

# 知っておきたいキーワード

## 局所画像特徴量

～SIFT, HOGを題材に～

庄野 逸<sup>†</sup>

<sup>†</sup>電気通信大学 大学院情報理工学研究所

"Local Features in Image; Using HOG and SIFT" by Hayaru Shouno (Graduate School of Informatics and Engineering, University of Electro-Communications, Tokyo)

キーワード：局所画像特徴量, HOG, SIFT

### 局所画像特徴量とは

デジタル画像を取り扱う場合、処理目的にもよりますが、画像をどのように表現するかということが問題になる場合があります。ここでは、パターン認識や位置合わせなどに用いられるHOG (Histograms of Oriented Gradients) とSIFT (Scaled Invariance Feature Transform) と呼ばれる特徴量について考えてみます。これらの特徴量は2000年位から提案されはじめ、現在においてもさまざまな拡張が提案されていますが、写真の中に写っているものを機械に判別させたいなどの問題を、“とりあえずやってみる”という立ち位置で研究・開発を始めるのに、コストパフォーマンスが良いと考えられています。

デジタル画像はピクセル値の集合なので、画像内に含まれる物体を探したり、識別したりするには、適切な表現が必要です。この画像表現を得ることを特徴抽出などと呼びます。例えば、ピクセル値をそのまま並べた数値ベクトルを特徴と捉えることができるし、画像全体に含まれるピクセル値の平均値をもとめ一つの値を、画像の特徴と

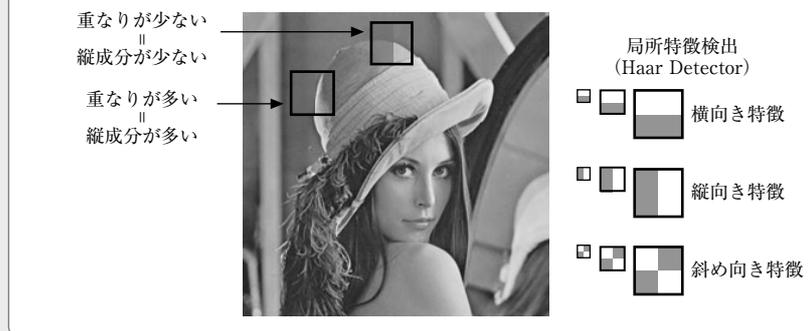


図1 局所特徴量の考え方

表1 局所特徴量の用途など

	照明の影響	回転の影響	拡大縮小の影響	用途
Haar	○	×	△	顔などの特定の形状
HOG	○	×	×	大まかな形状を把握したいとき
SIFT	○	○	○	細かい特徴の把握

することもできます。このように1ピクセルや画像全体といった極端な領域からではなく、画像の一部の領域から計算できる抽出できる量を局所特徴量と呼びます。

画像全体を一度に眺めるのではなく、小さな(エッジ成分のような)特徴で分類しようという試みは、画像をWavelet形式で表現する考え方に似ています。図1は、Lennaという画像を用いて、HaarWaveletと呼ばれる特徴

で表現することを表した図です。Haar特徴は画像から、縦成分、横成分、斜め成分を取出すために用いられます。図ではLenna画像に対し、縦成分のHaar特徴が、何処に、どのくらい含まれるかを模式的に表しており、帽子のエッジが縦にたった場所では、Haar縦成分の特徴とよく重なっているため、縦成分が大きな値をもちます。逆に帽子の上の部分では、横向きエッジが強く、Haar縦成分の

特徴と重なる部分が少ないため縦成分の値は小さくなります。

物体識別のような問題を考えた時に満たしているのが望ましい条件は、①環境の変化(環境光の違いなど)に対

応できるもの、②物体が画像内の移動に関して対応できるもの、③物体の拡大・縮小などに対応できるもの、などが考えられます。このような特徴量としてHOGやSIFTが使われていま

す。HOGやSIFTの比較的詳しい説明は、藤吉<sup>1)</sup>などによって詳しく述べられているので、ここでは、そのエッセンスの部分述べたいと思います。

### HOG

HOGは物体検出のための局所特徴量として提案されており<sup>2)</sup>、非常にポピュラーな特徴量として認知されています。HOGは、その名前のとおり局所的な画像勾配をヒストグラム化することで、画像を表現する特徴量です。計算アルゴリズムとしては、

- (1) 輝度勾配の算出
- (2) 局所ヒストグラム化
- (3) 局所ブロックによる正規化

からなります。図2はHOGの基本的な考え方を表現したものです。画像を、適当な大きさの“セル”と呼ばれる領域に分割し、この領域内から特徴量を求めます。また、セルの近傍のセル領域は“ブロック”と呼ばれます。HOGでは画像の勾配を求め、セル内部の勾

配の方向をヒストグラムで表現します。すなわち、セル内部でのエッジ成分がどちらの方向を向いているかを求めていることとなります。さらに、ブロック内でヒストグラムの数値を正規化することで、あるセルに対するHOG特徴量が得られます。図では、Lenna画像を16×16の大きさのセルに分割したHOGを考えており、ちょうど右眼に当たるところのセルを中心としたブロックを拡大表示しています。ブロック中央の右眼に該当するセルでは、さまざまな方向のエッジ成分を持っていることがわかりますが、中央のセルの左斜め上のセルでは、帽子にかかる成分から $\pi/4$ の勾配強度が非常に強く出ていることがわかります。一つのセルが持つ特徴の次元数は、ヒストグラム化した際の方向数と、ブロックの

大きさに依存します。図では、9方向のヒストグラムと、3×3の大きさをブロックとしていますので、一つのセルが持つ特徴量は $9 \times (3 \times 3) = 81$ 次元となり、画像全体の大きさは $256 \times 256$ なので、画像全体では、セルの個数は $16 \times 16$ 個で、近傍がない画像端の部分のセルを考慮に入れなければ、 $(16 - 2)^2 \times 81 = 15876$ 次元のベクトルで表すこととなります。

HOGは、写真中の対象物体の大きさ変化や回転の変化などには弱いですが、おおよそ、大きさが定まっている物体や、対象が一定の枠内に写っている画像等では、照明の変化などに強い特徴と考えられており、DalalらはHOGを人物検出の手法として用いて良好な成果を得ています<sup>2)</sup>。

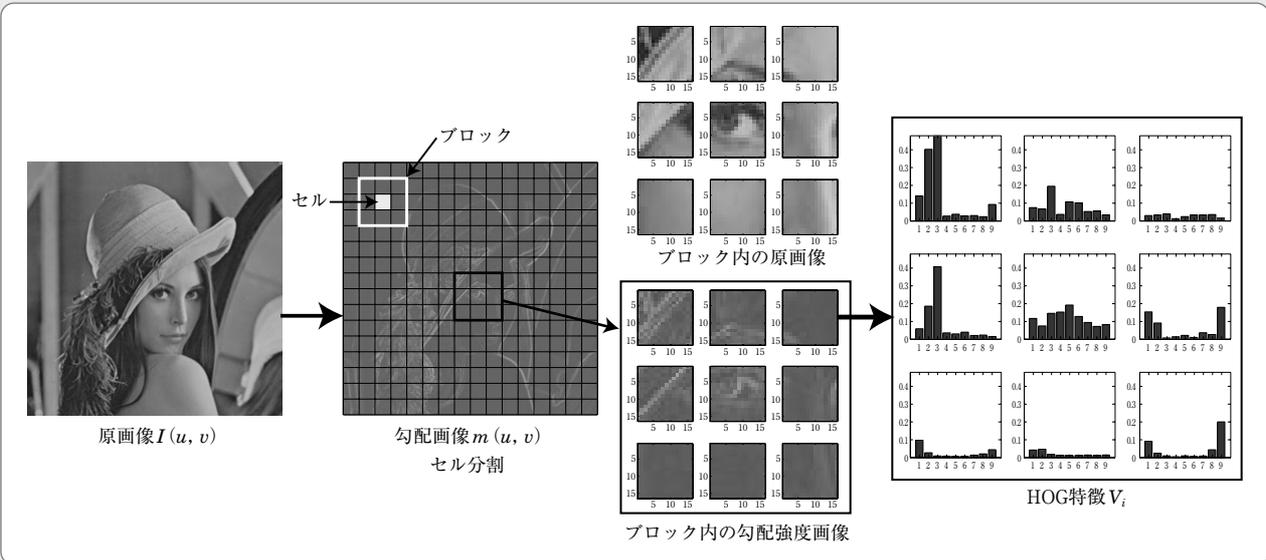


図2 HOG特徴量の計算手法

### SIFT

SIFTも、提唱されてすでに10年程経つ技術で<sup>3)</sup>、現実的な特徴抽出技術として広く受け入れられています(ただし特許が取得されているので利用には注意が必要です)。SIFTの処理は、

大きく分けて特徴点の検出と特徴量の記述の2段階からなります。特徴点の検出では、スケールの異なる平滑化画像の差分から画像特徴と思われる点(キーポイント)を決定します。その上で、各キーポイントの周りで周辺の画像の勾配情報を用いることで情報を

記述します。

図3は、SIFT記述子を取得する流れを表しています。まず、原画像をさまざまなスケールで平滑化を行います。図ではGaussianフィルタの広がり $\sigma$ をスケールパラメータとして平滑化のスケールを上から下へ徐々に

大きくして行きます。もっとも大きい平滑化領域では、非常にぼやけた画像になっていることがわかります。

次に、各スケール間で差分 (Difference of Gaussian: DoG) を取ることで、画像の変化 (物体と背景の境界など) が、どこに現れているかを算出します。この時、さまざまなスケールで計算するのが重要で、どのくらいの大きさの領域で変化が見えるのかを算出することが重要となります。

この変化が極大となる点が、SIFTの特徴点 (キーポイント) 候補となります。この点を探索するために、DoG画像を並べ、画像の  $(u, v)$  軸だけでなく  $\sigma_l$  方向を含めた空間で極値探索を行います。したがって、キーポ

イントが持つ情報としては、画像上の座標値だけでなく、どの位の大きさにわたって変化しているかというスケール情報を得ることができます。

図3中央のLenna画像に、SIFT特徴点をスケール情報を加えて表示しています。図中の円の半径がスケール情報を表し、どのくらいのスケールで特徴が存在しているのかがわかります。

SIFT特徴は、このキーポイントの周りでの画像勾配を記述することで得られます。まず、キーポイントのスケール円内部の画像勾配に重みを付けて平均をとることで、キーポイントがどちらの方向を向いているかを算出します。図中では、円の中心から伸びる線分がキーポイント方向を示していま

す。キーポイントの方向を計算することは、取出した特徴を特定方向に揃えて考えることができることを意味します。すなわち画像が回転した場合でも、特徴の形状に変化がない限り同じものと捉えることができるわけです。

キーポイントの方向を決定できれば、この方向にそって、スケールに応じた格子を考え、各格子内部に含まれる勾配をHOGの時と同様にヒストグラムとして考えることで、SIFT特徴量を得ることができます。

通常、SIFT特徴は  $4 \times 4$  の格子を考え、その格子内部の勾配を8方向のヒストグラムとして表現するので、各キーポイントにつき  $(4 \times 4) \times 8 = 128$  次元の特徴量が得られます。

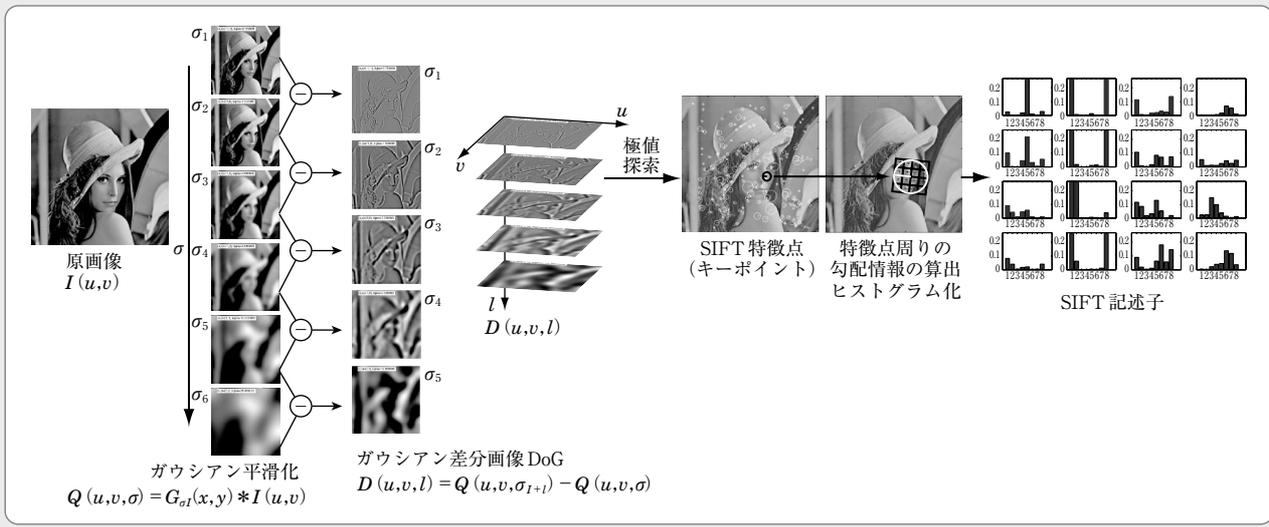


図3 SIFT特徴量の計算手法

## むすび

局所特徴量は比較的単純な計算手続きで得ることができます。HOGの実装は、OpenCVなどに移植されて

おり、計算機があれば比較的簡単に試すことができます<sup>4)</sup>。一方のSIFTは、Loweによって書かれたでもコードがWebサイトから入手できます<sup>5)</sup>。パテントフリーな局所特徴量として

SURFなどがあります<sup>6)</sup>。これにより、読者の皆さんが、なにかに使ってみようという気になっていただければ幸いです。

(2012年12月11日受付)

## 参考文献

- 1) 藤吉弘直: "Gradientベースの特徴抽出: SIFTとHOG", 情処研報, CVIM-2007-87 (2007)
- 2) N. Dalal and B. Triggs: "Histograms of oriented gradients for human detection", International Conference on Computer Vision & Pattern Recognition, 2, pp.886-893 (June 2005)
- 3) D.G. Lowe: "Distinctive image features from scaleinvariant key points", 60, pp.91-110 (2004)
- 4) <http://opencv.willowgarage.com/>
- 5) D.G. Lowe: "Demo software", Sift keypoint detector, <http://www.cs.ubc.ca/~lowe/keypoints/>
- 6) <http://www.vision.ee.ethz.ch/~surf/>



しょうの 庄野 逸 はやる 1994年、大阪大学大学院基礎工学研究科博士前期課程修了。1994年、大阪大学基礎工学部助手。2000年、奈良女子大学助手。2001年、山口大学工学部助教授。2008年より、電気通信大学大学院情報理工学研究科准教授。視覚神経回路モデル、ニューラルネットワークモデル、医療画像処理などの研究に従事。博士(工学)。