

2次元フーリエ変換を用いた文書画像の領域判別法

長谷雅彦

(日本電信電話公社)

星野坦之

横須賀電気通信研究所)

1. まえがき

データ処理技術の発展に伴い、印刷文書から大量のデータを入力することが必要となってきている。印刷文書からデータを入力するためには、印刷文書から、文字領域、図形領域、写真領域を切り分けること、次に、文字領域の中での文字列の切り出し、文字列からの文字の切り出し、および、その文字を認識することが必要である。

現在、文書画像内の文字、図形、写真領域の切り分け方法については、種々の方法が提案されており、⁽¹⁾⁽²⁾⁽³⁾ 今後も各種方法が模索されると考えられる。

本報告では、印刷文書画像中の文字領域、図形領域、写真領域の判別を目的として、レンズ系により光学的にも処理可能である2次元フーリエ変換による方法を試み、その有効性を確認したので以下に報告する。

2. 文書画像の領域判別処理系

2.1 文書画像の特徴

一般の印刷文書画像は、文字領域の部分、図形・表領域部分、および写真

領域部分が複雑に組みあわされて、構成されている。文字領域は、本文および、文字列ピッチの異なるあらまし、図表説明文から構成され、文字列群、文字列、文字部分と階層構造をとる。

図形領域は、一般にグラフ部分および表の部分からなり、それぞれ文字部分を含んでいる。また、写真等の部分は多値の濃淡画像であることが多い。

(図1参照)

図2に、一般的文書画像の例を示す。図中の文字領域、図形領域、写真領域を判別するためには、次に述べる知識を用いることが有効であると考えられる。

- (1) 文字領域では、一般に黒画素が行列に対応する周期を有する。
- (2) 図形領域では、黒画素の密度が比較的少なく、周期性は少ない。
- (3) 写真領域は、一般的に黒画素のある領域での密度が、文字、図形領域に比較して大きい。
- (4) 図形の説明文の中の文字のピッチは、文字領域におけるピッチと異なることが多い。

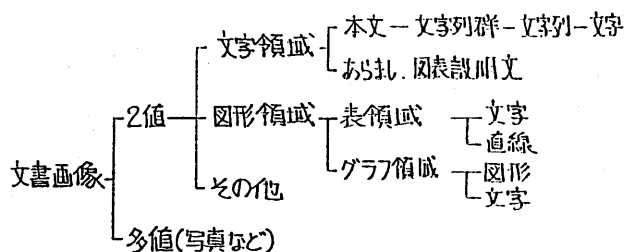


図1. 文書画像の階層構造



図2. 文書画像の一例

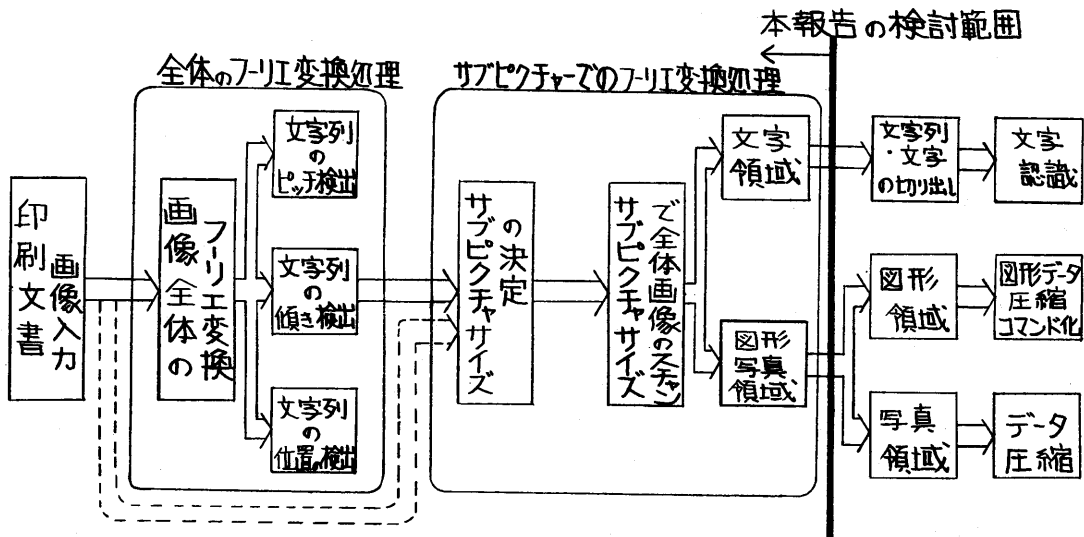


図3 印刷文書画像の領域判別処理系

2.2 文書画像の領域判別処理系

図3に、文書画像の領域判別処理系の概略を示す。まず文書画像全体の2次元フーリエ変換を行い、その結果を用いて文字列のピッチ、文字列の傾き、行数、文字列が存在する位置を検出する。次に、文字列のピッチの検出結果より、最適なサブピクチャのマトリックスサイズを決定する。このマトリックスサイズで画像全体をスキャンする際に、2次元フーリエ変換処理を行い、変換面でのピーク値の有無によって、文字領域と図形・写真領域の判別を行う。

文字領域は、その後、文字列の切り出し、さらに文字の切り出しを行い、最終的に、文字認識処理を行う。図形領域、写真領域は濃度レベル等により切り分けを行い、図形領域はコマンド化、およびデータ圧縮処理、また、写真領域も圧縮処理を行う。

文字列の角度、ピッチ等の情報が既知の場合(2枚目の文書画像処理)には、サブピクチャのみを用いての処理方式も考えられる。

3. 全体の2次元フーリエ変換処理⁽⁴⁾

図3に示した全体画像の2次元フーリエ変換処理での、各部の検出方法とその結果について述べる。

3.1 2次元フーリエ変換

2次元フーリエ変換は、画面上の各点の濃度情報から、その画面上の周期性に対応した情報を抽出するものであり、画面上の座標 (x, y) の濃度分布を $f(x, y)$ とすると、ディスクリートな2次元フーリエ変換は次のようにあらわされる。

$$F(u, v) = \frac{1}{N} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(x_i, y_j) \exp[-j2\pi(ux_i + vy_j)] \quad (1)$$

ただし、 u, v はそれぞれ x, y 方向の空間周波数、 N は文書画像の x, y 方向の画素の分割数、 $x_i = (i/N)\ell$ 、 $y_j = (j/N)\ell$ であり、 ℓ は画像の大きさである。 u, v はそれぞれ n/ℓ ($n = 0, \dots, N-1$)の不連続な値をとる。

(図4参照)

3.2 文字列ピッチ検出

(1) 2次元フーリエ変換結果

印刷文書画像(一般の英文雑誌)を

試料として、TVカメラを用いて計算機内に 128×128 の分解能で取り込み、閾値を設けて二値化し、その二値化情報に基づき、式(1)を用いて二次元フーリエ変換を行う。入力文書画像およびその二次元フーリエ変換強度を図4に示す。図4より、原画およびv軸上にピークが生じていることがわかる。これは、入力文書において、文字列の周期入に対応する空間周波数成分が顕著であることを示している。そこで、v軸上に着目し、 $|F(0, v)|$ とvの関係をもつ2つの例について図5に示す。図5より、一般に、文字列周期に対応したピークが生じていることがわかる。したがって、逆に、フーリエ変換後のピーク的位置より文字列のピッチを抽出することが出来る。

(2) 図形領域の影響

文書画像において、図形領域が二次元フーリエ変換のピーク値に与える影響について述べる。図形領域が増加していった場合のフーリエ変換のピーク値の変化を図6に示す。その結果、ピークの高さの値は文字列部分の面積に比例しており、ピーク的位置は、図形領域の増加に関係なく一定である。図6の結果より、図形領域の割合が5割以上を占めていても十分ピークが識別可能であり文字列ピッチを求めることが出来る。

(3) 文字列の周期の算出方法

二次元フーリエ変換面で、原画を除いて最も大きな値を持つ点(オリーブ点)が、変換面での隣接する点の値と比較して、差が小さくなる場合がある。これは文字列の周期数が整数でないことから生じる問題である。この場合の文字列の周期を求める方法を図7に示す。オリーブ点を ϕ_1 、オリーブ点の値を A_{ϕ_1} とし、 ϕ_1 に隣接する点のうち最大値を有する点を ϕ_2 、その点の値を A_{ϕ_2} とすると、真の行数 N および

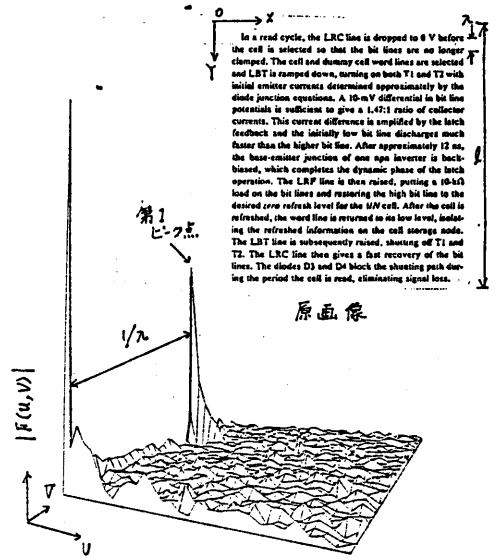


図4 入力文書画像と二次元フーリエ変換強度

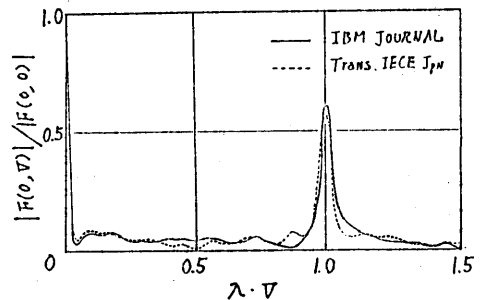


図5 $\lambda \cdot v$ と $|F(0, v)|$ との関係

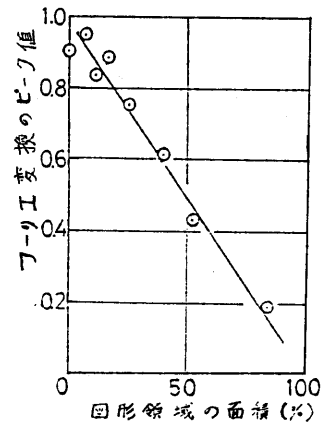


図6 図形領域に対するピーク値の変化

文字列のピッチ入は式(2)で求められる。

$$\frac{1}{\lambda} = x = i_1 + \frac{A_{i_2}}{A_{i_1} + A_{i_2}} \quad (2)$$

図4の文書画像を用いて、フーリエ変換する画像エリアを変化させていった場合の行数を、式(2)を用いて算出し、結果を図8に示す。画像エリアは120×120から128×128まで変化した。

図8より、行数は画素数の増加にしたがってリニアに変化しており、本方法が十分有効であることがわかる。

3.3 文字列の傾き検出

文字列が使っている場合の傾き θ と2次元フーリエ変換面との関係を図9に示す。図9に示すように、文字列の傾き θ は、変換面上の第1ピーク点の位置座標(u_1, v_1)を用いて式(3)で求められる。

$$\tan \theta = \frac{u_1}{v_1} \quad (3)$$

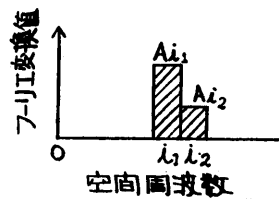
図4に示す原画像を、回転していった場合の実際の角度と、式(3)を用いて求めた角度との関係を図10に示す。

図10より、45°付近では少しばらつきがあるが、全体的によく一致しており、本方法の有効性が確認できた。

3.4 文字列の位置の検出

印刷文書中の文字列のピッチ、傾きは、2次元フーリエ変換面でのピーク点の位置から認識することが可能であるが、文字列の位置に付いては以下に述べる方法を用いて検出する。印刷文書中の文字列の傾きを補正した後の画像中の主な周期成分の濃度分布は次式であらわされる。

$$\begin{aligned} f(x) &= a \sin \omega x + b \cos \omega x \\ &= \sqrt{a^2 + b^2} \sin(\omega x + \theta) \end{aligned} \quad (4)$$



$$\frac{X - i_1}{i_2 - X} = \frac{A_{i_2}}{A_{i_1}} \quad X = \lambda_1 + \frac{A_{i_2}}{A_{i_1} + A_{i_2}}$$

図7 行数計算方式

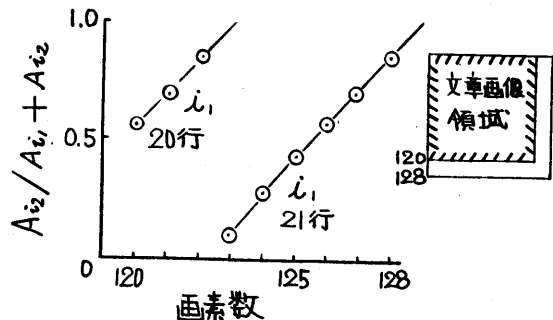


図8 行数の検出結果

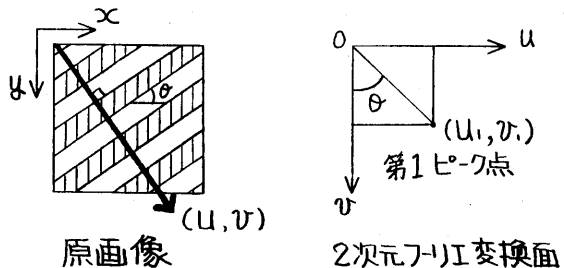


図9 文字列が傾いた場合の検出

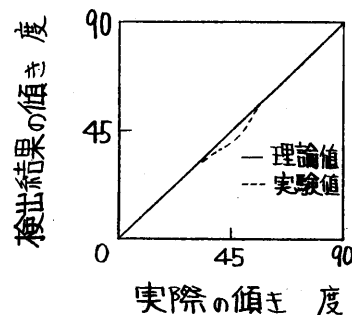


図10 文字列の傾き検出結果

$$\theta = \sin^{-1} \left(\frac{b}{\sqrt{a^2 + b^2}} \right)$$

a 、 b は一定値であり、 ω は角波数である。(図11①参照)

文字列のある位置を検出するために、式(4)が最大値 $\sqrt{a^2 + b^2}$ をとる部分、つまり、

$$\omega x + \theta = \frac{\pi}{2}, \frac{5\pi}{2}, \frac{9\pi}{2}, \dots; \frac{1}{2}(4n+1)\pi \quad (nは0以上の整数) \quad (5)$$

の x の値を求める必要がある。そこで式(4)の一次元のフーリエ変換処理は、

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) \exp(-j2\pi ux) \quad (6)$$

となり、式(4)を式(6)に代入すると、

$$F(u) = b - ja \quad (7)$$

となる。つまり、フーリエ変換処理の結果の実数部と虚数部の値より θ が求まり、最終的には、式(5)の x が求まる。

図11および表1に、文字列の位置を求める実験のためのデータとその検出結果を示す。表1の結果より、文字列の位置が精度よく検出できることが確認できた。

4. サブピクチャでの2次元フーリエ変換処理

4.1 最適マトリックスサイズの決定
3項で示したように、印刷文書全体の2次元フーリエ変換を行うことにより、文書画像中の文字列の行数、およびピッチが検出できる。このピッチを用いて最適なサブピクチャサイズを決定する。

文字列の周期を λ とすると(図12参照)、サブピクチャのマトリックスサイズは、 $\lambda \times n$ (n は0以上の整数)となる。 n を決定する要素としては次のことが考えられる。

- (1) 領域判別の分解能を考慮すると、 n は小さいほうがよい。

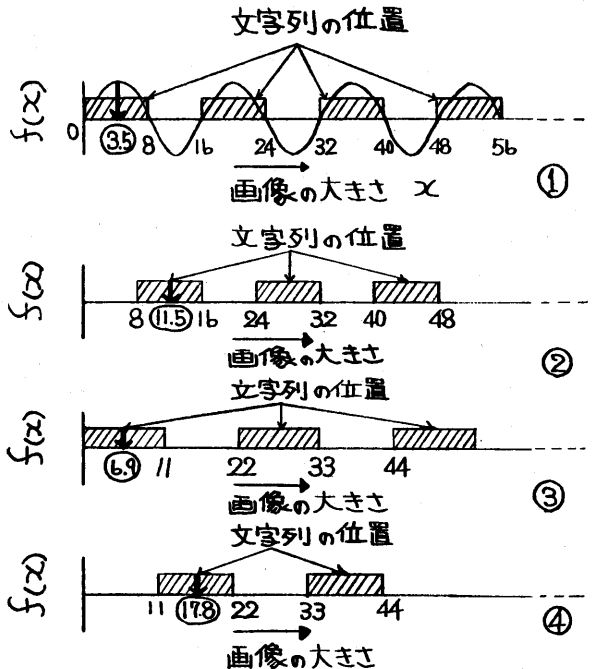


図11 文字列の実験データ

表1 文字列の位置の検出結果

	①	②	③	④
変換結果の実数部	80	-80	-149	149
虚数部	402	-402	352	-352
最初の文字列の位置	3.5	11.5	6.9	17.8

- (2) n が小さくなると(例えば $n=2$ の場合)他の雑音に弱くなり、識別の S/N 比が悪くなる。

S/N 比を考えれば、 $n=3$ 以上が適当であるが分解能を考慮して、本検討では $n=3$ とした。

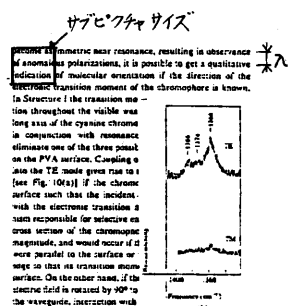


図12 サブピクチャサイズ

4.2 サブピクチャでのスキャンによる領域判別法

4.1項で決定されたサブピクチャのサイズで全体の画像をスキャンする際には、2次元フーリエ変換を行い、変換面でのピーク値が検出できるかどうかを判断する。

文書中に文字列が連続している場合には、フーリエ変換面でピーク値が検出できる。文書中の余白部分および図形・写真領域部分においては、行数に対応するピーク成分は弱い。一ライン方向に、サブピクチャでスキャンすることにより検出されるピークは、一般の文書画像を考慮した場合には、図13に示すような2つのパターンに分けられると考えられる。

図13(1)は、文字列領域が2つ以上に分かれている場合であり、図14(1)に示すように、中央に余白があり両側に文字列が存在している。

図13(2)は、図14(2)に示すように、文字列に続いて図形・写真領域がくる場合である。

図13に示すように、文字列領域を判定する分解能は、サブピクチャのサイズ以下にはならない。そこで、分解能を向上させる方法として、図15に示す方法が考えられる。

一ライン方向にサブピクチャでスキャンすることにより、図15の $g(x)$ が求まる。 $g(x)$ の周波数の α 、 β は、原画像上での文字列領域の開始点、終了点に対応している。そこで、 α 、 β 値を利用して、 $g(x)$ から $h(x)$ の周波数を導くことにより文字列の領域を判別することが可能である。

本方法を用いることにより、文字列領域が小さな余白を挟んで連続している場合にでも精度よく文字列の開始点と終了点を検出することが可能である。この例を図16に示す。

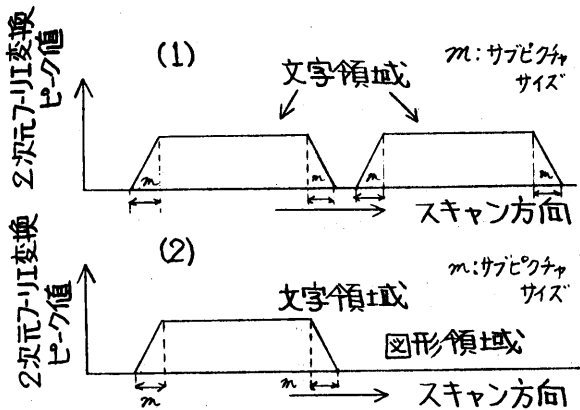
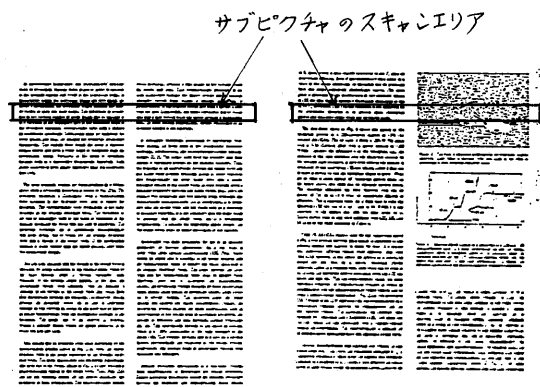


図13 ピーク値の2つのパターン



(1) 文字列のみの場合 (2) 図形領域を含む場合

図14 サブピクチャのスキャンエリア

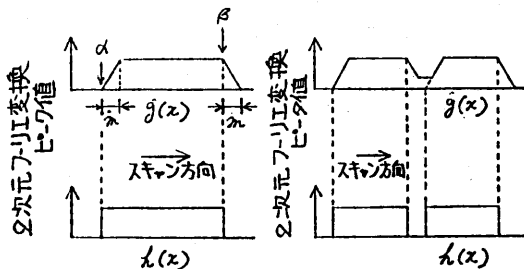


図15 分解能向上方式

図16 文字列領域判別結果

5.2 行、列ともに周期性のある文書画像

行方向と列方向に周期性のある文書の一例として、特許文書に対して領域判別法を適用した場合の、特許文書画像とその領域判別結果を図19および図20に示す。

また、文字列領域内でのサブピクチャの2次元フーリエ変換結果の一例を図21に示す。

図20に示すように、特許文書においても文字列領域と図形領域の切り分けが正確に行われていることがわかる。

図21より、特許文書は一文字、一文字離れており、かつ、縦、横の文字ピッチが同じであるために、2次元フーリエ変換した平面で、縦横の文字ピッチが明白にわかる。逆にいうと、3,4項の手法を行、列方向に行うことにより一文字の切り出しも容易にできる。

6. おわりに

印刷文書の領域判別に2次元フーリエ変換を適用し、次の結果を得た。

(1) 印刷文書の文字列領域には、そのフーリエ変換面に文字列のピッチに対応するピークが存在し、その位置・位相から文字列の周期、文字列の存在箇所が認識できる。

(2) サブピクチャをフーリエ変換し、フスキャンすることにより、文字列領域と非文字列領域が判別できる。

今後は、文字の切り出し、計算処理時間の短縮について検討を進める。

謝辞 日頃御指導頂き、日本電信電話公社種別電気通信研究所、高野陸男プリンタ研究室長ならびにプリンタ研究室の方々に感謝いたします。

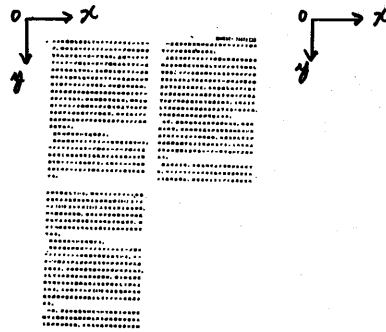


図20 特許文書の判別結果

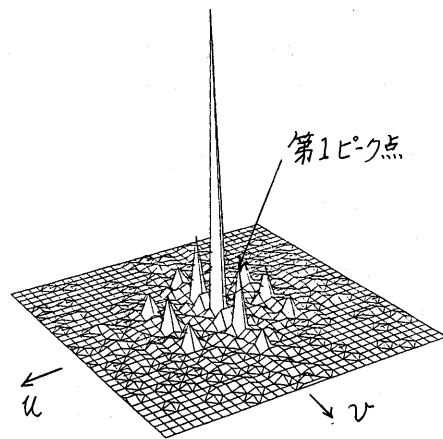


図21 サブピクチャ内の2次元フーリエ変換結果

参考文献

- (1) 秋山、増田：“印刷物の記事領域における文字の切り出し”，信学技報 PRL 80-70 (1981-01)
- (2) 坂井利文：“人工知能への期待”，情報処理 21, 12, PP1213~1212 (昭 55-12)
- (3) W. Scherl: "Automatic separation of text, graphic and picture segments in printed document", Pattern Recognition in Practice, P213~221, 1980
- (4) 長谷、星野：“印刷文書の周期的特徴”，信学論 (D), J65-D, 2, PP298~299 (昭 57-01)