

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-97609

(P2010-97609A)

(43) 公開日 平成22年4月30日(2010.4.30)

(51) Int.Cl.

G06T 1/00 (2006.01)

F I

G06T 1/00 200D

テーマコード (参考)

5B050

審査請求 未請求 請求項の数 4 O L (全 17 頁)

(21) 出願番号 特願2009-235288 (P2009-235288)
 (22) 出願日 平成21年10月9日 (2009.10.9)
 (31) 優先権主張番号 12/251,689
 (32) 優先日 平成20年10月15日 (2008.10.15)
 (33) 優先権主張国 米国 (US)

(71) 出願人 596170170
 ゼロックス コーポレイション
 XEROX CORPORATION
 アメリカ合衆国、コネチカット州 068
 56、ノーウォーク、ビーオーボックス
 4505、グローバー・アヴェニュー 4
 5
 (74) 代理人 100075258
 弁理士 吉田 研二
 (74) 代理人 100096976
 弁理士 石田 純
 (72) 発明者 フローラン ペロニン
 フランス ドメヌ ルー ジャン ジョ
 レス 107

最終頁に続く

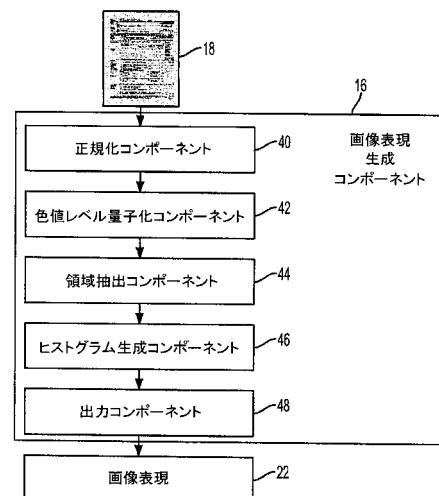
(54) 【発明の名称】 ランレンジスヒストグラムを用いて文書の表現を生成する方法及び装置

(57) 【要約】

【課題】 スキャン画像のような画像の表現を自動的に生成する。

【解決手段】 文書画像の画像表現を生成するためのコンピュータ装置であって、受け取った文書画像を格納するメモリと、前記文書画像内の複数の領域を抽出する領域抽出コンポーネント44と、前記文書画像の前記各領域についてランレンジスヒストグラムを抽出し、それら異なる各領域に対応するランレンジスヒストグラムを組み合わせたヒストグラム生成コンポーネント46と、組み合わせられたランレンジスヒストグラムに基づく画像表現を出力する出力コンポーネント48と、を備える。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

画像の表現を生成する方法であって、
表現を生成すべき画像を受け取り、
前記画像の複数の領域の各々についてランレンジスヒストグラムを計算し、
前記ランレンジスヒストグラムに基づき画像表現を生成し、
前記画像表現を出力する、
方法。

【請求項 2】

請求項 1 に記載の方法であって、前記ランレンジスヒストグラムの計算では、複数の画
素色値の量子化レベルの各々についてランレンジサブヒストグラムを計算し、それらサ
ブヒストグラムを組み合わせる、ことを特徴とする方法。

10

【請求項 3】

請求項 2 に記載の方法であって、更に、
前記ランレンジスヒストグラムの計算の前に、画像の画素色値を複数の量子化レベルへ
と量子化する、
ことを特徴とする方法。

【請求項 4】

文書画像の画像表現を生成するためのコンピュータ装置であって、
受け取った文書画像を格納するメモリと、
前記文書画像内の複数の領域を抽出する領域抽出コンポーネントと、
前記文書画像の前記各領域についてランレンジスヒストグラムを抽出し、それら異なる
各領域に対応するランレンジスヒストグラムを組み合わせるヒストグラム生成コンポーネ
ントと、
組み合わせられたランレンジスヒストグラムに基づく画像表現を出力する出力コンポーネ
ントと、
を備えるコンピュータ装置。

20

【発明の詳細な説明】**【技術分野】****【0001】**

30

この技術は、画像処理に関する。特に、ランレンジスヒストグラムを用いて画像の表現
を生成する装置及び方法との関連で応用可能である。このような装置や方法は、検索やカ
テゴリ分け、クラスタリングなどのアプリケーションに実装することができるが、そのよ
うな応用に限定されるものでないことは理解されるべきである。

【背景技術】**【0002】**

データ処理における便利な処理として、自動又は半自動の画像のカテゴリ分け、自動又
は半自動での類似画像検索や画像クラスタリングがある。例えば、組織化されていない文
書のスキャン画像（スキャンにより得られる画像）のデータベースが与えられた場合、そ
れら画像を文書の種類などのような分類ごとにソート又はカテゴリ分けすることが便利で
ある。納税書類、医療記録などのように、現状、文書を人手により種類ごとにソートして
いる応用分野は多く、このような分野では、文書が特定のフォーム（定型書式）の種類に
該当するか、又はあるフォームのページに該当するかを、特定の光学文字認識を用いるこ
となく自動的に判定できれば便利であろう。その文書又はページについての後続の処理は
、その判定に基づいて行うことができよう。ある関連の応用分野では、文書の画像が与え
られると、画像のデータベースから類似の画像を識別し検索することが便利であろう。

40

【0003】

そのような技術を自動又は半自動で実行できるように、画像の内容に基づき自動で画像
特徴解析（特徴付け）を行う仕組みが望まれる。デジタル画像は本質的には、一般手に数
百万もの多くの画素の各々についての画素値、すなわち色値(colorant values)、の集ま

50

りの形であるため、画像特徴解析技術は、一般に、パッチと呼ばれる画像の小さな部分に基づき、画像から特徴を抽出することに依拠している。1台又は1組の分類器を多数の訓練用画像から抽出された情報を用いて訓練することにより画像をカテゴリ分けする技術が発展してきた。訓練用の画像は、それぞれ、人、風景、動物、建物などのようなあらかじめ定められた被写体カテゴリの集合のなかの1以上のカテゴリに人手でラベル付けされる。分類器は、新たな画像をどのように特徴付けするかを、その画像から抽出された特徴（フィーチャー）とラベル付けされた各画像の抽出特徴とに基づき、学習する。しかしながら、このような技術は、訓練段階で多くの人での作業を要し、多数の画像に人手で各分類（これら分類を見分けられるよう分類器は訓練される）のラベルを付けることがしばしば要求される。さらには、そのような技術は、文書が上下逆さまに読み取られる場合もあるスキャン文書の場合には即座に適用することはできなかった。

10

【先行技術文献】

【特許文献】

【0004】

【特許文献1】米国特許出願公開第2007/0005356号明細書

【特許文献2】米国特許出願公開第2007/0258648号明細書

【特許文献3】米国特許出願公開第2008/0069456号明細書

【特許文献4】米国特許第4949392号明細書

【特許文献5】米国特許第5038381号明細書

【特許文献6】米国特許第5335290号明細書

【特許文献7】米国特許第5822454号明細書

【特許文献8】米国特許第5832118号明細書

【特許文献9】米国特許第6141464号明細書

【特許文献10】米国特許第7124149号明細書

【非特許文献】

【0005】

【非特許文献1】CSURKA, et al., "Visual Categorization with Bags of Keypoints", ECCV Workshop on Statistical Learning in Computer Vision, 2004

【発明の概要】

【発明が解決しようとする課題】

20

30

【0006】

スキャン画像のような画像の表現を生成する自動的な、容易に実装可能な方法が望まれる。

【課題を解決するための手段】

【0007】

1つの側面では、画像表現を生成するための方法では、画像表現の生成対象の画像を受け取り、その画像の複数の領域の各々について、ランレンジスヒストグラムを計算する。画像表現は、計算されたそれら各領域のランレンジスヒストグラムに基づき生成され、出力される。

【0008】

この方法は、画像を複数の領域に区分するステップを更に含んでもよい。

【0009】

この方法では、複数のスケールで領域を区分してもよく、それら領域のいくつかは部分的に重なり合ってもよい。

【0010】

この方法は、ランレンジスヒストグラムを計算するのに先立ち、画像を正規化するステップを含んでもよい。

【0011】

ランレンジスヒストグラムの計算では、更に、画素色値の複数の量子化レベルの各々についてランレンジサブヒストグラムを計算し、それらランレンジサブヒストグラムを

40

50

結合してもよい。

【0012】

この方法では、ランレンジスヒストグラムの計算の前に、画像の画素色値を複数の量子化レベルへと量子化してもよい。

【0013】

ランレンジスヒストグラムの計算では、領域ごとに、複数の方向のそれぞれについてランレンジスサブヒストグラムを計算してもよい。

【0014】

各ランレンジスヒストグラムは、一組の量子化されたランレンジス（ラン長）のそれぞれについて、画像色値レベルが同じ画素のランの数のヒストグラムを含んでいてもよい。

10

【0015】

ランレンジスの量子化は、例えば対数的でよい。

【0016】

画像表現の生成では、複数のランレンジスヒストグラムを連結してもよい。

【0017】

この方法では、更に、ランレンジスヒストグラム群、又はそれらランレンジスヒストグラムを連結したものを正規化してもよい。

【0018】

この方法では、前記画像は、テキスト文書をスキャンした画像を含んでいてもよい。

【0019】

この方法では、文書には、フォーム（帳票）、ビジネス記録、ID（身分証）カード、名刺のうちの少なく1つが含まれてもよい。

20

【0020】

この方法では、画像は例えばビットマップであってもよい。

【0021】

この方法では、画像表現は、複数の値を含むベクトルであってもよい。

【0022】

この方法では、更に、メモリ内で画像表現を画像と対応づけてもよい。

【0023】

この方法では、更に、第2の画像の複数の領域の各々について第2のランレンジスヒストグラムを計算し、第2の画像について計算されたランレンジスヒストグラムに基づいて第2の画像の画像表現を生成し、先に計算した第1の画像の画像表現と今回計算した第2の画像の画像表現とに基づき、第1の画像と第2の画像との類似度を計算してもよい。

30

【0024】

この方法は、画像を分類する方法に用いてもよい。画像を分類する方法では、複数の学習用画像についての画像表現を、上述した画像表現の生成方法を用いて生成し、生成した各学習用画像の画像表現と各学習用画像の分類（クラス）ラベルとに基づき分類器を学習させ、学習済みの分類器を用いて、新たな画像に対し、この画像の画像表現に基づいて分類（クラス）を割り当ててもよい。

【0025】

コンピュータに上述の画像表現の生成方法を実行させるための命令群を表したプログラムを提供してもよい。

40

【0026】

コンピュータに上述の画像表現の生成方法を実行させるための命令群を記憶したメモリと、メモリと接続されそれら命令群を実行するプロセッサと、を備えるコンピュータ装置を提供してもよい。

【0027】

別の側面では、文書画像の画像表現を生成するためのコンピュータ装置を提供する。この装置は、受け取った文書画像を格納するメモリと、前記文書画像内の複数の領域を抽出する領域抽出コンポーネントと、前記文書画像の前記各領域についてランレンジスヒスト

50

グラムを抽出し、それら異なる各領域に対応するランレングスヒストグラムを組み合わせるヒストグラム生成コンポーネントと、組み合わせられたランレングスヒストグラムに基づく画像表現を出力する出力コンポーネントと、を備える。領域抽出コンポーネントは、2以上の異なるスケールで領域抽出を行ってもよい。

【0028】

この装置は、文書画像を正規化する正規化コンポーネントを更に備えていてもよい。

【0029】

この装置は、文書画像中の画素色値を複数の画素色値量子化レベルへと量子化する色値レベル量子化コンポーネントを備えていてもよく、前記ヒストグラム生成コンポーネントは、複数の画素色値量子化レベルの各々についてヒストグラムを生成してもよい。

10

【0030】

別の側面では、画像表現を生成する方法では、文書をスキャンして文書画像を生成し、画素色値をNレベルの色値量子化レベルへと量子化し、文書を複数の異なるスケールで複数の領域へと区分し、Nレベルの色値量子化レベルの各々について且つL個の方向の各々について、複数の値を有するランレングスヒストグラムを計算する。ここで、N、Lはそれぞれ少なくとも2である、文書画像の表現は、ランレングスヒストグラムに基づいて生成される。

【図面の簡単な説明】

【0031】

【図1】 デジタル画像の表現を生成するための装置の例を示す機能ブロック図である。

20

【図2】 一実施例における図1の装置の画像表現生成器の機能ブロック図である。

【図3】 図1の装置を用いて実行することができる、デジタル画像の表現を生成する方法の一例を示すフローチャートである。

【図4】 画像の例を異なるスケールで複数の領域へと分割する処理を説明するための図である。

【図5】 図1の装置に組み込まれるか又はこれに通信可能に接続され得る分類器の例を図式化して示す図である。

【発明を実施するための形態】

【0032】

この例示的な実施の形態は、画像のカテゴリ分け、画像検索、画像のクラスタリングなどを含む様々な画像処理操作に適用可能な画像の表現（画像を表現する表現情報）を生成するための、装置、方法及びコンピュータプログラムに関する。この方法及び装置は、文書処理、特にフォーム（定型）処理に適しており、光学文字認識処理（OCR）を必要としない。

30

【0033】

様々な側面において、文書画像はランレングスヒストグラムを用いて全体として符号化される。文書画像内の複数の領域が1つ又は複数のスケールで抽出される（例えば、文書画像を1つの寸法の領域ごとに分けて抽出する、又は異なる寸法の各々について文書画像を当該寸法の領域ごとに分けて抽出するなど）。個々の領域ごとに、1つ又は複数のランレングスヒストグラムが生成される。すべての抽出されたヒストグラムは、後で連結されて文書画像表現を形成する。

40

【0034】

この方法で特徴付けされる原画像は、JPEG、GIF、JBIG、BMP、TIFF又は画像に用いられるその他の一般的なファイルフォーマットのようなどのようなファイルフォーマットの形でこの装置に受信されてもよく、そのフォーマットは場合によっては処理の前に他の適切なフォーマットに変換されてもよい。画像は単独の画像であってもビデオ画像であってもよく、またスキャンされた画像、写真、グラフィックス、テキスト画像、又はそれらの組み合わせなどであり得る。大まかに言えば、入力されるデジタル画像は、当該画像を形成する画素の配列（アレイ）についての画像データを含み、例えばビットマップの形式となっている。画像データは、単一の色分版について（例えば白黒画像

50

について)、又はRGBなどの一組の色分版について、グレースケールなどの色値を含んでもよく、また、異なった色を表現できる他の色空間で表現してもよい。一般的には、色値は、(RGB, $L^*a^*b^*$, YCbCr等々のように)どのように表現したとしても、単一の色チャンネルの光学的濃度値であり得る。

【0035】

1つの実装例では、画像は、ビットマップ形式のスキャンされた文書画像であり、例えば医療記録や納税書類、名刺、銀行小切手、請求書などをスキャンした画像のように全体がテキスト(文字)であるかテキストが支配的な文書の画像である。1つの側面では、スキャンされた文書は、ユーザが入力する様々な欄を有するテンプレートフォーム(例えばビジネスの書式)などのように、何らかの構造を少なくとも有している。それら文書の欄に(ユーザが記入すること)より、スキャンする前にテンプレートを個々のユーザに固有のものにすることができるとともに、この実装例では、それら文書の欄が少なくとも部分的にユーザに記入されることにより、そのテンプレートフォームの各コピーのスキャン結果は、互いに微妙に異なっているが依然として同じ全体構造を持ったものとなる。それら各欄の内容は、例えば手書き、タイプ打ち、オンライン入力により生成されたものであってよい。1つの文書がいくつかのページを含んでもよく、この場合スキャンにより各ページがそれぞれ1つの文書画像を形成する。

【0036】

ランレングスヒストグラムは、ランレングスの発生頻度のヒストグラムである。ランレングスは、スキャン画像において、指定された方向に沿って連続する、同じ量子化間隔に属する(すなわち、量子化により同じ値を持つ)画素の数である。一つの例では、1からその方向についての最大画素数までの生じ得るランレングスの各々についてそれぞれ1つのヒストグラム・ビンを用意してもよいが、後述するように、範囲を限られた数のインターバルへと量子化すればもっと便利である。

【0037】

図1には、画像表現を生成するための装置の一例が例示されている。この装置は、図示したコンピュータ10のような1以上の電子処理装置に実装することができる。このコンピュータは、表示又はその他の人間が知覚可能な出力を生成するためのディスプレイ12その他の出力装置と、ユーザからの入力を受け取るキーボード14その他の入力装置とを備える。電子処理装置10は、画像表現生成器16を実装するように構成されている。電子処理装置10は、図示したコンピュータであってもよいし、文書スキャンシステム(例えばXerox Xpimプラットフォーム)のプロダクション・イメージング・マネージャやネットワークサーバ、インターネットベースのサーバ、PDA(Personal Data Assistant)、携帯電話などのような他の電子処理装置であってもよい。

【0038】

コンピュータ10は、表現を生成すべき原画像18を受信してデータメモリ20などのメモリにその画像を格納するとともに、画像表現生成器16で生成されたその画像の表現22の格納及び/又は出力又を行うように構成されている。なお、「A及び/又はB」という記載は、Aのみの場合と、Bのみの場合と、A及びBの場合と、のいずれであってもよいことを示す。画像18は、ワークステーション、データベース、スキャナ、ファックス装置、又はディスクやカメラメモリなどのメモリストレージ装置、などのような適切な画像ソース24であればどのようなものからでも入力され得る。

【0039】

例示する実施形態では、画像ソース24は、ケーブル、電話線、ローカルエリアネットワーク、又はインターネットなどのワイドエリアネットワークなどの有線又は無線のリンク26により、モデム、USBポートなどの適切な入出力(I/O)コネクション28を介してコンピュータ10に対し通信可能に接続されたスキャナである。

【0040】

この技術分野で知られているように、典型的なスキャナでは、文書のシートが照明され、その文書シートから反射された光がCCD(Charge Coupled Device)やCMOS(Com

10

20

30

40

50

plementary Metal Oxide Semiconductor) アレイ (接触式画像センサ (C I S : Contact Image Sensor) としても知られている) などの光検知デバイスにより記録され、デジタル画像データへと変換される。文書シートが文書取り扱い部を通して移動するにつれて、又は文書シートが置かれたプラテンに対して相対的に光検知デバイスが移動するにつれて、文書シートの細長い領域が次々に照明される。そして、文書画像のそれら細長い領域がソフトウェアにより組み立てられ、元の文書を表す完全な画像が形成される。

【0041】

画像表現生成器 16 は、ハードウェア又はソフトウェア又はそれらの組み合わせとして具現化することができる。図示の通り、表現生成器 16 は、コンピュータの主メモリ 30 内に記憶されたソフトウェアコンポーネントの形となっており、そのコンピュータの中央演算装置 (C P U) 等のような関連のプロセッサ 32 により実行される。コンピュータ 10 のコンポーネント 20, 30, 32 は例えばデータ制御バス 34 を介して通信する。メモリ 20, 30 は別々のものであっても一体となったものであってもよく、ランダムアクセスメモリ (R A M)、リードオンリーメモリ (R O M)、磁気ディスク又は磁気テープ、光ディスク、フラッシュメモリ、ホログラフィックメモリ、又はそれらの適切な組み合わせなどのような、どのような種類のコンピュータ読み取り可能な媒体の形態をとってもよい。

【0042】

例示する画像表現生成器 16 は、画像 18 の表現を生成するための命令群を有している。図 2 に図式的に示すように、表現生成器 16 は、画像を処理するための様々なコンポーネントを備えている。例えば、文書画像を正規化する正規化コンポーネント 40、文書画像中の画素の色値 (例えばグレーレベル) を量子化する色値レベル量子化コンポーネント 42 (二値画像には不必要)、文書画像中の複数 (一般的には多数) の領域を、例えば 2 つ又はそれより多くの異なるスケールで抽出する領域抽出コンポーネント 44、各領域の欄連レグスヒストグラムを抽出し、それら異なる各領域に対応するヒストグラムを連結して 1 つのベクトルを生成するヒストグラム生成コンポーネント 46 である。ただし、正規化コンポーネント 40 及び色値レベル量子化コンポーネント 42 は必須ではない。出力コンポーネント 48 は、連結されたヒストグラム群に基づく画像の表現 22 を出力する。表現 22 は、したがって、連結されたヒストグラム (すなわち 1 つより多い数のヒストグラム)、又はその連結されたヒストグラムに正規化その他の修正を施して得られる派生値を含んでいてもよく、そのような情報は画像全体を表すものとなる。表現生成器 16 のコンポーネント 40, 42, 44, 46, 48 の機能は、後で更に詳しく説明される。これらのコンポーネントは、適切に組み合わせることができる複数のソフトウェアコンポーネントであってもよく、複数のサブコンポーネントに分割できるようなものであってもよい。

【0043】

画像表現を生成する方法の一例が、図 3 に概略的に説明される。この方法は S 100 で始まる。

【0044】

S 102 で、特徴付けすべき文書のスキャン画像 18 が受け取られる。この画像は、例えば、処理の間メモリ 20 に格納される。その処理対象の文書は、ユーザにより、例えばキーボードを介して選択されてもよいし、例えばその文書の画像とともに送られてくる命令に応じて、自動的に処理されてもよい。

【0045】

必須のことではないが、S 104 で、受け取られた文書画像を正規化してもよい。正規化は、例えばスキュー (傾き) の修正 (deskew)、サイズ変更などを含んでもよい。

【0046】

S 106 では、画像がまだ例えば二値画像の場合のように量子化されていなければ、文書画像中の画素の色値 (例えばグレーレベル) を N 個の離散的なレベルへと量しかしてもよい。ここで、N は、例えば少なくとも 2 で、例えば約 100 までの値でもよく、また例えば約 30 以下の値であってもよい。

10

20

30

40

50

【0047】

S108で、文書画像内の複数の領域が抽出される。

【0048】

S110で、領域ごとに、ランレングスヒストグラムが抽出される。

【0049】

S112で、それら異なる各領域に対応するヒストグラムが連結されて画像ヒストグラムを形成する。連結の前又は後のどちらかに、それらヒストグラムの正規化を行ってもよい。結果として得られる画像ヒストグラムは、したがって、値のベクトル ($v_1, v_2, v_3, v_4, \dots, v_n$) の形で記憶され、文書画像18全体（場合によってその正規化されたもの）を表す。

10

【0050】

S114では、前述の抽出されたランレングスヒストグラムに基づき例えばメモリ20に格納されている画像の表現22が、例えば更なる処理のために、出力される。

【0051】

S116で、そのようにして得られた画像表現22を利用するコンピュータに実装された処理が実行されてもよい。例えば、クラスタリング処理、分類処理、又は類似性判定などが実行される。1つの画像についての画像表現は、クラスタリング又は類似性判定のために、同じ方法で求められた1以上の他の画像の画像表現と比較することができる。他の例では、ある画像についての画像表現を、文書タイプにより分類された人手でラベル付け済みの一組の訓練用画像の画像表現に基づき学習した分類器に入力することにより、その画像を分類してもよい。そして、その文書から、その文書のタイプに応じた特別のデータを抽出してもよい。

20

【0052】

この方法は、S118で終了する。

【0053】

この例示の方法は、1台又は複数の汎用コンピュータ、特定用途コンピュータ、プログラムされたマイクロプロセッサ又はマイクロコントローラ及び周辺集積回路要素、ASICその他の集積回路、デジタルシグナルプロセッサ、ディスクリット要素回路のようなハードワイヤードの電子又は論理回路、PLD、PLA、FPGA又はPALなどのようなプログラマブルロジックデバイスに対して実装してもよい。一般的には、図3に示したフローチャートを実装することが可能な有限状態機械を実装可能な装置であれば、どのような装置であっても、画像表現を生成するための上記方法を実装するのに用いることができる。

30

【0054】

図3に例示された方法は、コンピュータで実行されるようなコンピュータプログラム製品の形で実施してもよい。コンピュータプログラム製品は、制御プログラムを記録したディスク、ハードドライブなどのような有形のコンピュータ読み取り可能な記録媒体であってもよいし、その制御プログラムをデータ信号として表す伝送可能な搬送波であってもよい。コンピュータ読み取り可能な媒体の一般的な形態には、例えば、フロッピー（登録商標）ディスク、フレキシブルディスク、ハードディスク、磁気テープ、その他の磁気媒体、CD-ROM、DVD、その他の光学的媒体、RAM、PROM、EPROM、FLASH-EPROM、その他のメモリチップ又はカートリッジ、音響波又は光波など、無線及び赤外線データ通信の際に生成される波などの送信媒体、又はその他のコンピュータが読み取って使用することが可能な媒体が含まれる。

40

【0055】

例示する方法及び装置の更なる詳細について以下に説明する。

【0056】

A. 画像の正規化（オプション）（S104）

オプションとして、入力された画像を、以下に示す正規化思慮のうちの1以上を実行することで正規化してもよい。

50

【0057】

1. すべての画像についての、クロッピング（トリミング）及び／又は共通のサイズへと拡大又は縮小。文書のエッジ（縁）を探してそれらエッジを取り囲む最小サイズのバウンディングボックスを求め、画像をそのバウンディングボックスのサイズへと切り取る自動クロッピング技術が利用可能である。切り取られた画像は、その後、選ばれたサイズへと拡大又は縮小してもよい。クロッピング処理は、文書が大きすぎるときや文書が小さい場合に実行してもよい（後者の場合はパディングにより）。文書の集合中での最大の文書の実際のサイズ又はそれに近いサイズまで、文書をパディング（空白などで埋めること）する。

【0058】

クロッピングは、異なる複数のフォームファクタ（形状や大きさを決める要因）の画像群が到来するような場合に便利であろう。特に、文書のサイズがそれぞれ異なっている場合がある。米国のレターサイズのフォームは、小切手又はIDカードとはサイズが異なる。このランレングス技術を有効なものとするために、このシステムは、同じサイズ及び同じ縦横比（長さ：幅）の要素を処理するときにもっとも効率的である。

【0059】

自動クロッピングは、スキャナ24で実行してもよい。しかし、ある種のスキャナの自動クロッピング機能は、その機能が作動するか否かによって全く異なる画像をもたらすことがある。例えば、IDカード又は名刺などの小さな文書がスキャンのために白紙のシート上に装着又は配置されているような場合には便利とはいえない。このような場合、自動クロッピングツールはその小さな文書が装着されたページの境界を見つけるだけかもしれないからである。背景スキャンにおける他のノイズ源により、文書が正しく自動クロッピングされるのが妨げられる場合もある。

【0060】

2. 重心へのセンタリング。これはIDカードなどの小さなハードコピー文書の複製やファクシミリ出力を取り扱う場合に便利である。そのような場合、スキャンされた文書は画像中のほんの小さな部分を占めるに過ぎず、その位置がきわめて重要になるかもしれない。その画像の重心は、例えば、ある位置の両側で水平方向及び垂直方向についてアクティブ（すなわち白でない）画素の数が等しくなるような位置を識別することにより見つけてもよい。これは、自動クロッピングツールがクロッピングすべき文書に対応する画像の部分信頼性をもって特定することができない場合などにも有益である。

【0061】

3. スキューの修正。スキュー補正には、エッジ検出と回転によりスキャンされた文書のエッジ群が水平方向及び垂直方向により近い方向になるようにする方法、又はテキストの各行が水平方向に整列するように回転する方法、等の様々な方法が存在する。

【0062】

元のハードコピー文書がすべて固定サイズであり、スキャナが文書を毎回ほぼ同じ位置に繰り返し置いてスキャン画像を生成する場合もあり得るが、そのような場合には正規化は不要であり、避けるべきである。

【0063】

B. 画素色値レベル量子化（S106）

白黒のハードコピー文書の場合ですら、スキャンすると、スキャナは画素についてのグレーレベルが毎回異なったものになり得る。8ビットシステムでは、256の離散的なグレーレベルが認識される。すなわち、デジタル画像のピクセル値は0から255にわたる。複数の色分版（例えばRGBスキャナの場合は3）を持つカラー画像の場合、各色分版が256の離散的な画素値を有する。したがって、デジタル符号化画像はすでに量子化（グレーレベルについて256レベルの量子化、カラー画像については $256^3 = 16,777,216$ 色）されているものの、ランレングス符号化は比較的レベル数が少ない場合に効率がよい。白黒文書の画像については、量子化レベルを2にすることが有益であり、これは画像を二値化することと等価である。カラー画像については、1つの実装では、色

10

20

30

40

50

分版のうちの1つだけ（例えばRだけ）を考慮し、画素値の範囲をNレベルへと量子化する。ここで、Nは正の整数であり、例えば2～10の値である。画像中の各画素は、N個の量子化レベルのうちの1つにより表現されることとなる。量子化レベルのレベル数を多くするほど情報はより多くなるが、ノイズが増える結果を招く可能性があることを理解されたい。RGBのうちの1つのみを用いた初期の試験ではよい結果が得られた。別の例では、2又はそれ以上の色分版を量子化する。例えば、RGB成分の各々に2つの色値量子レベルの1つを割り当てれば、 $N = 2 \times 2 \times 2 = 8$ の量子化値を得られる。

【0064】

C. 領域抽出 (S108)

図4に例示したように、次に画像中の領域50の集合が抽出される。それら領域は画像を区分することにより得られる。画像を規則的に、例えば 4×4 の領域へと区分してもよい。またこの代わりに、例えばすべての領域が（ほぼ）同じ数の画素を有することを保証する区分処理など、不規則的な区分を行ってもよい。

【0065】

画像を、異なるスケール（寸法）で、複数の異なる分け方で区分してもよい。この場合、画像中のいくつかの部分が複数の領域に含まれることとなる。図4は、異なるスケールで分割されたスキャン画像18の例を示している。以下の区分が例示されている。すなわち、 1×1 （画像全体）、 2×2 、 4×4 であり、それらは順に符号52、54、56が付されている。これは、 $1 \times 1 + 2 \times 2 + 4 \times 4 = 21$ の領域に対応する。それら異なるスケールの領域は互いに重なり合っており、それぞれ、より大きい領域の部分領域となっている。他の例では、それら領域のうちの少なくともいくつかは、部分的に重なり合う。例えば、画像を 3×3 個の領域に区分することにより得られる領域は、 2×2 に区分したときに得られる領域に対して部分的に重複することになる。代わりに、同じサイズ又は異なるサイズの領域をオフセットし（ずらし）てもよい。大略的には、それら複数の領域が一緒になって（場合によっては正規化されている）文書画像全体を網羅する。例示した領域は矩形形状のブロックであるが、三角形、正方形、六角形、不整形(irregular shape)、それら形状の組み合わせなどの、他の様々な形状の領域を用いてもよいことを理解されたい。

【0066】

画像18の異なる各領域50（場合によっては異なるスケールで区分されたもの）についてのヒストグラムをそれぞれ生成することにより、文書タイプ（種類）に関連する画像間の相違が符号化に反映される。構造化文書を取り扱う場合、文書はそれぞれ異なったレベルの構造を有してもよい。例えば、フォームの参照番号が左上の象限のどこかに現れ、タイトルが最上部の中央近傍に現れてもよい。しかし、スキャンの処理により文書の内容がいくぶんシフトすることになる可能性があり、スケールが小さい場合（例えば画像が多数の領域に分割された場合）、タイトルのような特徴が同じ領域に常に現れるとは限らないかもしれない。また、タイトルは、文書の1ページのみには現れるかもしれない。したがって、特徴が異なれば、スケールが異なる場合だけでなく位置が異なる場合にもヒストグラム群に対して異なる影響を及ぼす。ランレングスは文書を複数の領域に分割しなくても（すなわち画像全体を用いても）計算することができるものの、結果は、異なったスケールで区分された領域群を考慮することはフォームなどの構造化されたコンテンツ（内容）に対して特に有益であることを示唆している。

【0067】

D. 領域のランレングス符号化 (S110)

各領域50にて、ランレングスヒストグラムが抽出される。詳細には、ランレングスは、当該領域内の各画素ラインを選ばれた方向にスキャンして画素のランを識別することにより抽出される。ランは、選ばれた方向（例えば水平、垂直、対角線、反対角線）に連続する同じ色値を持つ画素の連鎖のことである。ランの長さ（ランレングス）は、その連鎖が含む画素の数である。二値画像の場合、各ランはそれぞれすべて黒画素（量子化値=0）であるか、又はすべて白画素（量子化値=1）である。より一般的には、2以上の色値

レベルが用いられる場合、1つのラン内の画素はすべて同じ量子化値を持つ。ランレンジスヒストグラムは、M個のランレンジス（ラン長）の各々についてのランの数のヒストグラムである。一例では、ヒストグラム・ビンの数を減らすためにランの長さを量子化する。量子化は、例えば線形のものでよく、例えばあり得るランレンジスの範囲を同サイズの複数の間隔へと分割するものでよい。あるいは、量子化は、対数的な分割その他不均一な分割により発生頻度の高いランレンジスをより弁別できるようにするものでもよい。対数的な方法では、量子化は例えば以下になる。

【0068】

【表1】

量子化 レベル／値	1	2	3	4	5	6	7	8	9	10
画素 ラン レンジス	1	2	3-4	5-8	9-16	17-32	33-64	65-128	129-256	>256

10

【0069】

これにより結果として10の間隔（ $M=10$ ）が得られる。了解されるように、ランは、選ばれた方向（例えば、画像の行方向、列方向、又は対角線方向）に沿った画素の総数より長くなることはできない。例示した10レベルの量子化は、100Kのオーダーの画素数を含む画像に適している。より解像度が高い画像には、画素の最大ラン（ラインの長さ）が大きいので、最高量子化レベルについての最大間隔セパレータ（最大間隔区切り）が256を超えるように選択されるようにしてもよい。

20

【0070】

1つの実装例では、黒と白のラン（二値画像の場合）のヒストグラムが複数の方向、例えば水平、垂直、対角線、反対角線方向、についてそれぞれ計算される。これにより、8個のサブヒストグラム（すなわち、水平の黒、水平の白、垂直の黒、垂直の白、対角線の黒、対角線の白、反対角線の黒、反対角線の白）が求められる。したがって、1つの領域がこれら8個のサブヒストグラムを連結（結合）したものにより記述される。したがって、画素量子化レベルが10の場合の例では、得られる領域のヒストグラムは $10 \times 8 = 80$ の値を含む。一般的な場合では、領域のヒストグラムに含まれる値の数は、 $M \times N \times L$ である。ここで、Mは画素ランレンジス量子化レベルの数であり、Nは色値量子化レベルの数であり、Lは方向の数である。M、N及びLはすべて整数である。

30

【0071】

領域のヒストグラムは正規化してもよい。正規化の方法には例えば以下の3つがある。

1) 領域のヒストグラムを、 $N \times M \times L$ の値（この例では80）の総和が1となるように全体的に正規化する。

2) 各方向及び各画素量子化値（例えば黒か白か）に対応するサブヒストグラム群を、それぞれ独立に正規化する。

40

3) 各方向を個別に正規化する。すなわち、 $N \times M$ の値の総和が1（この例では、黒と白のサブヒストグラムに対応する 2×10 の値の総和が1）になるように正規化する。

【0072】

一般的にはそれら方法のうちのただ1つが用いられるが、複数の正規化方法を用いることも考慮される。正規化はいくつかの場合には便利であるが、必ずしも必要ではない。

【0073】

E. グローバル（全体）画像表現（S112）

最終的な画像表現22は、各領域に対応する領域ヒストグラムを単に連結（結合）したものでよい。あるいは、例えば主成分解析などを用いることで、固定の要素数まで次元を

50

減らしてもよい。

【0074】

F. アプリケーション（応用）

1. 画像検索：2つの画像同士の距離は、それら画像のヒストグラム表現22同士の距離として定義することができる。ヒストグラム同士の類似度／距離についての適切な尺度としては、内積(dot product)、バタチャリヤ(Bhattacharyya)距離、カルバック-ライブラー情報量(Kullback-Leibler divergence)（対称化したものであっても、そうでなくてもよい）、ジェンセン-シャノン情報量(Jensen-Shannon divergence)、カイ二乗距離等があり、これらは例えば類似度を計算するのに用いてよい。例えば、画像18の大きな集合S、及びそれら画像に対応するヒストグラム表現22の集まりが、メモリに格納される。ユーザは、新たな画像を入力するか、又は前述の集合の中から1つを選び、それをターゲット画像として用いる。ターゲット画像のヒストグラム表現に対する各画像のヒストグラム表現の類似度に基づき、その集合の中から類似画像を自動的に検索される。例えば、この装置は、閾値以上の類似度を持つヒストグラム表現を持つ画像を検索する。また別の例では、類似度値が最も高い10個（又は他の選ばれた数の）画像を集合から検索してもよい。画像検索のためのコンポーネントは、画像表現生成器内に含まれていてもよいし、画像表現生成器の別のコンポーネントであってもよいし、別のコンピュータであってもよい。上述した方法の例は、このアプリケーションに有効であることがわかった。

10

【0075】

2. 画像分類：画像のヒストグラム表現は、図5に模式的に示したように、ほとんどのような判別分類器60に供給してもよい。例示する分類器は、スパースロジスティック回帰(sparse logistic regression)、線形判別分析(linear discriminant analysis)、ニューラルネットワーク、サポートベクターマシン(support vector machines)、ナイーブベイズ(naive Bayes)などのように、判別モデル(discriminative models)に基づくものであっても、生成モデル(generative models)に基づくものであってもよい。例えば、税務書類の場合のように、分類器は、クラスの集合の中の1つに手作業で割り当てられた画像群を用いて学習させてもよく、例えば、1040-ページ1、1040-ページ2、1040a-ページ1、1040b-ページ2、スケジュールA、スケジュールBなどに対応する約20のクラスがある（「1040」等は、米国の納税フォームの名称）。分類器は、新たな画像のベクトル表現に基づき、その画像に分類（クラス）ラベル62を割り当てる。分類された画像は、ユーザの閲覧のために、例えばディスプレイ12に表示してもよい。分類器60は、画像表現生成器16に含まれていてもよいし、同じコンピュータ10内の別のコンポーネントであってもよいし、通信可能に接続された別のコンピュータであってもよい。当該方法の分類への応用（アプリケーション）の例を以下に示す。

20

30

【0076】

3. 画像クラスタリング：画像表現26は非負なので、確率的潜在意味論解析(PLSA: Probability Latent Semantic Analysis)クラスタリングなどのような非負行列因子分解(non-negative matrix factorization)技術に適している。コンピュータ10、又は通信可能に接続されたコンピュータは、画像クラスタリングのためのコンポーネントを有していてもよい。そのようなコンポーネントは、集合内の画像同士の距離の計算結果の行列を生成してもよい。この行列は、画像群をクラスタリングするために、例えばPLSAを用いて処理される。

40

【0077】

生成的な視覚情報に基づいた分類(GVC: Generic Visual Classification)などの既存の分類手法と比べたときの本方法の利点は、ランレングスヒストグラムの方がフィッシャーカーネル(Fisher Kernel)に基づくGVC表現よりも一般に4～5倍程度高速に計算できるという点である。

【0078】

以下に示す例は、ある分類の問題に対してこの方法を適用する例である。

【0079】

50

〔例〕

納税フォーム、IDカード、契約書及び医療フォームなどの文書画像を含んだデータベースについてこの方法をテストした。この方法が文書を分類する能力をシミュレートするために、データベース内の画像群のうちの一部のものを、それぞれ一組の分類（クラス）のうちの1つに人手によりカテゴリ分けした。これらラベル付けした画像は、（スパースロジスティック回帰により）分類器に学習させるのに用いた。そして、学習済みの分類器をラベル付けされていない文書画像を分類するのに用い、割り当てられた分類を人手で精査することによりこの方法の精度を評価した。以下の3つのデータベースについての結果を例として示す。

【0080】

データベース1：このデータベースは、20種類の所得税フォームの5590個の画像を含んでいた。1カテゴリあたり10個の画像（合計200個の学習用画像）をランダムに抽出して学習に用いた。残りの画像はテストに用いた。

【0081】

データベース2：このデータベースは、IDカード、契約書などのような、ある営利企業により通常処理される種類の様々な種類の10000個以上の画像を含んでいた。

【0082】

データベース3：このデータベースは、別の企業で典型的に処理される、本質的に種類が異なる複数のフォームについての様々な種類の文書の3000個の画像を含んでいた。

【0083】

それぞれの場合において、10foldの交差検定(cross-validation)をそのデータベースに対して用いた（実験は、10の異なるサンプリングを用いて10回繰り返した）。それら3つの方法を評価した。

【0084】

方法1（比較）は、生成的視覚的分類（GVC）に基づくものであり、「GENERIC VISUAL CLASSIFICATION WITH GRADIENT COMPONENTS-BASED DIMENSIONALITY ENHANCEMENT（勾配成分に基づく次元の強調を用いた生成的視覚的分類）」と題するFlorent Perronninによる米国特許出願公開2007/025648号明細書に記載されている。この方法では、画像分類システムは、複数の画像分類（クラス）に対応する複数の生成モデル(generative model)を含んでいる。各生成モデルは、一般的な視覚的語彙(visual vocabulary)と画像分類に特有の視覚的語彙とのマージ（併合）を実現する。勾配に基づくクラス類似度モデラーは、各生成的モデルに対応する画像のモデルフィッティングデータを生成するモデルフィッティングデータ抽出器と、各生成的モデルについての、当該生成的モデルにより定義されるベクトル空間内でのモデルフィッティングデータの勾配に基づくベクトル表現を生成する次元強調器と、を有する。画像分類器は、勾配に基づくクラス類似度のベクトル表現に基づき画像を複数の画像クラスに分類する。

【0085】

方法2：この方法では、この実施形態で説明した方法により得たランレングスヒストグラムを用いた。ただし、単一スケール（すなわち画像を複数の領域に分けない）で行った。

【0086】

方法3：この方法では、図4に例示したような領域分割によりこの実施形態で説明した方法により得た複数スケールのランレングスヒストグラムを用いた。

【0087】

表2は、上記3つの方法により得られた結果（誤り率。すなわち謝って分類された画像のパーセンテージ）を示す。

【0088】

10

20

30

40

【表 2】

	データ ベース 1	データ ベース 2	データ ベース 3
方法 1 (GVC) (比較)	0.00%	3.55%	11.69%
方法 2 (単一スケールランレングス)	0.39%	8.63%	18.76%
方法 3 (複数スケールランレングス)	0.00%	3.96%	11.80%

10

【0089】

結果から分かるように、実施形態の方法によれば、複数スケールヒストグラムを用いた場合に GVC に近い結果が得られた。データベース 1 については、単一スケールのランレングスでもよい結果が得られたが、データベース 2 及び 3 については、複数スケールランレングスヒストグラムの方が単一ランレングスヒストグラムよりも効果が顕著であった。

【0090】

了解されるように、これらの方法では限られた数の画像のみを分類器の学習に用いた。実際の運用では、もっと多くの学習用画像を用いてもよく、これにより誤り率をより低くすることができるであろう。

【0091】

20

実施形態の方法のいくつかの利点を GVC との比較により以下に説明する。

【0092】

1. 実施形態の画像表現では分類器がモデルを学習する必要がないが、GVC は「視覚的語彙」の学習が必要である。実施形態の方法において最適なランレングスインターバル（間隔）の学習を行うことも考えられる。しかし、表 2 に示したように、そのような付加的な処理を行わなくても良好な性能が得ることができる。

【0093】

2. この実施形態の画像表現を計算するコストは、(フィッシャーカーネルに基づいて) GVC 表現を計算するコストと比して著しく低い。例えば、2.4 GHz の Opteron コア CPU のマシンでは、複数スケールランレングスの計算は 1 画像あたり約 50 ms を要するのに対し、GVC では 175 ms のオーダーの時間を要する。

30

【0094】

3. 画像の表現を一度計算すると、90, 180, 270 度回転させた同じ画像の表現を計算するのは容易である。単にヒストグラムのいくつかの順序を変えればよい。これによりスキャン方向が不明の場合でも非常に高速に処理できる。GVC の場合、スキャンした画像から 4 つの表現を別々に生成する必要があるだろうし、それら 4 つの表現が分類問題のために用いられるであろう。

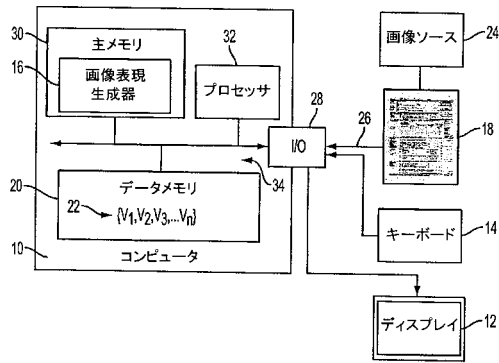
【符号の説明】

【0095】

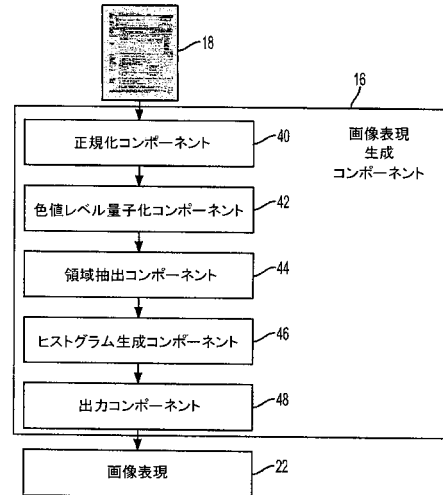
16 画像表現コンポーネント、40 正規化コンポーネント、42 色値レベル量子化コンポーネント、44 領域抽出コンポーネント、46 ヒストグラム生成コンポーネント、48 出力コンポーネント、22 画像表現。

40

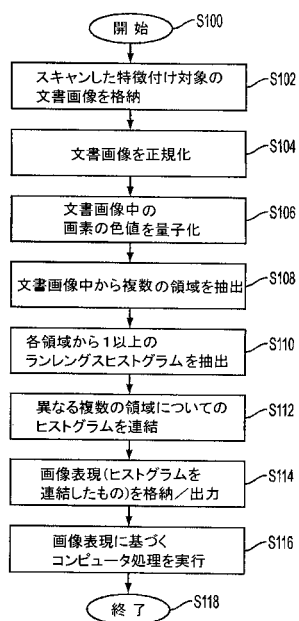
【図 1】



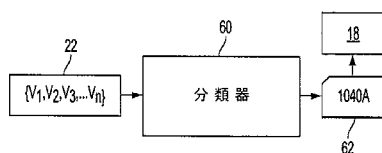
【図 2】



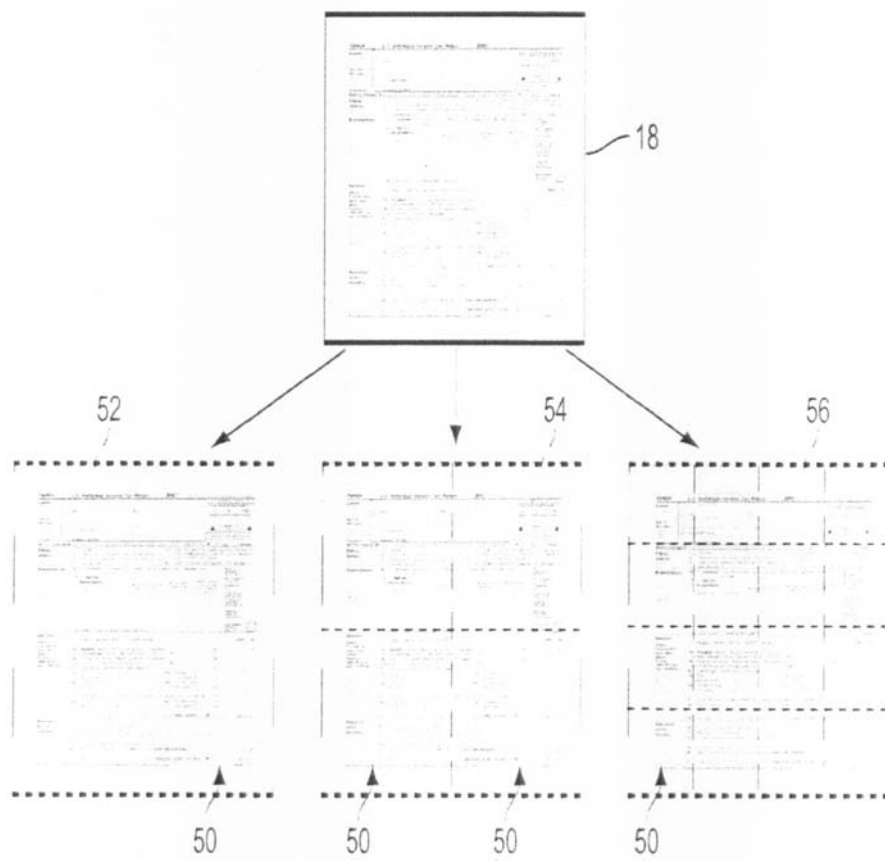
【図 3】



【図 5】



【図 4】



フロントページの続き

(72)発明者 ダミアン クラメ
フランス グルノーブル ルー サン ジャック 3 2

(72)発明者 フランソワ ラニエ
フランス ヴェノン プレ ペロー

Fターム(参考) 5B050 BA10 BA16 CA07 EA08 EA09 EA18 FA02 FA19 GA08