

画像の類似検索におけるマルチベクトル特徴空間方式と Earth Mover's Distance 方式の比較

山根 康男 星合 忠 津田 宏 片山 薫 太田 学 石川 博
富士通研究所言語処理研究部 東京都立大学大学院工学研究科

概要 画像の類似検索における代表的な距離である二次形式距離では、似ていない画像が似たものと判定される場合がある。この問題を解決するために、擬ユークリッド空間と斜交基底に基づくマルチベクトル特徴空間方式(MVPO)をすでに提案している。この方式では、画像を複数のベクトルが表す言わば立体で表現する。各ベクトルは画像の個々の特徴に対応する。立体間の距離としてはすでにD距離を提案している。類似した代表的な方式に Earth Mover's Distance (EMD) があり、二次形式距離を含めた他の方式より実験的に精度がよいと言われている。この EMD が MVPO における距離として定式化できることを示し、二次形式距離よりも優れている 1 つの根拠を示す。また D 距離との違いを述べる。

Comparison Between Multi-Vector Feature Space Method and Earth Mover's Distance Method In Similarity Searches of Images

Yasuo Yamane Tadashi Hoshiai Hiroshi Tsuda Kaoru Katayama Manabu Ohta Hiroshi Ishikawa
Language Processing Laboratory Graduate School of Engineering
Fujitsu Laboratories Ltd. Tokyo Metropolitan University

ABSTRACT There are some cases where dissimilar images are judged to be similar in using quadratic-form distance, a representative one in similarity searches of images. To address this problem, we proposed a multi-vector feature space method based on pseudo-Euclidean space and an oblique basis (MVPO). In this method, an image is represented by a solid consisting of multiple vectors; each vector corresponds to each feature. We also proposed D-distance as a distance between solids. A representative method similar to ours is Earth Mover's Distance (EMD), which is said to experimentally outperforms others including quadratic-form distance in precision. We show that EMD can be formalized as a distance in MVPO, and a reason why it outperforms quadratic-form distance. We also mention the difference between D-distance and EMD.

1. はじめに

画像の検索では、Google のイメージ検索のようにキーワード指定による検索が行なわれている。ただし、キーワードでは、ユーザが望む画像を十分に表現できない、キーワードを考えて指定するのが面倒といった問題がある。そこで、ユーザが指定した画像に類似したものを検索する類似検索への期待が高まっている。類似検索では、画像から自動的に抽出する色や形、テキストチャなどの特徴量と呼ばれる数値データに基づいて検索を行なう。

二次形式距離とその問題点

類似検索では検索速度とともに精度の向上が

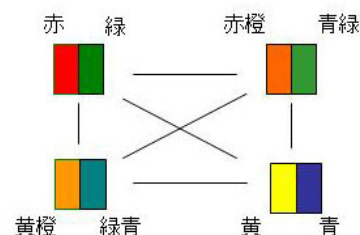


図1.1 識別性の消失
各線は距離が0であることを示す。

主要な課題である。類似検索では、画像は一般にユークリッド空間内の特徴ベクトルと呼ばれるベクトルで表現され、2つのベクトル間の距離が近いほど似ていると考える。しかし、単純にユークリッド距離で測ると、人間には似ているものが似ていないと判定される場合がある。この問題を

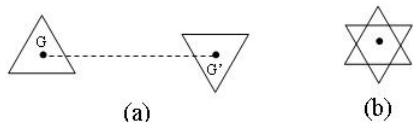


図1.2 三角形の間の距離

解決するために、各特徴間の類似性を距離に反映させる研究がなされてきた。その代表的なものが二次形式距離[Hafner95]である。

ただし、二次形式距離では、逆に似ていないものが似たもの、あるいは同一のものと判定される場合がある。図 1.1 はこの問題を示す図である。図 1.1 の4つの画像は、後で述べる条件のもとで、みな同一のものと判定されてしまう。これは人間の感覚とは明らかに違う。この現象を「識別性の消失」と呼ぶことにする。

マルチベクトル特徴空間への動機

三角形を例にして比喩的に説明することにする。図 1.2(a)における2つの三角形は離れている。これらの三角形の間の距離を、それぞれの重心 G , G' 間の距離として近似することにはさほど異論はないであろう。しかし、図(b)のように重心が重なった場合あるいは近い場合はどうか。このように三角形の大きさに比べ、重心が近くなると、重心間の距離では三角形の空間的な違いを表しているとは言えなくなる。マルチベクトル特徴空間を考えた動機はこの例に似ている。

識別性の消失の問題は、後述するように、この重心が重なる場合に対応している。二次形式距離は重心間の距離にあたる。図形としては重なっていないにもかかわらず、重心間の距離が0となるため、同じ画像と判定されてしまうのである。これが「識別性の消失」の原因である。

マルチベクトル特徴空間

この問題を解決するために、我々は擬ユークリッド空間と斜交基底に基づくマルチベクトル特徴空間方式をすでに提案している[Yamane04]。この特徴空間を、MVPO (multi-vector feature space based on pseudo-Euclidean space and oblique basis)と呼ぶことにする。この方式では、図 1.3 に示すように、画像は「ベクトル集合」と呼ぶ単一特徴ベクトルの集合として表される。

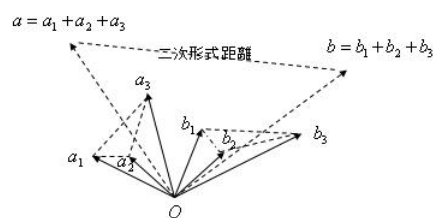


図1.3 MVPOにおけるベクトル集合

「単一特徴ベクトル」とは、画像の個々の特徴(たとえば、赤い色の割合)を表すベクトルである。

図 1.3 では、MVPO における2つのベクトル集合 $A = \{a_1, a_2, a_3\}$, $B = \{b_1, b_2, b_3\}$ の例を示している。 A , B の単一特徴ベクトルは、それぞれ a_i , $b_i (i=1,2,3)$ である。 a_i や b_i は後で説明するように、単一特徴間の類似性を反映するように、すなわち似た単一特徴間ほど近くなるように方向が決められている。2つのベクトル $a = a_1 + a_2 + a_3$ と $b = b_1 + b_2 + b_3$ は従来の特徴ベクトルであり、MVPO はこれらの間の距離が二次形式距離になるという性質を持っている。また、 a は a_1, a_2, a_3 の重心と同様なものと考えられる。というのは、 a に次元数の逆数を掛けると重心と一致するからである。 b も同様である。すなわち、 a や b はそれぞれベクトル集合 A , B の近似であり、画像を正確に表すのは A や B のベクトル集合、幾何的に言えば、複数のベクトルが表す言わば立体と考えられる。ベクトル集合間の距離としてはいろいろな候補が考えられるが、その1つとしてD距離と呼ぶものをすでに提案している。

Earth Mover's Distance

もう一つの代表的な距離に Earth Mover's Distance (EMD) [Rubner98]があり、画像の色やテキスチャに適用されている。EMD は5章で説明するように、MVPO 同様、集合間の距離に基づくものである。また、二次形式距離を含めた他の方式よりも良い精度を出している[Rubner01]。しかし、[Levina01]によれば、EMDの正当性を理論的に示すものはほとんどなかった。EMDは基本的に部分マッチに基づいている。そして輸送問題のアルゴリズムによって計算される。

この論文では、EMD を MVPO における1つの距離として定式化する。このことにより、



図 2.1 色相環

二次形式距離と EMD の幾何的な関係が明らかになる。そして、なぜ EMD が二次形式距離よりも精度的に優れているかの根拠の 1 つが説明できる。また、EMD と D 距離の違いを比較できる。

擬ユークリッド空間

あまり馴染みのない概念なので、説明を加えておく。この論文では、 n 次元の擬ユークリッド空間を $R^m \times iR^{n-m}$ ($0 \leq m \leq n$) として定義する。 R は実数の集合、 i は虚数単位である。 $m = n$ の場合、通常のユークリッド空間となる。擬ユークリッド空間での距離は、見かけ上ユークリッド距離を定義する式と全く同じ式で定義される。ただし、性質は一般にユークリッド距離とは異なるため、この論文では擬ユークリッド距離と呼ぶことにする。擬ユークリッド空間はパターン認識の分野や特殊相対論で時空を表すために用いられている。

2. 基本事項

2.1 色

以降、画像の色、特に色ヒストグラムを用いて論じる。その際、色の違いが図 2.1 に示す色相環の色の間の距離に比例して認識されるものとする。色や色の名前は覚えやすさを優先させて通常のものとは若干違えてあるので注意されたい。

色ヒストグラムは、1 つの画像の色の分布を表す。画像の色は n 色に分類され、色ヒストグラムの各柱が各色の特徴量を示す。色の特徴量としては、特に断らない限り、各色のピクセル数の画像全体のピクセル数に対する割合を考えることにするが、各色のピクセル数も使われる。通常、 n はたとえば 64 というように比較的大きな値であるが、説明をわかりやすくするため、例では小さい値を用いる。

また、以下の記法を用いる。 $\langle C \rangle$ は単一の色 C からなる画像を表す。たとえば、 $\langle \text{赤} \rangle$ は赤一色からなる画像を表す。 $\langle C_1, C_2 \rangle$ は互いに補色の関係にあるピクセル数が同じ 2 つの色 C_1, C_2 からなる画像を表す。 $\langle \text{赤}, \text{緑} \rangle$ は、半分が赤で、もう半分が緑である画像を表す。

2.2 基底

この節では、類似性の観点からベクトル空間における基底について説明する。

直交基底

1 つの画像の特徴量を実数 c_1, c_2, \dots, c_n で表す。たとえば、 c_1 は赤色の割合、 c_2 は赤橙色の割合といった具合である。 n を特徴空間の次元と呼ぶ。

このとき、画像は特徴ベクトル $f = (c_1, c_2, \dots, c_n)$ で表せる。ここで、 $'v$ はベクトル v を転置したベクトルを表す。ベクトルとしては、列ベクトルを使うことにする。正規直交基底 (今後単に直交基底と呼ぶ) を e_1, e_2, \dots, e_n とすると、特徴ベクトル f は特徴量と基底ベクトルの線形結合として、

$$f = c_1 e_1 + c_2 e_2 + \dots + c_n e_n \quad (2.1)$$

とも表せる。各 e_i を基底ベクトルと呼ぶ。 e_i は i 番目の単一特徴に対応している。図 2.2 に直交基底による特徴ベクトルの例を示す。この例では、緑、青緑、赤の特徴量がそれぞれ 0.3, 0.2, 0.5 の場合の特徴ベクトルを表している。

(2.1) による特徴ベクトル間の距離を、ユークリッド距離で定義すると、人間の目には似ている画像が似ていないと判断されるという問題が起きる [Jean90, Hafner95]。図 2.2 でこのことを説明する。ここでは 3 つの色だけを考える。基底ベクトル e_1, e_2, e_3 はそれぞれ緑、青緑、赤に対応する。このとき、3 つの単一色の画像 $\langle \text{緑} \rangle$, $\langle \text{緑青} \rangle$, $\langle \text{赤} \rangle$ の特徴ベクトルは、 e_1, e_2, e_3 とそれぞれ一致する。2 つの基底ベクトル間の距離はどれも

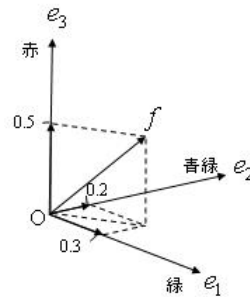


図 2.2 直交基底による表現

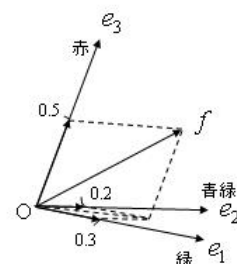


図 2.3 斜交基底による表現

$\sqrt{2}$ であるが、人間には<緑>は<赤>よりも<青緑>に似て見える。このように、問題が起きる原因は、類似度によらず、基底ベクトル間の距離が同じであるためとも考えられる。

斜交基底

図 2.3 は斜交基底を用いて、図 2.2 と同じ特徴量をもつ特徴ベクトルを表したものである。この図に示すように、もし2色間の類似性を反映させた斜交基底を用い、特徴ベクトルを(2.1)によって定めれば、上記の3つの単一色の画像の特徴ベクトルを人間の感覚に合うようにできる。

2.3 二次形式距離

まず類似行列と呼ばれる正方行列 S を

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix}$$

とし、各要素 s_{ij} は i 番目と j 番目の単一特徴間の類似性を表すものとする。 $s_{ij} = s_{ji}$ であるので、 S は対称行列である。このとき、ベクトル v, w 間の二次形式距離 $d(v, w)$ は

$$d(v, w)^2 = (v - w)^t S (v - w) \quad (2.2)$$

と定義される。 S が単位行列の場合、ユークリッド距離に等しくなる。

図 1.1 で示した識別性の消失の問題は、類似行列が図 2.1 に示す色相環における距離関係に基づくときに起きるものである。

3. MVPO の準備

この章では、MVPO を説明するために必要な斜交基底と識別性の消失の原因が単一特徴ベクトルから特徴ベクトルを合成することにあることを説明する。

3.1 斜交基底

斜交基底が満たすべき条件

まず、直交基底と同様に次のことを斜交基底に対して要請することにする。

- (C1) $\|e_i\| = 1 \quad (1 \leq i \leq n)$
- (C2) e_1, e_2, \dots, e_n は線形独立
- (C3) $(e_i, e_j) = s_{ij} \quad (1 \leq i \leq j \leq n)$

$\|v\|$ はベクトル v のノルムを表し、 (v, w) はベクトル v と w の内積を表す。条件(C1), (C2)は直交基底が満たすべき条件と同じである。

直交基底の場合と同様に、特徴量を c_1, c_2, \dots, c_n とし、斜交基底を e_1, e_2, \dots, e_n としたとき、特徴ベクトルを、特徴量と基底ベクトルの線形結合として、(2.1)と同じ式で表すことにする。

ここで重要なことは、こうして定義した特徴ベクトル間の擬ユークリッド距離は、二次形式距離に等しくなるということである[Yamane04]。このことは、二次形式距離に幾何的イメージを与えてくれる。我々が斜交基底を使う理由も、このためである。また、MVPO では単一特徴ベクトルを表す基盤となるという重要な役割を果たす。

なお、類似行列 S に対応する斜交基底の求め方については、[Yamane04]を参照されたい。

3.2 識別性消失の原因

ここでは、識別性消失の原因が単一特徴ベクトルの合成にあることを示す。

例 3.1 2つの補色からなる画像

円を 12 等分する各点に、図 2.1 の 12 色を順に対応させることにする。このとき、これらの点の間の距離関係を反映した類似行列は以下に示すものとなる。まず類似行列が Toeplitz 行列であることに注意されたい。Toeplitz 行列とは左上から右下へ斜めに行列の要素を見たときに、同じ値がならぶ行列のことである。したがって、類似行列を定義するには、第一行の要素を定義すればよく、その値は次のとおりである。

$$s_{1j} = \frac{1}{2} (1 + \cos(\frac{\pi}{6} (j - 1)))$$

ここで、円の直径、すなわち $d(e_i, e_{i+6})$ は $\sqrt{2}$ で

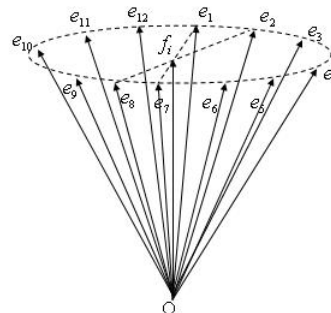


図 3.1 12色の斜交基底と補色の合成

あると仮定している。図 2.1 における色を順に C_1, C_2, \dots, C_n とする。このとき、画像 $\langle C_i, C_{i+6} \rangle$ に対応する特徴ベクトルは、

$$f_i = 0.5e_i + 0.5e_{i+6} \quad (1 \leq i \leq 6) \quad (3.2)$$

と表される。図 3.1 は多次元の擬ユークリッド空間における各基底ベクトル間の距離関係のイメージを表したものである。 f_i と $f_j (i \neq j)$ 間の距離は 0 となる。ただし、図 3.1 では一致しているように描いているが、斜交基底が線形独立であるため、 f_i と $f_j (i \neq j)$ はベクトルとしては一致しない。二次形式距離ではこういうことが起きる。正確には、類似行列の固有値に負または 0 のものがある場合である。重心間の距離が 0 になることの証明は省くが、(2.2) および Toeplitz 行列の性質、関係式

$$s_{ij} = (2 - d(e_i, e_j)^2) / 2$$

を用いて証明できる。

一方、特徴ベクトル f_i は 2 つの単一特徴ベクトル $0.5e_i$ と $0.5e_{i+6}$ を合成したものである。 $0.5e_i$ と $0.5e_{i+6}$ に注目すると、これらのベクトルは単一特徴間の類似性も特徴量も、また識別性も保持している。したがって、識別性の消失が起きる原因は、単一特徴ベクトルから特徴ベクトルを合成する操作(2.1)にあると言える。

4. MVPO

[Yamane04]で提案した MVPO について説明する。識別性消失の原因が単一特徴ベクトルの合成にあることから、合成前のベクトル集合に注目し、これが画像を表しているものと考えるのが、基本的な考え方である。

ここで、ベクトル集合間の距離をどう定義すべきかが問題となる。その 1 つとして、[Yamane04]で D 距離を提案した。ただし、その定義は数学的で直感的にはわかりにくいきらいがある。そこで、EMD の輸送コストによる定義(5章参照のこと)にならって、立体の変換コストという別の観点から再定義し、わかりやすい説明を試みることにする。なお、以降 $d(v, w)$ はベクトル v, w 間の擬ユークリッド距離を表すものとする。

D 距離の立体間の変換コストによる定義

ベクトル集合 A, B を考える。 A, B は、それぞれ原点から伸びる何本かのベクトル (有向線分)

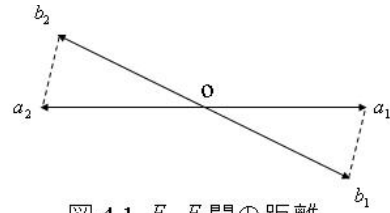


図 4.1 F_1, F_2 間の距離

からなる構造物あるいは立体とも考えられる。この立体もそれぞれ A, B と呼ぶことにする。このとき、 A, B 間の D 距離を次のように定義する。立体 A を分解して立体 B を組み立て直すことを考える。この作業にかかる最小コストが D 距離である。ただし、各操作にかかるコストは次のように定め、そのコストの総和を D 距離とする。

(D1) コスト不要の操作

ベクトルの分割や平行移動、同方向を向いた 2 ベクトルの合成にはコストはかからない。

(D2) ベクトル変形のコスト

ベクトルの伸縮、方向の変更 (回転)、あるいはこれらの操作を同時に行う場合、ベクトルの始点を固定して行うものとし、終点の動く距離をそのコストとする。

従来の特徴ベクトル間の距離は終点間の距離であり、その考えが(D2)に反映されている。ある状態を別の状態に変換するという EMD の考え方は、集合間の距離を定義する場合に、わかりやすく、また 2 つのベクトル集合の要素数の違いに対応できるという柔軟性を与え、有効と思われる。

この定義に基づいて、[Yamane04]におけるいくつかの簡単な例を説明する。

例 4.1 2 つの補色からなる画像への適用

例 3. 1 について考える。次のようなこれらの単一特徴ベクトルの集合を考える。

$$F_i = \{0.5e_i, 0.5e_{i+6}\} \quad (1 \leq i \leq 6)$$

まず、 F_1 と F_2 の間の D 距離を考える。 $a_1 = 0.5e_1, a_2 = 0.5e_7, b_1 = 0.5e_2$ および $b_2 = 0.5e_8$ とすると、 $F_1 = \{a_1, a_2\}, F_2 = \{b_1, b_2\}$ と表せる。これらは図 4.1 に示されている。上記のコストの定義から、 a_1 を b_1 に、 a_2 を b_2 に変換するのが、最小コストとなる。すなわち、 F_1 と F_2 の間の D 距離 $D(F_1, F_2)$ は $D(F_1, F_2) = d(a_1, b_1) + d(a_2, b_2)$ となる。任意の i, j に対しては、

$$D(F_i, F_{i+j}) = 2\sqrt{1 - |\cos(\pi j / 6)|}$$

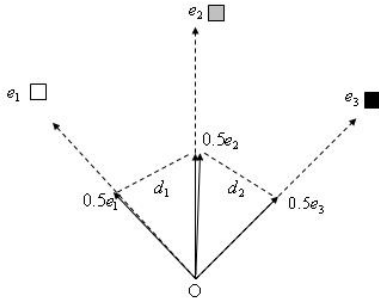


図4.2 等しい分割の例

であるから、

$$\begin{aligned} D(F_1, F_2) &= 0.732, \delta(F_1, F_3) = 1.414, \\ D(F_1, F_4) &= 2, D(F_1, F_5) = 1.414, \\ D(F_1, F_6) &= 0.732. \end{aligned}$$

である。ここで、 $|a|$ は a の絶対値を表している。これらの D 距離の値を見ると、識別性が保たれ消失の問題が解決されているだけでなく、値自体も適切なものと思われる。

例 4.2 ベクトルの分割が必要な場合

マンセルの色立体のような色立体において、白、黒、灰の3色について考える。色立体を地球にたとえれば、白、黒、灰はそれぞれ、北極、南極、地球の中心に対応する。すなわち、これらの3色は直線上に並ぶ。この論文では、赤、緑といった補色同士や白、黒といった対極にある色同士は独立した、すなわち基底ベクトルが直交する特徴と仮定している。したがって、白と黒の間の距離は $\sqrt{2}$ であり、白と灰、灰と黒の間の距離はともに $\sqrt{2}/2$ となる。これらの距離関係を表す類似行列は次のとおりである。

$$S = \begin{pmatrix} 1 & \frac{3}{4} & 0 \\ \frac{3}{4} & 1 & \frac{3}{4} \\ 0 & \frac{3}{4} & 1 \end{pmatrix}. \quad (4.4)$$

対応する基底ベクトルは次のように求められる。 e_1 が白、 e_2 が灰、 e_3 が黒に対応している。

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} \frac{3}{4} \\ \frac{\sqrt{7}}{4} \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ \frac{3\sqrt{7}}{7} \\ \frac{\sqrt{14}}{7}i \end{pmatrix}$$

純虚数が e_3 に現れている。このような基底ベクトル間の直線的な距離関係はユークリッド空間では表せない。これが擬ユークリッド空間を

用いる大きな理由である。擬ユークリッド空間では任意の類似行列に対応できる。

$F_1 = \{0.5e_1, 0.5e_3\}$, $F_2 = \{e_2\}$ とする。すなわち、2つの画像<白, 黒>、<灰>を考える。2つの集合の要素の数は異なっている。そこで、 e_2 を $0.5e_2$ と $0.5e_2$ に分割する。このとき、 $0.5e_1$ を $0.5e_2$ に、 $0.5e_3$ を $0.5e_2$ に変換するのが最小コストとなる。すなわち、

$D(F_1, F_2) = d(0.5e_1, 0.5e_2) + d(0.5e_3, 0.5e_2)$ である。この様子を示したのが、図 4.2 である。この場合が最小であることは、次の例 4.3 で示す方法と同様の方法で示される。

例 4.3 長さが等しくない分割の例

$F_1 = \{0.7e_1, 0.3e_3\}$, $F_2 = \{e_2\}$ とする。図 4.3 に示すように、 e_2 が灰、 e_1 が灰よりわずかに白よりの色、 e_3 が灰よりわずかに黒よりの色を表しているものとする。すなわち、 e_1, e_3 と e_2 は大変近く、かつ $d(e_1, e_2) = d(e_2, e_3)$ とする。したがって、対応する画像は極めて似ていることになる。すなわち F_1, F_2 間の距離は 0 に近くあるべきである。しかし、例 4.2 のようにベクトルを二等分すると、変換にかかるコストは $d(0.7e_1, 0.5e_2) + d(0.3e_3, 0.5e_2)$ すなわち、約 0.4 になってしまう。この場合、図 4.3 に示すように、 e_2 を $0.7e_2$ と $0.3e_2$ に分割し、 $0.7e_1$ を $0.7e_2$ に変換し、 $0.3e_3$ を $0.3e_2$ に変換する場合が最小コストとなる。そして、 D 距離は

$D(F_1, F_2) = d(0.7e_1, 0.7e_2) + d(0.3e_3, 0.3e_2)$ となる。この値は、図 4.3 からわかるように 0 に近い値となり、人間の感覚に合うものとなる。なお、図 4.4 は e_2 を 0.7 対 0.3 に分割するときが最小となることを示す図である。点 C' は $\angle OBC' = \angle BOC$ かつ $BC' = OC$ となる点であ

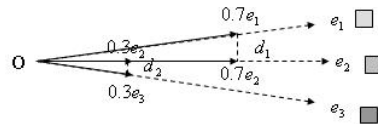


図4.3 等しくない分割の例

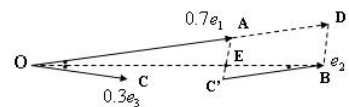


図4.4 最小距離となる理由

る。 $AE + EC'$ が変換コストとなるので、 A, C を直線で結んだときの OB との交点を E としたときが最小コストとなる。このとき、 AD の長さは0.3であり、 AC' と DB が平行であることに注意すると、 E が OB を0.7対0.3に分割していることがわかる。この方法は、光の反射の経路を求める方法と同様である。

5. D 距離と EMD との比較

EMDはMVPOの1つの距離として定式化できることを示す。これは、EMDがベクトル集合間の距離であることを意味し、このことがEMDが二次形式距離より優れている1つの根拠と考えられる。また、定式化したEMDとD距離を比較する。

5.1 EMD の説明

EMDの定義を、[Rubner98]にある土盛りと穴の比喩を使って簡単に説明することにする。2つの画像に対応するシグニチャ(signature)と呼ばれる距離の定められている空間内の点と特徴量の対の集合をそれぞれ、

$$x = \{(p_1, x_1), (p_2, x_2), \dots, (p_s, x_s)\}$$

$$y = \{(q_1, y_1), (q_2, y_2), \dots, (q_t, y_t)\}$$

とする。 x が土盛りの集合、 y が穴の集合に対応する。 s と t は異なっても構わない。この柔軟性がEMDの特徴の一つである。図5.1は $s=3, t=2$ の場合を示している。点 p_i には x_i の量の土が盛られ、点 q_j には y_j の大きさの穴が掘られている。なお、土の総量は穴を埋めるために十分あるものとする。 p_i, q_j 間の距離を d_{ij} とする。このとき、穴を全部埋めるために土をどう運んだら最小コストですむかというのがEMDの概略である。正確には、次のとおりである。

f_{ij} を p_i から q_j へ運ばれる土の量とする。このとき、全部の穴を埋めるための輸送コスト

$$\sum_{i=1}^s \sum_{j=1}^t d_{ij} f_{ij}$$

を最小にする f_{ij} を求め、 x, y 間のEMDを

$$EMD(x, y) = \frac{\sum_{i=1}^s \sum_{j=1}^t d_{ij} f_{ij}}{\sum_{i=1}^s \sum_{j=1}^t f_{ij}}$$

と定義する。分母は正規化のためにあり、総量が少ないシグニチャが選ばれやすくするのを防ぐ。

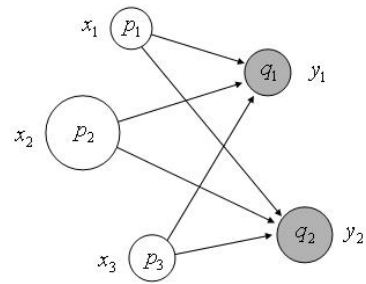


図5.1 Earth Mover's Distance

特徴量を各色のピクセル数とした場合は、2つのシグニチャの総量が異なりうる。総量が異なる場合は、EMDは部分マッチにおける距離を表す。そして、この場合、距離が0になった場合は、総量の小さい方の画像が大きい方の画像に部分的に完全にマッチすることを意味する。一方、D距離は総量のいかんにかかわらず、画像全体同士の全体マッチにおける距離を表す。

5.2 EMD の MVPO による定式化

EMDにおける空間内の点(5.1における p_i, q_j)全体の集合を $R = \{r_1, r_2, \dots, r_n\}$ とする。また、 r_i, r_j 間の距離を改めて d_{ij} とする。もし、距離が定義されていない場合は、0ではない適当な正の値を定めるものとする。そして、各 r_i に斜交基底ベクトル e_i を対応させる。このとき、 d_{ij} によって決まる類似行列を e_i を定めるために用いる。そして、シグニチャ x, y を

$$x = \{(r_1, x_1), (r_2, x_2), \dots, (r_n, x_n)\}$$

$$y = \{(r_1, y_1), (r_2, y_2), \dots, (r_n, y_n)\}$$

と表す。 r_i に対応する p_k がない場合は、 $x_i = 0$ とし、同様に、対応する q_k がない場合は、 $y_i = 0$ とする。そして、 x, y に対応するベクトル集合

$$X = \{x_i e_i \mid 1 \leq i \leq n\}$$

$$Y = \{y_i e_i \mid 1 \leq i \leq n\}$$

を考える。このとき、D距離の定義と同様、EMDをベクトル集合が表す立体間の変換として次のように定義しなおすことができる。EMDの場合は X が表す立体を分解して、 Y の一般には全体ではなく、部分を構築するコストとなる。

(E1) コスト不要の操作

D距離の場合と全く同様である。

(E2) ベクトルの変形にかかるコスト

ベクトルの伸縮はできない。方向の変更(回転)はでき、コストはD距離と同様、始点を固定した際の終点の移動距離で定める。

すなわち、D距離とは異なり、ベクトルの変形に

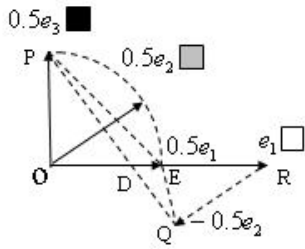


図5.2 D距離とEMDの比較

関しては等長変換だけが許されている。

5.3 D距離とEMDとの違い

今までに述べてきたことからD距離とEMDの違いを以下にまとめる。

- (1) EMDが部分マッチを基本としているのに対して、D距離は全体マッチを基本としている。
- (2) EMDはベクトルの変形において、方向を変えることは許されているが、長さを変えることはできない。これに対して、D距離では、長さも方向も変えることができる。
- (3) EMDには、正規化の因子で割るという操作が付加されている。

次に部分マッチと全体マッチが一致する場合、すなわち、特徴量の合計が2つのベクトル集合の間で等しい場合には、D距離とEMDを数量的にも比較することが可能となる。この場合について以下考える。今、一般性を失うことなく、特徴量の合計が1であると仮定する。この場合の違いは、(D2)と(E2)のみとなる。EMDの方がベクトルの変形に関する制限が厳しいので、一般には、D距離の方が数値的には小さいことになる。このことを示すのが、図5.1である。

2つのベクトル集合 $A = \{1.0e_1\}$ と $B = \{0.5e_2, 0.5e_3\}$ を考える。 e_1 は白、 e_2 は灰、 e_3 は黒に対応している。これら3つの基底ベクトルは、図5.1に示すように、距離 $d(e_1, e_3)$ 、 $d(e_1, e_2)$ を保つように、平面上に描くことができる。このとき、 A, B 間のD距離は、 $PD + DQ = PQ$ である。一方、 A, B 間のEMDは $PE + EQ$ である。ここで、点Dは線分PQとORの交点、点Eは線分ORの中点である。図からも明らかなように、D距離の方が短い。なお、点Dを求める方法は、例4.3と同様である。

このようにEMDよりもD距離の方が一般に短く、数学的な立体間の距離という意味ではいいかもしれない。しかし、EMDは同じ量あるいは割合で対応させており、その画像における意味づけがわかりやすいのに対し、D距離ではベクトル

の長さ、方向とも変更を許しており、逆に画像における意味づけがはっきりしないきらいがある。

6. まとめ

この論文では、すでに提案しているMVPOおよびD距離をわかりやすく説明し、またEMDがMVPOの一つの距離として定式化できることを示し、EMDが二次形式距離よりも精度において優れている根拠の1つを示した。また、特徴量の総量と同じ場合、D距離とEMDの違いがベクトルの変形部分だけであることを示した。

D距離とEMDのどちらが優れているかはまだ結論付けられないが、EMDよりもD距離の方がこの論文で示したように短い。ただし、D距離には画像における意味づけがはっきりしないきらいがある。今後の課題としては、この意味づけも含めた評価と、D距離の効率的なアルゴリズムを見つけることが上げられる。なお、文書に対するD距離の適用についても検討する予定である。

謝辞

貴重な助言を頂いた富士通研究所の増本主任研究員、上原さん、馬場さんに謝意を表します。

参考文献

- [Hafner95] James Hafner, et al., Efficient Color Histogram Indexing for Quadratic Form Distance Functions, IEEE Trans. Pattern Anl. Machine Intell. 17(7), pp.729-736, (1995)
- [Jean90] J.S.N. Jean, A New Distance Measure for Binary Images, Proc. IEEE ICASSP '90, 4, pp.3-6 (1990); Paper#: M5.19
- [Levina01] E. Levina, et al., The Earth Mover's Distance is the Mallows Distance: Some Insights from Statistics, Proc. ICCV 2001, pp.251 - 256 vol.2, (2001)
- [Rubner98] Yossi Rubner, et al., A Metric for Distribution with Applications to Image Databases, Proc. IEEE Intl. Conf. On Computer Vision, pp.59-66, (1998).
- [Rubner01] Yossi Rubner, et al., Perceptual Metrics for Image Database Navigation, Kluwer Academic Publishers, (2001).
- [Yamane04] 山根、他 類似検索のためのマルチベクトル特徴空間方式の提案、第15回データ工学ワークショップ (DEWS2004), 2004年3月