

# 密なエッジサンプリングに基づく 局所不変特徴量による対応付け

市村 直幸

産業技術総合研究所 脳神経情報研究部門

nic@ni.aist.go.jp, <http://staff.aist.go.jp/naoyuki.ichimura/>

あらまし 画像の対応付けのために局所不変特徴量を抽出する際、画像内に記述子を計算する局所領域を設定する。この局所領域は、LoG(Laplacian of Gaussian) フィルタ等を用いた特徴点抽出による方法で設定できる。本論文では、特徴点抽出による方法において、特徴点抽出フィルタの処理範囲と局所領域の差異に起因し、対応付けに有効な輝度変化を有する部分に局所領域が設定されないことを指摘する。そして、輝度変化をより活用するために、密なエッジサンプリングによる局所領域の設定方法を提案する。密なエッジサンプリングにより、対応付けに有効な情報を有する局所領域数が大きく増加することを、記述子のエントロピーを用いて定量的に明らかにする。GPU による実装を通じた実験により、隠れのあるシーンやテクスチャが乏しいシーンにおいて、頑健な対応付けが可能になることを示す。

## Matching with Local Invariant Features Based on Dense Edge Sampling

Naoyuki ICHIMURA

Neuroscience Research Institute,

National Institute of Advanced Industrial Science and Technology (AIST).

nic@ni.aist.go.jp, <http://staff.aist.go.jp/naoyuki.ichimura/>

**Abstract** Detecting local regions in which descriptors are computed is necessary to extract local invariant features for image matching. The filters for feature point extraction such as LoG (Laplacian of Gaussian) have been used to find the appropriate positions of the local regions in an image. In this paper, we point out on local region detection that the portions of an image with intensity variations useful for image matching are not used as the local regions due to the difference between the sizes of the filters and the local regions. In order to take full advantage of intensity variations, we propose a method to detect the local regions based on dense edge sampling. Using the entropies of descriptors, we quantitatively show that the number of local regions with intensity variations useful for image matching is greatly increased by dense edge sampling. Experimental results obtained by a GPU-based implementation demonstrate the robustness of the proposed method to scenes with occlusions and less textures.

### 1 まえがき

画像の対応付けは、複数の画像間で共通部分を見出す処理であり、多視点画像処理や物体認識、ロボットビジョン等の基本を成す。対応付けでは、視点の移動や照明条件の違いにより、画像間でシーンの見えに変化が生じても、その影響を受けずに対応を見出す機能が要求される。また、視野逸脱や遮蔽による隠れの下で、部分的な対応を得ることも重要な機能である。このような機能を実現するために、局所不変特徴量が幅広く用いられている [1, 2, 3, 4]。局所不変特徴量は、(1) 局所領域の設定、(2) 局所領域の画像特徴を表す記述子の計算、の2段階の処理を通じて抽出される。スケールスペースに基づく特徴点抽出や局所座標系の導入等を通じ、上記(1),(2)の処理結果が画像の幾何学的変換や輝度変化に

対し不変になるようにする。また、特徴の局所性により、隠れへの対応が可能となる。図3に局所領域の例を示す。図中の正方形が、記述子を計算する局所領域を表す。このような局所領域内で計算された記述子の比較により、画像の対応付けを行う。

局所領域の設定は、特徴点抽出により行える。特徴点抽出には、次の機能が必要である：(i) 輝度変化を含む領域が抽出可能なこと、(ii) 特徴点のスケール情報を与えること。前者は対応付けには輝度変化が必要であるため、後者は対応付けをスケール不変にするために要求される。これらの機能をもった特徴点抽出の結果、特徴点の位置、および、固有スケール(characteristic scale)[5]が得られる。局所領域は、特徴点の位置を中心とし、固有スケールに比例した大きさをもつように設定される

(図3の正方形の中心と一辺の長さに対応)。

上記の特徴点抽出による局所領域の設定方法は、対応付けにおいて多用されてきた。その一方、局所特徴に基づく一般物体認識では、異なる方法も採用されている。一般物体認識におけるスケールスペース内での局所領域の設定方法を大別すると、特徴点抽出、グリッドサンプリング、ランダムサンプリング、および、エッジサンプリングとなる [6, 7, 8, 9, 10, 11]。Nowak は、bag-of-features の枠組の下で特徴点抽出とランダムサンプリングの比較を行い、ランダムサンプリングの方が認識性能が高かったと報告している [10]。他の論文でも、認識性能向上のために、サンプリングによる方法を使用している [7, 8, 9, 11]。また、特徴点抽出による方法では、認識性能向上のための重要なファクターとなる局所領域数が不足するとの指摘もある [10]。これらのことは、特徴点抽出では、シーンを記述する情報を画像から十分に引き出せていないことを示唆している。

本論文の目的は、サンプリングによる局所領域の設定方法の、対応付けにおける有効性を検討することである。まず、特徴点抽出による方法では、特徴点抽出フィルタの処理範囲と局所領域の差異に起因し、対応付けに有効な輝度変化を有する部分に局所領域が設定されないことを指摘する。そして、輝度変化をより活用するために、密なエッジサンプリングによる局所領域の設定方法を提案する。提案方法により設定された局所領域が、対応付けに有効な情報を有するかどうかを、記述子のエントロピーにより評価する。そして、密なエッジサンプリングにより、対応付けに有効な情報を有する局所領域数が大きく増加することを示す。また、GPUによる実装を通じた実験を行い、提案方法の対応付けにおける有用性を確認する。

本研究のように、記述子のエントロピーを用い、サンプリングによる方法を定量的に評価した研究は、著者が知る限りにおいて過去に見当たらない。また、密なサンプリングによる計算量の増加に対処するための、GPUによる実装結果を示している点も、本研究の新規性と考えている。以下、各節において研究の詳細を示す。

## 2 特徴点抽出による局所領域設定の問題点

特徴点抽出による局所領域の設定では、Harris-Laplacian に代表されるように、コーナー検出フィルタと LoG に基づくフィルタを併用する [3, 4]。多重解像度のコーナー検出と LoG により、局所領域内部のエッジ形状の制約と、コントラストに基づく特徴点の固有スケールの検出を行う。局所領域は、特徴点を中心とし設定される。また、その大きさは固有スケールの数倍から十数倍に取られることが多い。これは、特徴量の計算に用いる画素数を確保するためである。特に、ヒストグラム特徴のように、ある量の発生頻度を特徴量とする場合

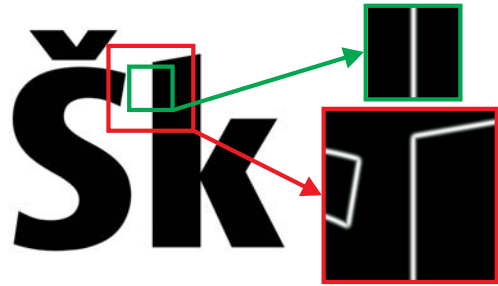


図1: フィルタの処理範囲と局所領域の差異による、同一位置におけるエッジ情報の変化。小さな正方形がフィルタの処理範囲を、大きな正方形が記述子を計算する局所領域を表す。

には、画素数の確保は重要となる。

上記のように局所領域の大きさを決定すると、ほとんどの場合、局所領域はフィルタの処理範囲より広がる。図1に一例を示す。2つの正方形が同一位置に示してある。小さい方がフィルタの処理範囲を表し、大きい方が局所領域を表す。フィルタの処理範囲で観測される輝度変化は、対応付け結果が不定となる垂直エッジに基づくものである。よって、この位置に特徴点は抽出されず、局所領域も設定されない。しかし、局所領域で観測を行うと、いくつかの方向を有するエッジにより、対応付け結果を一意に定めるのに有効な輝度変化が得られる。この例が示すように、特徴点抽出では、フィルタの処理範囲と局所領域の差異に起因し、対応付けに有効な輝度変化を有する部分であっても、局所領域が設定されるとは言えない。一般物体認識において、特徴点抽出よりも種々のサンプリングによる局所領域の設定が高い認識性能を与えるのは、特徴点抽出では上記の問題により、シーンを記述する情報を画像から十分に引き出せていないことが一要因であると思われる。

## 3 密なエッジサンプリングによる局所領域設定

前節で述べた特徴点抽出による方法の問題点から、対応付けにおいても、サンプリングによる局所領域の設定は検討に値すると考えられる。まず、グリッドサンプリングおよびランダムサンプリングを検討する。これらの方法では、設定された局所領域には必ずしも大きな輝度変化が存在するとは言えない。物体の種類によっては(例えば空などでは)輝度変化が少ないことが他の物体との識別に重要な情報となるため、一般物体認識においてはこれらのサンプリングが有効に働くと考えられる。しかし、対応付けでは、輝度変化が少ない部分是对応点を一意に定めるには不相当であり、これらのサンプリングは有効ではないと考えられる。よって、輝度変化の検出に基づくエッジサンプリングを用いる。

本研究では、スケールスペースにおいて、空間  $3 \times 3$  近傍でのエッジの極大点を全て用いる密なエッジサンプリングを提案する。この密なサンプリングにより、特徴

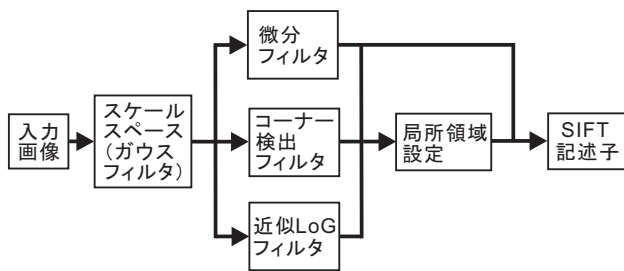


図 2: 局所不変特徴量抽出処理のフローチャート。密なエッジサンプリングによる局所領域, および, 特徴点抽出による局所領域, それらを併用して記述子を計算する。

点抽出では得られなかった輝度変化を対応付けに利用できると考えられる。また, スケールの変化によって自然に生じるエッジの極大点の変化を密に観測することが, シーンの形状を表すのに適切であると考えている。

図 2 は, 密なエッジサンプリングと特徴点抽出を併用した, 局所不変特徴量抽出アルゴリズムのフローチャートである。まず, 入力画像から, ガウスフィルタによりスケールスペースピラミッドを生成する。そして, そのスケールスペースピラミッドに対し, 3 種類のフィルタを適用する。微分フィルタ [12] では, エッジ抽出を行い, 輝度勾配の大きさと方向を計算する。近似 LoG フィルタ [13] とコーナー検出フィルタ [14] は, 特徴点の位置と固有スケールを得るために使用する。

局所領域の設定は, 次の 2 つの方法で行う。1 つ目の方法は, 密なエッジサンプリングによる方法である。微分フィルタの処理結果の全スケールにおいて, 微分フィルタの出力がしきい値以上, かつ, 空間  $3 \times 3$  近傍の極大点を探索する。そして, その条件に合致する全てのエッジ点を中心に局所領域を設定する。2 つ目は, 特徴点抽出による方法である。近似 LoG フィルタの出力を使用し, スケールスペースの極値点を探索する。極値点において, コントラストを表す近似 LoG フィルタの出力, および, コーナーらしさを表すコーナー検出フィルタの出力, この 2 つの出力がしきい値以上ならば, その点を中心として局所領域を設定する。この処理により, 主として blob を特徴とした局所領域を得る。このように 2 つの方法を併用することにより, エッジと blob という性質の異なる特徴に基づいた局所領域の設定が行える。設定された局所領域内部において, 輝度勾配の大きさと方向を用いて SIFT 記述子 [2, 15] を計算する。

次節において, 上記の密なエッジサンプリングによる局所領域の設定により, 対応付けに有効な輝度変化を有する局所領域数が大幅に増加することを示す。

#### 4 記述子のエントロピーによる局所領域評価

本節では, 密なエッジサンプリングにより設定された局所領域から, 対応付けに有効な情報が得られるかどう

かを評価する。図 3 の上段に, 評価に使用した画像を示す。図 3(a) は, テクスチャがなく輝度変化はシャープなエッジ上のみ存在する画像である。図 3(b) は, ロボットの自己位置推定等に利用される中心窩レンズ [17] により撮像された廊下の画像であり, テクスチャが乏しい。図 3(c) は, テクスチャやエッジを比較的豊富に含むシーンである。

上記の画像から, 初期スケールを 1.6 とし, 図 3(a),(b) では 4 オクターブ, 図 3(c) では 5 オクターブのスケールスペースピラミッドを生成した。各オクターブは 5 枚の画像を含む。それを元に, 図 2 のアルゴリズムで記述子を計算した。局所領域の大きさは, 中心画素のスケールの 15 倍とした。微分フィルタ, 近似 LoG フィルタ, コーナー検出フィルタの出力に対するしきい値は, それぞれ, 10, 10, 100 である。

図 3 の中段および下段は, それぞれ, 特徴点抽出による局所領域, および, 密なエッジサンプリングと特徴点抽出の併用による局所領域を表す。これらの図は, 密なエッジサンプリングにより, 局所領域数が明らかに増加することを示す。ここで重要なことは, 増加した局所領域が対応付けに有効な情報を含むかどうかである。この点の評価するために, 局所領域で計算された SIFT 記述子のエントロピーを用いる。SIFT 記述子は, 輝度勾配の方向のヒストグラムに基づく。そのため, そのエントロピーが大きいほど, 局所領域内部に複雑な輝度勾配の方向変化が存在することを意味する。これは, 様々な方向をもつエッジにより, 対応付け結果を一意に定めるのに有効な情報を局所領域が有するものと解釈できる。よって, もし, 記述子のエントロピーが大きい局所領域が増加すれば, 対応付けに有効な情報を抽出するために密なエッジサンプリングが有用であると言える。

図 4 に, 記述子のエントロピーによる局所領域の評価結果を示す。図 4(a),(b),(c) は, それぞれ, 図 3(a),(b),(c) に対応している。上段の棒グラフは, 各エントロピーの値の範囲における局所領域数を表す。“LoG” は特徴点抽出による方法, “Edge” は密なエッジサンプリングによる方法を示す。第 2, 第 3, 第 4 段は, 局所領域内部のエッジを表すパッチの例である。各段が各エントロピーの値の範囲に対応する。

各画像において, 特徴点抽出でのエントロピーの値は, その大多数が 4 以上に分布している。従来の研究において, 特徴点抽出による局所領域から, 対応付けに有効な特徴量が得られることが確認されている [1, 2, 3, 4]。よって, SIFT 記述子ではエントロピーが 4 以上であれば, 局所領域は対応付けに有効な情報を有していると考えられる。実際に, 図 4(a) のパッチを見ると, エントロピーが 4 未満では対応付け結果を一意に定めるのに有効な輝度変化が存在しているとは言えない。また, 図



# Škoda

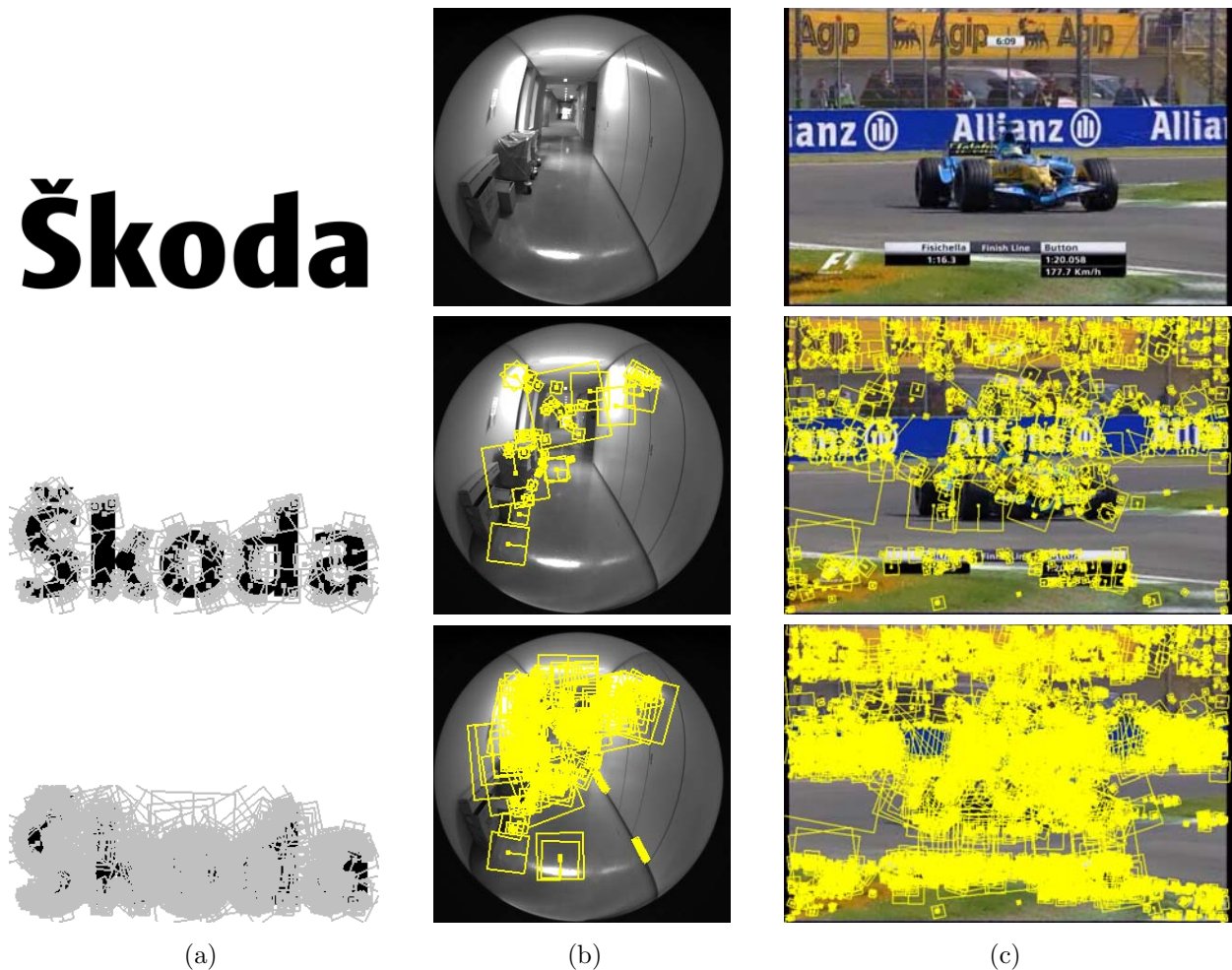


図 3: 局所領域の設定結果．図中の正方形が局所領域を表す．(a) 自動車メーカーŠkoda[16]のロゴマーク．(b) 中心窩レンズ [17]により撮像された廊下のシーン．(c) F1の放送映像 [18]．上段：原画像，中段：特徴点抽出による局所領域，下段：密なエッジサンプリングと特徴点抽出の併用による局所領域．

4(b),(c)では，そもそもエントロピーが4未満となる局所領域は出現していない．

密なエッジサンプリングでのエントロピーの分布を見ると，全画像において，値が4以上の局所領域数が大幅に増加している．パッチには対応付けに有効な輝度変化が存在していると言えることから，密なエッジサンプリングにより，特徴点抽出では得られなかった対応付けに有効な情報が画像から引き出せると言える．

次節では，密なエッジサンプリングの対応付けにおける有用性を，GPUによる実装を通じた実験により確認する．

## 5 GPUによる実装と実験結果

前節で示したように，密なエッジサンプリングにより局所領域数が増加する．これに伴い，記述子を求めるための計算量も増加する．この計算量の増加に対処するために，本研究ではGPUを用いた並列処理により図2に示すアルゴリズムを実装した．Parallel-Forベースの

方法により，画像処理フィルタでは注目画素の近傍領域を単位とし，また，記述子の計算では局所領域を単位として並列化を図った．開発環境として，NVIDIA社のCUDA[19]を使用した．計算に用いたGPUは，NVIDIA GeForce GTX 280である．また，比較対象のCPUとして，Intel Quadcore Xeon (3.16GHz/12MBL2)を用いた．OSはFedora8である．この実装の一部の詳細については，文献 [13, 20]を参照されたい．

表1に，図3の結果を得るためのGPUとCPUによる計算時間，および，その比を示す．GPUにより，CPUと比較して，特徴数が少ない場合には約9倍，多い場合には約18倍の高速化が達成できた．この結果から，密なエッジサンプリングによる計算量の増大に対応するために，GPUによる実装は有効な手段と考えられる．

図5に，得られた特徴量を用いた対応付け結果の一例を示す．図5(a),(b)は，放送映像 [18, 21]でのロゴの検出結果である．記述子間のユークリッド距離を用いた最近傍法により得られた対応点から，モデル画像とシー

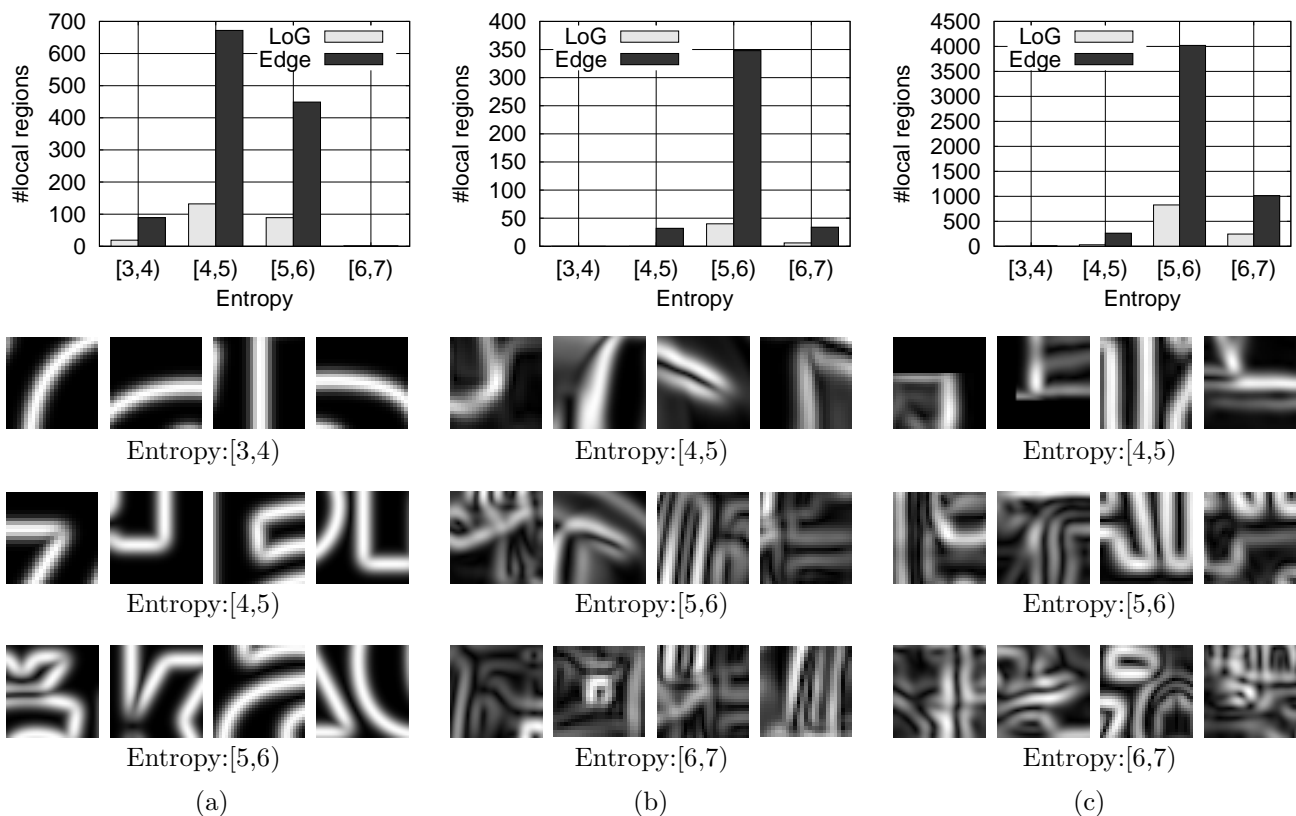


図 4: 図 3 の各画像における, 記述子のエントロピーによる局所領域の評価. この図の (a),(b),(c) は, それぞれ, 図 3(a),(b),(c) に対応している. 上段: 各エントロピーの値の範囲における局所領域数. "LoG" は特徴点抽出による方法, "Edge" は密なエッジサンプリングによる方法を示す. 第 2, 第 3, 第 4 段: 局所領域内部のエッジを表すパッチの例. 各段が各エントロピーの値の範囲に対応する. スケールによってパッチの大きさは変化するが, ここでは正規化して表示している.

ン画像内のロゴ間の射影変換行列を求め, ロゴの位置を表示している [22]. 上段と下段の画像は, 密なエッジサンプリングの導入前後の結果を表す. この結果から, 密なエッジサンプリングにより, 隠れに対してより頑健な対応付けが実現されていることがわかる. 図 5(c) は, 中心窩レンズにより得られた, 前進運動前後の画像の対応付け結果である. 対応付けは最近傍法により行い, 最近傍と 2 番目に近いデータの距離の比が 0.7 以下の対応点のみを表示している. 密なエッジサンプリング導入前後の左右の画像を比較すると, テクスチャの乏しいシーンでも明らかに対応点数が増加している. 対応点が多いことは, 特徴の消失や再出現がある状況でロボットの自己位置を推定する場合に, 処理の頑健性を向上させる一要因となるため, このような結果は望ましい. 以上の結果から, 密なエッジサンプリングの対応付けにおける有用性を確認できた.

## 6 まとめ

本論文では, 密なエッジサンプリングによる局所不変特徴量抽出について述べた. スケール空間において, 空間  $3 \times 3$  近傍でのエッジの極大点を全てサンプリングし, それらを中心として局所領域を設定する方法を

提案した. 提案方法により, 対応付けに有効な情報を有する局所領域数が大幅に増加することを, 記述子のエントロピーを用いて定量的に明らかにした. GPU による実装を通じた実験により, 密なエッジサンプリングが頑健な対応付けの実現に有用なことを示した.

今後は, 提案方法の応用を図ると共に, 対応付けの高速化を含めた研究を進める予定である.

## 謝辞

中心窩レンズにより撮像した画像を提供頂いた, 産業技術総合研究所知能システム研究部門の喜多伸之主任研究員に感謝致します. 本研究の一部は, 科学研究費補助金, 課題番号 18500145 の助成の下で行われた.

## 参考文献

- [1] C. Schmid and R. Mohr. Local greyvalue invariants for image retrieval. *IEEE Trans. PAMI*, Vol. 19, No. 5, pp. 530–535, 1997.
- [2] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vis.*, Vol. 60, No. 2, pp. 91–110, 2004.
- [3] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. J. Comp. Vis.*, Vol. 65, No. 1/2, pp. 43–72, 2005.

表 1: 局所不変特徴量抽出に必要な計算時間．単位は [ms]．OS の非リアルタイム性を考慮し，100 回の処理の平均値を示す．

Image	skoda			hallway			F1		
Image Size	356×132			320×320			720×480		
Processor/Time ratio	GPU	CPU	Ratio	GPU	CPU	Ratio	GPU	CPU	Ratio
Time	36.806	597.667	16.2	47.060	441.740	9.39	150.004	2788.497	18.6
#feature	1452			460			6405		

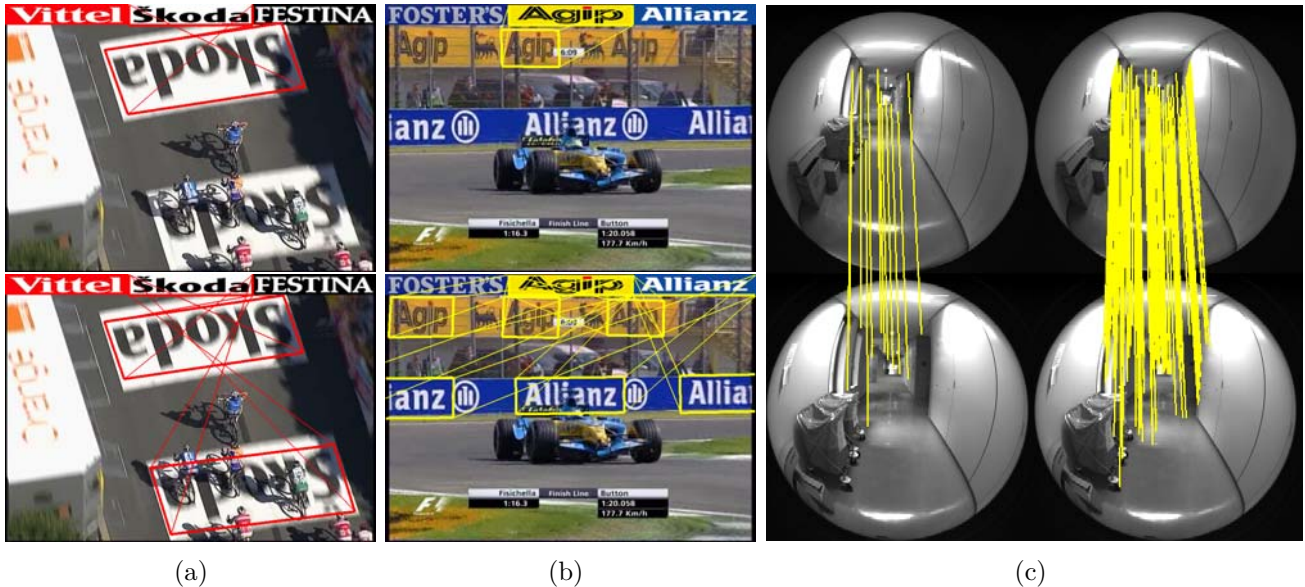


図 5: 画像の対応付け結果．(a),(b) 放送映像 [18, 21] でのロゴの検出結果．(c) 中心窩レンズ [17] により撮像された画像の対応付け結果．上段，左：特徴点抽出による結果．下段，右：密なエッジサンプリングを併用した結果．密なエッジサンプリングにより，隠れに対しより頑健な対応付けがなされている．また，テクスチャの乏しいシーンでも，より多くの対応点が得られた．

- [4] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. PAMI*, Vol. 27, No. 10, pp. 1615–1630, 2005.
- [5] T. Lindeberg. Feature detection with automatic scale selection. *Int. J. Comp. Vis.*, Vol. 30, No. 2, pp. 79–116, 1998.
- [6] C. Ssurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Proc. Workshop on Statistical Learning in Computer Vision*, pp. 1–22, 2004.
- [7] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Proc. Int. Conf. Comp. Vis. Patt. Recog.*, Vol. 2, pp. 524–531, 2005.
- [8] X. Ma and W. E. Grimson. Edge-based rich representation for vehicle classification. In *Proc. Int. Conf. Comp. Vis.*, Vol. 2, pp. 1185–1192, 2005.
- [9] K. Mikolajczyk, B. Leibe, and B. Schiele. Multiple object class detection with a generative model. In *Proc. Int. Conf. Comp. Vis. Patt. Recog.*, Vol. 1, pp. 26–36, 2006.
- [10] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. In *Proc. European Conf. Comp. Vis.*, pp. 490–503, 2006.
- [11] 柳井啓司. 一般物体認識の現状と今後. *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. 48, No. SIG 16, pp. 1–24, 2007.
- [12] S. Ando. Consistent gradient operators. *IEEE Trans. PAMI*, Vol. 22, No. 3, pp. 252–265, 2000.
- [13] 市村直幸. 近似 LoG フィルタを用いた局所不変特徴量の抽出 – GPU による実装 –. *情報研報*, No. 2008-CVIM-165, pp. 243–250, 2008.
- [14] M. Trajkovic and M. Hedley. Fast corner detection. *Image and Vision Computing*, Vol. 16, pp. 75–87, 1998.
- [15] 藤吉弘巨. Gradient ベースの特徴抽出 – SIFT と HOG –. *情報研報*, No. 2007-CVIM-160, pp. 211–224, 2007.
- [16] Škoda: <http://new.skoda-auto.com/>.
- [17] 若宮孝一, 千賀達也, 伊三木一皇, 山村則夫, 潮嘉次郎, 喜多伸之. 高性能中心窩レンズの開発. *ビジョン技術の実用化ワークショップ予稿集*, 2005.
- [18] フジテレビ 721 において放送された映像を使用している .
- [19] CUDA Zone: [http://www.nvidia.co.jp/object/cuda\\_home\\_jp.html](http://www.nvidia.co.jp/object/cuda_home_jp.html).
- [20] 市村直幸. 局所不変特徴量に基づく複数広告看板の認識 – GPU による特徴量抽出 –. *動的画像処理実用化ワークショップ講演論文集*, pp. 59–64, 2009.
- [21] J SPORTS において放送された映像を使用している .
- [22] N. Ichimura. Recognizing multiple billboard advertisements in videos. In *Proc. Pacific-Rim Symp. on Image and Video Technology (PSIVT)*, pp. 463–473, 2006.