

SIFT 特徴量のクラスタリングに基づく特定物体認識手法

大関 陽裕[†] 大野 将樹[‡] 獅々堀 正幹[‡]

[†]徳島大学 先端技術科学教育学部 システム創生工学専攻 知能情報システム工学コース

[‡]徳島大学 大学院ソシオテクノサイエンス研究部

〒770-8506 徳島県徳島市南常三島町 2-1

あらまし 近年、スマートフォンやデジタルカメラの普及に伴い、多くの画像データを高速かつ正確に処理することが一般的となってきた。本研究では、局所特徴量のクラスタリングを行う事で、大量の画像から特定の物体が写った画像を検索する手法を提案する。特定物体の検索には、通常 Bag of Features のような手法が一般的に用いられるが、対象の物体以外とマッチングしてしまう点が問題点とされる。そこで本手法では局所特徴点は物体ごとに集中して出現すると仮説を立て、局所特徴点にクラスタリングを施し、その結果を元に検索対象の画像を細かく分割することで認識精度を向上させる。約 50 万枚の画像を対象とした実験を行い、結果の考察と本手法の有用性を示す。

キーワード 画像処理, 画像検索, 特定物体認識, クラスタリング, 局所特徴量

1. はじめに

近年、スマートフォンやデジタルカメラの普及に伴い、画像データを取り扱うことが一般的となってきた。大規模な画像データを効率的に扱う方法は、画像認識として広く研究されており、大きく特定物体認識と一般物体認識に分かれる。特に特定物体認識は、人物認識や欠陥検出、画像ライブラリの構築、画像検索など様々な分野の改善につながることから、本研究では特定物体認識を取り上げ、改善に取り組む。

特定物体認識では、SIFT 特徴量[1]などの局所特徴量を用い、Bag-of-Features[2]のようなベクトル量子化手法を用いて認識をするのが一般的である[3]。しかし、対象となる物体以外の特徴点とマッチングすることが問題となる。そこで本研究では、あらかじめ局所特徴量を出現位置によってクラスタリングし、局所特徴量の出現位置の分布に基づいて検索対象画像を分割し、分割画像毎に特徴点とのマッチングを行うことで、誤検出を減らす。

特定物体認識は、一般物体認識より容易であるとされているが、交通道路標識の認識[4]や機械用基板における物体認識[5]など、需要は様々であり、画像認識分野において幅広く研究されている。[3]においては、特徴点の位置に着目した手法を用いており、対応点と基準となる特徴点の位置関係を比較し、誤検出を減らす工夫がなされている。認識率も 89%と高く、実用化が期待されている。

2. 提案手法

提案手法の流れを図 1 に示す。動画から抽出したキーフレーム画像から特徴量を抽出し、データベースに登録する。クエリ画像からも特徴量を抽出し、データベ

ースに登録された特徴量との比較を行う。

提案手法では局所特徴量を画像から検出し、特定物体の写っている画像の検索を行う。その際に、画像ごとに類似度を定める必要があるが、本手法では物体領域を分割するために、検出した特徴点を出現分布によってクラスタリングする。その結果を元に画像を分割することで、一つの画像を局所特徴量の集まった物体ごとに分割する(図 2)。クラスタリング手法としては、データマイニングの分野において広く用いられている K-Means++法[6]を用いて行う。

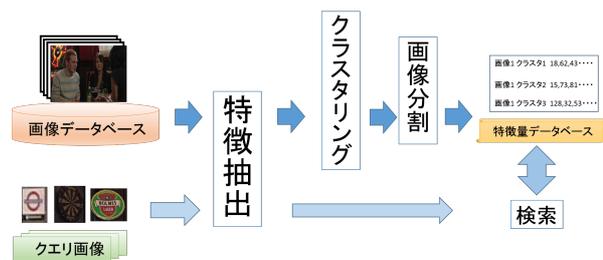


図 1: 提案手法の流れ

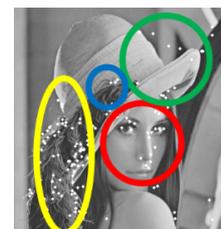


図 2: クラスタリングのイメージ

データベースには分割したクラスタごとに局所特徴量を登録する。登録部のイメージ図を図 3 に示す。局

所特徴量には SIFT 特徴量を用いるため、画像数×クラス数×128 次元で構成される。



図 3:データベース登録のイメージ

類似度として、SIFT 特徴量のユークリッド距離の合計を計算した数値を用いる。最終的な結果として、システムによって類似度が高いと出力された画像に検索物体が写っているかどうか精度の基準となる。

3. 評価

3.1. 実験方法

精度比較のため、2 種類の実験を行った。

- ・実験 1(既存手法)
クラスタリングを施さず、そのままデータベースに登録し 1 枚の画像ごとに類似度を求める。
- ・実験 2(提案手法)
クラスタリングによる画像分割を行い、分割画像ごとに類似度を求める。

クラスタリング時に用いる K-Means++法においては K=8, K=10 としてデータベースを作成した。なお、クラスタの個数はあらかじめ予備実験を行い、比較的分割が上手くなされると予想した個数を用いている。

実験に使用するクエリ画像は、以下の 4 種類とし、種類ごとに写り方と大きさの異なる 4 枚を用意し、計 16 枚の画像をクエリ画像とする。実際に使用したクエリ画像 4 種類の例を図 4～図 7 に示す。クエリ画像にはそれぞれマスク画像を用意し、クエリ画像にマスク画像を適用する。マスク画像内から得られた特徴点のみを入力し、検索する。

データベース画像として、映像コンテンツの解析や検索の高度化を目的とした競争型プロジェクトである TRECVID[7]で用いられる動画を短い時間で分割した画像を用いる。TRECVID では様々なタスクが与えられるが、本手法で用いたデータベースはインスタンスサーチタスクで実際に使用される動画データベースを元に作成している。画像データベースの枚数はおよそ 50 万枚となり、対象となる物体が写っていても、画質や角度が異なっている。

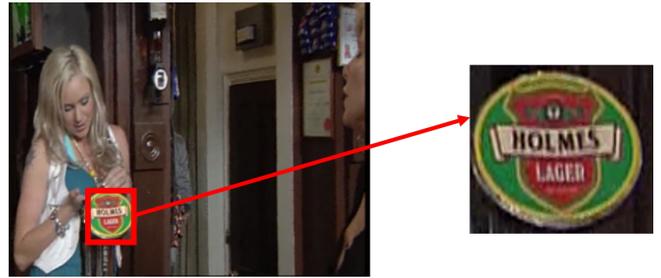


図 4:クエリ画像 1



図 5:クエリ画像 2



図 6:クエリ画像 3



図 7:クエリ画像 4

評価方法として、システムによって類似度が高いと出力した上位 200 件の再現率と適合率を元に評価を行う。使用した計算式を以下に示す。

$$\text{適合率}(\%) = \frac{\text{正解画像枚数(検索結果中)}}{200} \times 100$$

$$\text{再現率}(\%) = \frac{\text{正解画像枚数(検索結果中)}}{\text{正解画像枚数(データベース中)}} \times 100$$

3.2. 実験結果

全画像の再現率と適合率の平均値を図 8 に示す。また、4 種類のクエリ画像ごとの再現率と適合率の平均値を図 9、図 10 に示す。

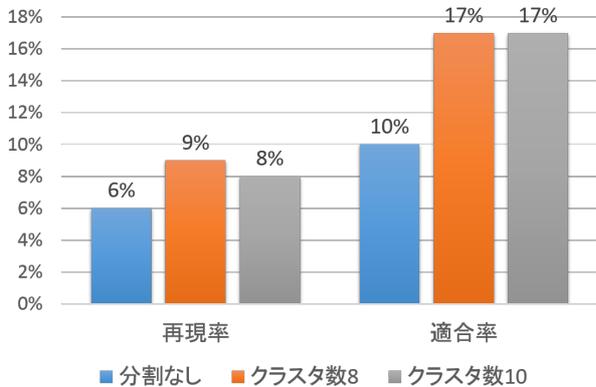


図 8: 適合率と再現率の平均

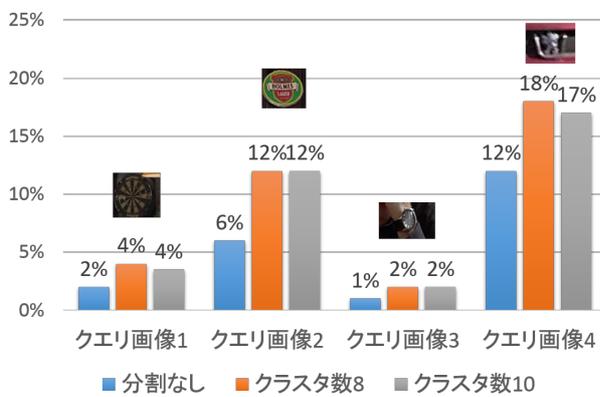


図 9: 再現率の平均

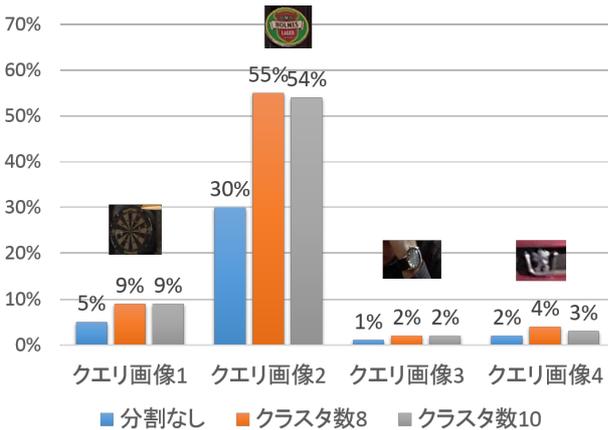


図 10: 適合率の平均

従来手法と提案手法を比較すると、適合率を最大で 25%、再現率を 6% 向上させることができた。このことより、従来手法と比較して特徴点の誤対応は減ったと考えられる。

精度の良かった例として、クエリ画像 2 は検索対象がはっきり写っており、更に比較的大きな物体であったため適合率の向上が見られた。

精度低下の原因としては、クラスタが検索対象その

ものを分割したり、クラスタを大きくとりすぎてしまったりといった例(図 11)が見られたことが原因として考えられる。現システムでは K-means++法ではクラスタ数は静的に決めるしかなく、更にクラスタ数を経験的に決めており、クエリごとに調整していくことはできていない。以上のようなことが精度低下の理由であると考えられる。

また、適合率と再現率をプロットしたグラフを参照し、考察していく。クエリ画像 2 の結果を図 12、クエリ画像 3 の結果を図 13 に示す。全体の画像サイズに対してクエリのサイズが比較的大きいクエリ画像 2 に関して、クラスタのサイズが大きい K=8 の結果が優れている事が分かる。また、クエリサイズの小さいクエリ画像 3 の結果を見てみると、クラスタの細かい K=10 の結果の方が優れている事が分かる。このような傾向がクエリ画像全体として見られることから、クラスタの大きさが検索対象の大きさに比例していると考えられる。

また、クエリ画像ごとに精度が大きく異なっているが、特徴量の性質上クエリの大きさや向き、画質の差に大きく依存しているため、色情報の追加や特徴量の見直しを行う必要がある。



図 11: 分割ミスの例

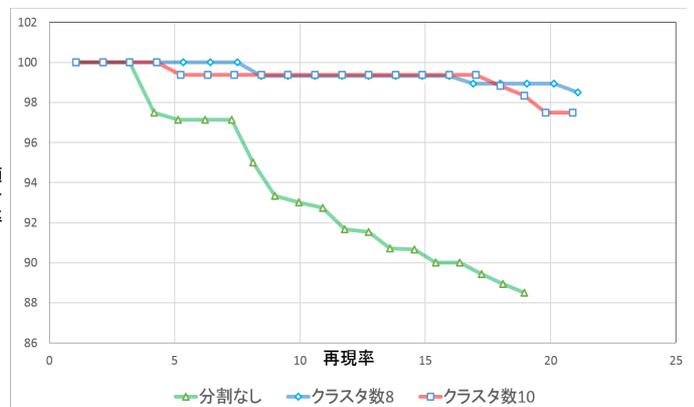


図 12: クエリ画像 2 の結果

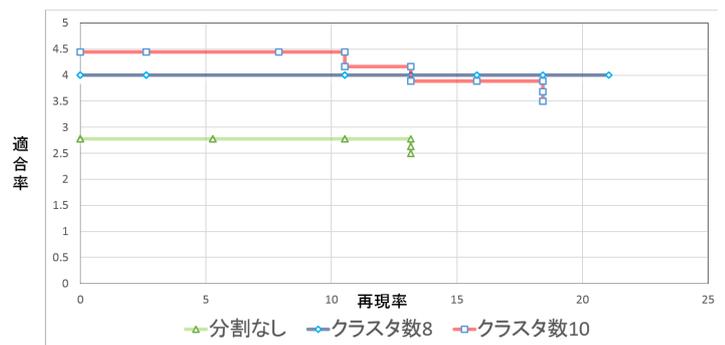


図 13: クエリ画像 3 の結果

4. 改善手法の提案

上記の問題点を解決するため、階層的クラスタリング手法である完全連結法を用いてクラスタリングを行い、実験を行った。

4.1. 階層的クラスタリング

階層的クラスタリングは、ユークリッド距離やミンコフスキー距離などの距離尺度を類似度とし、類似度の高い順に融合していくクラスタリング手法である。融合の手法に関しても、クラスタ内のメンバー間距離を元に融合する最近隣法や最遠隣法、全てのデータ対間の距離の平均を用いて融合していく群平均法がある。

最終的にはクラスタが 1 つになるまで処理を施し、融合の過程がデンドログラムとして得られる。デンドログラム(図 14)の枝の高さやクラスタの個数を指定して区切る事で処理の途中結果も得られる。

K-means のような非階層的クラスタリングでは、データ全体のバランスを考慮してクラスタリングを行うのに対して、階層的クラスタリングはデータ間の距離をより重視して分割できると考え、実際に予備実験を行った。

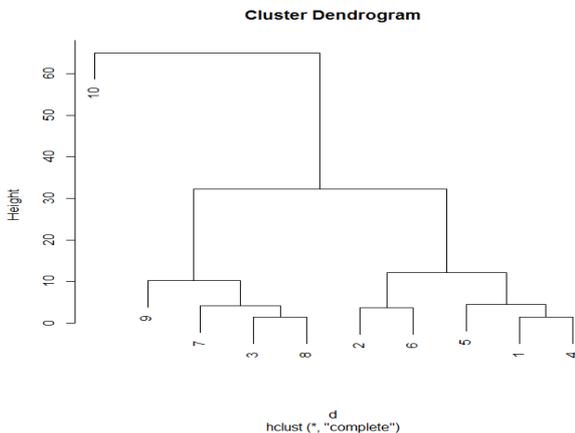


図 14:デンドログラムの例

4.2. 評価実験

3章で使用した画像データベースに対して以下の実験を行い、異なるクラスタリング手法の結果の比較を行う。

・実験 1(既存手法)

クラスタリングを施さず、そのままデータベースに登録し 1 枚の画像ごとに類似度を求める。

・実験 2(非階層的クラスタリング)

K-means++法を用いてクラスタリングを行う。

・実験 3(階層的クラスタリング)

完全連結法を用いてクラスタリングを行う。

階層的クラスタリング手法において、最近隣法のような手法は一定方向にクラスタが長くなってしまい、チェーン状のクラスタが形成されやすくなるのが分かっているため、今回はクラスタ融合の基準としてクラスタ間の最遠隣距離を用いる完全連結法を用い、距離尺度にはユークリッド距離を用いた。非階層的クラスタリング手法には K-means++法を用いた。また、使用するクエリ画像に関しても 3 章と同じもの(クエリ画像 1~4)を用いて比較を行う。K-means++法におけるクラスタ数は、クエリ画像 3 において最も精度が優れていた 8 個を指定した。完全連結法におけるクラスタ数は、最遠隣距離の和に閾値を設け、閾値を超えた時点で融合を止めることで決める。

評価に関しては、3章で行った実験と同じ式を用いて計算した適合率と再現率を元に行う。

4.3. 実験結果

全画像の再現率と適合率の平均値を図 15 に示す。また、4 種類のクエリ画像ごとの再現率と適合率の平均値を図 16、図 17 に示す。

結果を見ると、比較的階層的クラスタリングの結果が良く、上手く分割できていることがわかる。階層的クラスタリングの中でも完全連結法は最遠隣距離を元にクラスタリングするため、距離が離れたクラスタが生成されやすい。精度が良い事からも、クラスタが物体を分割するような例は完全連結法を適用することで解消されたと考える。

非階層的クラスタリングは、分散や重心を元にクラスタリングを行っており、全体のバランスを考慮しながらクラスタリングするため、離れたオブジェクトであっても同じクラスタであると判断してしまう事が多くあった(図 18)。特に K-means 法は超球にかつ各クラスタが同じサイズになるようにクラスタリングするという性質があり、K-means++法でもその性質は維持したままであるため、イメージ通りにクラスタリングできない事が多かったことが精度低下の原因であると考える。

また、今回の実験でも経験則的に閾値を決める必要があるため、引き続き十分な検討が必要である。

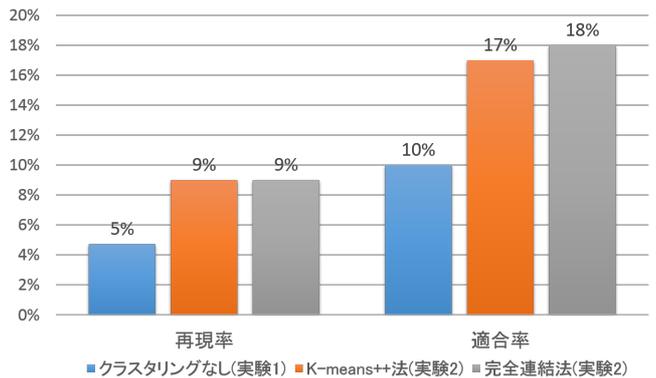


図 15:適合率と再現率の平均

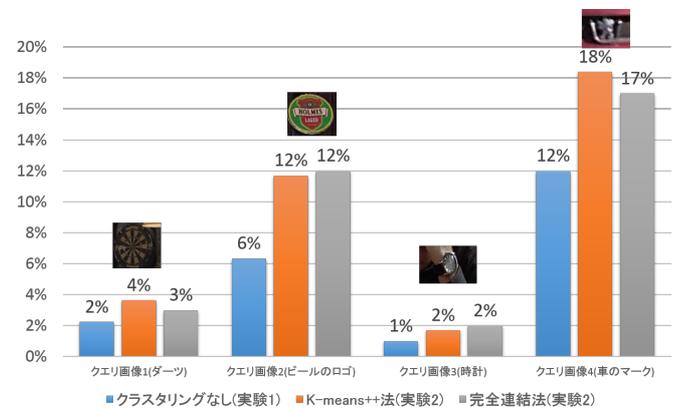


図 16:再現率の平均

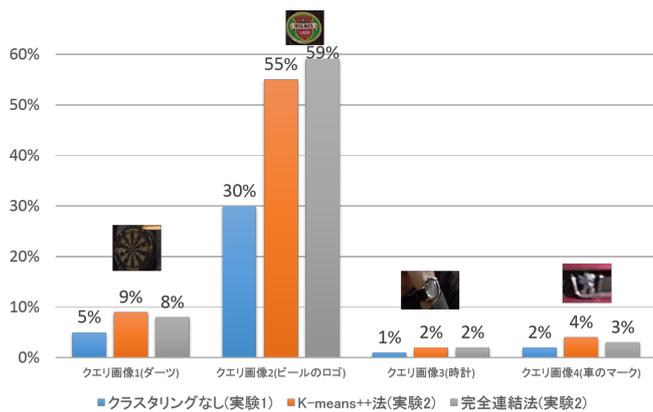


図 17:適合率の平均

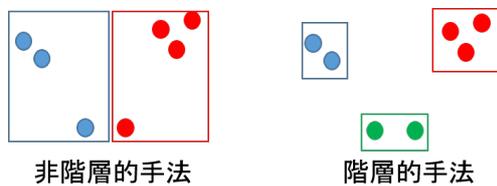


図 18:分割方法の違い

5. まとめ

特定物体認識手法として、局所特徴量のクラスタリングによって画像を分割し、分割した画像に対して特定物体の検索を行い、従来手法との比較を行った所、精度の向上が見られた。クラスタの大きさと検索対象の大きさに相関が表れる傾向にある事も示した。

また、階層的手法と非階層的手法の比較として実験を行い、階層的手法の可能性も示した。

今後はクラスタ数の動的決定手法に関して更なる検討を行うと共に、色情報の追加などの特徴量に関する考察を深める事で、より精度の高い認識手法を目指したい。

参考文献

- [1] D.G.Lowe, "Object recognition from local scale-invariant features", Proc.of IEEE International Conference on Computer Vision (ICCV), pp.1150-1157, 1999.
- [2] J.Zhang, M.Marszalek, S.Lazebnik and C.Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study", International Journal of Computer Vision , 2006.
- [3] Z.Zhao, Y.Zhao, Y.Hua, W.Wang, D.Wan, G.Jia, Z.Li, F.Su and A. Cai: "Bupt-mcprl at trecvid 2012", TRECVID 2012 Workshop Notebook, 2012.
- [4] 高木雅成, 藤吉弘亘, "SIFT 特徴量を用いた交通道路標識認識", 第 13 回画像センシングシンポジウム SSI07, LD2-06, 2007.
- [5] 三吉建尊, 越野亮, 木村春彦, "HOG と Bag of Keypoints を用いた工作機械用制御盤内における物体認識に関する研究", 第 27 回ファジィシステムシンポジウム講演論文集, pp.1183-1188, 2011.

[6]Arthur,D. and Vassilvitskii,S. , "K-means++:the advantages of careful seeding", Proc. of the eighteenth annual ACM-SIAM symposium on Discrete algorithm, 1027-1035, 2007

[7]TRECVID, <http://trecvid.nist.gov>