

確信度付き仮説群の相互作用に基づく複数対象追跡

杉本 晃 宏[†] 谷内 清 剛[†] 松山 隆 司[†]

画像中の対象に対して生成された仮説間の相互作用に基づく複数対象追跡手法を提案する。本手法では、追跡対象のアピアランスモデルを用意し、各仮説を、モデルのパラメタが構成する空間における確信度分布で表現する。そして、そのパラメタ空間中で確信度分布の伝播と再構成を繰り返すことで対象の追跡を実現する。さらに、誤検出を排除するとともに、正しい検出結果を保持するように、仮説間に相互作用を引き起こす。この相互作用により、仮説と画像中の対象とを 1 対 1 に対応させることができる。したがって、対象数が増減したり対象間でオクルージョンが起きるような状況下でも、安定かつ頑健に複数の対象を追跡することができる。実環境下での実験によって本手法の有効性を確認したので、その結果もあわせて報告する。

Multi-object Tracking Based on Interaction between Hypotheses with Certainty

AKIHIRO SUGIMOTO,[†] KIYOTAKE YACHI[†]
and TAKASHI MATSUYAMA[†]

We propose a method for multi-object tracking, where interaction between object hypotheses plays a key role. In our method, we first prepare an appearance model for target objects. Secondly, we generate hypotheses for an object in the image and represent them as certainty distributions in the model-parameter space. We then propagate and reform hypotheses over time in turn. In addition, we bring about interaction between hypotheses to eliminate the hypotheses denoting false positives and, at the same time, to maintain the hypotheses denoting objects. This allows us to identify the one-to-one correspondence between hypotheses and objects in the image. Consequently, the system tracks multiple objects stably even if occlusions occur and the number of objects in the image changes during tracking. Experimental results show the effectiveness of our method.

1. はじめに

移動する対象、とくに人物、を検出し追跡する技術は、コンピュータビジョンの分野で最も重要な課題の一つである^{2),6),7)}。人物を検出しその動きを視覚的に認識することは、視覚監視システム^{3)~5),12)}や遠隔講義¹³⁾などの自動撮影システムをはじめとする多くの応用システムにおいて重要な課題となっているからである。グループを成したり離れたたり、また、すれ違ったりするというような複雑な状況を想定した人物追跡には、人間の様々な動きやオクルージョン、あるいは、撮影環境の変化などに対して頑健な手法が要求される。

頑健な複数対象追跡を実現する手法として、最近、Condensation^{8),10)}が提案され、その有効性が報告さ

れている^{9),11)}。Condensationは、統計的な振る舞いを確率論の枠組に組み込んで、追跡を実現する手法である。そこでは、追跡中に対象数は変わらないという前提の下で、画像中の点をサンプリングして対象の見え方を評価し、多峰性の確率密度関数の形で、対象の仮説を複数同時に保持する。また、確率密度関数を時間方向に伝播させ、新たに得られた画像をサンプリングして確率密度関数を再構成する。そして、確率密度関数の伝播と再構成を繰り返すことで追跡を実現している。サンプリングの数が多くなるに従って、確率密度関数の各峰は対象の位置に漸近的に収束することが保証されている。

Condensationでは、画像上の各点で得た対象の見え方を表す密度を画像全体で正規化し、確率密度関数を得ている。この正規化は、画像上の各点で得られた対象の見え方を同一の尺度で相対評価していることを意味する。各対象ごとに、その見え方をひとつの尺度を用いて画像上で相対評価するのは有効

[†] 京都大学大学院情報学研究所
Graduate School of Informatics, Kyoto University
現在, 国立情報学研究所
Presently with National Institute of Informatics

であるが、複数の対象に対して共通の尺度で相対評価するのは問題である。なぜなら、非常に高い密度をもつ対象が存在した場合、他の対象は、たとえ正しく検出されていても、正規化の影響で、低い密度をもつことになり、対象として認識されない可能性が高くなってしまふからである。ある対象が画像上に存在するか否かは、本来、他の対象の存在とは独立であるべきである。したがって、画像上での、単純な密度の正規化は、複数対象追跡を行う上では問題があるといえる。これに加えて、Condensation には、追跡中に対象数が変化する場合に対処できないという問題点がある。

複数対象追跡では、画像中の対象数は未知であり、多くの場合、追跡中に対象数は増減する。したがって、システムは、生成した仮説がどの対象に対応するかを同定する必要がある。ひとたびこの対応関係が得られれば、Condensation のような確率論的な枠組は有効である。しかし、この対応関係が得られていないときは、対象数が変化する可能性のある画像（映像）を一つの確率密度関数とその伝播で議論することはできない。画面への対象の出入りやオクルージョンがある状況下で、複数の対象を追跡する、頑健で柔軟なシステムを構築するためには、複数の異なる仮説を同時に保持し、かつ、生成された仮説と対象とを 1 対 1 に対応させる機構を備えた手法が必要である。

本稿では、各対象に対して複数の仮説を生成し、それらの相互作用に基づいて、複数の対象を追跡する手法を提案する。本手法では、複数の異なる仮説を同時に保持するとともに、仮説間の相互作用を通して、仮説と対象とを 1 対 1 に対応させる。この相互作用には、仮説から得られる画像情報の時間的、空間的連続性と一貫性が利用される。本手法は次のように特徴づけられる：(1) 追跡対象のアピアランスモデルを用意し、各対象に対して、モデルのパラメタ空間中に複数の仮説を生成する。ここに、各仮説は、パラメタ空間中の確信度分布として表現される。(2) 仮説を時間方向に伝播させ、新たに得られた画像をサンプリングすることで確信度分布を再構成する。(3) 画像上で近い仮説に対して、仮説と対象とが 1 対 1 に対応するように相互作用を引き起こす。(4) 仮説間の整合性と競合に基づいて、各仮説の確信度を評価し、誤検出を排除するとともに、正しい検出結果を保持する。

提案手法では、仮説の、画像上での周辺の情報に基づいて、仮説間の相互作用の度合いを決定する機構（空間的連続性の利用）、時々刻々に得られる各仮説の確信度を累積することでその仮説の確信度を評価する機構（時間的連続性の利用）、を組み込むことで、上述の特徴

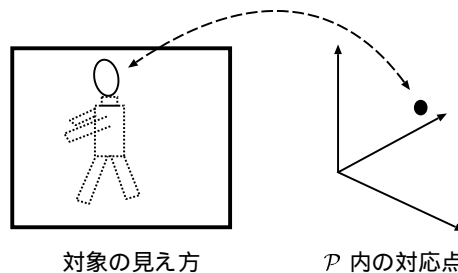


図 1 対象の見え方とそれに対応するパラメタ空間内の点
Fig. 1 A model of an object appearance and its corresponding point in the parameter space.

(3)、(4) を実現している。これらの機構により、システムは、最終的に、仮説と対象とを 1 対 1 に対応させることができる。したがって、オクルージョンが生じたり、対象数が変化しても、頑健な複数対象追跡が実現される。実環境下での実験によって、本手法の有効性が確認されている。なお、本稿では、単一の固定カメラを用いた対象追跡を扱い、照明条件は、撮影中に変化せず、シーン中でほぼ一様であると仮定しておく。

2. 対象の仮説とその表現

2.1 確信度と対象の仮説の表現

追跡対象のアピアランスモデルを定めると、画像上での対象の見え方の多様性は、そのモデルを記述するパラメタの集合として捉えられる。そして、対象の見え方全体は、モデルを記述するのに必要なパラメタ数の次元の空間（パラメタ空間とよぶことにする）を構成する。例えば、人物頭部のアピアランスモデルを楕円とすると、モデルは 5 個のパラメタをもち、画像中の頭部の見え方が与えられると、楕円を表す 5 個のパラメタの値が決まる（図 1）（この例では、パラメタ空間は 5 次元となる）。以下では、パラメタ空間を P で表すことにする。

画像中の対象の見え方は、 P 内の 1 点と対応する。完全に正確に対象の見え方が抽出されていない限り、その見え方に対応する P 内の 1 点を一意に決めるには不確実性が存在する。したがって、 P 内の各点に対し、その点が対象の見え方に対応する不確実性（より正確には、確実性）を、次の性質を満たす確信度と名づける絶対的な尺度で表現することにする。

(1) P を定義域とし、 $[0, 1]$ を値域とする。

定義域を P としてあるが、実際には、 P 内には、確信度が定義されない点が存在する。後に述べるように、 P 内をサンプリングして得られた点に対して確信度を計算し、確信度分布を得るので、その分布でカバーされない点に対しては、確信度は定義されないことになる。

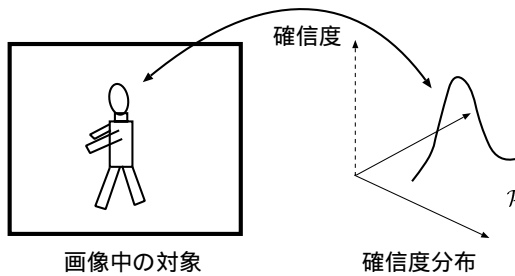


図 2 画像中の対象とそれに対する、 P 内の確信度分布として表現された仮説

Fig. 2 An object in an image and its hypothesis as the certainty distribution in P .

- (2) 各点における値は、その点が対象の見え方に対応する確実性を表す。
- (3) 対象の見え方に正確に対応するとき、その点における値は 1 となり、誤検出に対応するとき、0 となる。

さらに、対象の見え方に対して、 P における表現能力を頑健にするために、確信度に基づいた P 上の分布 (確信度分布とよぶことにする) を導入する。すなわち、対象の見え方の検出結果 (誤検出を含む) に対して、パラメタ空間内に確信度分布を生成し、その確信度分布はその対象の仮説を表現するとみなす (図 2)。

2.2 特徴量に基づく確信度の評価

画像が与えられたとき、 P 内の点に対して、それに対応する画像領域を F 個の異なる特徴量 i ($i = 1, 2, \dots, F$) で対象らしさを評価する。特徴量として、例えば、色ヒストグラムや濃淡ヒストグラム、明度勾配があげられる。

画像中の対象は、この対象らしさの評価値が大きい画像領域に存在している可能性が高いと考えられる。しかし、この評価値によって対象の存在を断定できるわけではなく、そこには不確実性が残っている。そこで、特徴量 i による評価結果を確信度に変換する。(なお、第 2.1 節にあげた性質を満たす限り、どのように確信度を定義してもかまわない。) これにより、 P 内の点における、特徴量 i に基づく確信度が得られる。本稿では、特徴量 i に基づくこの確信度の評価を確信度評価 f^i と呼ぶことにする。すなわち、確信度評価 f^i は、 $f^i(p) = \psi^i(e^i(p))$ と表される。ここに p は P 内の点であり、 e^i は特徴量 i による対象らしさの評価である。また、 ψ^i は、 e^i による評価結果を確信度に変換する関数である。

f^i を統合することで、 p を全ての特徴量で評価して

得られる確信度が定義される:

$$f(p) := \kappa \otimes_{i=1}^F f^i(p). \quad (2.1)$$

ここに、 \otimes は f^i を統合する演算を表し、 κ は値域を正規化するための定数である。 \otimes の例として、積や重みつき線形和があげられる。

確信度への変換関数 ψ^i は、例えば、追跡対象を (人手で) 正確に抽出した際に得られる値を用いて定めることができる。

2.3 確信度分布を用いた追跡

Condensation では、確率密度関数を時間方向に伝播させ、新たに得られた画像をサンプリングして確率密度関数を再構成することを繰り返すことによって追跡を実現しているが、本手法でもこれと同じ枠組で追跡を実現する。すなわち、システムは、対象の見え方に対して仮説を生成し、それを確信度分布として表現する。そして、確信度分布を時間方向に伝播させ、次に得る画像上での対象の見え方を予測する。次に、その予測と新たに得られた画像を用いて仮説を再構成する。再構成した仮説をさらに伝播させる。この伝播と再構成を繰り返すことで追跡を実現する。

2.3.1 仮説の生成と確信度分布による表現

追跡を開始した直後には、画像中の対象の位置や大きさに関する情報は何も得られていない。したがって、対象を検出するための探索範囲は、画像全体になる。初期検出に要する計算量を抑えるためには、背景差分を利用することが考えられる。背景差分により、対象が存在する可能性がある領域を絞り込むことができるからである。

背景差分で検出された領域は、対象の位置情報を与える。この情報は、 P 内の制約領域として表現される。そこで、その制約領域内の点をランダムにサンプリングし、サンプリングした点 p を式 (2.1) で評価する。予め与えたしきい値より $f(p)$ が大きい場合、点 p を仮説とみなし、 p を中心とする確信度分布によってこの仮説を表現する。

仮説を表現する P 内の確信度分布は、 p を中心とする限りどのような形の分布でもよい。以下では、議論を進めるために、 p_{init} を仮説としたとき、次のように確信度分布を定義することにする。すなわち、

さらに P に制約を加えられるとき、サンプリングする範囲をさらに制限することができる。

$$C_{\text{init}}(\mathbf{p}) := \begin{cases} \bigotimes_i f^i(\mathbf{p}_{\text{init}}) \exp\{-\frac{1}{2}g_{\text{init}}^i(\mathbf{p})\} & (\min_i \|g_{\text{init}}^i(\mathbf{p})\| \leq L) \\ 0 & (\min_i \|g_{\text{init}}^i(\mathbf{p})\| > L) \end{cases} \quad (2.2)$$

である。ここに、 i は特微量の名前、 L は与えられた定数であり、

$$g_{\text{init}}^i(\mathbf{p}) := (\mathbf{p} - \mathbf{p}_{\text{init}})^T \Sigma_{\text{init}}^i (\mathbf{p} - \mathbf{p}_{\text{init}})$$

である。また、 Σ_{init}^i は与えられた共分散行列を表す。各特微量の性質をより正確に反映するように、共分散行列は、特微量ごとに与えている。これにより、ある特微量に対して非常に良い評価値を得る \mathcal{P} 内の点を見逃さないようにすることができる。確信度分布の裾野は L によって切り取られているが、これは、分布の広がりパラメタ空間内で有限になるようにするためである。

このサンプリングを繰り返すことで対象の仮説を生成し、各仮説を確信度分布で表現することができる。以後、仮説 h を表現する確信度分布を C^h で表す。なお、画像中の対象一つに対して複数の仮説が生成されること、誤検出を含む可能性があるため、生成された仮説は必ずしも実際の対象に対応するとは限らないこと、を注意しておく。

2.3.2 確信度分布の遷移

対象の仮説 h に対して、それを確信度分布で表現した。しかし、この確信度分布は、時間に対して静的ではない。対象の移動に伴って、動的に変化する。すなわち、確信度分布は時間に依存する。

仮説 h を表現する確信度分布は、現在の画像に対するものである。新たな画像が得られたとき、その画像での仮説 h を表現する確信度分布を求めなければならない。新たに得られた画像に対する確信度分布を求めるために、現在の画像に対する確信度分布を時間方向に伝播させる。確信度分布には、中心、分散、スケールの三つのパラメタがあるので、この伝播は、三つのステップによって行われる。それらは、ピーク伝播、分散伝播、スケール伝播である(図3)。本節では、伝播における時間の前後関係がわかるように時刻を明示して記述する。離散的なタイミングで画像は撮影されるので、時刻は、 t_k ($k = 1, 2, \dots, n, n+1, \dots$) で表され

ここでは、確信度評価 f^i の分布は正規分布の形をしているが、これは単なる例である。どのような分布にしてもかまわない。理論的には、確信度分布は \mathcal{P} 上で無限に広がる分布となる。計算コストを抑えるためにこの制約を設けている。しかし、この制約によって、以後の議論が影響を受けることはない。

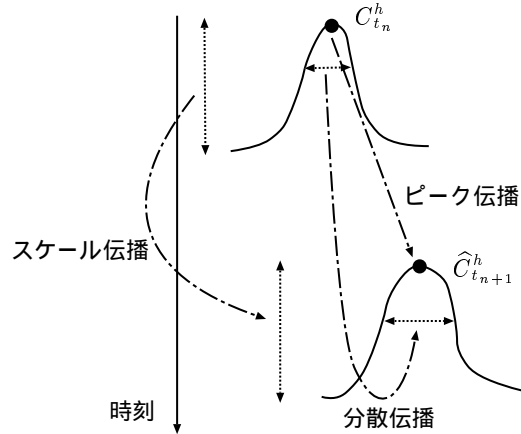


図3 確信度分布の伝播

Fig.3 Propagation of the certainty distribution.

る。なお、現在の画像は、時刻 t_n に撮影されたとする。

1. ピーク伝播

現在の時刻までに得られた確信度分布の中心を用いて、次の時刻における確信度分布の中心を予測するのがピーク伝播である。

いま、時刻 t_k において仮説 h を表現する確信度分布 $C_{t_k}^h(\mathbf{p})$ に対し、 $C_{t_k}^h(\mathbf{p})$ を最大にする \mathbf{p} を $\mathbf{p}_{\text{peak}}^h(t_k)$ とする:

$$\mathbf{p}_{\text{peak}}^h(t_k) := \arg \max_{C_{t_k}^h(\mathbf{p}) > 0} C_{t_k}^h(\mathbf{p}). \quad (2.3)$$

$\mathbf{p}_{\text{peak}}^h(t_k)$ は、確信度分布 $C_{t_k}^h(\mathbf{p})$ の中心である。

各時刻 t_k ($k = 1, 2, \dots, n$) において、 $\mathbf{p}_{\text{peak}}^h(t_k)$ を計算することができるので、 $\{\mathbf{p}_{\text{peak}}^h(t_k) \mid k = 1, 2, \dots, n\}$ が得られる。このとき、ピーク伝播は、 $\{\mathbf{p}_{\text{peak}}^h(t_k) \mid k = 1, 2, \dots, n\}$ を用いて $\mathbf{p}_{\text{peak}}^h(t_{n+1})$ を予測することである。

ピーク伝播のためにどのような予測規則を用いてもよい。予測のための関数を K とすると、 $\mathbf{p}_{\text{peak}}^h(t_{n+1})$ の推定値 $\widehat{\mathbf{p}}_{\text{peak}}^h(t_{n+1})$ が

$$\widehat{\mathbf{p}}_{\text{peak}}^h(t_{n+1}) := K(\{\mathbf{p}_{\text{peak}}^h(t_k) \mid k = 1, 2, \dots, n\})$$

により得られる。具体的には、線形予測やカルマンフィルタを用いることができる。

2. 分散伝播

確信度分布 $C_{t_k}^h(\mathbf{p})$ に対して、確信度評価 f^i での分散を $\Sigma^i(t_k)$ とする。このとき、分散伝播は、時刻 t_n までに得られた確信度分布の分散の履歴 $\{\Sigma^i(t_k) \mid k = 1, 2, \dots, n\}$ を用いて、時刻 t_{n+1} における分散 $\Sigma^i(t_{n+1})$ を予測することである。

ピーク伝播のときと同様にどのような予測規則を用

いてもよい。分散の推定値 $\widehat{\Sigma}^i(t_{n+1})$ は、注目している仮説に対応する対象に対して、その対象が存在していると推定される、モデルパラメタ空間中の領域の大きさを表している。したがって、推定される分散の値は、 $p_{\text{peak}}^h(t_{n+1})$ の推定の正確さに応じて変化させることが望ましい。

3. スケール伝播

確信度分布のスケール、すなわち、式 (2.3) の $p_{\text{peak}}^h(t_k)$ に対する確信度も、やはり、伝播させなければならない。ここでも、分布の値域を $[0, 1]$ に保つ限り、どのような伝播規則を用いてもよい。もっとも単純な規則は、スケールをそのままにしておくことである。

以上、三つのステップを組み合わせることで、確信度分布 $C_{t_n}^h$ を伝播させることができる。この伝播によって推定された確信度分布を $\widehat{C}_{t_{n+1}}^h$ で表す。

2.3.3 確信度分布の再構成

伝播によって推定された確信度分布 $\widehat{C}_{t_{n+1}}^h$ は、時刻 t_{n+1} に実際に撮影して得られた画像の情報を反映していない。新たに得られた画像の情報を反映させるためには、画像から得られる特徴量を用いて、伝播によって推定された確信度分布を再構成する必要がある。

仮説 h を表現する (伝播によって推定された) 確信度分布から、 \mathcal{P} 内の各点における仮説 h の確信度が得られる。そこで、この確信度に比例する確率で \mathcal{P} 内の点をサンプリングし、得られたサンプルを \mathbf{p}_s^h ($s = 1, 2, \dots, S$) とする。次に、新たに得られた画像を用いて、各特徴量ごとにこのサンプルの確信度を評価する。すなわち、全てのサンプル $s \in \mathcal{S} := \{1, 2, \dots, S\}$ と特徴量 $i \in \{1, 2, \dots, F\}$ に対して、 $f^i(\mathbf{p}_s^h)$ を計算する。

時刻 t_{n+1} に得られた画像を用いた確信度評価の結果を使って、確信度分布の再構成に用いるサンプルを選択する。これは、サンプル全てを用いて確信度分布を再構成するのに比べ、確信度評価値の大きいサンプルのみを用いて確信度分布を再構成する方が、より正確な確信度分布が得られるからである。確信度分布の再構成に用いるサンプルを選択するために、まず、上で得られたサンプルのうち、確信度評価 f^i (定義は第 2.2 節) の値がもっとも大きいものを $\mathbf{p}_{\text{max}}^{hi}$ とする：

$$\mathbf{p}_{\text{max}}^{hi} := \arg \max_{s \in \mathcal{S}} f^i(\mathbf{p}_s^h).$$

そして、確信度評価 f^i の値が $\frac{R^i}{100} \cdot f^i(\mathbf{p}_{\text{max}}^{hi})$ より大きいサンプルのみを選ぶ。ここに、 R^i は、0 以上 100 以下の定数である。以下、選ばれたサンプルを $\mathbf{p}_{s^*}^{hi}$ で

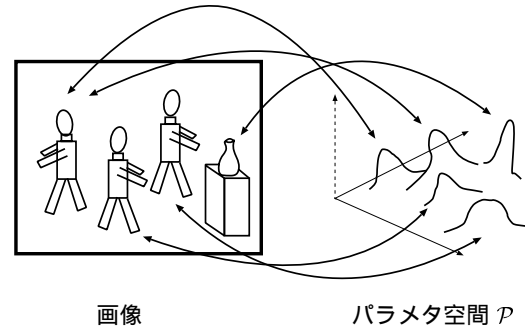


図 4 \mathcal{P} 内の確信度分布と画像内の対象との関係

Fig. 4 Relationship between certainty distributions in \mathcal{P} and objects in the image.

表す。

選ばれたサンプルの集合 $\{\mathbf{p}_{s^*}^{hi}\}$ を用いて、中心が $\mathbf{p}_{\text{max}}^{hi}$ となる確信度分布を再構成する。そのためにまず、 $\{\mathbf{p}_{s^*}^{hi}\}$ に対して、共分散行列を計算し、 Σ^i を得る ($i = 1, 2, \dots, F$)。次に、この Σ^i を用いて

$$c^{hi}(\mathbf{p}_{s^*}^{hi}) := f^i(\mathbf{p}_{\text{max}}^{hi}) \exp \left\{ -\frac{1}{2} g^{hi}(\mathbf{p}_{s^*}^{hi}) \right\}$$

を定義する。ここに、

$$g^{hi}(\mathbf{p}_{s^*}^{hi}) := (\mathbf{p}_{s^*}^{hi} - \mathbf{p}_{\text{max}}^{hi})^\top \Sigma^i (\mathbf{p}_{s^*}^{hi} - \mathbf{p}_{\text{max}}^{hi})$$

である。

最後に、 c^{hi} を統合して、仮説 h を表現する確信度分布を再構成する。すなわち、

$$C_{t_{n+1}}^h(\mathbf{p}) := \begin{cases} \bigotimes_i c^{hi}(\mathbf{p}) & (\min_i \|g^{hi}(\mathbf{p})\| \leq L) \\ 0 & (\min_i \|g^{hi}(\mathbf{p})\| > L) \end{cases} \quad (2.4)$$

とする。ここに、 $\mathbf{p} \in \mathcal{P}$ である。ここでもやはり、分布の広がりパラメタ空間中で有限になるように、 L によって分布の裾野を切り取ってある。

3. 仮説群の相互作用

3.1 仮説と対象との対応

複数対象追跡では、画像中の追跡対象の数は未知である。さらに、対象数は時間とともに変化する可能性がある。すなわち、画面から消えていく対象や新たに画面に入ってくる対象が存在する可能性がある。対象数が変化する状況で、頑健に対象を追跡するためには、仮説とそれに対応する画像上の対象との関係を確立する必要がある (図 4)。加えて、誤検出に対応する仮説を同定し、誤検出を排除しなければならない。

追跡中, システムは, 生成した仮説を (より正確にはそれを表現する確信度分布を), 対応する対象を同定することなく, 伝播させ再構成している. さらに, 誤検出に対応するかどうかも同定していない. 画像上の対象と対応する仮説であっても, 対応する対象が仮説間で同一であるか否かさえわかっていない.

複数対象追跡では, 誤検出に対応する仮説を排除し, 同一の対象に対応する仮説を同定することが不可欠である. 次の二つの節では, システムにこれらの機構をどのように組み込むかを論じる. 以下の議論のために, 保持している仮説を三つのタイプに分類しておく.

タイプ A: ある対象 A に対応する仮説.

タイプ \bar{A} : A ではない, 別の対象に対応する仮説.

タイプ F: 誤検出に対応する仮説.

システム自体は, まだ, どの仮説がどのタイプに属するかが分かっていないことを注意しておく.

3.2 最大確信度の累積による誤検出の排除

各仮説に対して, それを表現する確信度分布は, 各フレームごとに求められている. そこで, 仮説 h に対して, 時刻 t_k における最大確信度 $\gamma_{t_k}^h$ を定義する. すなわち,

$$\gamma_{t_k}^h := C_{t_k}^h(p_{\text{peak}}^h(t_k)) \quad (3.1)$$

を定義する. ここに, $C_{t_k}^h$ は, 時刻 t_k において仮説 h を表現する確信度分布であり, $p_{\text{peak}}^h(t_k)$ は, 式 (2.3) で定義されたものである. $\gamma_{t_k}^h$ は, 時刻 t_k における仮説 h の, もっとも楽観的に考えた場合の, 対象らしさの確信度とみなすことができる. この対象らしさの確信度を時間方向に累積すること, すなわち, 対象らしさの確信度の時間的連続性を利用すること, によって, 誤検出に対応する仮説を排除し, 画像中の対象に対応する仮説のみを保持することを考える.

仮説 h が実際の対象に対応する場合, すなわち, タイプ F に属さない場合, 各時刻 t_k において $\gamma_{t_k}^h$ の値は大きい. つまり, 対象らしさの確信度は大きい値になっている. そこで, 対象らしさの確信度を時々刻々加算することを考えると, それは $\gamma_{t_k}^h$ を t_k に関して累積することに相当し, その値は時間とともに急速に増加する (図 5 (a)). そして, 対応する対象が画面から消えない限り, 増加の速度は鈍くならない. 一方, 仮説 h が誤検出に対応する場合, 各時刻 t_k において $\gamma_{t_k}^h$

対応する対象が画面から消えた場合, 対象が消えたタイミングと同期して, 新たに加算される $\gamma_{t_k}^h$ の値は小さくなるので, $\gamma_{t_k}^h$ の累積値の増加は, 急速に鈍くなり始める. そして, 仮説 h は, 誤検出に対応する仮説となる.

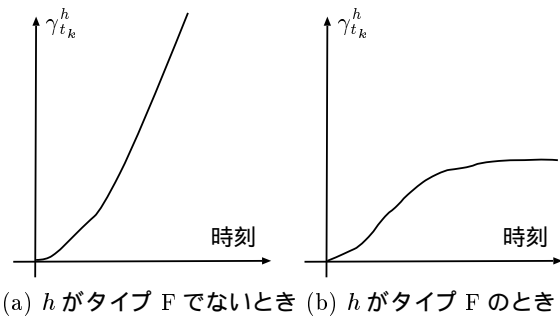


図 5 仮説 h の最大確信度 $\gamma_{t_k}^h$ の累積
Fig. 5 Cumulation $\gamma_{t_k}^h$ of hypothesis h .

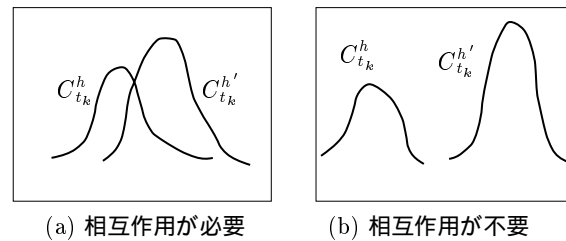


図 6 仮説 h と仮説 h' との相互作用の要不要
Fig. 6 Decision on interaction between hypotheses h and h' .

の値は小さい. したがって, $\gamma_{t_k}^h$ の累積値は, 小さい値の加算になるので, ゆっくり増加し, 場合によっては, ほとんど増加しなくなる (図 5 (b)).

最大確信度の累積に関するこのような性質を利用して, タイプ F に属する仮説であるか否かを判定することができる. すなわち, 最大確信度を時間方向に累積する. 一定時間累積した結果, 累積値が, 累積値に対するしきい値を越えない場合, その仮説は誤検出に対応すると判断し, その仮説を削除する. 累積値が, 累積値に対するしきい値を越えた場合, その仮説は画像中の対象に対応すると判断し, その仮説を保持する. このようにして, どのような場合にも, 誤検出に対応する仮説を削除し, 画像中の対象に対応する仮説のみを保持することができる.

3.3 相互作用を用いた仮説と対象との 1 対 1 の対応づけ

タイプ F に属さない二つの仮説を表現する確信度分布をそれぞれ画像上に投影したとき, 両者が十分近い場合, 仮説に対応する対象に応じて, 二つの場合が考えられる. 一方は, 仮説に対応する対象が同一の場合であり, 他方は, 仮説に対応する対象が異なる場合である. 前者の場合, 二つの仮説は同一の対象に対応

するので、一つの仮説に合併する必要がある。これに対し、後者の場合、異なる二つの対象が画像上で十分近くにいることになり、これは、二つの対象間でオクルージョンが生じようとしていることを意味する。したがって、二つの仮説をともに保持しなければならない。このような状況を適切に取り扱うために、本稿では、画像上で近くにある仮説に対して、相互作用を引き起こす機構を導入する(図6)。なお、相互作用は二つの仮説間に引き起こすものとする。三つ以上の仮説間での相互作用には、二つの仮説間の相互作用を組み合わせることで対応する。

二つの仮説間の相互作用の度合いを求めるために、画像上での空間的連続性、具体的には、仮説の周辺の情報、を利用する。ここで、仮説の周辺の情報とは、その仮説に対して最大確信度を与える \mathcal{P} 内の点に対応する対象の見え方を考えたときに、その見え方の(画像上での)周辺で得られる特徴量である。本稿では、この特徴量を付随特徴量とよぶことにする。付随特徴量としては、保持している仮説に対応する対象間の同定に有効であるような特徴量を選ぶことが望まれる。なぜなら、下に述べるように、二つの仮説に対して、それぞれの付随特徴量の類似度に応じて相互作用の度合いが決められ、その結果が、二つの仮説が同じ対象に対応するか否かを判断することに影響を及ぼすからである。人間の顔を追跡している場合、例えば、首や肩のライン、あるいは、身体から得られる特徴量を付随特徴量とすることができる。

付随特徴量を用いて、二つの仮説間の相互作用の度合いを求める方法を導入する。まず、付随特徴量の一つ選び、画像(フレーム)ごとに二つの仮説間のその付随特徴量ベクトルの距離(その付随特徴量の非類似度)を求め、その距離を評価して相互作用の度合いを定めることを考える。距離が小さいことは、それぞれの仮説に対応する対象が同一であることを支持するので、二つの仮説を一つに合併してよいことになる。一方、距離が大きいことは、それぞれの仮説に対応する対象が異なることを支持するので、二つの仮説をともに保持する必要がある。距離そのものは、各フレームでの相互作用の度合いと密接に関係する。そこで、距離を相互作用に度合いに変換する関数を用意し、これを用いて各フレームでの相互作用の度合いを求めることにする。そして、その時点までに得られているフレーム

あらゆる状況下において、万能な付随特徴量を具体的に示すことは困難である。むしろ、そのような付随特徴量は存在しないといってもよい。どのような付随特徴量を採用すればよいかは、状況に応じて変わる。

表1 仮説 h と仮説 h' との相互作用の望ましい結果
Table 1 The desired results after the interaction between two hypotheses h and h' .

h のタイプ	A	A	A	F
h' のタイプ	\bar{A}	A	F	F
保持すべきタイプ	A and \bar{A}	A	A	F
w の符号	+ (large)	-	\pm (small)	-

に対して、フレームごとに相互作用の度合いを求め、その重み付き平均値をその付随特徴量によって得られる相互作用の度合いであるとみなす。最後に、各付随特徴量で得られる相互作用の度合いを統合し、その時点における二つの仮説の相互作用の度合いを決定する。以上が、付随特徴量を用いて相互作用の度合いを求める方法の基本的な考え方である。なお、ここで得られた相互作用の度合いに応じた確信度が、各仮説の最大確信度の累積値に加えられる。

二つの仮説に対して、合併してよい場合には負の値をとり、ともに保持する場合には正の値をとるように、二つの仮説間の相互作用の度合いを求める関数 w を定義することを考える。このとき、 w に求められる性質は、表1にまとめられる。表1で、例えば、タイプAの仮説とタイプ \bar{A} の仮説とが相互作用を引き起こす場合(表1の2列目)、二つの仮説は各々異なる対象に対応しているため、仮説を両方とも保持しなければならないことを表す(したがって w の符号は正となる)。一方、二つの仮説がともにタイプAである場合(表1の3列目)、二つの仮説は同じ対象に対応しているため、それらは合併しなければならないことを表す。したがって、 w の符号は負となる。(他の場合についても同様である。)なお、二つの仮説には、誤検出に対応するものが含まれる場合があることを注意しておく。(表1において、 \pm は、誤検出に対応する仮説に応じて、 w の符号は正負いずれであってもよいが、その絶対値は十分小さいことが望ましいことを示す。なぜなら、確信度の累積を利用して誤検出に対応する仮説を削除するためには、相互作用の影響によって確信度を不必要に大きく変えないことが必要となるからである。)

このような性質をもつ w を具体的に定義することを考える。いま、仮説 h の時刻 t_k における ν 番目 ($\nu = 1, 2, \dots, N$) の付随特徴量ベクトルを $(\lambda_{t_k}^h)_\nu$ で表すことにする。仮説 h に対して得られる、時刻 t_n までの付随特徴量の履歴

$$A_{t_n}^h := \{(\lambda_{t_k}^h)_\nu \mid k = 1, 2, \dots, n; \nu = 1, 2, \dots, N\} \quad (3.2)$$

を定義する。このとき、 w は仮説 h と仮説 h' の付随特徴量ベクトルの履歴を引数とする関数として次のように定義される:

$$w(\mathbf{A}_{t_n}^h, \mathbf{A}_{t_n}^{h'}) := \bigotimes_{\nu=1}^N \left\{ \frac{1}{n} \sum_{k=1}^n \alpha_k \times \varphi_{\nu} \left(d((\lambda_{t_k}^h)_{\nu}, (\lambda_{t_k}^{h'})_{\nu}) \right) \right\}. \quad (3.3)$$

ここに、 $\alpha_k (\geq 0)$ は重みを表す係数であり、 $d((\lambda_{t_k}^h)_{\nu}, (\lambda_{t_k}^{h'})_{\nu})$ は、二つの仮説 h と h' に対して、時刻 t_k における ν 番目の付随特徴量同士の非類似度 (すなわち、距離) を計算する関数である。また、 φ_{ν} は、付随特徴量の非類似度を相互作用の度合いに変換する関数であり、その値域が $[-1, 1]$ であるような単調増加関数である。

φ_{ν} として例えば、 $\varphi_{\nu}(d) := \tanh(d - \kappa)$ とすると、付随特徴量ベクトルの距離 (非類似度) d と κ との大小で相互作用の符号が変わる。 $d < \kappa$ のとき (付随特徴量が類似しているので二つの仮説を合併してよいとき)、 $\varphi_{\nu} < 0$ となるので w は負の値をとる。一方、 $d > \kappa$ のとき (付随特徴量が類似していないので二つの仮説を保持するとき)、 $\varphi_{\nu} > 0$ となるので w は正の値をとる。このようにして、 w が満たすべき性質が満足される。

式 (3.3) の w の値に応じた確信度を、仮説の最大確信度の累積値に組み込むことで相互作用を実現する。まず、 $h' \neq h$ である二つの仮説 h と h' に対して、時刻 t_n において、 h' が h に与える相互作用の大きさを

$$I_{t_n}^{h' \rightarrow h} := w(\mathbf{A}_{t_n}^h, \mathbf{A}_{t_n}^{h'}) \cdot \gamma_{t_n}^{h'} \quad (3.4)$$

と定義する。なお、 $\gamma_{t_n}^{h'}$ は、仮説 h' の時刻 t_n における最大確信度である。次に、この相互作用の大きさを用いて、仮説 h の累積確信度を次のように定義する:

$$\Gamma_{t_n}^h := \sum_{k=1}^n \gamma_{t_k}^h + \sum_{\substack{h' \\ h' \neq h}} I_{t_n}^{h' \rightarrow h}. \quad (3.5)$$

w の値を正にする仮説は、仮説 h の累積確信度を大きくする影響を与える。これに対し、 w の値を負にする仮説は、小さくする影響を与える。 (w の値を零にする仮説は、仮説 h の累積確信度に影響しない。すなわち、仮説 h と相互作用を起こさない。)

時刻 t_n において、 $\Gamma_{t_n}^h$ が累積値に対するしきい \tilde{I} より大きいとき、仮説 h を保持する。小さいときは、仮説 h を削除する。なお、仮説を削除した場合 (例えば、時刻 t_{n^*} において仮説 h^* を削除したとする)、そのとき保持している仮説 h の累積確信度 $\Gamma_{t_{n^*}}^h$ から、時刻 t_{n^*} までに仮説 h^* が仮説 h に与えた相互作用の大きさ、i.e., $\sum_{k=1}^{n^*} I_{t_k}^{h^* \rightarrow h}$, を引いておく。これは、誤

検出に対応しない仮説の累積確信度が、相互作用の結果、不必要に小さくなることを防ぐためである。

このような仮説間の相互作用の機構によって、表 1 にあげたような仮説の合併、保持が引き起こされる。そして、最終的には、仮説と対象との 1 対 1 の対応が得られ、保持している確信度分布の中心は、画像上に存在する対象の見え方に対応する \mathcal{P} 内の点に収束することが期待される。このようにして、オクルージョンが起きたり、画像中の対象の数が減少したりする状況下でも、複数対象追跡が実現される。そして、保持している仮説の数を数えることで、システムは、その時点で画像上に存在する対象の数を知ることができる。

4. 対象数の増加に対する対処

追跡する対象の数が増加しない限り、これまで述べてきた手法は効果的に機能する。しかし、今のままでは、画面に新たに対象が現れたとき、その対象を検出する機構が備わっていない。ここでは、画面に新たに現れた対象を検出し、それに対応する仮説を生成する方法を述べる。

保持している仮説に対しては、既に、それぞれを表現する確信度分布をパラメタ空間 \mathcal{P} 上に形成している。そこで、時刻 t_k において、少なくとも一つの仮説に対しては、確信度が正となる \mathcal{P} 内の点を特定し、そのような点の集合を Ω_{t_k} で表す。 Ω_{t_k} は、次のように定義される:

$$\Omega_{t_k} := \{ \mathbf{p} \mid \mathbf{p} \in \mathcal{P}, C_{t_k}^h(\mathbf{p}) > 0 \text{ for } \exists h \}. \quad (4.1)$$

Ω_{t_k} に対応する画像上での領域は、現在保持している仮説群が存在する領域と考えられる。したがって、画像全体からその領域を除いた部分には、時刻 t_k において、対象が検出されていないことになる。すなわち、それは、対象が新たに現れる可能性がある画像領域である。

追跡中に新たに現れた対象を検出するためには、対象が新たに現れる可能性のある画像領域に対してのみ背景差分を実行すればよい。そしてその結果に基づいて、第 2.3.1 節に述べた方法で、新たに現れた対象に対応する仮説を生成し、それを確信度分布で表現する。このようにして、追跡中に対象数が増加した場合にも対処することが可能になる。

5. 複数対象追跡アルゴリズム

以上の議論に基づいて、複数の対象を検出し、追跡する手続きをアルゴリズムの形で記述する。

Step 0: 背景画像を撮影する。

- $k = 1$, すべての h に対して, $\Lambda_{t_k}^h := \phi$ とする.
- Step 1:** 時刻 t_k において画像を撮影する.
時刻 t_k において, $Region :=$ 画像全体 とする.
- Step 2:** $Region$ 内の領域に対して,
2.1: $Region$ に対して, 背景差分を適用する.
2.2: パラメタ空間 \mathcal{P} 内でサンプリングを行う. \mathbf{p}^h をサンプルとする (cf. 第 2.3.1 節).
2.3: 式 (2.2) の $C_{init}(\mathbf{p})$ を計算する. そして, $C_{t_k}^h := C_{init}$ とおく.
2.4: すべての h と ν に対して, 付随特徴量 $(\lambda_{t_k}^h)_\nu$ をもとめ, 式 (3.2) の $\Lambda_{t_k}^h$ に加える.
- Step 3:** すべての h に対して, $C_{t_k}^h(\mathbf{p})$ を伝播させ, $\hat{C}_{t_{k+1}}^h(\mathbf{p})$ を得る (cf. 第 2.3.2 節).
- Step 4:** $k := k + 1$ とする.
- Step 5:** 時刻 t_k において画像を撮影する.
- Step 6:** すべての h に対して,
6.1: $\hat{C}_{t_k}^h(\mathbf{p})$ に基づいて \mathcal{P} 内の点をサンプリングし, 確信度評価 f^i に基づいてサンプルを選択する. 選択されたサンプルを $\mathbf{p}_{s^*}^{h^i}$ とする (cf. 第 2.3.3 節).
6.2: 式 (2.4) の $C_{t_k}^h(\mathbf{p})$ を計算する.
6.3: 式 (3.1) の $\gamma_{t_k}^h$ を計算する.
6.4: すべての h と ν に対して, 付随特徴量 $(\lambda_{t_k}^h)_\nu$ を求め, それらと $\Lambda_{t_{k-1}}^h$ を式 (3.2) の $\Lambda_{t_k}^h$ に加える.
- Step 7:** 相互作用が必要な二つの仮説の組に対して, 式 (3.4) の $I_{t_k}^{h' \rightarrow h}$ を計算する.
- Step 8:** 式 (3.5) の $\Gamma_{t_k}^h$ を計算する.
- Step 9:** $\Gamma_{t_k}^h < \tilde{\Gamma}$ である場合, 仮説 h と $\Lambda_{t_k}^h$ を削除する.
すべての h' に対して, $\Gamma_{t_k}^{h'}$ から $\sum_{k'=1}^k I_{t_k}^{h' \rightarrow h'}$ を引く.
- Step 10:** 式 (4.1) の Ω_{t_k} を計算し, 画像中の対応する領域を特定する.
画像全体からその領域を除いた領域を $Region$ にセットし, Step 2 に戻る.

6. 人物頭部追跡実験

提案手法を用いて人物頭部追跡システムを構築し, 屋外環境下で実験を行った. その結果, 提案手法の有効性と頑健性, および, 実時間性を確認した.

6.1 人物頭部の検出

人物頭部を追跡対象とし, 頭部のアピアランスモデルを楕円とした^{1),14)}. 楕円は, 中心の座標, 長軸の長

さ, 傾き, 扁平率の 5 個のパラメタをもつ. しかし, 簡単のため, 実験では, 傾きと扁平率を固定した (実際には, 画像上の楕円の中心の位置に応じて, この二つのパラメタの値をいくつか用意し, それを適切に選択して用いた¹⁴⁾). したがって, 本実験におけるパラメタ空間 \mathcal{P} は, 楕円の中心位置 (x, y) と長軸の長さ a のみによって構成され, その次元数は 3 である.

人物頭部らしさの評価のために, 三つの特徴量を用いた. 楕円内部の色度のカラーヒストグラムとあらかじめ得ておいた人物頭部の色度のカラーヒストグラムとの重なり割合, 楕円周上の法線方向の明度勾配の平均値, 楕円内部の背景差分結果の平均値である. 各特徴量に対して, 人物頭部と楕円モデルとの評価値の分布をあらかじめ得ておき, これを用いて, 確信度評価を行った. その評価法を以下に示す.

(1) 楕円内部の色度のカラーヒストグラム

楕円内部の色情報による人物頭部らしさの評価関数 e^{color} は, 楕円のパラメタを $\mathbf{p} = (x, y, a)^\top$ として

$$e^{\text{color}}(\mathbf{p}) := k^{\text{color}} \times \sum_s \min \left(\frac{I(s)}{\sum_{s=1}^S I(s)}, \frac{M(s)}{\sum_{s=1}^S M(s)} \right)$$

とした. ここに, s はヒストグラム内の s 番目 ($s = 1, 2, \dots, S$) の色度を表し, I は楕円内部の色度のヒストグラムの度数, M はあらかじめ得ておいた人物頭部の色度のヒストグラムの度数を表す. k^{color} は, 無次元化のための定数である¹⁾. 次に, e^{color} とあらかじめ得ておいた人物頭部とそのモデルとの評価値の分布 Φ^{color} を用いて, 確信度を

$$\psi^{\text{color}}(e^{\text{color}}(\mathbf{p})) = \int_0^{e^{\text{color}}(\mathbf{p})} \Phi^{\text{color}}(x) dx \quad (6.1)$$

とした.

(2) 楕円周上の法線方向の明度勾配

楕円周上の明度勾配情報による人物頭部らしさの評価関数 e^{grad} は,

$$e^{\text{grad}}(\mathbf{p}) := k^{\text{grad}} \frac{1}{N} \sum_{s=1}^N |n(s) \cdot g(s)|$$

とした. ここに, N は楕円の円周上の画素の総数, $n(s)$ は円周上の画素 s における長さ 1 の法

人物頭部の色度のカラーヒストグラムは, 追跡中に得られた確信度分布のピークに対応する画像上の見え方の色度のカラーヒストグラムを一定時間にわたって平均して得た.

線ベクトル, $g(s)$ は s における明度勾配ベクトルである. また k^{grad} は無次元化のための定数である¹⁾. これを確信度に変換するときには, やはり, あらかじめ得ておいた人物頭部とモデルとの評価値の分布 Φ^{grad} を用いた. なお, 楕円周上の明度勾配情報は, 画像上での人物頭部の存在位置に依存する特徴量であるため, 画像上の領域ごとに人物頭部とモデルとの評価値の分布の形状が異なる. そこで, 画像を分布の形状がほとんど変わらない小領域 R_r ($r = 1, 2, \dots$) に分割し, それぞれの領域での評価値の分布 $\Phi_{R_r}^{\text{grad}}$ を用いて確信度に変換した. 具体的には, $\mathbf{p} = (x, y, a)^\top$ としたとき, $(x, y) \in R_r$ となるような $R_r \in \{R_r \mid r = 1, 2, \dots\}$ を求め, これを用いて,

$$\psi^{\text{grad}}(e^{\text{grad}}(\mathbf{p})) = \int_{R_r} \Phi_{R_r}^{\text{grad}}(x) \times |x - e^{\text{grad}}(\mathbf{p})| dx \quad (6.2)$$

とした. (これは評価値 $e^{\text{grad}}(\mathbf{p})$ と分布 $\Phi_{R_r}^{\text{grad}}$ との平均距離であり, 評価値の分布への帰属の度合いを表している.)

(3) 楕円内部の背景差分情報

楕円内部の背景差分情報による人物頭部らしさの評価関数 $e^{\text{bg-sub}}$ は,

$$e^{\text{bg-sub}}(\mathbf{p}) = k^{\text{bf-sub}} \frac{1}{N_{\text{in}}} \sum_{s=1}^{N_{\text{in}}} |in(s) - bg(s)|$$

とした. ここに, N_{in} は楕円内部の画素の総数, $in(s)$ は入力画像における s 番目の画素の明度値, $bg(s)$ は背景画像における s 番目の画素の明度値, $k^{\text{bg-sub}}$ は無次元化のための定数である. 背景差分情報は画像上での人物頭部の存在位置に依存するため, 楕円周上の法線方向の明度勾配の場合と同様に, 人物頭部とモデルとの評価値の分布を画像領域ごとに求めて, これを確信度への変換に用いた. すなわち, 画像を分布の形状がほとんど変わらない小領域 R_r ($r = 1, 2, \dots$) に分割し, それぞれの領域で評価値の分布 $\Phi_{R_r}^{\text{bg-sub}}$ を得ておく. そして, $\mathbf{p} = (x, y, a)^\top$ に対して, $(x, y) \in R_r$ となるような $R_r \in \{R_r \mid r = 1, 2, \dots\}$ を求め, これを用いて,

$$\psi^{\text{bg-sub}}(e^{\text{bg-sub}}(\mathbf{p})) = \int_{R_r} \Phi_{R_r}^{\text{bg-sub}}(x) \times |x - e^{\text{bg-sub}}(\mathbf{p})| dx \quad (6.3)$$

とした.

このようにして得られた三つの確信度 (式 (6.1), (6.2), (6.3)) を積によって統合した. そして, 背景差分によって検出された領域に対して, 1000 点をサンプリングし, 確信度を評価した.

サンプリングされた点を仮説として生成するか否かを決定する, 確信度に関するしきい値の設定は, カメラの設置状況に応じて調整する必要がある. はじめはしきい値を低く設定し, その状況下で検出される人物頭部のおおよその評価値を調べ, その値より少し小さい値をしきい値として設定すればよいことが経験的に得られている. これは, 設置状況において, 採用された特徴量によって求められた, 人物頭部の確信度を目安としてしきい値を設定すればよいことを意味している.

6.2 人物頭部の確信度分布

サンプリングによって生成された仮説に対して, それを式 (2.2) で与えられる確信度分布で表現した ($i = \text{colr}, \text{grad}, \text{bg_sub}$). ただし, この実験では, 計算を簡略化するため, 特徴量ごとに用いる共分散行列 Σ^i の非対角成分は全て零とした:

$$\Sigma^i = \begin{pmatrix} (\sigma_x^i)^2 & 0 & 0 \\ 0 & (\sigma_y^i)^2 & 0 \\ 0 & 0 & (\sigma_a^i)^2 \end{pmatrix},$$

ただし, $i = \text{colr}, \text{grad}, \text{bg_sub}$. これは各特徴量における評価において, パラメタ空間の三つの軸 (楕円の中心の x 座標, y 座標, 長軸の長さ) に相関がないと考えることを意味する.

確信度分布の伝搬は仮説ごとに独立に行った. 時刻 t_n における確信度分布から時刻 t_{n+1} における確信度分布を推定する伝搬規則は次のようにした. まず, ピーク伝播では, $\mathbf{p}_{\text{peak}}^h(t_{n-1})$ と $\mathbf{p}_{\text{peak}}^h(t_n)$ を用いて $\mathbf{p}_{\text{peak}}^h(t_{n+1})$ を線形に予測した. すなわち,

$$\mathbf{p}_{\text{peak}}^h(t_{n+1}) := \mathbf{p}_{\text{peak}}^h(t_n) + \frac{t_{n+1} - t_n}{t_n - t_{n-1}} \{ \mathbf{p}_{\text{peak}}^h(t_n) - \mathbf{p}_{\text{peak}}^h(t_{n-1}) \}$$

とした. また, 分散伝播には

$$\sigma_{uv}^i(t_{n+1}) := \sigma_{uv}^i(t_n) + \delta_{uv} M_{uv}$$

を用いた. ここに, $\sigma_{uv}^i(t_n)$ は, $\Sigma^i(t_n)$ の第 (u, v) 要素を表し, M_{uv} は定数である. また, δ_{uv} はクロネッカーのデルタ関数である. M_{uv} は, 伝播後の分布に, ある程度の広がりを持たせるための役割を担う. 一方,

スケール伝播では、スケールを不変とした。

なお、確信度分布を再構成する際のサンプリング数は、500 とした。

6.3 人物頭部の仮説の相互作用

相互作用の計算には、二つの付随特徴量を用いた。確信度分布のピークに対応する楕円の中心の画像上での移動速度（仮説の移動速度）と確信度分布のピークに対応する楕円の下部に作成した矩形領域内部の縦方向の微分値のヒストグラム（仮説下部の矩形領域のヒストグラム）である。二つの確信度分布が互いに重なる時、それらに対して、この二つの付随特徴量に基づいて相互作用の度合いを計算し、相互作用を引き起こした。実装した付随特徴量の評価法を以下に述べる。

(1) 仮説の移動速度

仮説に対応する人物頭部が同一の場合、その仮説の画像上での移動速度はほぼ等しくなる。この性質に基づいて、二つの仮説に関して、オクルージョンが起きているのか（表1の第2列目）同一の対象に対する仮説であるのか（表1の第3, 4列目）を区別することができる。そこで、二つの仮説 h, h' の、時刻 t_n における画像上での移動速度をそれぞれ、 $v_{t_n}^h, v_{t_n}^{h'}$ とし、この付随特徴量に基づき、時刻 t_n における相互作用の度合いを

$$\varphi_{\text{vel}}(d_{\text{vel}}(v_{t_n}^h, v_{t_n}^{h'})) = \tanh \left(a_{\text{vel}} \cdot \left(|v_{t_n}^h - v_{t_n}^{h'}| - b_{\text{vel}} \right) \right)$$

とした。ただし、 $a_{\text{vel}}, b_{\text{vel}}$ は定数である。 b_{vel} は、第3.3節で述べたように、相互作用の符号が変わる境界を表す。また、 a_{vel} の値が大ききときには、 b_{vel} の値付近での φ_{vel} の値が急激に変化するので、 a_{vel} は、その境界前後での人物頭部の区別の容易さを表す。 $(a_{\text{vel}}, b_{\text{vel}})$ の値を実際に決めるときには、このような意味をふまえて調整すればよい。

(2) 仮説下部の矩形領域のヒストグラム

仮説に対応する人物頭部が同一の場合、その仮説の下部に作成した矩形領域（正しい検出であれば人物の胴体部分に相当する）内部の微分値のヒストグラムは、ほぼ等しくなり、これによってオクルージョンをおきているかどうかを判定することができる。そこで、時刻 t_n において、仮説 h, h' の下部に存在する矩形領域内の縦方向の微分値ヒストグラムをそれぞれ $u_{t_n}^h, u_{t_n}^{h'}$ としたとき、この付随特徴量に基づき、時刻 t_n における相互作用の度合いを

$$\varphi_{\text{hist}}(d_{\text{hist}}(u_{t_n}^h, u_{t_n}^{h'})) = \tanh \left[a_{\text{hist}} \times \left(\sum_s \min \left\{ \frac{u_{t_n}^h(s)}{\sum_{s'} u_{t_n}^h(s')}, \frac{u_{t_n}^{h'}(s)}{\sum_{s'} u_{t_n}^{h'}(s')} \right\} - b_{\text{hist}} \right) \right]$$

とした。ここに、 $a_{\text{hist}}, b_{\text{hist}}$ は定数である。（ここでは、二つの仮説の下部に作成された矩形領域内の微分値ヒストグラムの重なり具合を特徴量間の距離としている。）

以上のようにして付随特徴量を評価し、相互作用の度合いを決定した。なお、式(3.3)における α_k に関しては、 t_n からみてあまりに古い情報は利用しないという観点から、どれだけ古い情報を利用するかを決定する定数 τ を用いて

$$\alpha_k = \begin{cases} \frac{1}{\tau} & (n \geq k > n - \tau) \\ 0 & (n - \tau \geq k) \end{cases}$$

とした。また、演算 \otimes は、総和とした。

6.4 実験結果とその評価

実験では、次のような状況を設定した。画面中には誰もいない状態が続き、その後、人物 A が画面に現れ（#34）、カメラの前を通りすぎて画面から消え去る（#59）。次に、人物 B が画面に現れ（#97）、画面の中心付近で立ち止まる（#110）。その後、人物 C が画面に現れ（#134）、B の後ろを通り抜けて（#140）、画面から消える（#146）。次に、人物 D が画面に現れ（#183）、B に接近する（#224）。D は、B から離れた（#243）あと、画面から消え（#251）、最後に B も画面から消える（#257）。（数字はフレーム番号を表す。）なお、実験中、人物は、カメラの前方 2~4m あたりを自然な速度で歩いている。また、追跡中、同時に保持する仮説の最大数を 10 とした。実際、実験では、10 個の仮説が同時に生成されることはなかった。

実験で得られた画像の一部を図7に示す。各フレームには、生成された仮説を表現する確信度分布のピークに対応する楕円を、検出された人物頭部として重ねがきした。また、各楕円には、追跡中に生成された仮説の名前を表すラベルを付与した。図7から、仮説 b, c, d は、それぞれ、人物 B, C, D の頭部に対応していることがわかる。また、人物頭部の数の変化やオク

追跡中に対象数が変化する、このような状況に対しては、Condensation では対処できない。Condensation では、追跡中に対象数は変わらないという前提があるからである。

背景差分によって検出された領域の約 10% が人物頭部に相当し、さらに、人物頭部内から得られるサンプルの約 10% のみが、仮説として生成されたからである。



図 7 屋外環境における対象追跡実験の結果 (数字はフレーム番号を表す)

Fig. 7 Images (with frame numbers) in the outdoor sequence and detected human heads.

ルージョンに影響されず、仮説 b , c , d は、それぞれ対応する頭部をほとんど正しく追跡していることが読み取れる。

追跡中に生成された仮説の累積確信度の変化を図 8 (a) に示す。9 個の仮説が生成され、そのうち 5 個が誤検出に対応していることがわかる。誤検出に対応する仮説の累積確信度は、期待した通り、大きい値にならず、全て正しく削除されている。そして、仮説と画像中の人物頭部とは、1 対 1 に対応するようになっている。また、画面からある人物が消えたあと、その人物の

頭部に対応する仮説も正しく削除されている。(なお、仮説 h が相互作用を引き起こしていない場合、 $\Gamma_{\ell_k}^h$ に上限を設定し、不必要に値が大きくなるようにした。)

図 8 (b) は、追跡中に、仮説 b とその他の仮説との間で引き起こされた相互作用の度合いを示す。ここに、 b , c , d 以外の仮説 ($f1 \sim f5$) は誤検出に対応していて、仮説 $f1$ は仮説 c の近くに、仮説 $f2 \sim f5$ は全て仮説 b の近くに生成されていた。例えば、仮説 b と仮説 $f2$ との相互作用では、 w の値は負になっている。これ

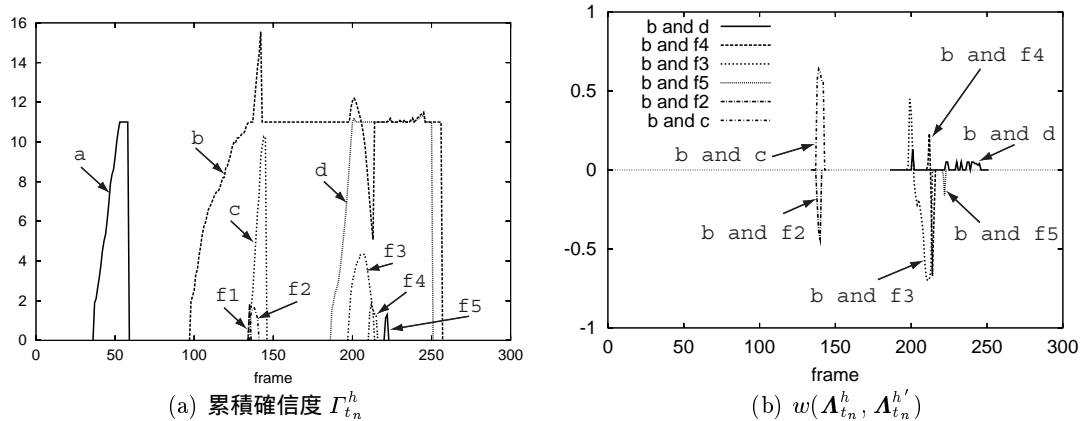


図 8 累積確信度と仮説間 (仮説 b とその他の仮説) の相互作用の度合い
 Fig. 8 Cumulative certainty and the degree of interaction between hypotheses (b and others).

に対し、仮説 b と仮説 c に対しては、 w の値は正である。これは、仮説 b と仮説 f_2 は一つに合併され、仮説 b と仮説 c はともに保持されるように相互作用が機能していることを示す。これは、正しい動作である。実際、仮説 f_2 の累積確信度は仮説 b によって抑制され、その後、誤検出に対応すると判断されている。一方、仮説 b と仮説 c は、オクルージョンが起きている #140, #141, #142 フレーム前後では、互いに累積確信度を大きくしあっている。この結果、システムは、仮説 b と仮説 c が近接しても両者をともに保持している。他のフレームに対しても同じような結果が観測され、それらは全て正しい動作であった。

導入した相互作用の有効性をさらに検証するために、相互作用の部分を除き、相互作用を引き起こさないシステムを実装し、そのシステムによって同一条件下で人物頭部追跡実験を行った。この実験での追跡中に得られた、仮説の累積確信度の変化を図 9 に示す。設定条件は同一であっても、サンプリングが行われる点が同一になることはない、生成された仮説の数が異なっていることに注意しておく。この場合、全部で 7 個の仮説が生成された。仮説 a, b, c, d は、それぞれ、人物 A, B, C, D に対応していた。また、仮説 f_1, f_2, f_3 は誤検出に対応していた。仮説 f_1, f_2 は、人物 A の検出に対して発生した誤検出であり、仮説 f_3 は、人物 B の検出に対して発生した誤検出であった。なお、これらの誤検出は、全て頭部付近に発生していたため、各フレームにおける確信度は、誤検出のわりには大きかった。それゆえ、確信度の累積のみでは削除されなかった。図 9 からわかるように、誤検出に対応する仮説は一度生成されると、対応するとその仮説が思い込んでいる人物頭部が画面から消えるまで消滅することはない。

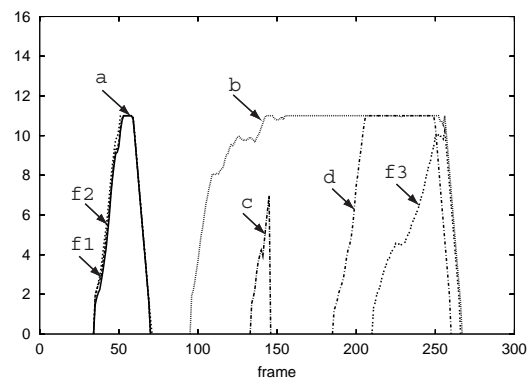


図 9 相互作用がない場合の累積確信度
 Fig. 9 Cumulative certainty without any interaction.

表 2 頭部の位置の検出誤差

Table 2 Errors in detecting human head positions.

仮説	平均 [pixel]	標準偏差 [pixel]
a	6.156	2.995
b	5.716	2.417
c	10.61	3.160
d	13.45	3.338

い。すなわち、人物頭部の方からみると、正しい仮説と誤検出が常にはりついている状態になる。一方、システム側からすれば、画面中の人物頭部の数を誤って認識していることになる。これに対し相互作用を導入すると、図 8 でみたように、誤検出が発生しても、仮説間の相互作用によって仮説の合併が引き起こされ、誤検出はすぐに削除されている。このように、相互作用が誤検出の削除に有効であることがわかる。

提案手法による検出・追跡の精度を定量的に調べるために、画像中の人物頭部に手で楕円を当てはめ、その結果得た楕円の中心座標と長軸の長さを、それぞれ、

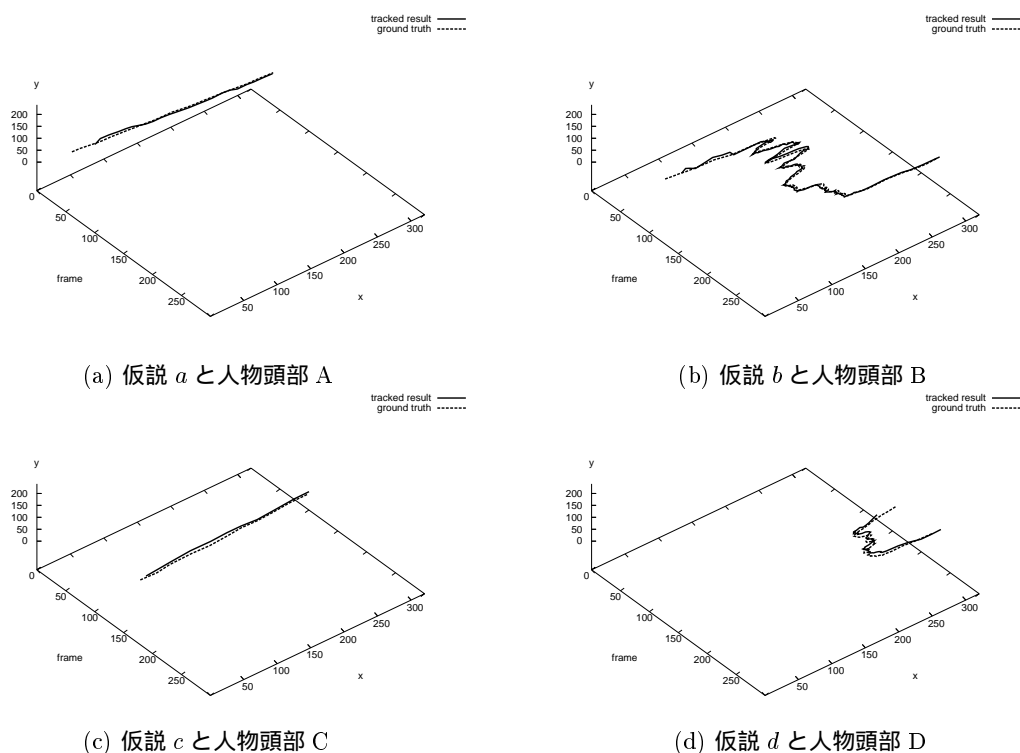


図 10 仮説の位置の軌跡と対応する人物頭部の位置の軌跡
 Fig. 10 Loci of hypotheses and their corresponding human heads.

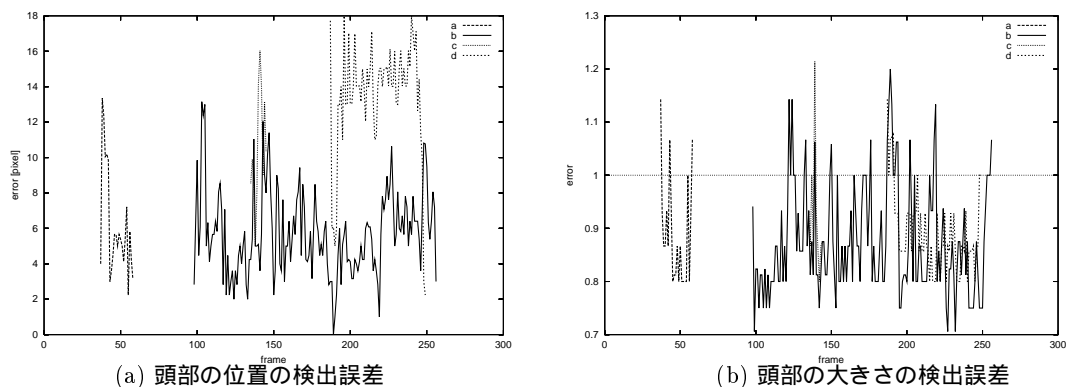


図 11 人物頭部の検出誤差
 Fig. 11 Detection errors of human heads.

人物頭部の真の位置, 真の大きさとみなし, 追跡中に得られた仮説の位置や大きさと比較した. 各仮説に対しては, それを表現する確信度分布の, 最大確信度を与える点に対応する楕円を考え, 画像上での, その楕円の中心の座標, 長軸の長さを, それぞれ, その仮説の位置, 仮説の大きさとみなした. 図 10 に, 仮説の位置とそれに対応する人物頭部の真の位置の画像上での軌

跡を示す. また, 仮説の位置とそれに対応する人物頭部の真の位置の誤差を図 11 (a) に, 仮説ごとに求めた位置の誤差の平均, および, 標準偏差を表 2 に示す. さらに, 人物頭部の真の大きさに対する仮説の大きさの比を図 11 (b) に, この比と 1 との差の絶対値の平均と標準偏差を表 3 に示す.

図 10 から, 仮説 *a*, *b*, *c*, *d* は, それぞれ対応する人

表 3 頭部の大きさの検出誤差
Table 3 Errors in detecting human head sizes.

仮説	平均 [pixel]	標準偏差 [pixel]
a	0.136	0.0603
b	0.143	0.0756
c	0.121	0.0569
d	0.120	0.0540

物頭部 A, B, C, D をほぼ正確に追跡していることがわかる。これを定量的に調べた図 11 (a), および、表 2 をみると、追跡中、人物 A, B に関する検出位置の誤差は、ほぼすべてのフレームにわたって、10 画素以下であり、人物 C, D に関しては、およそ 10 画素前後であることがわかる。これを実際の長さでみると、人物 A, B に関しては 5cm 程度であり、人物 C, D に関しては 10cm 程度であった。人物 C は、人物 A, B に比べ、カメラからの距離が遠かったため、検出位置の誤差が大きくなったと考えられる。また、人物 D の検出位置の誤差が特に大きいのは、襟の部分の明度勾配が大きいため、確信度の高い部分が頭部の真の位置よりやや下になってしまっているためであると考えられる。人物がカメラの前方 2m ~ 4m あたりに現れたことを考えると、この結果は、提案手法によって、人物の増減やオクルージョンに対する頑健な追跡が実現されていることを示すといえる。一方、図 11 (b) から、検出された頭部の大きさの誤差は、真の値の 2 割程度になっていることがわかる。また、表 3 から、検出された頭部の大きさは、安定していることが読み取れる。画像中の人物頭部に当てはまる楕円の長軸の長さは約 15 画素であったので、大きさの誤差は、実際には、5cm 程度となる。全体的に、仮説の大きさは、真の大きさに比べ、やや小さめになっている。これは、人物の顔と髪の境界部分に沿うような楕円に対して、楕円周上の明度勾配に対する評価が高くなってしまったためであると考えられる。しかし、この程度の誤差は、追跡結果に影響を及ぼすほどの大きさではなく、頭部の大きさは、十分正確に、しかも安定して検出できているといえる。

一方、320 × 240 画素のカラー画像 1 枚あたりに要する頭部検出の処理速度は、およそ 80 ミリ秒であった。これは、1 秒あたり約 13 フレームを処理することに相当し、提案手法はほぼ実時間で追跡を実現しているといえる。

これ以外にもいくつか場所を変え、あるいは状況を変え、実験を行った。既に述べたように、カメラの設置状況に応じて、仮説を生成するかどうかを判定する、

確信度に関するしきい値の調整が必要であったが、いずれの場合にも、上にあげた実験例と同様、人物頭部をほとんど正しく、しかも安定に追跡している結果が得られた。このことから、提案手法は、カメラ設置状況やカメラの前を通行する人物の状況に依存せず、有効性であることがわかる。

7. おわりに

複数対象追跡では、システムは、複数の異なる仮説を同時に保持しなければならない。その際、仮説と対象との 1 対 1 の対応関係を確立していない場合、仮説を、相対評価ではなく、絶対評価しなければならない。なぜなら、ある対象が画面中に存在するか否かは、他の対象の存在に依存しないからである。このような問題意識に基づいて、本稿では、確信度を導入した。

追跡中に生成された仮説と対象とを 1 対 1 に対応させるためには、仮説間の相互作用の導入が必須である。仮説から得られる画像情報の時間的、空間的連続性と一貫性を用いた仮説群の相互作用によって、誤検出に対応する仮説を排除するとともに、画像上の対象に対応する仮説のみを保持することができる。これにより、仮説と対象とを 1 対 1 に対応させることができ、対象数が変化したり、オクルージョンが起きたりする状況下でも、安定で頑健な複数対象追跡を実現することができる。

本稿では、撮影中、環境は変動しないことを前提としていた。しかし、本手法の応用が期待される交通管制システムなどの監視システムでは、一般に、照明や背景が大きく変動する。このような環境の変動が起こる状況下でも、頑健に複数対象追跡を実現することができるように、提案手法を拡張することが今後の課題である。また、追跡の機能を高めるためには、能動カメラを用いて、カメラの視線方向を制御し、対象を追跡することが考えられる。さらに、ズームを制御することも有効である。本稿で提案した手法と、カメラパラメタの能動的な制御とを組み合わせる複数対象追跡を実現する手法も、今後、検討していく予定である。

謝 辞

京都大学松山研究室 木村充宏氏には、実験データの収集・整理を手伝っていただいた。ここに記して謝意を表す。なお、本研究の一部は、科学研究費補助金 特定領域 13224051、および、基盤研究 (A) 13308017 の補助を受けて行った。

参 考 文 献

- 1) S. Birchfield: Elliptical Head Tracking Using Intensity Gradients and Color Histograms, *Proc. of CVPR '98*, pp. 232-237, 1998.
- 2) T. J. Cham and J. M. Rehg: A Multiple Hypothesis Approach to Figure Tracking, *Proc. of CVPR*, Vol. 2, pp. 239-245, 1999.
- 3) Y. Cui, S. Samarasekera, Q. Huang and M. Greiffenhagen: Indoor Monitoring via the Collaboration between a Peripheral Sensor and a Foveal Sensor, *Proc. of the IEEE Workshop on Visual Surveillance*, pp. 2-9, 1998.
- 4) L. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Hariatoglu and M. J. Black: Visual Surveillance of Human Activity, *Proc. of the 3rd ACCV*, Vol.2, pp. 267-274, 1998.
- 5) D.M. Gavrilu: The Visual Analysis of Human Movement: A Survey, *Computer Vision and Image Understanding*, 73 (1999), 1, pp 82-98.
- 6) I. Haritaoglu, D. Harwood and L. S. Davis: W⁴S: A Real-Time System for Detecting and Tracking People in 2 $\frac{1}{2}$ D, *Proc. of the 5th ECCV*, Vol. 1, pp. 877-892, 1998.
- 7) I. Haritaoglu, D. Harwood and L. S. Davis: An Appearance-based Body Model for Multiple People Tracking, *Proc. of the 15th ICPR*, Vol. 4, pp. 184-187, 2000.
- 8) M. Isard and A. Blake: Contour Tracking by Stochastic Propagation of Conditional Density, *Proc. of the 4th ECCV*, Vol. 1, pp. 343-356, 1996.
- 9) M. Isard and A. Blake: ICondensation: Unifying Low-Level and High-Level Tracking in a Stochastic Framework, *Proc. of the 5th ECCV*, Vol. 1, pp. 893-908, 1998.
- 10) M. Isard and A. Blake: Condensation-Conditional Density Propagation for Visual Tracking, *Int. J. of Computer Vision*, 29 (1998), 1, pp. 5-28.
- 11) J. MacCormick and A. Blake: A Probabilistic Exclusion Principle for Tracking Multiple Objects, *Proc. of the 7th ICCV*, pp. 572-578, 1999.
- 12) T. Matsuyama: Cooperative Distributed Vision —Dynamic Integration of Visual Perception, Action, and Communication —, *Proc. of Image Understanding Workshop*, pp. 365-384, 1998.
- 13) M. Minoh and Y. Kameda: Distance Learning Environment Based on The Interpretation of Dynamic Situation of Lecture Room, *Proc. of the Third International Workshop on Coopera-*

tive Distributed Vision, pp. 283-301, 1999.

- 14) K. Yachi, T. Wada and T. Matsuyama: Human Head Tracking using Adaptive Appearance Models with a Fixed-Viewpoint Pan-Tilt-Zoom Camera *Proc. of the 4th International Conference on Automatic Face and Gesture Recognition*, pp. 150-155, 2000.

(平成 12 年 2 月 4 日受付)

(平成 12 年 5 月 11 日採録)



杉本 晃宏 (正会員)

1987 年東京大学工学部計数工学科卒業。1989 年同大学院工学系研究科修士課程修了 (数理工学専攻)。同年、日立製作所基礎研究所に入社。1991 年～95 年、ATR に出向。1999 年京都大学講師、2002 年国立情報学研究所助教授、現在に至る。博士 (工学)。視覚情報処理や離散システム・アルゴリズムなどに興味をもち、数理的手法に基づいたコンピュータビジョンの研究に従事。2001 年情報処理学会論文賞。



谷内 清剛

1999 年京都大学工学部電気電子工学科卒業。2001 年同大学院情報学研究科知能情報学専攻修士課程修了。同年、ソニー (株) 入社。現在、ブロードバンドアプリケーション研究所勤務。実時間対象追跡の研究に従事。



松山 隆司 (正会員)

1976 年京大大学院修士課程修了。京大助手、東北大助教授、岡山大教授を経て 1995 年より京大大学院電子通信工学専攻教授。現在同大学院情報学研究科知能情報学専攻教授。2002 年学術情報メディアセンター長、評議員。工博。画像理解、人工知能、分散協調視覚の研究に従事。1980 年情報処理学会創立 20 周年記念論文賞、1990 年人工知能学会論文賞、1993 年情報処理学会論文賞、1994 年電子情報通信学会論文賞、1995 年第 5 回国際コンピュータビジョン会議 Marr Prize、1996 年国際パターン認識連合 Fellow、1999 年電子情報通信学会論文賞、2000 年画像センシングシンポジウム優秀論文賞。情報処理学会理事。