

研究テーマ 人物追跡に関する研究

研究者	富永将史	財団法人ソフトピアジャパン	雇用研究員
	Jean-Christophe TERRILLON	財団法人ソフトピアジャパン	雇用研究員
	渡辺博己	財団法人ソフトピアジャパン	雇用研究員
	石井洋平	財団法人ソフトピアジャパン	雇用研究員

フェーズ I

本研究は、基盤となる技術として「人物検出に関する研究」における存在確立を用いた視体積交差法による人物検出技術を基にフェーズ II において行った。基盤技術の詳細は「人物検出に関する研究」を参照のこと。

フェーズ II

1 研究の概要

「人物検出に関する研究」におけるパーセプトルルームの構築で培った技術を基に、人物追跡技術を確認し、応用モデルとして自動マーケティング情報システムを構築した。

自動マーケティング情報システムのプロトタイプとしてプラズマディスプレイに複数台のカメラを設置し、提示される広告に対する注目度および、顧客の性別・年代といった人物属性を広告毎に収集可能なシステムを実現した。更に、顧客に対しては注目度の高い広告の詳細な情報が提供できる。

また、自動マーケティング情報システムを室内空間に拡張し、顧客の動線を検出可能にした。入室から退室までの行動を把握することで、商品ブースにおける集客率を測ることができる。更に動線の累積頻度を求めることで、デッドスペースの検出にも繋がり、店舗レイアウトに活用することができる。

2 研究の目標

店舗内外に設置した広告媒体（ポスター、モニター、マネキン、商品自体の展示等）に対する顧客の注目行動を、自動的にかつ一定の評価基準により収集することができれば、多様な商品の市場調査が可能になると考えられる。そこで、提示情報に注目する人物の属性や注目度を分析することで、マーケティングに必要な情報を効率よく収集する“自動マーケティング情報システム（Automatic Marketing Information System）”の構築を目指し、プロトタイプとしてプラズマディスプレイ提示される広告に対する注目度および、顧客の性別・年代といった人物属性を広告毎に収集可能なシステムを実現した。本システムは複数台の固定カメラにより、視体積交差法により複数人物の位置追跡、顔検出、動作検出を行う。固定カメラから推定された位置情報を基に、複数台のパンチルトカメラにより詳細な顔画像を取得し、対象人物の属性や注目度を推定する。

また、複数の人物を同時に追跡することにより人物の動線を把握し、その移動履歴を分析することで、人口密度分布を得ることができ、店舗やオフィスのレイアウトへ反映することができると考えられる。更に行動パターンを細かく分析し、推定した性別・年代毎に集計することで、人物属性毎の商品開発におけるマーケティング情報としても有効なデータを収集することができる。そこで、店舗に訪れた顧客の動線から興味行動を分析する室内型 AMIS の実現を目指し、実時間での複数人物追跡および動線の検出が可能なシステムを構築した。

3 実施内容

3.1 自動マーケティング情報システム（プロトタイプ）

自動マーケティング情報システム（Automatic Marketing Information System）は広告媒体として情報提示装置を用い、提示する宣伝映像に対する注目度と注目人物の属性を収集する。そのため、調

査対象となる情報提示装置前の広い空間で、行き交う不特定多数の人物から、情報提示装置の宣伝映像に注目する人物を識別し、検出する必要がある。対象となる宣伝映像に注目する人物は、宣伝映像を見るために情報提示装置前で立ち止まり、顔を向ける。そこで本システムは、情報提示装置に対して静態する人物の位置を複数の固定カメラを用いて検出する。更に、その人物の属性や注目度を推定するためには、検出位置の情報を基にしてパンチルトカメラを制御し、その人物の顔を拡大して捕らえる。以上の目的を達成するため、自動マーケティング情報システムは、

- ・複数対象人物の位置検出
- ・顔検出とパンチルトカメラによる顔画像の獲得
- ・顔向き推定
- ・性別・年代推定

の4つの処理により成る。

3.1.1 システム構成

自動マーケティング情報システム（AMIS）では、宣伝映像を見る人物の属性および注目情報を詳細に捉えることを目的とし、提示用のプラズマディスプレイに複数のカメラを備える。図1にシステム構成を示す。画像提示用の50インチプラズマディスプレイに3台の固定カメラと4台のパンチルトカメラ配置した。固定カメラは、ディスプレイ上部（床より高さ170cm位置）に60cm間隔、光軸を一致させ、下方10度を向くよう設置した。この固定カメラ画像は、画面合成スイッチャーを通すことで1台のセンシングPCにより、同期の取れたビデオレートのカラー画像（各画像320×240）として取得し、人物位置の推定および顔検出に用いる。プラズマディスプレイ上下に設置したパンチルトカメラは、それぞれ1台のカメラPCによりビデオレートのカラー画像（各画像640×480）を取得する。パンチルトカメラは、固定カメラの位置推定結果を基に、パン、チルトおよびズームの制御を行うことで、対象人物の高精度な顔画像が取得できる。市場調査に有効な人物属性および注目情報は、この顔画像より推定する。メインPCは宣伝映像の提示と、顧客情報の集計を行う。なお、各PCはTCP/IPによりカメラ制御および顧客情報等の通信を行う。図2にプロトタイプとして構築したシステムの外観を示す。本システムの構築にあたり、Pentium4 - 2.53GHz、512MB、Windows2000のPC、固定カメラとしてSONY Progressive 1/2 3CCDカラービデオカメラDXC-9000 + ズームレンズVCL-0716BXA、パンチルトカメラSONY EVI-D100を使用した。なお、本システムでは2台以上の固定カメラで撮影可能な（PDP前方約1mから、5m未満）範囲を調査対象とする。

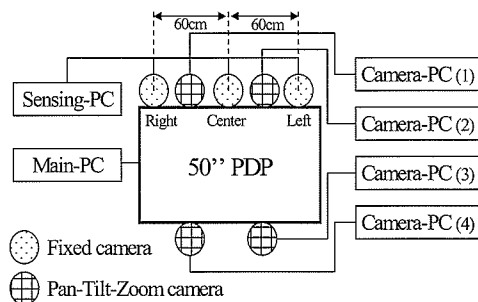


図1. AMIS システム構成

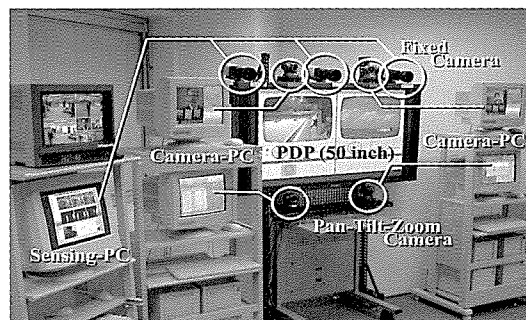


図2. AMIS システム外観

3.1.2 一方向からの複数カメラを用いた視体積交差による複数人物追跡手法の開発

複数人物の追跡および動作推定に、視体積交差法を用いる。視体積交差法によって検出したボクセルは空間的な人物行動の推定に利用することかできる。事前に取得した背景画像との差分、2値化、ラベリングを行い、カメラ画像毎に人物候補領域を検出する。検出された人物候補領域において視体積交差法を行う。なお、視体積交差法を用いる際、全てのカメラパラメータ（カメラ位置、パン・チルト角、焦点距離等）は固定で既知である。

視体積交差法によりボクセル空間に投票したボクセルに対し、ラベリングを行うことで固まり（ブロック）毎に分ける。このラベリングされたブロックのうち、前フレームでの人物重心位置に最も近い重心位置のブロックを、人物のブロック（人物ブロック）とし、このブロックの重心位置

を現在の人物位置とみなす。

各カメラ画像において、背景差分により得られた人物領域（画素）の同一人物同定を、視体積交差法による投票結果を基に推定する。図3に示すように、カメラから投影される画像の各人物候補領域の中心 $(x_\alpha, y_\alpha, z_\alpha)$ 方向に、カメラ位置 $(x_\beta, y_\beta, z_\beta)$ ($\beta=1, 2, 3$: カメラ番号)から伸ばした直線

$$\begin{pmatrix} x - x_\alpha \\ x_\beta - x_\alpha \end{pmatrix} = \begin{pmatrix} y - y_\alpha \\ y_\beta - y_\alpha \end{pmatrix} = \begin{pmatrix} z - z_\alpha \\ z_\beta - z_\alpha \end{pmatrix} \quad (1)$$

が最初に接触した（カメラ位置に最近の）ブロックの座標 $(x_\gamma, y_\gamma, z_\gamma)$ ($\gamma=A, B, \dots$: ブロックラベル)のラベル γ を領域に対応付ける。

$$region_label = \gamma \quad s.t. \left[\min \left((x_\beta - x_\gamma)^2 + (y_\beta - y_\gamma)^2 + (z_\beta - z_\gamma)^2 \right) \right] \quad (2)$$

以上のような処理を全カメラ画像の人物候補領域に対して行うことで、各カメラ画像間での同一人物の対応付けを行うことができる。実際には、人物候補領域の中心のみからその領域全体のラベルを一意に求めるのではなく、領域内の画素全て（もしくは数画素間隔）に対して求め、各画素における人物同定を行う。

図4にPDPを注目する4名の検出例を示す。図4(a)に示す各カメラ画像の同ラベル（A~D）は同定された同じ人物である。図4(b)は各画素におけるボクセル空間による推定された距離を示す。図中、近い距離を白で、遠い距離を黒で示す。図5はボクセル空間中の対応する人物ブロックである。なお、ボクセル空間サイズは（80, 80, 40）、分解能5cmである。この人物ブロックの出現、移動、消失により、人物の調査範囲内での行動（歩行、静止等）が推定できる。

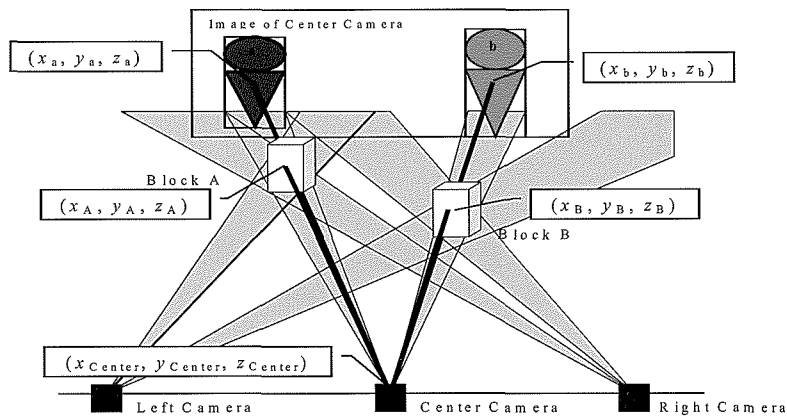


図3. 人物ブロックと人物領域の対応

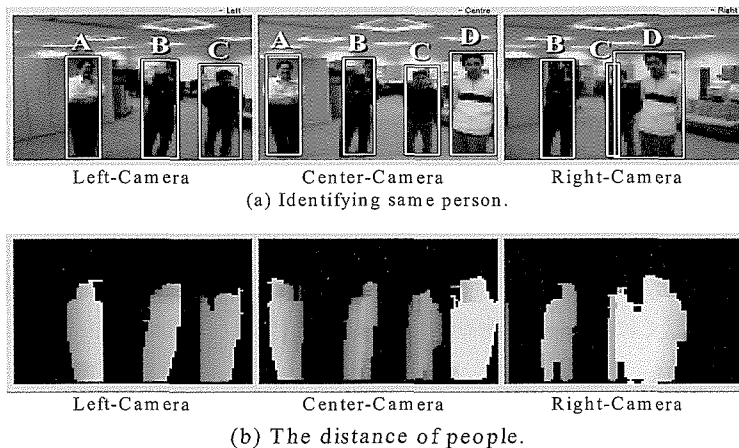


図4. 固定カメラによる人物検出結果

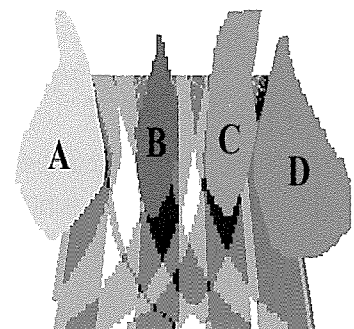


図5. 人物ブロック検出結果

閉空間では各ボクセルにおける平面位置の存在確率は常に対象空間内に人物がいるため、前フレームの位置を基に設計することができる。しかし、本研究では、情報提示装置前の開空間を対象とするため、対象エリアへの入退や、カメラ配置の制約により、整った存在確率のモデルを構築することは困難である。そのため、開空間における平面位置での存在確率は、人物追跡を考慮し、以下の条件を基に定義することとした。

条件1：ボクセル位置 (x, y) において、現地点 $(x, y) = (x_t, y_t)$ と前フレームの人物位置（人物ブロックの重心） (x_{t-1}, y_{t-1}) の距離が限定範囲 AR 以内である。

$$C_{-1}(x, y) = 1 \quad \text{s.t.} \quad |(x_t, y_t) - (x_{t-1}, y_{t-1})| \in AR \quad (3)$$

条件2：ボクセル位置 (x, y) において、属する人物ブロックの体積が、平均人物サイズ AV 以下である。

$$C_{-2}(x, y) = 1 \quad \text{s.t.} \quad \text{Size} \leq AV \quad (4)$$

条件3：ボクセル位置 (x, y) において、過去にブロックが存在したフレームよりの経過時間 t が制限時間 LT 以内である。

$$C_{-3}(x, y) = (LT - t) / LT \quad \text{s.t.} \quad 0 \leq t \leq LT \quad (5)$$

条件4：ボクセル位置 (x, y) において、PDP からの絶対距離が制限範囲 LD 以内である。

$$C_{-4}(x, y) = (LD - D) / LD \quad \text{s.t.} \quad 0 \leq D \leq LD \quad (6)$$

移動限定範囲 AR は、人間の平均移動速度（1 m/sec）および人間の標準サイズ（厚さ 40 cm）より、通常 10 cm の移動に際して 100 msec（1.5 フレーム）必要となることから、1 フレームで移動できる限界として 60 cm と設定した。以上の条件より存在確率を

$$PM(x, y) = \prod_{i=1}^4 C_{-i}(x, y) \quad (7)$$

と定義する。

次に、高さ方向の確率モデルに関して述べる。現フレームのボクセル投票結果から、図 6 に示すように、平面的な各座標位置 (x, y) において、ブロックの存在するボクセルの高さ $V(x, y, z) | z$ （ボクセル空間サイズ $K \times L \times M$ ）の累積

$$HF(x, y) = \sum_{k=1}^M V(x, y, k) |_k \quad \text{s.t.} \quad V(x, y, k) \neq 0 \quad (8)$$

を求める（以降、高さ累積マップと呼ぶ）。この高さ累積マップに対し、存在確率マップを掛け合わせ、

$$EM(x, y) = PM(x, y) \times HM(x, y) \quad (9)$$

閾値 t 以下となる座標位置 (x, y) に存在するボクセルを除去

$$\sum_{i=1}^K \sum_{j=1}^L \sum_{k=1}^M V(i, j, k) = 0 \quad \text{s.t.} \quad EM(i, j) \leq th \quad (10)$$

することで、雑音の除去を行う。高さ累積マップを適用することで、空中に浮かんで現れる身体の一部（腕など）は除去されにくく、床近くに存在する物や人物の影などによる雑音は除去され、人物同士の分離が可能となる。

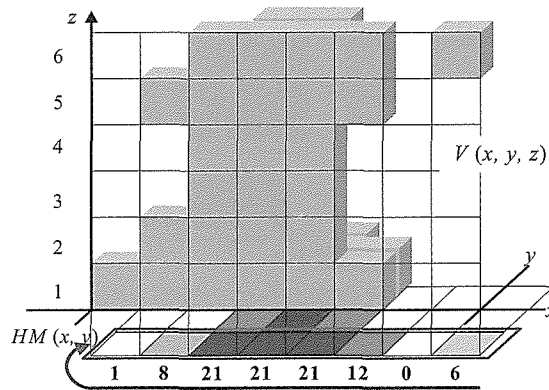


図 6. 高さの累積

3.1.3 顔検出と性別・年代推定

AMIS において、顧客の属性情報となる性別や年代の推定に至る技術は、他の研究テーマで培った手法を用いている。詳細は「顔画像データベースの構築」、「顔及び頭部の検出に関する研究」、「性別および年齢の推定に関する研究」、「顔向き推定に関する研究」にて述べているので、本節では簡単な紹介に留める。

AMIS において、提示広告に注目する人物の顔を検出するため、四方向面特徴 (FDF) とサポートベクタマシン (SVM) を用いた正面顔検出を行った。SVM は処理コストがかかる手法であるが、対象となる領域を視体積交差法により検出した人物領域に限定することで、軽減することが可能である。FDF は入力画像の水平、垂直、右斜、左斜の、4つの方向のエッジ特徴であり、パターンの識別に有効な特徴量の一つである。この FDF は、 8×8 ピクセルの低解像度に変換して処理を行うことで、環境光に対し、ロバストな特徴をもち、更に高速な処理が可能である。

獲得した高解像度顔画像より、注視方向を推定する。瞳検出による注視方向の推定には更に高解像度の画像が必要となるため、本システムでは顔向きを求めることで、注視方向の推定とした。顔向きは顔領域の中心位置と各顔部品のなす角度から推定する。顔部品検出には先に述べた FDF によるテンプレートマッチングと、弛緩整合法を用いる。

事前に収集した 10 人の顔画像を学習データとし、これより顔部品毎の FDF を求め、平均をとることでテンプレート (図 7) を生成する。このテンプレートに対し入力顔画像のマッチングを取り、類似度を正規化し初期対応度とする。個人差や顔向きに柔軟に対応するため、各顔部品間に図 8 に示すようなばね接続の概念を導入する。顔部品間の位置関係からガウス分布を基に周辺の適合確率を求め、顔部品の対応確率を更新する。これを繰り返す弛緩演算により、適切な顔部品位置を検出することができる。

性別・年代の推定には、FDF を線形判別分析 (LDA) で学習した識別器を用い、推定する。性別や年代の推定には、顔全体の大まかな特徴量とともに、目鼻口の詳細な特徴量を用いて判断することが重要と考えられる。そのため図 9 に示すように、顔全体と顔部品の 2 つの FDF を生成し、これらを用い判別を行うこととした。

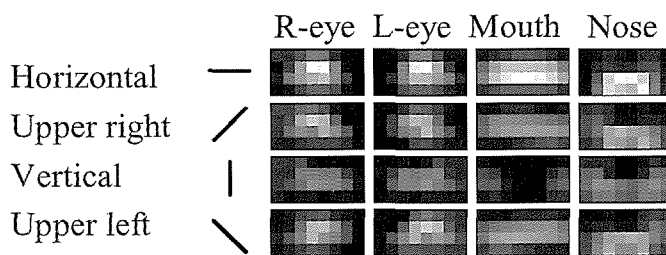


図 7. 顔部品毎の FDF テンプレート

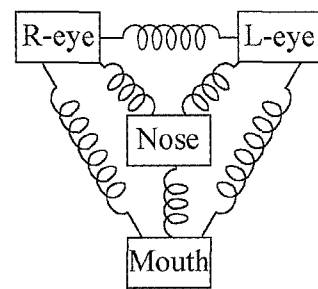


図 8. 接続概念

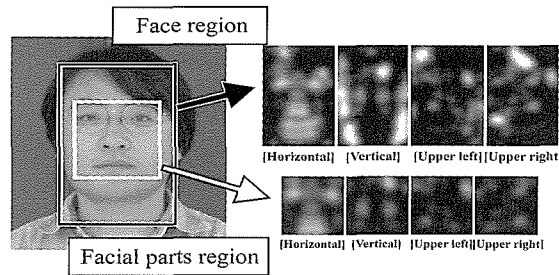


図 9. 性別・年代推定のための顔部品領域 FDF

3.2 室内型自動マーケティング情報システム

コンビニや小型店舗、デパートや百貨店内の小売店などにおいて、顧客の一連の行動を得る事ができれば、有効な商品展開や店舗配置が可能となる。そのため複数の人物を実時間で追跡し、動線を検出するシステムを構築した。室内型自動マーケティング情報システム（室内型 AMIS）においても、これまでに確立した視体積交差法による人物追跡技術をもとに開発を行った。視体積交差法には投票空間の分解能とカメラからの距離によって、多重投票や処理速度がかかるといった問題がある。そこで、画像解像度と投票空間の感度を考慮し、事前に求めたカメラ画像と投票空間の対応関係から、効率よく投票を行うことで、高速な人物位置の検出を実現した。また、先に述べたプロトタイプ AMIS の構築における研究では、視体積交差法に人物同士の接触を抑制する特性をもつ存在確率の概念を導入することで、ロバストに人物同士のセグメンテーションが可能な人物追跡を実現している。これを踏まえ、室内型 AMIS の構築では、投票する空間をこの存在確率の範囲に個人毎に構築することで、更なる高速化を実現した。

3.2.1 システム構成

図 10 に複数人物検出のために構築したシステムのカメラ配置を示す。本システムでは 4 台のカメラを設置した。図 10(a)に上面図を示す。図 10(b)は、カメラ 2 からカメラ 4 の方向に向かって見た側面図（カメラ 1 から部屋中央までの半分）である。部屋は 1 辺 460cm である。視体積交差法による人物検出では（カメラ画像の解像度が十分であれば）室内サイズをより大きく、広い空間をカバーできるが、今回は小売規模の店舗を想定し、このように構築した。カメラは四隅の高さ 250cm の位置に下方 45° を向くよう設置した。カメラ自体は一般のコンビニ等に監視カメラとして用いられているものと同様であるが、水平画角 110°、垂直画角 81.57° の魚眼レンズを装着した。部屋の角にカメラを設置することで、（カメラ直下から 18cm の範囲は見えないが）室内全体をカバーできる。図 10 のように各カメラを設置すると、部屋の全ての位置において、4 台のカメラから同時に人物を撮像することができる。4 台のカメラからの入力、4 ポート入力可能なキャプチャボード（Ituner Networks Corp 製 Spectra 8）を通じて 1 台の PC に接続されており、それぞれビデオレートのカラー画像（320×240）が取得できる。なお、Xeon 3.06 GHz, dual processor の PC を用いた。

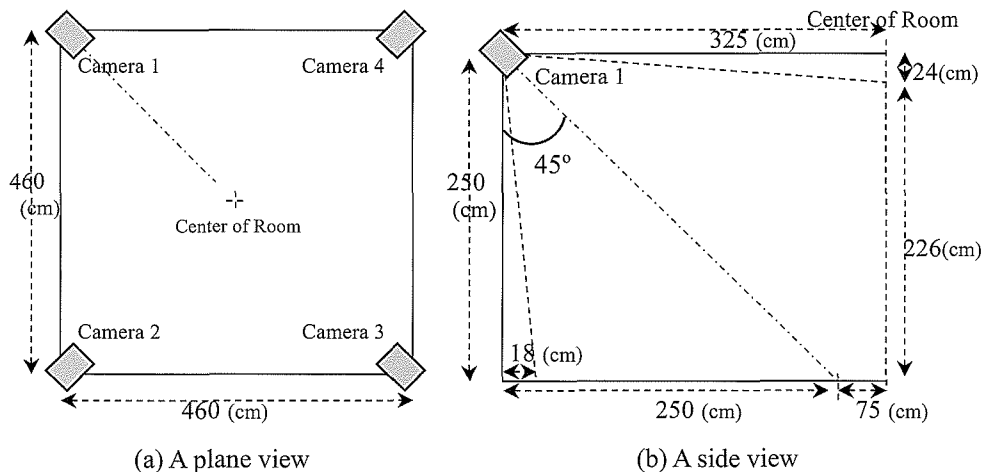


図 10. カメラ配置

本システムでは魚眼カメラを用いているため、視体積交差法の際に歪曲を補正して投票を行う必要がある。歪曲率 D は、

$$D = (Y - H)/H \times 100 \quad (\%) \quad (11)$$

とした。ここで、 Y は実像高、 H は理想像高である。人物検出が目的であり、ミリ単位の精度が必要な形状復元を目的としないことから、一般的なモデルである $f\theta$ を用いる。通常のカメラ $f \tan \theta$ に対する歪曲率 D は、

$$D = (f\theta - f \tan \theta)/f \tan \theta \times 100 \quad (\%) \quad (12)$$

となる。それゆえ、原画像 $F(i, j)$ は、光軸と画像中心座標 (h_i, h_j) が一致していると仮定すると、

$$G\left(h_i + \left(\frac{i - h_i}{1.0 + D \times 100}\right), h_j + \left(\frac{j - h_j}{1.0 + D \times 100}\right)\right) \quad (13)$$

のように補正できる。

図 11(a)に歪曲率のモデルを示す。図 11(b)に原画像を、図 11(c)に補正後の画像を示す。なお、視体積交差を用いる際、補正画像を生成するわけではなく、この歪曲モデルをパラメータとした投票を行う。

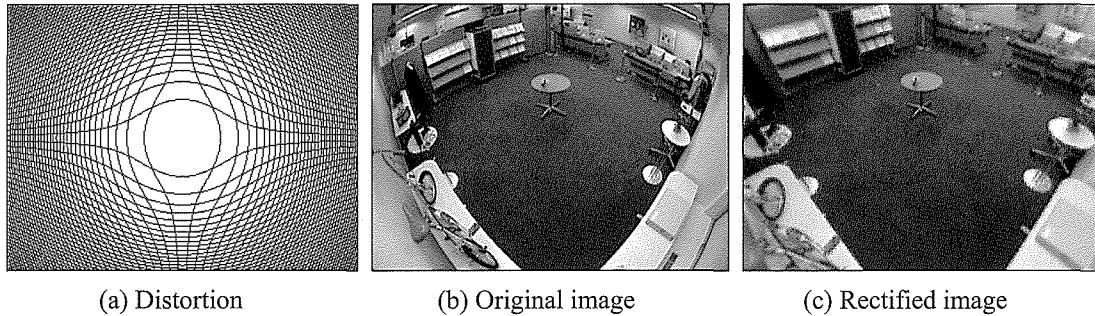


図 11. 魚眼カメラ画像の歪曲補正 (110°)

3.2.2 視体積交差における問題解決と複数人物追跡手法の開発

視体積交差法を用いることで、複数人物の動きや行動を 3 次元空間で推定することが可能になる。更に、それぞれのカメラ画像中の対応する人物領域を推定することができ、人物認識や人物行動の把握に有効な情報として利用することができる。

視体積交差法では、まず複数のカメラ画像において、事前に取得した背景画像との差分によりシルエット画像を得る。このシルエット画像とカメラ位置の関係に基づいて、視体積の交差を投票により求めることで投票空間に 3 次元的人物ボクセルの検出ができる。

ボクセルはカメラ位置とシルエットの各画素を結ぶ直線状に投票される。視体積交差法による形状復元では、精度を高めることが重要であり、空間分解能を高める必要がある。しかし、本研究の目的である複数人物の検出では形状の精度は重要ではなく、人物同士の分離と、処理速度が重要となる。そのため、投票空間 (空間サイズ $K \times L \times M$) の分解能は低く (粗く) 設定することとなる。このような場合、投票空間の分解能に対して、カメラ画像の解像度が高くなり、カメラに近い位置のボクセルへ多重投票といった問題が起きる。そのため、投票時には多重投票の回避を行うなどのチェックが必要となる。なお、画像解像度は、カメラから最も遠い位置での投票空間の感度と対応させることで、ボクセルに投票されないことが回避できる。

毎フレーム毎にカメラ画像から検出できるシルエットを元に投票ボクセルを求めるのに対し、投票空間の位置における対応画素を事前に求めておくことで、多重投票および回避チェックにかかる計算コストを減らすことができる。図 12 に投票空間と画像の関係を模式的に示す。各ボクセルはカメラからの距離によって、画像の対応する領域のサイズが大きく異なる。つまりカメラに近い位置のボクセルに対応する範囲は大きく、カメラから遠い位置のボクセルに対応する範囲は小さくなる。

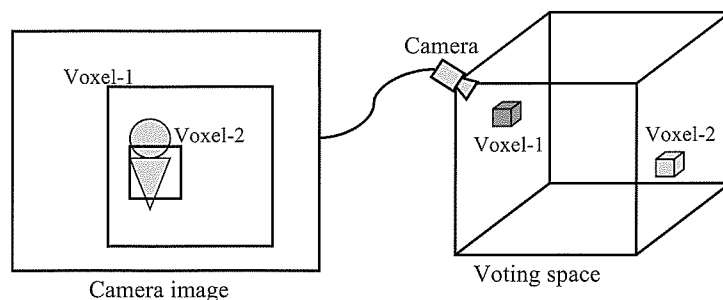


図 12. 投票空間とカメラ画像の関係

各ボクセルに対応する画像領域内に存在するシルエットの割合により、そのボクセルが人物か否（雑音）かを判断するための閾値が定められ、ボクセルへの投票が行われる。位置 (x, y, z) に対するあるカメラ C からの投票度数 $V_c(x, y, z)$ は、画素 F_c の平均値

$$V_c(x, y, z)_{ave} = \frac{1}{K} \sum_{i,j}^K F_c(i, j) \quad (14)$$

s.t. $F_c(i, j) \in Tv(x, y, z)$

より、

$$V_c(x, y, z) = \begin{cases} 1 & \text{if } V_c(x, y, z)_{ave} \geq th1 \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

として求まる。ここで $Tv(x, y, z)$ は、ボクセル $V(x, y, z)$ に対応する画像領域を示し、 K は画素数である。図 13 にカメラからの距離と画素数の関係（投票空間解像度 5.0cm）を示す。カメラに近いほど画素数は極端に増すが、画像領域の大きさは異なっても、空間分解能は一定であることから、閾値 $th1$ は対象の大きさによって一意に決定できる。人物検出を目的とした場合、空間分解能が人物に対して十分小さければ、高く設定することができる。

ボクセル V への投票度数は、 N 台のカメラから、

$$V(x, y, z) = \frac{100}{N} \sum_{c=0}^N V_c(x, y, z) \geq th2 \quad (16)$$

として求められ、閾値 $th2$ によって人物ボクセルとして判断される。閾値 $th2$ はカメラの台数より決定できる。

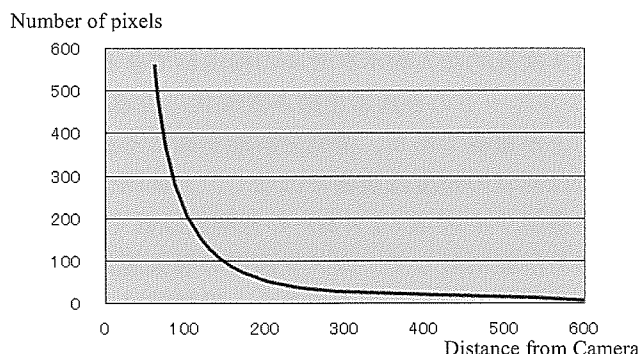


図 13. カメラからの距離と画素数の関係

視体積交差法において、雑音除去と人物同士の分離問題がある。これら問題に対し、存在確率マップを用いることで解決する。

各ボクセルにおける平面位置の存在確率は、サンプリング間隔が人物の移動に対して充分短ければ、前フレームの位置に近いほど高く、離れるに従い低くなると仮定できる。さらに、人物同士が

近接する場合、他の人物の存在確率が高い位置では、自分の存在できる確率はその分抑制されると考えられる。そのため、ここでは相手の存在によって各自の存在確率は減少すると仮定する。このような仮定から、ある人物 i が位置 (x, y) で存在する確率 $P_i(x, y)$ を、

$$P_i(x, y) = \left(\sqrt{(fx - x)^2 + (fy - y)^2} / R \right) \times 100\% \quad (17)$$

のように定義する。ここで、 R は存在確率を適用する半径、すなわち人物が1フレーム間に移動可能な範囲である。また、前フレームでの人物位置 (fx, fy) は半径 R の円の中心 (Rx, Ry) に等しい。存在確率の分布には正規分布等を用いたモデルを適用することで、人物位置の推定信頼度を向上させ得ると考えられる。しかし、本研究では、実時間で的人物追跡と人物領域検出を目指しているため、分解能を粗く設定し、計算量を考慮した単純な分布モデルが好ましい。そのため、本研究では、前フレームの人物位置を基準に単調減少に設計した。

さらに、存在確率に移動方向を考慮した変形を行う。人物の推定位置を基準にして、確率分布の適用範囲をスライドするように設計した場合、人物の急激な移動方向の変更に対応できなくなる。また、推定位置や推定方向に存在する雑音や他の人物に強く影響をうける可能性がある。そのため、存在確率分布の適用範囲を、現在の位置から1フレーム以内に人物が移動可能な距離を十分に含むように設定し、存在確率のみを推定位置を基準に変形する。すなわち、現在の位置から移動可能な存在確率の適用範囲に相当する円錐の底辺は移動速度に関わらず一定とし、存在確率の最高位置である頂点を次フレームの推定位置 (dx, dy) になるよう移動速度・加速度を反映させた変形を行う。以上により、分解能の低いボクセル空間においても複数人物の分離に有効に働くと考える。人物位置の推定では、ボクセル空間の分解能を粗くした場合、瞬間速度と加速度を用いると1フレームでの移動ボクセル数が1ボクセルに満たないことがある。安定した推定を行うため、過去5フレームの平均速度 (vx, vy) と平均加速度 (ax, ay) を算出し、推定位置 (dx, dy) を求める。存在確率 $P_i(x, y)$ は、位置 (x, y) と $t=5$ フレーム後の予測人物位置

$$(dx, dy) = \left(fx + v_x t + \frac{1}{2} a_x t^2, fy + v_y t + \frac{1}{2} a_y t^2 \right) \quad (18)$$

を結ぶ直線と、存在範囲（半径 R ）の円の交点から求まる。交点が存在しない場合、

$$P_i(x, y) = 0\% \quad (19)$$

となる。交点数が2： $(x1, y1)$ 、 $(x2, y2)$ 、又は1： $((x1, y1)=(x2, y2))$ の場合、推定位置 (dx, dy) から位置 (x, y) までの距離を $D = \sqrt{(x-dx)^2 + (y-dy)^2}$ 、 $(x1, y1)$ までの距離を $D_1 = \sqrt{(x1-dx)^2 + (y1-dy)^2}$ 、 $(x2, y2)$ までの距離を $D_2 = \sqrt{(x2-dx)^2 + (y2-dy)^2}$ としたとき、

$$P_{i_1}(x, y) = ((D_1 - D) / D_1) \times 100\% \quad (20)$$

$$P_{i_2}(x, y) = ((D_2 - D) / D_2) \times 100\% \quad (21)$$

より、推定位置 (dx, dy) から円の中心までの距離を $D_R = \sqrt{(Rx-dx)^2 + (Ry-dy)^2}$ 、位置 (x, y) から $(x1, y1)$ までの距離を $D_3 = \sqrt{(x1-x)^2 + (y1-y)^2}$ 、 $(x2, y2)$ までの距離を $D_4 = \sqrt{(x2-x)^2 + (y2-y)^2}$ とすると、

$$P_i(x, y) = \begin{cases} P_{i_1}(x, y) & \text{if } (D_R \geq R \ \&\& \ D_1 \geq D_2) \\ & \parallel (D_R < R \ \&\& \ D_1 \geq D_2 \ \&\& \ D_2 < D_4) \\ & \parallel (D_R < R \ \&\& \ D_1 < D_2 \ \&\& \ D_1 \geq D_3) \\ P_{i_2}(x, y) & \text{if } (D_R \geq R \ \&\& \ D_1 < D_2) \\ & \parallel (D_R < R \ \&\& \ D_1 \geq D_2 \ \&\& \ D_2 \geq D_4) \\ & \parallel (D_R < R \ \&\& \ D_1 < D_2 \ \&\& \ D_1 < D_3) \end{cases} \quad (22)$$

となる。ある人物の存在する地点において、他の人物の存在できる確率は抑制されると仮定する。それにより、人物 i の存在確率を、

$$P_i(x, y) = P_i(x, y) - \left(\sum_k^N P_k(x, y) \mid k \neq i \right) \quad (23)$$

とする。室内に存在する全ての人物によって求められる存在確率の分布

$$EM(x, y) = \max_{1 \leq k \leq N} P_k(x, y) \times 2 - \sum_{i=1}^N P_i(x, y) \quad (24)$$

を存在確率マップと呼ぶ。存在確率が 0 となる場所は、視体積交差により投票されたとしても、ノイズもしくは人物以外のボクセルであり削除される。そのため、人物位置をグローバルな空間座標で保持し、投票空間を個人毎にローカルに持つことで、投票時のロス無くすることができる。図 14 に 2 名を例に模式的に表した存在確率の分布を示す。図 15 に存在確率マップを示す。各人物の存在確率 P の山は同じ高さの位置で交差するため、人物間で原理上 0 の値を持ち、人物同士が接触し同位置に存在しない限り、人物同士のボクセルの分離が可能である。

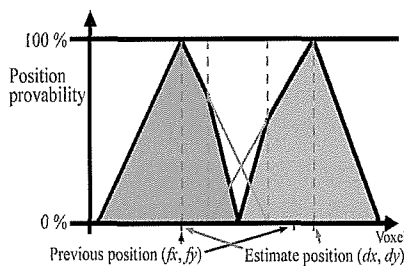


図 14. 存在確率の分布

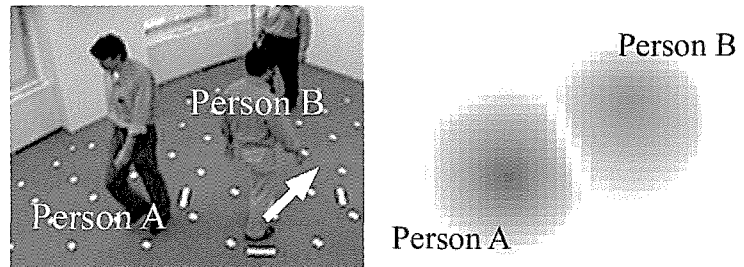


図 15. 存在確率マップ

投票時には、各人物の存在確率に合わせて半径 R のサイズの投票空間を個人毎に構築する。人物 i の投票空間 V_i は、室内全体の空間 V に対応する座標を保持する。投票度数は式(16)より、

$$V_i(x, y, z) = \frac{100}{N} \sum_{c=0}^N Vc(x, y, z) \times P_i(x, y) \geq th2 \quad (25)$$

となる。

フレーム間において、同一人物の追跡が適切にされているか人物特徴を用いて判定する。人物特徴には投票によって検出された人物ボクセルにおける色分布を利用した。検出された各ボクセルと対応するカメラ画像の画素は自明である。しかしながら、フレーム間で対応する部位のボクセルを推定することは、身体の向きが変わることから困難であり、各ボクセルの値をそのまま扱うことはできない。そのため、図 16 に示すように、人物の高さ（足先から頭頂まで）毎の色分布を人物特徴として採用した。この人物特徴を各フレームにおいて人物毎に計算し、前フレームと比較する。相関値が低ければ同一人物でない可能性が高いため、追跡に失敗しているとし、人物検出を再度行う。

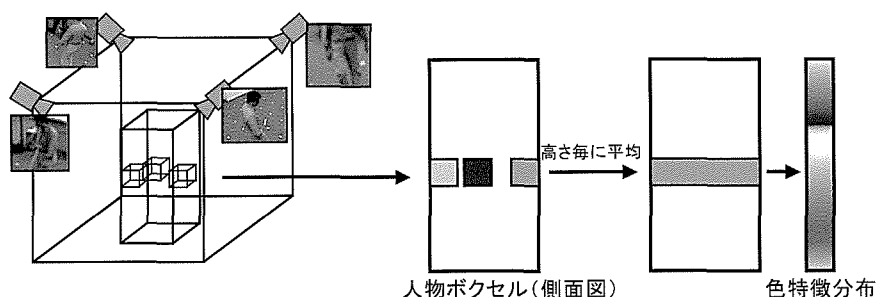


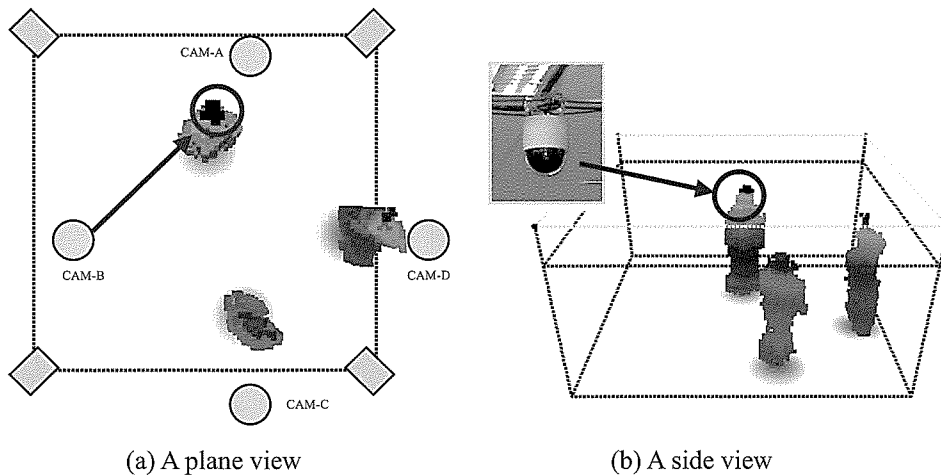
図 16. 同一人物判定のための人物特徴

3.2.3 顔画像の取得と人物属性の推定

マーケティング情報として有効な人物属性情報の推定、即ち性別や年代の推定、顔向きを検出といった顔画像処理では、ある程度以上の画像解像度を必要としており、遠方より撮像する室内全体を見渡すような固定カメラでは精度の高い推定は困難である。視体積交差法による人物追跡手法では、人物の像を3次的に検出できることから、頭部の3次元位置が容易に推定できる。従って、推定した頭部位置に別のカメラを向けることで、解像度の高い顔画像を取得することができ、性別や年代の推定精度を向上することができる。

本システムにおいては、解像度の高い顔画像を取得できるよう、ドーム式コンビネーションカメラを天井に設置した。4台を図17(a)のように配置することで、正面±45°以内の顔画像が取得できる。

ドーム式コンビネーションカメラの制御はRS232Cによって制御を行っている。人物検出を行ってから頭部にカメラを向けるという一連の制御にかかる時間は短時間ではあるものの、人物の移動速度に追いつくことは困難であり、移動中の顔画像処理の精度は低下する。そこで本システムでは、マーケティング情報として性別や年代の推定結果を得るため、室内への入室時の他、商品や展示物など対象物を注目している際に、ドーム式コンビネーションカメラをパン・チルト・ズーム制御し、顔画像処理に必要な解像度の顔画像を取得することとした。



(a) A plane view

(b) A side view

図17. ドーム式コンビネーションカメラによる顔の撮影

ドーム式コンビネーションカメラにより得られた顔画像から、肌色情報を用い、顔領域を検出する。この顔領域からテンプレートマッチングと弛緩整合方による部品検出を行い、目、鼻、口といった顔部品を検出する。これら顔領域、顔部品領域の情報を元に四方向面特徴を用いた線形判別によって性別と年代を推定する。以上の人物属性を推定するための顔画像処理技術はプロトタイプAMIS同様、「顔画像データベースの構築」、「顔及び頭部の検出に関する研究」、「性別および年齢の推定に関する研究」、「顔向き推定に関する研究」で確立した技術を用いている。

4 結果

4.1 自動マーケティング情報システム（プロトタイプ）

図18に示す2つの観光地宣伝映像を提示し、情報を収集するAMIS（プロトタイプ）を構築した。同時に最大4人までの注目度と人物属性を収集する。プロトタイプシステムでは、システムの収集状況が確認できるよう、推定した人物の属性を、似顔絵により注目している宣伝映像の下に表示する。なお、図18では50歳前後、40歳前後の人物が各宣伝映像に注目している。

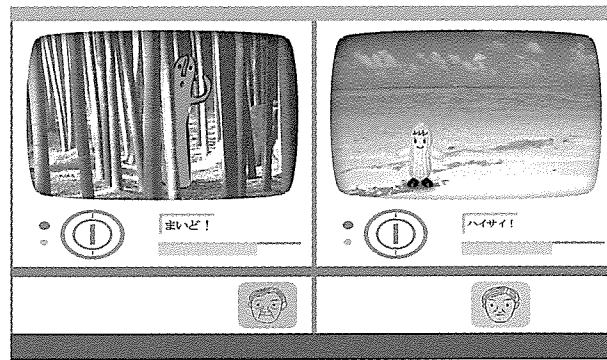


図 18. AMIS による広告映像の提示例

図 19、図 20 に複数人物の追跡結果を示す。4 名の人物が順に PDP の前に現れる動作を対象に行った。図 19 において、ラベルは同一人物であることを示す。また、図 20 は PDP の前空間における人物位置を推定した追跡結果である。フレーム 368~667 において人物 B は、人物 A の背後を通るため消失する。またフレーム 877~985 において人物 D も同様に消失する。本実験において、短時間の消失のため追跡可能であった。また、フレーム 733 において人物 C は、人物 B との距離があまりに近すぎるため、ボクセル空間で結合してしまうが、この場合も再度離れることで、再検出が可能であった。なお、本実験において、処理時間は平均 400msec であった。

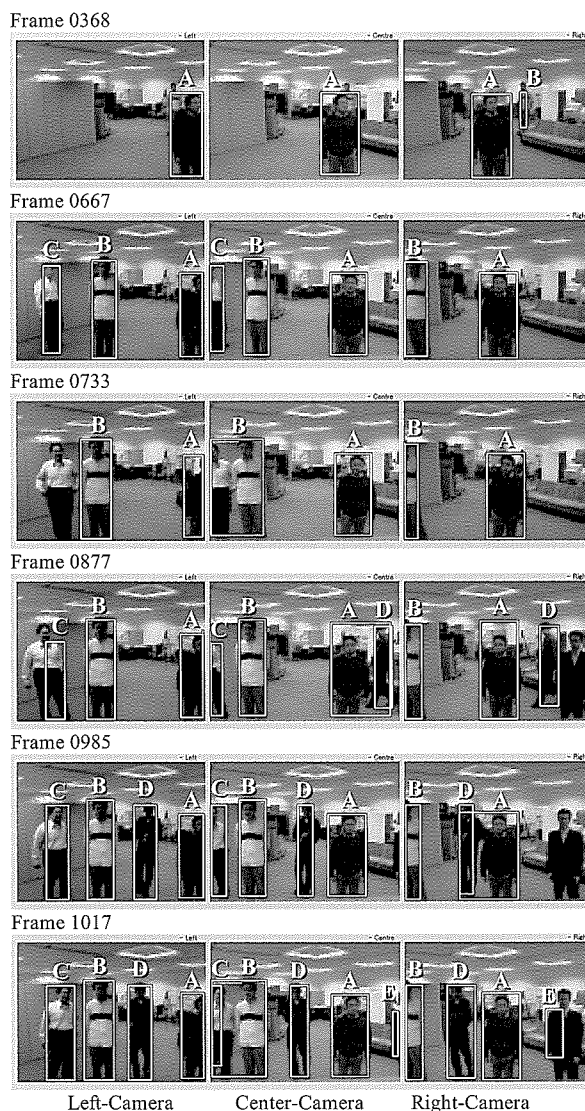


図 19. フレーム間での人物同定結果

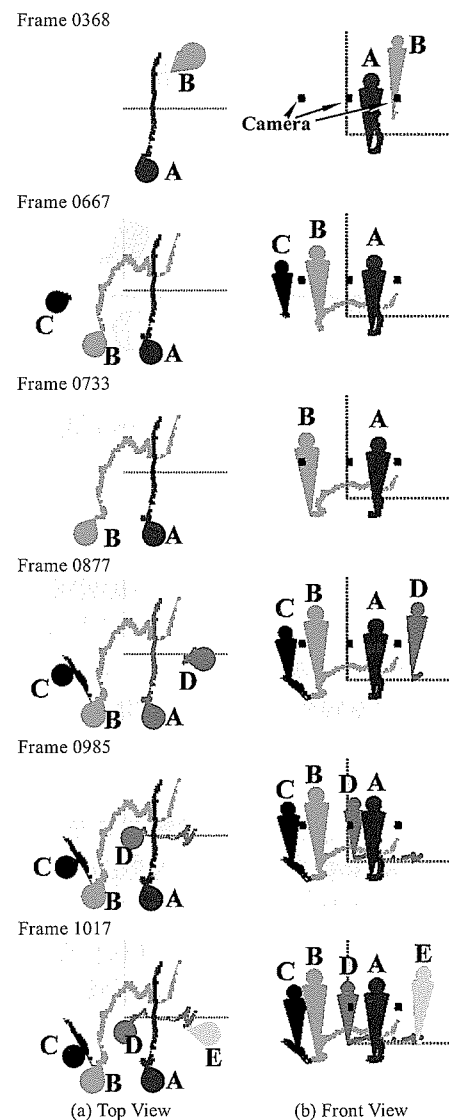


図 20. 人物追跡結果

図 21 に中央カメラの人物領域に対し、顔検出を行った結果を示す。図 22 に顔領域内の画素の平均距離から空間位置を求め、パンチルトカメラ (PT-カメラ) を制御した結果を示す。各人物の高解像度顔画像が取得できていることが確認できた。また、図 23 に顔部品 (両目、鼻、口) の検出結果を示す。

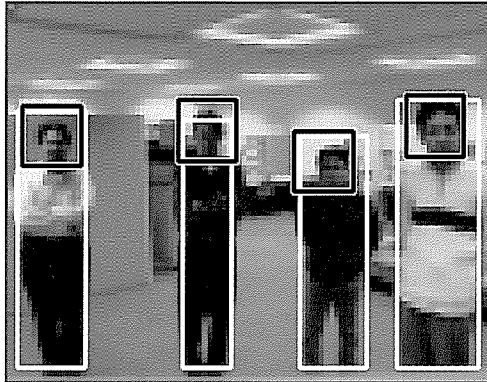


図 21. FDF と SVM による顔検出結果

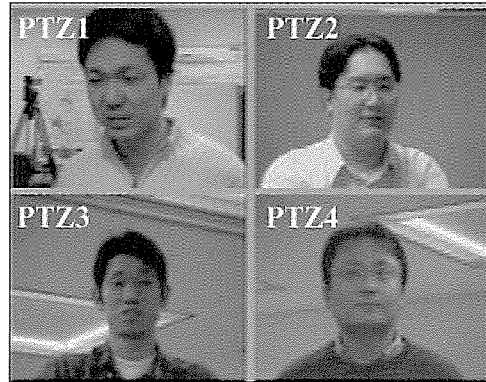


図 22. PT-カメラによる顔画像

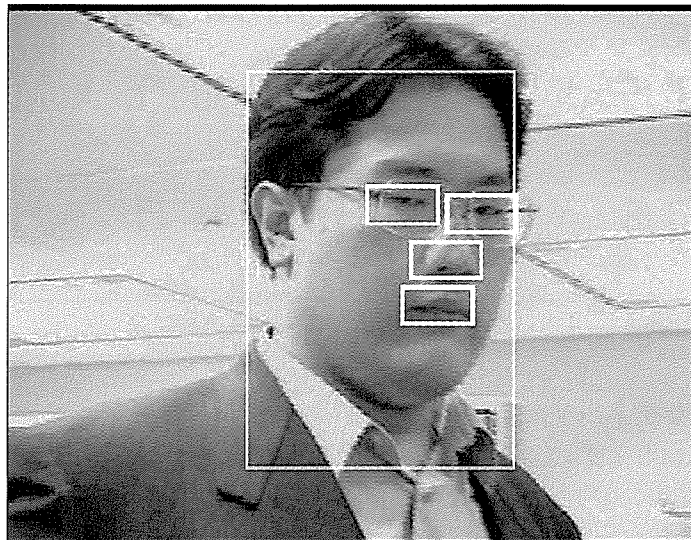


図 23. 顔部品検出結果

また、AMIS に搭載する性別・年代の推定部において、HOIP 顔画像データベース (300 名、15 歳～64 歳、5 歳間隔に男女各 15 名ずつ) の正面顔に対し実験を行った。性別推定実験においては、年齢順に並べた 300 名を奇数番および偶数番の 150 名ずつのグループにわけ、一方を学習人物に、もう一方を未知人物とした。結果、男女混在分類において、顔領域で学習人物 95.4%、未知人物 85.3% を、顔部品領域で学習人物 98.0%、未知人物 90.3% の推定率を得た。

年代推定実験においては、データセットの各クラスを均等に 5 等分し、4 セット (80%) を学習人物、残りの 1 セット (20%) を未知人物とし、Leave-one-out 法に基づき評価した。識別クラスを 10 歳区分としたところ、顔領域で学習人物 73.8%、未知人物 35.7% が、顔部品領域で学習人物 68.8%、未知人物 33.3% の推定率が得られた。

図 24 は AMIS により実験室内で約 2 ヶ月間情報を収集した結果である。各映像がどのような性別、年代 (10 歳間隔で 20 歳前後から 60 歳前後まで) の人物に注目されているか、似顔絵とともに、総注目時間をもって逐次確認ができる。今回は総注目時間を示したが、収集した情報をもとに多様な分析が可能である。更に、顧客への付加価値として、収集結果をもとに顧客に応じた広告展開が可能である。プロトタイプ AMIS では、一定の注目が集まると注目度の高い側の詳細な情報を提示するように構築した。

本システムでは、一方向からの撮像のため、オクルージョンが頻繁に発生する。そのため、ある一定時間以上消失すると、人物追跡は困難になり、再び現れた場合新たな人物として検出する。今後、顔検出に用いた特徴や、性別・年代推定結果を基に、消失後に同一人物かどうかの検証を行うことで追跡精度を向上することは可能であると考えられる。

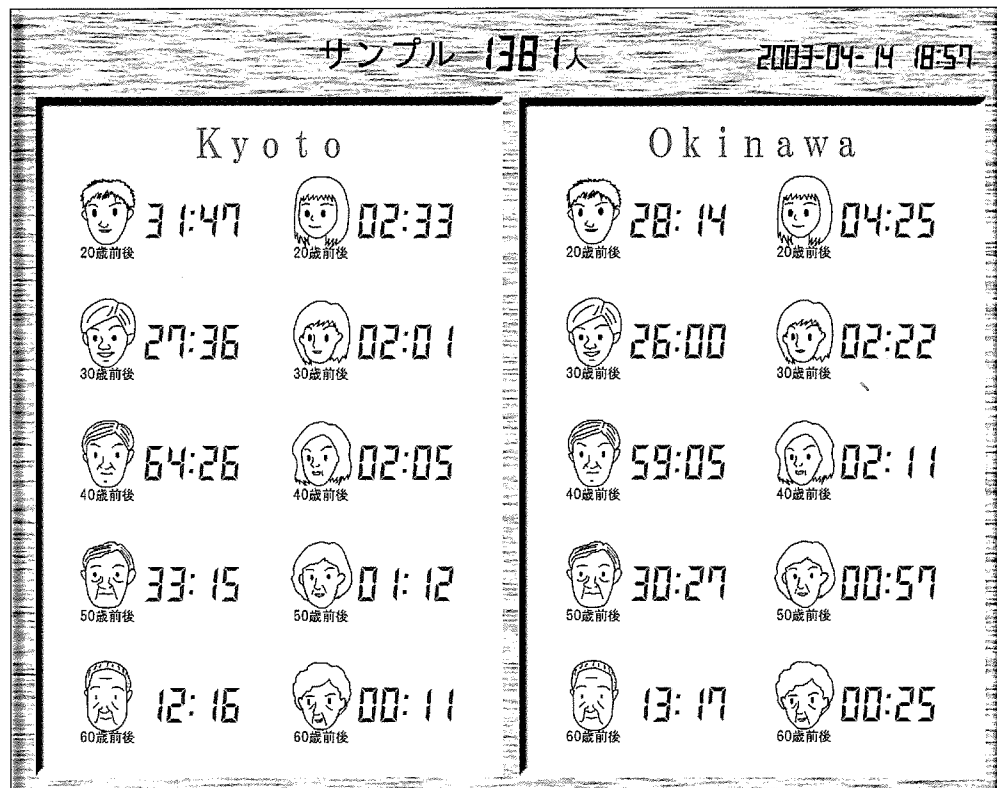


図 24. 情報収集結果例

4.2 室内型自動マーケティング情報システム

構築した室内型 AMIS において、実時間での複数人物追跡の実験を行った。図 25(a)にカメラ 3 からの映像を、図 25(b)に人物検出結果を示す。視体積交差法を用いることで、床面上の位置だけでなく、人物の三次元的な像が検出できる。

図 26 に複数人物追跡結果を示す。床面に示す線は各人物の動線である。人物同士が近づき交差しても、良好な追跡が確認できた。図 27 に Frame579 における各カメラ画像の人物領域同定結果を示す。各ラベルは同じ人物を示す。カメラ 2 にて別の人物によって遮られている人物 D や人物 G の検出も確認できた。人物 G は座ったり立ったりする動作を行ったが、このような動作に対しても良好に追跡することができた。

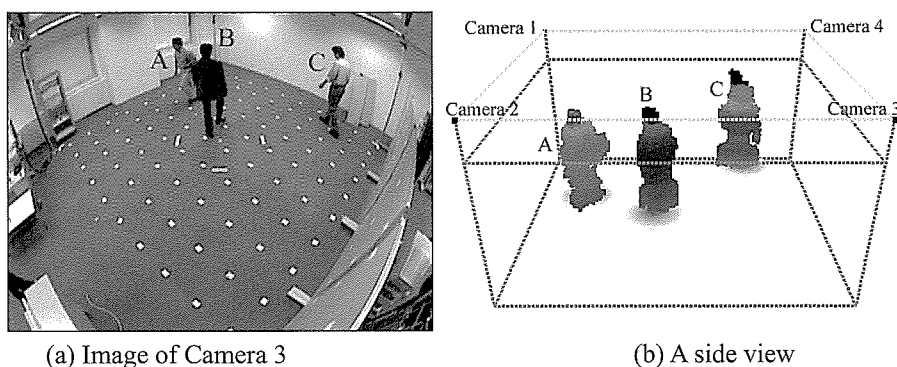


図 25. 複数人物検出結果

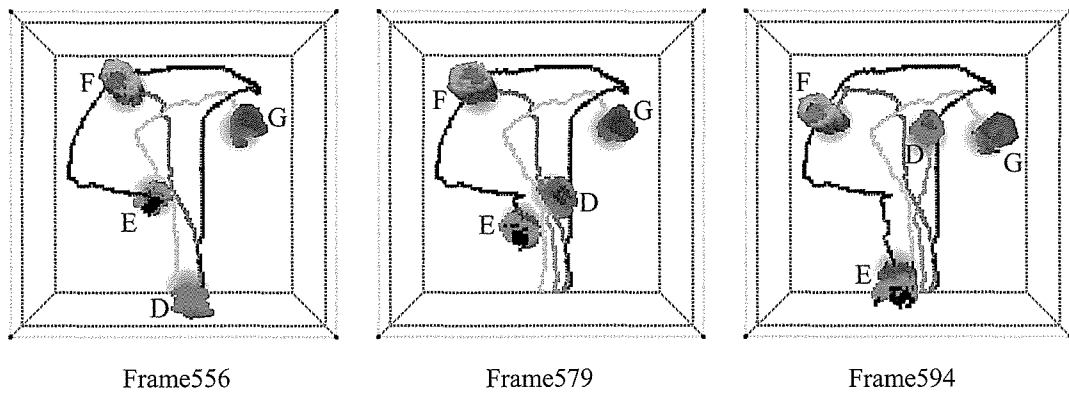


図 26. 複数人物追跡結果

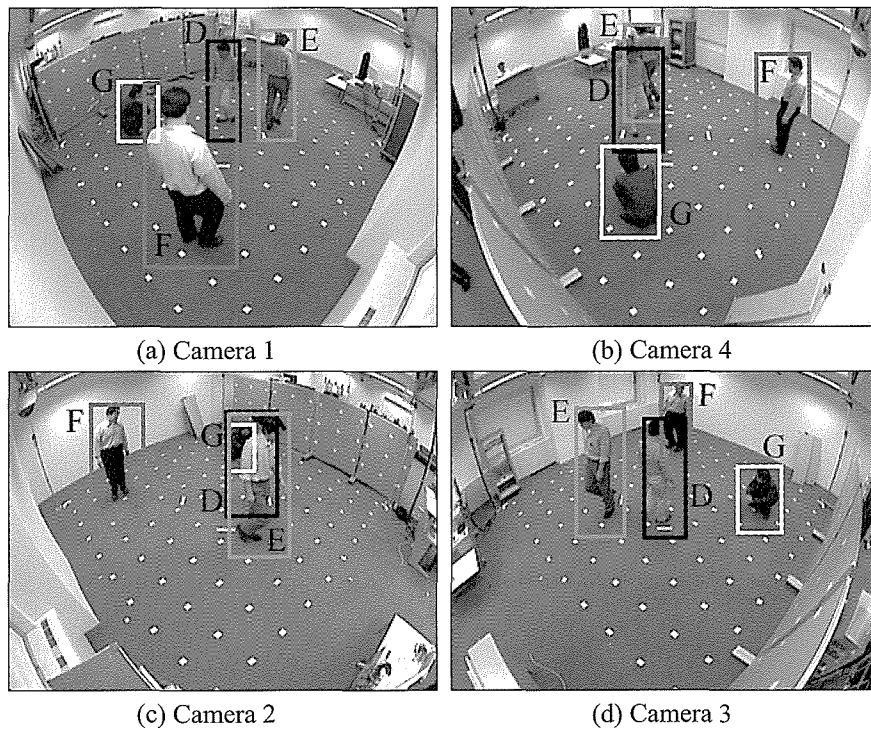


図 27. 人物領域同定結果 (Frame 579)

室内に幾つかの品物を配置し実験を行った。図 28 は品物をレイアウトした実験室の様子である。手前入口両側に処理を行う PC を配置し、室内には、左壁面にぬいぐるみを、左手奥に自転車を、右手奥に処理結果の映像とマネキン、右壁面にパンフレットを、部屋中央に円形のテーブルが配置してある。



図 28. 室内レイアウト

図 29 にドーム式コンビネーションカメラを頭部位置へ制御し、顔画像を取得した結果を示す。移動方向や肌色情報から正面方向のカメラを推定し、撮像するカメラを選出することもできるが、図 12 には 4 台全てのカメラにより一人の人物の顔画像を集中して取得した結果を示している。図 13 に得られた正面方向の顔画像から性別・年代を推定した結果を示す。この実験では、男性・30 歳前後と良好な結果が得られた。

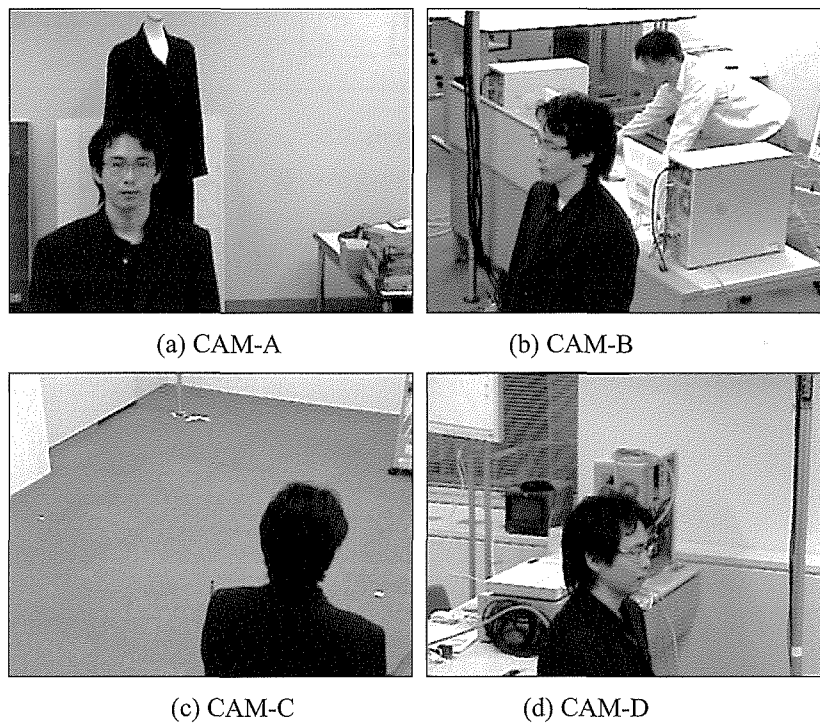


図 29. 顔画像の取得

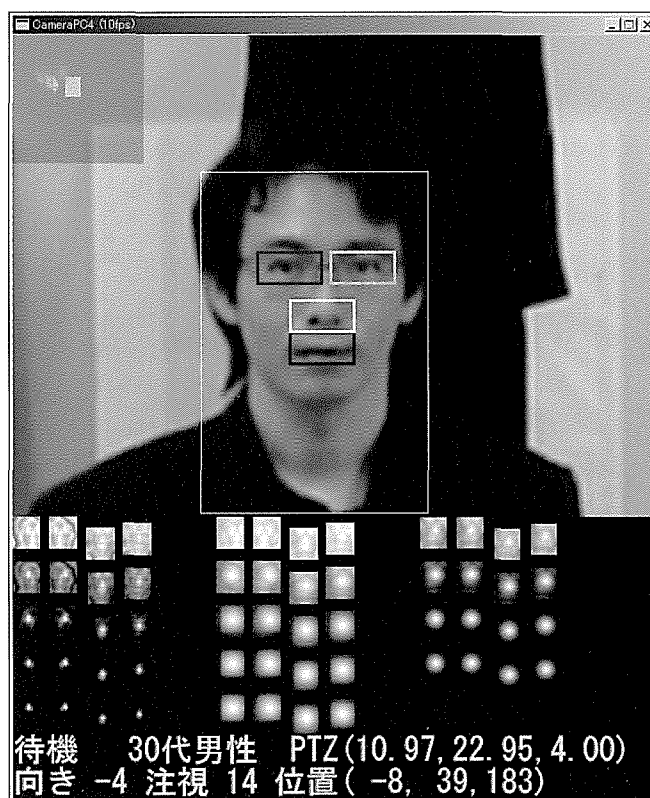
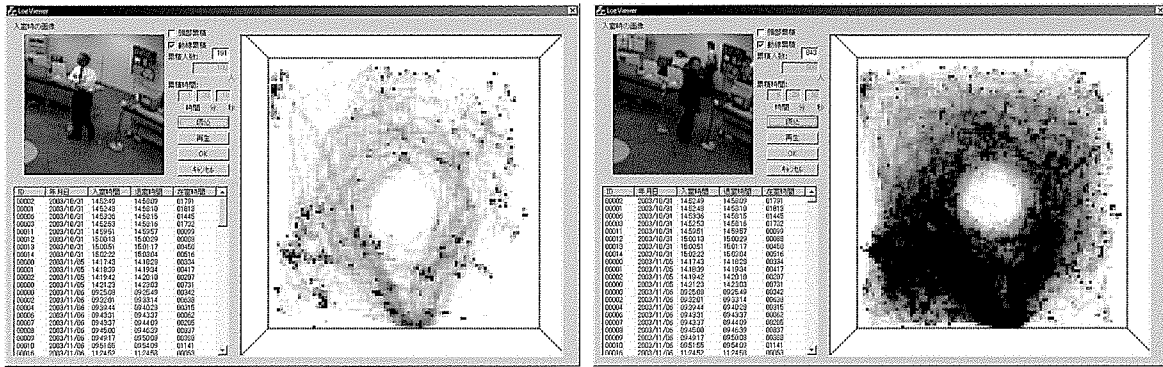


図 30. 顔部品抽出と年代・性別の推定結果

期間約5ヶ月間計測を行った。図31(a)は100人(累積時間約1.5時間)の動線を累積した結果であり、図31(b)は500人(累積時間約8.5時間)の動線を累積した結果である。画像下側が入口になり、青の濃淡値が人物の存在頻度を表す。100人の時点から、入口およびぬいぐるみ周辺に人が集中していることが分かる。500人の時点では、何も置いていないパンフレットの手前(右下のコーナー)がデッドスペースになっていることが確認できた。リアルタイムの複数人物追跡と共に、このような統計情報を収集することで、室内レイアウトに反映し、スペースの有効利用や、人物流動量の計測など様々な活用ができると考えられる。

構築したシステムでは、4名の人物を対象に、1フレームあたり約0.2秒の処理速度で安定した人物追跡および属性情報の収集が可能であった。今後、更に多くの人物に対応する予定である。



(a) 100人(累積時間約1.5時間)

(b) 500人(累積時間約8.5時間)

図31. 動線の累積(人物の存在頻度)

フェーズ III

今後の取り組み

これまでの研究において、人物追跡技術を確立し、実利用を目指した応用モデルとしてプロトタイプ自動マーケティング情報システムおよび室内型自動マーケティング情報システムを構築し、実験と共に有効性を示した。今後の取り組みとして、より多くの人物に同時に対応するとともに、照明条件や背景変動へ対応することで、更に現実的なものとしていくことが必要と考えられる。また、近年個人情報保護の問題が活発に論議されるようになり、収集した情報の管理とともに、どこまでの画像処理、どこまでの人物情報の収集が許されるかなどを検討していく必要がある。