

スティックと楕円の統合モデルによる 外見的特徴の少ない非剛体群追跡法

北島利浩[†] 浮田宗伯^{†‡} 木戸出正継[†]
奈良先端科学技術大学院大学情報科学研究科[†]

本研究では、追跡対象のサンプル画像集合から、対象の骨格線を表すスティックモデルと対象の領域を表す楕円モデルを統合した対象モデルを自動的に作成し、そのモデルを用いて複数の対象を同時に追跡する手法を提案する。遮蔽などに頑健な追跡手法として、各瞬間における対象追跡結果を確率的に表現する手法が近年注目を集めている。しかし、従来の手法は対象のテクスチャ、色、形状などの外見的特徴の検出結果に強く依存しており、そのような外見的特徴が少ない物体に対しては適用できない。また、同様の手法により対象の位置、領域だけでなく詳細な姿勢も推定する研究も行われているが、多くの従来法では追跡対象の剛体性が仮定されており、非剛体（骨格は存在するが、関節の位置が明確にはわからない物体とする）の姿勢までは推定できない。本研究の対象モデルは一対象の領域を複数部位の連結集合として表現しているため、非剛体の姿勢変化に対応でき、さらに各部位の検出信頼性を評価しているため外見的特徴が少ない物体に対して適用した際にも致命的な推定失敗が起こりにくい。提案手法の有効性を示すために、外見的特徴の少ない非剛体として金魚を選んで評価実験を行った。実験では、まず金魚のサンプル画像集合から対象モデルを自動的に作成し、複数の金魚による遮蔽が発生する環境下での有効性を確認した。

Tracking Non-rigid Objects without Sufficient Appearances based on the Stick and Ellipse Model

Toshihiro Kitajima[†], Norimichi Ukita^{†‡} and Masatsugu Kidode[†]

Graduate School of Information Science, Nara Institute of Science and Technology[†]

We propose a non-rigid objects tracking method with the object model generated automatically from a set of sample images. Our model consists of multiple sticks and ellipses each of which represent the skeleton and the region of an object, respectively. Previous tracking methods cannot track an object without characteristic appearances (e.g., texture patterns, shape, and so on) enough to be detected easily because they have to extract the appearance features. Moreover, it is difficult for the previous methods to estimate the posture of a non-rigid object. With the proposed model, on the other hand, our method can estimate the posture of a non-rigid object without sufficient appearances because (1) each portion of the model can fit each part of a non-rigid object and (2) the reliability of each portion is evaluated. To confirm the effectiveness of the proposed method, we conducted several experiments with goldfishes. The tracking system generated the model of a goldfish automatically and then could track some goldfishes that cause occlusions.

1 はじめに

近年、作業時間や労力を大幅に削減する目的でビジョンシステムを用いた監視、観察の研究が数多く行われている。侵入者や不審者を検知するセキュリティシステム [1]、スポーツシーンの解析 [2]、道路環境における交通監視システム [3]、動物の生態観察など様々な研究が行われている。観測画像を解析する

ことにより追跡対象の数、軌跡、速度などを得ることができれば、シーンの状況、対象の行動特性、対象の意志、対象の行動予測などの情報を自動的に取得することが可能になる。さらに観測対象の姿勢まで推定することができれば、より詳細な対象に関する情報を自動的に得ることができるため、システムの応用範囲を大きく広げることができる。監視、観察において追跡対象の数が1つであれば、追跡対象の情報は比較的簡単に取得できる。しかし、複数の

[†]科学技術振興機構さきがけプログラム (PREST, JST)

追跡対象が存在するシーンにおいては、対象の移動に伴って相互遮蔽が発生するために、追跡対象の数、軌跡などの情報の取得が困難になる。その問題に対処するために、遮蔽に対して頑健な複数対象追跡手法がいくつか提案されている。中でも、対象の存在の可能性を確率論の枠組みに組み込んで追跡を実現するという手法 [4, 5] が近年提案され注目を集めている。この手法では、追跡対象の位置、姿勢に関する仮説を1つだけ保持するのではなく、仮説を確率密度分布の形で表現し、複数の仮説を同時に保持することにより、遮蔽に頑健な追跡を可能にしている。この手法は追跡対象の遮蔽に頑健かつ汎用性が高いので、この手法を用いた対象追跡の研究は数多い。しかし、それらの従来研究では、取り扱っていない重要な問題点がいくつか存在する。それを以下に示す。

I. 非剛体の姿勢は推定できない：
従来研究では、追跡対象の領域をシンプルかつ単一の形状（楕円・矩形など）として表現しているため単純な形状の剛体、または複雑な形状をした対象の一部分しか追跡することができない。よって、追跡対象が非剛体（骨格は存在するが、関節の位置が明確にはわからない物体）の場合は、対象の全領域を1つの形状で表現することが困難であるので、その姿勢は推定できない。

II. 外見的特徴が少ない物体には対応していない：
追跡をする際に、物体のテクスチャ、色、形状などの外見的特徴が用いられている。例えば、人間の追跡では、人間の肌の色、頭部の形状のように他の領域や背景との識別が容易な特徴を利用して追跡を行っている [4, 6]。従来研究では、対象の検出率の高い外見的特徴をあらかじめ人が判断して、システムに知識として与えている。すなわち、追跡対象に対して検出率の高い外見的特徴を見つけていることができれば追跡はうまくいくが、追跡シーンの環境や追跡対象そのものが変わってしまった場合には、また検出率の高い特徴を人間が見つけなければならない。さらに、発見した特徴が、別途定義される追跡手法における対象識別に適していることを保証するのは困難である。また本質的な問題として、そのような検出率の高い外見的特徴が無い物体、またはそのような特徴が発見されにくい物体も数多く存在する。動物などはテクスチャの特徴が少なく、ほぼ単色で表現されるようなものが多い。そのような物体に対して、従来研究の手法をそのまま適用しても追跡そのものが不可能になりうる。

そこで本研究では、問題 I, II に対処するために以下のような処理をする。

まず、問題 I に対処する処理について述べる。

I - A. 非剛体の関節と仮定できる位置で非剛体を分割し、分割された各領域に対して部位モデル（ここでは、単一の形状で表現できるモデルとする）を自動生成する。

I - B. 関節箇所での各部位モデルの角度変化を許容した部位モデル集合を統合することにより、非剛体を1つの対象モデルとして表現する。

I - C. 非剛体の骨格線（ここでは、物体領域の中心を表す線とする）を用いて姿勢を推定する。

つぎに、問題 II に対処するために以下のような処理をする。

II - A. 少ない外見的特徴でも対象らしさを評価でき、どの物体にも存在するような以下の特徴量を用いる。

- (a) 骨格線情報
- (b) 色ヒストグラム
- (c) 面積

II - B. 追跡時に用いる特徴量に基づいてモデルを構成することにより、生成されるモデルや特徴量が追跡手法における対象識別に適していることを保証する。

II - C. 信頼できる部位モデルを基準にその他の部位モデルの位置を定める。これにより、特徴量が少なく位置が特定できない部位モデルも対象の各部位と正確な対応が得られる可能性が高くなる。

上記の問題対処法に基づいて、追跡に用いるモデルを作成する。本研究では、対象物体の骨格線を表し関節と仮定できる箇所で分割されているスティックモデルと、色ヒストグラムと面積の情報を表すための楕円モデルの2つのモデルを作成し、それらを組み合わせた対象モデルを提案する。この対象モデルは対象が存在するサンプル画像の集合から自動生成することができる。追跡時には、生成された対象モデルと観測画像からの検出結果を当てはめることにより、各検出対象の位置・姿勢を逐次推定する。本研究では、対象の初期検出には既存の手法を用いるのみとし、その後の時系列的な追跡の手法に主眼をおいて研究を行った。

2 対象モデルの生成

ここでは、サンプル画像集合（動画）から、対象モデルを生成する処理について述べる。処理の簡単化のため、サンプル画像集合には、追跡対象が1体だけ存在するものとする。また、画像集合中には、対象の様々な姿勢の画像が含まれている必要がある。

対象モデルは、スティックモデルと楕円モデルから成る。スティックモデルには、以下のような役割がある。

- 対象の姿勢変化に対応できるように、各スティック（ここでは、関節間の領域を表す）の長さ、すべてのスティックの接続関係の情報を保持し、関節箇所各スティックの角度変化を許容する。
- 特徴量としての骨格線情報を表す。

楕円モデルには、以下のような役割がある。

- 特徴量としての対象領域の面積を表す。

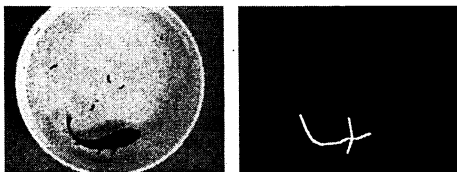


図 1: サンプル画像 図 2: スティックモデル

- 特徴量としての色ヒストグラムの情報を表す。

対象の関節箇所で分割された部位が細長い形状であったり、凹凸の激しい形状である場合には、1つの楕円でその部位全体を表現できない。そのために楕円同士が隣接しない箇所が発生するので、楕円モデルのみでは、それぞれの部位を連結させることができず、対象の各部位の接続関係を明確にすることができない。そのため、楕円モデルだけでは対象の姿勢変化に対応できない。逆にスティックモデルだけでは、面積、色ヒストグラムの情報を保持できないので、2つのモデルを統合して対象モデルとした。

2.1 スティックモデルの生成

追跡の対象が人間のように関節情報が自明である物体であれば、あらかじめスティックモデルを生成しておくことも可能である [7]。しかし、本研究では関節の位置が明確にはわからない非剛体を対象としているので、関節の位置、関節間の長さなどを対象画像から推定しなければならない。そこで、骨格を形成する骨の部分は直線で近似できることに着目し、対象領域に対して細線化処理を行い見かけ上の骨格線を抽出する。その骨格線に対して2次元の直線 Hough 変換を行うことにより直線、つまり剛体で近似できる部分を抽出し、その部分をスティックとして表現する。一つの対象に対して複数のスティックが存在するので、それらを連結し接続点において各スティックの角度変化を許容することによりスティックモデルが生成される。図1のサンプル画像からは図2のスティックモデルが生成される。1枚のサンプル画像からスティックモデルが1つ生成されるが、スティックの数、スティックの接続関係はサンプル画像中に映っている対象の姿勢によって異なる。様々な姿勢に対応できるようにサンプル画像集合の中から複雑な姿勢を表す、すなわち連結されるスティックの数が最も多いスティックモデルを選択し、それを最終的な対象のスティックモデルとして決定する。本手法では、スティックモデルの各スティックの重心のうち、対象領域の重心に最も近いものを基準スティックとし、その他のスティックにはそれぞれ ID 番号が割り当てられる。

2.2 スティックの信頼度

追跡対象の骨格線は対象領域の中心線を抽出することにより求めているので、追跡対象の姿勢によっては、自己遮蔽が発生して、スティックモデル中のスティックに対応する骨格線が検出できない場合もある。

また、対象領域に対して細線化処理を行うときに雑音の影響で実際には存在しないスティックをモデル化してしまう場合もある。そのような場合、骨格線にマッチしないスティックや誤った骨格線にマッチしてしまうスティックも存在するので、2.1項で決定したスティックモデルの各スティックをすべて同じ重みで信頼してしまうと追跡失敗の原因になる可能性がある。本手法では、スティックごとに信頼の度合いを表す重みを算出することにより、この問題に対処する。以下、信頼の度合いを表す重みをスティックの信頼度と呼ぶ。

ここで、基準スティックのパラメータを $\nu^{std} = (x, y, \theta^{std})^T$ として、位置 (x, y) と角度 θ^{std} の3つのパラメータで表す。他のスティックはスティック同士の接続関係を保つ目的で、隣接するスティックとの接続角度のみをパラメータとする。スティックモデルのスティックの数を n として、スティックの ID 番号が i ならば、そのパラメータは $\theta^i (i = 1, 2, \dots, n) (i \neq std)$ として表される。また、スティックモデル全体のパラメータは $\Phi = \{\theta^1, \theta^2, \dots, \nu^{std}, \dots, \theta^n\}$ で表す。

各スティックの信頼度を求めるために、サンプル画像集合に対して、2.1項求めたスティックモデルをマッチングさせる。すべてのスティックが対象領域に納まるようにスティックモデルの位置、各スティックの角度を変化させてマッチングを行う。マッチングの際に評価する照合度は以下の二つである。

- 対象領域上の画素とスティック上の画素の照合度

L は対象領域上の画素、 i はスティックの ID 番号、 S^i は i のスティック上の画素を表す。スティックの数を n として、すべてのスティックの照合度 $M_{ter}^i (i = 1, 2, \dots, n)$ の合計値 M_{ter} を式 (2) によって求める。

$$M_{ter}^i(\Phi^i) = S^i(\Phi^i) \cap L / S^i(\Phi^i) \quad (1)$$

$$M_{ter}(\Phi) = \sum_{i=1}^n M_{ter}^i(\Phi^i) \quad (2)$$

- 対象の骨格線上の画素とスティック上の画素の照合度

B は対象の骨格線上の画素を表し、すべてのスティックの照合度 $M_{skel}^i (i = 1, 2, \dots, n)$ の合計値 M_{skel} を式 (4) によって求める。

$$M_{skel}^i(\Phi^i) = S^i(\Phi^i) \cap B / S^i(\Phi^i) \quad (3)$$

$$M_{skel}(\Phi) = \sum_{i=1}^n M_{skel}^i(\Phi^i) \quad (4)$$

M_{ter} の値に対して閾値処理を行い、スティックモデル中のすべてのスティックが対象領域に納まるようにパラメータを限定する。限定されたパラメータの中で M_{skel} の値が最も高いパラメータを対象領域にマッチングするパラメータとして決定する。そのパラメータにおける骨格線との照合度 M_{skel}^i の値に対し

て閾値処理を行い、 i のスティックに対応する骨格線が存在するかどうかを判定する。この処理をスティックモデルのすべてのスティックに対して行い、骨格線の存在を判定する。すべてのサンプル画像に対して同じ処理を行い、各スティックについて骨格線が存在したサンプル画像の数 S_{ex}^i をサンプル画像の総数 S_n で割ることによって、各スティックの骨格線の存在確率 γ_{skel}^i を求める。各スティックの信頼度 W_{stck}^i はすべて合計すると確信度の最大値である 1.0 になるように、 γ_{skel}^i から式 (6) を用いて求める。

$$\gamma_{skel}^i = S_{ex}^i / S_n \quad (5)$$

$$W_{stck}^i = \gamma_{skel}^i / \sum_{i=1}^n \gamma_{skel}^i \quad (6)$$

2.3 楕円モデルの生成

本手法では、対象を表現する特徴量として骨格線情報に加えて色ヒストグラムを面積の情報を用いる。特に面積の情報は、対象の各部位の領域をできるだけ忠実に表現できた方が特徴量として有効になるので、対象の輪郭線までの領域を正確に取得する必要がある。そこで、任意の形状に適用可能、かつ対象の輪郭線までの領域をできるだけ広く取得するために、輪郭線に内接するような楕円を部位モデルとして用いる。そのために楕円膨張法 [8] を用いる。楕円膨張法は輪郭線に内接する円を描き、その円の長軸、短軸、角度、中心の位置を輪郭線に楕円がフィットするように変化させる手法である。本研究では、楕円を膨張させるときは長軸と短軸のみを変化させるものとした。輪郭線までの領域を広く取得しようとする、必然的に楕円の中心点は対象領域の中心線を表すスティック上の点となり、角度はスティックの角度とほぼ等しくなるので中心の位置と角度は変化させずに、中心点はスティック上の点に、角度はスティックと同じ角度に固定した。

もう1つの特徴量である色ヒストグラムによってより有効な対象情報を表現するためには、楕円モデルの領域内に特徴的な色情報を多く含むことが必要になる。そこで、楕円内部に特徴的な色情報が多く存在するような中心点を選択することにする。まず、対象領域を H S V 表色系で表現して色相 (H) のヒストグラムを作成する。次に頻度の高いものから順に削除していき、総頻度の 7、8 割程度を削除することにより、対象領域の中で稀な色情報、つまり特徴的な色情報を抽出することができる。金魚に対して適用した場合、図 3 のように目の黒い部分とヒレと尻尾の薄い赤色の部分が特徴的な色となる。楕円の中心点はスティック上の点で、角度はスティックの角度と同じであるので、対象領域に対するスティックモデルのパラメータを決定しておく必要がある。そこで 2.2 項の手法を用いて決定する。楕円の中心点はスティック上の点であれば、どの点でもよいので候補が複数存在する。それぞれの点において楕円を膨張させ、楕円内部の特徴的な色情報が存在する画素の数をカウ

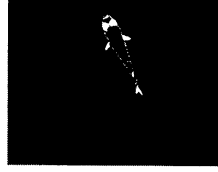


図 3: 特徴的な色相領域

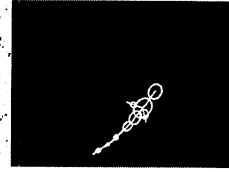


図 4: 対象モデル

ントする。そのカウント数が最多の点を楕円の中心点として決定し、その楕円の大きさを楕円の長軸、短軸として決定する。1枚のサンプル画像のみの情報では様々な姿勢に対応できない恐れがあるので、すべてのサンプル画像において決定した各スティックの楕円の中心点、長軸、短軸を平均することによって楕円モデルを完成させる。完成した楕円モデルをスティックモデルと組み合わせると図 4 のようになる。

特徴的な色情報を多く含む楕円は対象と正確な対応が得られる可能性が高いので、特徴量を計算するときの信頼度 W_{ci} を楕円内部に含まれる特徴的な色画素の数に基づいて高く設定する。

2.4 色情報、基本姿勢の取得

対象の追跡時に特徴量として用いる色ヒストグラムは、追跡を開始する前にあらかじめ学習しておく必要がある。そのため、すべてのサンプル画像に対して、2.2 項のマッチング方法を用いて、対象モデルをマッチングさせる。色ヒストグラムは、H S V 表色系の色相 (H) の部分だけをヒストグラムにしたものである。色ヒストグラムは様々な状況に対応できるように、マッチングした楕円内部の色ヒストグラムをすべてのサンプル画像について平均をとる。そうすることにより、追跡時に用いる色ヒストグラムの情報が得られる。また、対象の基本姿勢と各スティックの接続角度の最大値、最小値が得られると仮説生成のときに探索範囲が限定され、処理時間の短縮につながる。これらの情報もあらかじめ学習しておく。基本姿勢はすべてのサンプル画像に対して、各スティックの接続角度の平均値を求めることによって決定される。

3 追跡

本研究では、追跡手法として、確信度分布を用いた追跡手法 [4] の枠組みを用いる。追跡の処理は、画像観測によって現在の画像を取得するところから始まり、仮説生成、確信度分布内のパラメータをサンプリングして確信度を計算、確信度分布の再構成、確信度分布の時間方向への伝播、画像観測と続く。

3.1 確信度の評価

追跡をするには、パラメータ空間中の各パラメータが観測画像から検出された対象情報に対応付けられる可能性を表す確信度が必要になる [4]。ここで i をスティックの ID 番号、 i のスティックのパラメータを $P^i = (x, y, \theta)^T$ とする。後述するが、ここでは、2.2 項で設定したパラメータと異なり、すべてのスティックが位置と角度の 3 つのパラメータをもつ (3.3 項参

照)。確信度は以下の3つの評価によって決定される。

- (a) スティックモデルと骨格線のマッチング
外見的特徴の少ない物体に対して、対象の骨格線の法線方向において精度良くパラメータを推定し、さらに非剛体の姿勢を検出するためにスティックモデルと骨格線のマッチングを行う。スティックモデルと骨格線による追跡対象らしさの評価関数 E_{skel}^i を

$$E_{skel}^i(P^i) = \{A_{stck}^i(P^i) \cap A_{skel}\} / A_{stck}^i(P^i) \quad (7)$$

とした。 A_{stck}^i は、 i のスティックの面積を表し、 A_{skel} は対象の骨格線の面積を表す。

- (b) 楕円内部の色相ヒストグラム
外見的特徴が少ない物体においても特徴的な色をもった部位は色情報を用いると精度よく推定することができる。その部位を特定するために色相ヒストグラムを用いる。追跡対象の楕円内部の色相ヒストグラムの関数 g_{color}^i は、

$$g_{color}^i(P^i) = \sum_{s=1}^S \min[I^i(s), M^i(s)] \quad (8)$$

とした。ここに、 S はヒストグラムにおける色相の数を表し、 s はヒストグラム内の s 番目 ($s = 1, 2, \dots, S$) の色相、 I^i は i のスティックの楕円内部の色相ヒストグラムの度数、 M^i はあらかじめ得ておいた色相ヒストグラムの度数を表す (2.4 項参照)。また、あらかじめ得ておいた色相ヒストグラムの度数の合計値を ϕ_{color}^i とした。 g_{color}^i と ϕ_{color}^i を用いて、楕円内部の色相ヒストグラムによる追跡対象らしさの評価関数 E_{color}^i を求める際、2.3 項で算出した楕円の信頼度を用いることにより、特徴的な色情報をもつ楕円をより信頼するようにする。そこで評価関数 E_{color}^i を

$$E_{color}^i(P^i) = W_{color} \times \{g_{color}^i(P^i) / \phi_{color}^i\} \quad (9)$$

とした。

- (c) 楕円の面積と対象領域の面積
対象物体の各部位における領域の形状がそれぞれ異なる場合には、検出結果とモデル間で各部位の面積を照合させることによって各対象の位置・姿勢推定に有効な情報が得られる。楕円モデルと対象領域の面積による追跡対象らしさの評価関数 E_{area}^i を

$$E_{area}^i(P^i) = \{A_{elli}^i(P^i) \cap A_{objct} - A_{elli}^i(P^i) \cap \bar{A}_{objct}\} / A_{elli}^i(P^i)$$

とした。ここに、 A_{elli}^i は i のスティックの楕円の面積を表し、 A_{objct} は、対象領域の面積を表す。図5に示すように、 $A_{elli}^i \cap A_{objct}$ は照合した領域の面積、 $A_{elli}^i \cap \bar{A}_{objct}$ は、はみだした楕円モデルの面積を表す。

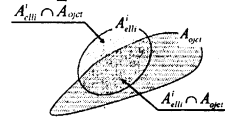


図5: 楕円と物体領域の面積

1つのスティックにおける追跡対象らしさを表す確信度 C^i は、上記の3つの評価関数を合わせて

$$C^i(P^i) = W_{skel} E_{skel}^i(P^i) + W_{color} E_{color}^i(P^i) + W_{area} E_{area}^i(P^i) \quad (10)$$

とした。 $W_{skel}, W_{color}, W_{area}$ は各評価関数の重みを表している。

また、対象モデル全体の確信度を表す評価関数 C は、2.2 項で求めた各スティックの信頼度 W_{stck}^i を用いて、

$$C(P) = \sum_{i=1}^n W_{stck}^i C^i(P^i) \quad (11)$$

とした。ここで $P(P^0, P^1, \dots, P^n)$ は各スティックのパラメータすべてを表す。

3.2 仮説生成

まず、追跡を開始した初期フレームにおいては、前フレームからの予測が使えないため、1枚の画像のみから、対象のパラメータを推定する必要がある。初期フレームにおいて追跡対象と高い確信度によって対応付けられているパラメータを仮説 (パラメータ空間における確信度分布) として複数保持し、その仮説を用いて、その後のフレームで追跡を行う。

初期フレーム以外においても、画面に対象が新たに現れる場合がある。その場面に対応するためには、新たに対象が現れた領域を特定し、その領域に仮説を生成する必要がある。そこで、ある時刻 t_k の画像中から抽出された全検出領域から、仮説が存在する領域を除いた部分を新たに対象が現れた可能性がある領域と見なし、その対象に対応する仮説を生成する。本手法では、初期フレームにおける対象物体の領域と以降のフレームで新たに現れた画像領域を対象未検出領域と呼び、2つの領域に対して同じ処理を行うこととする。

対象未検出領域に対して、追跡対象に対応する確実性が高いパラメータを推定する。マッチングには、2.2 項の手法を使う。ここで、2.4 項で求めた基本姿勢と各スティックの接続角度の最大値、最小値を用いることにより、探索範囲が限定されて処理時間が短縮される。マッチングにより、推定されたパラメータを Φ とする。

Φ は、基準スティックは3つ、他のスティックは1つのパラメータで表されたものであるが、後の処理のために、すべてのスティックが3つのパラメータをもつように、スティックの接続関係を用いて変換する。変換されたものを P とする。

3.3 確信度分布の再構成

対象モデル生成時におけるパラメータ Φ は、基準スティックは3つ (x, y, θ)、他のスティックは1つのパラメータ (θ) で表していたが、追跡時には、すべてのスティックが3つのパラメータをもつパラメータ P を用いている。追跡時に対象モデル生成時と同じパラメータ設定を用いない理由は確信度分布を利用する点にある。

追跡時に Φ のパラメータ設定を用いると、基準スティック以外のスティックは接続しているスティックのパラメータによって位置が変わってしまう。すなわち、全スティックの全パラメータが依存関係にあるため、次元が全パラメータ数 ($3+n-1$, ただし n はスティックの数) の高次元空間内に確信度分布が構成される。その確信度分布内のパラメータは1つのスティックの位置、姿勢に関しては、重複している部分が多く、冗長性が高い。そのために処理時間が多くかかるという問題が発生する。

一方、 P のパラメータ設定の場合は、各スティックのパラメータを独立して扱うことができるので、3次元のパラメータ空間が、スティックの数と同じ数だけ存在する。このことにより、1つのスティックの位置、姿勢に関して、重複する部分が多くなるので、冗長性が低くなり、処理時間も短縮されるためリアルタイム追跡に適している。また、スティックごとに確信度分布の分散を変化させることができるため、スティックごとの探索範囲の拡大縮小が可能になる。

伝播によって推定された確信度分布は、現在の時刻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させる必要がある。従来の手法では、確信度が最大値になるパラメータを確信度分布の中心にして、確信度分布の再構成を行っている。しかし、本研究では追跡時に、各スティックの x, y, θ を変化させるパラメータ P を用いているために、各スティックにおける確信度の最大値を分布の中心にして再構成をすると、スティック同士の連結性が保てなくなり、バラバラになってしまう可能性がある。そのためにスティックの連結性を保ったまま確信度分布を再構成させる必要がある。

あるスティック i を信頼できるスティックであると仮定して信頼スティック rt とする。信頼スティック rt の確信度分布中で確信度が最も高い点に対応するパラメータを信頼スティック rt の代表パラメータとして決定する。すると信頼スティック rt に隣接するスティック a は、 rt と連結しなければならないという制約から確信度分布内のパラメータが限定される。このとき、図6のように信頼スティック rt を親ノードとして、それに隣接するスティック a の限定されたパラメータ $a1, a2, a3$ を子ノードとして木構造で記述する。ノードにはそのパラメータにおけるスティックの確信度の情報が含まれる。次にスティック a に隣接す

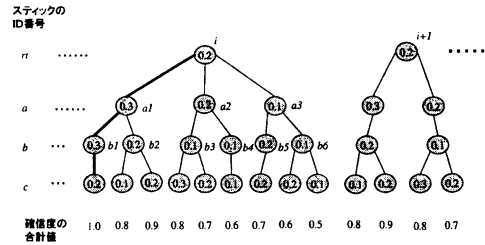


図6: 木構造

るスティック b のパラメータも同様に限定する。 $a1$ のパラメータを親ノードとし、 $a1$ と連結する制約から限定されるパラメータ $b1, b2$ を子ノードとして木構造で記述する。以下、順に $a2, a3$ に連結するパラメータを限定し、木構造で記述していく。この処理をすべてのスティックが木構造で記述されるまで繰り返す。そうすると信頼スティックを根として、それぞれ隣接するスティック同士の連結性が保たれたパラメータの木が完成する(図6参照)。この木は信頼スティック rt のパラメータを決定した段階で、接続関係のみから限定されるパラメータ群を表している。

次に $i+1$ のスティックを信頼スティック rt として木を作成し、さらに $i+2$ を根とした木を作成していき、すべてのスティックを根とした複数の木を作成する。

木がすべて完成すると木の根から深い方へ順に、対応するパラメータの確信度を加算していく。最深部までいくと確信度の合計値が求まる。図6の太線で表される探索路のように、複数の木のすべての探索路の中で確信度の合計値が最も大きい探索路のパラメータは、すべてのスティックの連結性から限定されたパラメータ群、言い換えると対象を関節箇所で途切れることなく表現できるパラメータ群の中で最も正確に対象に対応していると考えられることができる。よってこのパラメータを中心に確信度分布の再構成を行う。

この方法により、スティックのつながりを保ったまま確信度分布を再構成させることができる。さらに、一部のスティックに遮蔽が発生して確信度の最大値が実際に対象が存在する位置とは違う位置に発生してしまっても、他のスティックの情報を信頼することにより、後のフレームにおいて遮蔽が発生したスティックのパラメータを修正することが可能になる。すべてのスティックを信頼スティックに選択するのは、特定のスティックを信頼スティックとして選択しようとすると不確実性が残ってしまうことと、信頼スティックとして選択されたスティックの領域に遮蔽が発生したときに追跡失敗の可能性が高くなるという二点の理由からである。

3.4 確信度分布の時間方向への伝播

再構成させた確信度分布を1フレーム後の画像に伝播させるには、対象の移動予測に基づいて伝播させる。各スティックの連結性を保ったまま伝播させるので、基準スティックは3つ、他のスティックは1つの

パラメータで表されるパラメータ設定を用いる。ここで各スティックの中で最も確信度が高いものを基準スティックとして選択した。本項では、伝播における時間の前後関係が分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、 $t_k(k=1, 2, \dots, n-1, n, n+1, \dots)$ で表される。なお、現在の画像は時刻 t_n に撮影されたとする。各スティックの時刻 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には式(12)を用いた。確信度分布の伝播には、分布のピークを伝播させる必要があるので、 P_{peak}^i を i のスティックにおける分布のピークとして線形に予測した。

$$P_{peak}^i(t_{n+1}) = P_{peak}^i(t_n) + \frac{t_{n+1} - t_n}{t_n - t_{n-1}} \{P_{peak}^i(t_n) - P_{peak}^i(t_{n-1})\} \quad (12)$$

分散は確信度分布の再構成のときに変化させるので、時間方向への伝播のときは、変化させずにそのまま用いた。スケールも変化させずにそのまま用いた。

4 評価実験

提案手法の有効性を確認するために、外見的特徴の少ない非剛体として金魚を選択して屋内環境下で実験を行った。

4.1 対象モデルの生成

対象モデルを生成するために用いるサンプル画像の集合は、30fpsで10秒間撮影した300枚の静止画像(YUV422の16ビット、640×480pixelのカラー画像)を用いた。サンプル画像には、金魚の様々な姿勢が取得できるように、金魚が尻尾を曲げているものや色々な向きを向いているものを用いた。

本研究では、これらのサンプル画像から、金魚の赤い色を利用して色情報から対象領域を抽出した。対象領域に対して対象モデルを生成した結果、9本のスティックで表される図4のようなモデルが生成された。

スティックの信頼度は金魚のヒレの部位を表すスティックの値が極端に低くなった。これはヒレの部分が金魚の身体に隠れてしまう自己遮蔽が頻繁に起こったためにヒレの部分が画像上に存在しない場合が発生したためである。楕円の信頼度は金魚の頭を表すスティックの値が極端に高くなった。これは金魚の目玉の部分が黒く、他の部分の赤色に比べて稀な存在であったために重みが高くなったと考えられる。

4.2 追跡実験

対象モデル生成時に利用したサンプル画像集合とは異なる動画(30fps、2匹の金魚を観測)を撮影して追跡実験を行った。追跡中の各フレームには、生成された仮説の確信度分布のピークに対応する対象モデルを観測画像に重ね書きした。実験結果画像の観やすさを考慮して、対象モデルの楕円モデルは描かず、スティックモデルのみを描いた。また、確信度 C^i を求める式(10)における各評価関数の重み $W_{color}, W_{skel}, W_{area}$

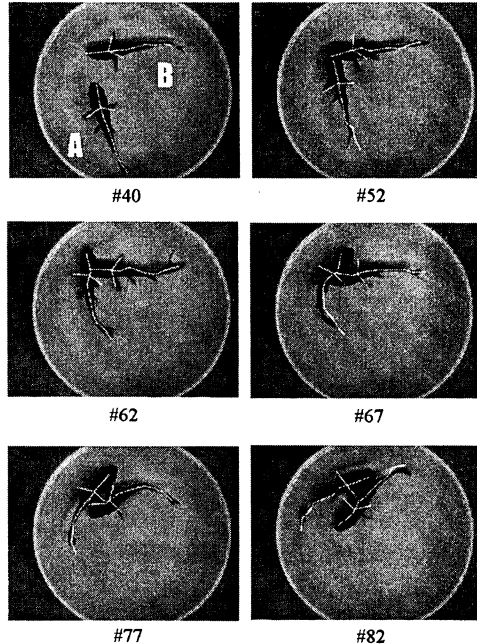


図7: 追跡中に遮蔽が発生する状況における追跡結果

は、色ヒストグラムと領域の面積の評価関数の重みを高くして、 $W_{color} = W_{area} = 2W_{skel}$ とした。

追跡対象を追跡しているときに、対象同士の遮蔽が発生するような状況で実験を行った。実験で得られた画像の一部を図7に示す。実験で撮影された映像は10秒間、つまり300フレームの画像であるが遮蔽が発生する直前のフレームを初期フレームとして追跡を行った。

初期フレームの#40において仮説Aと仮説Bが生成される。#52において対象同士の遮蔽が発生する。#62では、仮説A、Bの両方とも正しく対象の姿勢を検出しているが、#67では仮説A、Bともに遮蔽の影響で、頭部において正しく姿勢が推定されていないことがわかる。#77において、仮説Aでは、遮蔽の影響を受けていなかったスティックの情報を信頼することにより、頭部の姿勢が正しく修正されている。仮説Bも#82において頭部の姿勢が正しく修正されている。このように遮蔽の発生により対象の姿勢が正しく推定されなくても、その後姿勢を正しく修正することにより、遮蔽の発生する状況において正しく追跡が行える。

追跡中の仮説A、Bの確信度の推移に注目して、図8、9にそれぞれの確信度分布の最大値を折れ線グラフで表した。グラフの背景において色が明るくなっているフレームは、そのフレームにおいて追跡が失敗していることを示す。追跡失敗の判定は人間によって行い、スティックが対象領域に納まっているか、スティックは対象の姿勢を表現できているか、の2点の

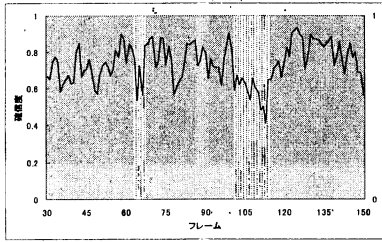


図 8: 仮説 A の確信度の推移

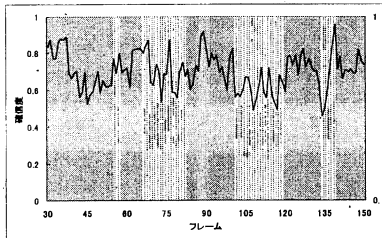


図 9: 仮説 B の確信度の推移

判断基準に基づいて判定を行った。

図 10 に追跡失敗の例を示す。図 10 の#66 では、仮説 A は対象の姿勢変化と遮蔽の影響により追跡が失敗している。そのため図 8 の#66 付近において確信度の値が低くなっているが、後のフレームにおいて遮蔽の影響を受けていないスティックの情報を基に修正できていることがわかる。#108 では仮説 A、B がお互いの遮蔽の影響により追跡が失敗し、#136 では仮説 B が姿勢変化についていけずに追跡が失敗し、確信度の値が低くなっている。しかし、後のフレームにおいて失敗を修正していることが図 8、9 から確認できる。#55 では遮蔽の影響で仮説 B の姿勢推定が失敗しているが、仮説 A の金魚の頭部に対応したため、確信度は低くならなかったと考えられる。

この実験において追跡成功率は、仮説 A が 87 %、仮説 B が 74 %、平均で 81 % であった。追跡成功、失敗の判定は上記の方法と同様に人間によって行い、10 秒間の映像、つまり 300 フレームの映像について行った。また、その映像では金魚の頭部同士の遮蔽と頭部と尻尾による遮蔽が合計 50 フレームにわたり発生した。仮説 A と B で追跡成功率に差が生じているのは、仮説 B の金魚が仮説 A の金魚によって遮蔽されたので、仮説 B が対応する対象を見失う場面が多く発生したためと思われる。

また、カメラのズーム機能を用いることにより対象の大きさが変化した状況において追跡実験を行った。従来研究では、対象の大きさが変化した場合には人手によってモデルを生成し直す必要があるため非常に労力を要するが、本研究では大きさが変化した対象のサンプル画像から、大きさの変化に対応した対象モデルを自動生成することができた。その状況においても同様の追跡結果が得られ、本手法の有効性が確認された。

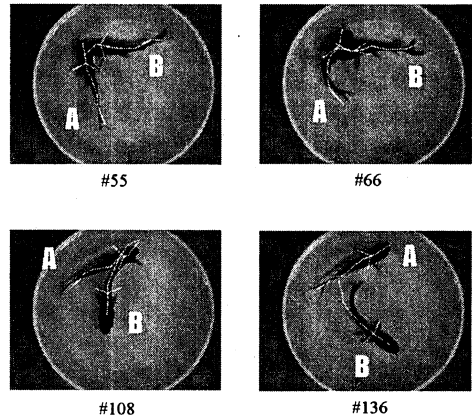


図 10: 追跡失敗例

5 おわりに

本研究では、スティックと楕円を統合した対象モデルをサンプル画像からの自動生成し、外見的特徴が少ない非剛体に対して適用ができる追跡手法を提案した。また、外見的特徴の少ない非剛体として金魚を選択し、本手法を用いて対象同士の遮蔽が発生する状況において追跡実験を行い有効性を確認した。今後の課題としては、対象の骨格線の接線方向に対する姿勢推定の精度の向上や、実時間追跡に向けての処理の高速化などが挙げられる。また追跡、姿勢推定の精度を高めるために、能動カメラを用いてカメラの視線方向、ズームを制御し、継続的に詳細な対象画像を観測して追跡を行うことも検討していく。本研究は JST の PREST プログラムによる支援を受けている。

参考文献

- [1] 伊藤, 上田: 広域監視用親カメラと追尾監視用子カメラによる協調監視システム; MIRU2002 Vol. II 229-234 (2002)
- [2] 須藤, 小沢: サッカーのシーン解析; MIRU2000 Vol. II 223-228 (2000)
- [3] 吉田, 影澤, 塘中, 池内: 局所特徴認識アルゴリズムによる車両の認識; 信学技報; PRMU2001-86 (2001)
- [4] 杉本, 谷内, 松山: 確信度付き仮説群の相互作用に基づく複数対象追跡; 情処学論, Vol.43 No.SIG 4(CVIM) 69-84 (2002)
- [5] Isard, M. and Blake, A: Condensation - conditional density propagation for visual tracking; IJCV, 28(1):5-28 (1998)
- [6] 馬場, 松尾, 大橋, 乃万, 松尾, 江島: HeadFinder 単眼視動画像を用いた複数人追跡; 画像センシングシンポジウム 2001, 363-368 (2001)
- [7] 大垣, 岩井, 谷内田: 形状モデルと動作モデルによる人物の姿勢推定; MIRU 98 Vol. II 429-434 (1998)
- [8] 原, 加藤, 井口: だ円膨張法を用いた動物シルエット画像の記述と認識; 信学論 D-II Vol. J74 No.3 366-375 (1991)