

コンピュータビジョンを用いた書籍のデジタル化に関する検討

川原子 淳^{*1}, 山下 淳^{*2}, 金子 透^{*2}, 三浦 憲二郎^{*2}

^{*1} 静岡大学大学院理工学研究科 〒432-8561 静岡県浜松市城北 3-5-1

^{*2} 静岡大学工学部機械工学科 〒432-8561 静岡県浜松市城北 3-5-1

Tel/Fax: 053-478-1070

Email: kaneko@eng.shizuoka.ac.jp

文書情報のデータベース化を目的としたコンピュータ文書画像処理においては、既存の書籍のデジタル化も重要な課題である。そしてこの具体策としては、紙面をスキャナーで入力し、これを文字認識することによりテキストファイル化する方法が考えられる。しかし分厚い本を対象とする場合には、スキャナー入力画像中に影の部分が生じてしまい、認識処理に十分な画質が得られないという問題点がある。そこで本研究では画像入力方法として、高精細のデジタルカメラにより紙面をステレオ撮影する方法を採用することとし、画像処理により本の湾曲や傾きを補正した後に文字認識を行う方法について検討する。予備実験により良好な結果を得た。

キーワード 画像処理, 文書解析, 書籍のデジタル化, ステレオ視, 文字認識

A study on the digitization of books using computer vision

Atsushi Kawarago^{*1}, Atsushi Yamashita^{*2}, Toru Kaneko^{*2}, Kenjiro T. Miura^{*2}

^{*1}Graduate School of Science and Engineering, Shizuoka University

3-5-1 Johoku, Hamamatsu-shi, 432-8561 Japan

^{*2}Faculty of Engineering, Shizuoka University

3-5-1 Johoku, Hamamatsu-shi, 432-8561 Japan

Tel/Fax: 81-53-478-1070

Email: kaneko@eng.shizuoka.ac.jp

Abstract

In document processing by computer, aiming at the construction of document database, digitization of existent printed books is an important issue. One solution is to convert a scanner image of each page into a text file by character recognition. However, when applying to thick books, shading occurred in images makes it difficult to obtain enough image quality for recognition. In this paper, we propose a method where stereo camera images are processed to adjust the curvatures and inclinations of the sheet prior to character recognition. A preliminary experiment has shown its effectiveness.

key words Image processing, Document analysis, Book digitization, Stereo vision, Character recognition

1. 目的

コンピュータを用いた文書画像処理の目的の1つに、文書情報のデータベース化がある。既存の書籍をデジタルファイルに変換する方法としては、次の2つの方法がある。紙面をそのまま画像ファイルに変換する方法と、スキャナ等を用いて入力し、文字認識することによりテキストファイルに変換する方法である。ファイル容量の小ささや編集のしやすさ、出力の美しさから後者の方がデータベース化に適している。しかし、分厚い本を対象とする場合、スキャナ等を用いた入力では、文字認識が出来ないという問題点がある。分厚い本では綴り部分が湾曲してスキャナ面から離れてしまい、綴り部分に影が生じるために読み取ることが出来ないからである。

画像中の陰影情報から3次元形状の復元する問題は、Shape from Shading と呼ばれる。この解決法の1つ[1]を用いて、陰影情報より紙面形状を推定し補正した上で、文字認識することが考えられる。しかし、この手法では、多くの基礎データが必要となるほか、紙質などに大きく影響される。また、ページをめくる毎に、分厚い本を持ち上げてスキャナ面に置き直すといった手間が必要となる。

そこで、本研究では、本を上向きに置くことで、動作はページをめくるのみとし、本の種類や紙面の明るさによらずテキストファイル化を可能にすることを旨とする。具体的には、高精細のデジタルカメラにより紙面をステレオ撮影し、画像変換を用いて湾曲を補正した後に、文字認識を行う方法について検討する。

2. 処理手順

本方式は、Fig.1 に示すような4つのステップからなる手順で処理を行う。今回は、入力画像中の文字のみを処理対象とする。また、本はデジタルカメラに対して、平行に置くものとする。

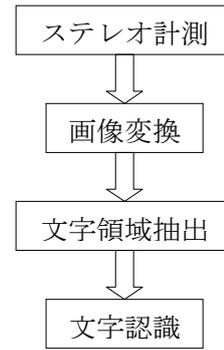


Fig.1 処理手順

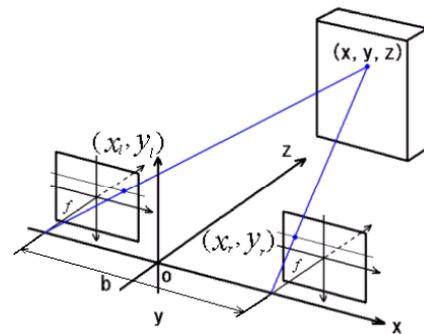


Fig.2 平行ステレオモデル

2.1 ステレオ計測

本研究では、平行ステレオにより文書画像を取得する。Fig.2 に平行ステレオモデルを示す。

3次元空間中の測定点 (x, y, z) が、左右の画像上で位置 (x_l, y_l) , (x_r, y_r) に投影されているとする。測定点の3次元座標値 (x, y, z) は、(1)式により算出することが出来る。ここで f は、レンズ中心から画像面までの距離であり、 b は基線長である。

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{b}{x_l - x_r} \begin{pmatrix} (x_l + x_r)/2 \\ y_l \\ f \end{pmatrix} \quad (1)$$

紙面形状は、ステレオ計測により紙面上の全ての点について、3次元座標値を算出することで復元することが出来る。そのためには、取得した左右の画像上で対応点を検出する必要がある。しかし、全ての点について正確に対応点を検出することは一般に困難である。

そこで、画像上の代表的な特徴点についてのみ対応点検出を行い、その点のステレオ計測を行う。特徴点としては、エッジ点などを用いる方法もあ

るが、文字列の中心線上の各点を特徴点として対応点検出を行う。以下に理由を示す。

- (1) 今回の対象は局所的に急激な変化がなく、一部の3次元座標値から紙面形状が予想出来る。
- (2) 各点の厳密な3次元座標値を必要としない。
- (3) 処理時間が短縮出来る。

始めに、取得画像から文字領域と図形領域に分割する。文字領域では文字認識を行うのに対し、図形領域では線画処理を行うためである。分割方法は、原画像を2値化・膨張処理により小領域に分割する。その後、小領域毎に輪郭線抽出を行い、領域の大きさ及び形状を計測して文字領域と図形領域に分割する(Fig.3)。文字列領域は横に連なるといった特徴から、分割することが可能である。



Fig.3 文字領域

次に、文字領域から文字列中心線を検出する。中心線検出は、文字領域輪郭線の垂直方向の中心点集合を対象に、文字列の中心線候補として最小二乗法を用いて3次曲線を求める。求めた曲線を文字列の中心線とする(Fig.4)。

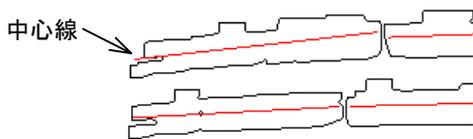


Fig.4 文字列の中心線

対応点検出は、文字列の中心線上の各点について、グレースケールの原画像での正規化相互相関(NCC)を用いたテンプレートマッチングにより求める。平行ステレオの場合は、左右画像上で同じ走査線上で対応点検出を行えば良い。左右画像上の対応点の座標値からステレオ計測を行い、文字列中心線の3次元座標値を求める。

以上の処理を各文字列について行うことで、紙面形状を復元することが可能である。

2.2 画像変換

文字列中心線の3次元座標値から紙面が平面となる画像に変換する。このとき、変換後の再配列処理には、逆変換に基づいた内挿法を用いて行う。

2.3 文字領域抽出

変換後の画像から文字幅を検出して、1文字毎の文字領域を抽出する。認識精度の向上、及び処理時間の短縮を図るために、フォントの文字幅により、Table 1のように文字を4種類に分類する。

Table 1 文字の分類

狭 ↑↓ 広	i j l t
	f r s
	a b c d e g h k n o p q u v x y z
	m w

文字幅の検出は、文字の最小幅から最大幅までを検索範囲として1文字毎に行う。まず検索範囲の最初の一列について調べる。1画素ごとの濃度値 g を累積し、文字領域の高さ v で割り、一列目の平均濃度 \tilde{g} を調べる。平均濃度が、文字以外の領域の平均濃度である背景濃度に近い場合、背景濃度との差の2乗和 W_i を算出する。これを検索範囲の最後まで順次算出する。算出後、検索範囲内で背景濃度との差が最小である列 W を文字の切れ目とし、文字幅として検出する((2)式) (Fig.6)。

$$W = \min \{W_i\} - W_0$$

$$W_i = \sum_{j=0}^v \{g(i, j) - \tilde{g}\}^2 \quad (2)$$

ここで、 $\tilde{g} = \sum_{j=0}^v g(i, j) / v$, v : 文字領域の高さ

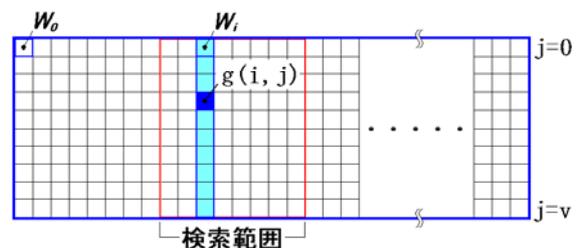


Fig.6 文字幅の検出

2.4 文字認識

一般的な文字認識では、2 値化された原画像の文字候補に対して、テンプレートマッチングや構造解析を行う。しかし、明るさの変化の影響がある場合には、正しく認識することが出来ない。

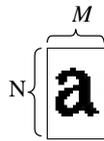
そこで、明るさの変化に対応するため、グレースケールの原画像を対象に、正規化相互相関(NCC)を用いたテンプレートマッチングにより文字認識を行う。

相互相関係数 C は、次式で与えられる。

$$C = \frac{\sum_{j=1}^M \sum_{i=1}^N \{I(i, j) - \mu_I\} \{T_c(i, j) - \mu_T\}}{\sqrt{\sum_{j=1}^M \sum_{i=1}^N \{I(i, j) - \mu_I\}^2 \sum_{j=1}^M \sum_{i=1}^N \{T_c(i, j) - \mu_T\}^2}} \quad (3)$$

ただし、 $\mu_I = \frac{1}{NM} \sum_{j=1}^M \sum_{i=1}^N I(i, j)$,

$$\mu_T = \frac{1}{NM} \sum_{j=1}^M \sum_{i=1}^N T_c(i, j)$$



ここで、 I は対象画像であり、 T_c はテンプレート画像である。

文字認識は、マッチング領域内の各位置で、各テンプレートについてマッチングを行う。マッチング領域は、文字幅の検出により算出した1文字毎の領域に、文字領域の抽出誤差を加えた領域である。このとき、最も相関が大きい文字を認識結果として、テキストファイルに出力する。

3. 実験結果

実験では、カメラの個体差による影響を避けるため、1台の高精細デジタルカメラを使用した。カメラは、本に対して平行に移動させて撮影した。

3.1 文字認識方法の比較

本方式(NCC)の有効性を確認するため、2 値化された原画像に対して、残差逐次検定法(SSDA)を用いたテンプレートマッチングと比較した。

文字認識は、小文字「a」～「z」を対象とした。

文字認識方法の比較結果を Table 2 に示す。

Table 2 より、SSDA 法よりも本方式(NCC)の方が

有効であることが確認された。これは、本方式では、2 値画像のような文字輪郭部の崩れがなく、また明るさの変化の影響を受けないためである。

Table 2 文字認識方法の比較結果

	総数	認識	認識率
NCC	123	93	76%
SSDA	123	59	48%

3.2 湾曲の大きい画像

Fig.8 に示すような湾曲の画像について実験を行った。この程度の湾曲した紙面をスキャナ等で読み込むと、中央部分が黒くなり文字認識は困難である(Fig.9)。実験では、撮影した画像(2400×1800 画素)の部分画像(580×380 画素)を使用した。

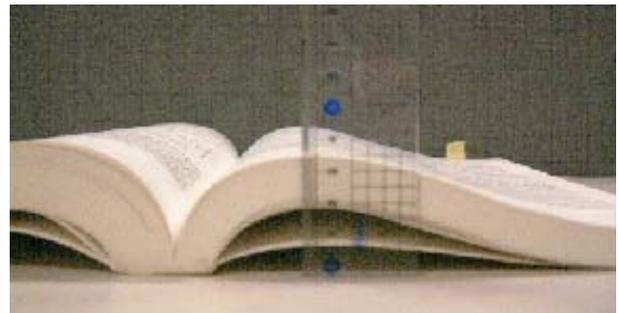
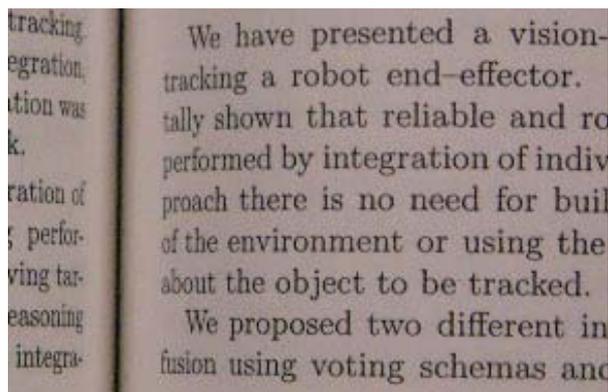


Fig.8 撮影対象

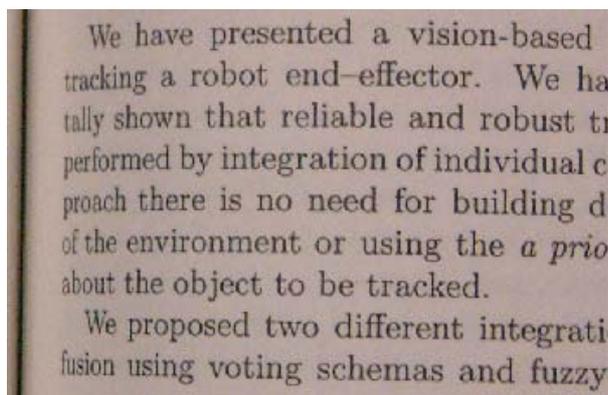


Fig.9 スキャナ入力画像例

カメラによる取得画像を Fig.10 に示す.



(a) 左画像



(b) 右画像

Fig.10 取得画像

ステレオ計測による文字列中心線の3次元座標値の算出結果を Fig.11 に示す. 紙面形状の実測を行っていないため比較はできないが, おおよその形状は, 復元出来ているものと思われる.

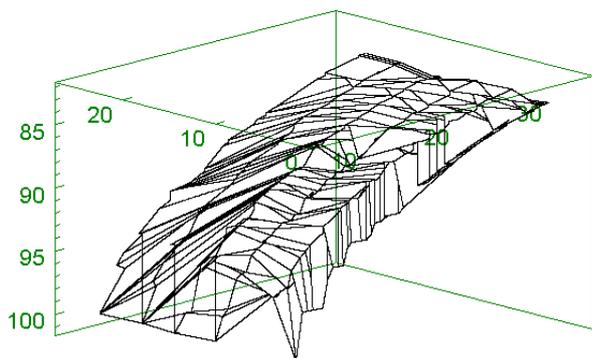


Fig.11 算出結果

画像変換の結果を Fig.12 に示す. 画像の再配列処理には, 最近隣内挿法を用いた.

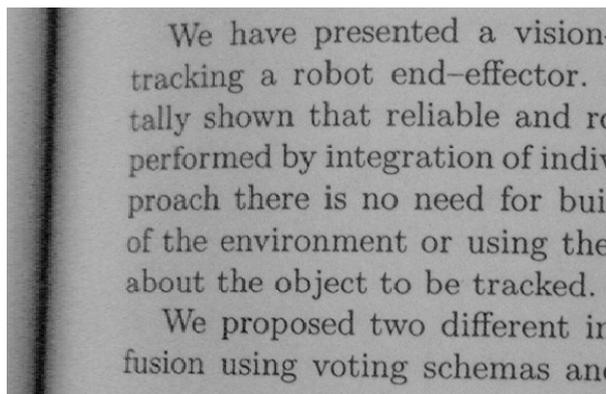


Fig.12 変換画像

画像変換結果の一部を Fig.13 に示す.

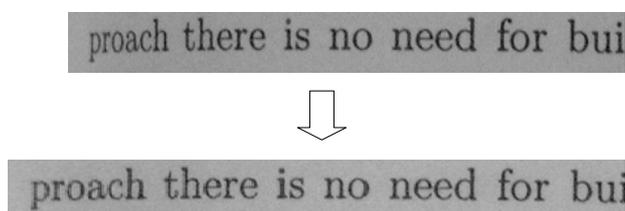


Fig.13 画像変換結果 (一部)

画像変換の結果は, Fig.12, Fig.13 より目視においては良好である.

この実験では, 3.1 の実験において誤認識が多かった「c」, 「e」と, 湾曲部に多く含まれていた「a」についてのみ文字認識を行った. Table 3 は, 画像変換前後の「a」, 「c」, 「e」の文字認識結果である. 画像変換により文字認識率が大きく向上した. これは, 湾曲の大きい部分での認識率が著しく向上したためである.

Table 3 文字認識結果

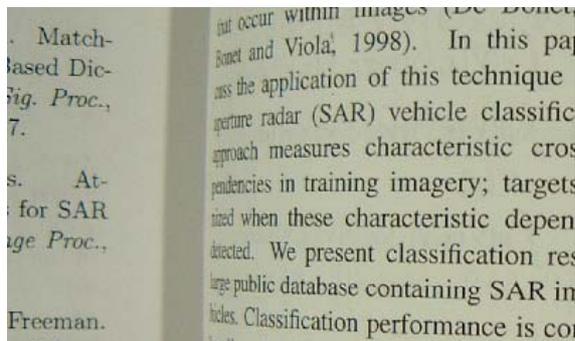
	総数	認識	認識率
変換前	a: 14	a: 10	64%
	c: 6	c: 4	
	e: 29	e: 24	
変換後	a: 14	a: 14	100%
	c: 6	c: 6	
	e: 29	e: 29	

3.3 湾曲のさらに大きい画像

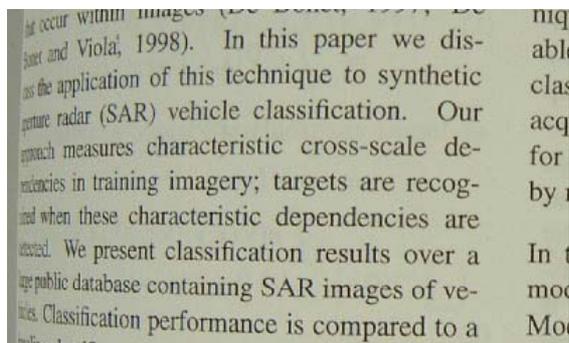
次に, 湾曲のさらに大きい画像について実験を行った. この画像は, 目視においても文字認識は

困難である。実験では、撮影した画像(2400×1800画素)の部分画像(640×430画素)を使用した。

カメラによる取得画像を Fig.14 に示す。



(a) 左画像



(b) 右画像

Fig.14 取得画像

画像変換の結果を Fig.15 に示す。

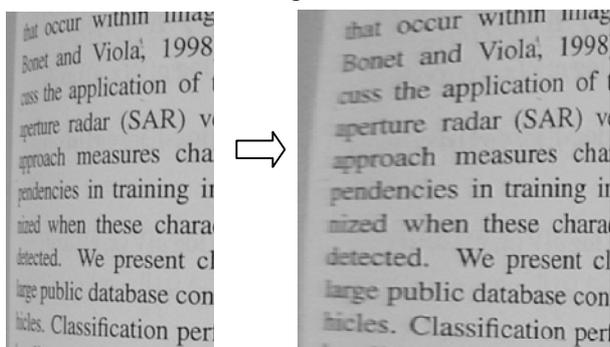


Fig.15 画像変換結果

画像変換結果の一部を Fig.16 に示す。

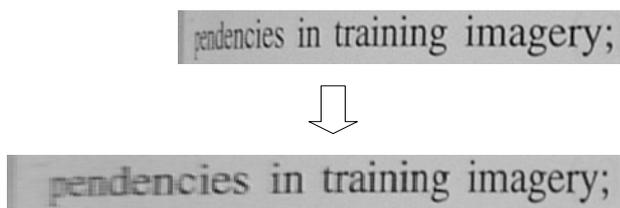


Fig.16 画像変換結果 (一部)

Fig.15, Fig.16 に示すように、湾曲の非常に大きい部分では、変換後の文字がかなりぼやけている。このようにぼけたのは、画像の再配列処理にアルゴリズムが単純な最近隣内挿法を用いたためであると思われる。3次補間法などの精度の高い補間を行うことで、改善することが可能である。

Table 4 は、画像変換前後の「c」、「e」の文字認識結果である。画像変換により認識率が向上した。しかし、湾曲のもっとも大きい部分の文字については、認識することは出来なかった。

Table 4 文字認識結果

	総数	認識	認識率
変換前	c: 6 e: 14	c: 1 e: 1	10%
変換後	c: 6 e: 14	c: 2 e: 12	70%

4. 結論

本論文では、高精細のデジタルカメラにより撮影した紙面の画像から、画像処理により本の湾曲を修正した後、文字認識を行う方法について検討した。具体的には、ステレオ撮影により取得した画像からの対応点検出、3次元座標値からの画像変換、正規化相互相関による文字認識である。そして、実際の書籍による実験を行い、本方式によって湾曲の修正が出来ること、及び文字認識が可能であることを示した。

今後の課題としては、現在手動で作成している文字テンプレートの自動作成方法の検討、及び湾曲の非常に大きい画像における文字認識精度の向上を図ることである。また、画像中に存在する図表、写真などの認識も検討していく予定である。

参考文献

- [1] 浮田浩行, 小西克信, 和田俊和, 松山隆司: “固有空間法を用いた陰影情報からの書籍表面の3次元形状復元”, 電子情報通信学会論文誌, D-II, Vol.J83-D-II, Vol.12, pp.2610-2621, 2000.