Vol.2013-CVIM-186 No.19
2013/3/15

局所的な形状特徴量とEMDを用いた類似画像検索手法

星賀 郁仁[1,a)]　樋口 達哉[1]　中島 佑真[1]　獅々堀 正幹[1]

概要：画像の局所的な特徴量である SIFT 特徴量を用いた類似画像検索が近年活発に研究されている. SIFT 特徴量を用いた検索手法として Bag-of-keypoints が有名であり広く普及している. ただし画像全体を固定長のベクトルに落とすため SIFT 特徴量の位置情報が考慮されない．そこで色情報を用いて領域分割を行い，各領域内の SIFT 特徴量から固定長のベクトルを作る方法が考えられる. しかしながら色情報を用いた領域分割を行うと分割数が画像によって変動するので距離尺度としてユークリッド距離を用いることができない. そこで距離尺度として Earth Mover's Distance (EMD) を適用し, 重み付きの特徴量で Bag-of-keypoints を構成することで, 従来の検索手法よりも検索精度を向上させる手法を提案する.

キーワード：Bag-of-keypoints, SIFT, EMD, コンテンツ型類似画像検索

A method of similar image retrieval system using EMD and SIFT

Hoshiga Fumito[1,a)]　Higuchi Tatsuya[1]　Nakajima Yuma[1]　Shishibori masami[1]

Abstract: The content-based image retrieval methods using the SIFT features which is the local features of a image have been studied actively in recent years. The Bag-of-keypoints is very famous as the retrieval technique using the SIFT features. However, in order to quantize the whole SIFT features extracted from the image to a fixed-length feature vector, the positions of each SIFT in the image can not be taken into consideration. This method applys color segmentation module in order to separate the corresponging image into some regions which have same color pixels. And then, this method makes the corresponding fixed-length feature vector form SIFT features in each region area. However, t is impossible for this method to use the Euclidean distance measure, because the number of color segmentation areas of the image is not fixed value, as a result, the lenght of vector also changes. In order to solve this problem, this mehod applys the Earth Mover's Distance (EMD) as the distance measure instead of the Euclidean distance.

Keywords: Bag-of-keypoints, SIFT, EMD, Content-based image retrieval methods

1. 背景と目的

近年, インターネットの高速化に伴い, 子供から老人と幅広い年齢層でパーソナルコンピュータ及び携帯電話でのインターネット接続が行われるようになってきた. 同時に SD カード, メモリースティックといった外部記録メディアの大容量化なども急速に進み, 画像, 映像, 音楽といった大容量データがデジタル化されている. これらを人の手で分類し, 検索することは困難であり, 自動的に分類, 検索できるシステムの構築の必要性が高まっている.

本論文ではコンテンツ型画像検索と呼ばれる手法の中で, 画像内の形状特徴量を用いた類似画像検索手法の精度向上を目標としている. 用いる形状特徴量は SIFT 特徴量と呼ばれる, 照明, スケール, 回転の変化に頑強な特徴量である.SIFT 特徴量は画像ごとに何百もあり, それらを一対一で比較していては計算が膨大になってしまう. そこで Bag-of-keypoints 手法と呼ばれる, 画像を数次元の特徴ベクトルとして表現し, 検索を行う手法が提案されている. しかしこの手法では, 画像内の形状は考慮されるが, 画像内の SIFT 特徴量の位置や色情報は一切考慮されないという問題点がある. そこで色情報を用いて領域分割を行い，各領域内の SIFT 特徴量から固定長のベクトルを作る方法が考

[1] 徳島大学大学院先端技術科学教育部システム創生工学専攻
[a)] hoshiga-fumito@iss.tokushima-u.ac.jp

Vol.2013-CVIM-186 No.19
2013/3/15

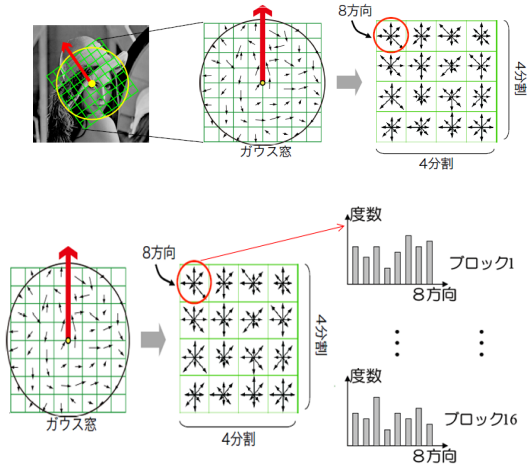

図 1 特徴量記述

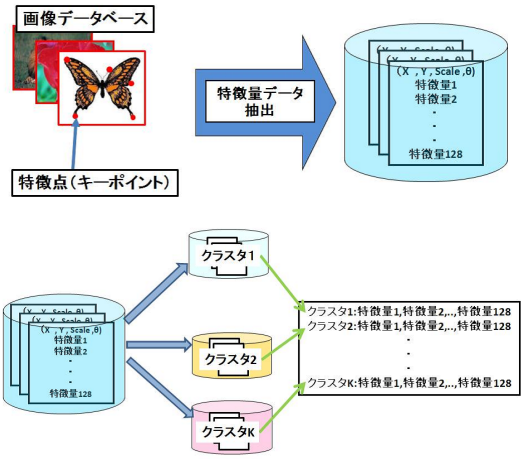

図 2 特徴量の抽出とクラスタリング

えられる. しかしながら色情報を用いた領域分割を行うと分割数が画像によって変動するので距離尺度としてユークリッド距離を用いることができない. そこで距離尺度として Earth Mover's Distance (EMD) を適用し, 重み付きの特徴量で Bag-of-keypoints を構成することで, 従来の検索手法よりも検索精度を向上させる手法を提案する.

2. Bag-of-keypoints 手法

Bag-of-keypoints とは, 画像を局所特徴量の集合として捉えた手法である. 膨大なデータを持つ特徴量をベクトル量子化することで, 精度をある程度保ったまま高速な検索が可能となっている. 今回使用した局所特徴量は,SIFT(Scale Invariant Feature Transform) 特徴量である.

2.1 SIFT 特徴量

SIFT は Lowe [1] によって提案された特徴ベクトルの抽出法である. 名が示す通り, 画像の拡大縮小, 回転や視点のズレに対してロバストであるという特徴を持つ. この特徴のため, イメージモザイク等の画像マッチングや物体認識に用いられている. 特徴ベクトルは 128 次元の整数値のベクトルで表される (図 1).

2.2 Bag-of-keypoints 表現

抽出された特徴量をクラスタリングし, クラスタごとに visual words と呼ばれる代表的な特徴ベクトルを生成し, 画像内の特徴ベクトルを最も類似する visual words で置き換える. そして各画像に visual words のヒストグラムで表現する. 数百の 128 次元の特徴ベクトルを数次元に量子化することで, 精度を保ったまま検索速度を向上させることができる (図 2). 検索には visual words のヒストグラムを, 距離尺度にはユークリッド距離を用いる.

3. 改良手法

Bag-of-keypoints 表現では, 画像の形状的な特徴に基づいて検索を行った. また SIFT 特徴量で用いられる 128 次元の特徴ベクトルは画像をグレースケールとして捉えて抽出するため, 画像内の色情報は用いられない. このため従来法では形は似ているが, 色の異なる物体の検索が不可能である. そこで色情報を用いた EMD(Earth Mover's Distance) を取り入れることで, 色の違いによる検索を可能とする.

3.1 EMD

Earth Mover's Distance(EMD) とは, 線形計画問題の 1 つである輸送問題の解に基づいて計算される距離尺度である. これは 2 つの離散分布において, 一方の分布を他方の分布に変換するための最小コストとして定義される.

EMD を計算するために必要な輸送問題とは, 一定の供給量を持つ複数の供給地と一定の需要量を持つ複数の需要地を設定し, 各供給地から需要地までの単位輸送コストを与えた場合, 需要地の需要を満たすように供給地から需要地へそう輸送コストが最小となるように荷物を輸送する輸送方法を探す問題である.

まず, m 個の供給地を持つ供給地集合, n 個の需要地を持つ需要地集合 $\boldsymbol{P}$, $\boldsymbol{Q}$ をそれぞれ以下のように表す.

$$\boldsymbol{P} = \{(\boldsymbol{p}_1, w_{p_1}), ..., (\boldsymbol{p}_m, w_{p_m})\} \quad (1)$$

$$\boldsymbol{Q} = \{(\boldsymbol{q}_1, w_{q_1}), ..., (\boldsymbol{q}_n, w_{q_n})\} \quad (2)$$

ここで $\boldsymbol{p}_i$ は i 番目の供給地を表す特徴ベクトルであり, w_{p_i} は i 番目の供給地の供給量を示す. 同様に, $\boldsymbol{q}_j$ は j 番目の需要地を表す特徴ベクトルであり, w_{q_j} は j 番目の需要地が必要とする需要量を示す. そして $\boldsymbol{P}$, $\boldsymbol{Q}$ の各要素である供給地 i, 需要地 j 間の単位輸送量あたりの輸送コスト (d_{ij}) を定義する. 単位輸送コストは解く問題によって様々に定義可能であるが, 一般的には単位輸送コストとして各要素

の特徴ベクトル $\boldsymbol{p}_i$, $\boldsymbol{q}_j$ のユークリッド距離が用いられ,

$$d_{ij} = ||\boldsymbol{p}_i - \boldsymbol{q}_j|| \tag{3}$$

として定義されることが多い.

次に, 供給地 i と需要地 j のすべての組み合わせの輸送量とそれに応じた輸送コストを考慮し，総輸送コストを計算する. 総輸送コストは, 供給地 i から需要地 j への輸送量 (フロー) ($F = \{f_{ij}\}$) を決定する以下の輸送問題を用いて計算する. 任意の供給地・需要地の組み合わせによる総輸送量 (WORK) は,

$$\mathrm{WORK}(\boldsymbol{P}, \boldsymbol{Q}, F) = \sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij} \tag{4}$$

と表す.

この目的関数は, i, j 間の輸送量に単位輸送コストを掛けて和をとることで総輸送コストが計算されることを表している. ただし総輸送コストを計算する場合, 以下の制約条件 (式 (5)〜式 (8)) を満たすものとする.

- 制約条件: 供給地から需要地の一方向にしか輸送されない

$$f_{ij} \geq 0, \ (1 \leq i \leq m, 1 \leq j \leq n) \tag{5}$$

- 制約条件: 供給地 i から供給できる容量は供給量 w_{p_i} を超過しない

$$\sum_{j=1}^{n} f_{ij} \leq w_{p_i}, \quad (1 \leq i \leq m) \tag{6}$$

- 制約条件: 需要地 j が受け取れる容量は w_{q_j} は以下であること

$$\sum_{i=1}^{m} f_{ij} \leq w_{q_j}, \quad (1 \leq j \leq n) \tag{7}$$

- 制約条件: 供給地から移動する輸送量 (総フロー)

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_i}\right) \tag{8}$$

最終的に EMD($\boldsymbol{P}$, $\boldsymbol{Q}$) は，上の輸送問題の最適値（総輸送コストの最小値）$\min(\mathrm{WORK}(\boldsymbol{P}, \boldsymbol{Q}, F))$ を総フローで割って,

$$\mathrm{EMD}(\boldsymbol{P}, \boldsymbol{Q}) = \frac{\min(\mathrm{WORK}(\boldsymbol{P}, \boldsymbol{Q}, F))}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}} \tag{9}$$

と計算できる.

EMD の計算方法の例が図 3 である. 供給地がトラック, 需要地が□, 供給量・需要量がみかんである.

類似画像検索では, 画像を色領域に分割して考える. クエリ画像の色領域が供給地であり, データベース画像の色領域が需要地である. 色領域の画素を供給量・輸送量とする. 単位輸送コストは, 色領域の色情報 (赤, 緑, 青) と重心 (X 座標,Y 座標) を特徴ベクトルとし, ユークリッド距離として定義する (図 4).

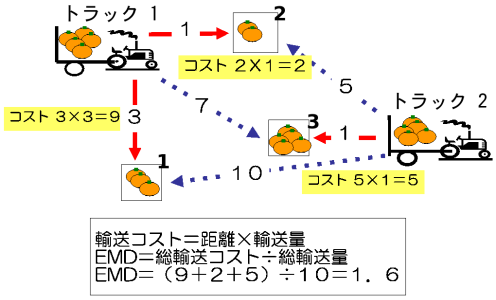

図 3 EMD の計算例

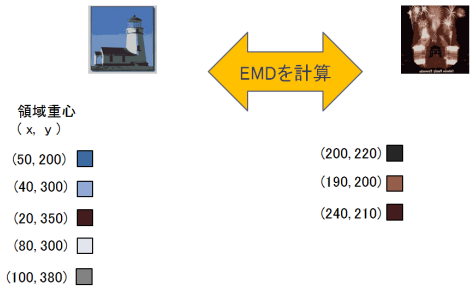

図 4 類似画像検索における EMD の利用

3.2 Bag-of-keypoints + EMD

EMD を用いた類似画像検索によって画像は色領域に分割されるが, 色領域ごとに Bag-of-keypotins 手法を用いて特徴ベクトルを作成する. 前述した色領域の情報 (重心の X,Y 座標, 色領域の赤緑青) と特徴ベクトルを合わせたものが, 提案手法における画像の特徴ベクトルとなる. 全体の手順は以下の通り.

1. 全画像から特徴量を抽出する
   openCV2.4.2 の cv::SiftFeatureDetector と cv::SiftDescriptorExtractor を使用して SIFT 特徴量を抽出した.

2. 特徴量をクラスタリングし, visual words を求める
   クラスタリングには k-means を使用した.

3. 全画像を減色処理し, 色領域に分割する
   今回は ImageMagick を使用し, 色上限を指定して減色した. 画像によっては上限に満たないこともある (図 5).

4. 色領域ごとにヒストグラムを作成する
   図 6 は色領域が 5 つ, visual words が 7 つの場合である.

5. 画像間の EMD を計算する
   図 4 と同じく, EMD を用いる (図 7). 単純なユークリッド距離を用いた場合だと, 画像により色領域の数が異なり, 計算することができないが, EMD のだと問題なく計算できる.

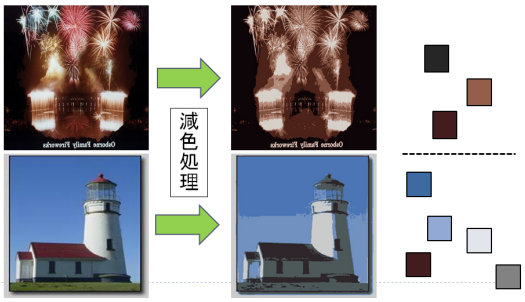

図 5 減色処理

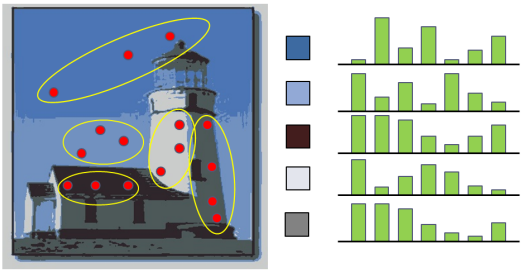

図 6 色領域ヒストグラム作成

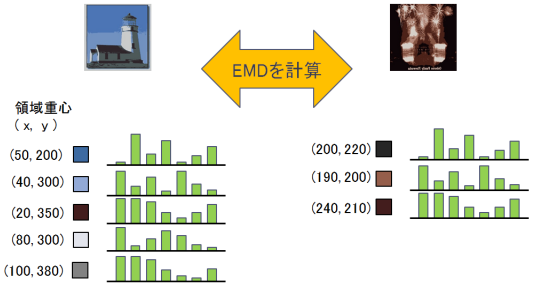

図 7 色情報とヒストグラムを用いた EMD

4. 評価

改良を加えた Bag-of-keypoints+EMD と, 形状特徴量だけの Bag-of-keypoints, 色情報だけの EMD と比較した. 条件は以下の通り.

- データベースは Caltec256 から選出した 10 のカテゴリ (表 1) から, 0001 から 0090 までの 90 枚を使用した. データベースの全画像数は 900 枚.

表 1 Caltec256 から選出した 10 カテゴリ

015.bonsai-101	盆栽
016.boom-box	ラジオ
023.bulldozer	ブルドーザー
036.chandelier	シャンデリア
072.fire-truck	消防車
073.fireworks	花火
092.grapes	ぶどう
132.light-house	灯台
213.teddy-bear	テディベア
251.airplanes-101	飛行機

- 従来法の色情報 EMD では減色数を 1 色から 24 色まで, 24 の環境に変化させた. ただし減色数を 5 色に設定したからといって, すべての画像が 5 色にはならない. 夜景や海が大部分を占める画像では色情報が乏しく, 5 色に満たないこともある.
- 従来法の Bag-of-keypoints 手法では, visual-words の数を 2 個から 24 個まで, 23 の環境に変化させた.
- 改良を加えた Bag-of-keyoitins+EMD では, 減色数を 1 色から 24 色まで, 24 の環境に変化させると同時に, visual-words の数を 2 個から 24 個まで変化させた. 24 色 ×23 個で計 552 個の環境を比較する.
- 入力画像はデータベース内の 900 枚の画像を使用する.

4.1 結果

900 枚の画像ごとに各 3 手法で最も良い平均適合率を持つ環境を比較し, 画像ごとにどの手法が良いか調べた (表 2).

表 2 手法比較

	提案手法	Bag-of-keypoins	EMD
最良画像数 (900 枚中)	428 枚	389 枚	83 枚

またカテゴリごとにどの手法が良いかも調べた (表 3). 1 つのカテゴリにつき, 90 枚の画像がある. 提案手法が大幅に優位である場合○を, 僅差で優位の場合は△を最後の列に入れた. 提案手法は前述したように, 552 個の環境がある

表 3 カテゴリごとの手法比較

カテゴリ	提案手法	Bag-of-keypoins	EMD	
盆栽	20	58	12	
ラジオ	45	44	1	△
ブルドーザー	55	28	7	○
シャンデリア	16	62	12	
消防車	46	40	4	△
花火	76	9	5	○
ぶどう	16	45	29	
灯台	59	26	5	○
テディベア	33	49	8	
飛行機	62	28	0	○

(24 色 ×23 次元). それらの中で, どの環境がよいかを 3 次元グラフに表した (図 8). 900 枚の入力画像において, 552 個ある環境の中で最も良い平均適合率がどれかを表したものである.

5. 考察

表 2 を見るに, 全体としては向上していると言える. しかし表 3 では, カテゴリによっては従来の Bag-of-keypoints に劣る点も見られた. 良好な結果を残したのはブルドーザー, 花火, 灯台, 飛行機で, これらに共通して言えることは

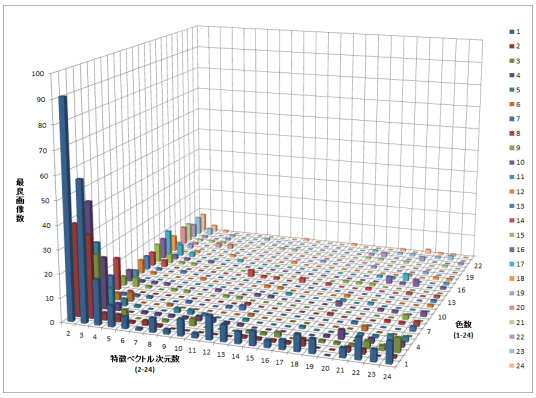

図 8 最適な色数とヒストグラム

背景が色的に似通っている点である. ブルドーザーは灰色の地面と空が, 花火は夜景が, 灯台は海と空が, 飛行機は飛行するものは空, 着陸しているものは芝生の上にあるものが多く, 物体と検索したというよりは, 似通った背景から検索を行ったと考えられる. 不良な結果を残したのは, 盆栽, シャンデリア, ぶどう, テディベアとなった. 不良な結果を残した理由は二つあると考えている. 一つ目は盆栽, ぶどうといったカテゴリは, 形状的に似通ったものが多く, 従来法の Bag-of-keyoituns で十分だった点. 二つ目はシャンデリアのカテゴリにおいてだが, 背景が白と黒に大別できた点で, 提案手法に取り入れた EMD で白と黒は EMD が大きくなり, 同カテゴリであっても白背景と黒背景を別の物体だと認識し, 検索精度が悪化したと考えられる.

また, 特に注目したいのは, 900 枚の画像のうち, より良い結果になったのが減色数が 1 色, visual words の数が 2 個の場合であった点である (図 8). 色情報と形状情を組み合わせて検索を行う場合, 画像の色の雰囲気 (赤っぽい, 黒っぽい) と, 頻出する visual words とそれ以外の visual words のヒストグラムを用いることで, 精度の高い検索が行えることを示している.

6. まとめ

本論文では, Bag-of-keypoints 手法を用いての類似画像検索を改良し, 検索精度の向上を示した. 画像を色領域で分割し, 領域ごとにヒストグラムを作成して検索を行う場合, 従来の Bag-of-keypoints 手法では, 画像ごと色数が異なった場合, 距離尺度にユークリッド距離を用いていたため検索できないという問題点があったが, 本論文では距離尺度に EMD を用いることで, 色数の異なる画像間の類似度を数値化し, 検索することができた. 結果としては, 大部分において成功していると言える. 問題点として, 画像の種類, 特に背景の色情報によって精度が低下することも確認できた. これは画像内における物体の画素の割合より, 背景の割合のほうが多いため, EMD を用いる場合は特に背景の色の違いが如実に現れた結果となった.

今後の課題としては, データベースの画像数を増やすことと, 検索速度の向上を考えている. また SIFT 特徴量よりも適した特徴量がないかも検討したい.

参考文献

[1] Lowe, D.G : Object recognition from local scale invariant features, Proc. of IEEE InternationalConference on Computer Vision, pp. 1150-1157(1999)