \quad w_l \quad w_a \quad w_a \quad w_a \quad w_a \quad w_a \quad w_a]^T \quad (30)$$

【0 1 1 8】

ここで、 I_N は $N \times N$ の単位行列であり、 $O_{N \times M}$ は $N \times M$ の零行列である。状態遷移モデルの性能を決定する重要なパラメータは、加えるガウス雑音 w_c, w_z, w_l, w_a の分散である。これらは処理対象の乱雑さに応じて決定する必要がある。ある節点の絶対座標 u_t の遷移が、各ガウス雑音によって次式 (31) によって与えられるとする。また、各雑音が u_t に与える、次式 (32) のような各影響 を考える。

【数 1 6】

$$u_t = f(u_{t-1}, w_c, w_z, w_l, w_a) \quad (31)$$

$$\begin{aligned} \rho_c &= |f(u_{t-1}, w_c, w_z, w_l, w_a) - f(u_{t-1}, 0, w_z, w_l, w_a)| \\ \rho_z &= |f(u_{t-1}, w_c, w_z, w_l, w_a) - f(u_{t-1}, w_c, 0, w_l, w_a)| \\ \rho_l &= |f(u_{t-1}, w_c, w_z, w_l, w_a) - f(u_{t-1}, w_c, w_z, 0, w_a)| \\ \rho_a &= |f(u_{t-1}, w_c, w_z, w_l, w_a) - f(u_{t-1}, w_c, w_z, w_l, 0)| \end{aligned} \quad (32)$$

【0 1 1 9】

以下において、上述の影響 の大小関係の決定について説明する。まず、各節点の相対的な位置関係は、 z, l, a によって決まるので、 $c > z, l, a$ とすることによって、前時刻 ($t-1$) における節点の位置関係を保持することができる。また、スティックの長さが急に变化する確率は低いと考えられるので、 $a > z, l$ とする。ただし、 w_a については、すべてのスティックの角度 $a_{p,q}$ に同じ分散を用いる必要はない。例えば、胴体の角度 $a_{b,g}$ は手の角度 $a_{b,d}$ に比べて大きく遷移する可能性は少ない。そのため、 $a_{b,g}$ の遷移に用いるガウス雑音の分散を、 $a_{b,d}$ の遷移に用いるガウス雑音の分散よりも小さく設定する。 z, l に関しては、前述したように、前時刻におけるスティック間の長さの比を保つため、 $z > l$ とする。

【0 1 2 0】

10

20

30

40

50

上記をまとめると、 $w_c > w_a > w_z > w_1$ となる。この関係式を満たすように、 w_c 、 w_z 、 w_1 、 w_a の分散を調整することによって、状態遷移された各粒子が、歩行者のスケルトンモデルとして実現可能性の高い方向に分散し、逆に、実現性の低い状態には遷移しにくいように、状態遷移に制約をかけることができる。

【0121】

(粒子の尤度の算出)

尤度 $t^{(n)}$ の算出手法には、コンデンセーションによる輪郭トラッキングで用いられている手法を拡張した、距離変換画像における極大値の探索に基づく手法を用いる。以下、これらを詳細に述べる。すなわち、尤度を算出する工程で用いる距離変換画像は、歩行者領域を表す黒の輪郭部から輪郭内部に行くほど高輝度と成るように白黒濃度で表現した画像とし、予測された各スケルトンモデルの尤度は、スケルトンモデルと距離変換画像とを重ねた状態で、スケルトンモデルを構成するスティック上の複数の点から、各点における該スティックの法線方向に存在する距離変換画像の輝度極値点、例えば、輝度極大点までの距離を用いて算出する。

10

【0122】

上述のような尤度の算出方法によれば、スケルトンモデルの各スティックと、距離変換画像における高輝度部分とは、共に人体の骨格を反映したものであり、これらの重なり具合を輝度極大点までの距離によって見積もることができるので、推定されたスケルトンモデルの尤度を、精度良く算出することができる。例えば、このような輝度極大点までの距離の和が小さいほど重なり度合いが大きく、尤度が大きいとされる。

20

【0123】

上述のように、コンデンセーションによる輪郭トラッキングで用いられている尤度算出の考えを応用し、粒子の尤度を算出する。コンデンセーションによる輪郭トラッキングでは、予測されたB-スプライン曲線上の数点について、入力画像におけるエッジを探索し、予測された輪郭とエッジとの距離によって粒子の尤度を評価している。そこで本実施形態に係る歩行者トラッキング方法では、予測されたスケルトンモデルと観測データであるモルフォロジカルスケルトン(画像から画像処理によって得られたスケルトンモデル)との比較によって、尤度を算出する。

【0124】

すなわち、スケルトンモデルを構成するスティック上の数点に対し、スティックの法線方向に最も近いモルフォロジカルスケルトン上の点を探索し、予測されたスケルトンモデルとの距離に基づいて、そのスケルトンモデルに対応する粒子の尤度を算出する。なお、探索の始点とするスケルトンモデル上の点を、探索基点と呼ぶことにする。

30

【0125】

(距離変換画像)

図11(a)はトラッキング対象の歩行者領域画像を示し、図11(b)は前記画像に距離変換を施して得られた距離変換画像の例を示す。モルフォロジカルスケルトンは、2値画像を距離変換した距離変換画像として得られる。距離変換画像は、2値画像において、注目画素から最も近い画素値0の画素までの距離を、その注目画素の画素値とする変換によって生成される。注目画素の画素値が0の場合の距離は0とされる。ここで、2値画像において画素値が0である領域をBとすると、距離変換画像における座標(x, y)を有する画素の画素値D(x, y)は、次式(33)で表される。

40

【数17】

$$D(x, y) = \begin{cases} 0 & (x, y) \in B \\ \min_{(x_i, y_i) \in B} \sqrt{(x - x_i)^2 + (y - y_i)^2} & (x, y) \notin B \end{cases} \quad (33)$$

【0126】

(単純な尤度算出)

50

図12(a)は距離変換画像に予測されたスケルトンモデルを重ねた状態を示し、図12(b)は前記の重ね合わせた図から尤度を求める方法を示す。本実施形態では、距離変換前の画像として、歩行者候補領域の画素値を1とした2値画像を用いる。1を白、0を黒としている。この2値画像は、トラッキングの前段階として得られている。

【0127】

予測ステップにおいて、図12(a)に示すようなスケルトンが予測されたとする。距離変換画像におけるモルフォロジカルスケルトンの探索は、予測されたスケルトンモデル上の探索基点から法線方向に最も近い距離変換画像における輝度の極値で、例えば、極大点を求めることで達成される。図12(a)中の脚のスケルトンにおける探索例を、図12(b)に示す。

10

【0128】

スケルトンモデル上のi番目の探索基点から探索された極大点までの距離を $\delta_i^{(n)}$ とする。粒子nの極単純な尤度算出方法では、次式(34)により与えられる $e^{(n)}$ を算出する。ここで、rはスケールングのための定数であり、 μ は探索範囲を表す定数である。式(34)はコンデンセーションによる輪郭トラッキングにおける尤度の算出式(7)に倣っている。このようにして算出される $e^{(n)}$ を、式(35)のように正規化することにより、粒子nの尤度 $\pi^{(n)}$ を算出する。

【数18】

$$e^{(n)} = \exp \left\{ -\frac{1}{r} \sum_i \left(\min(\delta_i^{(n)}, \mu) \right)^2 \right\} \tag{34}$$

20

$$\pi^{(n)} = \frac{e^{(n)}}{\sum_{k=1}^N e^{(k)}} \tag{35}$$

【0129】

(より高精度化した尤度算出)

30

図13(a)(b)は歩行者トラッキング方法において尤度を求める方法の他の例を示す。上述の単純な尤度算出方法は、次の2点が問題点として挙げられる。(1)手のスティックと胴体のスティックが重なる場合に尤度が高くなる。(2)スティックが短いほど、尤度が高くなる。以下、この2点の改良について説明する。

【0130】

(1.手と胴体のスティックの重なり対策)

手のスティック ST_{bd} などの上の点を探索基点として、距離変換画像の極大点を探索する場合、胴体に由来する極大点と手に由来する極大点の区別がつかない。そのため、予測された手のスティックが、胴体のスティック ST_{bg} に重なるようなスケルトンモデルが予測された場合、両者のスティック上の点を探索基点とした探索により、共に胴体に由来する極大点が検出され、高い尤度を得るという問題がある。

40

【0131】

そこで、手のスティック上の探索基点では、図13(a)に示すように、中心角が $\theta/2$ の円周の上方から下方に向けて極大点を探索する。最初に検出された極大点における角度から求められる探索線上の探索基点までの長さ L_j に基づいて、尤度を算出する。つまり、上述の問題を解決するため、水平方向に近い極大点を優先することとする。これは、垂直方向に近い角度において検出される極大点は、胴体に由来するものである可能性が高い、という事実に基づく。このような探索によって、第j番目の探索基点から得られた値を $\pi_j^{(n)}$ とする。

【0132】

50

(2 . 短スティックの高尤度対策)

スティックの長さが短ければ短かいほど、探索基点が密集する。そこで、密集した点付近に距離変換画像における極大点があれば、高い尤度が算出されることになる。このため、状態遷移を繰り返す度に、各スティックの長さが短くなるという問題が起こり得る。そこで、節点 a、節点 i、節点 k の 3 点を探索基点として、図 13 (b) に示すように、そのスティックの方向にモルフォロジカルスケルトンの終端 (距離) を探索することとする。このような探索によって、第 k 番目の探索基点から得られた値を $\zeta_k^{(n)}$ とする。上述の問題を解決するため、終端までの距離を尤度に加味し、スティックの長さが短い場合に ξ_k が大きくなるようにする。

【 0 1 3 3 】

(改良された尤度)

以上のようにして得られた $\delta_i^{(n)}$ 、 $\varepsilon_j^{(n)}$ 、 $\zeta_k^{(n)}$ を用いて、次式 (3 6) により $e^{(n)}$ を求める。ここで、 ξ_i 、 ξ_j 、 ξ_k は重み係数である。このようにして算出された次式 (3 6) の $e^{(n)}$ を前出の式 (3 5) により正規化して尤度 $e^{(n)}$ が求まる。

【 数 1 9 】

$$e^{(n)} = \exp \left[-\frac{1}{r} \left\{ \sum_i \xi_i \left(\min(\delta_i^{(n)}, \mu) \right)^2 + \sum_j \xi_j \left(\min(\varepsilon_j^{(n)}, \mu) \right)^2 + \sum_k \xi_k \left(\min(\zeta_k^{(n)}, \mu) \right)^2 \right\} \right] \quad (36)$$

【 実施例 】

【 0 1 3 4 】

以下では、上述した歩行者トラッキング方法を、種々な画像シーケンス (時系列画像) に適用し、評価する。まず、実施方法を示す。次に、本発明の実施形態に係る歩行者トラッキング方法 (以下、提案手法と記す) と、比較対象とするトラッキング方法とを、人工的に作成したシーケンス、及び実シーケンスにそれぞれ適用し、その結果に基づいて、本発明の有効性を示す。

【 0 1 3 5 】

(実施方法)

一般に歩行者トラッキングが失敗する原因として、例えば、オクルージョンや歩行者の速度の急激な変化、トラッキングターゲットである歩行者に似たオブジェクトの存在などがある。本発明の歩行者トラッキング方法の性能を評価するには、これらの原因を含む、多様な実シーケンスを用いて評価するのが望ましい。しかしながら、例えば、トラッキングターゲットの速度などのパラメータを細かく制御した実シーケンスを得ることは難しい。また、評価には真値が必要であり、これらの手入力は効率的でない。そこで、まずソフトウェアにより人工的に作成したシーケンスを用いて提案手法の特性を調べる。次に数種類の実シーケンスに提案手法を適用し、トラッキング精度を評価する。また、粒子数と 4 パターンのスケルトンモデル (a) (b) (c) (d)、を比較した上で、既存手法と比較検討する。

【 0 1 3 6 】

以下では、比較対象とするトラッキング手法について説明し、その後、評価基準について説明する。最後に、評価に用いる人工シーケンスと実シーケンスについて、それぞれ説明する。また、歩行者領域の誤検出を、次の 2 種類に分類する。1 つは、前景の静止物体などにより、歩行者領域の一部もしくは全体が欠ける誤検出であり、これを静的誤検出と呼ぶ。他の 1 つは、トラッキングターゲット以外の移動物体などにより、本来の歩行者領域とは異なる領域が、歩行者領域として検出される誤検出であり、これを動的誤検出と呼

10

20

30

40

50

ぶ。また、これらの誤検出の原因物体を、それぞれ静的誤検出、動的誤検出と呼ぶことにする。

【0137】

(比較対象トラッキング手法)

提案手法に対する比較対象のトラッキング方法として、カルマンフィルタによる重心のトラッキング方法、及び、コンデンセーションによる輪郭トラッキング方法を取りあげる。まず、これらの実装について簡単に述べる。カルマンフィルタによる重心のトラッキングでは、カルマンフィルタにより予測された重心の位置から、最近傍の歩行者候補領域内の画素を探索し、その画素に連続する領域内の重心座標を求め、カルマンフィルタによりこれを補正してトラッキングを行う。

10

【0138】

コンデンセーションによる輪郭トラッキングでは、トラッキングを開始するフレームの直前15フレームの真値を基に主成分分析による学習を行なう。輪郭を近似するB-スプライン曲線の節点数は、32とした。尤度算出(計測)に用いるエッジは、3×3のソベル演算子を用いて検出する。

【0139】

コンデンセーションによる輪郭トラッキングは、一般にどのような形状のオブジェクトであっても、主成分分析により事前に学習することによってトラッキング可能とされる。そこで、コンデンセーションによる輪郭トラッキングを、歩行者トラッキング専用に拡張し、精度を向上させた手法を実装し、比較対象の方法に加える。以下では、この手法をコンデンセーションによる歩行者輪郭トラッキングと呼ぶ。

20

【0140】

以上より、評価の対象は、次の4つの手法である。(1)提案手法、(2)カルマンフィルタによる重心のトラッキング、(3)コンデンセーションによる輪郭トラッキング、(4)コンデンセーションによる歩行者輪郭トラッキング。

【0141】

(歩行者輪郭トラッキング)

以下では、比較対象のトラッキング方法の1つである、コンデンセーションによる歩行者輪郭トラッキングについて説明する。コンデンセーションによる歩行者輪郭トラッキングでは、主成分分析により状態遷移を制約する代わりに、トラッキングターゲットを歩行者に限定することで状態遷移に制約をかける。まず、歩行者の輪郭を近似するB-スプライン曲線の32点の各制御点をそれぞれ、頭部、胴体、四肢の6つのグループに分類する。B-スプライン曲線の制御点を表現する座標系としては、提案手法と同様に基点を定め、各節点はこの基点からの距離とx軸に対する角度により表現する。グループIに属し、グループIにおける節点のインデックスがjである節点の角度を θ_{Ij} とする。このとき、図14に示すように、角度 θ_{Ij} を各グループ内で共通の角度 θ_I と、そこからの相対角度 ϕ_{Ij} の和として表現する。すなわち、 $\theta_{Ij} = \theta_I + \phi_{Ij}$ となる。

30

【0142】

角度 θ_{Ij} の状態遷移は2段階に分けて行なう。まず、各グループにおける共通の角度 θ_I を2次の自己回帰モデルにより遷移させ、次に、相対角度 ϕ_{Ij} を1次の自己回帰モデルにより遷移させる。すなわち、次式(37)の状態遷移モデルを用いる。

40

【数20】

$$\begin{aligned} \theta_{Ij,t} &= \Phi_{I,t} + \phi_{Ij,t} \\ &= 2\Phi_{I,t-1} - \Phi_{I,t-2} + w_g + \Phi_{Ij,t-1} + w_l \end{aligned} \quad (37)$$

【0143】

ここで、ガウス雑音 w_g の分散に比べ、 w_l の分散を小さくすることで、グループ内の節点間で相対角度の関係が保たれることになり、効率的な歩行者トラッキングが実現可能であると期待される。また、基点の絶対座標、基点から各節点までの距離の状態遷移モデ

50

ルは、それぞれ2次の自己回帰モデル、1次の自己回帰モデルを用いる。

【0144】

(評価基準)

図15は歩行者トラッキング方法におけるトラッキング精度の評価のための評価基準を示す。提案手法によるトラッキング結果は、歩行者のスケルトンによって示され、比較対象のカルマンフィルタによる重心のトラッキング結果は、重心の座標によって示され、他の比較対象であるコンデンセーションによる輪郭トラッキングやコンデンセーションによる歩行者輪郭トラッキングの結果は、歩行者の輪郭によって示される。

【0145】

これらの結果を統一的に比較検討するため、図15に示すように、トラッキングにより推定されたオブジェクトを包含する四角形に基づいて、(1)重心誤差 g 、つまり、推定された四角形の重心位置と真値とのユークリッド距離、及び(2)スケール誤差、つまり、推定された四角形の高さ d_1 と真値の高さ d_0 との比、を求めて、これらの数値を評価基準とする。

10

【0146】

重心誤差 g の単位はピクセルである。重心誤差が小さいほど、トラッキングの精度がより高いことを示し、スケール誤差が1に近いほど、歩行者の形状がより精度良く推測できていることを示す。ところで、モンテカルロフィルタによる状態推定は、確率的手法であり、毎回違ったトラッキング結果が得られるので、歩行者トラッキングを各シーケンスに対し100回ずつ行ない、これらの平均として上述の評価用数値を算出する。

20

【0147】

(人工シーケンス)

人工的にシーケンスを生成するソフトウェアをC言語により実装した。生成されるシーケンスは、歩行者候補領域の画素値が1である2値画像であり、この画像に距離変換をかけることで、提案手法への入力とする。画像のサイズは720×480とし、フレーム数は200とする。このソフトウェアは、入力用シーケンスの他に、評価や初期値の設定、事前学習に用いる真値を出力することが可能である。以下ではこのソフトウェアについて説明する。

【0148】

図16(a)~(i)は、歩行者トラッキング方法の評価に用いられる画像シーケンスの一部を示す。これらの画像は、静的誤検出、動的誤検出などの無い、最も単純な場合のシーケンスの一部について、歩行者の領域を拡大した画像である。これらの図のように、8個の楕円により人型のオブジェクトを構成した。各フレームにおける各楕円の中心の座標、長軸、短軸の長さ、長軸の傾きをそれぞれ計算し、楕円内の領域の画素値を1とすることで、人型オブジェクトの画像を生成した。

30

【0149】

上述の人工シーケンス生成ソフトウェアは、人型オブジェクトの初期位置、速度、加速度などは制御可能である。また、静的誤検出、動的誤検出の両方を追加でき、その形状および位置なども制御可能である。そこで、以下に示すシーケンスを評価用に生成した。

【0150】

(ノーマルシーケンス)

静的誤検出、動的誤検出を加えていないシーケンスを評価に用いる。以下では、このシーケンスをシーケンスNと呼ぶ。

40

【0151】

(逆再生シーケンス)

シーケンスNの途中でシーケンスを逆再生し、歩行者の速度が反対方向に変わるシーケンスを評価に用いる。以降では、このシーケンスをシーケンスRと呼ぶ。シーケンスRでは、59フレーム目、107フレーム目でシーケンスを逆再生した。

【0152】

(ガードレール型静的誤検出を加えたシーケンス)

50

図17(a)(b)は歩行者トラッキング方法の評価に用いられるガードレール型静的誤検出評価のための画像シーケンスの一部を示す。これらの図に示すように、2パターンのガードレールに見立てた静的誤検出GRを、シーケンスNに加えたシーケンスを用意した。以下では、これらのシーケンスをそれぞれシーケンスG1, G2と呼ぶ。これらのシーケンスは、18フレーム目以降、歩行者の領域Gの一部が静的誤検出により欠ける。

【0153】

(ボール型静的誤検出を加えたシーケンス)

図18(a)(b)(c)は歩行者トラッキング方法の評価に用いられるボール型静的誤検出評価のための画像シーケンスの一部を示す。これらの図に示すように、3パターンのボールに見立てた静的誤検出PをシーケンスNに加えたシーケンスを用意した。これらのシーケンスをそれぞれシーケンスP1, P2, P3と呼ぶ。シーケンスP1は、9-40, 72-92, 124-144フレーム目において歩行者の領域の一部が静的誤検出により欠ける。シーケンスP2は、9-42, 72-95, 124-152フレーム目において歩行者の領域の一部が静的誤検出により欠ける。シーケンスP3は、9-42, 48-70, 77-96, 101-160フレーム目において歩行者の領域の一部が静的誤検出により欠ける。

10

【0154】

(自動車車型動的誤検出を加えたシーケンス)

図19(a)(b)は歩行者トラッキング方法の評価に用いられる動的誤検出評価のための画像シーケンスの一部を示す。これらの図に示すように、2パターンの自動車に見立てた動的誤検出Cのそれぞれに、自動車の進行方向を歩行者と同じ方向、逆方向として、計4パターンの動的誤検出をシーケンスNに加えたシーケンスを用意した。これらのシーケンスをそれぞれシーケンスC1, C2, C3, C4と呼ぶ。シーケンスC1, C3は、車の進行方向は歩行者と逆向きであり、シーケンスC2, C4は、車の進行方向は歩行者と同じ向きである。シーケンスC1は、18-37, 96-109, 169-180フレーム目において、歩行者の領域Gと動的誤検出された領域が重なる。シーケンスC2は、9-34, 141-158フレーム目において、歩行者の領域と動的誤検出された領域が重なる。シーケンスC3は、18-41, 91-113, 163-185フレーム目において、歩行者の領域と動的誤検出された領域が重なる。シーケンスC4は、2-40, 128-165フレーム目において、歩行者の領域と動的誤検出された領域が重なる。

20

30

【0155】

(実シーケンス)

図20(a)(b)(c)は歩行者トラッキング方法の評価に用いられる実画像シーケンスの一部を示し、図20(d)は図20(a)から得られた2値画像を示す。評価に用いる実シーケンス、およびトラッキングターゲット座標の真値として、PETS (International Workshop on Performance Evaluation of Tracking and Surveillance: <http://www.cvg.cs.rdg.ac.uk/slides/pets.html>) から提供されているデータセットを用いた。評価に用いる3つのシーケンスの一部を、図20(a)(b)(c)に示す。図中の四角Gは、トラッキングターゲットである歩行者を示す。以下では、図20(a)に示すシーケンスをシーケンスU、図20(b)に示すシーケンスをシーケンスS1、図20(c)に示すシーケンスをシーケンスS2と呼ぶ。これらのシーケンスは固定カメラから撮影されたものである。シーケンスUの画像サイズは768×576であり、シーケンスS1, S2の画像サイズは720×576である。トラッキングするフレーム数は200とする。

40

【0156】

シーケンスUでは、トラッキングターゲットの動きは単純だが、135-140フレーム目において、前景の電柱により、オクルージョンが発生する。シーケンスS1, S2では、トラッキングターゲットの動きは複雑で、トラッキングターゲットに似た他のオブジェクトが画像中に多く存在する。さらに、シーケンスS2では、12-23フレーム目、

50

1 1 2 - 1 2 1 フレーム目、1 5 4 - 1 6 6 フレーム目にてトラッキングターゲットと他のオブジェクトがほぼ完全に重なり、トラッキングが極めて難しいシーケンスとなっている。

【0157】

(初期状態ベクトル、距離変換画像)

提案手法は、初期状態ベクトル x_0 、及び、距離変換画像の入力が必要である。初期状態ベクトル x_0 は、手で与えることとする。距離変換画像は、背景差分法によって得られる差分画像を基に生成する。座標 (x, y) における原画像の画素値を $I_{orig}(x, y)$ 、背景画像の画素値を $I_{bg}(x, y)$ とするとき、差分画像の画素値 $I_{sub}(x, y)$ は、次式(38)のように、原画像と背景画像の差分絶対値として得られる。

10

【数21】

$$I_{sub}(x, y) = |I_{orig}(x, y) - I_{bg}(x, y)| \quad (38)$$

【0158】

2値画像の画素値 $I_{bin}(x, y)$ は、差分画像から次式(39)に示す閾値処理により算出される。これにより、原画像において移動物体が存在する領域の、画素値が1となった2値画像が得られる。ここで τ は閾値である。

【数22】

$$I_{bin}(x, y) = \begin{cases} 1 & (I_{sub}(x, y) \geq \tau) \\ 0 & (I_{sub}(x, y) < \tau) \end{cases} \quad (39)$$

20

【0159】

図20(a)に示すシーケンスUから背景差分法により得られた画像を、図20(d)に示す。このようにして得られた2値画像に対し、式(33)に示した距離変換によって距離変換画像を生成した。

【0160】

(評価結果)

30

以下では、上述した評価方法を用いた評価結果をシーケンス毎に示し、提案手法の有効性について述べる。評価結果を示す図21~図35において、曲線a, b, c, dは、それぞれ以下の手法の結果に対応する。

提案手法(曲線a)、

カルマンフィルタによる重心のトラッキング(曲線b)、

コンデンセーションによる輪郭トラッキング(曲線c)、

コンデンセーションによる歩行者輪郭トラッキング(曲線d)。

【0161】

(シーケンスN)

まず、提案手法をシーケンスNに適用し、上述した各スケルトンモデルに対して粒子数を変化させた場合におけるトラッキング精度について、全フレームの平均として算出した重心誤差により評価する。評価結果を表1に示す。

40

【表 1】

表 1 : シーケンス N における評価結果 (重心誤差)

粒子数	100	1,000	10,000
スケルトンモデル(a)	7.192	3.365	2.248
スケルトンモデル(b)	9.004	7.425	6.416
スケルトンモデル(c)	10.129	7.469	5.935
スケルトンモデル(d)	8.986	7.421	6.411

10

【0162】

この結果から、粒子数を増やすことによるトラッキング精度の向上が確認できた。また、シーケンス N では手が検出され、また肩幅が無いことから、スケルトンモデル (a) が最も適すると期待される。結果から、この期待通りの結果が得られていることが確認できた。以下の人工シーケンスにおける評価では、粒子数を 1000 とし、スケルトンモデル (a) を用いることとする。

【0163】

図 21 (a) はシーケンス N に対する重心誤差の評価結果を示し、図 21 (b) はスケール誤差の評価結果を示す。横軸は時間を示し、その単位はフレームである。

20

【0164】

この結果から、カルマンフィルタによる重心のトラッキング (曲線 b) は、ほぼ誤差の無い性能を示した。これは、雑音が全く無いシーケンス N ではトラッキングターゲットの重心を正確に抽出可能なためである。提案手法 (曲線 a) は、コンデンセーションによる輪郭トラッキング (曲線 c)、コンデンセーションによる歩行者輪郭トラッキング (曲線 d) よりも優れた性能を示しているが、重心誤差に周期的なばらつきが見られる。この周期は、歩行の周期と同一であり、図 15 (e) (f) のようなフレームで誤差が大きくなる。これは、提案手法では膝の関節をモデル化していないので、粒子の尤度算出精度が低下するためだと考えられる。コンデンセーションによる輪郭トラッキング (曲線 c) では、スケール誤差の増大が確認できる。これは主成分分析による状態遷移への制約が適切でないことを示す。スケール誤差の増大により、各粒子の尤度が減少し、図 21 (a) のようにトラッキングターゲットを見失う。また、コンデンセーションによる輪郭トラッキング (曲線 c) のような汎用的なトラッキング手法では、一度誤差が大きくなると再び誤差が小さくなる可能性は低い。一方、提案手法 (曲線 a) では、誤差が一時的に大きくなったとしても、再び誤差を小さくすることが可能であることが確認できた。

30

【0165】

(シーケンス R)

図 22 (a) はシーケンス R に対する重心誤差の評価結果を示し、図 22 (b) はスケール誤差の評価結果を示す。シーケンス R では、ほぼシーケンス N と同様の結果が得られた。これは、このシーケンスは各フレーム間におけるトラッキングターゲットの移動距離が小さいため、急な逆再生でもほぼ影響が無いと思われる。

40

【0166】

(シーケンス G1, G2)

図 23 (a) (b) はシーケンス G1 に対する評価結果を示し、図 24 (a) (b) はシーケンス G2 に対する評価結果を示す。シーケンス G1, G2 では、静的誤検出により歩行者の領域が分割されるため、シーケンス N に比べてカルマンフィルタによる重心のトラッキング (曲線 b) の精度低下が目立つ。一方、提案手法 (曲線 a) ではそれほど大きな精度の低下は見られず、コンデンセーションによる歩行者輪郭トラッキング (曲線 d)

50

も提案手法とほぼ同等なトラッキング精度を示した。シーケンスG2では、提案手法（曲線a）は、60 - 120フレーム目に大きな誤差が見られる。これは、長期に渡ってスケルトンの脚における終端が静的誤検出により検出されず、全体的に（-y）方向にずれるためである。

【0167】

（シーケンスP1, P2, P3）

図25（a）はシーケンスP1に対する評価結果を示し、図26（a）（b）はシーケンスP2に対する評価結果を示し、図27（a）（b）はシーケンスP3に対する評価結果を示す。シーケンスP1, P2, P3のどのシーケンスにおいても、提案手法（曲線a）は、静的誤検出により一時的に大きくなった重心誤差を再び小さくすることができ、最も優れた性能を示した。コンデンセーションによる歩行者輪郭トラッキング（曲線d）は、シーケンスG1, G2では提案手法とほぼ同等の性能を示していたが、シーケンスP2, P3で特に大きな誤差が見られる。これは、静的誤検出により、検出されたトラッキングターゲットとの境界に輪郭が検出されるためだと考えられる。また、カルマンフィルタによる重心のトラッキング（曲線b）では静的誤検出により、重心位置が不安定となり、トラッキングに失敗した。

10

【0168】

（シーケンスC1, C2, C3, C4）

図28（a）（b）はシーケンスC1に対する評価結果を示し、図29（a）（b）はシーケンスC2に対する評価結果を示し、図30（a）（b）はシーケンスC3に対する評価結果を示し、図31（a）（b）はシーケンスC4に対する評価結果を示す。シーケンスC1, C2, C3, C4のどのシーケンスにおいても、提案手法（曲線a）は最も優れた性能を示した。特に、シーケンスC3, C4ではトラッキングターゲットのほぼ全域が一時的に動的誤検出された領域と重なるが、提案手法ではトラッキングに成功している。一方、他の手法（曲線b, c, d）では、動的誤検出により大きな誤差が確認された。コンデンセーションによる歩行者輪郭トラッキング（曲線d）では、特に歩行者と動的誤検出の移動方向が同一の場合、スケール誤差の増大に示されるように歩行者の輪郭の形状が大きく乱れ、トラッキングに失敗した。カルマンフィルタによる重心のトラッキング（曲線b）では、動的誤検出された領域を誤ってトラッキングした。

20

【0169】

（実シーケンスU）

まず、提案手法をシーケンスUに適用し、各スケルトンモデルに対して粒子数を変化させた場合における精度について、全フレームの平均として算出した重心誤差により評価する。評価結果を表2に示す。

30

【表2】

表2：シーケンスUにおける評価結果（重心誤差）

粒子数	100	1,000	10,000
スケルトンモデル(a)	8.505	6.966	6.537
スケルトンモデル(b)	7.461	6.588	6.392
スケルトンモデル(c)	9.418	8.459	7.698
スケルトンモデル(d)	9.720	9.427	8.263

40

【0170】

この結果から、粒子数を増やすことによるトラッキング精度の向上が確認できた。また、シーケンスUからは手が検出されず、また肩幅がほぼ無いことからスケルトンモデル（

50

b) が最も適すると期待される。評価結果から、この期待通りの結果が得られていることが確認できた。

【0171】

図32(a)(b)はシーケンスUに対する評価結果を示す。各手法の粒子数を1,000とし、また提案手法ではスケルトンモデル(b)を用いた。この結果、提案手法(曲線a)は最も優れたトラッキング精度を達成したが、カルマンフィルタによる重心のトラッキング(曲線b)、コンデンセーションによる歩行者輪郭トラッキング(曲線d)も提案手法とほぼ同等の精度を示した。

【0172】

(実シーケンスS1)

まず、提案手法をシーケンスS1に適用し、各スケルトンモデルに対して粒子数を変化させた場合における精度について、全フレームの平均として算出した重心誤差により評価する。評価結果を表3に示す。

【表3】

表3：シーケンスS1における評価結果(重心誤差)

粒子数	100	1,000	10,000
スケルトンモデル(a)	71.738	41.222	26.238
スケルトンモデル(b)	52.440	27.688	6.820
スケルトンモデル(c)	53.786	26.761	11.584
スケルトンモデル(d)	48.727	27.633	8.659

【0173】

この結果から、粒子数を増やすことによるトラッキング精度の向上が確認できた。また、シーケンスS1からは手が検出されないフレームが多く、また肩幅があるフレームが多いことからスケルトンモデル(d)が最も適すると期待されるが、結果からスケルトンモデル(b)が最も優れた性能を示した。これは、シーケンスS1におけるトラッキングターゲットの領域が小さいため、肩幅の影響があまり問題にならないことを示している。

【0174】

図33(a)(b)はシーケンスS1に対する評価結果を示す。各手法の粒子数を10,000とし、また提案手法ではスケルトンモデル(b)を用いた。この結果、提案手法(曲線a)とコンデンセーションによる歩行者輪郭トラッキング(曲線d)がほぼ同等の性能を示している。一方、カルマンフィルタによる重心のトラッキング(曲線b)はトラッキングターゲットの急な速度の変化に対応できず、20フレーム目以降、違うオブジェクトを誤ってトラッキングした。

【0175】

(実シーケンスS2)

まず、提案手法をシーケンスS2に適用し、各スケルトンモデルに対して粒子数を変化させた場合における精度について、全フレームの平均として算出した重心誤差により評価する。評価結果を表4に示す。

【表 4】

表 4 : シーケンス S 2 における評価結果 (重心誤差)

粒子数	100	1,000	10,000
スケルトンモデル(a)	108.847	32.693	12.333
スケルトンモデル(b)	75.874	12.270	7.801
スケルトンモデル(c)	110.883	36.250	23.430
スケルトンモデル(d)	58.349	20.794	7.847

10

【0176】

この結果から、粒子数を増やすことによるトラッキング精度の向上が確認できた。また、シーケンス S 2 では前半は手が検出されないフレームが多く、後半は手が検出されるフレームが多い。またトラッキングターゲットの肩幅はほぼ無いフレームが多いことからスケルトンモデル (a) もしくは (b) が最も適すると期待される。結果から、スケルトンモデル (b) が最も良い性能を示すことが分かった。これは前半において誤差が大きい場合、後半に誤差が伝搬するためだと考えられる。

【0177】

(適応的スケルトンモデル)

図 3 4 は歩行者トラッキングに適応的スケルトンモデルを用いた場合の重心誤差の評価結果を示す。手が検出されない 97 フレーム目までにおいてスケルトンモデル (b) を用い、97 フレーム目以降、スケルトンモデル (a) に切り替えることにより、トラッキング精度が更に向上すると考えられる。このモデルを適応的スケルトンモデルと呼ぶ。適応的スケルトンモデル、スケルトンモデル (b) を用いた場合における重心誤差の比較を図 3 4 に示す。粒子数はそれぞれ 10,000 とした。この結果、適応的スケルトンモデルはスケルトンモデル (b) よりも後半において優れた性能を示した。今回、スケルトンモデルの切り替えは手動で行なったが、トラッキング実行中に自動的にスケルトンモデルを切り替えることにより、より精度向上を図ることができる。

20

30

【0178】

図 3 5 (a) (b) はシーケンス S 2 に対する評価結果を示す。各手法の粒子数を 10,000 とし、また提案手法ではスケルトンモデル (b) を用いた。この結果、提案手法 (曲線 a) は最も優れたトラッキング精度を示した。しかし、図 3 5 (b) に示すようにスケール誤差が大きい。これはシーケンス S 2 ではトラッキングターゲットが画面奥から手前に移動することによるオブジェクト領域の拡大を正確に推測できていないことを示す。コンデンセーションによる歩行者輪郭トラッキング (曲線 d) では、トラッキングターゲットと他のオブジェクトが重なる 20 フレーム目前後に大きなスケール誤差がある。これは推定された輪郭の形状が大きく乱れたことを意味し、これによってトラッキングに失敗した。カルマンフィルタによる重心のトラッキング (曲線 b) では、120 フレーム目以降、他のオブジェクトを誤ってトラッキングした。

40

【0179】

(まとめ)

以上に示したように、カルマンフィルタによる重心のトラッキング (曲線 b) は、雑音の少ない状況では良い精度を示したが、しばしば動的誤検出された領域を誤ってトラッキングした。コンデンセーションによる輪郭トラッキング (曲線 c) では、歩行者の輪郭の遷移に対して、主成分分析による制約が難しいため、スケール誤差が増大し、トラッキングに失敗した。コンデンセーションによる歩行者輪郭トラッキング (曲線 d) は、トラッキングターゲットが動的誤検出された領域と重なる場合に、推定された輪郭が大きく乱れ、著しい精度の低下を示した。

50

【 0 1 8 0 】

一方、提案手法（曲線 a）は、従来手法（曲線 b, c, d）ではトラッキングに失敗するようなシーケンスにおいても、安定したトラッキングを実現した。4 パターンのスケルトンモデル（a）～（d）を比較した結果、スケルトンモデル（b）はどのシーケンスにおいても安定した精度を見せた。また、提案手法のトラッキング精度が、粒子数を増やすことにより、向上することを確認できた。

【 0 1 8 1 】

なお、提案手法、すなわち、本発明の歩行者トラッキング方法及び歩行者トラッキング装置は、シーケンス G 2 のような静的誤検出に対して、式（36）における重み係数を適応的に変化させて粒子の尤度を算出することにより改良可能である。また、拡大するオブジェクトへの対処に対しては、式（22）に示したズームパラメータ z の算出手法に移動方向に応じた要素を反映させることにより対処可能である。また、提案手法は、初期状態ベクトルを画像から自動生成するようにすることができる。提案手法は、画像中の複数の歩行者に対してそれぞれ独立に適用することができる。これにより、複数の歩行者を同時にトラッキング可能である。この場合に、相互に、近隣の歩行者の動きを尤度の算出に反映させることにより、より効率的に複数の歩行者をトラッキングできる。

【 0 1 8 2 】

また、図 8 および図 10 に示したスケルトンモデルにおける節点 b の座標の分散²を算出し、当該分散が一定の値を超える場合には、追跡対象は歩行者でないとして判定し、追跡を終了するようにしてもよい。節点 b の座標を x, y（座標値 x_i, y_i; i = 0 ~ N、その平均を x_m, y_m）とすると、分散²は次式（40）で与えられる。

【 数 2 3 】

$$\sigma^2 = \frac{\sum_{i=0}^N \left\{ (x_m - x_i)^2 + (y_m - y_i)^2 \right\}}{N} \quad (40)$$

【 0 1 8 3 】

上述のように、予測工程によって予測された各スケルトンモデルにおける、胴体スティックと頭長スティックとの節点 b の座標の分散²を算出し、当該分散²が一定の値を超える場合には追跡を終了する工程、をさらに備えることにより、非歩行者を誤って歩行者領域として抽出してしまった場合であっても、節点 b の座標の分散²から、追跡対象が歩行者でないことを判別でき、誤った追跡を終了することができる。

【 0 1 8 4 】

また、本発明は、上記構成に限られることなく種々の変形が可能である。例えば、歩行者トラッキングを、前段の候補領域の分類にフィードバックすることにより、その精度の向上に利用できる。また、尤度を算出するため、スケルトンモデルと比較する対象となるモルフォロジカルスケルトン（画像から画像処理によって得られたスケルトンモデル）として、上述した距離変換画像を用いる方法の他に、図 7（b1）～（b3）に示すような、収縮演算（A B）と切断演算（A B）とを用いて得られるモルフォロジカルスケルトンを用いることができる。また、本発明の歩行者トラッキング方法及び歩行者トラッキング装置は、歩行者に限らず、移動物体、例えば、動物やロボットなどの種々のオブジェクトに対し、それぞれスケルトンモデルによりモデル化して、適用することができる。

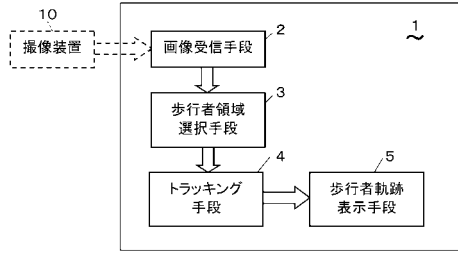
【 0 1 8 5 】

なお、本発明の歩行者トラッキング方法及び歩行者トラッキング装置は、CPU やメモリや外部記憶装置や表示装置や入力装置などを備えた一般的な構成を備えた電子計算機上のプロセス又は機能の集合を用いて構成することができる。

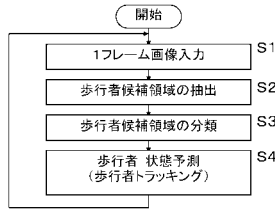
【 0 1 8 6 】

この出願は 2006 年 7 月 10 日付けの特許出願に基づいて優先権主張を行う。その出願の内容の全体が参照によって、この出願に組み込まれる。

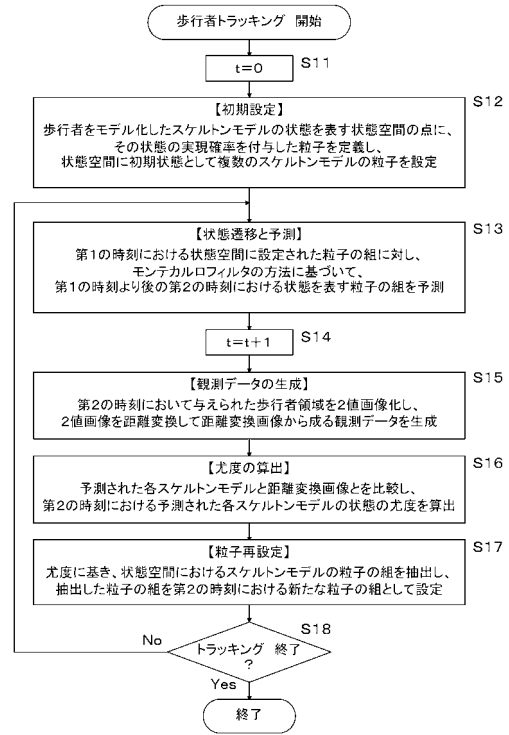
【図1】



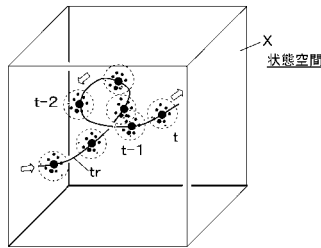
【図2】



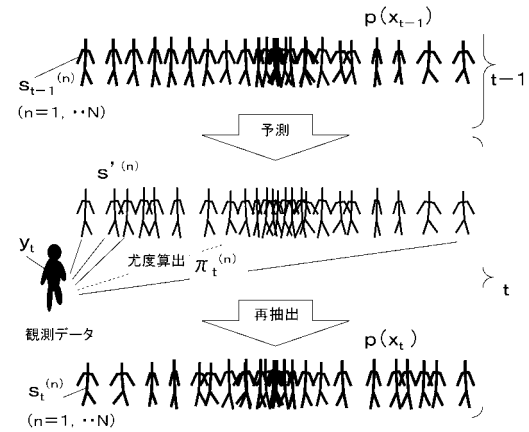
【図3】



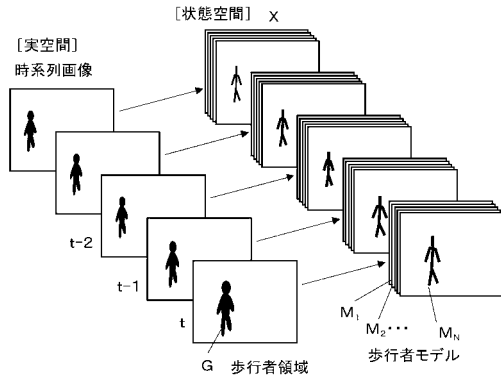
【図4】



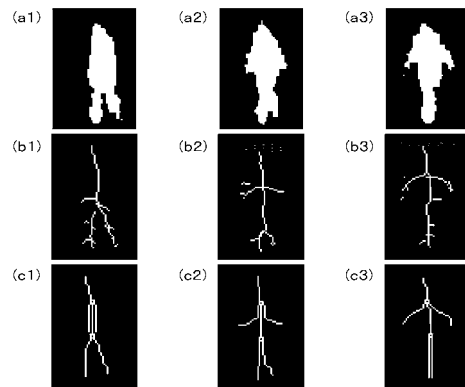
【図6】



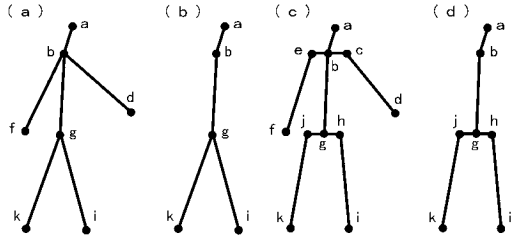
【図5】



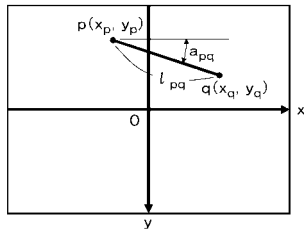
【図7】



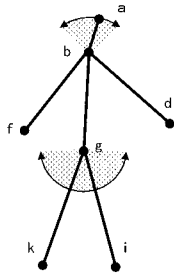
【 図 8 】



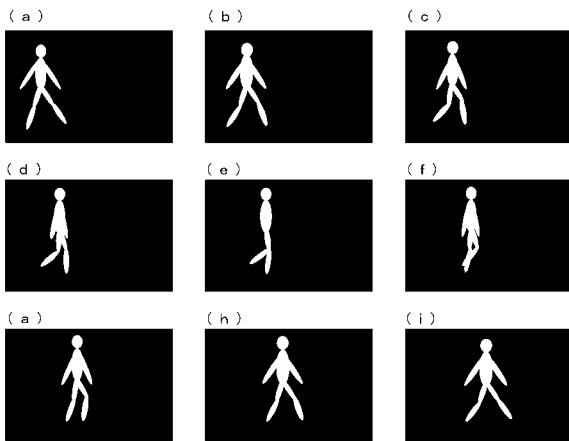
【 図 9 】



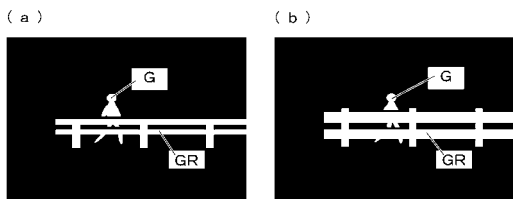
【 図 10 】



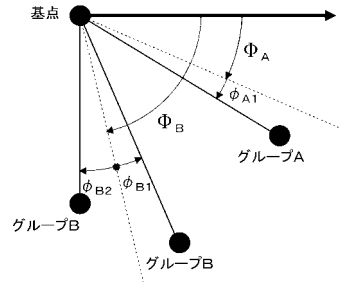
【 図 16 】



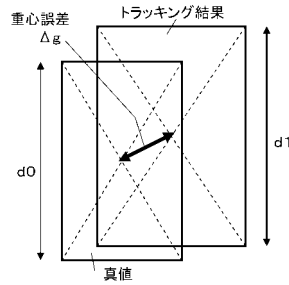
【 図 17 】



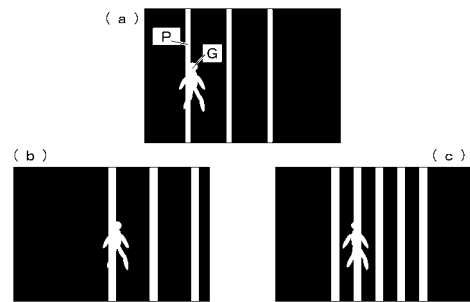
【 図 14 】



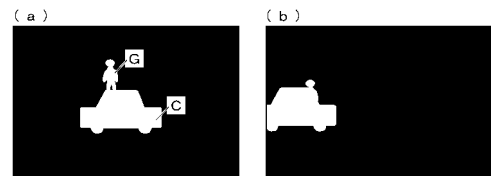
【 図 15 】



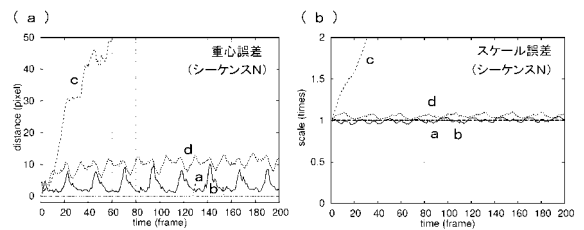
【 図 18 】



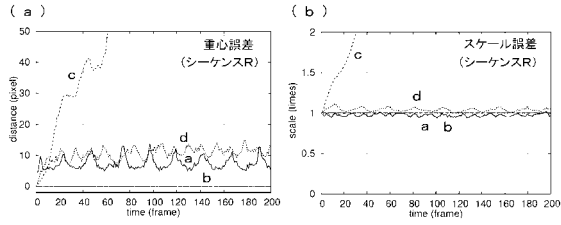
【 図 19 】



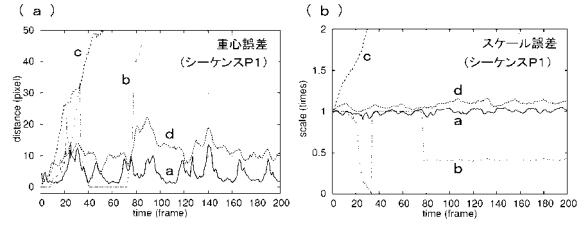
【 図 21 】



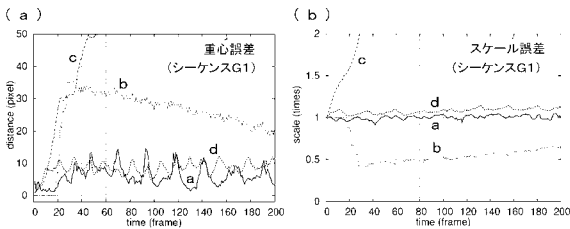
【図 2 2】



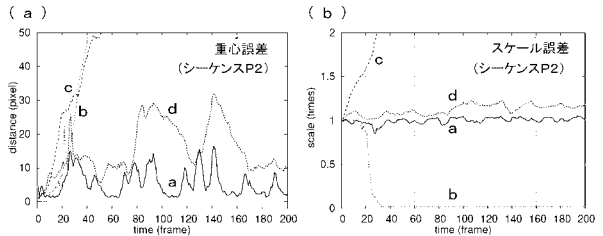
【図 2 5】



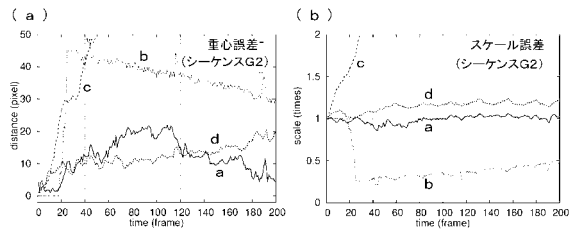
【図 2 3】



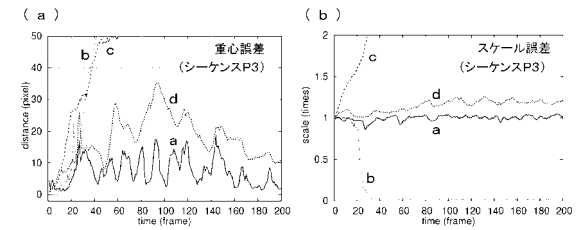
【図 2 6】



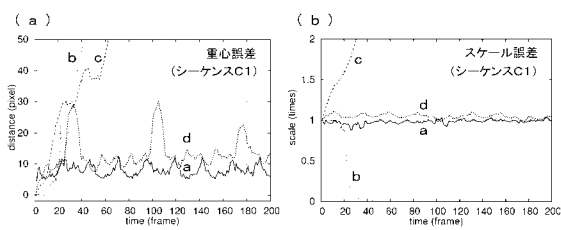
【図 2 4】



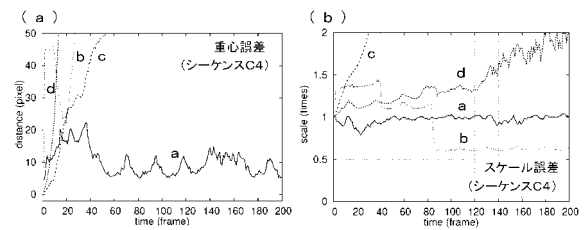
【図 2 7】



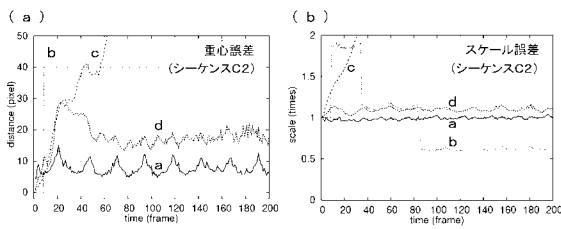
【図 2 8】



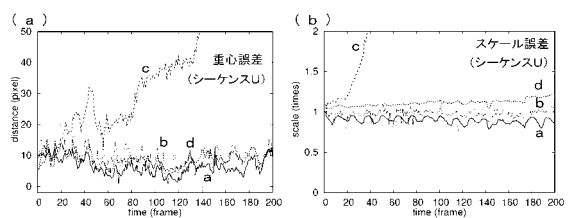
【図 3 1】



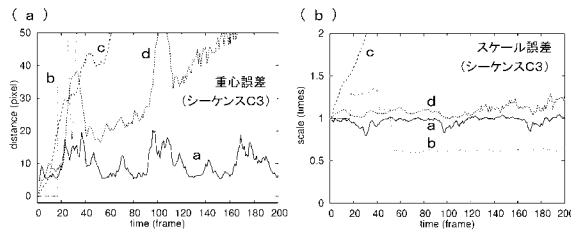
【図 2 9】



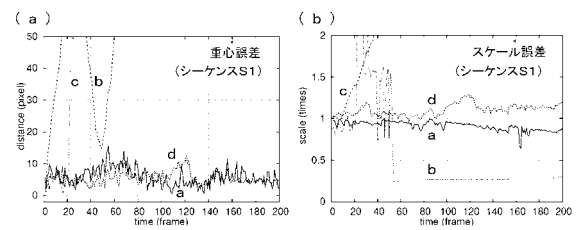
【図 3 2】



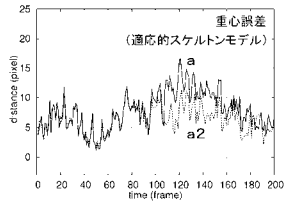
【図 3 0】



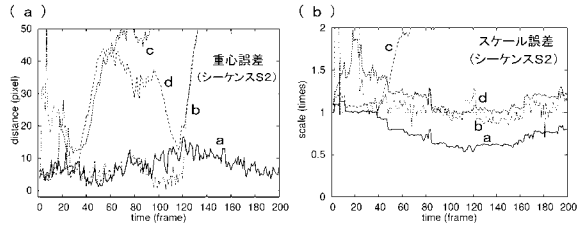
【図 3 3】



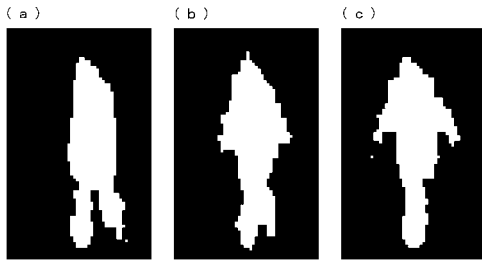
【 図 3 4 】



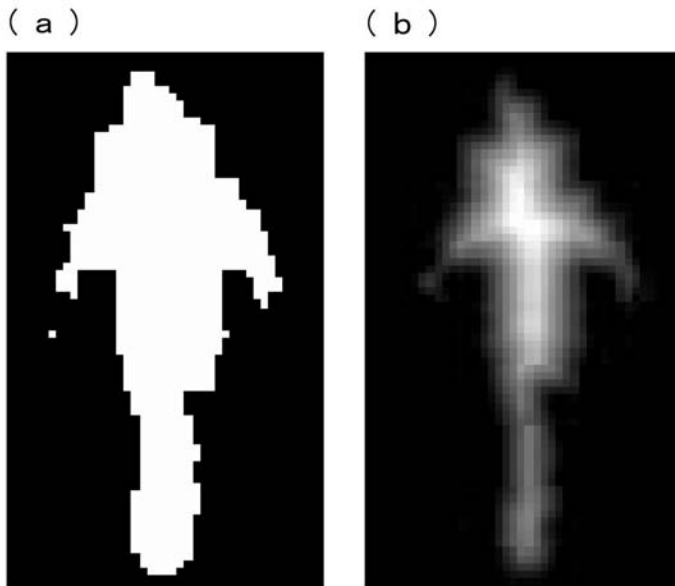
【 図 3 5 】



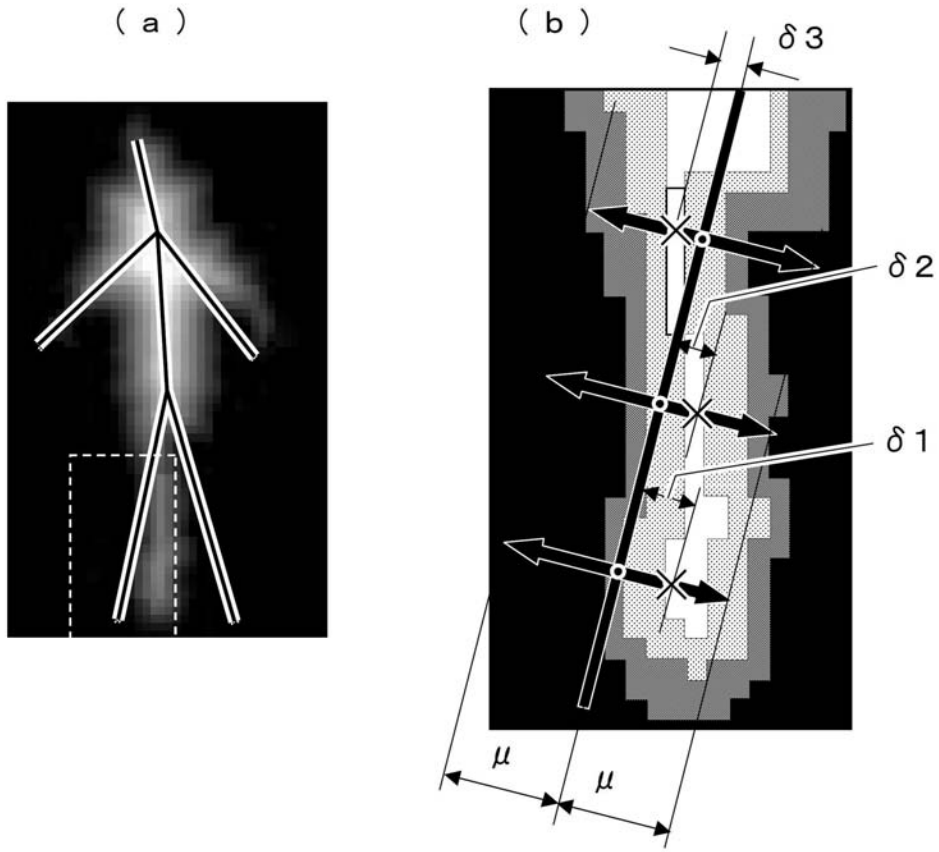
【 図 3 9 】



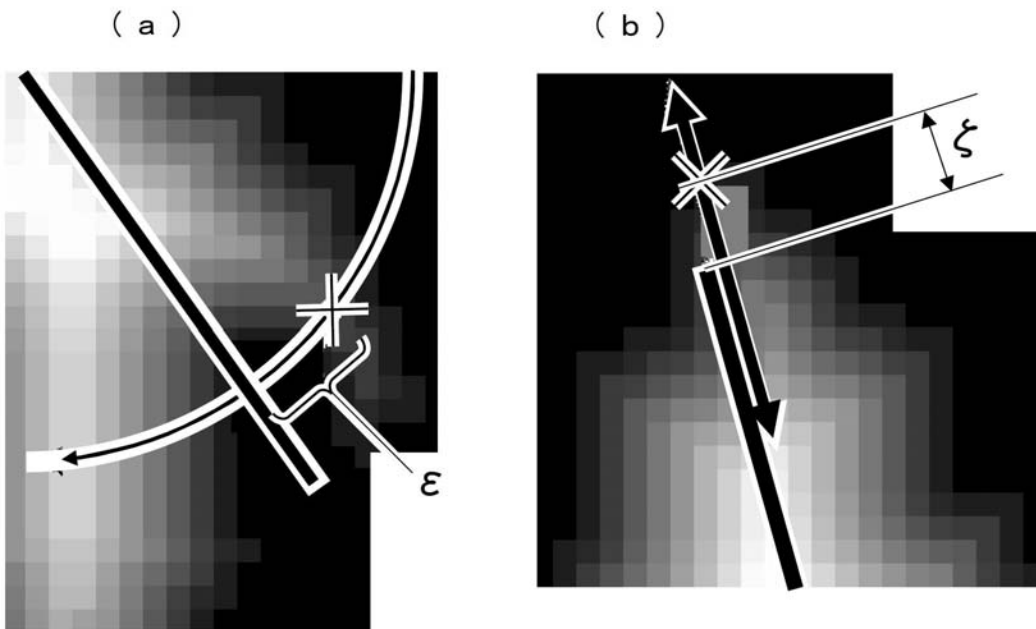
【 図 1 1 】



【 図 1 2 】

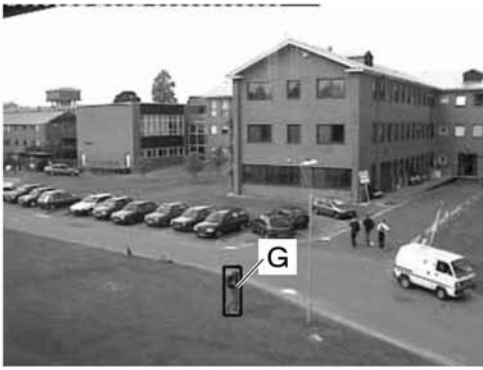


【 図 1 3 】



【 図 20 】

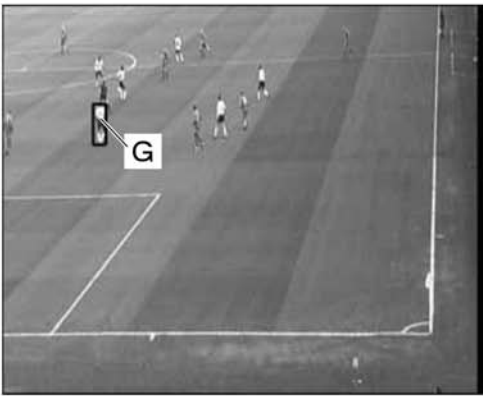
(a)



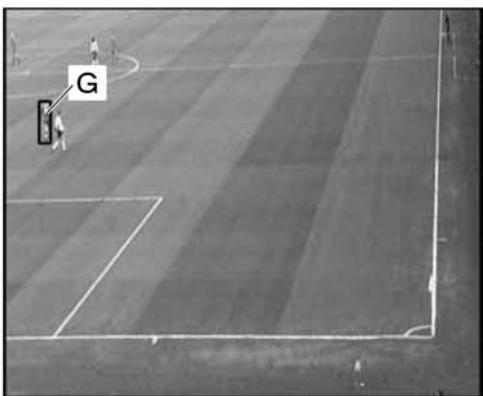
(d)



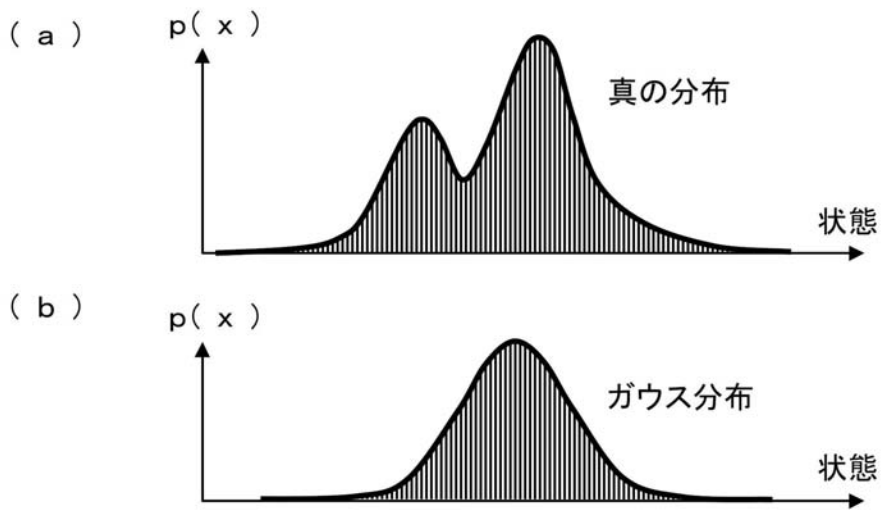
(b)



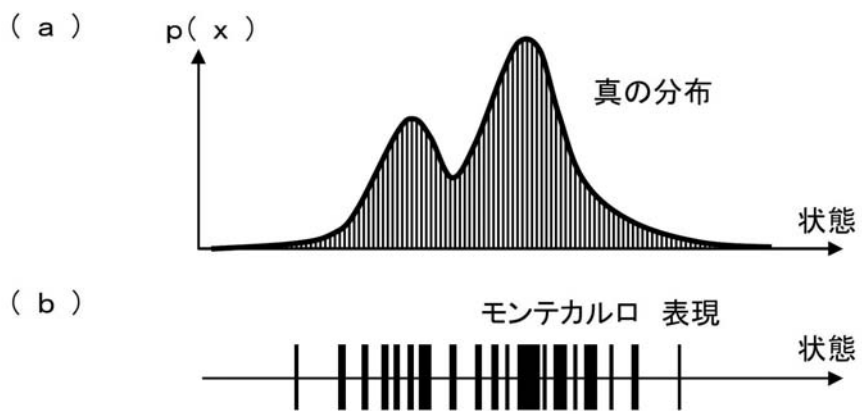
(c)



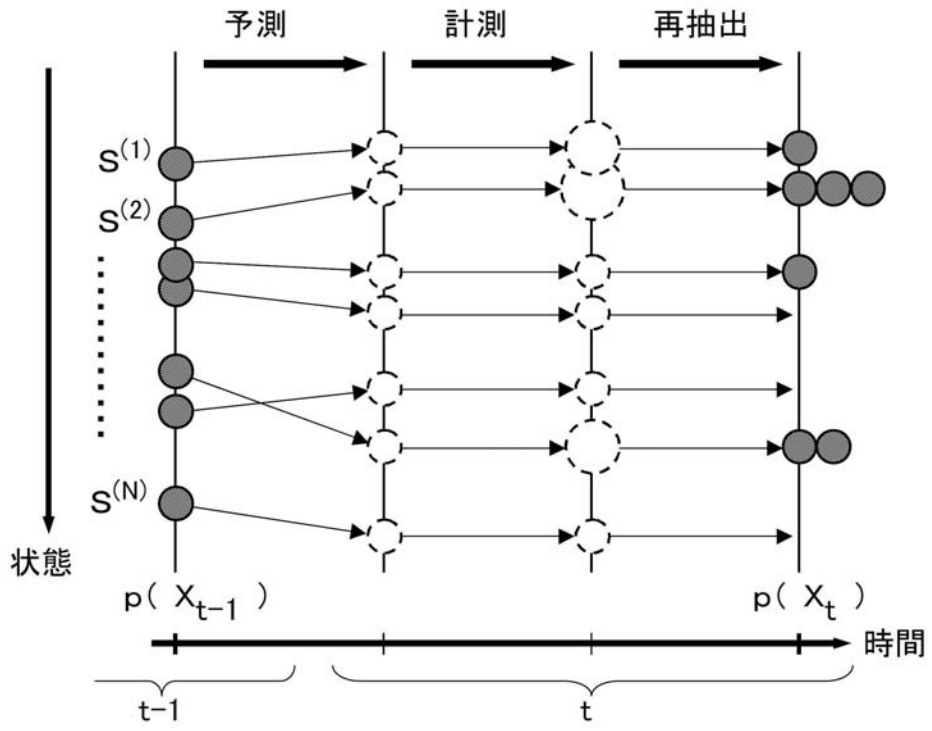
【図36】



【図37】



【 図 3 8 】



フロントページの続き

審査官 松尾 俊介

- (56)参考文献 特開2005-339100(JP,A)
特開2005-165688(JP,A)
特表2009-514111(JP,A)
樋口知之, 粒子フィルタ, 電子情報通信学会誌, 日本, 社団法人電子情報通信学会, 2005年
12月1日, 第88巻, 第12号, p.989-994
池田浩雄, 外2名, 行動に基づく関節自由度の制約を利用した人体姿勢推定, FIT2005第
4回情報科学技術フォーラム 一般講演論文集 第3分冊 画像認識・メディア理解 グラフ,
日本, 社団法人情報処理学会, 社団法人電子情報通信学会, 2005年9月22日, p.115-11
6
Raquel Urtasun 外3名, Priors for People Tracking from Small Training Sets, Computer V
ision, 2005. ICCV 2005. Tenth IEEE International Conference on, 米国, IEEE Computer soci
ety, 2005年10月17日, Vol.1, p.403-410
Isard, Brake, CONDENSATION: Conditional Density Propagating for Visual Tracking, Interna
tional Journal of Computer Vision, 米国, Springer, 1998年8月, Vol.29, No.1, p.5-
28

(58)調査した分野(Int.Cl., DB名)

G06T 7/00~7/60

H04N 7/18