

デジタルカメラで撮影した展示パネル画像の認識

Poster Contents Recognition in Scene Image

岩田 達明† 石谷 康人† 亀山 研一†
Tatsuaki Iwata Yasuto Ishitani Kenichi Kameyama

1 はじめに

デジタルカメラの普及に伴い、看板や自動車ナンバープレートなど紙媒体以外の物体の上に存在する文字情報を撮影・デジタル化してコンピュータに取り込むことが可能となった[1-3]。その発展的な例として、展示博覧会などでよく見られる説明パネル（以後、展示パネルとする）をデジタルカメラで撮影し、パネル中の文章をOCRでコード化することが考えられる。展示パネルには展示物の名称や特徴など重要な情報が記載されているため、その電子化と再利用に対するニーズが非常に高いことが予想される。

実世界における展示パネルは、不均一な光源、スポットライト、他の物体による影などの影響を受けていることが多く、デジタルカメラで撮影した画像には多様な輝度ムラが生じる。また、展示パネルを常に正面から撮影できるとは限らないため、撮影画像に図1のような透視的な歪みが生じることがある。従来のOCRでは、光源が一様であることと、文書が撮像デバイスと正対していることを仮定しているため、様々な条件の下で撮影された展示パネル画像を対象とした場合には文字認識精度が著しく低下する。

そこで本論文では、輝度ムラと透視歪みを伴う展示パネル画像を従来のOCRを用いて高精度に認識することを目的として、Super-sampling法による低解像度画像を用いた階層的2値化方式と、展示パネル中の文章領域を対象とした透視歪み補正方式を提案する。また本論文では、様々な展示パネルに対して提案方式を適用して得た画像を用いて文字認識実験を行うことにより提案方式の有効性を評価する。

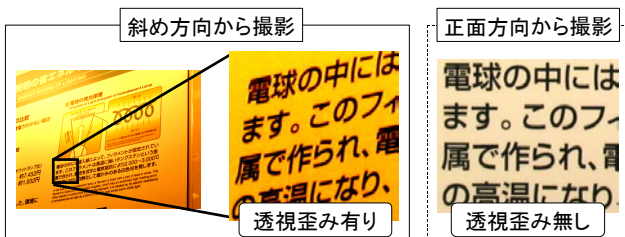


図1 透視歪みを伴う展示パネル画像の例

2 Super-sampling 法による低解像度画像を用いた階層的2値化方式

画像の輝度ムラに対処して、撮影画像から高品位な2値画像を生成するためには、局所領域ごとあるいは画素ごとに2値化のためのしきい値を決定する必要がある[1][4]。しかしこの場合、処理手順が複雑になるため処理コストが大幅に増大し、文字ストロークに劣化や欠損が生じて文字認識精度が低下することがある。本論文では、このような問題点を解決するために、Super-sampling法を用いて高品位な低解像度画像を生成し、低解像度画像の画素ごとにしき

い値を決定した後、そのしきい値を元の画像に適用する階層的な2値化方式を提案する。

(1) Super-sampling 法による低解像度画像の生成

$W \times H$ の入力濃淡画像を、 $M \times N$ へと低解像度化する場合、一旦 $\sigma \times \rho$ の画像に圧縮したあと、そこから $M \times N$ の低解像度画像を生成する。このとき、 $W > \sigma > M$, $H > \rho > N$ とする。コンピュータグラフィックスの分野では、このように段階的に低解像度化を行うことで、画像の劣化が軽減することが知られている。

(2) 低解像度画像を用いた適応的しきい値決定

$M \times N$ の低解像度画像において、Niblack[4]の式： $T(x, y) = m(x, y) + k \cdot s(x, y)$ を用いて画素単位にしきい値を決定する。ここで、 $T(x, y)$ を画素 (x, y) におけるしきい値、 $m(x, y)$ と $s(x, y)$ を画素 (x, y) を中心とする $P \times P$ の部分領域内の平均と標準偏差、 k を重み係数とする。

(3) 階層的2値化処理

低解像度画像の各画素で決定したしきい値を用いて、それに対応する入力画像の部分領域(横 W/M 画素×縦 H/N 画素)を順次2値化することにより、入力画像から2値画像Iを生成する。

3 展示パネルの文章領域を対象とした透視歪み補正方式

透視歪みを伴う画像Iは、実際の展示パネルOを透視変換行列Mを用いてデジタルカメラの撮像面に透視投影したものと見なすことができる。したがって、透視変換行列の逆行列 M^{-1} を画像Iに適用することにより実際の展示パネルと等価な透視歪みのない画像I'を生成することができる。このとき、画像Iにおける複数の点の集合P(図2(a))と、画像I'における複数の点の集合P'(図2(b))の間で1対1写像が成立すると仮定できれば、それらの点の位置座標を透視変換の関係式[5]に代入することにより、透視変換行列Mを推定することができる。

しかし、本方式では画像I'が未知であるためP'が一意に定まらない。そこで本方式では、展示パネルOにおいて複数の直線状の文字行が平行に分布しており、それらが整然と配置されている(文字行が長方形のテキストブロックに内接する)と仮定する。このとき、画像I上で図2(a)のような4つの点P1~P4を持つ四辺形を推定することができ、その四辺形が画像I'上では図2(b)のような4つの点P1'~P4'を持つ長方形に相当すると見なすことができれば、透視変換行列Mの推定が可能となる。

提案方式では、まず2値化処理で得られた画像Iにおいて文章領域を抽出し、文章領域を構成する文字行の分布を分析して点P1~P4を推定する。そして、P1~P4を用いて展示パネルOにおける点P1'~P4'を推定した後、P1~P4とP1'~P4'を用いて透視変換行列Mを求める。最後に、 M^{-1}

† (株) 東芝研究開発センター, Corporate R&D Center, TOSHIBA Corporation

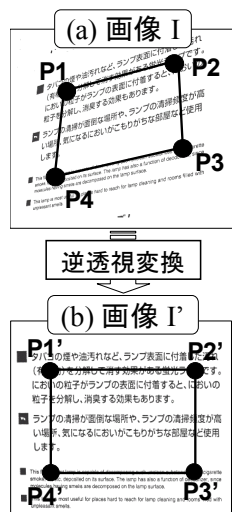


図2 画像の透視歪み補正の例

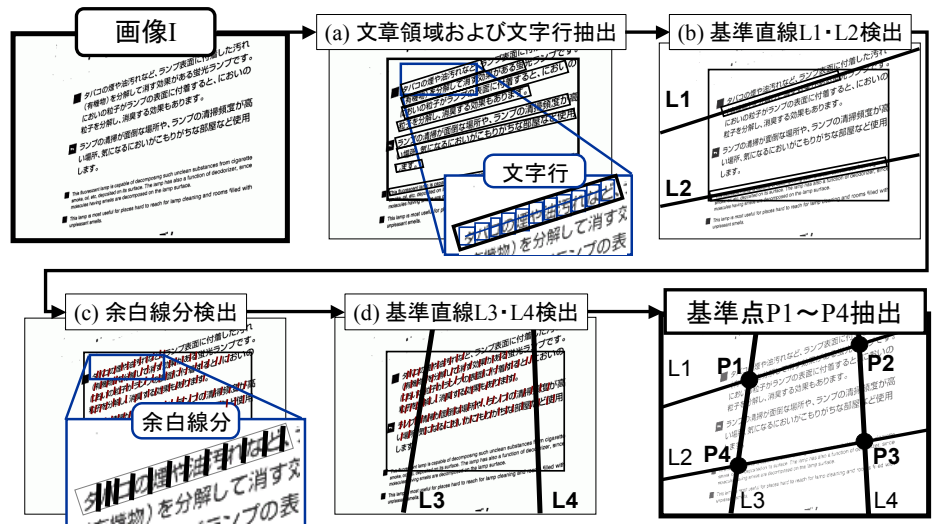


図3 透視変換行列推定のための基準点 P1~P4 の抽出

を画像 I に適用することで透視歪みの補正を行う。以下では、画像 I における文章領域の抽出と文章領域での P1~P4 の抽出について説明する。

(1) 文章領域および文字行の抽出

画像 I に対してラベリング処理を施して黒画素の連結成分を抽出した後、矩形表現された連結成分の縦幅値と横幅値に対して統計処理を施すことにより文字成分を抽出する。そして文字成分が最も密集する $\delta \times \phi$ の部分領域を抽出して、それを文章領域と見なす。また、文章領域において隣接する文字成分をリンクすることで図 3(a) のような文字行を抽出する。このとき、それぞれの文字行において、文字成分を内接して面積が最小となる外接矩形を抽出する。

(2) 画像 I 上における P1~P4 の抽出

文字領域中の文字列において、行間距離が ϕ 以上で、文字成分数の合計が最大となる 2 つの文字行の組を選択する。これら 2 つの文字行に対して直線近似を行うことで 2 つの基準直線 L1, L2 を求める(図 3(b))。このあと、文章領域内のすべての文字行において、外接矩形内を走査して文字成分の黒画素と交差しない図 3(c) のような余白線分を検出する。そして、文章領域内の余白線分の集合を対象として統計処理を行うことで基準直線 L3, L4 を検出する(図 3(d))。このようにして得られた基準直線 L1~L4 の交点を画像 I における基準点 P1~P4 と見なす。

4 実験

デジタルカメラで撮影した東芝科学館の展示パネル画像(解像度 1600×1200)を用いて、本論文で提案した 2 値化方式と透視歪み補正方式を評価した。

4.1 2 値化方式の評価

実験では、透視歪みの無い展示パネル画像 9 枚(総文字数 1499 文字)を用いて提案方式と Niblack の方式[4]を比較評価した。この際、それぞれの方式の処理時間と、それぞれの方式を用いて生成した 2 値画像に対して OCR を適用した際の文字認識率を計測した(表 1)。その結果、本方式において認識率が 1.33% 向上しており、また処理時間が 22.5 倍高速化していることが分かった。

表 1 2 値化方式の実験結果

	従来方式	提案方式
認識率(%)	97.67	99.00
処理時間(ms)	74323	3309

4.2 透視歪み補正方式の評価

それぞれ異なる透視歪みを伴った展示パネル画像 6 枚(総文字数 666 文字)を用いて、提案方式により透視歪みを補正した場合と手動で透視歪みを補正した場合の文字認識精度を比較した。提案方式を実施した場合には、歪み補正処理時間 12.4 秒、文字認識率 97.6% という性能が得られた。一方、手動で透視歪みを補正した場合の文字認識率は 98.3% となった。なお、透視歪みを補正しない場合は、全ての画像でレイアウト解析が失敗したため、認識結果を得るに至らなかった。以上の結果から、本方式により透視歪みが自動的に補正され、透視歪みによる文字認識精度の劣化が十分に抑えられていることが実証された。

5 まとめ

輝度ムラや透視歪みを伴う様々な展示パネル画像に対して、本論文で提案した 2 値化方式と透視歪み補正方式を段階的に適用したところ、97.6% の文字認識率を得ることができた。この結果、従来の OCR の処理対象に関する制約を取り払うことができ、応用範囲を広げることができた。

参考文献

- [1] 塩昭夫, “情景中文字の検出のための動的 2 値化処理法”, 信学論 D, Vol.71-D, No.5, pp.863-873, May, 1988.
- [2] 内藤貴志他, “撮像位置にロバストなナンバープレート認識方法”, 信学論 A, Vol.J81-A, No.4, pp.536-545, Apr. 1998.
- [3] 萩原敏浩他, “歪みを有する活字文字列における形状補正手法”, 信学技法, PRMU2001-190, pp.101-108, Dec. 2001.
- [4] W.Niblack, “An Introduction to Digital Image Processing,” pp.115-116, Englewood Cliffs, N.J.: Prentice Hall, 1986.
- [5] 佐藤淳, “コンピュータビジョン—視覚の幾何学—”, コロナ社.