

北海道大学工学部、岡山理科大学* ○早川 洋平, ラシキア・ワクタンク*, 金子 俊一, 五十嵐 悟

要旨

ロボットビジョンなどへの応用を指向して、一般画像内の文字候補領域の切り出しを、認識を経ずに行う手法を開発した。フーリエスペクトルに基づいて基本的特徴量を定義し、それを用いてインスタンス投票法を実現した。基本的実験の結果を報告する。

1. はじめに

ロボットビジョンにおける物体認識機能や画像パターン識別機能の頑健(ロバスト)性, 柔軟性, 操作性などをより向上させるためには, 現実の現象をより忠実にモデル化し, 現象の様々なゆらぎ(撮像条件の変動, 様相変化, 遮蔽, 変形など)にロバストに対応する認識手法を準備する必要がある。更に, 見本提示による学習方式に基づくことによりパラメタ調整などを回避でき, より高い操作性を獲得することが期待できる。我々は「インスタンス投票法」と呼ぶ独自の学習認識アプローチの提案[2]を行っており, これは上記のような目的を指向している。本報告で述べる, インスタンス投票法の応用としての文字領域(候補)の自動抽出は, 画像理解, ビークルナビゲーション, 文字認識などへの広い応用が期待できる基本的認識機能と考えることができる。

2. インスタンス投票法

この方法は元来故障診断などを指向して S.V.Aleshin によって一部定式化された[1]。我々はそれを一般的なパターン認識問題を扱えるように拡張すると共に, 明度ヒストグラムに基づく特徴量により文字抽出を行った[2]。今回は, 次節に詳述するフーリエスペクトルに基づく画像特徴を用いる。厳密な定義および説明は文献[2]に譲る。ここでは基本操作とその意味について説明する。図1に処理の流れを示す。学習訓練用の複数の原画像を入力とし, 特徴抽出処理に入る。この中では, 分割, FFT, 特徴量計算, 量子化を経てコード化に至る。コード化2値化インスタンス(以下, CBI)は0と1を要素とするビット列である。CBIは各々文字(S)あるいは非文字(N)領域の教示情報をもつ。「プライムテスト(Pテスト)」と呼ぶ部分ビット列(不連続でも可)を抽出する。SのいかなるCBIのPテスト部分も, NのCBIの対応部分と同じビット列をもたない。即ち, SとNを区別するために必要十分な局所的な部分ビット列をPテストと定義する。Pテストは複数存在する(Pテスト表)。教師によって与えられた訓練画像群は一般に一様ではなく, 種々の区別の仕方が求められることに対応する。認識対象画像から得られるCBIをPテスト表に照らして照合計算を行う。特徴的なのは, Pテスト表のすべてのCBIにわたって, かつすべてのPテストにわたって照合(投票)を行うことである。対象のCBIのあるPテスト部分がある实例のCBIに一致すれば1票獲得するという具合である。得票を集計しSかNか多い方に識別する。

Pテストは人間の部分的な記憶に似ている。部分的な記憶との照合を投票とみなし, 得票結果を最頻選択(当

選)という形で統合することにより, ロバストな識別メカニズムが得られる[3]。

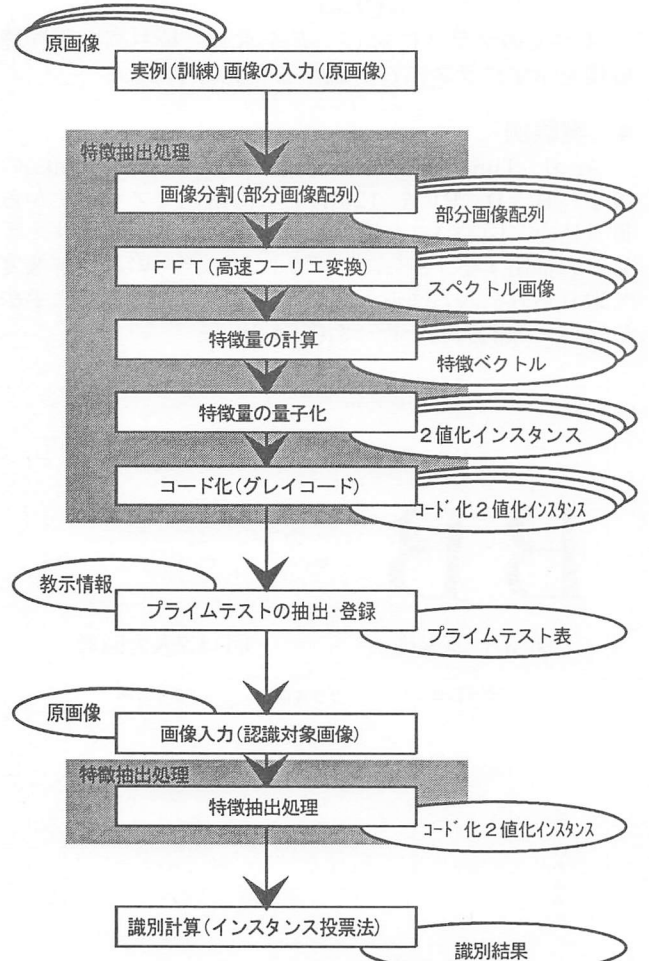


図1 インスタンス投票法による文字領域抽出

3. フーリエ特徴

サイズ $N \times M$ の原画像 $f(x, y)$ を $L \times L$ (実験では $L=64$) の部分画像の配列に分割し, 各部分画像を独立に FFT 処理しスペクトル画像 $F(u, v)$ を得る。フーリエスペクトルは高速計算可能であり, また種々の正規化により不変量とすることができる。次の2種類(4個)の特徴量を定義する。

(1) ピッチ特徴量(横, 縦) C_{pu}, C_{pv} :

$$C_{pu} = \max(A_3^*, \dots, A_8^*)$$

$$\text{ここで、 } A_p^* = \sum_{m=1}^3 A_p(m \cdot u_0), \quad \left(u_0 = \frac{64}{p} \right)$$

$$A_p(u) = \{F(u,0) - F(u-b,0)\}$$

$$- h_1 \cdot \{F(u+b,0) - F(u-b,0)\}$$

$$+ h_2 \cdot F(u,0)$$

$$b = \begin{cases} 1 & ; F(u+1,0) \geq F(u-1,0) \\ -1 & ; F(u+1,0) < F(u-1,0) \end{cases}$$

p : ピッチの大きさ
 h_1, h_2 : 正の定数

C_{pv} についても同様に定義する。

今回は特に漢字を例としたが、拡張は可能であると考えられる。ゴシックフォントなどで顕著であるが、文字は一定間隔に配列した一定幅線分の集合と近似できる。ピッチ特徴量 $C_{pv}(v)$ は一定間隔の配置を検出することを目的とする。

(2) 偏差特徴量 (横, 縦) C_{su}, C_{sv} :

$$C_{su} = \sqrt{\frac{\sum_{u=-32}^{31} \{u^2 \cdot F(u,0)\}}{\sum_{u=-32}^{31} F(u,0)}}$$

C_{sv} についても同様に定義する。

壁などの大きな一様背景に比べて、文字部分は局所的で一般に高周波数成分を多く含んでいる。この分散特徴量 $C_{su}(v)$ は周波数成分にわたる散布度の違いを調べている。

一般に「文字パターンは様々な形状、サイズ、姿勢をもつ」と考えられているが、我々の周囲に注意すると、文字と隣接背景間にはある種の規則がある。サイズは様々であるが、姿勢は鉛直方向が殆どであり、これは文字の目的は本来可読性にあることによる。

4. 実験

図 2 (a) は部分画像配列 (実画像), 同 (b) は対数変換したスペクトル画像を示す。文字領域のスペクトルの広がりが見える。表 1 は訓練画像の CBI (グレイコード) である。幾つかの P テストは確認できるが、完全な P テスト表は現在作成中である。



図 2 (a) 部分画像配列 (実画像) [4]

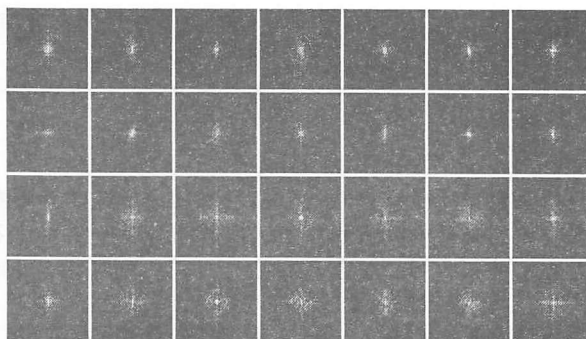


図 2 (b) 対数変換したスペクトル画像

表 1 訓練画像の CBI (グレイコード)
(a) 非文字領域 (b) 文字領域

i	Cpy	Csy	Cpx	Csx	i	Cpy	Csy	Cpx	Csx
01	011	100	000	110	16	100	100	111	010
02	011	111	001	101	17	111	101	100	100
03	001	110	000	001	19	101	111	010	111
05	010	101	001	000	20	111	110	111	100
08	000	011	011	111	21	100	111	001	010
09	011	001	011	101	23	110	111	110	111
11	011	111	000	000	24	100	111	001	110
13	000	000	001	111	28	010	110	011	101
14	000	011	001	111					
15	111	101	000	000					

i: 画像番号

(注) 上記は代表的な部分画像のみの CBI. 左上の部分画像が $i=1$ で右方向に向かって $i=2, 3, \dots$ と定義する。

5. まとめ

フーリエ特徴を用いたインスタンス投票法による、文字候補領域の自動抽出手続きを提案した。実画像を用いた処理において、P テストを確認した。今後は P テスト表の作成、投票手続きの実現を行わなければならない。更に特徴量の増加および質の向上、可変部分画像の定義などを進める予定である。

参考文献

- [1] S.V. Aleshin: "Metriization of Discrete Procedure of Recognition", Trudi Seminara po Diskretnoi matematike i eyo prilijenyam, Moscow (1989).
- [2] 馬場直人, ラシキア・ワクタング, 金子俊一, 本多庸悟: "インスタンス投票法による文字領域の抽出", 第 2 回画像センシングシンポジウム講演論文集, 31-36 (1996).
- [3] P.J. Rousseeuw et al.: "Robust Regression and Outlier Detection", John Willy & Sons, New York (1987).
- [4] DOS/V マガジン vol.5, 82 (1996)