

代表色選択を色ヒストグラムに適用した一般物体認識

鈴木 慎人[†] 若原 徹[†]

法政大学 情報科学部[†]

1. はじめに

近年 SNS や写真共有サイトなどの Web サービスが誕生したことにより、インターネット上には日々、膨大な数の画像がアップロードされている。これらのほとんどはテキスト等のメタデータを持たないビッグデータであり、こういった画像を検索するためには CBIR(Content-Based Image Retrieval) が必要となる。CBIR は「画像検索」という研究分野で主に扱われてきたが、一方でこうした大規模データに対して意味的な分類を行うことを目指す「一般物体認識」という研究分野が近年注目を集めている。一般物体認識には、データベースが大規模かつ共通であることや、定量評価が可能といった特徴がある。

本論文では、Swain らが提案した画像検索手法である色ヒストグラム[1] に、代表色選択と GLVQ(Generalized Learning Vector Quantization)[2] を組み合わせた新しい手法を提案する。これを一般物体認識のタスクに適用し、評価実験を行った。

2. データセット

対象とするタスクは、”Places2 Challenge 2016”[3] におけるシーン分類タスクである。データセットは、画像データベース Places2[3] の中から、名詞によって定義された 365 個のカテゴリに属する画像を抽出したものであり、学習用画像の数によって 2 種類に分かれている。画像数は表 1 に示すとおりである。なおテスト用画像は共通で、カテゴリごとに 100 枚存在する。本研究では処理速度を鑑み、Standard において画像サイズが 256x256 に前処理されているものを用いた。

表1. データセットの学習用画像数.

名称	Standard	Challenge
全数	1,803,460	8,026,628
カテゴリごと	3,068~5,000	3,068~40,000

An Application of Representative Colors to Color Histogram in Generic Object Recognition

[†]Masato Suzuki, Faculty of CIS, Hosei University

[†]Toru Wakahara, Faculty of CIS, Hosei University

3. 提案手法

代表色を用いた画像検索については先行研究が存在する[4]。しかしこれらの研究では、用いる色を発見的に決定しているという問題があった。本論文で提案する手法では、メディアンカット法[5] を用いて代表色を決定し、GLVQ[2] を用いた参照ベクトルの学習によって認識を行う。提案手法の流れを図 1 に示す。

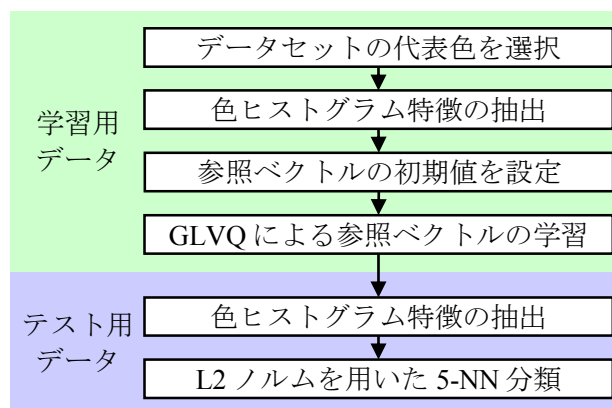


図1. 提案手法の流れ.

3.1. データセットの代表色の選択

まず色ヒストグラムのビンに対応する代表色を求める。一般的には均等量子化によって求めるが、ここでは全学習用データに対してメディアンカット法[5] を適用することで、画像の近似誤差をより小さくする代表色を選択する。

3.2. 色ヒストグラム特徴の抽出

代表色を求めた後、各学習用データに対して近似を行い、色ヒストグラムを求める。画像 $P = \{p_i\}_{i=1}^M$ に対し、 K 色の代表色 $Q = \{q_k\}_{k=1}^K$ を用いて近似を行った際の色ヒストグラム $H = \{h_k\}_{k=1}^K$ における要素 h_k は、以下の式で求められる：

$$h_k = \frac{1}{M} \sum_{i=1}^M g(k, i^*)$$

$$\text{where } i^* = \underset{1 \leq k \leq K}{\operatorname{argmin}} \|p_i - q_k\|^2 \quad (1)$$

$$g(x, y) = \begin{cases} 1 & : x = y \\ 0 & : \text{otherwise} \end{cases}$$

3.3. 参照ベクトルの作成

分類に用いる参照ベクトルを、カテゴリごとの学習用データに対して K-means クラスタリングを行い L 個ずつ生成する。その後、GLVQ を用いた学習を T 回を行い、カテゴリを代表する最適な値に修正する[2]。 L および T はあらかじめ指定する。

3.4. テスト用データの 5-NN 分類

最後に、参照ベクトルとの L_2 ノルムを用いて分類を行う。入力されたテスト用画像に対して属すると予想されるカテゴリを、その確率が高いほうから 5 個、 $l_j (j = 1, \dots, 5)$ として出力する。このとき、対象タスクにおけるアルゴリズムに対する評価指標は、以下のとおりである。

N 枚のテスト用画像に対し、Ground truth として画像ごとに 1 つのカテゴリ $g_i (i = 1, \dots, N)$ があらかじめ付与されており、全テスト用画像に対するエラー率 Err (Top-5 Error) は、

$$Err = \frac{1}{N} \sum_{i=1}^N \min_j d(g_i, l_j) \quad (2)$$

where $d(x, y) = \begin{cases} 0 & : x = y \\ 1 & : otherwise \end{cases}$

として定義される。

また、認識率 (Top-5 Accuracy) は、1 から Err を引くことにより求められる。本論文においても同様の指標に基づき、認識率を求める。

4. 評価実験

まず、パラメータ間での比較実験を行った。用いたパラメータおよび認識率を表 2 に示す。なお本論文では、すべての実験において RGB 色空間を用い、色数は $K=512$ とした。

表2. 実験結果.

No.	減色手法	分類手法	Top-5 Acc. (%)
(a)	Quantized	全数 k-NN	11.7
(b)	Median Cut	全数 k-NN	12.0
(c)	Quantized	$L=100, T=0$	9.6
(d)	Median Cut	$L=100, T=0$	9.8
(e)	Median Cut	$L=500, T=0$	10.8

表 2 において、“全数 k-NN” は全学習用データを用いた k-NN 分類を表し、 L および T はそれぞれ参照ベクトル数、学習回数を表し GLVQ による分類を表す。

次に、実験(d), (e) について、GLVQ による学習回数と認識率の関係を調べた。結果を図 2 に示す。

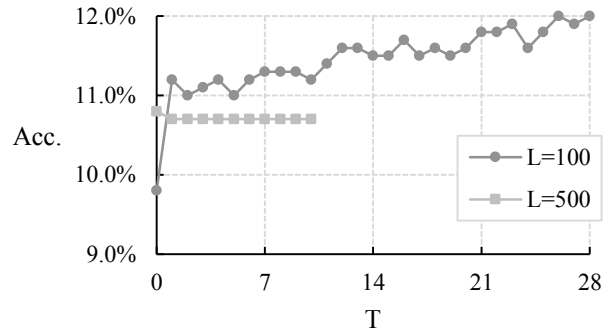


図2. 学習による認識率の推移.

5. 考察

表 2 の結果から、メディアンカット法を用いた方が、認識率が高くなることが示された。これは多くの予備実験でも同様であり、特に色数が少ないときに差が顕著であった。様々な場面において均等量子化の代わりに代表色選択を用いることで、性能向上が期待できると考える。

図 2 では、GLVQ の有効性が示されている。特に参照ベクトル数、すなわちテンプレート数が少ない場合に学習の効果が高く、 $L=100$ の場合には $T=26$ の学習回数で実験(b) と同じスコアとなった。これは、全数 k-NN と同じ精度で、より高速な分類が可能であることを意味する。

6. まとめ

本論文では、代表色選択を色ヒストグラムに適用した手法を提案し、一般物体認識のタスクを対象とした評価実験によって、代表色選択が均等量子化よりも有効であることを示した。

文献

- [1] M. J. Swain, and D. H. Ballard, “Color Indexing,” *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.
- [2] A. Sato and K. Yamada, “Generalized Learning Vector Quantization,” *Advances in Neural Information Processing Systems*, vol. 8, pp. 423-429, MIT Press, 1996.
- [3] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba and A. Oliva, “Places: An Image Database for Deep Scene Understanding,” *arXiv:1610.02055*, Oct. 2016.
- [4] V. Ljubovic, and H. Supic, “A Compact Color Descriptor for Image Retrieval,” *2013 XXIV International Symposium on Information, Communication and Automation Technologies*, Oct. 2013.
- [5] P. Heckbert, “Color image quantization for frame buffer display,” *ACM SIGGRAPH Computer Graphics*, vol. 16, no. 3, pp. 297-307, Jul. 1982.