

回転方向の自己相関関数を用いた商標図形分類法 A Trademark Classification using Auto-correlation Functions for Rotated Images

餅原 道元†
Dougen Mochihara

佐野 利行†
Toshiyuki Sano

石田 崇‡
Takashi Ishida

平澤 茂一‡
Shigeichi Hirasawa

1. 研究目的

現在、特許電子図書館 [1] には 180 万件を越える商標図形が登録されており、この中から手作業で類似商標を検出するのは非常に困難である。その際、あらかじめ商標図形がカテゴリ分けされていれば、そのカテゴリ内での類似商標の検索を行うことにより類似商標の検出が容易となる。

商標図形の分類手法として、長嶋ら [2] により水平・垂直方向の2つの自己相関関数のグラフの形を近似した類似関数を用いる手法が提案されている。しかし、水平・垂直方向の2つの自己相関関数で分類しているため、図形の形状特徴を表現しきれっていない。また、特許庁の分類結果に基づくカテゴリを用いていない。

そこで本研究では、回転方向の自己相関関数による特徴量を提案する。さらに、提案手法を用いて特許電子図書館で公開されている登録済の商標 [1] に対し、特許電子図書館が提示するカテゴリに基づいて分類したとき、従来手法より分類精度が向上することを示す。

2. 準備

2.1 商標図形の種類

特許電子図書館のデータベースはウィーン図形分類リストに従って 29 のカテゴリに分けられている。本論文では、登録数の多い上位 5 件のカテゴリ（幾何、書体・数字、植物、動物、天体・自然現象・地図）の 2 値画像を対象とし、(1) 幾何 (2) 書体・数字 (3) その他の図形（植物、動物、天体・自然現象・地図）の 3 つのカテゴリに分類する。

2.2 水平・垂直方向の自己相関関数

自己相関関数とは、対象画像とその画像をある相関長（水平・垂直方向にずらした画素数）ずらしたときの黒画素の重なる割合を表す関数である。次式により水平方向の自己相関関数 $R_h(x)$ 、垂直方向の自己相関関数 $R_v(y)$ を定義する。ただし、 x, y は相関長である。

$$R_h(x) = \frac{1}{l_Y(l_X - x)} \sum_{Y=1}^{l_Y} \sum_{X=1}^{l_X - x} g(X, Y)g(X + x, Y), \quad (1)$$

$$R_v(y) = \frac{1}{l_X(l_Y - y)} \sum_{X=1}^{l_X} \sum_{Y=1}^{l_Y - y} g(X, Y)g(X, Y + y). \quad (2)$$

$x(y)$: 水平(垂直)方向の相関長
 $X(Y)$: 水平(垂直)方向の座標
 $l_X(l_Y)$: 外接矩形¹⁾の横(縦)の長さ
 $g(X, Y) \in \{0, 1\}$: 座標 (X, Y) における画素値

3. 従来手法 [2]

3.1 概要

長嶋ら [2] は、水平・垂直方向の自己相関関数に基づいて決定木により商標図形を分類する手法を提案してい

る。ただし、自己相関関数をそのまま用いるのではなく、以下のように近似した類似関数を作成し、類似関数の係数を特徴量として用いている。ただし、 A は振幅、 ω は角速度である。

$$S_1(x) = a_0 e^{-a_1 x}, \quad (3)$$

$$S_2(x) = -A \sin \omega (x - x_1) + b_0 + b_1 x, \quad (4)$$

$$S_3(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3. \quad (5)$$

類似関数を求める際には、相関長の区間を 3 つに分けて、各区間内において類似関数を求める。

3.2 従来手法のアルゴリズム

- a1. 水平方向の自己相関関数 $R_h(x)$ を求める。
- a2. 自己相関関数を最小二乗近似した 1 次式を求める。
- a3. 自己相関関数と a2 で求めた 1 次式の 2 つの交点により、区間を 3 つに分ける。
- a4. 式 (3)~式 (5) を用いて、各区間 $i(i=1,2,3)$ の類似関数 $S_1(x), S_2(x), S_3(x)$ を求める²⁾。
- a5. 類似関数の係数 $a_0, a_1, b_0, b_1, A, \omega, c_0, c_1, c_2, c_3$ の値を特徴量として決定木により商標図形を分類する。□

垂直方向についても同様に求める。

4. 提案手法

4.1 概要

従来手法では、書体・数字図形のうち、横長・縦長の図形が対象であれば、十分な精度を保つことができる。しかし、書体・数字図形には、外接矩形が正方形に近い図形も含まれており、これらの図形は水平方向のグラフの相関長が短くなるため、幾何図形に分類されてしまうという問題点がある。

一方、幾何図形は、外接矩形が正方形に近い図形が多く、図形に含まれている黒画素の割合が多い。そこで幾何図形と書体・数字図形の回転させたときの相関（回転させたときの黒画素の重なる割合）の違いに着目し、回転方向の自己相関関数を考える。

4.2 回転方向の自己相関関数

回転方向の自己相関関数は、図形を回転させたときの相関を表すものである。次式により回転方向の自己相関関数 $R_r(z)$ を定義する。ただし、 r は回転半径、 z は回転角度、 (r_ω, r_h) は図形を回転させる中心座標である。

$$R_r(z) = \frac{1}{\pi r^2} \sum_{Y=1}^{l_Y} \sum_{X=1}^{l_X - z} g(X, Y)g(P, Q), \quad (6)$$

$$P = (X - r_\omega) \cos \frac{\pi z}{180} - (Y - r_h) \sin \frac{\pi z}{180} + r_\omega, \quad (7)$$

†早稲田大学大学院創造理工学研究科

‡早稲田大学理工学術院

¹⁾ 図形を囲む最小の長方形

²⁾ 類似関数の係数を求めるアルゴリズムについては [2] を参照

$$Q = (X - r_w) \sin \frac{\pi z}{180} - (Y - r_h) \cos \frac{\pi z}{180} + r_h. \quad (8)$$

図1に幾何図形と書体・数字図形の回転方向の自己相関関数(実線)とそれを最小二乗近似した一次式(点線)のグラフを示す。図1より、幾何図形は書体・数字図形より回転方向の自己相関値が全体的に大きいことがわかる。

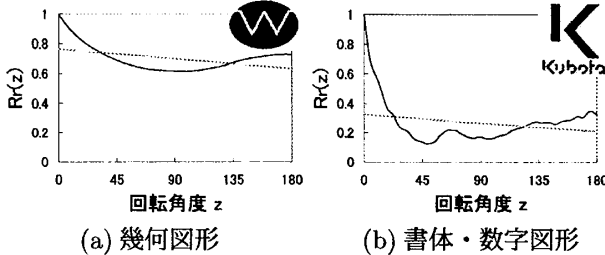


図1: 商標図形と回転方向の自己相関関数の関係

そこで、回転させたときの相関の違いにより(1)幾何(2)書体・数字(3)その他の図形(植物、動物、天体・自然現象・地図)を分類する手法を提案する。

また、従来手法では類似関数を求めているため、情報が損失していると考えられる。そこで、類似関数を用いるのではなく、自己相関関数の標本化・量子化を行う手法を提案する。なお、特徴量が多くなるのを防ぐために標本化を行い、分類には決定木ではなくSVM(Support Vector Machine)[3][4]を用いた。SVMは、教師あり学習を用いる識別手法の一つである。

4.3 提案手法のアルゴリズム

- b1. 従来手法と同様に特徴量を求める。(a1~a4)
- b2. 式(6)~式(8)で回転方向の自己相関関数 $R_r(z)$ を求める。
- b3. (a) 回転方向の自己相関関数を最小二乗近似した1次式を求める。
(b) 標本化・量子化する場合はb3.のステップは省略する。
- b4. (a) a5で得られた特徴量に加え、b3.(a)で求めた1次式の切片・傾きを特徴量として、商標図形を分類する。
(b) 標本化・量子化する場合は水平・垂直方向の自己相関値に加え、回転方向の自己相関値を特徴量として商標図形を分類する。□

5. 実験結果および考察

5.1 実験条件

表1: 実験に用いた商標図形の個数

商標図形のカテゴリ	学習データ	テストデータ
幾何	200	400
書体・数字	100	200
その他	100	200
合計	400	800

実験では、(1)幾何(2)書体・数字(3)その他の図形(植物、動物、天体・自然現象・地図)の3つに分類し、その分類精度を求めた。

実際の登録数の比率を考慮して特許電子図書館[1]から無作為に選んだ。実験に用いた各カテゴリに含まれ

る商標図形は表1の通りである。なお、分類には、決定木の作成にSee5[5]、SVMとしてTinySVM[6]、SVMでの多値分類にはone-vs-rest法を用いた。

5.2 実験結果と考察

表2に正しく分類された割合の実験結果を示す。

表2: 分類精度(再現率)

商標図形のカテゴリ	従来(決定木)	提案(決定木)	提案(SVM-量子化)
幾何	0.78	0.81	0.88
書体・数字	0.57	0.71	0.75
その他	0.57	0.52	0.60
全体	0.67	0.71	0.78

1. 提案手法による改善理由

回転させたときの相関を考慮した結果、幾何図形のカテゴリに誤って分類された書体・数字図形のうち、外接矩形が正方形に近く黒画素の含まれている割合が少ない図形を正しいカテゴリに分類することができた。これらの図形を正しく分類できたのは、回転方向の自己相関値が他の幾何図形と比べて全体的に低かったためである。

また、標本化の間隔を狭くすると精度が向上することを確認できた。

2. 提案手法による誤分類

従来手法で正しく分類された幾何図形が、回転させたときの相関を考慮した結果、書体・数字図形のカテゴリに誤分類された。これらの図形が誤分類されたのは、幾何図形であっても黒画素の含まれている割合が少ない図形が存在するため、回転させたときの相関が低く、回転方向の自己相関値が全体的に低くなったためである。

6. むすびと今後の課題

本研究では、水平・垂直方向の自己相関関数に回転方向の自己相関関数を組み合わせ、(1)幾何(2)書体・数字(3)その他の図形(植物、動物、天体・自然現象・地図)を効果的に分類する手法を提案した。実験結果より、分類精度が全体として0.11向上した。

今後の課題は、植物、動物、天体・自然現象・地図のカテゴリの特徴を考慮し、自己相関関数以外の他の物理的特徴量を用いてこれらのカテゴリを分類する手法の検討である。

謝辞: 本研究の成果の一部は電気通信普及財団の助成による。

参考文献

- [1] 特許電子図書館
<http://www.ipdl.ncipi.go.jp/homepg.ipdl>
- [2] 長嶋秀世, 椿伸一郎, 中島淳, “自己相関関数のグラフの形を利用した商標図形の分類,” 電気学会論文誌C, vol.123, no.9, pp.1547-1554, 2003年9月.
- [3] Corinna Cortes and Vladimir Vapnik, “Support vector networks,” *Machine Learning*, vol.20, no.3, pp.273-297, Mar.1995.
- [4] 山本寛康, 松本裕治, “Support Vector Machineの多値分類問題への適用法について,” 情報処理学会研究報告, 2001-NL-146, no.1, pp.35-49, 2002.
- [5] RuleQuest
<http://www.rulequest.com/index.html>
- [6] TinySVM
<http://chasen.org/taku/software/TinySVM/>