

局所不変特徴量に基づく複数広告看板の認識*

市村 直幸†

あらまし モータースポーツ等のイベントのスポンサーは、出資の見返りとしてイベント会場に広告看板を設置できる。出資効果の確認のため、放送映像上での広告看板の露出時間や露出面積等を調べることは重要であり、この確認作業の自動化には映像内の広告看板の認識が必要となる。本論文では、局所不変特徴量に基づく複数広告看板の認識方法を提案する。本方法では、2つの基本的な処理を用いる。1つ目は、局所不変特徴量を用いたモデル画像とシーン画像間の対応付けである。局所不変特徴量の使用により、認識対象の変形や輝度変化、隠れへの対処が可能となるが、背景による誤対応や複数の同一認識対象による対応点の混合が生じる。その対応付け結果から個々の認識対象の対応点を分離するために、2つ目の処理として、改良を施した RANSAC による射影変換の計算を用いる。これら2つの処理を使用した、4段階からなる認識方法を示す。実画像を用いた実験を通じ、モデル画像が1枚のみでも、種々の状況での認識が可能であることを示す。

Recognizing Multiple Billboard Advertisements Based on Local Invariant Features

Naoyuki ICHIMURA†

Abstract The sponsors for events such as motor sports can put billboard advertisements in return for investments. Checking how ads appear in a broadcast is important to see the effectiveness of investments and recognizing ads in videos is required to make the check automatic. We propose a recognition algorithm based on local invariant features, which can handle multiple ads. Two fundamental procedures are used. The first one is matching between model and scene images using local invariant features. We can obtain point correspondences even recognition targets in scene images have deformation, illumination change and occlusion thanks to local invariant features. False matches caused by a background and mixed point correspondences due to the same multiple targets, however, appear. To extract the point correspondences of each target from the result of matching, the second procedure, the calculation of a homography by a modified RANSAC, is used. The proposed algorithm consists 4 steps based on the procedures. We demonstrate by experiments using real images that the proposed algorithm recognizes ads in various situations even only one model image is given.

1 まえがき

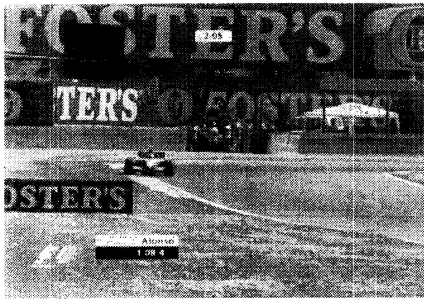
モータースポーツ等のイベントのスポンサーは、出資の見返りとしてイベント会場に広告看板を設置できる。出資効果の確認のため、放送映像上での広告看板の露出時間や露出面積等を調べることは重要であり、この確認作業を自動化するためには、映像内の広告看板の認識が必要となる。広告看板は、その大きさや設置場所、それをとらえるカメラの角度やズーム等により、様々な見え方で映像上に表れる。図1に、Formula 1 (F1) の映像における広告看板の例を示す [1]。図1 (a) では、広告看板の大

きさや設置場所の違いによるスケーリング、輝度変化、および、視野逸脱と他の物体による隠れが生じている。図1 (b) は、オンボードカメラの映像であり、スケーリングと斜めから見ることによる変形が生じている。これらの例が示すように、広告看板の認識では、認識対象の変形、輝度変化および隠れへの対処が問題となる。

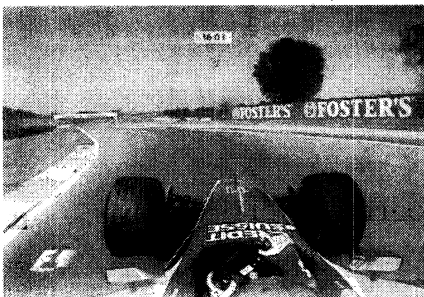
認識対象の変形、輝度変化および隠れに対処する方法の1つに、局所不変特徴量の使用がある。局所不変特徴量の構成は、認識対象の変形と輝度変化に対して不変となる、複数の局所領域の設定とその領域内の特徴を表す記述子 (descriptor) の計算により行われる。図2に、局所領域の例を示す。図中の四角形が、記述子を計算する局所領域を表す。複数の局所領域の使用により、認識対象の一部が隠れても、見えている部分の特徴量を用いて認識が行える。

*本研究の一部は、JSPS Overseas Research Fellowship の支援の下で行われた。

†産業技術総合研究所 脳神経情報研究部門, Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), nic@ni.aist.go.jp, <http://staff.aist.go.jp/naoyuki.ichimura/>



(a)



(b)

図 1: 広告看板の例。F1 において、サーキットに設置された広告看板。出資効果の確認のため、スポンサーは、広告看板が適切に設置されているかを現地で調べ、また、放送映像上での露出時間や露出面積等を調べる。(a) 広告看板の大きさや設置場所の違いによるスケーリング、輝度変化、および、視野逸脱と他の物体による隠れが生じている。(b) オンボードカメラの映像。スケーリングと斜めから見ることによる変形が生じている。広告看板の認識では、このような変形、輝度変化および隠れへの対処が問題となる。

局所不変特徴量には、局所領域の設定方法および記述子が異なるいくつかの種類がある。局所領域の設定方法には、スケールスペース内の特徴点抽出に基づく方法 [2, 3, 4, 5, 6, 7, 8, 9]、輝度の極値や領域分割に基づく方法 [6, 10, 11, 12]、エッジ抽出に基づく方法 [12, 13] 等がある。また、記述子には、輝度の微分 [3, 9]、局所領域形状を正規化した画像パッチ [5]、モーメント特徴 [10, 11, 12]、輝度勾配の方向ヒストグラム [4, 6, 7, 8, 13] 等が用いられる。これらの局所特徴量は、画像の幾何学的な相似変換もしくはアフィン変換、および、輝度のアフィン変換に対して不変である。

図 3 に、局所不変特徴量を用いた対応付けの例を示す。図中の左上に、認識対象のモデル画像を示している。この画像とシーン画像の記述子間の距離を求め、最近傍法により対応点を得た。特徴量の不変性により、認識対象のスケーリングや輝度変化、隠れに関わらず対応点は得られる。しかし、局所的に類似した部分が背景にあるため、誤対応が生じる。また、シーン内に複数の同一認識対象が存在するた

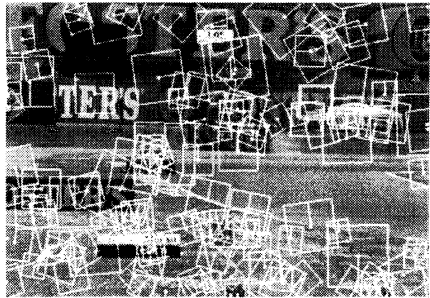


図 2: 局所領域の例。図中の四角形が、記述子を計算する局所領域を表す。複数の局所領域の使用により、認識対象の一部が隠れても、見えている部分の特徴量を用いて認識が行える。この画像では 4706 個の局所領域が得られるが、簡明さのため、3%だけを表示している。

め、複数の対象の対応点が混合される。本研究の目的は、このような状況から個々の認識対象の対応点を分離し、それを用いて認識を行うアルゴリズムの開発である。

1つの認識対象の対応点を背景や他の広告看板から分離することを、outlier (外れ値) を含んだ対応点に対するモデル当てはめ問題として取り扱う。ここでモデルとは、1つの認識対象の対応点が従う大局的な拘束を意味する。多くの広告看板は平面であるため、その大局的な拘束として射影変換 [15] が使用できる。よって、例えば図 3 において、最も大きな認識対象の対応点へ射影変換を当てはめると、背景とその他の広告看板の対応点を outlier として分離できる。

Outlier を含む対応点に対する射影変換の当てはめには、ロバスト推定的一种である RANSAC (Random SAmple Consensus) [14] が使用できる。しかし、outlier が多い場合、outlier を含むサンプルが偶然に最大の投票数を得て、誤った解が選択されることがある。誤対応を減少させる前処理は使用できるが [8, 12]、複数の同一認識対象から得られる混合された対応点は局所的に正しいため、大局的な拘束を当てはめる前にそれらの一部を取り除くことは困難である。よって、多くの誤対応を含んだ状態で射影変換の計算を行うことを、常に想定する必要がある。RANSAC において誤った解が選択されるのは、射影変換の自由度が高すぎ、視点位置や認識対象の姿勢の変化によっては起こり得ないモデル画像の変形まで表現されるためと考えられる。よって、正しい解を選択するためには、求めた射影変換によるモデル画像の変形を評価することが有効と言える。

本論文では、以下の 4 段階の処理からなる複数広告看板の認識方法を提案する: (1) 局所不変特徴量を用いた仮対応付け、(2) 射影変換による仮位置合せ、(3) 探索領域を制約した対応付け、(4) 最終的

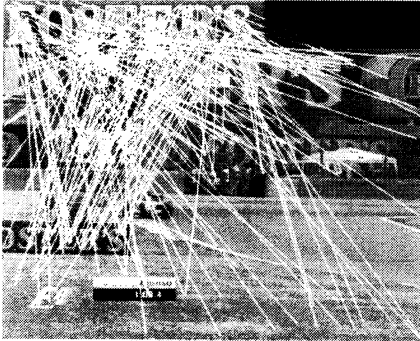


図 3: 局所不変特徴量を用いた対応付けの例。左上の画像は、認識対象のモデル画像である。この画像とシーン画像の記述子を、最近傍法により対応付けた。特徴量の不変性により、認識対象のスケールや輝度変化、隠れに関わらず対応点は得られる。しかし、局所的に類似した部分は背景にもあるため、誤対応が生じる。また、シーン内に複数の同一認識対象が存在するため、複数の対象の対応点が混合される。個々の認識対象の対応点を背景や他の広告看板から分離するために、1つの認識対象の対応点が従う射影変換を推定する。

な位置合せと検証。(1)では、対応点を得ると共に、記述子間の距離の比に基づく誤対応の除去[8]を行う。(2)では、RANSACにより射影変換を計算し、1つの認識対象のシーン画像内での大まかな位置を得る。RANSACで正しい解を選択するため、求めた射影変換によるモデル画像の変形が、視点位置や物体姿勢の変化によって生じ得るものかどうかを評価する。(3)では、(2)の仮位置合せの結果により探索領域を制約し、対応点を得る。この探索領域の制約により、1つの認識対象の周辺のみから対応点を得ることができる。(4)では、(3)で得られた対応点から射影変換を計算し、最終的な位置合せ結果を得る。この結果の検証のため、変換後のモデル画像とシーン画像を正規化相関を用いて比較する。(1)~(4)の処理を終了条件が満たされるまで繰り返し、複数広告看板を認識する。実画像を用いた実験を通じ、モデル画像が1枚のみでも、種々の状況での認識が可能なことを示す。

2 対応付けに用いる局所不変特徴量

本節では、対応付けに用いる局所不変特徴量について述べる。図4は、広告看板のモデル画像の例である。広告看板には文字が多く用いられ、テクスチャに乏しい。そのため、エッジに基づく局所不変特徴量を対応付けに用いる。

局所領域の設定には、Difference of Gaussian (DoG)によるスケール空間内の極値に基づく特徴点抽出[4, 5, 7, 8]を用いる。この特徴点抽出の結果、特徴点の位置と特徴点が発見されたスケールが得られる。記述子としては、輝度勾配の方向ヒ



図 4: モデル画像の例。広告看板には文字が多く用いられ、テクスチャに乏しい。そのため、エッジに基づく局所不変特徴量を対応付けに用いる。

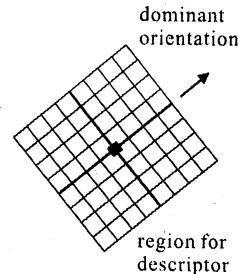


図 5: 記述子を計算する局所領域。特徴点を中心に、最も頻度の高い輝度勾配の方向に回転させた局所座標系を作る。特徴点のスケールに比例した大きさをもつ局所領域を設定し、それをいくつかのブロックに分割する。各ブロック内での輝度勾配の方向ヒストグラムを連結し、そのノルムを1に正規化して記述子とする。

ストグラム[4, 6, 7, 8, 13]を用いる³。DoGはエッジ抽出フィルタである Laplacian of Gaussian (LoG)を近似するため、この記述子は多重スケールエッジの方向ヒストグラムとも解釈でき、テクスチャの乏しい画像にも有効と考えられる。Mikolajczykらの比較実験[16]で、この記述子の高い不変性が示されていることも選択の理由である。

記述子の計算の概要を述べる。特徴点の近傍領域で輝度勾配の方向ヒストグラムを求め、最も頻度の高い方向 (Dominant Orientation) を得る。そして、図5のように、特徴点を中心として、その方向に回転させた局所座標系を作る。その座標系内で、特徴点のスケールに比例した大きさをもつ局所領域を設定し、それをいくつかのブロックに分割する。各ブロック内での輝度勾配の方向ヒストグラムを連結し、そのノルムを1に正規化して記述子とする。

スケール空間内の極値は並行移動、スケール不変であり、最も頻度の高い輝度勾配の方向を基準とした回転角は回転不変であるため、上記の局所特徴量は相似変換に対して不変となる。実際には、ある程度のアフィン変換に対しても不変性を保つ[8, 16]。また、ノルムの正規化により、輝度のアフィン変換に対しても不変となる。次節で、この局所不変特徴量を用いた認識アルゴリズムを示す。

³この選択により、使用した局所不変特徴量は Lowe によって提案された SIFT[4, 8] となる。

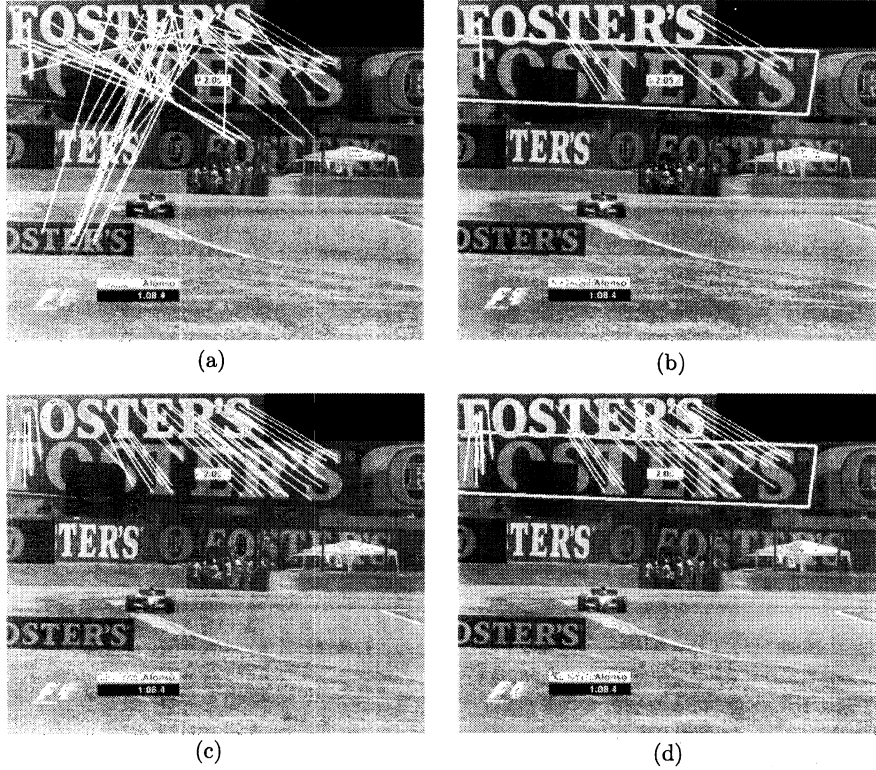


図 6: 認識処理の例。(b),(d)では, inlier のみを対応点として示している。また, 位置合せ結果は四角形で表す。(a) 式 (1) を用いた仮対応付け。最近傍法を用いた図 3 に比べ, inlier を保持しつつ outlier を減少できている。しかし, 複数の同一認識対象の対応点は混合されたままとなる。(b) 射影変換による仮位置合せ。モデル画像の変形を評価する RANSAC により, 1 つの認識対象の位置を得る。(c) 探索領域を制約した対応付け。(b) の結果から探索領域を制約し, 1 つの対象対象の周辺のみから対応点を得る。(d) 最終的な位置合せ。(c) で得られたより多くの対応点から射影変換を再度計算し, 認識結果を得る。複数の同一認識対象の存在に関わらず, 1 つの認識対象の分離に成功している。

3 認識アルゴリズム

本節では, 提案する認識アルゴリズムを, 以下の表現を用いて説明する。モデル画像の局所不変特徴量を $f_i^m = \{p_i^m, \sigma_i^m, d_i^m\}$, $i = 1, \dots, N^m$ と表す。ここで, p_i^m は同次座標で表した特徴点の位置, σ_i^m は特徴点が発見されたスケール, d_i^m は記述子, N^m は特徴量の個数である。同様に, シーン画像の局所不変特徴量を $f_j^s = \{p_j^s, \sigma_j^s, d_j^s\}$, $j = 1, \dots, N^s$ とする。特徴量間の距離として, 記述子間のユークリッド距離 $d_{ij} = \|d_i^m - d_j^s\|$ を用いる。

3.1 局所不変特徴量を用いた仮対応付け

最近傍法では, 特徴量 f_i^m に対し, $j_{1NN} = \arg \min_j d_{ij}$ のインデックスを持つ特徴量 $f_{j_{1NN}}^s$ を対応付ける。しかし, 図 3 に示すように, 最近傍法では多くの誤対応が生じる。この誤対応を減少させるため, 次式を満たす対応点のみを使用する [8]。

$$d_{ij_{1NN}}/d_{ij_{2NN}} < t, 0 \leq t \leq 1 \quad (1)$$

ここで, $j_{2NN} = \arg \text{secondmin}_j d_{ij}$, つまり, 2 番目に近い特徴量のインデックスであり, t はしきい値である。式 (1) は, 最近傍の距離 $d_{ij_{1NN}}$ が, 2 番目に近い特徴量の距離 $d_{ij_{2NN}}$ より t で規定されるだけ離れていることを条件としている。

式 (1) を用いて得られる対応点数は, t を大きくするにつれて増加し, $t = 1$ の最近傍法で最大となる。射影変換の計算には 4 点以上の inlier が必要であるため, それを確保するように t を定める必要がある。本論文では, 対応点数がある数 P_{min} に達するまで, 次式に従って t を増加させた。

$$t(k+1) = \alpha t(k) \quad (2)$$

$$\alpha = 1.01, t(0) = 0.80, k = 0, 1, 2, \dots$$

ここで, k は繰り返し回数, α は t の増加を制御する係数である。 P_{min} をいくつにすれば最低限必要な inlier を確保できるかは, シーンによって異なる。そのため, いくつかの P_{min} を用いて認識を行う。その詳細は 3.6 節で述べる。

図 6 (a) に $P_{min} = 20$ とした場合の例を示す。図 3 に比べ、inlier を保持しつつ、outlier を減少できている。しかし、複数の同一認識対象の対応点は混合されたままとなる。これらは局所的に正しいため、式 (1) を用いても取り除けないのである。ある認識対象の射影変換の計算において、その他の対象の対応点は誤対応として働くため、混合した対応点は RANSAC で誤った解が選択される要因になり得る。次節では、正しい解を選択するための、モデル画像の変形に関する評価を導入した RANSAC について述べる。

3.2 射影変換による仮位置合せ

対応点を $C = \{p_k^m, p_k^s\}$, $k = 1, \dots, P$ と表す。この対応点を関係付ける射影変換を H (3×3 行列) とし、変換誤差を次式で定義する。

$$e_k = \|p_k^s - Hp_k^m\|, k = 1, \dots, P \quad (3)$$

H は、以下に示す RANSAC [14] で計算できる。

(i) 集合 C から、4 つの対応点からなるサンプルをランダムに取り出す。

(ii) Direct Linear Transformation (DLT) アルゴリズム [15] と式 (3) の変換誤差の和を評価関数とする非線形最適化により、サンプルから H を求める。

(iii) 全対応点に対して式 (3) の変換誤差を計算し、次式を満たす inlier の数 (投票数) N_i を求める。

$$e_k < \varepsilon, k = 1, \dots, P \quad (4)$$

ここで、 ε はしきい値である。

(iv) (i)~(iii) の処理を繰り返し、最大の投票数をもつ inlier を得る。

(v) (iv) で得られた inlier より H を計算する。

上記のアルゴリズムで、outlier を含むサンプルが偶然に最大の投票数を得る場合がある。この誤りが生じる主たる原因は、射影変換の自由度が高すぎ、視点位置や認識対象の姿勢の変化によっては起こり得ないモデル画像の変形まで表現されることが考えられる。よって、解の選択の誤りを防ぐために、次の処理を (ii) の後に加え、射影変換によるモデル画像の変形を評価する。

(ii') H でモデル画像を変換した結果が、ねじれ四角形、もしくは、反転した四角形の場合には、(i) に戻る。そうでない場合は、(iii) に進む。

図 7 (a) はねじれ四角形、(b) は反転した四角形の例である。このような生じ得ない変形に対応する射影変換に対しては、式 (4) による投票を行わない。ねじれ四角形は、向かい合う 2 組の辺の変換後の交点位置が、変換後の頂点から構成される凸包内にあるかどうかで検出できる。反転した四角形は、符号付き面積 [17] の符号から検出できる。これら



(a)



(b)

図 7: RANSAC で投票から除外される変形の例。(a) ねじれ四角形。(b) 反転した四角形。このような変形は、視点位置や認識対象の姿勢の変化によっては生じない。

検出のための計算は非常に効率が良いため、(ii') の評価は、繰り返し計算を必要とする RANSAC に適している。

図 6 (a) の対応点より 10000 個のサンプルを得て、射影変換を計算した結果、9899 個の射影変換が、ねじれ、または、反転した四角形に対応した。このように、実際に RANSAC で数多くの生じ得ない変形が表れることから、(ii') の変形の評価は正しい解の選択のために有効であると言える。図 6 (b) に、式 (4) で $\varepsilon = 3$ [pixel] とした場合の仮位置合せ結果を示す。直線は inlier であり、四角形は射影変換によりモデル画像を変換した結果である。誤対応や混合した対応点が除去され、1 つの認識対象の位置が得られている。

3.3 探索領域を制約した対応付け

図 6 (b) に示すような仮位置合せの結果を用い、背景や他の広告看板の影響を排除した状態で、1 つの認識対象の周辺のみから対応点を得ることができる。仮位置合せで得られた射影変換 H を用い、次の予測位置を求める。

$$\hat{p}_k^s = Hp_k^m, k = 1, \dots, P \quad (5)$$

この予測位置を中心に半径 r の円形の探索領域を設定し、その中にある特徴量に対し式 (1) を用いて対応付けを行う。1 つしか対応点が見つからず、式 (1) が評価できない場合には、その対応点を用いる。

探索領域の半径 r は、図 8 に示す関数より決定する。横軸 a は、モデル画像と H により変換したモデル画像の面積比である。図 6 (b) では、モデル画像と位置合せ結果を表す四角形の面積の比となる。この関数により、シーン画像内の認識対象の面積が大きいほど、探索領域を大きくする。面積比の最小値、最大値を a_{min} , a_{max} とし、それぞれに対応する半径を r_{min} , r_{max} としている。

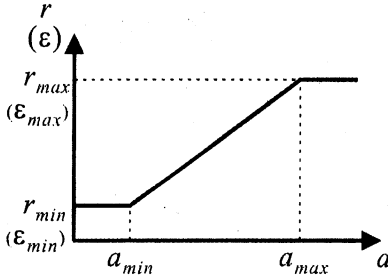


図 8: 探索領域の半径を決定する関数。横軸は変換前後のモデル画像の面積比 a ，縦軸は探索領域の半径 r である。この関数により，シーン画像内の認識対象が大きいほど，探索領域を大きくする。3.4 節で用いる，最終的な位置合せにおける RANSAC のしきい値 ϵ ，および，正規化相関を計算する局所領域の半径 r についても，同様に面積比の関数として変化させる。

図 6 (c) に， $a_{min} = 1$ ， $a_{max} = 4$ ， $r_{min} = 20$ [pixel]， $r_{max} = 80$ [pixel] (画像サイズは 720×480 [pixel]) として得た対応点を示す。図 6 (b) で $a = 1.65$ であったため， $r = 33$ [pixel] となった。この半径を使用した探索領域の制約により，背景や他の広告看板の影響が排除され，1 つの認識対象のみから対応点が得られている。

探索領域を制約した対応付けは guided matching と呼ばれ，射影変換や基礎行列の計算に使用されている [15]。提案しているアルゴリズムでは，guided matching は対応付けのためだけでなく，1 つの認識対象の対応点のみを抽出するセグメンテーションとしても働いている点が重要である。

3.4 最終的な位置合せと検証

探索領域を制約した対応付けにより得られる対応点を用い，3.2 節と同様にして射影変換 H を求める。式 (4) のしきい値 ϵ は，図 8 に示す関数により決定する。つまり，シーン画像内の認識対象が大きいほど，式 (3) の変換誤差を許容する。得られた射影変換によってモデル画像を変換し，最終的な位置合せ結果を得る。

その後認識対象の検証を行う。 N_i 個の inlier に対し，式 (5) と同様に予測位置を求め，図 8 より得られる半径 r をもつ局所領域を設定する。その局所領域の中で， H により変換したモデル画像とシーン画像の正規化相関 NCC_l ， $l = 1, \dots, N_i$ を求める⁴。そして，その平均値が次式を満たす場合に，最終的な位置合せは正しいと判断する。

$$\frac{1}{N_i} \sum_{l=1}^{N_i} NCC_l < \gamma \quad (6)$$

⁴正規化相関は輝度のアフィン変換に対し不変であるので，処理全体において輝度不変性は保たれる。

ここで， γ はしきい値である。正規化相関の計算を，画像全体ではなく局所領域内で行うため，隠れの影響を小さくできる。

図 6 (d) に， $a_{min} = 1$ ， $a_{max} = 4$ ， $\epsilon_{min} = 3$ [pixel]， $\epsilon_{max} = 12$ [pixel]， $\gamma = 1.0^5$ として得た位置合せ結果を示す。より多くの inlier から射影変換が計算され，認識対象の位置が正しく得られている。式 (6) の NCC_l の平均値は，2.32 であった。

以上の 3.1 節から 3.4 節の処理により，図 6 において，複数の同一認識対象が存在するにも関わらず，1 つの認識対象の分離に成功している。

3.5 アルゴリズムの終了条件

最終的な位置合せが正しいと判断された場合，認識済みの広告看板がある領域内 (図 6 (d) では，認識結果を表す四角形内) に位置する局所不変特徴量を取り除く。他の認識対象を認識するため，残った特徴量に対し 3.1 節から 3.4 節の処理を行う。これを以下の終了条件が満たされるまで繰り返す: (a) 対応付けにおいて， P_{min} 個以上の対応点が得られない，(b) 射影変換の計算において，4 点以上の inlier が得られない，(c) 式 (6) の条件が満たされない。それぞれの条件は，局所的な類似性がない場合，大局的な拘束を満たさない場合，位置合せ結果が誤っている場合，に対応する。

3.6 対応点数に対する複数仮説

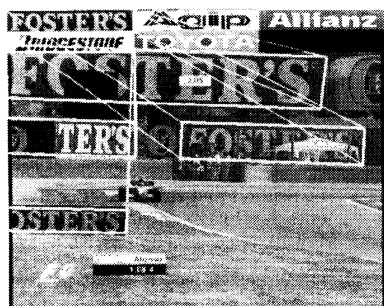
対応点数を P_{min} 個確保するため，式 (1) のしきい値 t を式 (2) により変化させる。 P_{min} をいくつにすれば 4 点以上の inlier が確保できるかはシーンに依存し，事前に予測できない。そのため， $P_{min} = 20, 60, 120$ の 3 つの場合に対し認識を行う。3 つの認識結果を得た後，認識された広告看板の数が最も多い結果を選択する。広告看板の数が同じ場合には，式 (3) の変換誤差の平均が最も小さい結果を選択する。このように，対応点数に対する複数の仮説を使用することにより，同一認識対象の数や隠れの度合いが異なる種々のシーンに対応できる。

4 実験

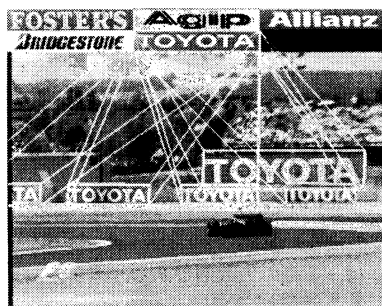
実験の題材として，広告看板の見え方の変化が比較的大きい F1 の映像を選択した。5 つの広告看板を認識対象とし，図 9 の各画像の上部にモデル画像を示している。各認識対象に対し，モデル画像は，図 9 に示した 1 枚のみを与えている。

記述子の次元数は，図 5 に示す局所領域内のブロック数，および，各ブロックでの輝度勾配の方向ヒストグラムの bin 数により決まる。本実験では， 8×8 ブロック，bin 数 16 とした。よって，記述子の

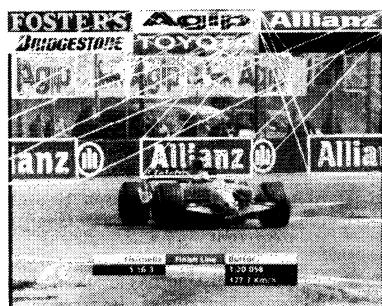
⁵正規化相関は R,G,B の各チャネルで計算したため，[0,3] の値を持つ。



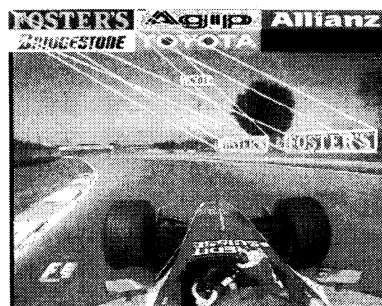
(a)



(b)



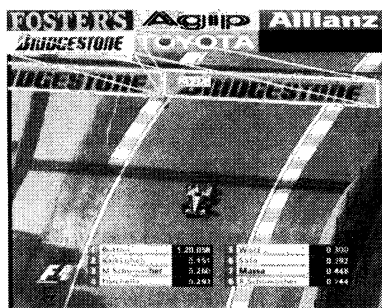
(c)



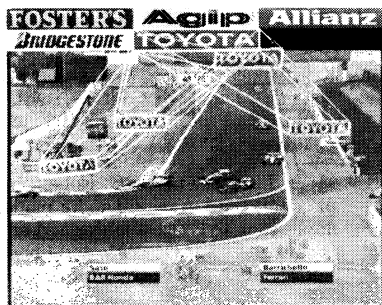
(d)



(e)



(f)



(g)



(h)

図 9: F1 の映像に対する認識結果。(a)~(f) は成功例である。(a),(b) スケーリング、輝度変化および隠れがある。(c) 金網による隠れや輝度変化、および、車体と視野逸脱による隠れがある。(d),(e) スケーリング、視点の位置による変形、輝度変化がある。(f) ヘリコプターのカメラからの映像。輝度変化と隠れがある。このように、モデル画像が 1 枚のみでも、提案方法によって種々の状況での認識に成功している。(g) と (h) は失敗例である。(g) 不変量でカバー可能な変形やスケーリングの範囲を逸脱したと考えられるため、手前の地面にある認識対象と遠くの小さい認識対象から対応点が得られなかった。(h) 隠れの度合いが大きいため、左上の認識対象の認識に失敗した。局所不変特徴量の改良、および、複数のモデル画像の使用が、このような失敗をなくす方法と考えられる。

次元数は $8 \times 8 \times 16 = 1024$ となる⁶。これ以外のパラメータは、図 6 に対する値を参照されたい。

図 9 (a)~(f) は認識の成功例である。それぞれのシーン画像には、スケーリングや視点位置による変形、輝度変化、隠れが見られるが、全ての認識対象の認識に成功している。この結果は、提案方法により、モデル画像が 1 枚でも、種々の状況での認識が可能なことを示している。

図 9 (g),(h) は、認識の失敗例である。図 9 (g) の手前の地面にある認識対象では、それをほぼ水平に見る位置に視点があるため、変形が非常に大きくなっている。また、遠くにある認識対象は、非常に小さく観測される。これらの認識対象の変形やスケーリングは、不変量でカバー可能な範囲を逸脱していると考えられ、対応点は得られず認識に失敗した。図 9 (h) では、左上の認識対象の認識に失敗した。この場合は、隠れの度合いが大きく、十分な数の対応点が得られなかった。

図 9(g),(h) の失敗例のように、局所不変特徴量を用いても対応点が得られなければ、当然認識に失敗する。このような状況を改善する 1 つの方法は、より変形や隠れに強い不変特徴量を開発することである。また、他の方法として、あらかじめ想定される認識対象の変形を含んだ、複数のモデル画像を使用することが挙げられる。今後は、この 2 つの方法について検討を行う予定である。

5 むすび

本論文では、局所不変特徴量に基づく複数広告看板の認識方法を提案した。基本的に、提案方法は、局所不変特徴量を用いた対応付けと、改良を施した RANSAC による射影変換の計算から構成される。局所不変特徴量の使用により、認識対象に変形や輝度変化、隠れが生じて、モデル画像との対応付けが行える。また、モデル画像の変形に対する評価を、RANSAC に組み込んだ。これにより、視点位置や認識対象の姿勢の変化では起き得ないモデル画像の変形を与える射影変換を、解の候補から除去することができる。この対応付けと射影変換の計算方法を用いた、4 段階からなる認識方法を示した。実画像を用いた実験の結果、モデル画像が 1 枚のみでも、種々の状況での認識が可能なことを確認した。今後の課題として、より不変性の高い特徴量や複数のモデル画像を使用した認識方法の開発、および、認識方法の定量的評価が挙げられる。

⁶他の論文 [4, 6, 8, 13] では $4 \times 4 \times 8 = 128$ 次元が多いが、広告看板に対しては正しい対応点が十分得られなかった。これは、テクスチャが少ないため、より細かくエッジの方向を見る必要があるためと考えられる。

参考文献

- [1] 本研究では、SKY PerfecTV フジテレビ 721 において放送された映像を使用している。
- [2] C. Harris and G. Giraudon: "A combined corner and edge detector," Proc. 4th Alvey Vis. Conf., pp.147-151, 1988.
- [3] C. Schmid and R. Mohr: "Local greyvalue invariants for image retrieval," IEEE Trans. PAMI, Vol.19, No.5, pp.530-535, 1997.
- [4] D. Lowe: "Object recognition from local scale-invariant features," Proc. Int. Conf. Comp. Vis., pp.1150-1157, 1999.
- [5] M. Brown and D. Lowe: "Invariant features from interest point groups," Proc. British Machine Vis. Conf., pp.656-665, 2002.
- [6] J. Sivic and A. Zisserman: "Video Google: a text retrieval approach to object matching in videos," Proc. Int. Conf. Comp. Vis., Vol.2, pp.1470-1477, 2003.
- [7] M. Brown and D. Lowe: "Recognising panoramas," Proc. Int. Conf. Comp. Vis., Vol.2, pp.1218-1225, 2003.
- [8] D. Lowe: "Distinctive image features from scale-invariant keypoints," Int. J. Comp. Vis., Vol.60, No.2, pp.91-110, 2004.
- [9] K. Mikolajczyk and C. Schmid: "Scale & affine invariant interest point detectors," Int. J. Comp. Vis., Vol.60, No.1, pp.63-86, 2004.
- [10] F. Schaffalitzky and A. Zisserman: "Viewpoint invariant texture matching and wide baseline stereo," Proc. Int. Conf. Comp. Vis., pp.636-643, 2001.
- [11] J. Matas, O. Chum, M. Urban and T. Pajdla: "Robust wide baseline stereo from maximally stable extremal regions," Proc. British Machine Vis. Conf., pp.384-393, 2002.
- [12] T. Tuytelaars and L. Van Gool: "Matching widely separated views based on affine invariant regions," Int. J. Comp. Vis., Vol.59, No.1, pp.61-85, 2004.
- [13] K. Mikolajczyk, A. Zisserman and C. Schmid: "Shape recognition with edge-based features," Proc. British Machine Vis. Conf., pp.384-393, 2003.
- [14] M. A. Fischler and R. C. Bolles: "Random Sample Consensus: A paradigm for model fitting with application to image analysis and automated cartography," ACM Graphics and Image Processing, Vol.24, No.6, pp.381-395, 1981.
- [15] R. Hartley and A. Zisserman: "Multiple view geometry in computer vision," 2nd edition, Cambridge University Press, 2003.
- [16] K. Mikolajczyk and C. Schmid: "A performance evaluation of local descriptors," Proc. Int. Conf. Comp. Vis. and Patt. Recog., pp.384-393, 2003.
- [17] T. Moller and E. Haines: "Real-time rendering," 2nd edition, A.K.Peters, 2002.