

二つの視点から撮影した同一情景の領域間対応づけ

王 彩 華[†] 阿 部 圭 一^{††}

異なる視点から同一情景を撮影した画像の間で特徴を対応づける問題は、多くのコンピュータビジョンの応用において、基本的な問題となっている。この問題に対して、従来、エッジを特徴とするアプローチが良く使われているが、本論文では、領域を特徴として用いる一つのアプローチを提案する。この方法は二つの段階、すなわちマッチング段階と検証・修正段階からなっている。マッチング段階では、注目領域とその近傍にある三つの領域でなす3種類の三角形によって表される領域間局部構造の類似度に基づいて、弛緩法を使って領域の対応づけを行う。検証・修正段階では、得られた領域対応からカメラの移動パラメータを推定し、それからカメラの移動によって生じるオプティカルフローを推定する。この全体的整合性のある推定オプティカルフローを用いて、前段階で得られた領域対応の中から誤った対応（つまり、全体的整合性に一致しない対応）を検出し、修正する。

Region Correspondence for Color Scene Images Taken from Different Viewpoints

CAIHUA WANG[†] and KEIICHI ABE^{††}

Image correspondence, that is, finding correspondences of various primitives between images of the same scene but taken from different viewpoints, is a basic task in many computer vision applications. In contrast to the widely used edge-based approach, we propose here a region-based matching approach for image correspondence. Our method consists of a matching stage and a validating and correcting stage. In the first stage, local ternary relations among regions, that is, three triangles composed of the centers of gravity of the region in question and its three nearest neighbors, are used together with chromatic features of regions to establish region correspondences. In the second stage, the optical flow generated by the camera movement which can be estimated using all obtained correspondences is used to validate individual correspondences, and the ones which do not obey the optical flow are detected as mistakes and then corrected using the estimated optical flow.

1. はじめに

ほぼ同じ情景を視点を変えて撮った2枚の画像のあいだで、同一の特徴どうしを対応づけるという画像のマッチングは、動画処理、ステレオ画像のマッチング、画像中の対象物の同定などの応用には重要な一環であり、良く研究されている。その代表的な手法として、エッジに基づく画像マッチング手法があるが^{1)~3)}、最近、領域に基づく画像マッチング手法も多く研究されてきた^{4)~7)}。その理由は領域に基づく手法がいくつかの利点を持っているからである。エッ

ジに比べて、領域はその個数も少ないし、ノイズや物体の隠れには相対的に安定であり、また、領域に対してはより多くの情報（例えば、領域の面積、色、形状、モーメントなど）を利用することもできる。これらの理由により、領域に基づく画像の対応づけはより信頼性高く、より効率的に行える。領域単位のマッチングでは大まかな情報しか得られないという欠点はあるが、この点は、領域のマッチングをガイド情報としたエッジなどのもっと詳細な特徴のマッチングにより解決できる。

領域に基づく画像マッチング手法として、いくつかの手法が提案されている。その一つは領域の特徴だけを用い、領域の間の関係を利用しない手法である^{4),6)}。この手法では、シーン中にいくつかの似ているものが存在したり、視点の変化や領域分割の不一致性によって一つの物体の両画像での像の特徴がかなり異なったりする場合、領域の対応づけの任意性が生じる。この

[†] 静岡大学 電子科学研究科
Graduate School of Electronic Science and Technology,
Shizuoka University

^{††} 静岡大学 情報学部情報科学科
Department of Computer Science, Faculty of Information,
Shizuoka University

問題に対して、領域の間の関係を用いる構造的な手法が提案されている^{5),7)}。Leeら⁵⁾は領域の特徴を用いて初期のマッチング確率を得た後、領域の重心の間の角度一致性に基づいて弛緩法で大域的なマッチングを行なった。すなわち、各画像にある各々三つの領域の重心で成す三角形を作り、2枚の画像にある対応できる二つの相似三角形(つまり、二つの三角形の三つの内角が同じであり、かつ三つの頂点がそれぞれ対応候補になっている)の間で弛緩処理を行っている。しかし、視点の変化や領域分割の不一致によって、対応する三つの領域から構成される二つの三角形が相似三角形にならないこともあるし、また、二つの三角形が相似三角形でありながら領域の対応として誤っていることもあり得る。

一方、Chenら⁷⁾は領域の隣接関係による構造的な手法を提案した。しかし、領域間の隣接関係は、一つの物体の領域(面)のあいだの関係を表現するには有効であるが、別々の物体に属する領域の関係としては不安定である。すなわち、視点の変化や領域分割の結果によって、画像中の物体の面や領域間の隣接関係が変わってしまうことが多い。また、同じ隣接関係でありながら、その周辺の領域の並び方の点で領域間構造が異なっていることもある。つまり、隣接関係だけでは領域の構造情報を十分に表せない。さらに、全体構造の最適マッチングを探索するため、膨大な計算をしなければならない。

本論文では、二段階からなる手法を提案する。第一の局所構造マッチング段階では、注目領域とその近傍にある三つの領域で成す三つの三角形で注目領域の局所構造を表し、この局所構造の類似度に基づいて領域のマッチングを行う。さらに、この局所構造よりもっと広い範囲での構造情報を利用するため、弛緩法を用いる。ここで、局所構造の類似度は、三つの三角形の類似度(角度の類似度、辺の長さの類似度および頂点領域の色類似度)を用いて計算する。その三つの三角形はもう一方の画像で多少変わっても、まだよく似ているという意味で、この局所構造はLeeらやChenらの提案よりも安定である。この局所構造は画像の移動や回転にも不変^{5),8)}であるし、それにより注目領域とその近傍領域の相対的な位置関係も完全に決められる。したがって、この局所構造と弛緩法を用いることによって優れたマッチング結果を期待できる。

しかし、他のすべての手法と同様、二つの画像間の領域の対応を全部正しく見つけるのは容易ではない。普通、得られた領域対応の中には誤りが存在していると考えられる。そこで、第二の検証・修正段階を用意

し、第一段階で得られた領域対応の全体的整合性を利用して、明らかに間違っている対応を検出し、修正する。情景は静止して、カメラだけが動いた場合には、二つの画像に写されている物体の像の画像中での移動は規則的である。つまり、像の移動はカメラの移動によって生じたオプティカルフローに従っている。ここで、領域の対応(その多数は正しくできていると仮定して)からカメラの移動パラメータを推定し、このパラメータにより定まるオプティカルフローを推定する。このオプティカルフローを利用して領域の対応づけの誤りを検出し、修正する。

2. 領域分割

我々はカラー画像に対して領域を単位とした対応づけ(領域マッチング)を行う。ここで、連結しているほぼ同一の色を持つ画素どうしが一つの領域となるように領域の分割を行う。カラーモデルとしては、Yagiら⁹⁾が提案したHSVの改良モデル2を用いる。ただし、S軸とV軸の比は2:1とした。

具体的な手順は以下の通りである。

まず画素の色に注目してHSV(色相・彩度・明度)の三次元空間でクラスタリングを行い、各画素を分類する。クラスタリングによって、各画素に一つのラベルが与えられるが、しかし、同じラベルの画素は必ずしも一つの連結領域になってはおらず、バラバラになっていることが多い。ここで、連結している同じラベルの画素が一つの領域として一つのラベルとなるように、再びすべての画素にラベルづけ直す。

次に面積がある閾値 T_1 以下のきわめて小さい領域を、色などの一致性を問わず、それと一番長い境界線を共有する領域に合併する。この処理を面積が小さい順に閾値 T_1 以下である領域がなくなるまで繰り返す。

面積が閾値 T_1 以上 T_2 以下($T_2 > T_1$)の領域(Aとする)に対して、次の色類似条件と境界条件を用いて合併対象領域(Bとする)を決め、その合併対象領域に合併する。

色類似条件: 領域AとBの色のHSV空間での距離はある閾値以下である。つまり

$$\sqrt{(V_A - V_B)^2 + S_A^2 + S_B^2 - 2S_A S_B \cos(H_A - H_B)} \equiv D_c(A, B) \leq T_c \quad (1)$$

境界条件: 領域AとBとの共通の境界線の長さはAの周長の1/3以上である。

この二つの条件を満たす領域AとBに対して、領域AをBに合併することができる。合併の順番は次のように決める。

- (1) 各領域 A について、 A との間で色類似条件と境界条件を満たし、もっとも長い境界線を共有する領域 A_m を求める。
- (2) $\{(A, A_m)\}$ の中で A の周長に正規化した共通境界線の長さが最大になる対を求める。すなわち、 $l_p(A)$ を A の周長、 $l_c(A, A_m)$ を領域 A と A_m の共通境界線の長さ、 $l_n(A, A_m) \equiv l_c(A, A_m)/l_p(A)$ とするとき、次の式

$$l_n(A^*, A_m^*) = \max_{A_m} l_n(A, A_m) \quad (2)$$

を満たす領域の対 (A^*, A_m^*) を求め、これらを合併する。

この手順を繰り返して、色類似条件と境界条件を満たす領域を合併して行く。

3. 領域の3項近傍関係に基づくマッチング

2枚の画像の領域分割の結果から得られた各領域に關しては、いくつかの特徴量（平均色、面積など）を計算することができる。これらの領域特徴に基づいて領域のマッチングが行えるが、しかし、似たもののあるシーンではマッチングの任意性が生じる。ここで、領域の間の関係構造の類似度を併用して、各領域の多数の特徴の似ている可能な対応領域から、周辺の領域との関係構造も最も似ている“最適”な対応領域を決める。

領域の関係構造を領域のマッチングに使うために、まず三つの領域の間の関係を表す領域の3項関係¹⁴⁾を導入する。次に、注目領域とその近傍領域との3項関係でなす領域の局所構造を定義し、領域とその対応候補との間の局所構造の類似度の計算法を与える。もっと広い範囲での領域間の関係を利用するには、領域の局所構造に基づく弛緩法を用いる。

3.1 領域の3項近傍関係

ここで用いる領域の関係は三つの領域の間の関係を表す領域の3項関係である。三つの領域の間の関係として、三つの領域の重心の作る三角形を考える（図1）。もしその三つの領域が正しく対応づけされているならば、二つの画像における三角形は大体似ているであろう。この三角形で表される関係として、辺の長さや角の大きさをを用い、頂点の絶対的な座標を用いないことにすれば、関係は平行移動や回転に依存しないという特徴ももつ^{5),8)}。

一方、ある注目する領域とその近傍の三つの領域を考えると、注目領域と各二つの近傍領域で3種類の三角形（図2の $\triangle AA_1A_2, \triangle AA_2A_3, \triangle AA_1A_3$ ）を作る。この3種類の三角形によって、それら四つの領域の空

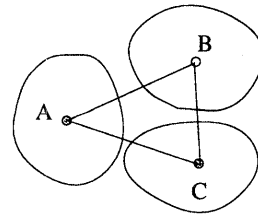


図1 領域の3項関係
Fig. 1 Ternary relations among regions.

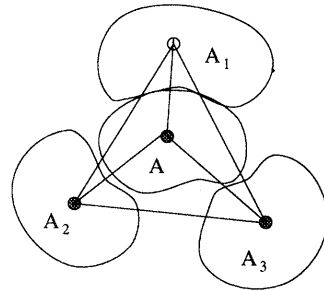


図2 領域の局所構造
Fig. 2 Local relational structure of regions.

間的配置の相対的な順番は完全に決まる。注目領域とその近傍の三つの領域の間の3種類の三角形で表す関係を注目領域での局所構造と呼ぶ。

3.2 局所構造の類似度

画像 A の領域 A の三つの近傍を (A_1, A_2, A_3) とする。 A の画像 B 中のマッチング候補領域を B^i とする。ここで、 A を B^i にマッチングしたとき、 A と (A_1, A_2, A_3) でなす局所構造は B^i の局所構造とどの程度一致するかという類似度を計算することを考える。

A を領域 B^i にマッチングしたと仮定し、 A の三つの近傍領域 (A_1, A_2, A_3) をそれぞれ画像 B 中の対応領域候補に対応させる。それによって、 A の局所構造を表す3種類の三角形もそれぞれ画像 B 中にある3種類の三角形に対応づけされる。この3種類の三角形の類似度の和を求め、 (A_1, A_2, A_3) をその特定のマッチング候補に対応させたときの A と B^i の局所構造の類似度とする。 (A_1, A_2, A_3) をすべての可能なマッチング候補に対応させ、その最大値を A と B^i との間の構造類似度とする。二つの三角形の類似度はその頂点・辺・角のそれぞれの類似度の積である。

具体的には以下ようになる。 A の三つの近傍 $A_n (n = 1, 2, 3)$ のマッチング候補を $B_n^k (k = 1, 2, \dots, m_n)$ とする。ここで、 m_n は A_n に対する画像 B 中のマッチング候補の個数である。

(A_1, A_2, A_3) の一つのマッチング $(A_1, A_2, A_3) \xrightarrow{P} (B_1^j, B_2^k, B_3^l)$ に対して、 A の構造類似度 $S(A, B^i, P)$

は次のように定義される。

$$\begin{aligned} S(A, B^i, \mathcal{P}) &= S_t(\Delta AA_1 A_2, \Delta B^i \mathcal{P}(A_1) \mathcal{P}(A_2)) + \\ & S_t(\Delta AA_1 A_3, \Delta B^i \mathcal{P}(A_1) \mathcal{P}(A_3)) + \\ & S_t(\Delta AA_2 A_3, \Delta B^i \mathcal{P}(A_2) \mathcal{P}(A_3)) \\ &= S_t(\Delta AA_1 A_2, \Delta B^i B_1^i B_2^i) + \\ & S_t(\Delta AA_1 A_3, \Delta B^i B_1^i B_3^i) + \\ & S_t(\Delta AA_2 A_3, \Delta B^i B_2^i B_3^i) \quad (3) \end{aligned}$$

ただし、領域の重心を領域と同じ記号で表している。また、 $S_t(\Delta abc, \Delta def)$ は二つの三角形 $\Delta abc, \Delta def$ の間の類似度で、次のように定義する。

$$\begin{aligned} S_t(\Delta abc, \Delta def) &= S_v(\Delta abc, \Delta def) \\ & S_e(\Delta abc, \Delta def) \\ & S_a(\Delta abc, \Delta def) \quad (4) \end{aligned}$$

$S_v(\Delta abc, \Delta def)$ は二つの三角形の頂点の類似度で、その三つの頂点領域の類似度の積である。最初は領域の色の類似度を領域の類似度とする。

$$\begin{aligned} S_v(\Delta abc, \Delta def) &= (1 - D_c(a, d))(1 - D_c(b, e)) \\ & (1 - D_c(c, f)) \quad (5) \end{aligned}$$

ここに、 $D_c(a, d)$ 等は式(1)で定義した領域の色の間の距離である。

$S_e(\Delta abc, \Delta def)$ は二つの三角形の辺の長さの類似度で、その三つの辺の長さの類似度の積と定義する。

$$\begin{aligned} S_e(\Delta abc, \Delta def) &\equiv f_i(\overline{ab}, \overline{de}) f_i(\overline{ac}, \overline{df}) \\ & f_i(\overline{bc}, \overline{ef}) \quad (6) \end{aligned}$$

$\overline{ab}, \overline{de}$ などは辺の長さで、 $f_i(\overline{ab}, \overline{de}) \equiv 4\overline{ab} \cdot \overline{de} / (\overline{ab} + \overline{de})^2$ とする。

$S_a(\Delta abc, \Delta def)$ は二つの三角形の角類似度で、次のように定義されている。

$$\begin{aligned} S_a(\Delta abc, \Delta def) &= 1 - (|\angle bac - \angle edf| + \\ & |\angle abc - \angle def| + |\angle acb - \angle dfe|) / 360 \quad (7) \end{aligned}$$

A の近傍領域 (A_1, A_2, A_3) をすべてのマッチング候補に対応させ、その中の最大構造類似度

$$S(A, B^i) = \max_{\mathcal{P}} S(A, B^i, \mathcal{P}) \quad (8)$$

を領域 A と (A_1, A_2, A_3) でなす局所構造と領域 B^i の局所構造の間の類似度とする。

普通、 A の最も近い三つの領域 (A_1, A_2, A_3) において、 A とその対応候補者 B^i の間の局所構造の類似度を計算すれば良いが、しかし隠れの多いシーンでは、 A_1, A_2, A_3 のどれかがもう一方の画像で隠れたら、 A とその真の対応領域との局所構造の類似度が低下してしまう。そこで、 A に最も近い四つの領域を選んで、その中の各々三つの領域において A と B^i の局所構造の類似度を計算する。その中の最大類似度を A と B^i との局所構造類似度とする。

3.3 マッチング手順

1. 各領域の面積、重心位置、色 RGB の平均値を計算し、それから HSV の値を求める。画像 A の各領域 A に対して、画像 B から面積と色の平均値が領域 A のと似ている（それぞれある閾値以内の）領域 B^i を選んで、マッチング候補とする。領域 A と候補領域 B^i の色類似度 ($1 - D_c(A, B^i)$) を領域類似度の初期値 $S^0(A, B^i)$ とする。画像 B に対しても同様に画像 A 内の候補領域と領域類似度の初期値を定める。具体的に、候補領域となる条件は次の通りである。

1) $1/T_a \leq a(A)/a(B^i) \leq T_a$, ここで $a()$ は面積を表し、 T_a は適当な閾値である。

2) $D_c(A, B^i) \leq T_c$, ただし、 $D_c()$ は式(1)で定義された HSV 空間中での距離、 T_c は閾値である。

2. 画像 A の各領域 A に対して、3.1 節で述べた方法で A をその対応候補 B^i にマッチングしたときの A と B^i の間の局所構造の類似度を計算する。ただし、 S_v の計算には前回の弛緩処理で得られた領域の類似度 $S^{n-1}()$ を用いる。つまり、

$$\begin{aligned} S_v(\Delta abc, \Delta def) &= S^{n-1}(a, d) S^{n-1}(b, e) \\ & S^{n-1}(c, f) \quad (9) \end{aligned}$$

画像 B にたいしても同じ処理を行う。

3. 二つの画像のすべての領域に対して、そのマッチング候補の中で類似度の正規化を行う。すなわち、一つの領域のすべてのマッチング候補に対する類似度の和を 1 にする。

4. 二つの画像からのマッチングの候補の類似度を統合する。つまり、画像 A から見た領域 A と画像 B のマッチング候補領域 B との類似度と、画像 B から見た領域 B と画像 A のマッチング候補領域 A との類似度の和を領域 A と B のあいだの類似度とする。

5. 統合した領域類似度を新しい領域類似度 $S^n(A, B)$ とする。画像 A の各領域にたいしてそのマッチング候補領域から一番類似度の高い候補領域をマッチング領域として選ぶ。

6. 2~5 を、5 で得たマッチング結果が変わらなくなるまで繰り返す。

図 4 に、図 3 の原画像に対して行った領域分割の結果と領域マッチング結果を示す。面積に対する閾値は $T_a = 4$, 色類似度の閾値は $T_c = 0.07$ と定めた。領域分割に使った閾値は、 $T_1 = 100$, 画像 a) に対しては $T_2 = 800$, 画像 b) に対して $T_2 = 4000$ である。これらの閾値は試行錯誤により選ばれた値である。それぞれの処理時間は 44.8 秒と 9.5 秒である (SUN ワークステーション SPARC-10 でのユーザ時間)。

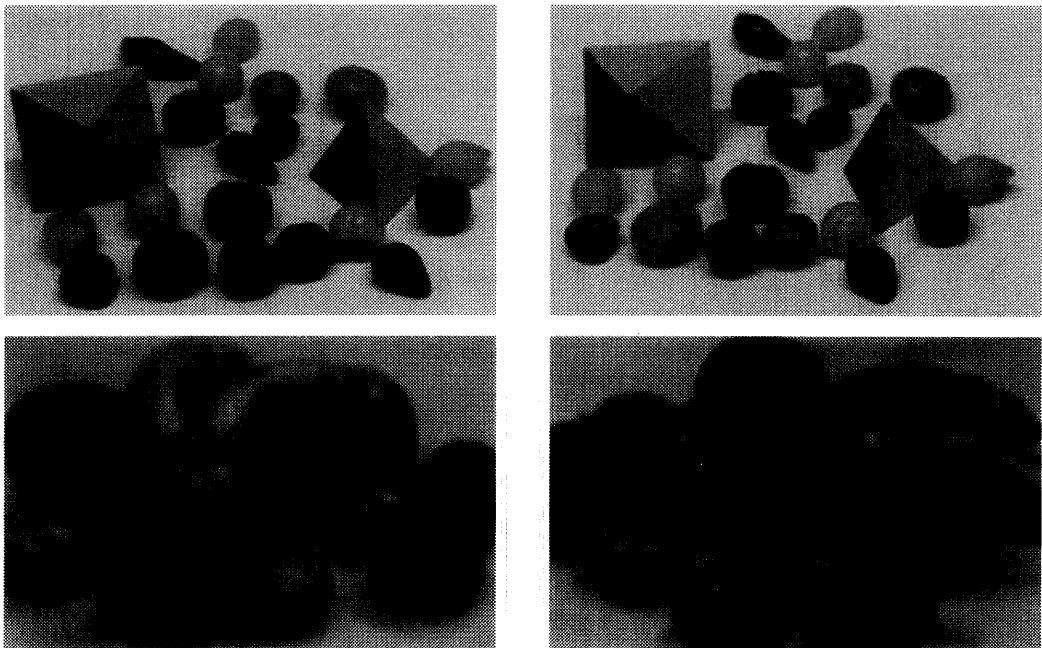


図 3 入力カラー画像 (明度成分を示す): a) 玩具 (上), b) 果物 (下)
 Fig. 3 Input images (only bright values are shown): a) Toys(above), b) Fruits(below).

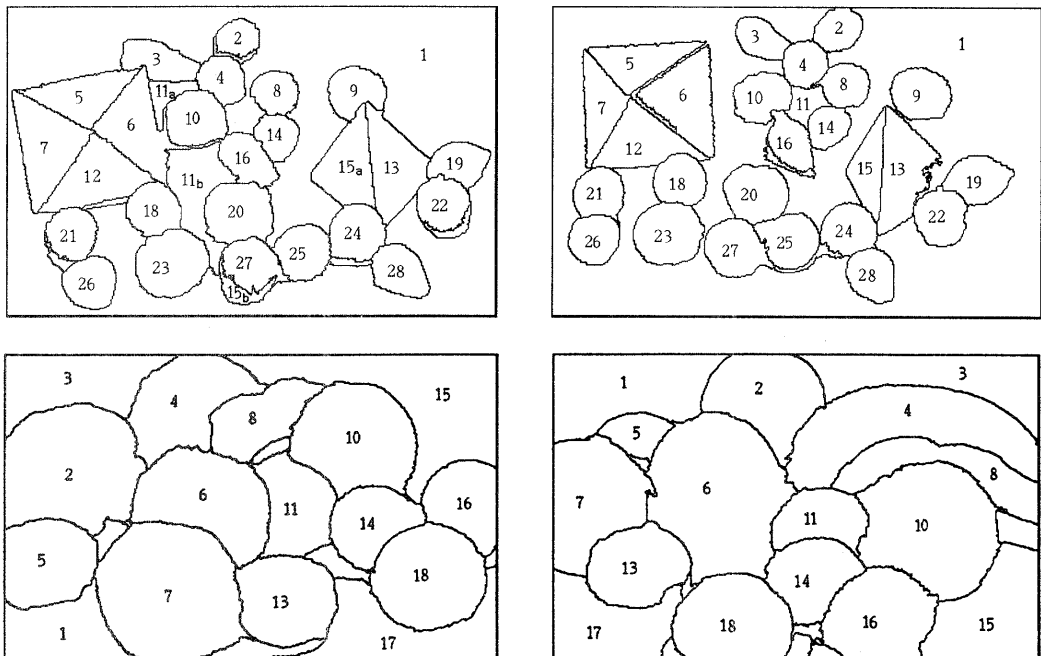


図 4 入力カラー画像の領域分割と領域マッチング結果: a) 玩具 (上), b) 果物 (下).
 同じ数字が対応付けられた領域である (多対一の対応の場合, a, b 等の添字で表す).
 Fig. 4 Results of segmentation and matching for input images in Fig. 3 a) Toys(above),
 b) Fruits(below). The same numbers show the matched regions(the multiple regions
 which correspond to a same region are subscripted by an alphabet).

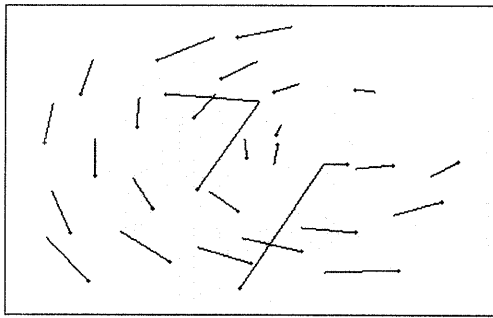


図5 図4(a)の対応領域の重心のずれ
Fig. 5 Locational shifts between corresponding regions in Fig. 4(a).

4. オプティカルフローによる誤対応の修正

前節で述べた手法によってよりいい領域マッチング結果を得られることが期待できるが、しかし、他のすべての手法と同様、二つの画像間の領域の対応を全部正しく見つけるのは容易ではない。普通、隠れや領域分割の不一致などによって、領域マッチング結果には誤対応が含まれていることが多い。図4(a)では領域分割の不一致によって、誤対応が生じている。ここで、得られた領域の対応の多数が正しいという前提の下で、それらの対応の全体的整合性を利用して誤対応を訂正することを考える。

情景は静止していてカメラだけが動いたと仮定すると、正しく対応づけされた領域に対して、その重心の画像間でのずれを求めれば、そのずれはカメラの移動によって生じた像の移動(オプティカルフロー)に一致していることが期待できる。図5には、図4(a)で得られた対応する領域の重心の画像間でのずれを示している。領域の変形、領域分割のされ方の違いにより通常のオプティカルフローよりもさらに大きなノイズが入っているけれども、その中の多数の正しい対応に対するずれは、かなり規則的である。この規則性を用いて、図5ではどの対応結果が誤っているか直観的に見てとれるであろう。以下は、これをアルゴリズムとして実行しようとする試みである。方針としては、得られたすべての対応からカメラ移動のパラメータを求め、全体として整合性のあるオプティカルフローを推定する。この推定オプティカルフローに従っていない対応を誤対応として検出し、推定オプティカルフローを利用してそれらの誤対応を修正する。

4.1 カメラ移動のパラメータの推定

カメラ移動のモデルとしては連続的なモデル¹⁰⁾と離散的なモデル^{11),12)}がある。ここでは、カメラの

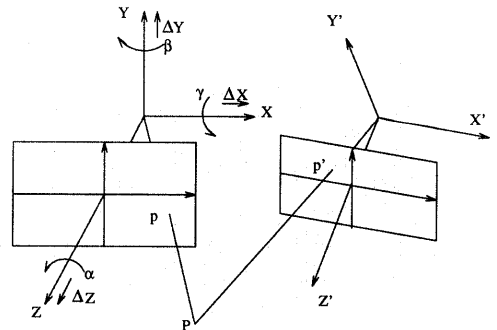


図6 カメラ移動のモデル
Fig. 6 Model for camera moving.

移動量が多い場合を想定しているの、離散的なモデルを用いる。また、上に述べたように、カメラ移動のパラメータの推定に使われるデータには、誤対応も含め誤差が大きいため、できるだけロバストな推定法を使う必要がある。

図6に示すように、カメラがまず順番に X , Y , Z 軸の回りにそれぞれ角度 γ , β , α で回転し、次に X , Y , Z 軸に沿ってそれぞれ $-\Delta X$, $-\Delta Y$, $-\Delta Z$ だけ平行移動したと仮定する。そうすると、3次元空間中の一点はカメラの移動に対して逆の方向に移動している。ここでは、回転角度の代わりにカメラの回転行列が分かれば十分であるから、回転行列 R によってシーンのカメラに対する回転移動を表す。同様にシーンの平行移動を T で表す。3次元空間の一点のカメラに対する移動 $(X, Y, Z) \rightarrow (X', Y', Z')$ は次式で表される。

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + T \quad (10)$$

ただし、

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \quad T = \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix}$$

一方の画像の一点 (x, y) がもう一方の画像で点 (x', y') と対応するとすれば、次の epipolar 制約条件が得られる。

$$\begin{pmatrix} x' & y' & 1 \end{pmatrix} W R \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0 \quad (11)$$

ただし、

$$W = \begin{pmatrix} 0 & -\Delta Z & \Delta X \\ \Delta Z & 0 & -\Delta Y \\ -\Delta X & \Delta Y & 0 \end{pmatrix}$$

上式の二つの行列の積 $WR \equiv E$ は essential matrix と呼ばれる。ノイズのない場合、いくつかの点の対応が分かれば、 E を定数倍の不確定性を残して定めることができる。 E から直接、平行移動量 $(\Delta X, \Delta Y, \Delta Z)$ の方向を決めることができ、 E の特異値分解によって回転行列 R が求められる¹¹⁾。

一般的な場合は、対応の数が多く、またノイズも入っている。そこで、最小二乗法で E を求め、 E の residual 分解で $(\Delta X, \Delta Y, \Delta Z)$ の方向と R を求める¹²⁾。さらに、各対応に対して

$$\min \left\| Z' \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} - ZR \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} - T \right\| \quad (12)$$

により、各対応の奥行き情報 Z と Z' ($(\Delta X, \Delta Y, \Delta Z)$ の大きさに対する相対値) を決める。

4.2 誤対応の検出

誤対応を検出するには、最も単純な手法として、推定奥行き情報に基づく検出手法がある¹³⁾。つまり、推定したカメラの移動パラメータでカメラが移動したと仮定して、対応領域の重心のペアの推定奥行き距離 Z または Z' が負の値になった（つまり、その領域がカメラの後ろにしかあり得ない）場合、誤対応と判断される。しかし、この手法では誤対応の検出に十分ではない。二つの対応領域の重心のずれが推定パラメータから予測されるオプティカルフローからかなり離れていても、その推定奥行き距離が負の値にならない場合は多い。一方、画像での動きの小さい領域対応に対して、領域の分割の不一致によって、その重心のずれがオプティカルフローと逆方向になった場合、一方の推定奥行き距離は負の値になるため、その対応を誤って誤対応と検出してしまう。

本論文では、誤対応を検出する2種類の統計的な手法について考察した。

一つの手法は推定パラメータの誤差分析に基づくものである。得られた全体の対応から推定したカメラの移動パラメータの値にはその真の値からの誤差がある。その誤差は二つの原因で生じると考えられる。一つの原因は、正しく対応された領域の重心の移動はカメラの移動パラメータで定められたオプティカルフローに従っていても、領域分割の不一致や見かけの領域形状の変化などのため、その重心の移動にランダムなずれ

を加えていることである。それは推定パラメータに第一種の誤差を与える。一方、間違った対応、または、対応が正しいけれども、その領域は一方の画像ではかなり隠されたり、かなりの部分が画像の枠からはみ出したりしている対応に関しては、その領域の重心の移動はカメラの移動パラメータで定められたオプティカルフローから大きく離れている。これらのカメラの移動パラメータに合わない対応が推定パラメータに第二種の誤差を与える。得られた対応の多数が正しいことを前提条件とし、第一種の誤差がランダムであると仮定すれば、それは最小二乗のさいに平均化によって抑えられ、第二種の誤差に比べてかなり小さいと考えてよい。また、最小二乗のさいの平均化効果のため、多数の正しい対応からその一つを抜いても、パラメータの推定にあまり影響を及ぼさない。一方、誤対応による第二種の誤差は推定パラメータの値を大きく左右する。これらの誤対応は少数なので、その一つ一つがパラメータの推定に大きい影響を与えている。そこで、次のような方法が考えられる。

もし全体の対応から一つの対応 P_i を除いてカメラの移動パラメータを推定し直し、その値と全体の対応を使って推定した値との差を計算し、それらの差を $\{d_i \mid i = 1, \dots, m\}$ (m は対応の個数) とするならば、 $\{d_i\}$ は二つのグループに分けられる。一つはオプティカルフローに従う対応に関するグループで、 d_i の値が小さい。もう一つはカメラの移動パラメータに合わない対応に関するグループで、 d_i の値も大きい。これらの可能な誤対応を簡単に誤対応と呼ぶ。

図4(a)と(b)の各対応を抜いたときの推定パラメータの変化 (d_i) を表1と表2に示す。 $\{d_i\}$ から、我々は Kittler ら¹⁵⁾ が提案した Kullback 情報距離の最小化による閾値選択手法で、誤対応を検出する。Kittler 法で決められた閾値も表1と表2に示されている。それらの閾値によって、誤対応の検出だけでなく、変形の大きい領域の対応も検出されている。この二種の区別は修正段階で行われる（後述）。

もう一つの手法では、推定したカメラの移動パラメータのもとで、各対応の画像平面での誤差を計算することによって、誤対応を検出する。一方の画像 I の一点 $p = (x, y)$ のもう一方の画像 I' での対応点を $p' = (x', y')$ とし、また、それぞれのカメラ原点を O と O' とすると、 (x', y') はカメラの移動パラメータと $p = (x, y)$ によって定められる epipolar 直線上にある。つまり、 (x', y') は O と p を通る直線の、両方の画像からともに見える部分を画像 I' に投影した直線分上にある。その線分的一端は直線 Op の無限遠点

表1 図4(a)の各対応を抜いたときの推定パラメータの変化

Table 1 Perturbations of estimated parameters by discarding each correspondence in Fig. 4(a).

抜かれた対応	パラメータの変化	抜かれた対応	パラメータの変化
1	0.0576	15 _a	0.0024
2	0.0133	15 _b	0.3040
3	0.0134	16	0.0025
4	0.0015	18	0.0003
5	0.0382	19	0.0031
6	0.0069	20	0.0027
7	0.0002	21	0.0007
8	0.0049	22	0.0097
9	0.0130	23	0.0019
10	0.0009	24	0.0035
11 _a	1.7139	25	0.0066
11 _b	0.1016	26	0.0058
12	0.0026	27	0.0077
13	0.0001	28	0.0241
14	0.0053	閾値: T = 0.030	

表2 図4(b)の各対応を抜いたときの推定パラメータの変化

Table 2 Perturbations of estimated parameters by discarding each correspondence in Fig. 4(b).

抜かれた対応	パラメータの変化	抜かれた対応	パラメータの変化
1	0.9623	11	0.7809
2	0.1814	13	0.1524
3	0.3418	14	0.5030
4	0.4400	15	1.5119
5	0.4039	16	1.5668
6	0.6556	17	0.2185
7	0.4949	18	0.0243
8	0.2815	閾値: T = 0.970	
10	0.8989		

の像 (x'_0, y'_0) である (図7)。

$$\begin{pmatrix} x'_0 \\ y'_0 \end{pmatrix} = \begin{pmatrix} r_1x + r_2y + r_3 \\ r_7x + r_8y + r_9 \\ r_4x + r_5y + r_6 \\ r_7x + r_8y + r_9 \end{pmatrix} \quad (13)$$

もう一端は両方の画像から見える Op 直線の最も近い点の像 (x'_1, y'_1) である。 $\Delta Z \geq 0$ の場合,

$$\begin{pmatrix} x'_1 \\ y'_1 \end{pmatrix} = \begin{pmatrix} (x'_0 + \Delta X)/(1 + \Delta Z) \\ (y'_0 + \Delta Y)/(1 + \Delta Z) \end{pmatrix} \quad (14)$$

その他のとき

$$\begin{pmatrix} x'_1 \\ y'_1 \end{pmatrix} = \begin{pmatrix} x'_0 + \Delta X \\ y'_0 + \Delta Y \end{pmatrix} \quad (15)$$

この (x, y) と (x', y') を対応領域の重心とすれば, 正しい対応に対して, (x', y') はその線分に近く, 誤対応の場合, (x', y') はその線分から遠く離れている

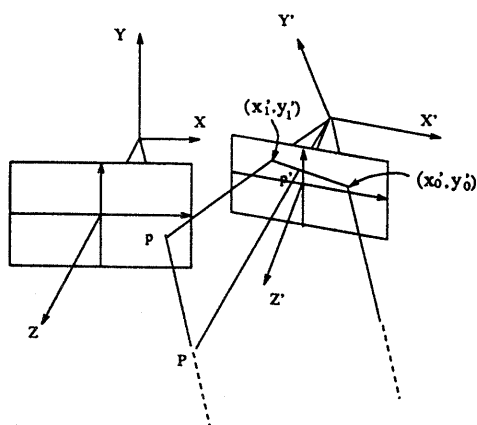


図7 Epipolar 直線で定めた点の対応範囲

Fig. 7 Corresponding segment on Epipolar line.

表3 図4(a)の各対応の推定パラメータに対する画像中での誤差

Table 3 Error of each correspondence in the image planes under the estimated parameters.

対応	誤差	対応	誤差	対応	誤差
1	0.0207	11 _a	0.0030	20	0.0044
2	0.0028	11 _b	0.0338	21	0.0199
3	0.0000	12	0.0156	22	0.0007
4	0.0000	13	0.0003	23	0.0162
5	0.0221	14	0.0006	24	0.0095
6	0.0122	15 _a	0.0029	25	0.0133
7	0.0254	15 _b	0.0804	26	0.0179
8	0.0003	16	0.0011	27	0.0132
9	0.0007	18	0.0092	28	0.0123
10	0.0016	19	0.0016	閾値: T=0.030	

はずである。 (x', y') からその線分までの距離をその対応の画像平面での誤差 $\{e_i\}$ とすると, 上の $\{d_i\}$ のように, $\{e_i\}$ も二つのグループになっている。

ここで, 図4(a)の各対応に対して計算した $\{e_i\}$ を表3に示す。 Kittler法で計算した閾値もその表に示している。 この手法では二つの誤対応が検出されたが, 残る一つは検出されなかった。

4.3 推定オプティカルフローによる誤対応の修正

上述の方法で検出された対応に対して, 次のステップで修正を行う。

1) 各対応に対して, 推定したカメラの移動パラメータと相対的な奥行き情報を使って, 一つの画像から他の画像へのあるべきオプティカルフローを計算する。 ただし, 誤対応に対しては, その奥行き距離を全体の対応の奥行き距離の平均値で代用する。 これを推定オプティカルフローと呼ぶことにする。 シーンの中の物体間の奥行き距離の変化範囲はそんなに大きくないと仮定すれば, 推定オプティカルフローはその領域の対応領域の大まかの位置を表しているであろうと考えら

れる。

2) 検出された誤対応に対して、次のように推定オプティカルフローを用いて、修正を行う。すなわち、カメラが移動する前の画像の領域 A に対して、移動後の画像の候補領域 B^i から推定オプティカルフローと一致しているもっとも類似度の高い候補領域を対応領域とする。ここで、類似度としては3章の方法で得られた最終的な類似度を用いる。候補領域が推定オプティカルフローと一致しているというのは、領域 A の重心と候補領域 B^i の重心とを結ぶベクトルが推定オプティカルフローから10度以内にあり、かつその長さも推定オプティカルフローの長さの2/3から3/2までの範囲内にあることとする（これらの値は経験的に決めた）。

3) 上のステップで修正できなかった誤対応に対して、その領域とその近傍にある正しい対応と2)で修正された対応に関する領域の近傍関係を使って、その対応を修正する。つまり、領域 A の二つの確実に対応ができた近傍領域 A_1 と A_2 を選んで、その三つの領域の位置関係と A_1 と A_2 の対応領域との関係を用いて A に対する対応領域の位置を推定する。この位

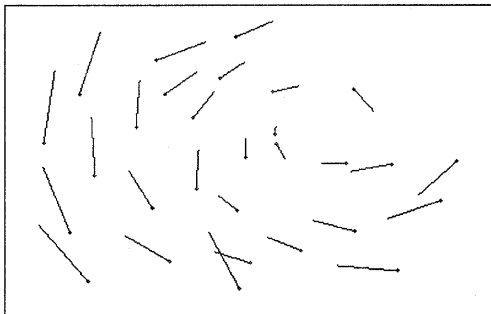


図8 図4(a)の対応から推定したオプティカルフロー
Fig. 8 Optical flow estimated from the region correspondences in Fig. 4(a).

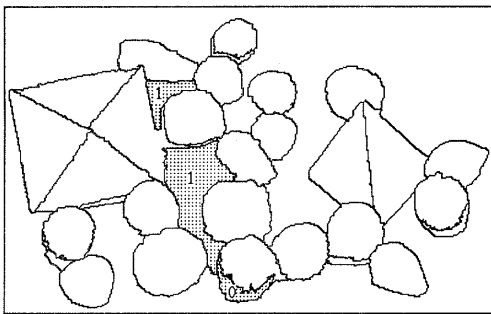


図9 図4(a)に修正された対応
Fig. 9 Corrected correspondences (for left image in Fig. 4(a)). A '0' means that no corresponding region exists in the other image.

置からある範囲で A の色と似ている領域を探し、見つけたら対応を付ける。見つけられない場合は、その領域はもう一つの画像では隠されたと判断する。

図4(a)のマッチング結果から計算した推定オプティカルフローを図8に示す。図9は、図4(a)のマッチング結果を修正した結果を示している。'0'はその領域に対する対応領域が見つけれなかったことを示している。ここでは、前節の第一の検出手法で誤対応を検出した。図4(a)の1と5の領域の対応と、図4(b)の15と16の領域の対応が誤って誤対応として検出されたが、修正ステップのステップ3)によって、元の正しい対応に戻された。オプティカルフローの推定と誤対応の修正に要した計算時間は、どちらの例でも1秒未満であった。

5. おわりに

領域マッチング手法ですべての領域の対応づけを正しく得るのは難しい。普通、得られた対応の多数は正しいが、誤った対応も含まれる。我々は、背景じたいは静的でカメラのみが移動する場合について、領域間の局所構造の類似度に基づいてマッチングを行った後、多数の正しい対応から、画像全体の領域対応の整合性を求め、これによってマッチング結果を修正することを考えた。本研究では、領域の対応からカメラの移動パラメータを推定し、このパラメータで定めたオプティカルフローを推定する。このオプティカルフローを利用して明らかに間違っている対応を検出し、修正する方法を示した。

領域マッチング方法としては、領域の3項関係に基づく領域の構造的なマッチング手法を提案した。この手法に用いられた領域の3項関係は近傍領域の重心の間の距離と角度を使うので、領域の隣接関係より安定であるし、画像中の領域の平行移動や回転にも依存しない。また、この3項関係によって注目領域の近傍領域の空間的配置の相対的な順番は完全に決まるから、隣接関係よりも領域間の構造情報を十分に表せる。したがって、上で仮定した「多数の対応は正しい」という前提を満たしやすい。

今後の課題としては、次のことが挙げられる。今まで調べたカメラの移動パラメータを推定する方法は、いずれも本研究のような大きな誤差を含んだデータには弱い。この問題に対してまだ検討する必要がある。今回の手法は領域の3項関係を用いたが、領域の間の透視変換に対する不変的な関係や、また複数の関係を用いるマッチング手法を考察する必要がある。また、今回はこの手法をほぼ静止しているシーンに適用した

が、本手法を静的な背景中に移動物体のある場合の領域マッチングと移動物体の検出に適用できないかと考えている。その場合には、誤対応であるか移動物体であるかを区別する方法が鍵になるであろう。

参考文献

- 1) Dhond, R. and Aggarwal, J.K.: Structure from Stereo - A Review, *IEEE Trans. on SMC*, Vol. 19, No. 6, pp. 1489-1510 (1989).
- 2) Ohta, Y. and Kanade, T.: Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming, *IEEE Trans. on PAMI*, Vol. 7, No. 2, pp. 139-154 (1985).
- 3) Horaud, R. and Skordas, T.: Structural Matching for Stereo Vision, *Proc. 9th Int'l. Conf. on Pattern Recognition*, pp. 439-445 (1988).
- 4) Marapane, S.B. and Trivedi, M.M.: Region-Based Stereo Analysis for Robotic Applications, *IEEE Trans. on SMC*, Vol. 19, No. 6, pp. 1447-1464 (1989).
- 5) Lee, H.J. and Lei, W.L.: Region Matching and Depth Finding for 3D Objects in Stereo Aerial Photographs, *Pattern Recognition*, Vol. 23, No. 1/2, pp. 81-94 (1990).
- 6) Fuh, C.S., Maragos, P. and Vincent, L.: Visual Motion Correspondence by Region-Based Approaches, *Proc. Asian Conf. on Computer Vision*, pp. 784-789 (1993).
- 7) Chen, Y. and Nakatani, H.: Image Region Correspondence by color and Structural Similarity, *IEICE Trans. on Information and Systems*, Vol. E76-d, No. 4, pp. 429-436 (1993).
- 8) Li, S.Z.: Matching: Invariant to Translation, Rotation and Scale Changes, *Pattern Recognition*, Vol. 25, No. 6, pp. 583-594 (1992).
- 9) Yagi, D., Abe, K. and Nakatani, H.: Segmentation of Color Aerial Photographs Using HSV Color Models, *Proc. IAPR Workshop on Machine Vision Applications*, pp. 367-370 (1992).
- 10) Bruss, A.R. and Horn, B.K.P.: Passive Navigation, *Computer Vision, Graphics and Image Processing*, Vol. 21, pp. 3-20 (1983).
- 11) Tsai, R.Y. and Huang, T.S.: Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces, *IEEE Trans. on PAMI*, Vol. PAMI-6, No. 1, pp. 13-27 (1984).
- 12) Weng, J. and Huang, T.S.: Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation, *IEEE Trans on PAMI*, Vol. 11, No. 5, pp. 451-476 (1989).
- 13) 王, 阿部: オプティカルフローによる画像領域マッチング結果の修正, 1994年度電子通信学会春季大会講演論文集, p. D-565 (1994).
- 14) 阿部, 王: 画像のマッチング, 平成5年度東海支部連合大会講演論文集, pp. S-67-S-68 (1993).
- 15) Kittler, J. and Illingworth, J.: Minimum Error Thresholding, *Pattern Recognition*, Vol. 19, pp. 41-47 (1986).

(平成6年11月11日受付)

(平成7年5月12日採録)

王 彩華 (学生会員)



1963年生。1983年中国人民大学情報工学科卒業。1986年同大学院修士課程修了。現在、静岡大学電子科学研究科博士後期課程在学中。画像理解・認識、文書画像処理に興味を持つ。電子情報通信学会学生会員。

阿部 圭一 (正会員)



1941年生。1963年名古屋大学工学部電気学科卒業。1968年同大学院博士課程満了。工学博士。名古屋大学工学部助手、静岡大学工学部講師、助教授を経て、現在静岡大学情報学部情報科学科教授。構造的パターン認識、画像処理アルゴリズム等の研究のほか、広く人間・機械インタフェース全般および計算機科学教育に興味をもつ。著書「ソフトウェア入門 第2版」(共立出版)。電子情報通信学会、人工知能学会、IEEE 会員。