構成要素のレイアウト情報を用いたカラー文書画像の方位検出

非会員 泉 井 宏 之 (関西大)

正 員 伊 藤 秀 隆 (関西大)

正 員 隈 元 昭 (関西大)

Detection of the Color Document's Direction
Using Layout Information of Document Components

Izui Hiroyuki, Non-member, Ito Hidetaka, Member, Kumamoto Akira, Member (Kansai University)

We present a method of detecting the color document's horizontal(or vertical) direction using the local layout information of document's components. Different from conventional string or string-line detections which have been intensively studied in an image-based document processing, the presented method deals with a color document with or without strings included as document's components. The given color document is analyzed following three steps of (a)color subtraction, (b)detection of two line segments which wrap appropriate two components in the form of a single convex hull and (c)generating the histogram based on the frequency of above line segments. The overall direction for a given document is basically computed using this histogram. The experiment is carried out for 92 color documents including those with or without strings. The result shows that a satisfactory direction is obtained for 90 documents tested.

キーワード：カラー文書，レイアウト，方位検出，要素間辺

1. はじめに

近年，与えられた文書画像の中から，文字列や文字列候補，あるいは，文字自体を抽出する研究をはじめとして，画像処理の手法に基づく文書処理の研究が盛んに行われている^{(1)~(8)}．しかし，文字や文字列が文書中で傾いている場合や，入力された文書画像自体が最初から傾いている場合などでは，文字あるいは文字列領域の抽出は一般に困難である．そこで，例えば文字の傾き検出では，投影法・ランレンクス法・極地点対応法などによって，文字位置の推定や方位の補正，あるいは，文字の認識などを実現する研究が行われている⁽⁹⁾．

ところが従来の研究では，対象とする文書画像が，予め2値画像に限定されていたり，画像中に含まれる要素が文字のみに限定されているなどの制限があるものが多い．また，文字のみからなる文書であっても，多くのフォントを含む文書や飾り文字などが存在するような多様な文書での研究事例は少ない．電子化された文書には，単なる文字のみを含む文書はもとより，ビジネスフォーム，あるいは，ちらしや広告などのように，デザイン・図・文字などを多様を含むものなど，多くのバリエーションがある．また，情報技術の進展に伴って，カラー文書の使用頻度が格段に増大

してきている．これらのカラー文書に対して，文書本来の方位を検出することは，例えば文字列検出に対する前処理としての役割も含めて，大きな意味をもつと考える．

一方，文書自体の方位を補正する試みとして，ペン型スキャナでの入力処理において，文書方位を自動的に補正する研究⁽⁷⁾も行われているが，これは，鏡像反転や90度回転された文字に対しては補正を行うものの，それ以外の任意の角度に傾いた文字に対しては対処できない．また，文字のみからなる2値画像を対象としているので，文字や図などを含む一般的なカラー文書へは直接には適用できない．

文書の方位を検出する代表的な手法である射影分布を用いる手法⁽⁸⁾は，通常 $\pm 15^\circ$ 程度の傾きを前提にしており⁽⁹⁾，極端に大きな傾きをもっている文書での適応例は少ない⁽¹⁰⁾．また，傾きの大きな文書を扱っている場合⁽⁹⁾，⁽¹¹⁾であっても，白黒2値画像の文書が処理対象になっていたり，文字領域と図や写真などの領域がはっきりと分かれている文書を扱っていたりしており，本論文が対象とするような背景カラー画像中に文字列やデザインなどが埋め込まれている場合はあまり考慮されていないようである．

そこで本論文では，カラー文書画像を対象として，文書中の要素（文字，絵，図などを一般に含む）の局所的な情報に基づいて，入力画像の全体的な方位を検出する手法を提

案し、実験によってその効果を検証する。本手法は、基本的には、与えられた文書中から、類似した色の連結閉領域を構成要素として多数検出し、構成要素相互の局所的なレイアウト情報を利用するボトムアップ手法である。したがって本論文の手法が適用可能な文書は、比較的類似した色で、大きさや距離の近い構成要素が整列しているようなものである必要がある。ここで、本手法は文字列の抽出を特に意識しているわけではないので、構成要素は文字であっても、図や写真中の要素であっても構わない。また文字同志が接触していても、接触した文字列と同じような色で大きさや距離の近い他の要素が整列しておれば、基本的には本手法が適用可能である。また、縦書き文書の場合には、縦方向に整列した文字列の方向性は検出できるが、文字認識をしないので縦書きである事実を検出するものではない。文字や絵などが混在している場合には、一般に文字列の方向性から文書の方位が検出される場合が多いが、文字列が少ない場合であっても絵の中に方位に関する情報（例えば建物中の整列した多数の窓など）が存在しておれば、ある程度方位を検出できる。

一般に、文書の大域的な情報を利用する手法に比べて局所的な情報を用いる場合には、文書構造に関する事前の知識が不要となり、多様な文書を扱えるといった利点がある⁽⁸⁾。本論文で提案する手法は、この文書の多様性に対処する目的で、ボトムアップ的な方式を採用している。局所的な情報の中には不要な情報も幾分かは必然的に含まれるものの、平均的に妥当な結論が最終的には得られることを期待するものである。

以下、2.で本提案の手法の概要を述べ、3.で詳細な手順を説明する。4.で代表的な実験結果を考察と共に示し、5.で結論と今後の課題を示す。

2. 方位検出の概要

提案手法では、文字・図・写真などを含むカラー画像に対して、画像内の構成要素の色・大きさ・レイアウト情報を利用することによって、画像の方位を検出する。

以下で用いる用語の定義は、つぎのとおりである。すなわち、「同じ色の画素からなる閉領域を構成要素」、「2個の構成要素を全体として1つの凸包へ併合する際に生じる、要素間をつなぐ2個の線分を要素間辺」、「ある代表色の画素値をもつ画素を代表色画素」、「代表色画素以外を透過画素」、「代表色画素と透過画素に分類した2値画像を代表色画像」、「構成要素のうち、4近傍のいずれかにおいて1つ以上の透過画素と隣接する代表色画素を境界線画素」と定義する。

入力画像の方位は次の3つのプロセスによって検出する。

- (1) 入力画像に対して減色処理を行う。
- (2) 構成要素を抽出する。
- (3) 要素間辺を検出し、方位を検出する。

減色処理によって、入力画像は複数枚の代表色画像(2値画像)へ分解される。減色処理の目的は、次の2点である。

1) 1色に見える領域でもわずかな色の違いがある場合、構成要素は複数個として抽出されるが、それらを1個として抽出する。2) 画像の方位を検出する上で、ノイズと思われる小さい構成要素を除去する。

構成要素の抽出において、1) 構成要素の輪郭を検出し、2) 輪郭の大きく曲がる点(以下、頂点候補という)を求め、3) それらの頂点候補から1個の凸包を作成する、という手順を経て、構成要素の頂点と色を抽出する。

要素間辺は、1) 色が同じ、2) 大きさが同程度、3) 距離が近い、の3条件を満たすすべての構成要素の組合せに対して検出する。また、一部の要素間辺をノイズとして除去する。次いで、要素間辺の傾きのヒストグラムから、減色された画像(減色画像)の方位を検出し、その方位を入力画像の方位とする。

3. 処理方法

〈3・1〉 減色処理の方法 減色処理は次の2つのプロセスから成り立つ。

- (1) 減色後の色表であるコードブックを作成する。
- (2) コードブックから入力画像の全画素を最良近似候補へ近似する。

減色後の代表色を決定するコードブックの作成には、ベクトル量子化を用いた。

ベクトル量子化では、画素を、RGBを軸とする3次元ベクトルとして捉え、RGBそれぞれが同程度の値である画素を、すべて同一色に近似することによって減色する。量子化の方法として、再帰分割・階層型クラスタリング・LBGアルゴリズムの3種類が知られている⁽¹²⁾。

本研究では生成速度が優れている再帰分割を用いた。以下にその手順を示す。

- (1) 全ての色を包含する最小の立方体(領域)を作成し、領域数を1とする。
- (2) 領域内の色分布の重心を求める。
- (3) 各領域について、RGB各色成分ごとに色の分散幅を求め、最大の分散幅をもつ色成分軸上の重心でその色の分散幅を2分割する(領域数が2倍になる)。
- (4) 目的の領域数(色数)に達するまで(2)に戻って繰り返す。

このようにして、最終的に得られる領域の重心を代表色とすることによって、コードブックが作成される。

各画素をコードブックの中の最良近似候補に量子化するには、コードブックの全ての色と画素とを比較し、色空間(RGBの3次元)内のユークリッド距離がもっとも近くなる候補を選択する。

この一連の処理によって入力画像は減色処理される。減色画像は、減色処理の際に作成したコードブックによって、複数枚の代表色画像(2値画像)に分解することができる。

〈3・2〉 構成要素の抽出方法 構成要素の抽出処理は次の3つのプロセスから成り立つ。

- (1) 構成要素の輪郭を検出する。

図1 構成要素の頂点検出

Fig. 1. Extraction of component's vertexes.

- (2) 構成要素の頂点候補を検出する。
- (3) 構成要素の頂点 (構成要素の凸包の頂点) を検出する。

(1)の構成要素の輪郭検出では、すべての代表色画像に対してラスタ走査を行い、境界線追跡処理によって、境界線画素を検出する。

(2)の頂点候補の検出では、境界線が大きく曲がる境界線画素を検出し、これらを構成要素の頂点候補とする。

(3)の構成要素の頂点検出では、頂点候補から構成要素の凸包を作成し、頂点候補のうち、凸包の内部にあるものを除外して、残りを構成要素の頂点とする。構成要素の頂点を検出することによって、各構成要素の輪郭情報をできるだけ保持したまま、情報量を絞り込み、後の計算コストを軽減することができる。

以上の処理によって、減色処理された画像から、構成要素の頂点と色の、2種類の情報が抽出される(図1)。

〈3・3〉入力画像の方位の検出方法 方位検出処理は、次の2つのプロセスから成り立つ。

- (1) 要素間辺(図2)を検出する。
- (2) 要素間辺を用いたヒストグラムを作成する。

ここで、構成要素の大きさを、水平垂直方向の外接長方形上に存在する構成要素の頂点間の、距離の最大値とする。構成要素の中心は外接長方形の中心とし、中心間の距離を構成要素間の距離とする。

要素間辺は、

- 減色後のRGBすべての値が等しい
- 大きい要素の大きさが小さい要素の大きさの1.3倍以内
- 要素間距離が小さい要素の大きさの2倍以内

の3条件(図3)を満たすすべての構成要素の組み合わせに対して、次の手順で検出する。

- (1) 2個の構成要素(の凸包)から、それらを併合する凸包を作成する。
- (2) (1)で得られた新しい凸包を構成する線分のうち、もとの2個の構成要素間をつなぐ2本の線分を、要素間辺として検出する。

次に要素間辺を用いた方位検出の方法について説明する。

- (1) 検出した要素間辺のうち、長いほうから総数の1割をノイズとして除去する。
- (2) 要素間辺の傾きの角度を横軸、要素間辺の本数を度数とするヒストグラムを作成する。
- (3) ヒストグラムを平滑化処理する。
- (4) ヒストグラムのピークに対応する要素間辺の傾き

図2 要素間辺

Fig. 2. Two line segments wrapping a pair of components.

図3 要素間辺の検出条件
(×印は条件を満たしていないことを示す)

Fig. 3. Three conditions of detection.

を画像の方位として検出する。

長い要素間辺は、画像の背景のような比較的大きい構成要素から検出されるため、画像本来の方位を示さない場合が多い。そこで、特に長い要素間辺をノイズとして除去する。また、要素間辺の傾きは、入力画像の上方向を0度とし、時計回りに180度までの値をもつものと定める。

4. 実験および考察

カラー文書画像を対象とした方位検出実験を行い、本手法の有効性を考察する。ここでは、代表的な実験結果として、

- 文字のみの画像(A)
- 文字・絵の混在した画像(B)
- 文字を含まない画像(C)

の3つの画像に対する結果を示す。表1に示すデータのうち処理時間は、5回測定したうちの最短のものである。実行環境は、CPUがPentium3(750MHz)、メモリが256MBであり、使用言語はJAVA、OSはKondara2000を用いた。

〈4・1〉文字のみの画像 原画像(A)(図4, 1138×800 pixel)は、フォントがすべて異なる12の文字列からなる画像であり、一部の文字列にはドロップシャドウの飾りを施してある。背景の縦縞に着目すれば、原画像の文書の方位は0度または90度とも考えることができるが、ここでは、構成要素の辺の情報は利用していないので、原画像作成時に設定した文字列の傾き70度を正解とみなす。

32色に減色した場合の処理結果を、表1(A)と図5〜図8に示す。方位の検出結果は、表1に示すように70度であった。図5には、一部に背景と同一色となった文字列があるが、方位検出には影響していない。減色処理後の色数を少なくしすぎると、このように、背景と文字が同一色になっ

表 1 計算時間と検出結果
(a:減色処理後の色数, b:減色処理時間 (秒), c:構成要素の抽出処理時間 (秒), d:要素間辺の検出処理時間 (秒), e:処理合計時間, f:抽出した構成要素数 (個), g:要素間辺数 (本), h:検出方位 (度))
Table 1. Computing time and detected results.

(A)							
a	b	c	d	e	f	g	h
32	76.6	18.1	0.2	94.9	409	372	70
128	3.0×10^2	155.6	0.4	3.2×10^2	570	458	69

(B)							
a	b	c	d	e	f	g	h
16	9.1	1.7	0.9	11.7	947	872	19
128	52.8	38.2	0.9	91.9	1268	638	20

(C)							
a	b	c	d	e	f	g	h
16	15.5	5.3	1.6	22.4	1543	632	90
128	1.6×10^2	93.3	11.0	2.6×10^2	4233	1772	90

図 4 原画像 (A)
Fig. 4. Original image(A).

図 5 減色画像 (A)
Fig. 5. Color subtracted image(A).

てしまう場合がある。図 7 の黒色の線分が要素間辺であり、文字のフォントに関係なく要素間辺が検出できている。

〈4・2〉 文字・絵の混在した画像 原画像 (B)(図 9, 430×353 pixel) は、文字列の傾きが 20 度となるように回転させた後に切り取った画像である。

16 色に減色した場合の処理結果を、表 1(B) と図 10～図 12 に示す。図 10 の白色の線分が検出した要素間辺である。これらの要素間辺の中には、構成要素のレイアウトに起因

図 6 処理過程における拡大図 (A)
Fig. 6. Expanded view during the process(A).

図 7 要素間辺の検出結果 (A)
Fig. 7. Detected lines wrapping components(A).

図 8 要素間辺の傾きのヒストグラム (A)
Fig. 8. The angle histogram(A).

して文書本来の方位とは異なる要素間辺も検出されているが、最終的に検出された方位 (19 度) の決定に寄与している 2 度違い以内の傾き (17～21 度) をもつ要素間辺を抽出して、図 12 で黒色に描画して示す。なお、それ以外の要素間辺 (長いものはノイズとして除去済) は白色で示されて

図 13 原画像 (C)
Fig. 13. Original image(C).

全体図

拡大図

図 14 要素間辺の検出結果 (C)
Fig. 14. Detected lines wrapping components(C).

図 15 要素間辺の傾きのヒストグラム (C)
Fig. 15. The angle histogram(C).

レイアウト情報を利用することによって、文書の方位を検出する実験を行った。同様の実験を雑誌の表紙や広告、ちらしなど 92 枚のカラー文書画像に対して行い、90 枚から

全体図

拡大図

図 16 検出した方位に寄与した要素間辺 (C)
Fig. 16. Result of detected direction(C).

人間が認識する方位と等しい方位を検出することができた。方位を検出できなかった画像は、高度なレイアウト情報の画像 (図 17) と文字列の少ない画像であった。図 17 では、要素間辺の傾きのヒストグラムに大きなピークが 2 つあり、ピークの値の差が小さいことから、領域の局所的な方向性は検出しているが、URL 文字列に着目した文書全体に対する人間の判断のような正しい検出はできていない。筆者らは、構成要素の大きさや単体の構成要素の情報 (例えば、辺の長さ等) をさらに有効に利用することのほか、要素間辺を検出した後に、画像を領域分割し、ヒューリスティックに対応することなどによって、より妥当な結果が得られるものと考えており、構成要素の色、大きさ、配置などに着目した際の、意味理解を避けた人間の方位判断基準検出実験⁽¹³⁾にも一部取り組んでいる。

最後に、今後の課題として、文字列の少ない画像や、高度なレイアウトの画像への対応、および、人間の作成する文書の方位に対する一般的かつ暗黙の知識の研究などを挙げる。

検出できなかった画像

ヒストグラム

図 17 失敗例

Fig. 17. Example of failure.

(平成 13 年 3 月 26 日受付, 同 13 年 7 月 30 日再受付)

文 献

- (1) 劉詠梅, 大西昇, 山村毅, 杉江昇: シーン中の文字列領域の抽出について, 画像ラボ, Vol.10, No.3, pp.23-26 (1999)
- (2) 松尾賢一, 上田勝彦, 梅田三千雄: 情景画像からの看板文字列領域の抽出, 画像ラボ, Vol.9, No.1, pp.56-60 (1998)
- (3) 田中直樹: 表紙画像からの文字領域抽出方式, 画像ラボ, Vol.3, No.1, pp.19-23 (1998)
- (4) 篠川敏行, 長谷博行, 米田政明, 酒井充, 丸山博: 段階的弛緩法による文字列抽出, 画像電子学会誌, Vol.29, No.5, pp.471-479 (2000)
- (5) A. Jain and B. Yu, "Automatic text location in images and video frames", Pattern Recognition, Vol.31, No.12, pp.2055-2076 (1998)
- (6) Yimei Ding, 若林哲史, 木村文隆, 三宅康二: 手書き英単語の局所傾き推定と補正, 電気学会論文誌 C, Vol.120-C, No.12, pp.1854-1859 (2000)
- (7) 酒井憲一郎, 千葉広隆, 野田嗣男: ペン型スキャナ用文字方向自動補正技術, 1999 年電子情報通信学会総合大会, D-11-166 (1999)
- (8) 後藤英昭, 阿曾弘具: 文字行の局所的な直線性を利用した頑健・高速な文字行抽出, 電子情報通信学会論文誌 D-II, Vol.D-II-166, No.3, pp.465-473 (1995)
- (9) E. Kavallieratou, N. Fakotakis and G. Kokkinakis, "Skew angle estimation in document processing using Cohen's class distributions", Pattern Recognition Letters, Vol.20, Issue 11-13, Nov., pp.1305-1311 (1999)
- (10) 秋山照雄, 益田 功: 周辺分布, 線密度, 外接長方形特徴を併用した文書画像の領域分割, 電子情報通信学会論文誌, Vol.J69-D, No.8, pp.1187-1196 (1986)
- (11) Y. K. Chen and J. F. Wang, "Skew detection and reconstruction based on maximization of variance of transition-counts",

Pattern Recognition, Vol.33, pp.195-208 (2000)

- (12) 岸毅: 「ベクトル量子化のアルゴリズム」, CMAGAZINE 7 月号 pp.59-75, ソフトバンクパブリッシング (2000)
- (13) 村田教見, 渡邊朋宏, 伊藤秀隆, 隈元昭: 文書方位に関する人間の判断基準の検出実験, 第 45 回システム制御情報学会研究発表講演会, 2046, pp.195-196 (2001)

泉 井 宏 之 (非会員) 1976 年生. 1999 年 3 月関西大学工学部電気工学科卒業. 2001 年 3 月関西大学大学院工学研究科博士課程前期課程修了. 在学中, 画像処理に関する研究に従事. 現在, (株) 村田製作所に勤務.

伊 藤 秀 隆 (正員) 1994 年京都大学大学院博士後期課程修了. 同年, 関西大学工学部助手. 現在, 同専任講師. 京都大学博士 (工学). 非線形力学現象, 知識情報処理, 画像処理などに関する研究に従事. 電子情報通信学会, 人工知能学会, 計測自動制御学会各会員

隈 元 昭 (正員) 1971 年京都大学大学院修士課程修了. 同年, 関西大学工学部助手. 現在, 同教授. 京都大学工学博士. 知識情報処理, ネットワーク情報処理, 画像処理などに関する研究に従事. 電子情報通信学会, システム制御情報学会, 計測自動制御学会各会員