

# Earth Mover's Distance を用いた SIFT 特徴量に基づく類似画像検索手法

西野 裕一<sup>†</sup> 獅々堀 正幹<sup>‡</sup> 北 研二<sup>‡</sup>

<sup>†</sup> 徳島大学大学院 先端技術教育部 システム創生工学専攻

<sup>‡</sup> 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

E-mail: <sup>†</sup> {n\_yuichi, bori, kita}@is.tokushima-u.ac.jp

**あらまし** 近年、圧縮技術の発展により、大量のデジタル画像をコンピュータ上やネットワーク経由で扱うことが容易になった。それに伴い、大量の画像データを自動的に分類、検索するシステムの必要性が向上している。検索手法としては、画像の内容（色、形状）を解析することで多次元ベクトル化した特徴量を用いる手法、また、入力画像との類似性を計る距離尺度としては、ユークリッド距離を用いる手法が主流である。本稿では、画像から得られる特徴として Scale Invariant Feature Transform (SIFT) 特徴量を用い、距離尺度として Earth Mover's Distance (EMD) を用いる手法を提案する。SIFT は、特徴点の検出と特徴量の記述を行うアルゴリズムであり、EMD は、輸送問題における輸送コストの最適解である。本手法では、輸送問題における各供給地が有する資源量および輸送コストを SIFT 特徴量から算出することで、類似した画像を検索する。

**キーワード** EMD, SIFT, 類似画像検索

## Similar Image Retrieval Method based on SIFT Feature Using the Earth Mover's Distance

Yuichi NISHINO<sup>†</sup> Masami SHISHIBORI<sup>‡</sup> and Kenji KITA<sup>‡</sup>

<sup>†</sup> The University of Tokushima graduate school

2-1 Minami-Josanjima-cho, Tokushima-shi, Tokushima, 770-8506 Japan

E-mail: <sup>†</sup> {n\_yuichi, bori, kita}@is.tokushima-u.ac.jp

### 1. はじめに

近年、HDD など大容量メディアの普及に伴って大量のデジタル画像・写真などが用意に保存可能となった。しかし、それらを人手によって分類、検索することは困難であり、自動的に分類、検索するシステムへの必要性が増している。類似画像検索技術として、キーワード型検索と呼ばれる技術は、キーワード（タグ）を与えて検索するが、個々の画像にキーワード（タグ）付けする必要があるため手間や時間がかかるという問題点がある。一方、上記の問題を解決するため、画像の内容（色、形状）を解析することでデータを多次元ベクトル化し、多次元ベクトル間の距離を測ることで類似した画像を検索する内容型類似画像検索 (Content-based Image Retrieval) が研究されている。

近年、画像全体からの特徴量ではなく、画像から得られる局所特徴量を用い、画像間の類似性を判定する手法が注目されている。局所特徴量としては、SIFT (Scale Invariant Feature Transform) 特徴量 [1] を用いるものが

有名である。これは、画像から検出した数多くの SIFT 特徴量を Bag-of-keypoints 手法 [2] を用いて、固定長の多次元ベクトル表現に変更した後、画像間の類似性をユークリッド距離により判定するものである。しかしながら、上記の手法では、事前に計算コストの高いクラスタリングを行う必要があり、また最適なクラスタ数が登録画像数や種別により変動する点が問題となる。更に、SIFT 特徴量として検出可能な特徴点の位置情報及びスケール値を反映することができないといった問題点もある。

そこで本稿では、距離尺度に EMD (Earth Mover's Distance) [3] を用いることにより、上記の問題点を解決した SIFT 特徴量に基づく類似画像検索手法を提案する。本手法では、EMD を計るための特徴量に直接、個々の SIFT 特徴量を記述するため、事前にクラスタリングをする必要がない。また、EMD の特徴量に SIFT 特徴量の位置情報及びスケール値を反映させることができるため、高精度な検索が可能になる。

以下,2章にて従来手法である Bag-of-keypoints 手法について説明する.次に,3章にて EMD について,4章では EMD を用いた提案手法について述べる.その後,5章にて実験による評価を行い,6章でまとめと今後の課題にふれる.

## 2. 従来手法

### 2.1 SIFT とは

Scale Invariant Feature Transformation(SIFT)は David Lowe によって考案された. SIFT アルゴリズムは特徴点検出部と特徴点周辺の特徴量記述部に分けられる. 特徴点検出部では, DOG(Difference of Gaussian) により, 原画像を数段階に縮小したマルチスケール画像群から濃淡変化の大きな特徴的な点を多数検出し, 特徴量記述部において, その特徴点周辺の局所領域の特徴量を 128 次元特徴ベクトルで表現する. この特徴は同じ対象であれば異なる画像でも, 同じ場所が特徴点として検出される. 学習, テスト全画像に対して局所特徴を 1 枚の画像につき約 100~1000 個検出し, SIFT 特徴で表現する. SIFT 特徴の概要は以下のようになっている.

- ガウシアン差分を最大にすることによりスケールを決定
- 主要な勾配方向として局所方向を決定  
(すべての更なる計算をスケールと回転に対し不変にするために, このスケールと方向を使う)
- 勾配方向ヒストグラムを計算し, 128 次元ベクトル(4 × 4 × 8) で表現
- 強度変化に強くするために記述子を正規化

この特徴量は, アフィン変換に対して耐性のある類似性の検出が行えるようになるとされている. つまり, 拡大, 縮小, 回転, 歪み, 遮蔽, 3D 視点の変化に強く, ノイズが入った画像に安定して類似性が検出される.

図 5 において特徴点から特徴量を記述する例を示す. 得られた特徴点からは X 座標, Y 座標, スケール, オリエンテーション, 128 次元のデータが記述される.

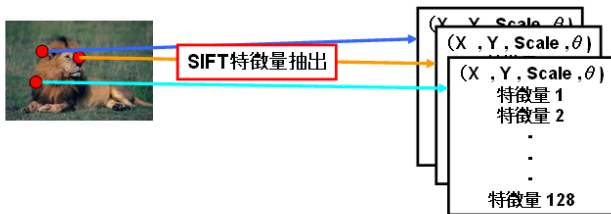


図 1.特徴点抽出例

### 2.2 Bag-of-keypoints 手法による類似画像検索

Bag-of-keypoints 手法とは, 画像を局所特徴の集合と捉えた手法である. 局所特徴の特徴ベクトルをベクトル量子化し, visual words と呼ばれる特徴ベクトルを生成する. それらをまとめたものを Code book と呼び, それを記述子として画像の特徴ベクトルを生成する. よって, 画像は visual word の集合(bag) として表現されるわけである. Bag-of-keypoints を用いた画像検索の流れは以下ようになる. (図 2.参照)

- (1) 学習画像・テスト画像データから特徴を抽出
- (2) それをベクトル量子化し, codebook を作成
- (3) codebook をもとに学習画像の特徴ベクトルを生成
- (4) テスト画像の特徴ベクトルも生成し, 学習画像との比較によって結果の出力

しかし, Bag-of-keypoints 表現を用いた検索手法においても問題点が存在する. まず, 事前に計算コストの高いクラスタリングをする必要がある. その際, 適切なクラスタ数が画像数や種類ごとに異なり, 最適なクラスタ数を設定することが困難である. また, 特徴量が Codebook に代替されているため, 実際の特徴量とのずれが生じ, 精度低下をまねく恐れがある. 更に SIFT 特徴量として抽出した特徴点の位置情報, スケール値を用いることができないといった問題点もある.

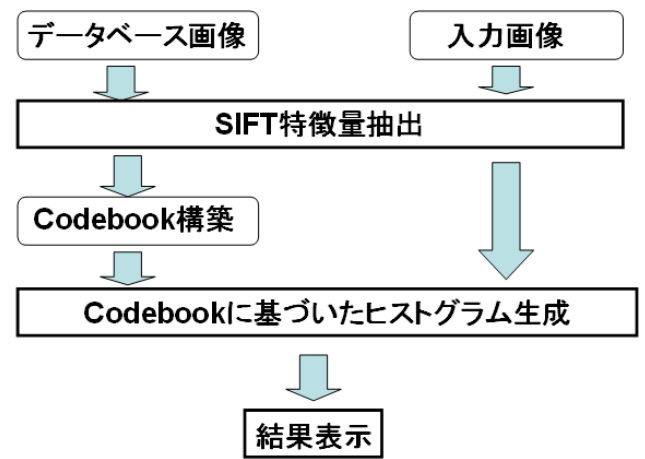


図 2. Bag-of-keypoints 手法を用いた画像検索の流れ

### 3. EMD とは

Earth Mover's Distance(EMD)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. 輸送問題とは一定の供給量をもつ複数の供給地と同じく一定の需要量を必要とする需要地を設定し, 各供給地から需要地までの輸送コストが与えられた際に, 需要地の需要を満たすように供給地から需要地へ最小の輸送

コストで荷物を輸送する輸送法要を探す問題である。次に,EMD の計算方法を説明する。

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

$$P = \{(p_1, w_{p1}), \dots, (p_m, w_{pm})\}$$

$$Q = \{(q_1, w_{q1}), \dots, (q_n, w_{qn})\}$$

ここで, $p_i$  は i 番目の供給地を表す特徴ベクトル, $w_{pi}$  は i 番目の供給地が有する供給量, $q_j$  は j 番目の需要地を表す特徴ベクトル, $w_{qj}$  は j 番目の需要地が必要とする需要量を示す。供給地  $p_i$  と各需要地  $q_j$  間の単位輸送辺りの輸送コスト ( $d_{ij}$ ) を定義する。

$$d_{ij} = \|p_i - q_j\|$$

一般に,輸送コストと各ベクトル要素  $p_i$  と  $q_j$  のユークリッド距離を用いる。次に  $p_i$  と  $q_j$  のすべての組み合わせを考慮し,総輸送コストを計算する。

総輸送コストは,P から Q への輸送量(フロー  $F = \{f_{ij}\}$ ) を決定する以下の輸送問題の解を用いて計算する。任意の供給地・需要地の組合せによる総輸送コスト (WORK) は次式のように表すことができる。

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

ただし,総輸送コストを計算する場合,以下の制約条件を満たすものとする。

- 制約条件 1:供給地から需要地の 1 方向にしか輸送されない。

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

- 制約条件 2:供給地 i から供給できる容量は供給量  $w_{pi}$  を超過しない。

$$\sum_{j=1}^n f_{ij} \leq w_{pi}, (1 \leq i \leq m)$$

- 制約条件 3:需要地 j が受け取れる容量は需要量  $w_{qj}$  以下である。

$$\sum_{i=1}^m f_{ij} \leq w_{qj}, (1 \leq j \leq n)$$

- 制約条件 4:供給地から移動できる最大総輸送量

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min \left( \sum_{i=1}^m w_{pi}, \sum_{j=1}^n w_{qj} \right)$$

最終的に EMD(P,Q)は,上の輸送問題の最適値 (総輸送コストの最小値)  $\min(WORK(P,Q,F))$  を総フローで割って以下のように求まる。

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

EMD の計算処理の例を図 4 に示す。図 3 において供給地が 4 つ,需要地が 3 つありそれぞれ図に示す資源が割り当てられる。供給地から需要地への距離はユークリッド距離にて求められ矢印上の値は供給地から運ばれる資源量である。右上(図 4)にて示す計算式によって輸送フローを表す。

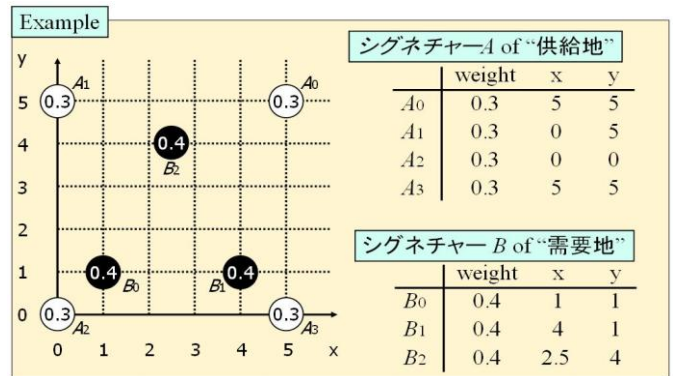


図 3.供給地,需要地にある資源量および位置

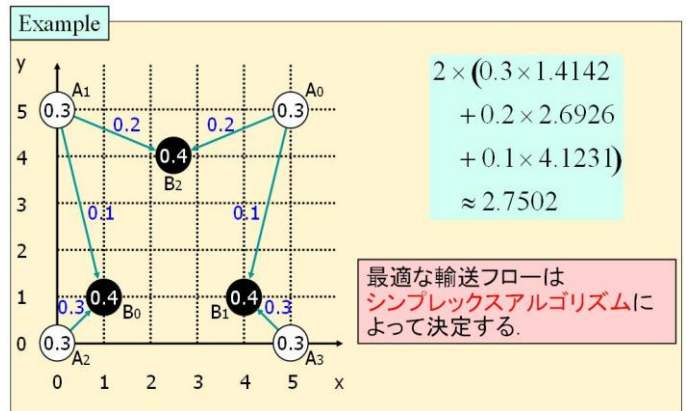


図 4. EMD の計算例

## 4. 提案手法

### 4.1. EMD 距離尺度における特徴量

本研究で用いる特徴量は,各画像からスケールの大きい SIFT 特徴点を上位 100 件抽出する。距離尺度に EMD を適用することで,輸送コスト,需要量及び供給量を定義する必要がある。輸送コストは,SIFT アルゴリズムによって得られる x 座標,y 座標の値と 128 次元の特徴量の計 130 次元の特徴量から計算される。また,各画像から得られる SIFT 特徴量のスケール値の大きさを式(1)により計算した値を需要量及び供給量とする。

$$\frac{100*(key+1-順位)}{\sum_{j=1}^{key} j} \quad (1)$$

key は画像から抽出できる特徴点数を示す。1~100 までの値を取る。

画像から抽出された特徴点数が key=100 の場合

- スケール値の最も大きいもの (1 位)は式(1)より 1,980 となる

$$\frac{100*(key+1-順位)}{\sum_{j=1}^{key} j} = \frac{100*(100+1-1)}{\sum_{j=1}^{100} j} = \frac{100*100}{5050} = 1.980$$

- スケール値の最も小さいもの(100 位)は point=1 として式(1)より 0.0198 となる。

$$\frac{100*(key+1-順位)}{\sum_{j=1}^{key} j} = \frac{100*(100+1-100)}{\sum_{j=1}^{100} j} = \frac{100*1}{5050} = 0.0198$$

このように,作成する方法としてポイント制を用いた EMD 特徴量と呼ぶことにする。

## 4.2 提案手法の特徴

従来手法である Bag-of-keypoints 手法の問題点を以下のように解決できる。

距離尺度に EMD を用いることで,

- 特徴点ごとの特徴量を用いるため,クラスタリングを行う必要がない
- 特徴点ごとに実際の特徴量を用いるため, Codebook に代替する必要がない
- 特徴点ごとの位置情報,スケール値を用いることができる

距離尺度として EMD を用い,特徴量として SIFT 特徴量を用いた類似画像検索手法を提案する。提案手法での画像検索の流れは以下ようになる。

- データベース画像から特徴量を抽出
- EMD 距離尺度を用いるために特徴量の変換
- EMD 距離尺度を用いて距離計算を行う
- 同様に入力においても特徴量抽出,変換を行い,データベースとの距離計算によって結果の出力

## 5. 評価実験

### 5.1. 実験条件

検索対象の画像データベースとしては Corel 画像集より,様々な画像のカテゴリ-100 種類,各カテゴリごとに 10 枚, 計 1000 枚を使用した。実験に用いた計算機

の OS は Linux, CPU は Intel Pentium 4 ,CPU 3.00GHz メモリは 270MB である。

### 5.2. 評価方法

提案手法と比較するため,色情報ヒストグラムを用いた検索手法,SIFT 特徴量より得られる Codebook を用いた検索手法を従来手法として用いた。また,評価尺度は適合率を用いた。

### 5.3. 実験結果

画像 DB から得られる 100 種類各 10 枚,計 1000 枚を入力画像とする。正解とする画像は同じカテゴリーに含まれる画像 9 枚と入力として与えた 1 枚の計 10 枚とする。表 1 は正解画像を検索したときの平均適合率の平均である。

表 1.各システムの平均適合率の平均

検索手法	色情報	Bag-of-keypoints	提案手法
平均適合率(%)	21.32	16.87	34.63

図 5~8 では,4 件の入力画像の平均適合率と提案手法による実際の検索結果の上位 10 件を表示したものである。図 5,6 は従来手法より検索精度が向上した例である。これらの画像では,入力画像と同じ位置に同じような特徴量をもつ SIFT 点が検出されており,位置情報を EMD 特徴量に導入することにより,精度が向上したと考えられる。

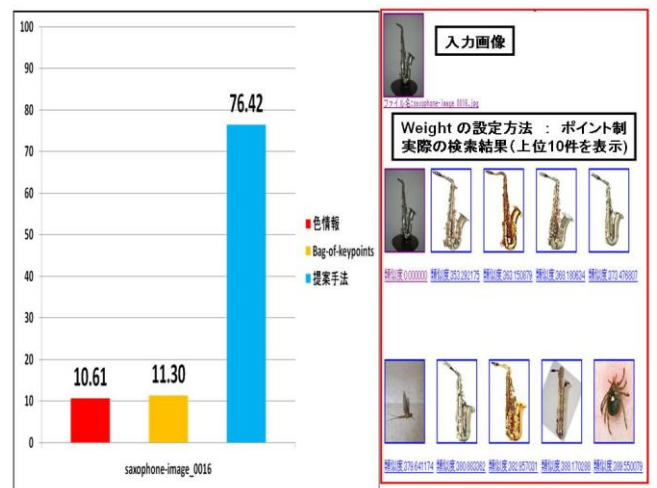


図 5. 評価実験の結果 1

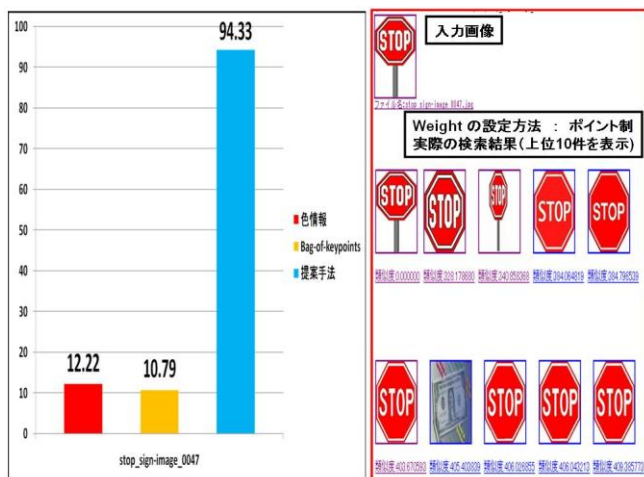


図 6. 評価実験の結果 2

図 7 は従来手法と比較するとほとんど同じ精度であった例である。これらの画像では、画像 DB に含まれる正解画像が適切でない。背景を持つ画像、持たない画像における特徴点の差があると考えられる。

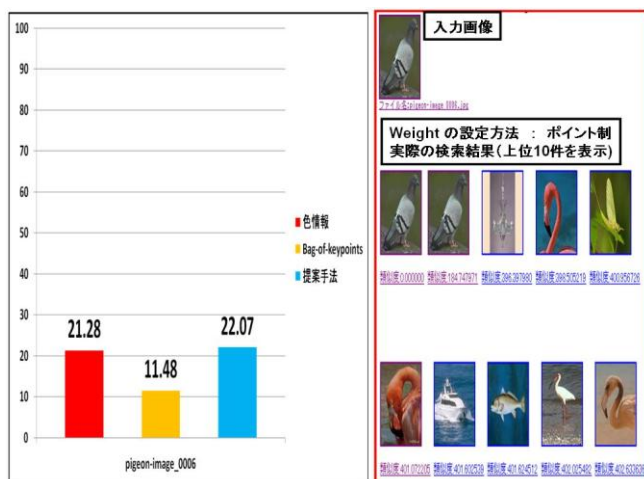


図 7. 評価実験の結果 3

図 8 は従来手法よりも提案手法の検索精度が低下した例である。入力画像と似た特徴点を検出された画像は検索されたが、正解画像にデザインが違うカメラがあるため検索精度が低下したと考えられる。

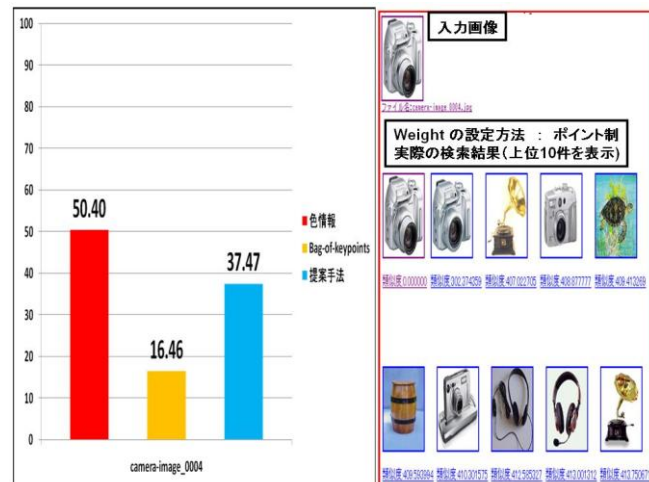


図 8. 評価実験の結果 4

## 6. まとめ

本論文では、距離尺度として EMD を用い、特徴量として SIFT 特徴量を用いた類似画像検索手法を提案した。評価実験においては、従来の検索手法である色情報を用いた検索手法、Bag-of-keypoints 手法との比較において検索精度の向上が確認できた。今後の課題としては、画像 DB の変更及び拡張の必要がある。また、各画像ごとに最適なシステムが違っているため、画像ごとに最適なシステムの構築、提案が必要である。

## 参考文献

- [1] Lowe, D.G.: Object recognition from local scale invariant features, Proc. of IEEE International Conference on Computer Vision, pp. 1150-1157(1999).
- [2] G. Csuka et al., "Visual Categorization with Bags of Keypoints", ECCV2004
- [3] Rubner, Y, Tomasi, C, and Guibas, I: The Earth Mover's Distance as a Metric for Image Retrieval, International Journal of Computer Vision, 14(3), pp.130{137, (2000).