

疎テンプレートマッチングとその実時間物体追跡への応用

松原 康晴[†], 尺長 健[†]

多数の画素で構成されるテンプレートから、あらかじめ少数の画素を選択できれば、テンプレートマッチングに必要な計算コストを大幅に削減できると考えられる。本論文では、これを実現するため疎テンプレートマッチングの基本的な枠組みを示すとともに、疎テンプレートの作成法を連続領域モデルおよび不連続領域モデルに基づいて提案する。次に、疎テンプレートマッチングの特性を活かして、実時間でロバストな物体追跡を実現する方法として、段階的ランダム探索法と condensation 法を示し、自然環境下での顔追跡などの実験結果を報告する。

Sparse Template Matching and Its Application to Real-time Object Tracking

YASUHARU MATSUBARA[†] and TAKESHI SHAKUNAGA[†]

Sparse template matching is proposed for realizing efficient object tracking. Key idea of our method is posed in the reduction of calculation cost by introducing sparse templates. Point selection methods are discussed with two template models, and reasonable solutions are proposed for them. The sparse template matching is stabilized by using the parallel sparse template matching. It also can collaborate with the condensation method for implementing a robust real-time tracking in compound scenes. Experimental results show effectiveness of the sparse template matching in both the parallel and the condensation methods for several image sequences.

1. はじめに

従来、画像情報に基づく動物体追跡は、対象物体の3次元モデルが既知の場合には、姿勢空間の概念を利用して効率的な追跡を実現できることが知られている⁶⁾。これに対し、3次元モデルが得られていない状況で、動物体追跡を行う場合、姿勢を厳密に取り扱うことが難しく、2次元での追跡が主体となるが、人物追跡などの実用を目指して活発に研究が行われている。その中で、Blackら¹⁾の eigen-tracking 法は代表的な研究であり、示唆に富む実験結果を報告している。この方法は固有空間、ロバスト推定、オプティカルフロー推定を要素技術として構成され、比較的変形が少ない対象に効果的である。しかし、eigen-tracking 法では、高速処理とロバスト性を同時に実現するには至っていない。eigen-tracking 法では、オプティカルフロー推定と固有空間を組み合わせることにより計算量を削減し

ているが、オプティカルフロー推定には画像サイズに比例した計算コストが必要である。これは、オプティカルフロー推定がノイズの影響を受けやすいため、精度の高い推定を行うには多くの点が必要となるからである。我々は、eigen-tracking 法のアプローチとしての限界が、オプティカルフロー推定の利用にあると考え、オプティカルフローを用いずに高速かつロバストな物体追跡を実現することを目指す。このため、本論文ではテンプレートマッチングに焦点を当てる。

ここで、テンプレートマッチングは画像処理の基本技術であり、2枚の画像間で各要素（画素）の類似度の和を計算することに基づいている。このため、テンプレートサイズや探索領域が大きい場合には、計算量が膨大となるという問題が生じる。この問題を解決するために、従来から広く研究が行われ、SSDA、ピラミッドモデル、固有空間の利用など様々な技術が提案されている^{1),3),8),9),16),18),19)}。これらの技術は各応用分野において有効な手法を提供しているが、実時間処理によるロバストな物体追跡系の実現するには至っていない。テンプレートマッチングを動物体追跡に適用するには、計算コストの削減と、様々なノイズに対するロバスト性の確保を同時に実現する必要がある。そ

[†] 岡山大学大学院自然科学研究科

Department of Computer Science, Graduate School of Natural Science and Technology, Okayama University
現在、キヤノン株式会社

Presently with Canon Inc.

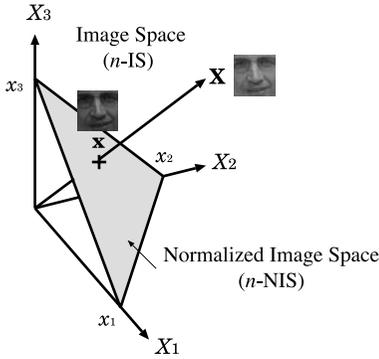


図1 正規化画像空間の概念

Fig.1 Concept of normalized image space.

ここで、本論文では、テンプレート内の点の数を少なくすることで計算コストの削減し、これをベースとして構成される追跡系に適度な冗長性を加えることにより、安定かつ高速に対象物体追跡法を実現する。

2. 疎テンプレートマッチング

2.1 正規化画像空間

照明変動にロバストな画像空間として、我々は図1に示す正規化画像空間^{(13),(15)}を提案し、顔認識などの議論を展開している^{(11),(12)}。本論文では、正規化画像空間でテンプレートマッチングを論じる。 n 画素からなる画像（以下、 n 次元画像と呼ぶ）を n 次元ベクトル $\mathbf{X} (\neq \mathbf{0})$ で表すとき、正規化画像 \mathbf{x} は次式により定義される。

$$\mathbf{x} = \mathbf{X} / (\mathbf{1}^T \mathbf{X}) \quad (1)$$

ここで、 $\mathbf{1}$ はすべての要素を1とする n 次元ベクトルである。この正規化により任意の画像 \mathbf{X} は $\mathbf{1}^T \mathbf{x} = 1$ の意味で正規化される。この正規化は、画像の総輝度を一定にするという意味で自然である。特に、入力画像 \mathbf{X} が影領域や鏡面反射領域などの非線形領域を含まない場合、正規化画像 \mathbf{x} は光源の明るさによらず一定の画像となる。一方、非線形領域が含まれる場合にも、線形領域の比率が高い場合には非線形領域の影響を受けにくいという特性を持つ。

2.2 テンプレートマッチングとパラメータ空間

テンプレートマッチングは画像処理の基本技術であり、画像の探索領域内で最もテンプレートと類似した（あるいは、相違度の小さい）位置を求める問題として定式化される。元来、並進のみを対象としていたが、回転やスケール変換などを加えたより一般的な変換へと拡張が行われている^{(1),(7)}。

本論文では、一般の変換を対象とするため、画像 \mathbf{Y} からテンプレート画像空間への変換を T で示す。し

たがって、 $T\mathbf{Y}$ は画像 \mathbf{Y} から変換 T により得られる n 次元ベクトルを表す。正規化テンプレートを \mathbf{x} 、入力画像を \mathbf{Y} とすると、変換の集合 $T = \{T\}$ を対象としたテンプレートマッチングは、次式で定義される変換 τ を求める問題となる。

$$\tau = \operatorname{argmin}_{T \in T, \alpha} \{(\alpha T\mathbf{Y} - \mathbf{x})^T (\alpha T\mathbf{Y} - \mathbf{x})\} \quad (2)$$

なお、式(2)で、各 T に対して α の最適値は解析的に次式で与えられる。

$$\alpha = \frac{\mathbf{x}^T \mathbf{x}}{\mathbf{x}^T T\mathbf{Y}} \quad (3)$$

よって、式(2)で示される最適化問題は、次式で示される T のみを対象とした最適化問題と等価になる。

$$T_{min} = \operatorname{argmin}_{T \in T} (\alpha T\mathbf{Y} - \mathbf{x})^T (\alpha T\mathbf{Y} - \mathbf{x}) \quad (4)$$

画像 \mathbf{Y} にオクルージョンや鏡面反射などのノイズが含まれる場合、 τ はその影響を大きく受ける。このような場合にもロバストなマッチングを実現するため、式(2)を次式のように一般化して考える。

$$\tau^* = \operatorname{argmin}_{T \in T, \alpha} \{\hat{\rho}(\alpha T\mathbf{Y} - \mathbf{x})\} \quad (5)$$

ここで、 $\mathbf{x} = [x_1 \cdots x_n]^T$ とすると、 $\hat{\rho}(\mathbf{x}) = \sum_{j=1}^n \rho(x_j)$ である。 $\rho(x)$ はロバスト関数であり、本論文では、次式に示すGeman-McClure関数を用いる。

$$\rho(x) = \frac{x^2}{c^2 + x^2} \quad (6)$$

ここで、 c は定数であり、適切な値を設定することで式(5)により、ノイズの影響を適切に軽減できる。正規化テンプレートでは、画素の平均値が $1/n$ になることから、この定数倍を c とするのが妥当と思われる（予備実験の結果、本論文では $c = 0.3/n$ を用いる）。

一方、Sakaueら⁽¹²⁾はロバスト射影における残差の取扱いを、従来の絶対残差ではなく相対残差にすることを提案し、その有効性を示している。これに従うと、相対残差を用いたテンプレートマッチングは次式で表される。

$$\tau^{*'} = \operatorname{argmin}_{T \in T, \alpha} \{\hat{\rho}_2(\alpha T\mathbf{Y} - \mathbf{x}; \mathbf{x})\} \quad (7)$$

ここで、 $\mathbf{x} = [x_1 \cdots x_n]^T$ 、 $\mathbf{y} = [y_1 \cdots y_n]^T$ とすると、 $\hat{\rho}_2(\mathbf{x}; \mathbf{y}) = \sum_{j=1}^n \rho(x_j/y_j)$ ($y_j \neq 0$)である。ただし、 $y_j = 0$ のとき、 $\rho(x_j/y_j) = 0$ とする。なお、関数 $\hat{\rho}_2$ に関してはテンプレート内の画素ごとに正規化が行われていることから、 $\mathbf{1}$ の定数倍を c とするのが妥当と思われる（予備実験の結果、本論文では $c = 1$ を用いる）。相対残差を使用することにより、照明変動に対してもロバストなマッチングを実現できる。

なお、式 (5) および (7) による最適化は式 (2) とは異なり、 T と α のあらゆる組合せに関する比較を必要とし、計算量の観点からは無視できない問題である。この取扱いについては、一般化して 2.3 節で述べる。

残差として相対残差、絶対残差のいずれを使用するかは、計算量とロバスト性のトレードオフによる。照明変動が小さい場合、両者はほとんど同じ結果を与える。一方、照明変動が激しい場合には相対残差を用いた方がよりロバストな結果を得る。本論文では、以下、特に示さない限り相対残差を用いる。

2.3 部分テンプレートマッチング

テンプレートマッチングではテンプレート内の全点を平等に取り扱うことにより、計算手続きの単純化と、ある程度のロバスト性の確保を行っているが、マッチングに用いる領域を限定することにより、式 (5) や式 (7) によるロバスト計算よりも正確なテンプレートマッチングを行うことができる場合がある。また、領域を限定することにより、計算量を大幅に削減できる。ここではまず、部分領域の決定方法に関係なく、一般的に部分テンプレートマッチングを取り扱う枠組みを示す。

部分領域を示す方法としては、部分射影¹¹⁾ の議論で導入した領域指定行列 P を用いる。 P は各対角要素が 1 または 0 となる $n \times n$ 対角行列であり、テンプレートの各画素が部分テンプレートに含まれる（有効）かどうかを示す。もし j 番目の対角要素 p_{jj} が 1(0) であれば j 番目の画素は有効（無効）である。領域指定行列 P が与えられると、式 (5) および式 (7) に対応する部分テンプレートマッチング問題は次式で表される。

$$\tau_P^* = \operatorname{argmin}_{T \in \mathcal{T}, \alpha} \{\hat{\rho}(P(\alpha TY - \mathbf{x}))\} \quad (8)$$

$$\tau_P^{*l} = \operatorname{argmin}_{T \in \mathcal{T}, \alpha} \{\hat{\rho}_2(P(\alpha TY - \mathbf{x}); \mathbf{x})\} \quad (9)$$

以下の議論のため、この結果を与える評価値をそれぞれ次のように陽に定義しておく。

$$\epsilon_P^* = \min_{T \in \mathcal{T}, \alpha} \{\hat{\rho}(P(\alpha TY - \mathbf{x}))\} \quad (10)$$

$$\epsilon_P^{*l} = \min_{T \in \mathcal{T}, \alpha} \{\hat{\rho}_2(P(\alpha TY - \mathbf{x}); \mathbf{x})\} \quad (11)$$

なお、式 (10) および (11) による最適化は T と α のあらゆる組合せに関する比較を必要とする。計算量の観点から実用的には、式 (13) によって α の近似値を求め、これを初期値として、残差（絶対残差あるいは相対残差）による外れ値（ノイズ領域）検出、外れ値の推定値への置換え、式 (13) による α の更新を繰り返すロバスト推定¹²⁾ により解を求める近似解法をとる。ここで、残差を相対残差とする場合、ノイズ領

域は次式で定義される $d_j(P(\alpha TY - \mathbf{x}); \mathbf{x})$ により検出する。

$$d_j(P(\alpha TY - \mathbf{x}); \mathbf{x}) = \begin{cases} 1 & \text{if } p_{jj} = 1 \\ & \text{and } \left| \frac{\{\alpha TY\}_{j-\{\mathbf{x}\}_j}}{\{\mathbf{x}\}_j} - \hat{d} \right| \geq d_\theta \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

式 (12) において、 d_θ は閾値であり、 \hat{d} は $\frac{\{\alpha TY\}_{j-\{\mathbf{x}\}_j}}{\{\mathbf{x}\}_j}$ の中央値である。 $d_j(P(\alpha TY - \mathbf{x}); \mathbf{x})$ は TY の第 j 要素に一定量以上のノイズが含まれているかどうかを示す。検出されたノイズ領域については、画素値を推定値 \mathbf{x} に置き換える。これによりノイズの影響を早期に低減し、計算量を削減する。また、繰返し処理の各段階でノイズ領域が閾値（実験では 30% とした）以上になった場合、ロバスト推定の信頼性は著しく低下するため、ロバスト推定不能と判定し、繰返し処理を打ち切る（結果として求められる評価値は大きくなる）。特に、本論文のように疎で少数の点による推定を行う場合は、ノイズの影響が大きいと考えられるため、この判定は厳しく行う。

なお、テンプレートを正規化固有空間に拡張した場合、式 (13) は同次固有空間への部分射影に一般化され、 P が既知の場合には、線形射影に帰着できる¹¹⁾。

$$\alpha = \frac{\mathbf{x}^T P \mathbf{x}}{\mathbf{x}^T P T \mathbf{Y}} \quad (13)$$

2.4 疎テンプレートマッチングの定義

領域指定行列 P が疎な点集合を表すとき、部分テンプレートマッチングを、特に、疎テンプレートマッチングと呼ぶ。疎テンプレートによる物体追跡については、坂上ら¹⁰⁾ がその可能性を示唆しているが、具体的な疎テンプレート作成法については論じていない。以下では、これについて議論する。

3. 疎テンプレートの作成法

3.1 疎テンプレートが満たすべき性質

疎テンプレートを用いて対象物体を追跡する場合、疎テンプレートが有効に動作するためには、テンプレートと背景の識別、および、テンプレート内での識別について、次の性質が満たされる必要がある。

(1) テンプレートと背景の識別

疎テンプレートは背景と識別可能でなければならぬ。すなわち、対象物体から生成される疎テンプレートと背景部分の相違度（ $\hat{\rho}_2$ あるいは $\hat{\rho}$ ）は十分に大きいと仮定する。なお、この仮定は対象物体の像がいくつも背景に含まれるような特殊な状況や、たとえば、

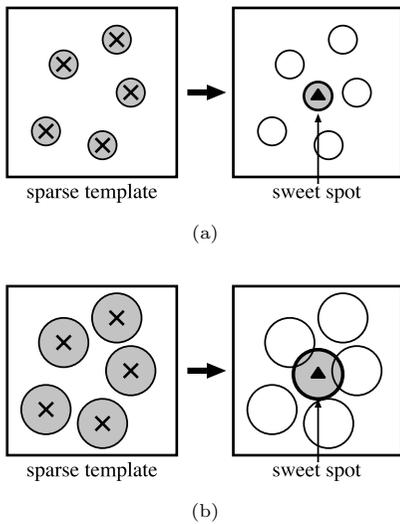


図 2 疎テンプレートとテンプレート内部との関係

Fig. 2 Relationship between the sparse template and the full template.

「曇天の雪原中に転がる白いボールを追跡する」というような、背景と対象物体を画像情報によって識別不可能（あるいは困難）な場合を除外するためのものである。また、これほど極端でなくても、テンプレートが背景と類似している場合、テンプレートマッチングは難しくなるため、適切な特徴を事前に選択することが必要になる場合がある。特に、背景が移動する追跡では、場合によっては実時間での特徴選択が必要である²⁾が、本論文では、これらの特徴選択の問題は取り扱わない。

(2) テンプレート内部での識別

疎テンプレートはまた、テンプレート内部の近傍領域においても、十分に識別可能でなければならない。たとえば、図 2 (a) に示すようにテンプレート内の輝度が一定である小領域の中心点の集合で疎テンプレートを構成する場合と、(b) に示すように、輝度一定である大領域から同様に疎テンプレートを構成する場合を考えよう。このとき、微小な変位に対して、(a) では $\hat{\rho}$ あるいは $\hat{\rho}_2$ で定義される相違度が大きくなることから精度の高い追跡が可能である。すなわち、スイートスポット（＝相違度が正解位置と同程度である領域）が小さいことになる。一方、(b) では、微小な変位に対しては相違度が変化しないから、スイートスポットは (a) よりも大きい。したがって、疎テンプレート (b) によっては精度の高い追跡が難しくなる。スイートスポットの大きな疎テンプレートは、背景とテンプレートの識別が容易な場合、物体検出には有効と考えられる。一方、本論文で対象としている物体追跡には、ス

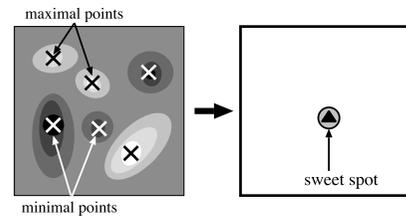


図 3 連続領域モデルに基づく点選択（極大・極小点）

Fig. 3 Point sampling criterion for continuous region model.

イートスポットがある程度小さな疎テンプレートを作成する必要がある。

なお、適切なスイートスポットサイズは、物体追跡を制御する系の特性に依存すると考えられる。condensation 法等の確率的探索と組み合わせる場合、探索範囲をどの程度絞り込めるかによって、適切なスイートスポットサイズが決まる。探索範囲を広くとる必要がある場合には、ある程度の大きさを持つスイートスポットを用いることで、安定性を確保することも考えられる。追跡精度と安定性のトレードオフは、今後の重要な検討課題であると思われる。一方、本論文では、できるだけ追跡精度を保つことを優先して、condensation 法への適用にあたっては、スイートスポットを小さくしつつ、安定性を確保することを目指す。

以上の考察を基にして、以下では、2 種類の領域モデルについて、物体追跡に有効な疎テンプレートの作成法を論じる。

3.2 連続領域モデル

第 1 の領域モデルは、図 3 に示す連続領域モデルである。画像が画像座標の連続関数で表される場合、輝度が極大・極小となる点のみから疎テンプレートを構成する。このとき、並進のみを仮定すると、正解位置において類似度は最大（相違度は最小）となり、微小なズレに対しては移動方向に関係なく類似度が低下することが分かる。すなわち、極大・極小点の集合によって、並進移動に対してロバストな疎テンプレートマッチングを実現できる。この関係は、回転やスケール変換に対しても一般化できるので、一般の微小変位に対して極大・極小点の集合で与えられる疎テンプレートが有効であることが分かる。また、この関係はすべての点が極大または極小点であるときにはつねに成り立つことから、極大・極小点の集合が連続領域モデルに対する解を与えることが分かる。

なお、実用上は点の間隔にも留意する必要がある。これは、選択される点がある部分に集中すると、部分遮蔽などの影響を受けやすくなるからである。たとえ

ば、5章で述べる実験では、極大・極小点選択を次のように行っている。まず、テンプレート内で8-近傍比較によって極大(極小)点をすべて求めた後、画素値で降順(昇順)に列べ、上位から順に点を、極大/極小から交互に選択する。ここで、すでに選択された点と一定の距離(実験では6画素)以内にある場合、その点は選択しない。これにより、テンプレート内で適当な点間隔を持つ極大・極小点集合を自動的に生成できる。

3.3 不連続領域モデル

第2の領域モデルは、複数の連続領域の間に境界が存在する場合をモデル化したものである。図4のように、2領域が境界によって分離されている場合、境界に直交し、かつ、境界から両側に等距離の2点を選択し、この2点組を境界ダイポールと呼ぶ。ここで、1組の境界ダイポールはダイポールの間に境界位置を制約する。図4(a)のように境界ダイポールを広い間隔で選択する場合、2次元テンプレートは大きなスイートスポットを持つ。一方、図4(b)のように境界ダイポールを狭く選択すると、スイートスポットは小さくなる。

複数の境界ダイポールをテンプレートから選択することにより、追跡に有効な疎テンプレートを作成できる。安定したテンプレートマッチングを実現するためには、異なる方向のダイポールを適切に組み合わせる必要がある。これにより、境界ダイポールの方向が偏った場合に生じるエッジ方向の不定性(aperture problemで生じる不定性と等価)を回避することができる。また、局所的な隠れや変形に対応するためには、境界ダイポールを適当に分散させることが望ましい。これらの各点を満たすようにするため、5章で述べる実験では、境界ダイポール選択を次のように行っている。まず、入力画像にラプラシアンガウシアンオペレータをかけ、ゼロ交差を求めることにより境界を検出する。ここで、各境界要素において、境界と直交する方向に両側2画素目の画素からなる画素ペアを求める。このようにして求めた画素ペアを、境界の方向別に2画素の輝度差で順序付けし、輝度差の大きなものから順に、境界ダイポールとして選択した。ここで、境界の方向は縦・横・斜め45度の4方向に分け、各方向から同じ数の境界ダイポールが選択されるようにしたが、該当する境界要素がない場合やペア間に輝度差が大きなものがない場合には、他の方向のもので代用した。また、点ペア相互間の距離についても、極大・極小点の選択と同様の方法で相互間の距離を用いた選択を行っている。

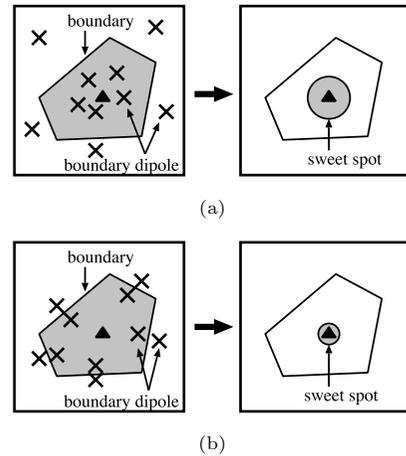


図4 不連続領域モデルに基づく点選択(境界ダイポール)
Fig. 4 Point sampling criterion for discontinuous region model.

3.4 疎テンプレート作成

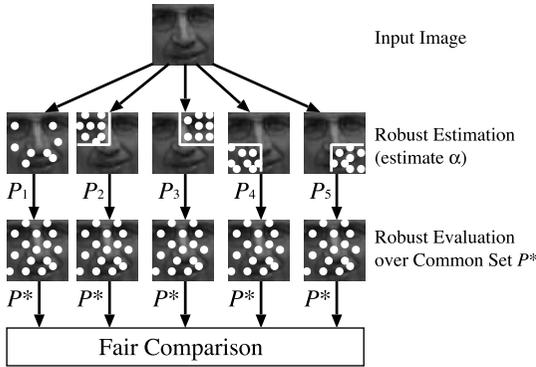
一般のテンプレートは、連続領域モデルと不連続領域モデルの組合せと考えることができる。したがって、有効な疎テンプレートの作成には、前節までに述べた2種類の方法を組み合わせるのが妥当である。この方法により、任意のテンプレートに対して、3.1節で述べた条件を満たす疎テンプレートを作成することができる。なお、上述の2種類の疎テンプレート作成法は、どちらもテンプレート内部を局所解析することによって実現できる。したがって、本論文で述べた点選択法は、文献(14)、(17)において述べられているような大局的な方法とは根本的に異なる。

4. 追跡系の安定化に関する検討

4.1 並列処理による追跡の安定化

疎テンプレートは高速計算を目的とするため、できる限り少数の画素から構成することが望ましい。一方、ノイズに対するロバスト性を確保するためには、系にある程度の冗長性を持たせる必要がある。両者を同時に実現するためには、適度な数の画素を含むように疎テンプレートを構成する必要がある。

しかし、単一の疎テンプレートではロバスト性に限界がある(break-down point problem: ノイズ量が臨界点以上の場合、確率的に破綻が生じることを回避できない)ため、複数の疎テンプレートを並列的に利用することが次に考えられる。このとき、異なる疎テンプレート間で公平な比較を行う必要が生じる。すなわち、異なる点集合 P_i と P_j にロバスト疎テンプレートマッチングを適用すると、式(11)により、 $\epsilon_{P_i}^*$

図5 並列処理における共通の点集合 P^* による評価Fig. 5 Parallel processes using evaluations over a common set P^* .

と $\epsilon_{P_j}^*$ の2つの評価値が得られる．ここで、 P_i と P_j に対応する点集合は共通な要素をほとんど含まない．また、両者に含まれる点の数が小さいため、シーンの変動などに起因するノイズの影響は一方にしか及ばない．したがって、2つの評価値の単純な大小比較により類似度を評価するのは不相当である．また、点の数が小さいために、個々の疎テンプレートに加わるノイズの影響を統計的に処理することも難しい．すなわち、点集合 P_i と P_j が異なる限り、式 (11) のままでは公平な比較を実現できないと考えられる．そこで、以下では、共通の点集合 P^* を導入することにより、公平な評価を実現する．

この問題は、本質的には、変換 T を固定 ($\mathcal{T} = \{T\}$) した場合の比較の問題に帰着される．すなわち、図5に示すように個々の点集合 P_i において式 (9) で示した $\tau_{P_i}^* = (T, \alpha_i)$ を求めたうえで、共通の点集合 P^* において評価値 $\epsilon_{P^*}^*$ を求めることにする．これにより、疎テンプレート間の不公平を解決し、並列処理による追跡の安定化を実現できる．具体的には、評価値は次式で表される．

$$\begin{aligned} \tilde{\epsilon}_{P_i}^* &= \hat{\rho}_2(P^*(\alpha_i T Y - \mathbf{x}); \mathbf{x}) \quad (14) \\ \text{where } \alpha_i &= \underset{\alpha}{\operatorname{argmin}} \{ \hat{\rho}_2(P_i(\alpha T Y - \mathbf{x}); \mathbf{x}) \} \end{aligned}$$

なお、2.3節で述べたように、式 (9) の計算は、ロバスト推定によって近似解を求める．すなわち、 $P = P_i$ とした場合に式 (13) によって与えられる α を α_i の初期値として、相対残差によるノイズ領域の抽出、ノイズ領域の推定値への置換え、式 (13) による α_i の推定を繰り返すロバスト推定を用いる．

4.2 疎テンプレートマッチングと condensation 法の統合

よりロバストな追跡を行うためには、確率論的な取

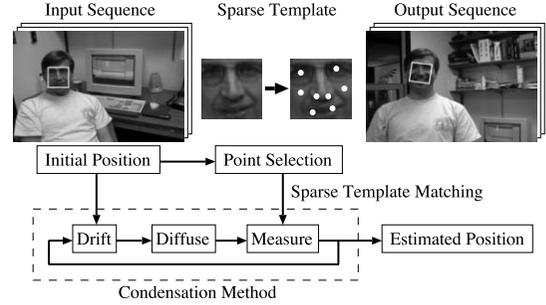


図6 見え方に基づくコンデンセーション追跡システムの概要

Fig. 6 Overview of appearance-based condensation tracker.

扱いが有効と思われる．本論文では、その一例として、condensation 法⁴⁾の利用を考える．図6に示すように、condensation 法と疎テンプレートマッチング法を統合することにより、アルゴリズムを構成する．ここで、従来の condensation 法は、drift, diffuse, measure の3段階で構成され、多数の粒子 (particle) をパラメータ空間内で伝播させることを基本としている．drift では、前時刻の粒子の中から尤度の高いものをランダムに選択し、diffuse では、drift で選択された粒子に対しランダムな値を付加することで、その時刻における粒子を生成する．measure では、diffuse で得られた粒子に対する尤度を画像から求める．ここで、尤度として疎テンプレートマッチングで得られる評価値 (式 (14)) の逆数を用いることにより、condensation 法と疎テンプレートマッチングを統合する．また、前節で述べた並列処理の観点を採り入れ、 P_i の選択をパラメータ空間に加えることで系を構成する．

以下、本論文では、姿勢パラメータとして並進 $2(t_x, t_y)$ 、回転 $3(\psi, \theta, \phi)$ 、スケール $1(c)$ の計6パラメータを考える．ここで、 ψ は画像の x 軸方向の回転、 θ は y 軸方向の回転、 ϕ は z 軸方向 (画像平面に対する回転) を表す．このとき、画像上の点 $\mathbf{m} = [a \ b]^T$ から $\mathbf{m}' = [a' \ b']^T$ への変換は次式で表される．

$$\mathbf{m}' = \mathbf{cR}_{2 \times 2} \mathbf{m} + \mathbf{t} \quad (15)$$

$$\mathbf{R} = \mathbf{R}(x, \psi) \mathbf{R}(y, \theta) \mathbf{R}(z, \phi) \quad (16)$$

$$\mathbf{R}(x, \psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & -\sin \psi \\ 0 & \sin \psi & \cos \psi \end{bmatrix} \quad (17)$$

$$\mathbf{R}(y, \theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \quad (18)$$

$$\mathbf{R}(z, \phi) = \begin{bmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (19)$$

$$\mathbf{t} = \begin{bmatrix} t_x & t_y \end{bmatrix}^T \quad (20)$$

ここで、 $\mathbf{R}_{2 \times 2}$ は回転行列 \mathbf{R} における左上 2×2 の部分行列である。また、diffuse では各パラメータに対して、ガウシアンノイズを独立に発生させて用いる。ここで、ガウシアンノイズの標準偏差は実験的に与えた。

従来の condensation 法⁴⁾ では、時刻 t における粒子集合 $\{s_t^{(1)}, \dots, s_t^{(L)}\}$ は時刻 $t-1$ における姿勢の近傍に生成される。本論文では、疎テンプレートの選択をパラメータとして追加する。したがって、時刻 t における l 個目の粒子 $s_t^{(l)}$ は以下のように表される。

$$s_t^{(l)} = \begin{bmatrix} t_x & t_y & \psi & \theta & \phi & c & i \end{bmatrix}^T \quad (21)$$

ここで、 i は疎テンプレート P_i を選択することを意味する。4.1 節で示した並列処理をすべての粒子において適用することも考えられるが、condensation 法には系の冗長性が含まれているため、計算コストの観点から各粒子における並列処理は行わない。すなわち、各粒子においては、5 種類の疎テンプレート (P_i) の中から 1 つをランダムに選択し、点集合 P^* で評価する方法 (式 (14)) を用いる。これにより、疎テンプレートの違いによる不公平は解消できる。

具体的には、時刻 t における粒子 $s_t^{(l)}$ で選択された疎テンプレート P_i に対して、式 (14) により評価値が計算される。この評価値を $\tilde{\epsilon}_{P_i}^{*(l)}$ で表すと、粒子 $s_t^{(l)}$ の確率密度 $\pi_t^{(l)}$ は次式で表される。

$$\pi_t^{(l)} = \frac{1/\tilde{\epsilon}_{P_i}^{*(l)}}{\sum_{j=1}^L 1/\tilde{\epsilon}_{P_i}^{*(j)}} \quad (22)$$

その他のモジュールは従来の condensation 法に従って構成する。ただし、実験では、粒子数 L を 1000 とし、そのうち drift の段階で選択される候補を上位 10 個のみとした。

5. 実験

5.1 極大・極小点による段階的ランダム探索法

本論文で提案した疎テンプレートマッチングの有効性を示すため、動画像を用いて追跡実験を行った。1 つ目の実験では、領域境界が明らかでない対象物体の追跡を行った。この実験では、連続領域モデルに基づいて疎テンプレートを作成した。図 7 に構成された 5 つの疎テンプレート ($P_1 \sim P_5$) および評価用疎テンプレート P^* を示す。 $P_1 \sim P_5$ は 46×54 pixel 中の 8 点、

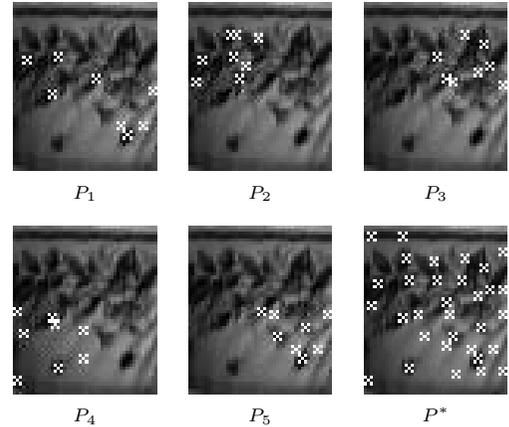


図 7 疎テンプレート (極大・極小点)

Fig. 7 Sparse templates made up by the first criterion.

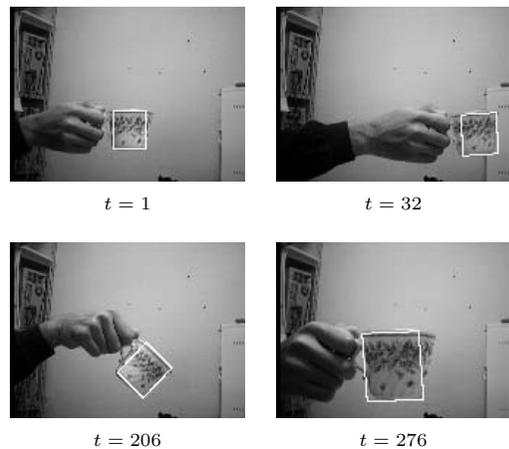


図 8 段階的ランダム探索法による追跡結果 (極大・極小点)

Fig. 8 Tracking results of the sequential random search using sparse templates as shown in Fig. 7.

P^* は 32 点の極大・極小点から構成されている。ここでは、4.1 節で示した並列処理により、ランダムに生成された粒子の中から最も評価の高いものを各時刻に選択する段階的ランダム探索法による (確定的な) 追跡を行った。図 8 に追跡結果を示す。この結果より、極大・極小点からなる疎テンプレートの並列処理によりうまく追跡が行えることが分かる。

5.2 境界ダイポールによる段階的ランダム探索法

2 つ目の実験では不連続領域モデルを用い、前節の実験と同様に段階的ランダム探索法による追跡を行った。図 9 に 5 つの疎テンプレート ($P_1 \sim P_5$) および評価用疎テンプレート P^* を示す。 $P_1 \sim P_5$ は 48×56 pixel 中の 8 点、 P^* は 32 点の境界ダイポールから構成されている。実験結果を図 10 に示す。この結果より、境界ダイポールで構成された疎テンプレートによりう

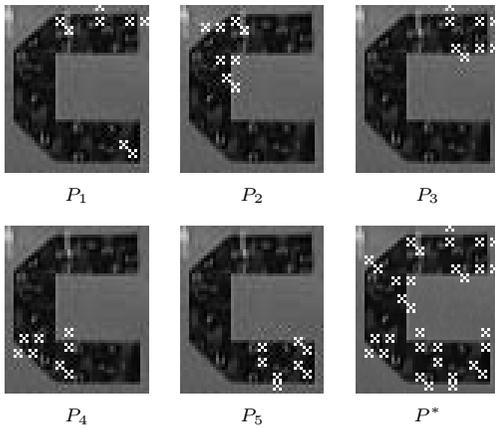
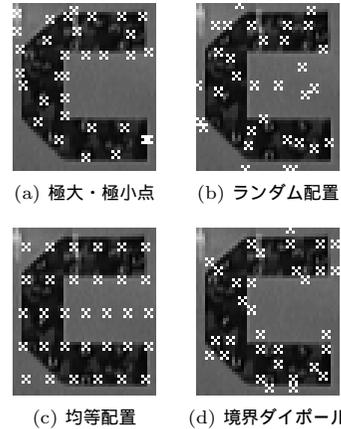
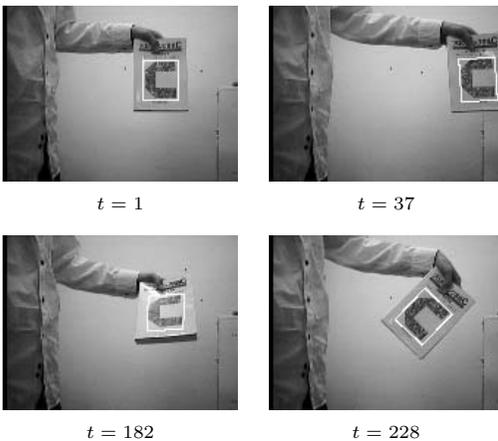


図 9 疎テンプレート (境界ダイポール)

Fig. 9 Sparse templates made up by the second criterion.

図 11 4 種類の疎テンプレート (P^*)Fig. 11 Four types of sparse templates (P^*).図 10 段階的ランダム探索法による追跡結果 (境界ダイポール)
Fig. 10 Tracking results of the sequential random search using boundary dipoles.

まく追跡できることを確認できた。

次に、疎テンプレートの作成方法を変化させて比較実験を行った。この実験では、前述の境界ダイポールと極大・極小点に加え、ランダム配置、均等配置（それぞれの疎テンプレート P^* を図 11 に示す。 $P_1 \sim P_5$ については省略）を比較した。ここで、各疎テンプレートにおけるサンプル点の数は境界ダイポールの場合と同じにした。定量的な比較を行うために、各フレームにおいてテンプレートの正解位置（シミュレーション実験ではないため正解は本来定義できないので、全画素からなるテンプレートを用いて十分に時間をかけてロバストマッチングを行った結果抽出される位置を正解位置とした）に対する平均誤差（4 隅の位置に関する推定誤差の平均値）を算出した。結果を図 12 に示す。また、 $t = 248$ における追跡状況を図 13 に

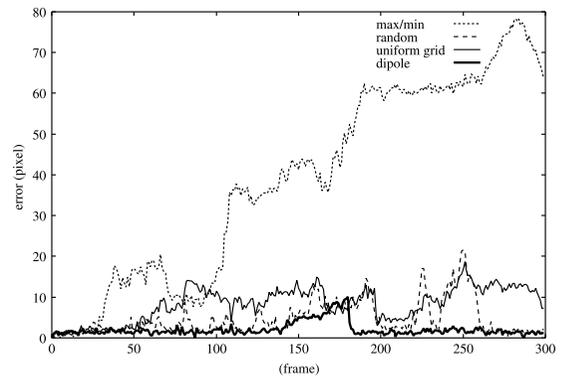


図 12 位置推定誤差の比較

Fig. 12 Comparison of errors in the estimated locations.

示す。この結果より、境界ダイポールを用いた場合が最も安定していることが分かる。また、全フレームの平均誤差は境界ダイポールの場合に 2.1 pixel であるのに対し、極大・極小点では 38.7 pixel、ランダム配置では 4.7 pixel、均等配置では 8.3 pixel であり、いずれも境界ダイポールよりも誤差が大きいたことが分かる。したがって、境界を含む物体に対しては境界ダイポールが有効であることが確認できた。

5.3 condensation 法による追跡実験

Jepson ら⁵⁾ で使用されている動画像に対し、condensation 法による追跡実験を行った。この動画像には顔の 3 次元回転やスケール変動、形状の変形や手によるオクルージョンなどが含まれている。この実験では、極大・極小点選択と境界ダイポール選択を組み合わせる疎テンプレートを作成した。図 14 に 5 つの疎テンプレート ($P_1 \sim P_5$) および評価用疎テンプレート P^* を示す。 $P_1 \sim P_5$ は 45×45 pixel 中の 8 点、 P^* は 32 点から構成されている。これらの疎テンプレートを

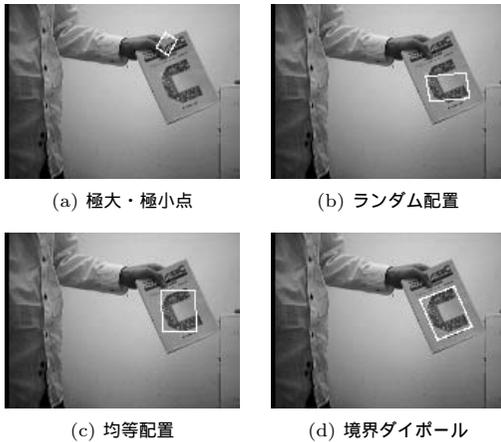


図 13 追跡状況の比較 ($t = 248$)

Fig. 13 Tracking results of four types of sparse templates ($t = 248$).

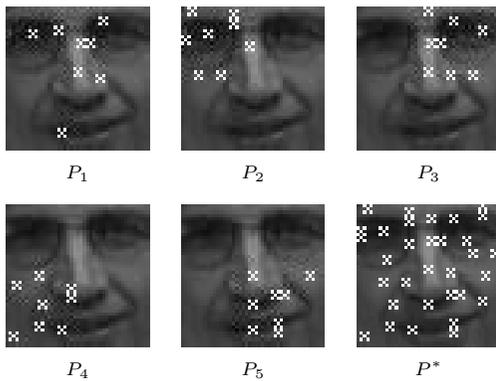


図 14 疎テンプレート (極大・極小点と境界ダイポールの併用)
Fig. 14 Sparse templates made up by a mixture of the first and second criterions.

用いて追跡を行った結果を 図 15, および, 図 16 に示す. また, 全フレームにおいて顔姿勢を示す回転パラメータ (ロール, ピッチ, ヨー) とスケールの変動を 図 17 に示す. なお, 図 17 中の縦線およびグラフの下に示した記号は, 図 15 の各フレームに付けた記号と対応する. また, 図 17 の上のグラフから, ロールの変動幅が $-20 \sim 15$ 度の範囲であるのに対し, ピッチ, ヨーの変動幅は $-40 \sim +30$ 度の範囲に及ぶことが分かる. また, 同図下のグラフから, スケールに関しては, 0.6 から 1.5 の範囲で変動していることが分かる. さらに, 図 16 は $204 \sim 214$ フレームにおけるオクルージョン (右手が顔の前を通過することによって生じている) 時の追跡結果を示しているが, コンデーンション法の枠組みで姿勢候補を保持できるため, 一時的な隠れに対しても正しく追跡を行えることを示している. 以上の結果より, 姿勢・スケール・隠れな

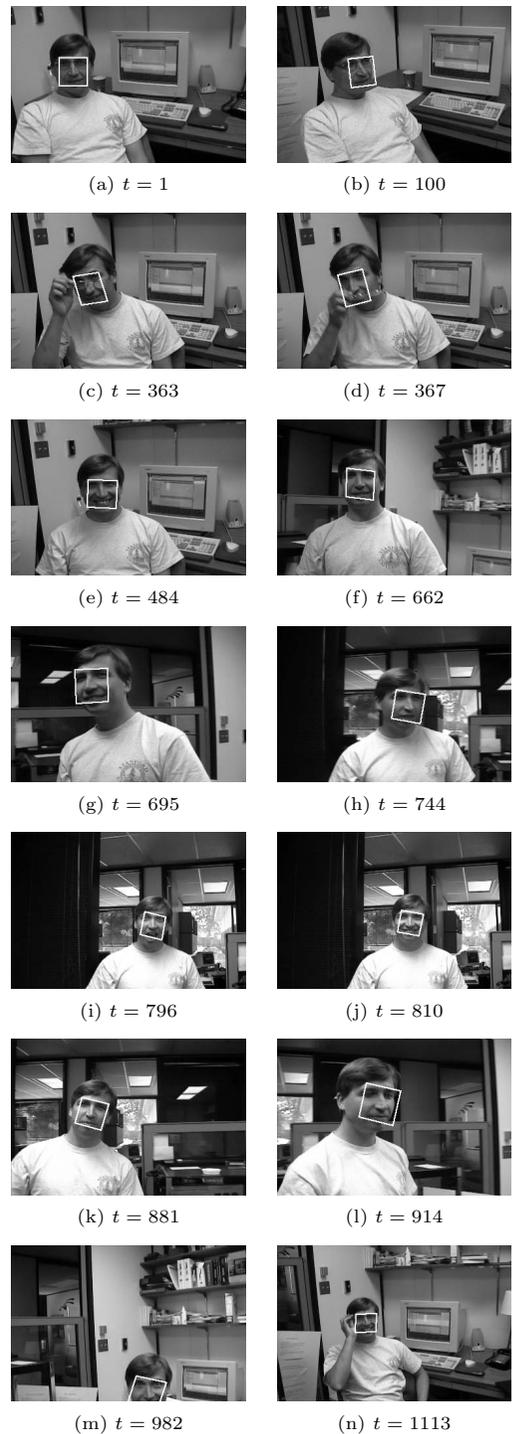


図 15 condensation 法による追跡結果 (極大・極小点と境界ダイポールの併用)

Fig. 15 Tracking results of the condensation method using mixture of maximal/minimal points and boundary dipoles.

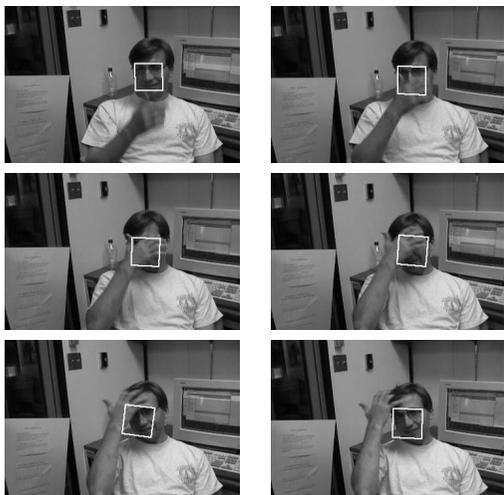


図 16 condensation 法による追跡 (204~214 フレーム)
Fig.16 Tracking results from 204th to 214th frames .

どの変動を含む場合にも、顔をうまく追跡できることを確認した。また、各時刻の各粒子に対するロバスト推定における推定回数の分布を表 1 に示す。ただし、ノイズと判定された領域が 30%以上の場合は評価の対象外として棄却した。この結果より、正解位置に近く、テンプレートと画像の類似度が高い場合には、1 回の推定（この場合、式 (13) で与えられる α を初期値として使用して、ノイズ領域検出/修正した結果から $\hat{\rho}_2$ を計算したものが結果として得られる）でロバスト推定が終了していることが分かる。一方、正解位置から離れている場合には、ノイズ領域が大きくなるため、繰り返し推定が行われるが、実験では繰返し回数は最大で 3 であった。また、各段階でノイズ領域が大きいために棄却される場合もある。なお、今回の実験では、推定回数の平均値は 1.25 であった。

この実験においても、疎テンプレートの作成方法を変化させて、追跡の安定性を比較した。図 18 に比較に用いた 4 種類の疎テンプレートを示す。この各々について、各 1,000 回の追跡実験（ランダム変数により、各粒子の発生を制御した）を行った結果、得られた追跡成功率および推定誤差の平均と標準偏差を表 2 に示す。ここで、追跡成功率とは、全 1,145 フレームにおける追跡結果の（正解位置に対する）誤差の平均が 5.0 pixel 以内である確率を示す。この結果より、極大・極小点と境界ダイポールの 2 種類の選択基準の組合せによって作成したテンプレートが最も安定していることが確認できた。この結果は、人物の顔については、2 種類の選択基準をうまく組み合わせることにより、安定した追跡が可能であることを示している。

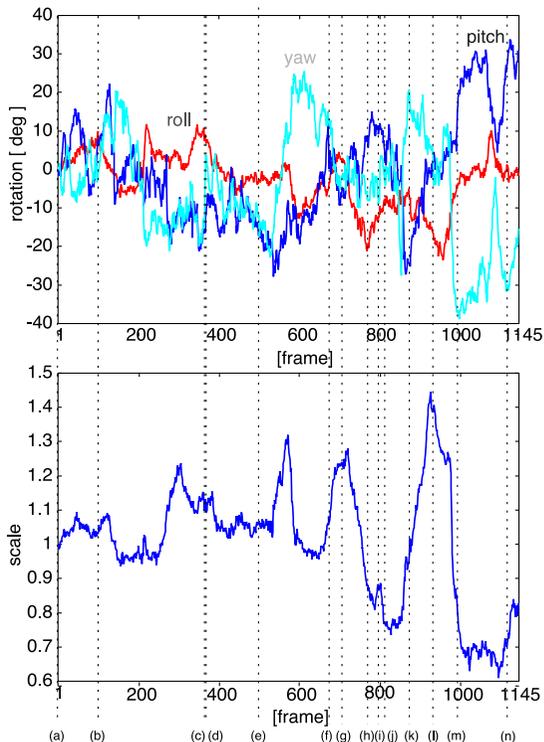


図 17 追跡時の回転パラメータとスケールパラメータの変動
Fig.17 Changes of estimated rotation/scale parameters during the tracking.

表 1 ロバスト推定における推定回数の分布

Table 1 Distribution of required estimation iterations in robust estimation.

推定回数	1	2	3	4
頻度	852,589	291,384	27	0
棄却数	226,953	93	0	0

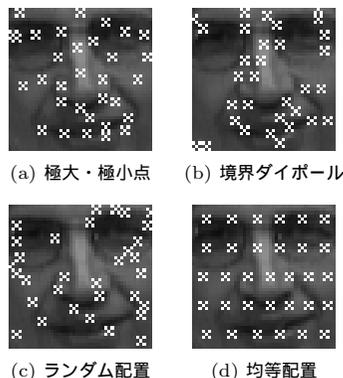


図 18 4 種類の疎テンプレート (P^*)

Fig.18 Four types of sparse templates (P^*).

表 2 各テンプレートによる追跡成功率および推定誤差の比較
Table 2 Comparison of successful tracking rates and estimation errors among five types of sparse templates.

テンプレート	追跡成功率 [%]	推定誤差 [pixel]	
		平均	標準偏差
極大・極小点	10.3	9.7	7.5
境界ダイポール	0.1	20.9	14.4
ランダム配置	3.4	22.5	16.2
均等配置	0.0	37.0	12.8
2種類の併用	98.0	3.2	2.1

表 3 残差の種類による追跡成功率および推定誤差の比較
Table 3 Comparison of successful tracking rates and estimation errors by two types of residuals.

残差	追跡成功率 [%]	推定誤差 [pixel]	
		平均	標準偏差
絶対残差	81.8	6.3	9.2
相対残差	98.0	3.2	2.1

表 4 追跡アルゴリズムの種類による追跡成功率および推定誤差の比較
Table 4 Comparison of successful tracking rates and estimation errors by two types of tracking algorithms.

追跡アルゴリズム	追跡成功率 [%]	推定誤差 [pixel]	
		平均	標準偏差
段階的ランダム探索法	44.1	12.3	14.2
condensation 法	98.0	3.2	2.1

また、同じ実験において、評価値として絶対残差と相対残差を用いた場合の比較を行った結果を表 3 に示す。この結果より、相対残差の方が安定した追跡を実現できることが分かる。これは、絶対残差による評価では暗い領域の影響を受けやすいためであると考えられる。結果として、相対残差による評価の有効性を確認できた。なお、Geman-McClure 関数のパラメータ c を変動させ、追跡成功率の変化を調べたところ、相対残差では $0.5 \leq c \leq 1.1$ で追跡成功率が 92.6~98.2%を示し安定していた。一方、絶対残差では $0.26 \leq cn \leq 0.36$ で追跡成功率は 73.1~81.8%となり、相対残差と比較して不安定であり、また、閾値 c の設定が難しいことが分かる。

次に、各時刻で多数の姿勢候補を保持する condensation 法と姿勢を 1 つに絞る段階的ランダム探索法との比較実験結果を表 4 に示す。この結果より、確率的な取扱いを行っている condensation 法の優位性を確認できた。

なお、この実験では Pentium4 2.8 GHz 搭載の PC を使用しており、1 フレームあたりの処理時間は condensation 法で約 15 msec (テンプレートの全点を使用した場合の処理時間約 2.8 sec の 175 分の 1) であっ

た。これは、ビデオレートで実時間処理によりロバストな追跡を実現できていることを示している。この実験では、初期フレームから自動的に生成されるテンプレート以外には、事前知識をまったく利用していない。このような条件においても、condensation 法と疎テンプレートマッチングの組合せだけで、モデルの更新なしに、ロバストでかつ高速な追跡を実現できている点は注目に値すると思われる。

6. まとめ

本論文では、高速なテンプレートマッチングを実現する方法として疎テンプレートマッチングを提案した。疎テンプレートは 2 種類のモデルを組み合わせることにより作成した。また、疎テンプレートマッチングを用いて安定な物体追跡を実現する方法をいくつか示した。照明変動や姿勢変動、隠れに強い方法として確率的枠組みの利用が必須と考えられる。本論文では、この一例として condensation 法との融合を試み、実験によりその有効性を確認した。疎テンプレートマッチング法は、テンプレートの作成を除いて事前知識を必要としないが、固有空間の利用¹¹⁾ など適切な事前知識を用いることにより、より高度の追跡を実現できると考えられる。

なお、疎テンプレートの設計法として、本論文では 2 つのモデルに基づく方法を提案したが、これらはスイートスポットをできるだけ小さくすることを目指したものである。一方、追跡目的および適用環境によっては、必ずしもスイートスポットを小さくせずに追跡の安定化を目指すアプローチもありうる。このようなアプローチで疎テンプレートマッチングを使用する場合には、疎テンプレートの設計法を別途検討する必要がある。また、本論文で一例を示した condensation 法等の確率的探索との組合せを考える場合、同時に探索可能な(姿勢)範囲との関係でスイートスポットの大きさを最適化する必要があると思われる。

本論文では、追跡対象の 2 次元画像に基づく追跡を議論し、この範囲においては提案手法が有効であることを示した。一方、3 次元物体モデルに基づく追跡においては、1 台のカメラしか用いない場合には、ピッチ・ヨーに関する推定精度が他の姿勢パラメータ(並進・スケール・ロールなど)に比べて著しく劣ることが知られている。疎テンプレートマッチングを 3 次元物体追跡に適用する際にも、この問題の対処法を検討する必要がある。

謝辞 本研究の一部は科学研究費補助金基盤研究(B)(2)(課題番号 15300062)によった。実験に協力

いただいた野口清志氏, 日頃討論をしていただく坂上文彦氏, 阿部圭佑氏に感謝する.

参 考 文 献

- 1) Black, M.J. and Jepson, A.D.: Eigentracking: Robust Matching and Tracking of Articulated Objects Using a View-based Representation, *International Journal of Computer Vision*, Vol.26, No.1, pp.63–84 (1998).
- 2) Collins, R.T. and Liu, Y.: On-line Selection of Discriminative Tracking Features, *Proc. International Conference on Computer Vision*, pp.346–352 (2003).
- 3) Hager, G.D. and Belhumeur, P.N.: Efficient Region Tracking with Parametric Models of Geometry and Illumination, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.20, No.10, pp.1025–1039 (1998).
- 4) Isard, M. and Blake, A.: Condensation-Conditional Density Propagation for Visual Tracking, *International Journal of Computer Vision*, Vol.29, No.1, pp.5–28 (1998).
- 5) Jepson, A.D., Fleet, D.J. and El-Maraghi, T.F.: Robust Online Appearance Models for Visual Tracking, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.25, No.10, pp.1296–1311 (2003).
- 6) Lowe, D.G.: Robust Model-based Motion Tracking through the Integration of Search and Estimation, *International Journal of Computer Vision*, Vol.8, No.2, pp.113–122 (1992).
- 7) Moghaddam, B. and Pentland, A.: Probabilistic Visual Learning for Object Representation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.19, No.7, pp.696–710 (1997).
- 8) Kaneko, S., Murase, I. and Igarashi, S.: Robust Image Registration by Increment Sign Correlation, *Pattern Recognition*, Vol.35, No.10, pp.2223–2234 (2002).
- 9) 奥田晴久, 橋本 学, 鷲見和彦, 佐々木和則: 部分テンプレート組み合わせ最適化に基づくロバスト画像照合, *電気学会論文誌 C*, Vol.124, No.3, pp.629–636 (2004).
- 10) 坂上文彦, 尺長 健: 正規化固有空間への部分射影の幾つかの応用について, *情報処理学会研究報告 CVIM-2003-136*, pp.163–170 (2003).
- 11) 坂上文彦, 尺長 健: 正規化固有空間への最適部分射影とその応用, *情報処理学会論文誌コンピュータビジョンとイメージメディア*, Vol.44, No.SIG 17 (CVIM 8), pp.100–108 (2003).
- 12) Sakaue, F. and Shakunaga, T.: Robust Projection onto Normalized Eigenspace Using Relative Residual Analysis and Optimal Partial Projection, *IEICE Trans. Inf. & Syst.*, Vol.E87-D, No.1, pp.34–41 (2004).
- 13) Shakunaga, T. and Shigenari, K.: Decomposed Eigenface for Face Recognition under Various Lighting Conditions, *Proc. Computer Vision and Pattern Recognition*, Vol.1, pp.864–871 (2001).
- 14) Shi, J. and Tomasi, C.: Good Features to Track, *Proc. Computer Vision and Pattern Recognition*, pp.593–600 (1994).
- 15) 重成一真, 坂上文彦, 尺長 健: 固有顔の直交分解と仮想化による照明変動に影響されない顔画像認識, *電子情報通信学会論文誌 (D-II)*, Vol.J86-D-II, No.7, pp.996–1004 (2003).
- 16) Tanimoto, S.L.: Template Matching in Pyramids, *Computer Graphics and Image Processing*, Vol.16, pp.356–369 (1981).
- 17) Torr, P.H.S. and Davidson, C.: IMPSAC: Synthesis of Importance Sampling and Random Sample Consensus, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.25, No.3, pp.354–364 (2003).
- 18) Turk, M. and Pentland, A.: Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71–86 (1991).
- 19) Yoshimura, Y. and Kanade, T.: Fast Template Matching Based on the Normalized Correlation by Using Multiresolution Eigenimages, *Proc. IROS'94*, pp.2086–2093 (1994).

(平成 16 年 8 月 20 日受付)

(平成 17 年 3 月 4 日採録)

(担当編集委員 鷲見 和彦)



松原 康晴

平成 14 年岡山大学工学部情報工学科卒業. 16 年同大学大学院自然科学研究科博士前期課程修了後, キヤノン株式会社に勤務. 在学中, コンピュータビジョンの研究に従事.



尺長 健 (正会員)

昭和 53 年京都大学大学院修士課程修了. 同年 NTT 入社. 平成 8 年より岡山大学教授. コンピュータビジョン, パターン認識, 人工知能の研究に従事. 工学博士. 共訳書『ロボットビジョン』(朝倉書店). 電子情報通信学会, IEEE 各会員.