ヒューマンインタフェース　95−11
コンピュータビジョンと
イメージメディア　129−11
（２００１．９．１３）

「人を見るシステム」のための人物追跡と頭部の分類

馬場 功淳† 大橋 健† 乃万 司† 松尾 英明†† 江島 俊朗†

†九州工業大学

††松下電器産業株式会社
九州マルチメディアシステム研究所

あらまし　我々は，人物を観察しその人物が望んでいることを先回りして示すことができたり，人物の統計情報を提供できる「人を見るシステム」を構築したいと考えている．このような「人を見るシステム」は，環境の変化にロバストで信頼性の高い人物追跡器と，その人物の情報を的確に得ることができる人物特徴推定器が必要である．我々は，この 2 つを HeadFinder と HeadClassifier というシステムで実現した．HeadFinder はフレーム間差分をベースとする人物追跡器であり，HeadClassifier は SVM や Boosting などの統計的認識手法を用いた人物特徴推定器である．本稿は，これら 2 つのコンポーネントについて述べ,「人を見るシステム」のアーキテクチャーについて考察する．

Person Tracking System and Head Classification System for Looking at People System

Naruatsu BABA† Takeshi OHASHI† Tsukasa NOMA† Hideaki MATSUO†† Toshiaki EJIMA†

† Kyushu Institute of Technology

††Kyushu Multi Media System Laboratory,
Matsushita Electric Industrial Co.,Ltd.

Abstract　Our goal is to build “looking at people system”. Looking at people system can act intelligently, and can get person’s statistics information. Looking at people system needs two faculties. One is a person tracking system that is robust to environmental change. This faculty has been realized by the system called HeadFinder using frame difference. Another is a system which has the ability to extract a key feature of the person. This faculty has been realized by the system called HeadClassifier using SVM and Boosting. In this paper, we propose an architecture of looking at people system.

1　はじめに

人を見る技術が注目されている [1]．カメラで撮られた映像をもとに，人が居るのか居ないのか，動いているかいないのか，あるいは笑っているのかいないのか等をコンピュータが自動的に判断する技術である．

このような，人を見る技術は，様々な活用例が考えられる．

- 人物データベース
  今日入場した人のうち 40 歳以上の男性をリストアップせよ，という命令が実行可能なシステム．

- 監視システム
  人物特定を行い，登録者以外の立ち入りを禁止するシステム．

- 人に快適な環境の自動調整
  その空間に子どもが多いとテレビを子ども向け番組に変更したり，上着をきている人が多いと部屋の温度を上げるなどが実現可能なシステム．

我々は，このような，人物を観察しその人物が望んでいることを先回りして示すことができたり，統計情報として有用なデータベースを作成することができる「人を見るシステム」を構築したいと考えている [11]．

上に挙げるような「人を見るシステム」を構築するためには，実時間で動作する２つのコンポーネントが必要である．長時間の使用に耐える環境の変化にロバストな人物追跡器と，その人物から様々な情報を抽出する特徴推定器である．我々はこれらの考えに基づいて，人物追跡システム HeadFinder[3][4] と頭部分類システム HeadClassifier[8][9][10][12] を作成した．HeadFinder は，フレーム間差分をベースとする人物追跡器である．対象が単数人であれば，屋内・屋外・近赤外線環境において 90 %程度の高い認識率で動作する．また，人物の重なりなどのオクルージョン時の人物同定も可能である．一方，HeadClassifier は，SVM[7] や Boosting[5] といった統計的認識手法を用いて頭部画像を分類する人物特徴推定器である．

本稿では，これら 2 つのコンポーネントについて述べ,「人を見るシステム」の実現の可能性について考察する．

2 人物追跡システム HeadFinder

人物追跡システム HeadFinder は、カメラ画像から入力された動画像データから動きのある円形の物体を抽出し、それを人物の頭部とみなして検出し、追跡するシステムである．このシステムはパーソナルコンピュータと単眼カメラから構成される．双方ともに一般に普及している安価な製品で充分であり、人物マーカなどの特殊な機器は一切必要としない．

処理の流れを述べる．まずカメラ出力を一定間隔でサンプリングし、それをもとにフレーム間差分によるエッジ情報を作成する．次にエッジの輪郭外形成分を抽出し、動き予測を考慮した Hough 変換を適用して、人物頭部領域を検出する．Hough 変換による投票数がもっとも高い領域を頭部と判断し、モニタに表示する．PentiumIII マシンで本システムを実行した場合、毎秒 22 フレーム程度の処理が可能である．個々の処理の詳細は次節で解説する．

2.1 フレーム間差分によるエッジ検出

フレーム間差分とは時刻 t の入力画像と時刻 $(t-1)$ の入力画像との差分である．人物のようなある程度の大きさを持つ物体が動いた場合にはフレーム間差分として人物領域と背景とのエッジ部分が検出される．この手法による対象のエッジ検出は照明や背景の変化の影響を受けにくいため、実環境では動きエッジの検出には非常に有効な方法である．反面、この手法には対象に動きがない場合にエッジ情報が得られないという欠点がある．ただし動きが検出されない場合は対象物体は以前の場所に留まっていると推測できるので、本システムにおいてその点は問題とならない．

2.2 Hough 変換による円形状検出

人物の輪郭は姿勢や個人差によって異なるため、全身像を輪郭情報から抽出することは難しい．しかし、姿勢や個人差による変化が比較的少ない頭部ならば輪郭からでも抽出が可能である．

HeadFinder は頭部検出の前提として、「動く円は頭部である」とみなす．円の抽出には Hough 変換を用いるが、しかし頭部は厳密な円ではないために抽出処理が失敗する可能性がある．この解決策として、HeadFinder では一定の幅を持つ円形の枠をテンプレートとする Hough 変換を行うことで、頭部の傾きや個人差による輪郭情報の正円からのずれを吸収している (図 1)．Hough 変換の投票空間は $(c^{(x)}, c^{(y)}, r)$ の 3 次元空間である．$(c^{(x)}, c^{(y)})$ は円の中心点、r は円の半径である．すなわち、画像中の (x, y) という場所に輪郭を構成する画素が見つかったとすると

$$(x - c^{(x)})^2 + (y - c^{(y)})^2 = r^2 \tag{1}$$

を満足する $(c^{(x)}, c^{(y)}, r)$ に投票する．ただし、

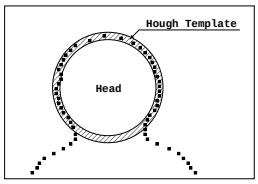

図 1: Hough 変換のテンプレート

処理の高速化を図るために予め、8 通りの半径 $(\{r_0, r_1, ..., r_7\})$ の幅のある円 (円枠) のテンプレートを記憶しておき、投票は各テンプレートの中心を見つけた画素 (頭部を構成する輪郭の候補点) に合わせて投票平面に足し込む作業になる．この作業は，PC のマルチメディア系命令の SIMD 型整数演算に適しており，高速に処理することができる．

2.3 輪郭情報

十分短い時間で処理されたフレーム間差分は，動物体のエッジを検出していると考えて良い．つまり人物であれば，輪郭情報だけでなく顔と髪の境界線や服のテクスチャにも反応してしまう．これは，Hough 変換の処理を行った場合に多数の円候補を生じることになり，検出率低下につながる．

本システムは，純粋な輪郭情報のみを検出するため，差分情報を x 軸・y 軸の全方向より走査し，最初に検出された差分情報を人物の輪郭とする．

このとき問題となるのが，どこから走査するのかである．人物を覆うような矩形を設定し，その外形線より走査するのが最適である．しかし，そのシーンに複数の人物が存在した場合，矩形をどこに設定するかが極めて難しい問題となる．差分情報のみから人物を覆うような矩形を設定する方法として，x 方向の差分の切れ目を利用する手法などが考えられるが，人と人が重なっていた場合困難である．

そこで，差分情報 $f(x)$ の最大値 $p(x)$

$$f(x) = (y_1, y_2, y_3, \ldots, y_n) \tag{2}$$

$$p(x) = max(f(x)) \tag{3}$$

に対して空間ローパスフィルターを適用し、「波高が高い部分には人がいる」という特性を利用する．つまり，波形の谷の位置で領域を分割する (矩形の一辺にする)．これにより，ある程度人が重なった場合においても対応できる (図 2)．

2.4 キューによる信頼度の計算

本システムは，前回の情報を基に次回の頭部位置を予測し，ある大きさのウィンドウで絞り込みを行なう．ウィンドウが小さいと処理すべき対象が減り，認識率の向上が見込まれる．しかし，前回の情報に誤りがあった場合，人物がいない場所のみを検索する可能性がある．

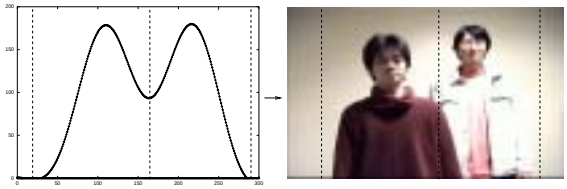

図 2: 領域の分割

適切に予測を行なうには，基となる情報がどれだけ信用できるかという指標が必要である．本システムは信頼度という概念を導入している．これは頭部位置の時系列的な連続性を仮定した上での推定量である。HeadFinder は人オブジェクト P_i が複数存在し，それぞれが人物 1 人の情報，すなわち頭部の画像中の位置 (x, y)，頭部の半径 r，および現在の状態 S を管理する．つまり，人オブジェクト P に一情報として信頼度 C を追加する．

信頼度は，人物の動きが空間的及び時間的に連続していることを利用する．よって，なんらかの形で過去の情報を保持しなければならない．本システムは人物の頭部の (x, y, r) 情報を要素とするキュー（queue）を採用した．キューの要素が空間的な近さを保証するように，キューの並びが時間的なつながりを保持するようにキューを構成すると，そのキューの長さが信頼度として利用できる．つまり信頼度 C の計算は，3 次元ベクトル $D = (x, y, r)$ を格納する長さ l のキュー Q を作成し，以下のアルゴリズムで Enque・Deque を行なう。

$$\begin{array}{ll} \texttt{if} & \min\{dist(D_{pre}, D_{new}) \mid D \in Q\} \ \leq \ T_d \\ \quad \texttt{then} & \textbf{ENQUE;} \quad (Q \leftarrow D_{new})^{w_1} \\ \quad \texttt{else} & \textbf{DEQUE;} \quad (Q \rightarrow D_{last})^{w_2} \end{array}$$

D_{new} は現在得られた頭部位置，D_{pre} はキューの中で最も新しい頭部位置，D_{last} はキューの中で最も古い頭部位置，$dist(D_1, D_2)$ は D_1 と D_2 の距離を返す関数，T_d は距離の閾値である．あるフレームにおいて動きが検出されない場合は無条件に Deque を行なう．ここで，キューに格納されているベクトル D の個数 n を信頼度 C と定義する．

この信頼度を基に予測ウィンドウの大きさを決める．つまり信頼度が高くなるにつれ予測ウィンドウを小さくする．信頼度が 0 であるなら，予測ウィンドウを一番大きく，すなわち人がいないとして画像全体をウィンドウサイズに決定する．また，w_1, w_2 は Enque・Deque を行なう場合の比率 $(R = w_1/w_2)$ である．R が大きいと，対象人物の動きの少ない場合や，特定の一人の人物を深く追跡する目的に適する．

2.5 複数人追跡

本システムは人オブジェクト P_i を複数保持しており，その P は 3 つの状態のいずれかに属する．

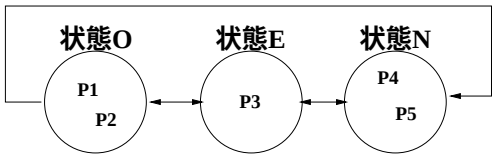

図 3: 各状態の結線

その状態とは「人がいない (状態 N)」「人がいる (状態 E)」「人と人が重なっている (状態 O)」であり，それぞれ図 3 のように遷移する．

あるシーンに人が入ってきた場合，システムは状態 N の中から任意のオブジェクト P を選び，その人と結びつけ状態 E に遷移させる．オブジェクト P はその人の情報 (位置情報・信頼度) を持ち，以降いなくなるまで (状態 N に遷移するまで) 保持される．また，信頼度がある閾値を下回った場合も状態 N に遷移する．

2.6 重なり時の追跡

複数人追跡を実現する場合，もっとも考慮しなければならないのは人と人の重なりである．本システムは基本的に重なり時の追跡は行なわない．しかし，重なりが解除された後の人物の同定に力をいれる．なぜなら,「人を見るシステム」のような高次な処理を考慮した場合，間違いなく同一人物として提供することが重要であるからである．

2.7 HeadFinder の性能評価

人物追跡システム HeadFinder は，表 1 に示す環境上に実装されている．平均フレームレートは単数人の場合は $22fps$ であり，複数人の場合は $16fps$ であった．また使用カメラは，一般の CCD カメラと近赤外線カメラを使用した．

2.7.1 歩行者の記録

「室内環境」「屋外環境」「近赤外線環境」という異なる三つの環境での実験を行い、以下の通行人検出率、記録画像適合率の二つの指標を用いてシステムの評価を行う．

通行人検出率　：＝　抽出人数/通行人数
記録画像適合率　：＝　頭部画像数/全抽出画像数

通行人検出率は人物検出に完全に失敗しなかった割合であり、記録画像適合率は全抽出画像の中で頭

表 1: HeadFinder の動作環境

計算機	AT 互換機
OS	FreeBSD-4.4
CPU	PentiumIII 750MHz
取り込み画像サイズ	300x200

表 2: 歩行者の記録実験結果

	屋内	屋外	近赤外線
通行人数 (a)	362	20	20
抽出人数 (b)	361	20	18
記録画像数 (c)	28280	559	1002
頭部画像数 (d)	25462	557	901
通行人検出率 (b/a)	99.7(%)	100.0(%)	90.0(%)
記録画像適合率 (d/c)	90.0(%)	99.6(%)	89.9(%)

部画像である割合である．通行人検出率における抽出人数は，人物がフレームインしてからフレームアウトするまでに一度でも検出に成功した場合は，通行人検出に成功したとしてカウントされる．なお，被験者にカメラの前を自由に歩くように指示をして，実験を行った．

結果を表 2 に示す． 結果，通行人検出率・記録画像適合率ともに 90 %程度の高い認識率であった．特に注目すべき点は，屋内環境実験の結果である．屋内環境はシステムを一度も停止せず，15 時間続けて実験を行った．この間，照明の ON・OFF が繰り返されたり，撮影領域に大きなダンボールが置かれたなど，環境の変化が激しい場所であったにもかかわらず，開始時と終了時の認識率に差はなかった．これは我々が目標とする「人を見るシステム」のコンポーネントとしての有効性を証明するものである．

2.7.2 エラー率と人物検出率

前節は，通行人記録システムとしての実験であった．このため，通行人検出率は人物検出に完全に失敗しなかった割合とした．つまり，ある人物がフレームインしている数十フレームのうち，1 フレームでもその人物を検出していれば成功としてカウントする．一方，人物がフレームインしている全フレームのうち，どれぐらいの割合で人物を検出しているかという人物検出率も重要な評価法である．通常，この人物検出率は，エラー率 (記録画像不適合率) と相関関係にあり，一方を上げれば一方が下がるという特性がある．このため，一意に人物検出率を示すのは適切でない．

よって，閾値などのパラメータを調整して，あるエラー率に対する人物検出率をプロットしたグラフ (Receiver Operating Characteristic(ROC) curve)[2] を用いて，HeadFinder の特性を示す．実験は，屋内環境，屋外環境で行った．結果を図 4 に示す．

2.7.3 複数人追跡時の重なり

人物がすれ違う行動には，図 5 で示すように 2 パターン存在する．完全にすれ違う行動 (行動 A) と，すれ違ってから戻る行動 (行動 B) である．実験は，この 2 つの行動を被験者 4 人で 20 回ずつ繰返し，計 40 回で評価する．

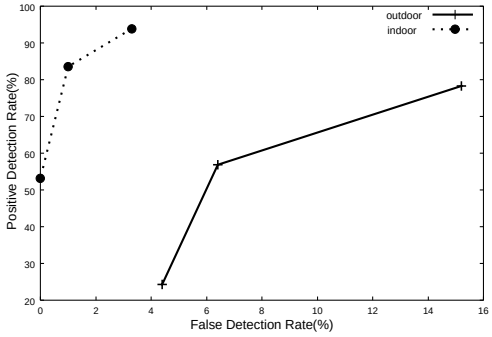

図 4: ROC curves

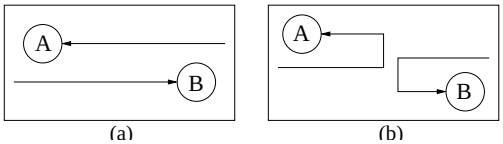

図 5: 上から見たすれ違いのパターン

結果は表 3 に示す通りである．行動 A は，重なりが発生する時間が短いため，検出率が良い．また人物の同定も，20 回中 1 回の誤りしか起こらなかった．行動 B は重なりあっている時間が長い場合が多く，差分未検出が続き評価値の低下をまねいた．その結果，検出率・同定率ともに低下した．

表 3: 実験結果

	頭部検出率 (%)	人物同定率 (%)
行動 A	98	95
行動 B	95	90

3 頭部分類システム HeadClassifier

小倉，松尾らが開発を行っている HeadClassifier は、HeadFinder から得られた頭部画像を様々な特徴を基に分類するシステムである．HeadFinder によって得られる頭部画像には背景などの雑多なノイズが多数含まれるために、従来のような繊細な特徴推定手法をそのまま用いることは不可能である。こうした悪条件のもとでもロバストな特徴推定を実現するため、SupportVectorMachine(SVM) や Boosting と呼ばれる統計的認識手法を用いて推定処理を行う．また，特徴量は，頭部画像のテクスチャそのものとする．具体的には，14 × 14 に正規化した 256 階調のグレースケール画像とした．

現在，顔の向き推定，顔非顔推定，男女推定の 3 つが実装されている．顔非顔推定は HeadFinder の誤認識を検出する目的を持っており，男女推定は顔の向き推定の結果が，正面を向いている時にのみ

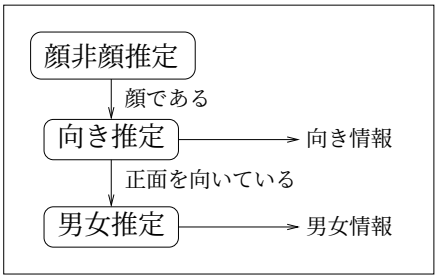

図 6: HeadClassifier の構成

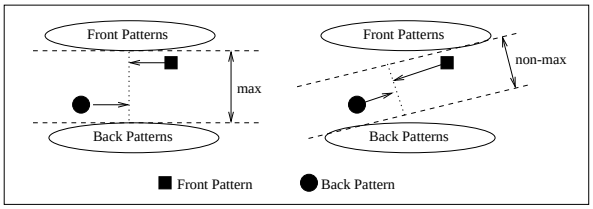

図 7: マージン最大化

行うよう実装されている．つまり，HeadClassifier は図 6 のような構成で動作する．

3.1 向き推定

前後，左右の顔の向きを識別する 2 つの推定器を用意し，その推定器の各々の出力が作るベクトルを顔の向きとする．つまり，入力頭部画像をベクトル x とし，前後を識別する推定器の出力を $g_h(x)$，左右を識別する推定器の出力を $g_v(x)$ とすると，これらを成分とするベクトル $\vec{d} = (g_h(x), g_v(x))$ を頭部の推定方向として出力する．これにより，前後左右に限らず，任意の向き方向を推定できる．

この手法は各推定器が入力頭部画像の方向に応じた尤度を返すという仮定のもとに成り立っている．これは，マージン最大化という特徴を持つ SVM や Boosting などの識別器において意味を持つ．与えられた特徴ベクトルは，分割面と垂直に位置する軸に射影され，その距離 (尤度) を返す．この時，マージンが最大となるような分割面を設定すると，$(g_h(x), g_v(x)) = (\cos\theta, \sin\theta)$ となることが期待できる (図 7)．

3.2 HeadClassifier の性能評価

HeadFinder によって得られた頭部画像 3600 枚を分類し，各推定器 (前後，左右，男女，顔非顔) で学習を行った．テストデータは，学習データに含まれていないものを 12145 枚選定し，実験を行った．なお，識別器は，SVM，AdaBoost[6] を用いて実装した．結果を表 4 に示す．本来，向き推定は 360 度以上の粒度で結果を返すが，ここでは前後左右の 4 方向に正規化し実験を行った．

結果は，SVM，AdaBoost ともに同程度の識別率であった．しかし，SVM は Support Vector の数が増加し計算コストが高くなる傾向があり，AdaBoost は SVM と比べて半分程度の計算コストであった [12]．

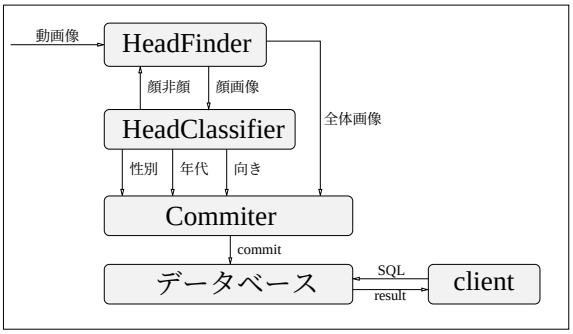

図 8: データベースとコンポーネント

表 4: HeadClassifier

	向き推定	男女推定	顔非顔推定
AdaBoost(%)	95.4	91.9	91.0
SVM(%)	95.7	92.3	91.2

4 「人を見るシステム」について

4.1 並列処理

前節までで「人を見るシステム」のコンポーネントとなる，HeadFinder と HeadClassifier について述べた．現在，この 2 つは顔画像などを通信することにより，各々独自に動作する仕組みとなっている．よって，通信の信頼性が保証されている場合，実行されるマシンが物理的に離れていても問題はない．これは負荷分散を容易に実現でき，実時間処理が必要な「人を見るシステム」において非常に重要な要素となる．

4.2 データベース

「人を見るシステム」を実現するために，HeadFinder や HeadClassifier などのコンポーネントから得られる情報をデータベースに格納する必要がある．この時，各コンポーネントは非同期で動作するので，人物に対する一貫性を保証する仕組みを与えねばならない．よって，データベースにコミットする専用のコンポーネントが必要であろうと考える．このコンポーネントを Commiter と名付けるならば，図 8 のような構成になる．

4.3 各コンポーネントの課題

1 節で挙げた「人を見るシステム」を実現するために必要な機能と，それに対する現状と課題を以下に述べる．

- 実時間処理
  HeadFinder が $15fps$ 程度，HeadClassifier が $20fps$ 程度の処理時間で動作する．すでに述べた通り，負荷分散が容易に実現できる構成なので，各々で実時間処理が保証されている場合は，全体としても実時間処理が可能である．

- 環境に依存せず，長時間の動作保証
  HeadFinder は，フレーム間差分をベースとした人物追跡器である．よって，環境の変化に比較的ロバストであり，長時間の動作には適している．しかし，車が通る道路を背景にした場合や，あまりにも人が込み合った空間などにおいての使用は，現在のところ不可能である．つまり，環境依存なシステムといえる．先に挙げた具体例の中には，現在追跡不可能な場所での使用を想定しているので，人物追跡器である HeadFinder の性能向上が望まれる．

- 信頼性の高い特徴推定
  HeadClassifier の性能を表 4 に示した．いずれも 90 %以上の高い識別率であった．「人を見るシステム」が実際に参照する場合は，voting やスムージングなどの工夫を施すことにより，さらに精度良く特徴推定が実現できるであろう．また，年代推定，個人特定については，HeadClassifier に実装する予定である．

5 まとめ

本稿は，まず,「人を見るシステム」の具体例を挙げ，我々が目指すシステムを明らかにした.「人を見るシステム」は，人物を観察しその人物が望んでいることを先回りして示すことができたり，統計情報として有用なデータベースを作成できるものと定義する．そして，その「人を見るシステム」のコンポーネントとなる人物追跡システム HeadFinder と，頭部分類システム HeadClassifier について述べた．HeadFinder は対象が単数人であれば，屋内･屋外・近赤外線環境において 90 %程度の高い認識率で動作することを確認した．また，HeadClassifier は，SVM や Boosting といった統計的認識手法を用いて頭部画像を分類する人物特徴推定器であり，顔の向き推定，顔非顔推定，男女推定において 90 %以上の識別率で動作することを確認した．最後に,「人を見るシステム」を構築するための課題，問題点を示した．

参考文献

[1] Alex Pentland “Looking at People: Sensing for Ubiquitous and Wearable Computing” *IEEE Transactions on Pattern Analysis and Machine Intelligence* , Vol. 22, No. 1, pp. 107-119, January 2000

[2] Anuj Mohan, Constantine Papageorgiou, and Tomaso Poggio, “Example-Based Object Detection in Images by Components” *IEEE Transactions on Pattern Analysis and Machine Intelligence* , Vol. 23, No. 4, pp. 349-361, APRIL 2001

[3] 馬場功淳、大橋健、乃万司、松尾英明、江島俊朗”HeadFinder:フレーム間差分をベースにした人物追跡” 第 6 回画像センシングシンポジウム２０００,pp329-334,2000

[4] 馬場功淳、松尾篤、大橋健、乃万司、松尾英明、江島俊朗”HeadFinder:単眼視動画像を用いた複数人追跡” 第 7 回画像センシングシンポジウム２００１,pp363-268,2001

[5] Yoav Freund and Robert Schapire”A Short Introduction to Boosting” *Journal of Japanese Society for Artificial Intelligence* ,14(5),771-780,1999

[6] Yoav Freund and Robert E “A decision-theoretic generalization of on-line learning and an application to boosting” *Journal of Computer and System Sciences* ,55(1),pp.119-139, August 1997.

[7] Nello Cristianini and Jhon Shawe-Taylor “An Introduction to Support Vector Machine and other Kernel-Based Learning methods” Cambridge University Press(2000)

[8] 松尾 篤、馬場功淳、大橋 健、江島俊朗 ” 人物追跡システムにおける顔の向き推定 “第 21 回バイオメカニズム学術講演会, 2B21 ,2000

[9] 郷原邦男、松尾篤、馬場功淳、江島俊朗”HeadFinder を用いた頭部の向き推定” 電気関係学会九州支部第５３回連合大会,pp.1127,2000

[10] 松尾 篤, 馬場 功淳, 乃万 司, 松尾 英明, 江島 俊朗”HeadClassifier:人物顔画像の実時間分類” 第 7 回画像センシングシンポジウム, pp411-416,2001

[11] 馬場 功淳, 大橋 健, 乃万 司, 松尾英明, 江島俊朗”「人を見るシステム」の構築-人物追跡と顔画像分類-” 電子情報通信学会 2001 年ソサイエティ大会,(掲載予定)

[12] 小倉 康伸, 榎田 修一, 馬場功淳、松尾 英明, 江島 俊朗” 人物顔画像の実時間分類” 電気関係学会九州支部第 54 回連合大会,(掲載予定)