H-024

BoVW を用いた特定物体認識における投票関数に関する一考察

On Voting functions for BoVW-Based Specific Object Recognition

内田 祐介 †　高木 幸一 †　酒澤 茂之 †
Yusuke Uchida　Koichi Takagi　Shigeyuki Sakazawa

1. はじめに

画像中の物体を，見えの変化に対して頑健に認識できることから，局所特徴量を用いた特定物体認識手法が注目を集めている．特に，大規模データベースを対象とした場合には，検索効率から Bag-of-Visual Words (BoVW) と転置インデックスを組み合わせた手法が広く用いられている．また，単純な BoVW に基づくマッチングの後に，局所特徴量間の距離を算出し，算出された距離に基づいてマッチングを高精度化を実現する手法も提案されている．本稿では，局所特徴量間の距離に基づいて高精度化を行う際に，特徴量の密度を考慮することで更に精度を改善する手法を提案し，公開データセットで評価を行う．

2. BoVW を用いた特定物体認識の高精度化

本節では BoVW を用いた特定物体認識手法および改良手法を概説する．BoVW を用いた特定物体認識手法 [1] では，画像から SIFT 等の局所特徴量を抽出し，予め定義される Visual Words (VWs) を用いてベクトル量子化を行い，各 VW の頻度に基づいて画像を表現する．画像間の類似度は，上記頻度のコサイン類似度で定義され，転置インデックスを用いることで効率的な類似度の算出が可能となる．

上記の手法では，同一の VW に割り当てられた局所特徴量同士が全てマッチしたものとして扱い，逆にそれ以外ではマッチしないものとして扱うため，VW の数に関して下記のようなトレードオフが存在する．すなわち，VW の数が小さい場合には無関係な特徴量同士がマッチすることで識別性が低下する．逆に，VW の数が大きい場合には，同一の局所特徴量がノイズ等によって変化し，マッチしなくなることで再現性が低下する．このトレードオフを改善するため，VW レベルでマッチした局所特徴量間の距離を算出し，その距離に基づいて投票する特徴量の数を制限することで，高精度化を図る手法が提案されている [2, 3, 4]．Hamming Embedding (HE)[2] では，局所特徴量をランダム直交射影変換および閾値処理によりバイナリ符号に変換し，ハミング距離を用いて特徴量間の距離を近似する．一方，[3, 4] では，直積量子化 (Product Quantization, PQ) を用いて特徴量間のユークリッド距離の近似値を算出する．その後，[2, 4] では，算出した距離が閾値以下の特徴量に対して投票を行う一方，[3] では，算出した距離で特徴量を昇順にソートした際の上位 k 件に対し投票を行うことを提案している．

† 株式会社 KDDI 研究所 メディアソリューショングループ

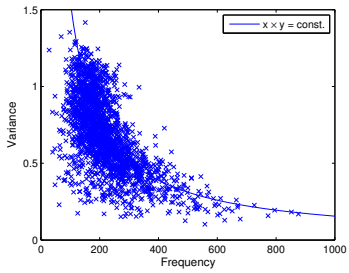
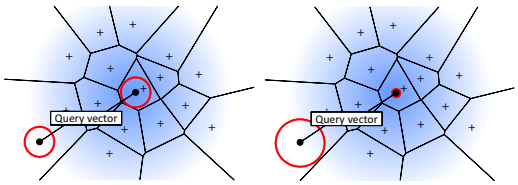


図 1　400 万個の訓練ベクトルから 20K 個の VW を作成した際の，VW に属する特徴量数（横軸）と VW の代表ベクトルからの平均二乗誤差（縦軸）を，ランダムに選択した 2K 個の VW についてプロットした

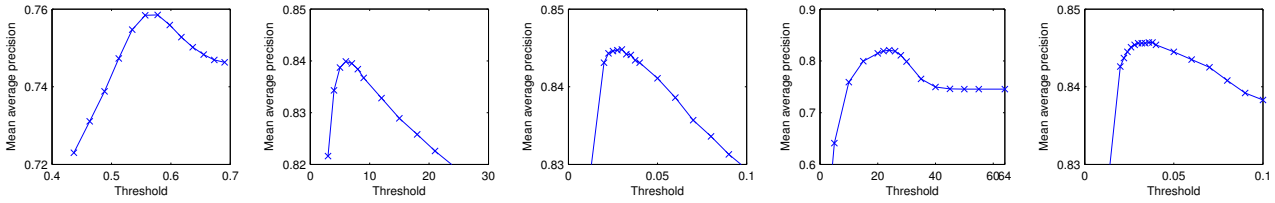


(a) 距離　　(b) k 近傍

図 2　距離および k 近傍を利用した場合のクエリベクトルにマッチする特徴ベクトルの範囲

3. 比率を用いた投票関数

2 節で概説した BoVW 改善手法には，下記のように，各 VW に属する特徴量の密度および各 VW のセルのサイズを考慮していない問題がある．図 1 に，各 VW に属する特徴量数と VW の代表ベクトルからの平均二乗誤差を示す．平均二乗誤差は大まかなセルの大きさに相当し，属する特徴量数が多い VW ほどセルが小さい傾向にあることが分かる．すなわち，VW 毎に特徴量の密度およびセルのサイズが大きく違うことが分かる．このとき，図 2 (a) に示すようにサイズの異なるセルに対して距離を閾値として投票を行うと，小さなセルではほとんどの特徴量が投票に利用される一方，大きなセルでは投票に利用される特徴量が少なくなってしまう問題が発生する．逆に，図 2 (b) に示すように k 近傍に投票を行った場合には，特徴量の密度が高い VW では距離に換算するとより近い特徴量にしか投票が行われず，ノイズ等による特徴量の変化に対応出来なくなる一方，特徴量の密度が低い VW では類似していない特徴量にも投票が行われる問題が発生する．上記の考察より，本稿では，VW レベルでマッチした特徴量を推定された距離でソートした際に，各 VW に属する特徴量の数に対して一定の比率の特徴量に対して投票を行うことを提案する．これは基準として距離を利用した場合と k 近傍を利用した場合の中間の効果があり，上記の問題が解消できると期待される．



(a) ユークリッド距離, th=0.58, MAP=0.756　(b) k 近傍, th=6, MAP=0.840　(c) セル内特徴数比, th=0.03, MAP=0.845　(d) ハミング距離, th=24, MAP=0.821　(e) セル内特徴数比, th=0.036, MAP=0.846

図 3　異なる指標を閾値を用いて投票を行った際の特定物体認識精度 (MAP)

4. 評価実験

4.1. 実験環境

画像データセットとして，本分野で広く利用されている Recognition Benchmark Images[*1] データセットを利用する．上記データセットは，2,550 個のオブジェクトを異なる 4 方向から撮影した画像，合計 10,200 枚から構成されている．各画像をクエリとし，オリジナルの画像を含む 4 枚の画像を正解として検索を行った際の Mean Average Precision (MAP) により性能評価を行う [2]．実装や実験条件は [2] を踏襲し，マッチングのスコアは VW の Inverse Document Frequency (IDF) の二乗で重み付けを行い，マッチング後の各画像のスコアは BoVW の L2 ノルムで正規化されることとする．また VW の数は 20K とし，PQ および HE ともに利用する符号長は 64bit とした．

4.2. 実験結果と考察

図 3 に，様々な基準を利用して投票を行った際の検索精度 (MAP) を示す．また，各基準において達成された最も良い精度 (MAP) とそのときの閾値の値を示す．(a) は PQ を用いて距離を推定し閾値以下の特徴量に対して投票を行った場合 [4]，(b) は PQ を用いて距離を推定し k 近傍に対して投票を行った場合 [3]，(d) は HE を用いて距離を推定し閾値以下の特徴量に対して投票を行った場合 [2] の精度を示す．(c) および (e) は，それぞれ PQ および HE を用いて距離を推定し，比率に基づいて投票を行った場合（提案手法）の精度を示す．図 3 より，PQ および HE どちらを用いて距離を推定した場合においても，VW に属する特徴量の数に対する比率に基づいて投票を行った場合に，最も高い認識精度を達成できていることが分かる．以下に個々の詳細な考察を示す．

(b) と (c) を比較すると，k 近傍に対して投票を行う手法は比較的良好な結果を示すことが分かる．これは，k 近傍を利用した場合には，特徴量の密度が高い VW では特徴量の再現性が比率を用いた投票と比較して低くなる一方，投票に利用される特徴量の数を比率に換算すると平均 1.7% 程度の特徴量に対してのみ投票を行っていることに相当し，無関係な特徴量への投票が少なく抑えられているためと予想される．

(a) と (d) を比較すると，近似最近傍探索手法として見た場合には PQ が HE と比較して優れている [3] にも関わらず，距離そのものを閾値として利用する場合には HE がより良好な結果を示している．これは，HE では特徴量を 0 と 1 が等確率かつ無相関に出現するバイナリ列に符号化しているため，ある特徴量からのハミング距離が閾値以内である特徴量の数は，VW に属する特徴量の数に対してほぼ一定の割合となり，比率を用いた投票と同様の効果が得られているためと考えられる．

(c) と (e) を比較すると，比率を用いた投票を行う場合には，距離推定に PQ および HE どちらを利用しても，達成される精度はほぼ同じであることが分かる．このことから，局所特徴量間の距離を用いて BoVW に基づく特定物体認識手法を高精度化する際には，最近傍探索手法の選択よりも投票方式の選択がより重要であると言える．

5. まとめ

本稿では，局所特徴量間の距離を用いて BoVW に基づく特定物体認識手法を高精度化する際に，各 VW に属する特徴量の数に対する比率を利用することで精度を改善する手法を提案し，公開データセットを用いた評価実験により提案手法の有効性を示した．今後は，LSH 等のハッシュ関数によって定義される VW における検証や距離学習との組み合わせについて検討を行う．

参考文献

[1] J. Sivic and A. Zissermane, “Video google: A text retrieval approach to object matching in videos,” in *Proc. of ICCV*, 2003, pp. 1470–1478.

[2] H. Jégou, M. Douze, and C. Schmid, “Improving bag-of-features for large scale image search,” *IJCV*, vol. 87, no. 3, pp. 316–336, 2010.

[3] H. Jégou, M. Douze, and C. Schmid, “Product quantization for nearest neighbor search,” *IEEE Trans. on PAMI*, vol. 33, no. 1, pp. 117–128, 2011.

[4] Y. Uchida, M. Agrawal, and S. Sakazawa, “Accurate content-based video copy detection with efficient feature indexing,” in *Proc. of ICMR*, 2011.

[*1] http://www.vis.uky.edu/~stewe/ukbench/