



画像処理における SIFT 局所特徴の応用

神戸大学 経済経営研究所
助教 陳 金輝

我々人間は物を見て音を聴き、脳で周囲の情報を処理したり、周囲の環境にあるものを認識したり、状況を理解・処理することができる。コンピュータなどの計算機器に同様のことを行わせ、機械自身搭載の知能アルゴリズムにより、周囲の状況を計算し、人間のような知識を学習しながら行動する機能を果たさせることを目的としたのが、人工知能という分野の研究である。「アルファ碁」を初めとした人工知能実用化の大きな成功に繋がった人工知能研究が近年注目されてきた。人工知能、特に知能メディアにおける主な研究課題は、特徴付け (Feature) と機械学習 (若しくは統計的モデリング) のアルゴリズムである。今回は「特徴付け」に注目する。

SIFT (Scale-Invariant Feature Transform) [1] [2]、SURF (Speed Up Robust Feature) [3]、と輝度勾配方向ヒストグラム HOG (Histogram of Oriented Gradient) [4] は、機械学習による画像検索・認識・分類タスクにおいて、最もよく用いられている優秀な局所特徴である。表現力が最も優秀な SIFT は、計算量も一番多いため、計算機器の高計算性能が必要とされる。SURF 特徴の抽出速度は最も速いが表現力が最も悪い。そのため、半導体・機器の搭載カメラを初めとした計算速度の遅い器械に活用されることが多い。上述の二つの画像特徴抽出アルゴリズムは、特徴点を中心にして、特徴点量 (特徴の記述子) の抽出を行うため、画像の細かい特徴情報を大事にするタスク (*e.g.*, 顔認識・文字認識など) に、重要な手法とされている。現在広く使用されている HOG の多くは、画像上で空間的に一様に正方領域のセルを配置し、セルをオーバーラッピングさせた局所正方領域上で、正規化処理が行われている。このため、HOG は特徴点ではなくて、特徴ブロックに注目して特徴量が抽出される。従って、HOG はブロックオブジェクトの認識・検出タスクに活用されることが多い。特に車載カメラに基づく通行人・車検出の応用に非常に役立っている。

CPU やコンピュータストレージ技術の発展とともに器械の計算能力と情報処理量が大幅に改善されてきた。この点から、今後、SIFT はさらに重視されてくると考えられる。

今月のニュースレターコラムではこの SIFT のアルゴリズムを、画像特徴 (記述) 抽出手法の代表として紹介する。SIFT は、回転・スケール変化・照明変化にロバスト (robust: 頑健) な局所特徴量であり、2段階の処理からなり、Difference-of-Gaussian (DoG) による特徴点とスケールの検出と特徴点の特徴量記述を行う [1] [2]。

1. Difference-of-Gaussian 処理

この処理により、画像中の特徴点の候補を検出する。まず、スケール (Scale) の異なるガウス関数 $G(x, y, \sigma)$ と入力画像 $I(x, y)$ を畳み込んだ平滑化画像 $L(x, y, \sigma)$ を用意する。入力画像とガウス関数を畳み込みの計算が行われたと、得られた畳み込み画像では、離散のピクセルにガウス分布の属性を加えて、個体ピクセルの差異ノイズを最大程度に抑えられるため、畳み込んだ平滑化画像に基づく抽出された特徴量は、生のピクセルにより得られた特徴量と比べて、統計学での予測は信用性が高くなる。

次に、得られた各平滑化画像の差分をとり、その差分画像である DoG 画像 $D(x, y, \sigma)$ を次式で求める：

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \otimes I(x, y) \quad (1)$$

k は増加率であり、 k 倍ずつ大きくした異なるスケール間で DoG 処理を行うことで、Fig. 1 に示すような複数の DoG 画像が得られる。この DoG 処理により、特徴点候補の検出とスケール探索を同時に行うことができる。

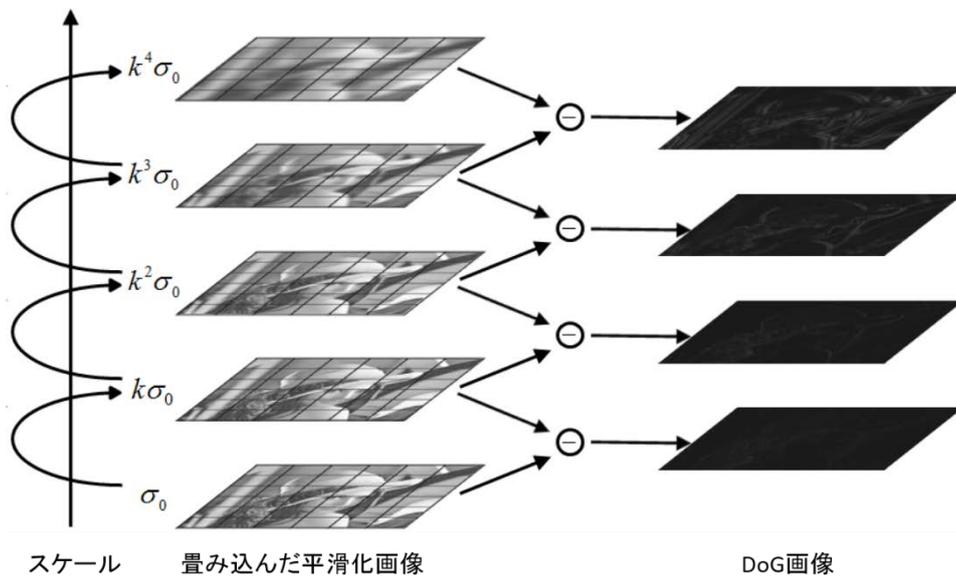


Fig. 1 . 異なるスケール空間の計算結果

2. スケールと特徴点の検出

得られた DoG 画像から極値を検出し、特徴点とスケールを決定する。Fig. 2 のように DoG 画像を 3 枚一組とし、注目画素の DoG 値と、上下を含む隣接する画素とその周りの 26 近傍を比較し極値を求める。検出された極値が特徴点となり、そのスケールが特徴点のスケールとなる。SIFT は特徴を最も含むスケールを自動的に決定するため、空間的に同範囲の領域から特徴量を記述する。よって、拡大・縮小に不変な特徴量となる。

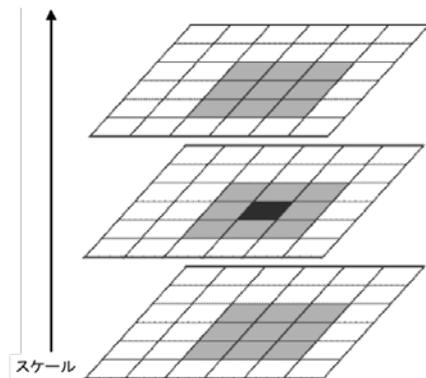


Fig. 2 スケール空間にて極値を検出する

3. オリエンテーションの算出

検出した各特徴点について、特徴点における方向を表すオリエンテーション (Orientation) を求める。まず、特徴点が発見された平滑化画像の各画素の勾配強度 $m(x, y)$ とその勾配方向 $\theta(x, y)$ を求める。得られた勾配強度 $m(x, y)$ と勾配方向 $\theta(x, y)$ を用いて、Fig. 3 に示すような重み付き勾配方向ヒストグラムを以下の式により作成する：

$$h_{\theta'} = \sum_x \sum_y \omega(x, y) \cdot \delta[\theta', \theta(x, y)] \quad (2)$$

$$\omega(x, y) = G(x, y, \sigma) \cdot m(x, y) \quad (3)$$

ここで、 h_{θ} は全方向を 36 方向に量子化したヒストグラムである。 $w(x, y)$ はある局所領域の画素 (x, y) の重みであり、特徴点のスケールサイズのガウス窓 $G(x, y, \sigma)$ と勾配強度より求められる。このガウス窓による重みを付けることで、特徴点に近い特徴量がより強く反映される。 δ は Kronecker のデルタ関数で、勾配方向 $\theta(x, y)$ が量子化した方向 θ' に含まれるとき 1 となる。この 36 方向のヒストグラムの最大値のオリエンテーションをその特徴点の代表オリエンテーションとして割り当てる。

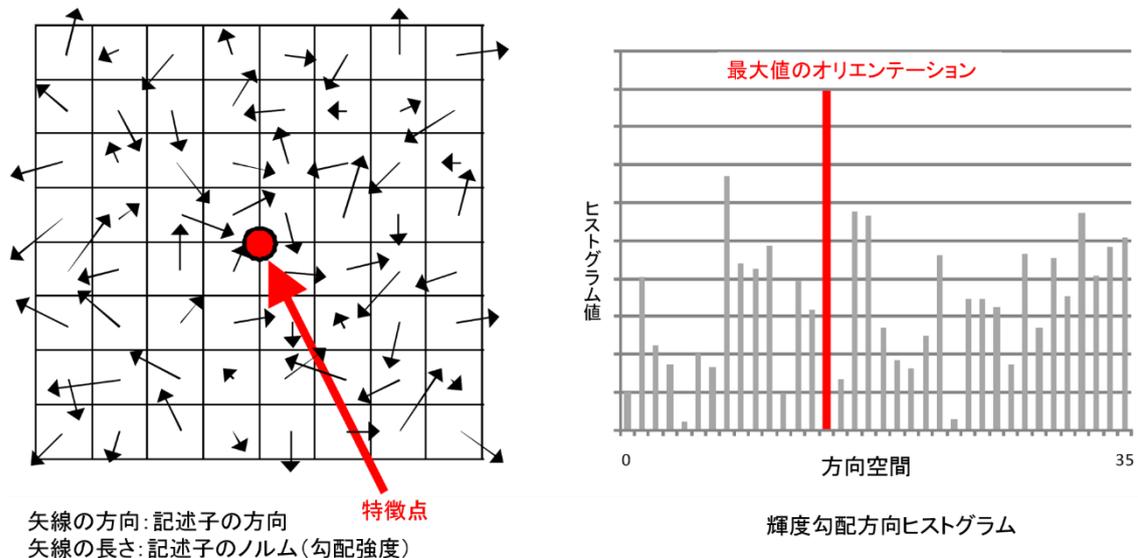


Fig. 3 重み付き勾配方向ヒストグラムを用いた SIF 記述子及び特徴量主方向の計算

4. 特徴量の記述

特徴点の周辺領域を 3 項で求めたオリエンテーション方向に回転する。このように、オリエンテーション方向を座標軸とした領域で特徴量を記述するため、回転への不変性が得られる。そして、この周辺領域を一辺 4 ブロックの計 16 ブロックに分割し、ブロック毎に 8 方向の勾配方向ヒストグラムを作成する。これにより、Fig. 4 のような 4 ブロック × 4 ブロック × 8 方向の 128 次元の特徴量が得られる。また、128 次元の各特徴ベクトルの長さはベクトルの総和で正規化するので、特徴点は照明変化に対してロバストになる。

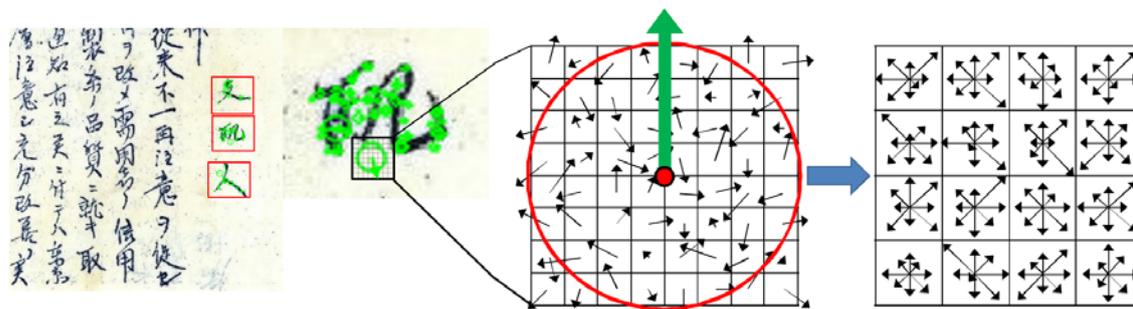


Fig. 4 鐘紡史料画像から抽出された SIFT 画像記述子

現在、最先端の画像分類システム（ディープラーニング系を除く）は大きく分けて3つの工程で構成される。SIFT (Scale-Invariant Feature Transform) による特徴抽出、BoVW (Bag of Features・統計的言語処理である BoVW (Bag of View Word) [5]を画像分類に適用した手法である) によるコード化、SPM (Spatial Pyramid Matching) などの手法によるプーリング処理である。こうして得られた各画像の特徴を並べて一つの特徴ベクトル（全局特徴）とする。その特徴ベクトルを用いて学習データより分類器を学習する。テストデータに対しては、学習した分類器を用いてクラス分類して認識・分類結果を出力する。検索の場合は、全局特徴をサポートベクターマシーン (SVM) に基づいており、画像クラスリストの統計的モデリングの学習を行う。クエリ (Query・検索画像) に対して、学習したモデリングを用いて、画像集合中の画像を計算して類似さにより画像リストを生成し、リストのトップ指定された数の画像を出力する。

但し、最近流行っているディープラーニングの処理の流れや理論は上述の手法と異なっている。そのため「アルファ碁」を代表としたディープラーニングの理論と技術については、今後のニュースレターコラムもしくはディスカッションペーパーにてまた紹介する。

参考文献

1. Lowe, D.G.: “Distinctive Image Features from Scale-Invariant Keypoints”, In: *Proc. Int. J. Comput. Vis. (IJCV)*, vol.60, pp. 91 - 110 (2004).
2. Lowe, D.: “Object Recognition from Local Scale-invariant Features”, In: *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, pp. 1150 - 1157 (1999).
3. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: “Speeded-Up Robust Features (SURF)”. *Comput. Vis. Image Underst. (CVIU)* vol.110, pp. 346 - 359 (2008)
4. Dalal, N., Triggs, B.: “Histograms of Oriented Gradients for Human Detection”, In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. vol. 1, 886 - 893 (2005).
5. Christopher D. M., Hinrich S.: “Foundation of Statistical Natural Language Processing”, *The MIT Press*, (1999).