

NAIST-IS-MT0351046

修士論文

ステレオビジョンによる人の検知手法に関する研究

笠島 尚志

2005年2月3日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(工学) 授与の要件として提出した修士論文である。

笠島 尚志

審査委員： 緒方 司郎 教授
横矢 直和 教授
千原 國宏 教授
諏訪 正樹 助教授

ステレオビジョンによる人の検知手法に関する研究*

笠島 尚志

内容梗概

本論文は、カメラから得られる画像データのみを用いて人を検知・追跡する手法に関するものである。近年、セキュリティに対する意識の向上から、監視カメラで捉えた画像から人を検知し、追跡するビジョンセンシング技術のニーズが高まっている。また、自動で画像内から人を検知し、移動軌跡を検出した結果は、小売店舗で必要とされるマーケティング情報に直結するものである。画像データから人を検出・追跡する技術はその多くが、主に低コストであるという観点から単眼カメラでおこなわれているのが現状である。単眼カメラによるビジョンセンシングでは、対象の見かけ変化や隠れが発生した場合に安定した結果を得ることが困難であることが従来から指摘されてきた。

そこで本論文では、3次元空間情報を利用するビジョンセンシング技術を提案する。具体的にはステレオカメラを用いて画像中の特徴点群から3次元空間情報を取得し、得られた3次元空間情報を人のアピアランスモデルをベースにクラスタリングすることで人検知を行う。さらにクラスタリング結果から、推定した人の3次元空間上での存在位置を統計処理を用いて時空間的に連結することで追跡を行う。輝度値を直接利用せず、ステレオ視により得た人の3次元情報を特徴量として利用することで、カメラ設置位置に対する制約が非常に少ない、ロバストな人検知および追跡の実現が可能になった。また、市販のカメラおよびコンピュータを用いることで、手軽で自由度の高いシステム実現を可能とした。

キーワード

ステレオビジョン, 特徴抽出, 人検知, 人追跡, カルマンフィルタ, ロバスト性

* 奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT0351046, 2005年2月3日.

Study of Human Detection using Stereo Vision*

Hisashi Kasajima

Abstract

This thesis discusses a stereo-based vision system for the human detection and tracking. Recently, needs of human detection and tracking technology based on vision sensing are rapidly rising with increasing awareness about security. In addition, it is directly linked to marketing research in the retail industry to analyze trajectories of the customers using vision sensor. Most conventional human detection and tracking systems have only a monocular camera from a viewpoint of low-cost. However, it has been pointed out that conventional vision based systems do not work well when the appearance change on the image or occlusion of the objects occur.

This thesis proposes a novel human detection and tracking method using 3D information. In our method, 3D space information is calculated using some feature points, which almost exist on the contours of each humans. Each person in the image is detected by our proposed algorithm, which is based on new human appearance model. Furthermore, detected persons can be tracked on the 3D space using statistical spacio-temporal method. Our proposed method realizes a robust human detection and tracking system and also can construct a low cost system with off-the-self cameras and PC.

Keywords:

stereo vision, feature extraction, human detection, human tracking, kalman filter, robustness

* Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0351046, February 3, 2005.

目次

1. 緒言	1
1.1 はじめに	1
1.2 ビジョンセンシングにおける人の検知および追跡手法	1
1.3 本研究の目的と論文の構成	2
2. ステレオビジョンによる人検知および追跡	4
2.1 従来の人検知および追跡手法	4
2.2 特徴点抽出	5
2.3 ステレオ画像処理技術	7
2.3.1 ステレオ画像処理原理	7
2.3.2 ステレオマッチング	8
2.4 クラスタリング手法	9
2.4.1 階層型分類	9
2.4.2 非階層型分類	11
2.4.3 ステレオビジョンによる人検知 [7],[9]	12
2.5 追跡手法	13
3. 人検知および追跡のロバスト性の向上	15
3.1 Harris Operator を利用した特徴点抽出	15
3.1.1 応答関数出力値 R と特徴点の関係	16
3.1.2 人輪郭特徴点の抽出	18
3.1.3 本手法の特徴	18
3.2 姿勢変化に対応した人の検知	20
3.2.1 頭部位置検出フィルタ	21
3.2.2 3次元追跡モデルを利用した時空間追跡	22
3.2.3 クラスタ毎の高さ閾値の更新	25
3.2.4 姿勢変化が発生した場合の人検知	26
3.2.5 本手法の特徴	27

4. 実験	28
4.1 概要	28
4.2 実験環境およびシステム構成	29
4.3 実験結果	30
4.3.1 通常歩行時	30
4.3.2 ジャンプアップ時	32
4.3.3 転倒時	34
4.3.4 多人数出現時	36
5. 考察	38
5.1 本手法の性能検証に対する考察	38
5.1.1 処理時間	38
5.1.2 位置追跡精度	38
5.2 従来手法 ([9]) との比較	39
5.2.1 処理時間	39
5.2.2 位置追跡精度	39
6. 結言	41
謝辞	42
参考文献	43

目 次

1	全体処理フロー	5
2	テンプレートマッチング	6
3	ステレオ画像処理原理	7
4	人物輪郭モデル	12
5	濃淡変化がない場合	17
6	勾配方向が一様な場合 (エッジ)	17
7	勾配が交わる場合 1	17
8	勾配が交わる場合 2	17
9	勾配が交わる場合 3	17
10	入力画像	19
11	特徴量を輝度値に変換した画像	19
12	人高さ閾値	20
13	検知追跡処理フロー	21
14	高さ方向重みの設定	22
15	Kalman Filter フロー	24
16	姿勢変化に伴う高さ位置の変動	26
17	通常歩行時の画像	30
18	xz 平面から見た軌跡 (提案手法)	31
19	xz 平面から見た軌跡 (従来手法)	31
20	xy 平面から見た軌跡 (提案手法)	31
21	xy 平面から見た軌跡 (従来手法)	31
22	yz 平面から見た軌跡 (提案手法)	31
23	yz 平面から見た軌跡 (従来手法)	31
24	ジャンプ時の画像	32
25	xz 平面から見た軌跡 (提案手法)	33
26	xz 平面から見た軌跡 (従来手法)	33
27	xy 平面から見た軌跡 (提案手法)	33
28	xy 平面から見た軌跡 (従来手法)	33

29	yz 平面から見た軌跡 (提案手法)	33
30	yz 平面から見た軌跡 (従来手法)	33
31	転倒時の画像	34
32	xz 平面から見た軌跡 (提案手法)	35
33	xz 平面から見た軌跡 (従来手法)	35
34	xy 平面から見た軌跡 (提案手法)	35
35	xy 平面から見た軌跡 (従来手法)	35
36	yz 平面から見た軌跡 (提案手法)	35
37	yz 平面から見た軌跡 (従来手法)	35
38	多人数出現画像	36
39	xz 平面から見た軌跡 (提案手法)	37
40	xz 平面から見た軌跡 (従来手法)	37
41	xy 平面から見た軌跡 (提案手法)	37
42	xy 平面から見た軌跡 (従来手法)	37
43	yz 平面から見た軌跡 (提案手法)	37
44	yz 平面から見た軌跡 (従来手法)	37

表 目 次

1	特徴点抽出パラメータ	28
2	人検知パラメータ	29
3	実験環境およびシステムのスペック	29
4	通常歩行時の位置追跡精度および処理時間	30
5	ジャンプ時の位置追跡精度および処理時間	32
6	転倒時の位置追跡精度および処理時間	34
7	多人数出現時の位置追跡精度および処理時間	36

1. 緒言

1.1 はじめに

近年，侵入・盗難などの防犯，財産保全および人身の安全確保などセキュリティに対する関心が高まってきており，不審者や侵入者を自動で検知する技術に対するニーズは今後さらに増えていくものと思われる．また，小売店舗におけるプロモーション戦略，商品戦略等の構築において，人の動作軌跡は重要なマーケティング情報の一つである．このように人の存在を検知し，その動きを追いかける技術は，様々なアプリケーションへの展開が考えられる重要な基盤技術である．本論文では入力系としてカメラを用いた人の検知・追跡技術について述べる．カメラで捉えた画像を処理（ビジョンセンシング）することで視野内にいる人を検知し，追跡する技術は，

1. エリア内に進入した人を検知し，追跡することで行動を認識する監視システム
2. 小売店舗内における人の移動軌跡を利用したマーケティング調査
3. 各施設での入退場者数計測
4. エレベータや駅，停留所等での待ち人数計測

など様々な応用が考えられる．またビジョンセンシング技術は，既存の赤外線センサなどと比較して，可視的な情報が得られることや，処理エリアが広いことから今後の普及が期待される分野である．また，録画が可能であり，目視による直接的なモニタリングが不要となり，ネットワークを利用することによって遠隔監視が可能であるという利点もある．

1.2 ビジョンセンシングにおける人の検知および追跡手法

ビジョンセンシングにおける現状の課題は検知性能に対する信頼性である．人間同士が画像上で重なるような状況下において正確に切り分けることができない，

影，照明変動，映りこみなど周囲の環境変動を誤って人と認識してしまうなどが挙げられる．

単眼カメラを用いて人検知を行うビジョンセンシング技術には，多くの既存技術が存在する．単眼輝度画像を用いることで人検知を行う [1], [2], [3] 等は，背景差分，フレーム間差分を基本とした手法により人検知を行っている．そして，オプティカルフロー等の動きベクトルに基づく手法で追跡を行う．森田 [4] は局所相関演算に基づいた動き検知オペレータで動物体を検知し，動き検知処理と動きベクトルを組み合わせることで追跡を行っている．Javed ら [5] は，複数の単眼カメラを視野が被さるようにシーケンスに並べ，人の検知および追跡を実現している．これらの単眼カメラを用いたビジョンセンシングによる人検知は，人の重なり，隠れ，照明変化等の環境条件の変動に対してロバストに対応できない．

ステレオビジョンでは対象の3次元情報を用いるため，照明変化等の環境条件の変動に対してロバストに対応できる．Beymer ら [6] は横方向と縦方向にカメラを設置して，視点依存性の少ないステレオ距離情報を得て人の検知，追跡を行っている．しかしながらカメラ設置位置に制限があり汎用的ではない．また，池谷ら [7], [9] は，交通流センサに用いられている自社開発技術SV(Silhouette Vision)を用いた人検知処理を実現している．ステレオビジョンを利用することで，環境の変化に依存しない，人同士が重なるといったオクルージョンに強い検知を実現している．しかしながら，3D空間中のある高さに固定した人のアピアランスモデルを利用した検知を行っているため，様々な姿勢の人を検知することが難しい．

1.3 本研究の目的と論文の構成

本論文では，重なりや照明変動等に強い高精度でロバストな人検知，追跡および設置位置に自由度を持たせた汎用的なシステム構築を目的とする．文献 [7], [9] で提案された手法をもとに，人の輪郭上に存在する特徴点を抽出する処理，姿勢変動が起こった場合にも対応可能な人検知処理，時空間的な統計処理による3次元的な人追跡処理手法を新たに提案する．

特徴点抽出においては，Harris Operator を利用して，画像信号の相関性と特徴点の関係に着目し，それらを閾値として与えることで，できるだけ人の輪郭上か

ら特徴点を抽出できるようにする．そして，得られた特徴点群からエピソード拘束に基づいてステレオマッチングを行い，3D 空間座標を算出する．高さ位置推定値と現在高さ位置に対し，頭部位置検出フィルタをかけることで次フレームでの高さを決定し，人の輪郭モデルに基づく最長距離法により人検出を行う．得られた人の重心空間座標を Kalman Filter を利用して追跡を行うことで 3 次元的な動きの追跡を実現する．高さ情報を追跡モデルに加えることで，3 次元的な追跡を可能とし，現在の追跡結果を一期先の人検知および追跡にフィードバックさせることで，人検知追跡の精度向上を図る．

次に本論文の構成を述べる．第 2 章ではステレオビジョンを利用した人検知追跡システムについて言及するとともに，ステレオビジョンを用いた人検知および追跡手法の概要について述べる．第 3 章では本研究で提案した手法について述べると共に理論的検討を行う．第 4 章でシステム概要および提案手法の有効性確認実験，およびその実験結果を記し，第 5 章で実験結果に対する考察について述べる．第 6 章で本研究の総括を記す．

2. ステレオビジョンによる人検知および追跡

本章では，ステレオビジョンを利用した人検知および追跡手法の概要について述べる．特徴点抽出，ステレオ画像処理，クラスタリング，追跡と各処理毎にその手法について言及する．

2.1 従来の人検知および追跡手法

ステレオビジョンでは，人間の視覚と同様に2つのカメラで対象を撮像し，三角測量の原理から奥行き位置情報を取得する．単眼画像を用いる場合，画像の輝度情報から直接対象を検知することが多いのに対し，ステレオビジョンでは対象の3次元情報を用いるため，環境変動およびオクルージョンの影響が少なく，かつモデル化が画像上でおこなう場合と比べて容易である．ステレオビジョンを適用した人検知追跡実用化例として，Censys3D [8]がある．Censys3Dは，Digiclops Stereo Vision を用いた人検知，追跡システムである．計測地点の上方に3眼ステレオカメラユニットを配置し，それから得られる3次元情報により人の検知と追跡を行う．ステレオ処理なので，影の影響を受けにくく実際の3次元物体の変化に対応した計測ができる．しかしながら，カメラ設置位置が天井に制限されているため，設置自由度が小さい．また，追跡は x, z の2次元的な情報を用いて追跡を行っているので3次元的な動きを追跡できないという欠点がある．また，設置位置に制約をなくした場合の例として林ら [10]の手法がある．カメラの設置位置に自由度を持たせた人検知アルゴリズムであるが，人の重なりに改善の余地があり，かつ追跡に関しては触れられていない．次節では，本研究で提案したステレオビジョンによる人の検知および追跡手法の概要について述べる．図 13 にその処理フローを示す．

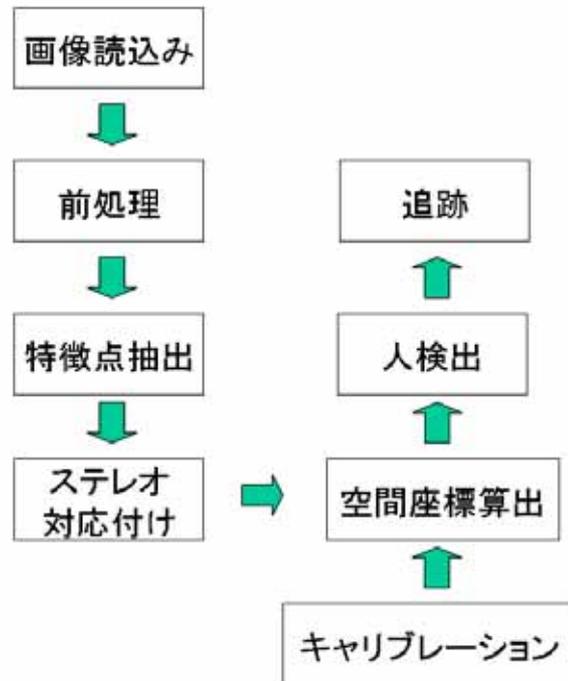


図 1 全体処理フロー

2.2 特徴点抽出

ステレオビジョンを用いて対象の3次元情報を算出する場合、2つのカメラ画素間の対応関係を求める（以後”対応づけ”と呼ぶ）必要がある。これには全画素の対応づけをおこなう場合と、あらかじめ設定したルールに基づいて粗くサンプリングした画素のみ対応づけをおこなう場合の二通りが考えられる。前者に基づく手法として、柴山ら [11] の提案手法がある。これは、ステレオビジョンで周期的に背景距離画像をあらかじめ導出し、背景差分により対象を抽出する。ハードウェア的に高速化しているため、専用のハードウェアが必要となり、汎用的であるとは言えない。本研究は後者に基づく方式であるため、3次元情報を算出したい点（特徴点）をあらかじめ抽出しておく必要がある。特徴点は、面と面の境界

であるエッジ上やコーナーなどの部分から抽出される場合が多い。

一般的な特徴点抽出手法として，ある画像上の点 p と別の画像上の点 q の残差平方和によるテンプレートマッチングで局所相関を取ることである。

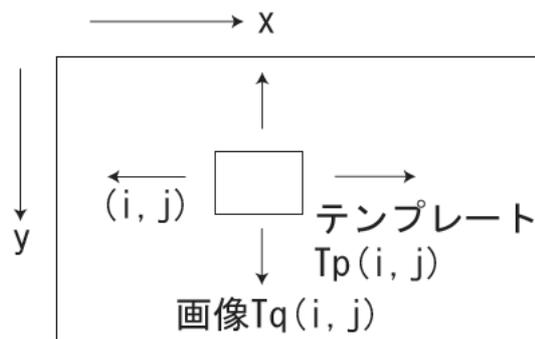


図 2 テンプレートマッチング

$$J(p, q) = \sum_{(i, j) \in N} |T_p(i, j) - T_q(i, j)|^2 \quad (1)$$

N はテンプレートの輝度値， T_p, T_q はそれぞれ点 p, q を中心とするテンプレートの輝度値を示す。これは，対応する画素間の色や明るさの違いの総和を求めるものであり，この値が小さければ対応する可能性が高い。しかし，点 p, q が一様な色や明るさ領域にあれば，テンプレートを移動しても残差平方和の値の変化が乏しく，対応を決定しにくい。よって，精度の良い対応を求めるにはその近傍で色や濃淡変化の激しい点を選ぶ必要となる [12]。このように点で高い値を与える関数を利用することで，特徴点の抽出を行う operator(作用素) を利用することが一般的である。

ステレオ画像処理で人を検知する際に必要な人らしい特徴として，画像中での人の輪郭が挙げられる。上記手法に基づき，画像上の輝度変化の大きいエッジ，コーナー点を抽出することで，人の輪郭を得る。

2.3 ステレオ画像処理技術

2.3.1 ステレオ画像処理原理

ステレオ画像処理は三角測量の原理に基づいた手法である。複数のカメラを用いて異なる視点から同一対象を見たとき、その画像は若干異なる。左右のカメラの相対的な位置、左右カメラの見ている対象の対応、および画像の違いを用いて距離を逆算し、対象の3次元情報を得るものである。

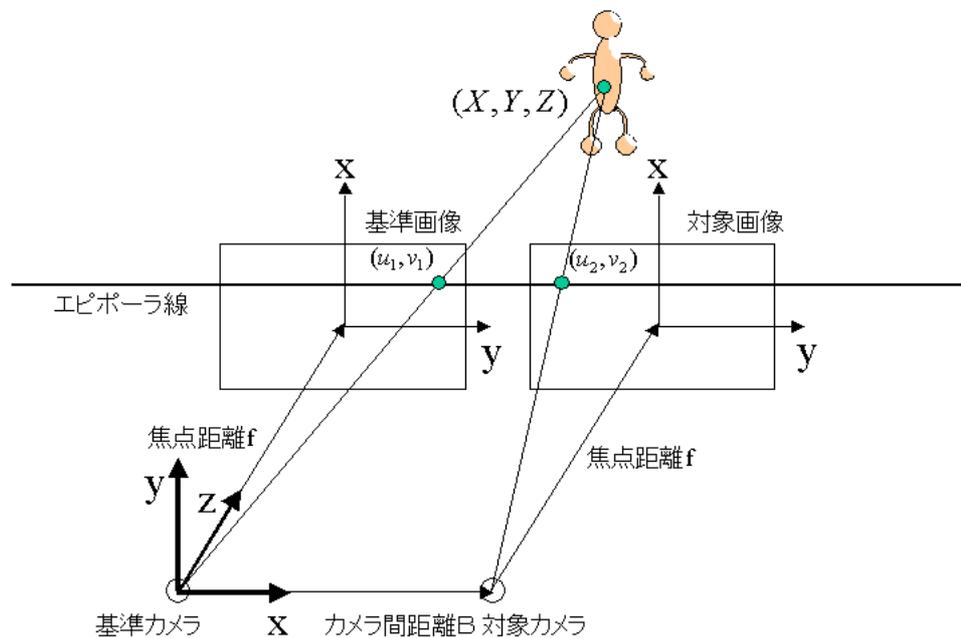


図3 ステレオ画像処理原理

その原理について記す。図3のように、空間の注目するワールド座標での位置 (X_c, Y_c, Z_c) が、それぞれ画像上の点 $(u_1, v_1), (u_2, v_2)$ に投影されたとすると、

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} \frac{u_1 B}{d} \\ \frac{v_1 B}{d} \\ \frac{f B}{d} \end{bmatrix} \quad (2)$$

ここで、 f は焦点距離、 $d = v_1 - v_2$ は視差、 B はカメラ間距離である。(2) は、 f, B が既知であれば、2つの画像上の投影座標から、もとの3次元位置が計算できることを意味する。ステレオ画像処理における重要な問題として、2つの画像における対応(視差)を精度よく算出することである。一方の画像上の点に対応する他方の画像上の点は、ある直線上に存在することが知られている。図3の場合、この直線は画像座標系で $y = v_1$ である。この直線をエピポーラ線と呼び、ステレオ画像点が満たすべき条件をエピポーラ拘束 [13], [14] とする。

本研究では、人検知および追跡を式(2)で示されるカメラ中心を原点とするカメラワールド座標系ではなく、カメラ中心の床面への垂線の足を原点とする地面ワールド座標系 (X_g, Y_g, Z_g) のもとでおこなっている。カメラワールド座標系と地面ワールド座標系との間には、以下の関係が成り立つ。

$$\begin{bmatrix} X_g \\ Y_g \\ Z_g \end{bmatrix} = \mathbf{R}_{cw} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} + \begin{bmatrix} 0 \\ H \\ 0 \end{bmatrix} \quad (3)$$

ただし \mathbf{R}_{cw} はカメラ俯角を含む回転行列、 H はカメラ高さである。 \mathbf{R}_{cw} 、 H に関しては、図1のキャリブレーションの工程でカメラの内部・外部パラメータと共に事前に算出しておく。

2.3.2 ステレオマッチング

2.3.1に基づいて、左右の画像の対応(視差)を算出する。本研究では、左画像を基準カメラ画像として、左画像上にウィンドウを設置し、各ウィンドウに対して次式の正規化相互相関を用いて右画像と濃淡値の相関を取ることで対応を算出する。

$$r_{cc} = \frac{\sum_{i=1}^N \sum_{j=1}^M (f(i, j) - \bar{f}(i, j))(g(i, j) - \bar{g}(i, j))}{\sqrt{\sum_{i=1}^N \sum_{j=1}^M (f(i, j) - \bar{f}(i, j))^2} \sqrt{\sum_{i=1}^{N-1} \sum_{j=1}^M (g(i, j) - \bar{g}(i, j))^2}} \quad (4)$$

ここで、ウィンドウサイズを N, M 、基準カメラ画像(左画像)および対象画像(右画像)の輝度列を $F = f(i, j)_{i=0, j=0}^{N, M}$ 、 $G = g(i, j)_{i=0, j=0}^{N, M}$ 、輝度値の平均を $\bar{f}(i, j)_{i=0, j=0}^{N, M}$ 、 $\bar{g}(i, j)_{i=0, j=0}^{N, M}$ とする。

2.4 クラスタリング手法

画像から特徴点を抽出し、ステレオ画像処理を行うことでそれらの3次元座標を導出した後に、特徴点間の類似性や非類似性をもとに分類する処理をクラスタリング、分類された群をクラスタと呼ぶ。ここで類似性、非類似性の指標は主に距離が用いられる。用いられる距離として、 n 個の特徴点について、 p 個の変量を持つとする。2つの特徴点 $\mathbf{X}_i^T = (x_{1i}, x_{2i}, \dots, x_{pi})$, $\mathbf{X}_j^T = (x_{1j}, x_{2j}, \dots, x_{pj})$ 間の距離を $D_{ij}, i, j = 1, 2, \dots, n$ としたとき

- ユークリッド距離

$$D(\mathbf{X}_i, \mathbf{X}_j) = \left\{ \sum_{k=1}^p (x_{ki} - x_{kj})^2 \right\}^{1/2} \quad (5)$$

- 絶対距離

市街地距離と呼ばれ、各変量毎に1変量でのユークリッド距離を計算しその総和を求める

$$D(\mathbf{X}_i, \mathbf{X}_j) = \sum_{k=1}^p |x_{ki} - x_{kj}| \quad (6)$$

- マハラノビス距離

分散と共分散を考慮した距離で

$$D(\mathbf{X}_i, \mathbf{X}_j) = (\mathbf{X}_i - \mathbf{X}_j)^T \Sigma^{-1} (\mathbf{X}_i - \mathbf{X}_j) \quad (7)$$

と表す。ここで Σ は p 変量の共分散行列である。

これらの距離が、類似性、非類似性の指標として主に用いられる。クラスタリング手法にはクラスタが体系化される階層型分類と体系化されずいくつかの最適な分類を行う非階層型分類がある。以下でこれら手法について説明する。

2.4.1 階層型分類

階層型の分類には、個々のデータをそれぞれ1つのクラスタとみなしてボトムアップ式に階層化を行う手法が一般的である。抽出した全特徴点に対し、上述し

た距離尺度に基づいて非類似性を求め、非類似性の小さい特徴点同士を融合してクラスタを形成する。代表的な階層型クラスタリング手法として以下のものがある [15], [16]。

クラスタ (a) とクラスタ (b) が最も距離が小さく、これらを融合してクラスタ (c) ができるとする。クラスタ (c) と任意のクラスタ (i) との距離を D_{ci} とする。クラスタ (k) は n_k 個の特徴点 $X_j^k = (x_{1j}^{(k)}, x_{2j}^{(k)}, \dots, x_{pj}^{(k)})$ で構成され、その重心は $\bar{X}^{(k)} = (\bar{x}_1^{(k)}, \bar{x}_2^{(k)}, \dots, \bar{x}_p^{(k)})$, ($k = a, b, c, i$), ($j = 1, 2, \dots, n_k$) であるとする。

- 最短距離法

2つのクラスタ間の距離を各クラスタに含まれる特徴点間の最短距離とする。このとき、クラスタ (c) とクラスタ (i) の距離 D_{ci} は

$$D_{ci} = \min(D_{ai}, D_{bi}) = \frac{1}{2}D_{ai} + \frac{1}{2}D_{bi} - \frac{1}{2}|D_{ai} + D_{bi}| \quad (8)$$

となる。この手法は、最も近い1つの特徴点を含むクラスタを逐次融合していくので、線状クラスタの検出に優れている。融合が行われる毎に距離が近づくため、空間が収縮する。

- 最長距離法

2つのクラスタ間の距離を各クラスタに含まれる特徴点間の最長の距離と定義する。このとき、(c) クラスタと (i) クラスタの距離 D_{ci} は

$$D_{ci} = \max(D_{ai}, D_{bi}) = \frac{1}{2}D_{ai} + \frac{1}{2}D_{bi} + \frac{1}{2}|D_{ai} + D_{bi}| \quad (9)$$

この手法は、最短距離法とは逆に融合が行われていないクラスタ (i) は融合が行われたクラスタ (c) から離れるため、空間が拡散される。よって分類の感度は高まる。また、クラスタ構成が密になり、クラスタは似たような大きさになる傾向がある。

- 群平均法

2つのクラスタ間の距離を各クラスタに含まれる特徴点間の全ての組合わせの距離を求め、その平均値をクラスタ間の距離と定義する。このとき、クラスタ (c) とクラスタ (i) の距離 D_{ci} は

$$D_{ci} = \frac{n_a D_{ai} + n_b D_{bi}}{n_a + n_b} = \frac{n_a}{n_c} D_{ai} + \frac{n_b}{n_c} D_{bi} \quad (10)$$

で表す．この手法は，空間の収縮や拡散が少なく，球状のクラスタを作る傾向がある．

- 重心法

各クラスタに含まれる特徴点の重心を求め，重心をクラスタの代表点とし，重心間の距離をクラスタ間の距離と定義する．このとき，クラスタ (c) とクラスタ (i) の距離 D_{ci} は

$$D_{ci} = \sum_{j=1}^p \left(\frac{n_a \bar{x}_j^a + n_b \bar{x}_j^b}{n_a + n_b} - \bar{x}_j^{(i)} \right)^2 = \frac{n_a}{n_c} D_{ai} + \frac{n_b}{n_c} D_{bi} - \frac{n_a n_b}{(n_c)^2} D_{ab} \quad (11)$$

で表せる．

- ウォード法

クラスタ内の偏差平方和を計算し，融合したクラスタの全体の偏差平方和の増加を最小にする2つのクラスタを融合する．すなわち，クラスタ (c) とクラスタ (i) が融合することによる偏差平方和の増加

$$\Delta S_{ci} = \frac{n_c n_i}{n_c + n_i} \sum_{i=1}^p (\bar{x}_i^{(c)} - \bar{x}_i^{(i)})^2 = \frac{1}{n_c + n_i} \{ (n_a + n_i) \Delta S_{ai} + (n_b + n_i) \Delta S_{bi} - n_i \Delta S_{ab} \} \quad (12)$$

を最小にするようにクラスタを融合することである．よって，クラスタ (c) とクラスタ (i) の距離 D_{ci} は

$$D_{ci} = \frac{1}{n_c + n_i} \{ (n_a + n_i) D_{ai} + (n_b + n_i) D_{bi} - n_i \Delta S_{ab} \} \quad (13)$$

と表せる．

2.4.2 非階層型分類

非階層的クラスタリングとは，初期状態として適当なクラスタ中心を設定し，そのクラスタ内のメンバを組み替えて少しずつよりよいクラスタを求めていく手法である．一般的に非階層型クラスタリングは，クラスタ数が既知であるか各クラスタ数にある程度まとまりがあるという条件が必要になる．よって，クラスタ数が未知のものに対して適用しようとした場合には結果に恣意性が入ってしまい，

汎用的であるとは言えない。

非階層型クラスタリング手法の代表的なものとして、K平均法、ISODATA法がある [17]。K平均法は、特徴点をクラスタリングする際に、事前にクラスタ数、および初期クラスタの中心値を与える必要があり、それらの設定によって得られるクラスタ形状が変わってくる。ISODATA法は、設定したクラスタ数の $\frac{1}{2}$ ~2倍の範囲で自律的に有効なクラスタ数を決めることができるが、初期クラスタ設定だけでなく、事前に調整しなければならないパラメータが多数存在する。

以上いくつかの代表的なクラスタリング手法について述べてきたが、本研究ではこの中から最長距離法を利用する。主な理由としては、(1) 人の輪郭から抽出された特徴点群は、3次元空間上で似たような広がりをもつこと (2) 特徴点群の分類感度は高いことが望まれる (一人一人の検知を確実におこないたい) ことが挙げられる。次節で最長距離法に基づく具体的なクラスタリング方法について述べる。

2.4.3 ステレオビジョンによる人検知 [7],[9]

本研究では、画像から抽出された特徴点群をステレオ画像処理を行うことで3次元空間に投影した後、それらをクラスタ数未知の条件下でクラスタリングすることで人領域を抽出する手法を利用している。クラスタリングの際には、3次元空間中での人特徴点の分布は、ある大きさの円錐内に分布すると仮定する。

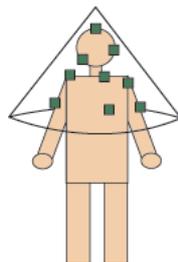


図 4 人物輪郭モデル

これは、人は歩行中、静止中の際には、手が肩を支点として下方に下がった状態にあり、図4に示すように、人間の上半身は円錐内に分布する。このモデルを前述したユークリッド距離を使った最長距離法の尺度に適用し、

$$D(i, j) = \sqrt{((x_i - x_j)^2 + (z_i - z_j)^2)} + \lambda \{ \min(y_i, y_j) - offset \} \quad (14)$$

とすることで、人の分解能性能を向上させている。ここで、 $offset$ は1[m]とし、 λ は重み係数で、 X, Z 平面上での一人の胴体部特徴点間の最大距離を L_{body} 、二人の頭部特徴点間の最大距離を L_{head} 、人間の平均身長を H とすると

$$\lambda = \frac{L_{body} - L_{head}}{H - offset} \quad (15)$$

で求められる。

2.5 追跡手法

追跡処理では、検知した人領域を時系列に3次元空間内で追跡する。計算コスト等により、導出した3次元空間情報をそのまま使わず、情報を落して平面的な追跡を行う。床面を X, Z 軸、高さを Y 軸としたとき、*censys3D* のようにカメラ設置位置が天井等の人の頭部位置よりも情報に場合は、高さ情報 (Y 座標) 用いずに平面的 (XZ 平面) な追跡をおこなう。森ら [18] は、多視点で得られた観測情報を統合して Kalman Filter を用いて平面的 (XZ 平面) な追跡をおこなっている。文献 [7],[9] では、前述のクラスタリングをにより得た人領域を xz 平面で追跡する。追跡を行う際の追跡注目点としては

- 最大値
- 最小値
- 重心値

等が挙げられるが、ノイズによる影響を軽減でき、安定した追跡をすることが可能であることから重心値を採用している。

追跡は過去のクラスタ位置情報から現フレームでの位置を，人の移動は局所的に等速直線運動であるという仮定の下で以下のように

$$x_{i+1} = x_i + r(x_i - x_{i-1}) \quad (16)$$

$$z_{i+1} = z_i + r(z_i - z_{i-1}) \quad (17)$$

クラスタ移動量から推測する．ここで， r はクラスタ位置の予測割合である．

3. 人検知および追跡のロバスト性の向上

2章では，ステレオビジョンを利用した人検知および追跡手法の概要について言及したが，ここでは，本研究で提案した手法の詳細について説明する．提案手法は，人検知および追跡をロバストに行う上で重要な観点，

- 人の輪郭を形成する特徴点を高精度に抽出
- 姿勢変化に依存しない人検知
- 空間的な人追跡

を実現する．上記観点を実現するアルゴリズムの流れについて簡単に説明する．まず高精度に人輪郭特徴点を抽出し，ステレオ対応付けにより視差を算出することで人の空間座標を求める．求めた空間座標から人検出を行い，追跡を行う（ボトムアップ追跡）．この際，追跡のモデリングにおいて人検出に必要な高さ情報を加えることで，3次元的な追跡を実現し，その結果を一期先クラスタリング情報にフィードバックする（トップダウン追跡）．以上の繰り返し処理により，ロバストな人検知および追跡を実現する．これら処理について説明すると共に，その理論的検証を行う．

3.1 Harris Operator を利用した特徴点抽出

ここでは，Harris Operator を用いて高精度に人輪郭特徴点を抽出する方法について述べる．本研究で利用する Harris Operator [19] とは，画像信号の相関性に基づいて特徴点を導出する手法で，エッジ，コーナー等の特徴点において相関出力値が大きくなる operator である．Schmid ら [20] によると，この operator は回転，拡大などの画像変形が発生した場合にも，同一の点に特徴点が検出される再現性が高いとされている．

入力画像上の座標 (x, y) における輝度値を $I(x, y)$ ，輝度勾配を $\nabla I(x, y) = \left(\frac{\partial I(x, y)}{\partial x}, \frac{\partial I(x, y)}{\partial y} \right)$ ，座標 (x, y) を中心とした幅 $r \geq \|(x, y) - (x_0, y_0)\|$ の正方形窓を $N(x_0, y_0, r)$ ，窓の移動量を $(\Delta x, \Delta y)$ とする．このとき，座標 (x, y) を中心とした窓 $N(x_0, y_0, r)$ と，

$(x + \Delta x, y + \Delta y)$ を中心とした $N(x_0 + \Delta x, y_0 + \Delta y, r)$ の最小残差平方和 $f(N_0)$ を求める .

$$f(N_0) = \min \sum_{(x,y) \in N_0} (I(x, y) - I(x + \Delta x, y + \Delta y)) \quad (18)$$

$I(x + \Delta x, y + \Delta y)$ は Taylor 展開を利用することにより

$$I(x + \Delta x, y + \Delta y) \approx I(x, y) + \nabla I(x, y)^T \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} \quad (19)$$

と近似が可能である . よって式 (18) は ,

$$f(N_0) \approx \begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix} A [\Delta x, \Delta y] \quad (20)$$

と近似ができる . ここで $\begin{bmatrix} \Delta X \\ \Delta Y \end{bmatrix}$, $[\Delta x, \Delta y]$ は一定値であるので , 正定値行列

$$A = \sum_{(x,y) \in N_0} \nabla I(x, y)^T \nabla I(x, y) = \sum_{(x,y) \in N_0} \begin{bmatrix} \frac{\partial I(x,y)}{\partial x} \\ \frac{\partial I(x,y)}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial I(x,y)}{\partial x} & \frac{\partial I(x,y)}{\partial y} \end{bmatrix} \quad (21)$$

が r 近傍の残差に相当する . Harris Operator の応答関数 R は , この正定値行列 A の固有値 λ_1, λ_2 から

$$R(x, y) = \det A - \kappa (\text{tr} A)^2 \quad (22)$$

で定義される . このとき , $\det A = \lambda_1 \lambda_2 = A(1,1)A(2,2) - A(1,2)^2$, $\text{tr} A = \lambda_1 + \lambda_2 = A(1,1) + A(2,2)$ で κ は曲率である .

3.1.1 応答関数出力値 R と特徴点の関係

正定値行列 A の固有値 λ_1, λ_2 の値によって窓 N_0 内の特性が変わってくる .

図5では , 窓内の濃淡変化がない場合で , このときは行列 A の固有値 λ_1, λ_2 が 0 になることは明らかである . 図6に示すように , エッジなど勾配方向が一様である場合は , 特異行列であるため行列 A の固有値 λ_1, λ_2 のどちらかが 0 になる . また図7,8にあるように , 勾配方向が交差する場合 (コーナー) においては , 行列

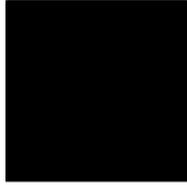


図 5 濃淡変化がない場合



図 6 勾配方向が一様な場合 (エッジ)

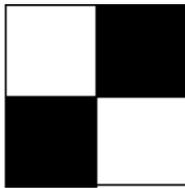


図 7 勾配が交わる場合 1

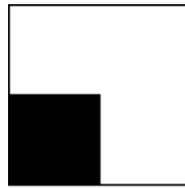


図 8 勾配が交わる場合 2

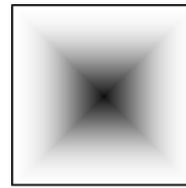


図 9 勾配が交わる場合 3

A はフルランクの正則行列であるので，固有値 λ_1, λ_2 とも 0 より大きくなる．図 9 の特異点の場合も同様のことが言える．

特徴点の性質と，固有値の関係を (22) に適用することによって，応答関数 $R(x, y)$ の出力値に応じて特徴点の種類を特定することができる．すなわち，応答関数 $R(x, y)$ と特徴点の間に

- $R(x, y) \cong 0$
窓内に特徴点が存在しない
- $R(x, y) < 0$
窓内にエッジが存在
- $R(x, y) > 0$
窓内にコーナーが存在

が成立する．応答関数の出力値に対する任意の閾値を設定することで，必要とする特徴点の抽出が可能となる．

3.1.2 人輪郭特徴点の抽出

しかしながら応答関数の出力値特性だけでは、必要としている人の輪郭を形成する特徴点を抽出することは困難である。画像中での人特徴点は輝度変化が大きくなる。よって、近傍での色や濃淡変化が激しい場合は応答関数 R の出力値が大きく(小さく)なることに着目し、

$$\text{コーナー} : R_n(x_n, y_n) > R_{n-1}(x_{n-1}, y_{n-1}) > \cdots > R_1(x_1, y_1) > T_c \gg 0 \quad (23)$$

$$\text{エッジ} : R_{n-1}(x_n, y_n) < R_{n-1}(x_{n-1}, y_{n-1}) < \cdots < R_1(x_1, y_1) < T_e \ll 0 \quad (24)$$

なる特徴点を抽出する。濃淡変化の激しいものから順に抽出することになり、これらを人物特徴点候補とする。ここで T_c, T_e は応答関数閾値、 n は抽出する特徴点番号である。さらに計算コスト削減も考慮し、本研究では幅 D の正方形窓 W_D 内で

$$\text{コーナー} : (x_f, y_f) = \arg \max_{(x_i, y_i) \in W_D} R(x_i, y_i) \quad (25)$$

$$\text{エッジ} : (x_f, y_f) = \arg \min_{(x_i, y_i) \in W_D} R(x_i, y_i) \quad (26)$$

なる (x_f, y_f) を最終的な特徴点として抽出する。図 10 に対して、上述した手法に基づき算出された特徴量を輝度値に変換した画像を図 11 に示す。

3.1.3 本手法の特徴

Harris Operator の応答関数出力値に着目して、特徴点の性質を分離し、人の輪郭を形成する特徴点を高精度に抽出可能とする。以下に本手法の特徴について列挙する。

- 特徴
 - Harris Operator 応答関数の出力値特性に基づいて特徴点抽出を行う
 - 特徴点間の距離尺度を組み合わせることにより、抽出点にばらつきをもたせ、できるだけ少なくかつ人を検知するのに十分な特徴点抽出が可能



図 10 入力画像



図 11 特徴量を輝度値に変換した画像

3.2 姿勢変化に対応した人の検知

2.4.3 節で述べた人検知技術 [7], [9] では, 人の輪郭形状モデル (図 4 参照) を用いることで, 人の検知をおこなう. この際, 直立状態にある人は, 高さ (Y 座標) 1 ~ 2[m] の間に上半身が入ると仮定し, この間にある特徴点群を人の候補とする. 式 (14) で表される, 高さ情報を利用した最長距離法に基づくクラスタリングにより人の検知をおこなう.

この手法は, 図 12 のように人物輪郭モデルの高さ閾値 $TH = 2.0, TL = 1.0$ をあらかじめ設定し, この領域内にある特徴点を人候補としている.

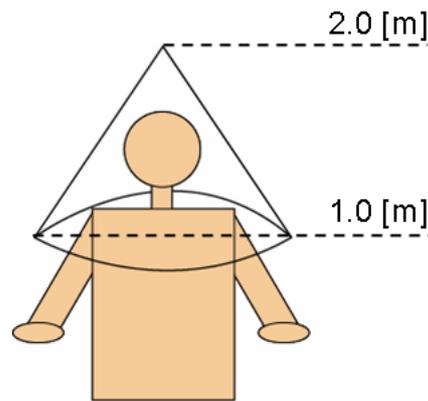


図 12 人高さ閾値

しかしながら, 座った場合, 倒れた場合など, 高さ (Y 軸) 方向の移動が発生したときに柔軟に対処できない. そこで本研究では, 高さ情報を加えた追跡を行い, その結果を一期先の高さ (Y 軸) 推測値としてフィードバックすることで, クラスタ毎に高さ閾値を動的に設定する方法を提案する. クラスタリングは動的に更新された閾値を用いておこなわれる. 本提案手法は, 追跡結果を検知処理にフィードバックさせることにより, 高精度な人検知を実現し, 検知結果を追跡に利用することでより安定した追跡をおこなえるのが特徴である. 処理のフローを図 13 に示す.



図 13 検知追跡処理フロー

3.2.1 頭部位置検出フィルタ

人に高さ (Y 軸) 方向の変化 (座る, 倒れる等) が生じた場合, 変動量が顕著に現れる箇所として, 頭部が挙げられる. したがって人のクラスタから頭部らしき位置が検出できれば, その軌跡から人に高さ (Y 軸) 方向の変化を追従できると考えられる.

従来の頭部位置の検出手法としては色情報を用いるものとして, 宮脇ら [21] は HSV 色相空間を利用して画像中から肌色領域を検出することで顔領域を抽出し, Y.Dai ら [22] は濃淡画像のテクスチャ情報を YIQ 基底で表現されたカラー画像の I 成分に適用して顔領域を検出している. しかしながら, 色情報は背景シーンの色彩成分の特性に大きく左右される.

本研究では, ステレオビジョンにより得た人特徴点群の 3D 情報のみを利用して頭部近辺位置の検出を行う. 人の頭部位置は, 座る・倒れる等の姿勢変動が発生した場合にも, 基本的には人の中心 (重心) 位置よりも上部に位置する. この点に着目し, 人クラスタ h_i 内の 3D 特徴点群の高さ $y_{h_i,j}$ 値に対して

$$y'_{h_i,k} = w_j y_{h_i,k}, (k = 1, 2, 3, 4), (k = 1, 2, \dots, n) \quad (27)$$

と重み付けした位置を重心値とは別に保持する. このとき重み w_j は, 図 14 のように設定する. 直立状態にある人の頭部位置が, 輪郭モデル内の高さ (Y 軸) 方向

に上から $1/2 \sim 2/3$ の間にあると仮定する．この位置にある w_j の重みを最大にすることで，頭部位置近辺にある特徴点の高さ (y) 座標値が検出される．

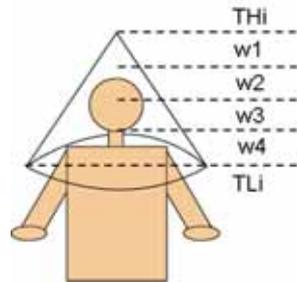


図 14 高さ方向重みの設定

そして，重み付け高さ $\bar{y}'_{h_i,j}$ と特徴点の高さ $y_{h_i,j}$ の重心 $\bar{y}'_{h_i}, \bar{y}_{h_i}$ を以下のように求める．

$$\bar{y}'_{h_i} = \frac{1}{n} \sum_{k=1}^n y'_{h_i,k} \quad (28)$$

$$\bar{y}_{h_i} = \frac{1}{n} \sum_{k=1}^n y_{h_i,k} \quad (29)$$

3.2.2 3次元追跡モデルを利用した時空間追跡

追跡処理では，検知処理部から送られてきた観測情報 (人の位置，時刻) を基に，状態量の最適な推定値を計算する．この計算には Kalman Filter を用いる [23], [24],[25]．Kalman Filter は観測信号を入力する線形ダイナミカルシステムであり，コンピュータにより容易に表現ができる．人の動きは局所的 (微小時刻間) に見た場合，等速直線運動とみなせる．したがって本手法の処理時間オーダー (数百ミリ秒) では，人の追跡は線形モデルで十分適用可能であると考えられる．

時刻 t における人 h_i の状態は， $\mathbf{X}_{h_i,t} = (x_{h_i}, y_{h_i}, \bar{y}'_{h_i}, z_{h_i}, \dot{x}_{h_i,t}, \dot{y}_{h_i,t}, \dot{y}'_{h_i,t}, \dot{z}_{h_i,t})$ の 8次元ベクトルで構成される．ここで， $x_{h_i}, y_{h_i}, z_{h_i}$ は，人クラスタの x, y, z の重心点を， \bar{y}'_{h_i} は，人クラスタ内特徴点群の高さ y_{h_i} 値に対して重み付けをし導出し

た重心値である．また， $\dot{x}_{h_i,t}, \dot{y}_{h_i,t}, \dot{y}_{h_i,t}, \dot{z}_{h_i,t}$ は，それぞれ時刻 t における X, Y, Z 軸方向への速度を表す．

人 h_i の時刻 $t+1$ の状態は，時刻 t における状態 $\mathbf{X}_{h_i,t}$ から推定することができる．これを状態モデルとし，

$$\mathbf{X}_{h_i,t+1} = \mathbf{F}\mathbf{X}_{h_i,t} + \mathbf{w}_t \quad (30)$$

で表す．ここで， \mathbf{w}_t は系に加わる 8 次元の外乱ベクトルで，平均が 0, 共分散が \mathbf{Q}_t とする．また， \mathbf{F} は

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & \Delta T & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \Delta T & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \Delta T & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \Delta T \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (31)$$

の 8×8 行列で， Δt はフレームレートの逆数である．

時刻 t における人 h_i の状態予測値 $\hat{\mathbf{X}}_{h_i,t+1}$ が，平均 $\hat{x}_{h_i,t}$, 共分散 $\hat{\mathbf{P}}_{h_i,t}$ のガウス分布で表現されるとすると，この分布に基づく時刻 $t+1$ における人 h_i の状態の予測分布平均 $\bar{\mathbf{X}}_{h_i,t+1}$, 共分散 $\bar{\mathbf{P}}_{h_i,t+1}$ は

$$\bar{\mathbf{X}}_{h_i,t+1} = \mathbf{F}\hat{\mathbf{X}}_{h_i,t} \quad (32)$$

$$\bar{\mathbf{P}}_{h_i,t+1} = \mathbf{F}\hat{\mathbf{P}}_{h_i,t}\mathbf{F}^T + \mathbf{Q}_t \quad (33)$$

で与えられる．

次に，時刻 t において人 h_i の重心位置が観測されたとし，観測点を $\mathbf{Z}_{h_j,t}$ とする．このときの観測モデルは

$$\mathbf{Z}_{h_j,t} = \mathbf{H}\mathbf{X}_{h_i,t} + \mathbf{v}_t \quad (34)$$

で表せる．ここで， v_t は 4 次元の測定誤差ベクトルで，平均が 0，共分散が R_t で既知である．また， H は

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (35)$$

の 4×8 行列である．

以上から，Kalman Filter を構成し，時刻 t における人 h_i の状態予測分布の平均 $\hat{X}_{h_i,t}$ ，共分散 $\hat{P}_{h_i,t}$

$$\hat{X}_{h_i,t} = \bar{X}_{h_i,t} + K_{h_i,t}(Z_{h_i,t} - H\bar{X}_{h_i,t}) \quad (36)$$

$$K_{h_i,t} = \hat{P}_{h_i,t} H R_t^{-1} \quad (37)$$

$$\hat{P}_{h_i,t} = (\bar{P}_{h_i,t}^{-1} + H^T R_t^{-1} H)^{-1} \quad (38)$$

を得る．Kalman Filter の追跡処理フローを図 15 に示す．ここで，式 (37) は Kalman Gain であり，状態 $X_{h_i,t}$ に関する情報の相対的な重みを表す．なお，初期値は $\hat{X}_0 = X_0, \hat{P}_0 = P_0$ とする．

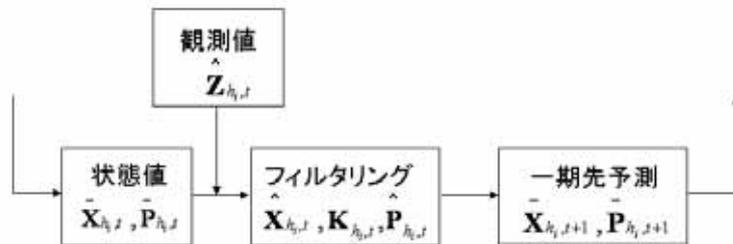


図 15 Kalman Filter フロー

3.2.3 クラスタ毎の高さ閾値の更新

時刻 t における人クラスタ h_i の高さ (Y 座標) 重心 $\bar{y}_{h_i,t}$ と, 3.2.1 節において言及した重み付けをした高さ (Y 座標) 重心 $\bar{y}'_{h_i,t}$ を, 3.2.2 節により求めた一期先位置の推定値を用いてクラスタ毎の高さ閾値を更新する.

頭部近辺にある特徴点の値が大きくなるようなフィルターをかけることにより, 重み付けをした高さ (Y 座標) 重心 $\bar{y}'_{h_i,t}$ は, 座る, 倒れる等の姿勢変動が起った場合により顕著に変化が表れる. $\bar{y}'_{h_i,t}$ と $\bar{y}_{h_i,t}$ の差 $diff$ は

$$diff = \bar{y}'_{h_i,t} - \bar{y}_{h_i,t} = \frac{1}{n} \sum_{i=1}^n (w_j - 1)y_i \quad (39)$$

となる. これから

- 直立状態の場合は $diff$ の値が大
- 座る, 倒れる等の姿勢変動が起った場合は $diff$ の値が小さい

と言える (p.26 の図 16 参照). 直立状態では, 人物輪郭モデルの頭部近辺の特徴点が多くなるため, $diff$ の値は大きくなり, 座る, 倒れる等の状態では頭部近辺の特徴点が少なくなり, $diff$ の値は小さくなる. 図 14(p.22) の頭部近辺の重みを $w_j = 2, (j = 2, 3)$ とすれば, 頭部付近の特徴点だけを求めることができる. すなわち, $diff$ の値は頭部付近にある特徴点の平均とみなしてよい.

この $diff$ の値に着目して, 図 12(p.20) の輪郭モデル閾値 TH, TL をクラスタ毎に動的に更新していく. 時刻 t における人 h_i の重み付け高さ (Y 座標) 重心 $\bar{y}'_{h_i,t}$ と高さ (Y 座標) 重心 $\bar{y}_{h_i,t}$, 追跡処理において導出した一期先重み付け高さ重心 $\bar{y}'_{h_i,t+1}$, 高さ重心 $\bar{y}_{h_i,t}$ から人クラスタ毎の高さ閾値を逐次的に更新していく. 時刻 t における $diff_t$ と時刻 $t+1$ における $diff_{t+1}$ の差分を取ることで, 頭部位置の変動量に応じて以下のように閾値を更新する.

$$TH_{h_i,t+1} = TH_{h_i,t} + (diff_{t+1} - diff_t) \quad (40)$$

$$TL_{h_i,t+1} = TL_{h_i,t} + (diff_{t+1} - diff_t) \quad (41)$$

人 h_i の時刻 t における閾値 $TH_{h_i,t}, TL_{h_i,t}$ に対し, バイアス演算を行うことで時刻 $t+1$ の閾値 $TH_{h_i,t+1}, TL_{h_i,t+1}$ を決定する.

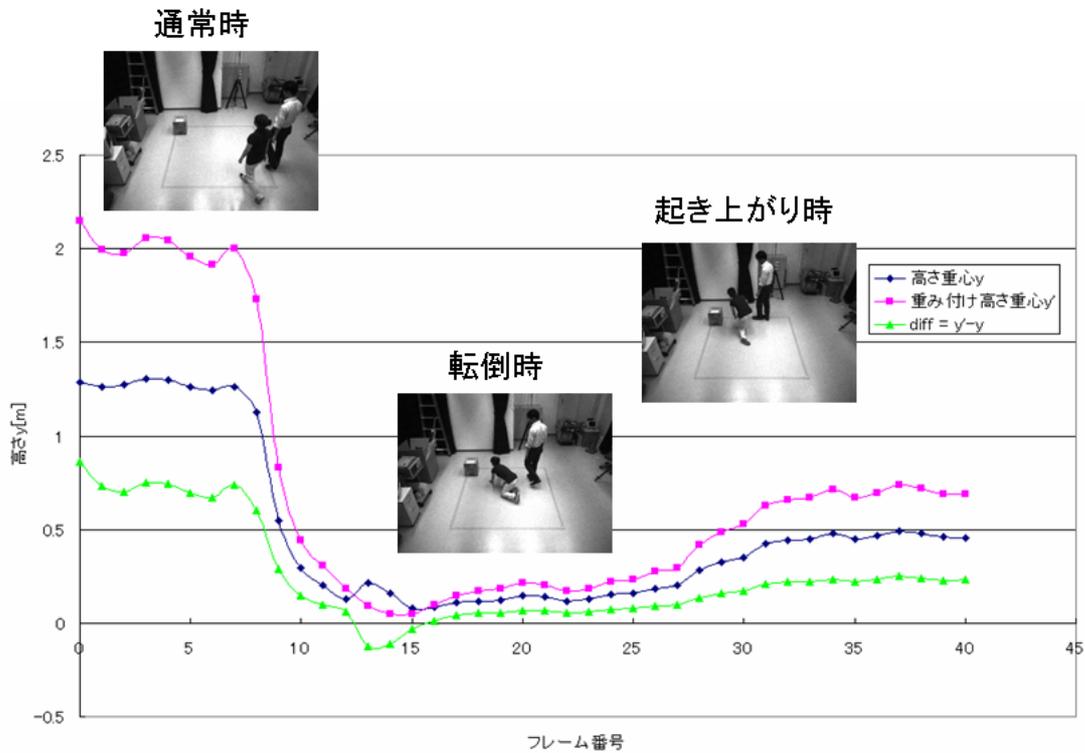


図 16 姿勢変化に伴う高さ位置の変動

3.2.4 姿勢変化が発生した場合の人検知

3.2.3 節において姿勢変化が発生した場合は，その変動量に応じて人クラス h_i 毎に輪郭モデル高さ閾値が更新される．時刻 $t + 1$ における，人 h_i の高さ閾値 $TH_{h_i,t+1}, TL_{h_i,t+1}$ である特徴点群を人物候補とし，これらの特徴点群に対して式 (14) を適用する．人物候補特徴点群が人クラス毎に更新された高さ閾値に応じて決定するので，姿勢変化が発生した場合でも柔軟に対処できる．なお，新規で表れた人 h_j に対しての高さ閾値は，本論文では $TH_{h_j,0} = 2.0, TL_{h_j,0} = 1.0$ とする．

3.2.5 本手法の特徴

本手法は、高さ情報を加えた3次元追跡モデルにより3次元的な動作の追跡を行える。また、追跡した結果を一期先の人検知処理に利用することで、姿勢変化等にも対応したロバストな人検知を行う。以下に本手法の特徴について列挙する。

- 特徴

- 色情報を利用せず、3次元空間情報で人の頭部近辺の高さ座標を導出している
- コンピュータで表現しやすい、線形ダイナミカルシステムを人物追跡に利用している
- 追跡処理により得た推定値を用いて、人クラスタ毎に一期先の人物輪郭モデルの高さ閾値を更新して
- クラスタ毎に人物輪郭モデルの高さ閾値を更新することで、姿勢変化起った場合にも対応している

- 長所

- 輝度情報、色情報を利用していないので、照明変動等の環境変化が起った場合でも安定した頭部近辺位置を導出できる
- 逐次的に検知追跡を行うので、データを全て蓄えておく必要がない
- 人の姿勢変化に対してロバストである

4. 実験

この章では、3章で提案した手法自身の性能を検証すると共に、従来手法 ([9]) との性能の比較実験を行う。3.2節にて提案した手法に対する性能検証を中心にを行う。

4.1 概要

提案手法の性能検証として、複数人が通常歩行を行った場合、人が倒れた場合、および人が飛びはねる動きをした場合における3次元的动作の位置追跡精度、および処理時間の検証を行う。ここで位置追跡精度 TR を、

$$TR = \frac{1}{N} \sum_{n=1}^N \left| \sqrt{x'_{t+1} \cdot x'_{t+1} + y'_{t+1} \cdot y'_{t+1} + z'_{t+1} \cdot z'_{t+1}} - \sqrt{x_t \cdot x_t + y_t \cdot y_t + z_t \cdot z_t} \right| \quad (42)$$

とする。 $x'_{t+1}, y'_{t+1}, z'_{t+1}$ は時刻 t における一期先の推定量を表す。追跡軌跡は、 x, y, z 座標の各重心値の軌跡である。次に、3.1節で提案した手法の初期パラメータを示す。

正方形窓サイズ	5×5 [pixel]
κ	0.04
抽出特徴点数	1000
エッジ抽出用応答関数閾値	-1500000
コーナー抽出用応答関数閾値	6000000
特徴点間距離	3×3 [pixel]
標準偏差	0.7[m]

表 1 特徴点抽出パラメータ

ここで、標準偏差は Gaussian による入力画像平滑化に用いる。

3.2 節で提案した手法の初期パラメータは以下の通りである。

高さ閾値 TH	2.0[m]
高さ閾値 TL	1.0[m]

表 2 人検知パラメータ

重み w_j は, 人物輪郭モデルの初期閾値を 2.0(TH) ~ 1.75, 1.75 ~ 1.5, 1.5 ~ 1.25, 1.25 ~ 1.0(TL) に 4 分割し, それぞれ $w_1 = 1.0, w_2 = 2.0, w_3 = 1.0, w_4 = 1.0$ として重みを割り当てた。

また, 本手法の有効性を確認する意味で, 従来手法 ([9]) との比較実験を行う。p12,13 の手法と, 追跡精度, 処理時間を中心に比較検証する。

4.2 実験環境およびシステム構成

観測空間は屋内空間に限らないが, 適当な広さの会議室程度の部屋を想定した。提案手法では天井等に特別なカメラ等を設置する必要がないので, 市販のカメラを使用している。ベース長が $B = 20[\text{cm}]$, レンズ焦点距離が $f = 4[\text{mm}]$ としている。以下表 4.2 にシステム構成について記す。

観測空間広さ	10×10×3[m][m]
処理エリア	4×4×3[m]
カメラの数	2 台
カメラの高さ	2.8[m]
カメラ解像度	1024×768
カメラ俯角	50[deg]
CPU	Intel Pentium 1.0 GHz
Memory	512 MB
OS	Windows 2000

表 3 実験環境およびシステムのスペック

4.3 実験結果

4.3.1 通常歩行時

本実験で使用した通常歩行画像の一部を図 17 に示す。



図 17 通常歩行時の画像

通常歩行時の位置追跡精度および処理時間結果を以下の表 4 に示す。なお、特徴点抽出処理に関しては従来手法においても 3.1 節で言及した手法を用いている。また、提案手法と従来手法の xy, yz, zx 各平面における追跡軌跡を図 18 ~ 23 に示す。

\	提案手法	従来手法
位置追跡精度	0.051[m]	0.084[m]
特徴点抽出処理時間	133.2[msec]	同左
クラスタリング処理時間	52.68[msec]	45.42
追跡処理時間	0.680[msec]	3.983×10^{-3} [msec]
その他処理 (前処理, 空間座標算出等)	135.0[msec]	同左

表 4 通常歩行時の位置追跡精度および処理時間

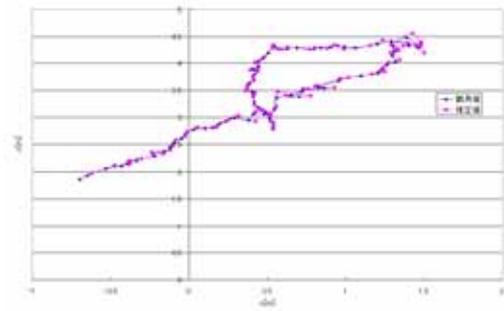
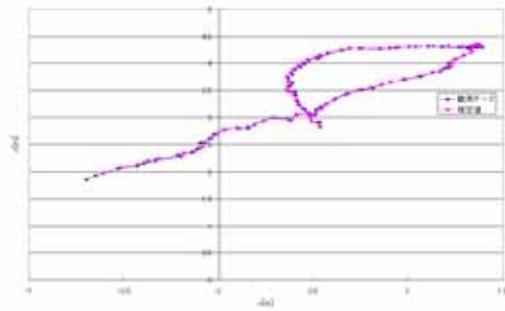


図 18 xz 平面から見た軌跡 (提案手法) 図 19 xz 平面から見た軌跡 (従来手法)

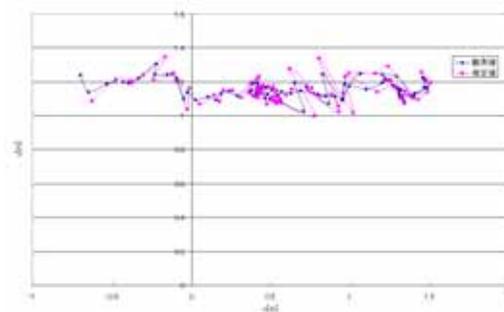
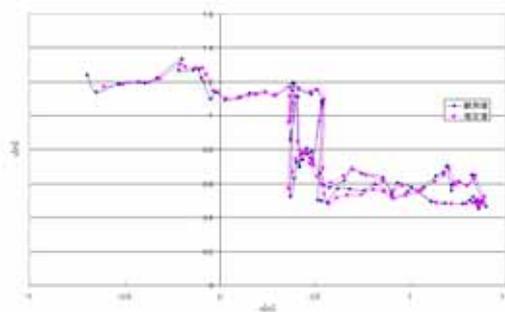


図 20 xy 平面から見た軌跡 (提案手法) 図 21 xy 平面から見た軌跡 (従来手法)

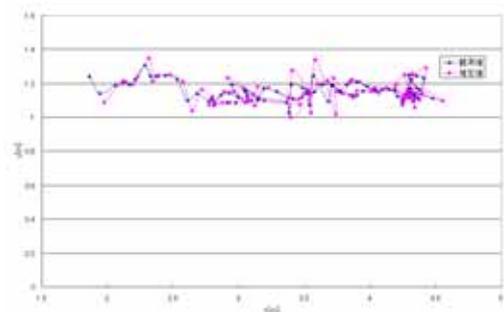
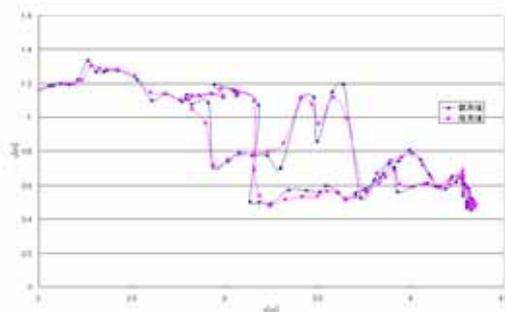


図 22 yz 平面から見た軌跡 (提案手法) 図 23 yz 平面から見た軌跡 (従来手法)

4.3.2 ジャンプアップ時

本実験で使用したジャンプアップ画像の一部を図 24 に示す。

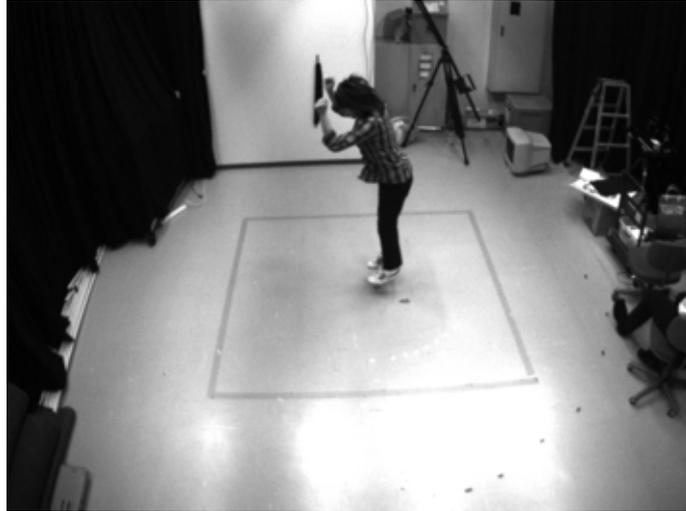


図 24 ジャンプ時の画像

ジャンプをしている人の位置追跡精度および処理時間を以下の表 5 に示す。なお、特徴点抽出処理に関しては従来手法においても 3.1 節で言及した手法を用いている。また、提案手法と従来手法の xy, yz, zx 各平面における追跡軌跡を図 25 ~ 30 に示す。

	提案手法	従来手法
位置追跡精度	0.052[m]	0.324[m]
特徴点抽出処理時間	124.2[msec]	同左
クラスタリング処理時間	43.46[msec]	43.88[msec]
追跡処理時間	0.279[msec]	2.428×10^{-3} [msec]
その他処理 (前処理, 空間座標算出等)	85.66[msec]	同左

表 5 ジャンプ時の位置追跡精度および処理時間

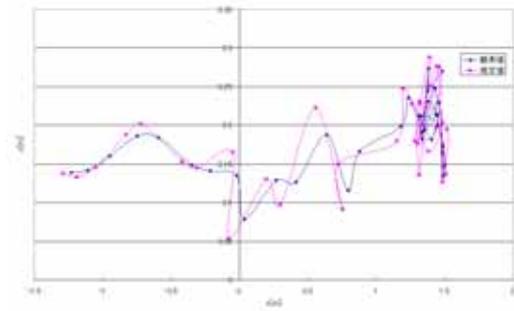
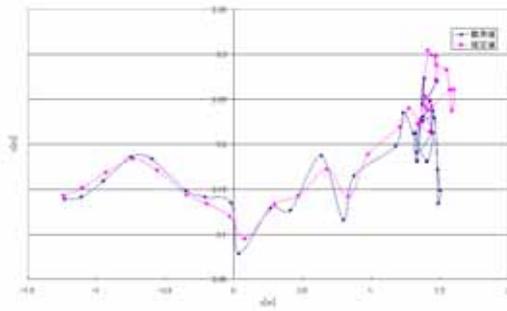


図 25 xz 平面から見た軌跡 (提案手法) 図 26 xz 平面から見た軌跡 (従来手法)

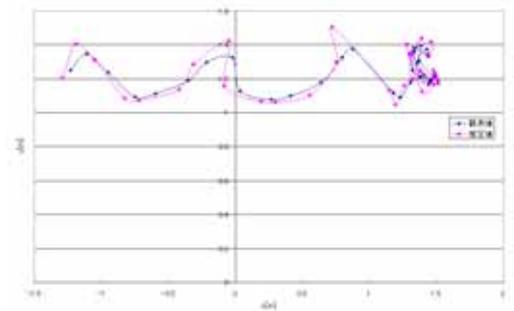
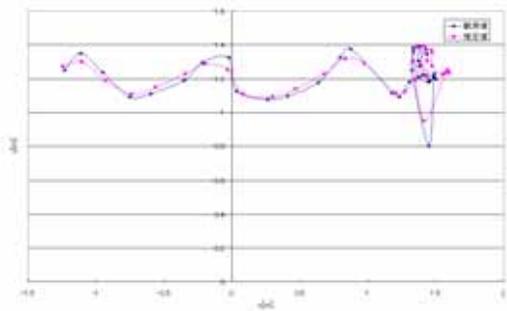


図 27 xy 平面から見た軌跡 (提案手法) 図 28 xy 平面から見た軌跡 (従来手法)

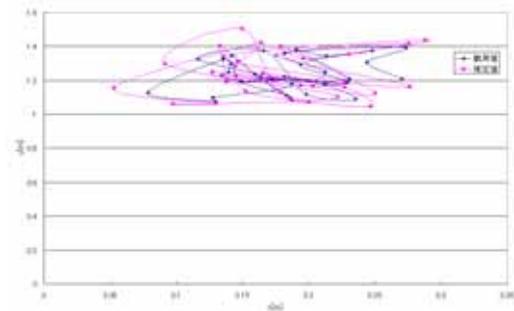
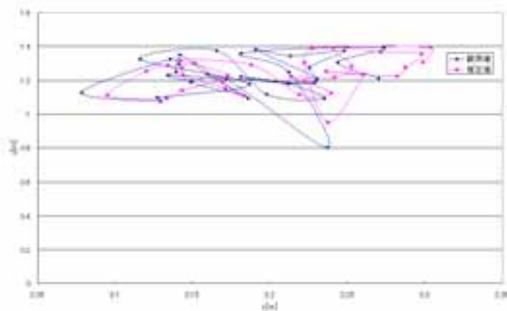


図 29 yz 平面から見た軌跡 (提案手法) 図 30 yz 平面から見た軌跡 (従来手法)

4.3.3 転倒時

本実験で使用した転倒画像の一部を図 31 に示す。

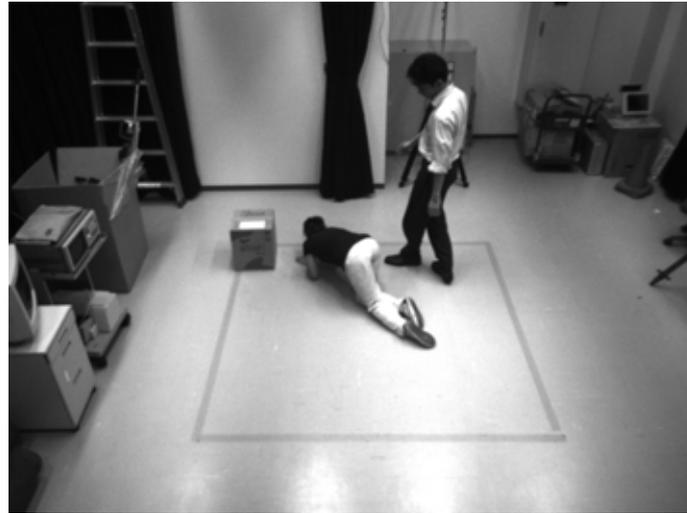


図 31 転倒時の画像

転倒した人の位置追跡精度および処理時間は以下の表 6 の通りである。なお、特徴点抽出処理、およびその他処理(前処理, 空間座標算出等)に関しては従来手法においても 3.1 節で言及した手法を用いている。また、提案手法と従来手法の xy, yz, zx 各平面における追跡軌跡を図 32 ~ 37 に示す。

\	提案手法	従来手法
位置追跡精度	0.058[m]	0.063[m]
特徴点抽出処理時間	127.2[msec]	同左
クラスタリング処理時間	44.62[msec]	45.34[msec]
追跡処理時間	0.439[msec]	3.242×10^{-3} [msec]
その他処理(前処理, 空間座標算出等)	159.1[msec]	同左

表 6 転倒時の位置追跡精度および処理時間

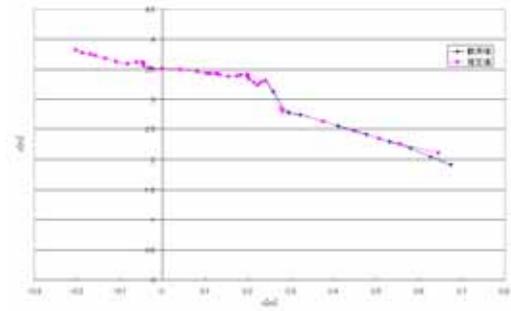
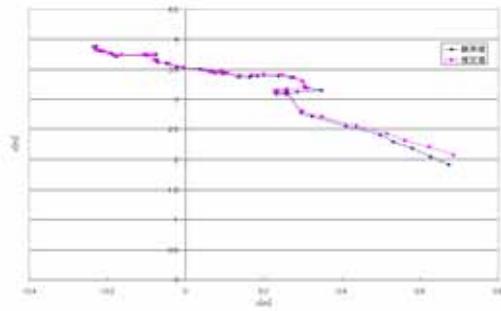


図 32 xz 平面から見た軌跡 (提案手法) 図 33 xz 平面から見た軌跡 (従来手法)

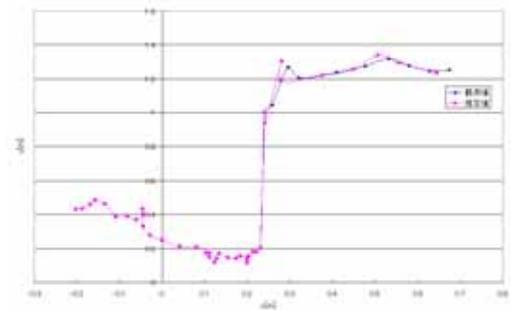
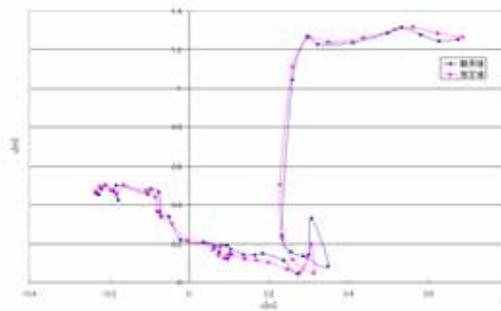


図 34 xy 平面から見た軌跡 (提案手法) 図 35 xy 平面から見た軌跡 (従来手法)

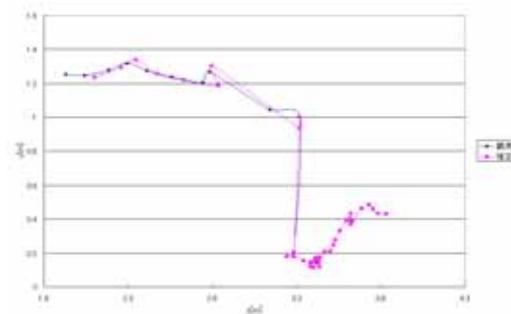
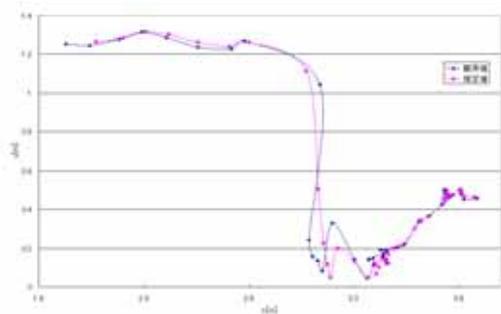


図 36 yz 平面から見た軌跡 (提案手法) 図 37 yz 平面から見た軌跡 (従来手法)

4.3.4 多人数出現時

本実験で使用した多人数出現画像の一部を図 38 に示す。



図 38 多人数出現画像

多人数出現時の位置追跡精度および処理時間は以下の表 7 の通りである。なお，特徴点抽出処理，およびその他処理（前処理，空間座標算出等）に関しては従来手法においても 3.1 節で言及した手法を用いている。また，提案手法と従来手法の xy, yz, zx 各平面における追跡軌跡を図 39 ~ 44 に示す。

\	提案手法	従来手法
位置追跡精度	0.020[m]	0.079[m]
特徴点抽出処理時間	125.6[msec]	同左
クラスタリング処理時間	68.88[msec]	65.34[msec]
追跡処理時間	1.092[msec]	2.678×10^{-3} [msec]
その他処理（前処理，空間座標算出等）	111.4[msec]	同左

表 7 多人数出現時の位置追跡精度および処理時間

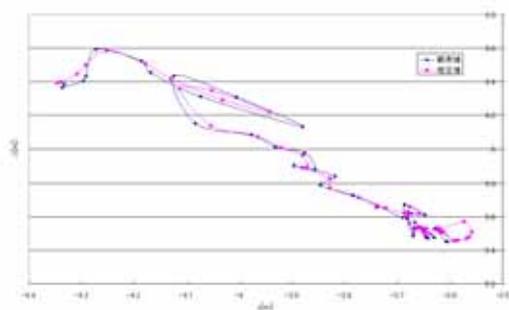


図 39 xz 平面から見た軌跡 (提案手法)

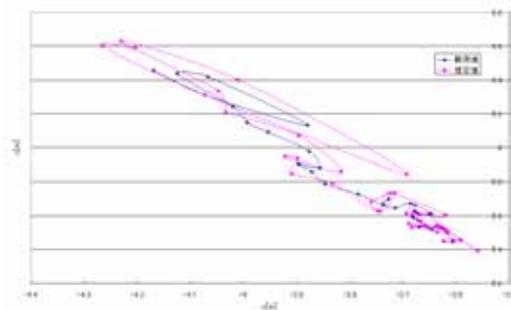


図 40 xz 平面から見た軌跡 (従来手法)

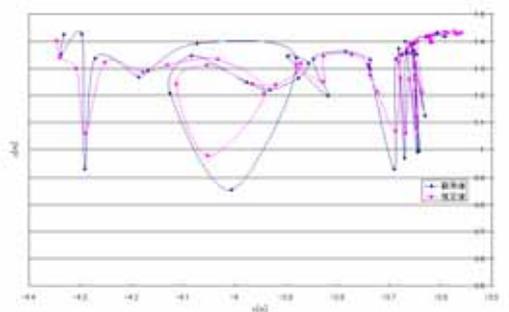


図 41 xy 平面から見た軌跡 (提案手法)

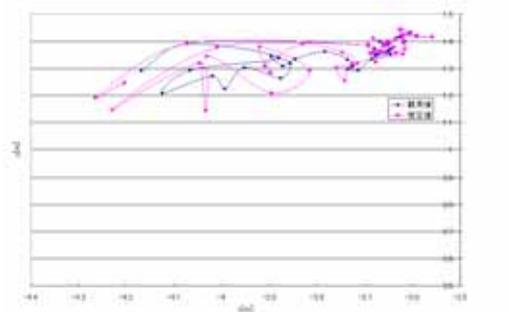


図 42 xy 平面から見た軌跡 (従来手法)

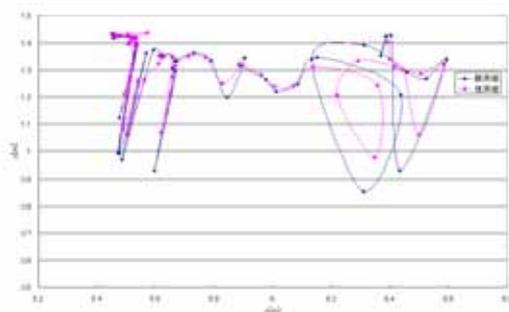


図 43 yz 平面から見た軌跡 (提案手法)

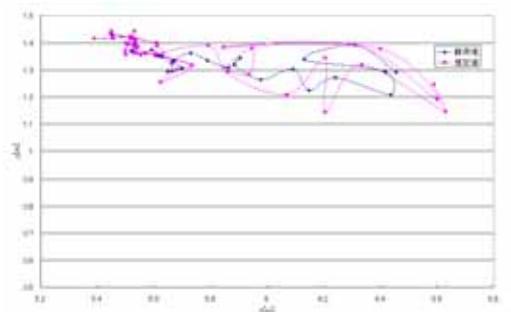


図 44 yz 平面から見た軌跡 (従来手法)

5. 考察

5.1 本手法の性能検証に対する考察

5.1.1 処理時間

特徴点抽出処理に関しては、画像全ピクセルに対して局所相関を取っていることで、抽出精度は高いが、他処理部と比較して処理時間が長い。また、クラスタリング部、追跡部に関してはエリア内に現れる人数に依存する。人数が増えた場合、処理時間は線形的に増加して行く。通常歩行および多人数出現時における実験では、同時に4人以上の人が現れたため、処理時間が他の姿勢のものに比べて長い。

特徴抽出処理が約130[msec]、人数に依存するが、エリア内に3~5名程度でクラスタリング処理が約50[msec]、追跡処理が約0.7[msec]で、ステレオ対応付けによる3次元空間座標算出処理、画像の読み込み、表示処理等を含めても1フレームの処理に平均1[sec]程度有するため、リアルタイム処理は実現できていない。しかしながら、人間の平均的な歩行速度を4[km/h]≒1.11[m/s]と考えた場合、観測空間が10×10×3[m]、処理エリアが4×4×3[m]であること、および計算機処理速度を考慮すると、オンライン処理として扱うのに十分な結果が得られたと考える。また本システムの表示機能をOFFにし、人の3D位置のデータロガーや人数カウンターとして使用すれば、100~400[msec]の処理時間で運用可能である。

5.1.2 位置追跡精度

位置追跡精度に関しては、各姿勢において0.05[m]程度と良好な結果が得られたと考える。ジャンプや転倒等、高さ(Y軸)方向の変動の大きいものに関しては、人物輪郭モデルを人クラスタ毎に更新して行く処理の性能が発揮されていると考える。Kalman Filterのような線形フィルタでは急な変動が起こった場合追跡が難しいが、図13(p.21)のように追跡結果を検知処理にフィードバックすることにより、高さ(y)方向に関しては急変動にも対応可能になった。フレームレートを上げることにより、更なる精度向上が見込まれる。

5.2 従来手法 ([9]) との比較

5.2.1 処理時間

クラスタ移動量追跡を 3 次元追跡処理 (本手法性能検証実験と同様に人クラスタ重心を追跡) に拡大し, 本手法との比較検証実験を行った. 処理時間は, クラスタリング処理, および追跡処理ともクラスタ移動量推定手法の方が短い. クラスタリング処理部に関しては, 個々の人クラスタに対して, 人物輪郭モデル閾値を動的に更新する処理を追加したためである. 現れる人数に依存するが, 4~5 名程度で 5[msec] 程度本手法の方が処理時間が長い. また, 追跡処理部に関しては, クラスタ移動量推定の場合, 重心位置の移動分のバイアス演算であるので高速である. 本手法と比較して, 10^2 オーダー処理時間差があるが, 処理時間全体に占める割合が極めて小さいので全く差し支えない. なお, 特徴抽出部に関しては提案手法を利用したので, 当該処理部の処理時間計測は行っていない. 5.1 節でも言及したように, 本手法はオンライン処理として差し支えないレベルであるので, 従来手法と比較しても遜色はないものとする.

5.2.2 位置追跡精度

本手法と提案手法における大きな違いは, 高さ (Y 軸) 方向に変動のある姿勢変化が起きた場合に安定して人を検出し, 追跡が行えるかどうかという点である. xz 平面で見た場合の軌跡 (図 19,26,33 参照) に関しては, 提案手法とクラスタ移動量推定手法で大差はない. しかしながら, 高さ (Y 軸) 方向に関しての安定した追跡ができていない. これは, 人物輪郭モデル閾値を固定にしていることに起因すると考える.

追跡軌跡は提案手法と大きく異なってくる. その理由として, 人候補特徴点群初期値は $TL \leq (x_f, y_f, z_f) \leq TH, (i = 1, 2, \dots, n)$ であるが, 本手法では人クラスタ毎に高さ閾値 TL, TH を逐次的に更新して行く. 高さ (Y 軸) 方向の姿勢変化が起きた場合は高さ閾値 TL, TH の値が小さくなり, 人候補となる特徴点群も変化する. 人特徴点群候補が変化することによって, 追跡点 (人クラスタの重心) 座標も変化し, 追跡軌跡に変化が出る.

6. 結言

本研究では，Harris Operator の出力値および，抽出した特徴点間の距離に着目することで，人物輪郭を形成する特徴点を高精度に抽出した．そして，ステレオビジョンにより特徴点群の空間情報を導出し，追跡結果を検知処理にフィードバックする人検知を実現することで，姿勢変化に対応したロバストな検知を実現した．また，高さ情報を加えた3次元追跡モデルにより，人の3次元的な動作追跡を可能とした．提案手法により実現可能となったことを以下に列挙する．

- Harris Operator の応答関数値に着目し，人検知のためのクラスタリングに有効な人輪郭特徴点抽出を実現した
- 高さ情報を加えた3次元追跡モデルを使い，人の3次元的な動きを逐次追跡する手法を実現した
- 追跡結果を利用し，人の頭部近辺高さ位置を導出することで，人の姿勢変化に対応した検知を実現した
- カメラ位置に依存しないアルゴリズムで，市販カメラおよびコンピュータで処理可能とした

本手法は，オンラインで行える人物検知追跡手法であり，監視システム等へ実用化が可能であると考えられる．今後は本研究では取り上げなかった，より複雑な挙動（手を回しながら移動，ほふく前進など）についてもさらに評価検証を進める．また応用としては，さらにリアルタイム性を向上させることにより，ロボットの視覚センサ等，様々な画像センサーへの適用が可能であると考えられる．

謝辞

本研究を進めるにあたり，数々の有益なご指導を下さいました情報科学研究科 視覚情報メディア講座 横矢直和教授に厚く御礼を申し上げます．

副指導教官としてご助言下さいました情報科学研究科 像情報処理学講座 千原國宏教授，光センシング講座 緒方司郎教授に感謝の意を表します．

研究の方向性について有益なご指導，ご助言を下さいました光センシング講座 諏訪正樹助教授，オムロン株式会社 來海雅俊主幹研究員および池谷崇氏に深謝致します．

中間発表等において適切なお助言を下さいました情報科学研究科 視覚情報メディア講座 山澤一誠助教授，神原誠之助手，佐藤智和助手に御礼申し上げます．そして，日頃からお世話になっておりました視覚情報メディア講座 博士後期および前期課程のみなさん，守屋知代秘書，光センシング講座 木田周作氏に心から感謝いたします．

参考文献

- [1] 株式会社富士通ゼネラル. 公開特許公報【侵入者監視装置】. 2003.
- [2] 塩昭夫, ジャックス克蘭スキー. ”動画像からの歩行者検出”. 情報処理学会, CV 75-5, pp.1-8, 1991.
- [3] 藤原秀人, 谷口博康, 黒田伸一. ”動き画像を用いたエスカレータ乗客転倒検出手法”. 情報処理学会, CV 78-5, pp.33-41, 1992.
- [4] 森田俊彦. 局所相関演算による動きの検知と追跡. 電子情報通信学会 (D-II), Vol.J84-D-II, no.2, pp.299-309, 2001.
- [5] O.Javed, Z.Rasheed, K.Shafique, and M.Shah. ”tracking across multiple cameras with disjoint views”. *Int'l Conference on Computer Vision*,, 2003.
- [6] D.Beymer and K.Konolige. Real-Time Tracking of Multiple People Using Continuous Deteciton. *Int'l Conference on Computer Vision*, 1999.
- [7] オムロン株式会社. 公開特許公報【移動体検出方法及び装置並びに移動体認識方法及び装置並びに人間検出方法及び装置】. 日本国特許庁, 1999.
- [8] Point Grey Research Inc. People tracking system. <http://www.ptgrey.com/>, 1997.
- [9] 池谷崇, 諏訪正樹, 麻生川佳誠, 来海雅俊. 重なりに強いステレオ型人間検知システム. 画像ラボ, VOL.15, No14, pp53-56. 日本工業出版, 2004.
- [10] 林健太郎, 橋本学, 鷺見和彦. ステレオ距離画像を用いた人物計数技術の提案. ”信学技報 PRMU2002-148, pp.13-18”, 2002.
- [11] 柴山憲文, 埜圭二, 太田勝. 踏切障害物検知用のステレオ画像装置の開発. 画像センシングシンポジウム 2003, pp.1-6, 2003.
- [12] 金澤靖, 金谷健一. コンピュータビジョンのための画像の特徴点抽出. 電子情報通信学会誌, VOL.87, No.12, pp.1043-1048. 2004.

- [13] 出口光一郎. ステレオの仕掛けを解き明かす. 情報処理学会誌, VOL.37, No.7, pp.662-670. 1996.
- [14] R.Szeliski and G.Hinton. Solving random-dot stereograms using the heat equation. *In Proc CVPR*, pp.284-288, 1985.
- [15] 鳥脇純一郎. 認識工学. コロナ社, pp.76-95, 1993.
- [16] 栗原浩次. データとデータ解析. 放送大学教育振興会, pp.130-155, 1996.
- [17] 柳熙泛, 宮永喜一. ロバストデータクラスタリングによる画像のセグメンテーションについて. 電子情報通信学会 (A), Vol.J86-A, no.7, pp.731-738, 2003.
- [18] 森大樹, 内海章, 大谷淳, 谷内田正彦, 中津良平. "非同期多視点画像による人物追跡システムの構築". 電子情報通信学会 (D-II), Vol.J84-D-II, no.1, pp.102-110, 2001.
- [19] C.Harris and M.Stephens. A Combined Corner and Edge Detector. *Proc.Alvey Vision Conf.*, pp.147-151, 1988.
- [20] C.Schmid, R.Mohr, and C.Bauckhage. Comparing and Evaluating Interest Points. *Proc.6th Int'l.Conf.on Computer Vision*, pp.230-235, 1998.
- [21] 宮脇隆志, 石橋聡, 岸野文郎. 色彩情報を用いたカラー画像の領域分割. 信学技報, IE89-50, 1989.
- [22] Y.Dai and Y.Nakano. Extraction of facial images from complex background using color information and SGLD matrices. *Proc.Int'l Workshop on Automatic Face and Gesture Recognition*, pp.238-242, 1995.
- [23] 片山徹著. "応用カルマンフィルタ". 朝倉書店, pp.83-132, 2002.
- [24] Y.Wu. Visual Tracking. *ECE510-Computer Vision Notes Series 7*, 2002.
- [25] M.Isard and A.Blake. CONDENSATION-conditional density propagation for visual tracking. *Int'l Journal of Computer Vision*, 29:5-28, 1998.