

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5214642号
(P5214642)

(45) 発行日 平成25年6月19日(2013.6.19)

(24) 登録日 平成25年3月8日(2013.3.8)

(51) Int.Cl.

G06K 9/72 (2006.01)

F 1

G 06 K 9/72

Z

請求項の数 5 (全 18 頁)

(21) 出願番号 特願2010-16733 (P2010-16733)
 (22) 出願日 平成22年1月28日 (2010.1.28)
 (65) 公開番号 特開2010-176672 (P2010-176672A)
 (43) 公開日 平成22年8月12日 (2010.8.12)
 審査請求日 平成25年1月18日 (2013.1.18)
 (31) 優先権主張番号 12/361,178
 (32) 優先日 平成21年1月28日 (2009.1.28)
 (33) 優先権主張国 米国(US)

早期審査対象出願

(73) 特許権者 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 06856、ノーウォーク、ピーオーボックス 4505、グローバー・アヴェニュー 45
 (74) 代理人 100079049
 弁理士 中島 淳
 (74) 代理人 100084995
 弁理士 加藤 和詳
 (72) 発明者 ホセ エー. ロドリゲス セラノ
 スペイン バルセロナ 08950 エスピルゲス デ ジョブレガット ベルヘ
 デ ラ メルセ 67 7 1
 最終頁に続く

(54) 【発明の名称】ベクトル系列用モデル基準比較指標及びそれを用いたワードスポットティング

(57) 【特許請求の範囲】

【請求項 1】

比較方法を実行するためにデジタル処理装置によって実行し得る命令を格納する持続性記憶媒体であって、

入力オブジェクトを表す順序化された入力ベクトル系列を生成するステップと、
入力ベクトル重みパラメータの順序化された系列を生成するために、前記入力オブジェクトの先駆的情報及び参照用の重みパラメータの順序化された系列を含むガウス混合モデル(GMM)モデル基準を有する参照用の半連続隠れマルコフモデル(SC-HMM)を利用して前記順序化された入力ベクトル系列をモデル化するステップと、

定量的な比較指標を生成するために、前記入力ベクトル重みパラメータの順序化された系列と、前記参照用重みパラメータの順序化された系列とを比較するステップと、

を含む比較方法を使用することを特徴とする持続性記憶媒体。

【請求項 2】

前記格納された命令は、GMMモデル基準を有するSC-HMMを利用して順序化された単一の参照用のベクトル系列をモデル化することによって前記参照用のSC-HMMを生成するように実行し得ることを特徴とする請求項1に記載の持続性記憶媒体。

【請求項 3】

前記格納された命令は、前記GMMモデル基準を有するSC-HMMを利用して複数の順序化された参照用のベクトル系列を含む訓練セットをモデル化することにより前記参照用のSC-HMMを生成するように実行し得ることを特徴とする請求項1に記載の持続性記憶媒体。

記憶媒体。

【請求項 4】

前記格納された命令は、入力画像から順序化された特徴ベクトルの系列を抽出しつつ前記 GMM モデル基準を有する SC - HMM を利用して前記順序化された特徴ベクトルの系列をモデル化することによって前記参照用の SC - HMM を生成するように実行し得ることを特徴とする請求項 1 に記載の持続性記憶媒体。

【請求項 5】

前記格納された命令は、(I) 訓練セットを生成するため対応する 1 つ以上の参照画像から 1 つ以上の順序化された特徴ベクトルの系列を抽出するとともに、(II) 前記 GMM モデル基準を有する SC - HMM を用いて前記訓練セットをモデル化することにより前記参照用の SC - HMM を生成するように実行し得ることを特徴とする請求項 4 に記載の持続性記憶媒体。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は情報処理技術、情報検索技術、オブジェクト分類技術、オブジェクト認識技術、オブジェクト比較技術、ワードスポットティング技術、及びそれらに関連する技術に関する。

【背景技術】

【0002】

20

手書き文書の自動処理は書類の処理を促進し、ある主題に関連する書類の検出等を可能とする。そのような処理は、例えば光学的スキャニングによって、手書き文書のデジタル画像を生成することから始まる。光学式文字読み取り (OCR) として知られる技術においては、英語のテキストコンテンツの場合のアルファベットや数字などの文字を識別するために、デジタル画像が処理される。タイプライタで印刷された、あるいはコンピュータで生成したフォントを用いてプリンタで印刷された、標準化されたテキストに対しては、OCR は一般的にテキストのほとんどあるいはすべてを識別することができる。しかし、手書きのテキスト内容に対しては OCR はあまり有効ではない。それは、タイプライタの標準化された铸造文字、あるいはコンピュータの標準化されたフォントに比べて、手書き文字は変形の度合いが大きいからである。一人の書き手の手書きに関してもかなりの変形があるが、数多くの異なる人によって書かれた文書に関しては、その変形の度合いがはるかに大きくなる。これは、ビジネス文書や官公庁書類に記入された書類を大量に処理する場合によくみられることである。

30

【0003】

ワードスポットティングは、OCR とは基本的に異なる手法を用いる。文書画像をテキスト化するために個々の文字を識別して OCR テキストを処理しようとするのではなく、ワードスポットティングでは、画像として語あるいは句のレベルで作用し、標準的あるいは代表的な単語と類似した単語を識別、すなわち「スポット」する（見つけ出す）。ワードスポットティングでは、単語の全体的な形状あるいは外形に関する情報をうまく利用するので、OCR に比べて手書きのテキストコンテンツに対してより正確である。更に、ワードスポットティングは語あるいは句の中の個別の文字を変換するのではないので、計算処理上の効率が高い。

40

【0004】

ワードスポットティングは例えば、文書中の 1 つか 2 つ、あるいは数個のキーワード、もしくはキーフレーズを識別して文書を並べ替えたり、分類するのに利用される。ここでは手書きのテキストについて述べたが、ワードスポットティングは、コンピュータ、あるいはタイプライタで作成したテキストにも、画像形式で適用できる。

【0005】

手書き文書に対するワードスポットティングの基本的なプロセスは以下のものを含む。（1）文書の各ページを光学的にスキャンするステップ。（2）個別の単語や句の画像を識

50

別するために、画像を分割するステップ。(3) 単語や句の画像のそれぞれの定量的表示を抽出するステップ。(4) その定量的表示を、スポットティング対象の単語や句に対応することがわかっている単語や句(あるいは単語や句の訓練用画像セット)から抽出された参照用表示との比較を行うステップ。抽出操作(3)の代表的なやり方としては、語あるいは句の画像に沿って左から右(あるいは右から左)へ選択された様々な点において特徴ベクトルが抽出され、それによって順序化された特徴ベクトル系列が生成される。次いで、比較操作のステップ(4)で、入力された特徴ベクトル系列と参照用の特徴ベクトル系列とを比較するために、任意の適切な比較指標が用いられる。

【0006】

通常利用される比較指標は、例えばSakoe(迫江)等による非特許文献1に述べられている、動的時間伸縮法(DTW)の距離指標である。ここでの距離は、2つの系列間の最適整合性の指標となるものであり、ベクトルとベクトルの比較を集積する。DTW距離指標は、計算する上でも効率が良く、手書きの比較の際によく見られる、手書きの語あるいは句のいびつな変形に対処できる。しかし、DTWは限られた情報、すなわち入力ベクトル系列及び参照ベクトル系列に含まれる情報を活用するだけである。またDTWは、語あるいは句の参照画像のトレーニングセットを導入するのに便利な方法ではない。

10

【0007】

Jebaraによる非特許文献2では、積率カーネル(PPK)に基づく距離指標が述べられている。この手法では、抽出操作(3)は連続隠れマルコフモデル(C-HMM)、つまり連続的な状態出力確率を持ったHMMを使用している。状態出力確率は一般に、ガウス混合モデル(GMM)である。入力及び参照の語あるいは句のモデルは従って連続隠れマルコフモデルである。次に、比較操作のステップ(4)では、PPKを利用して入力と参照用の語あるいは句の画像のC-HMM間の距離を計算する。このモデル基準の手法においては、参照モデルは参照用の単語や句の手書きサンプルのトレーニングセットに基づいて生成することができる。従って、DTW指標に比べて大きな利点がある。しかし、Jebaraの手法は、PPKの計算が状態間の、すなわちGMM間の距離の計算を本質的に必要とするので、計算が膨大となる。

20

【0008】

更にDTWもJebaraによる手法も、手書き一般に関する先駆的な情報は組み込まない。つまりこれらはいずれも、手書きの語や句のそれぞれが、アルファベットや他の先駆的に知られた記号の組から選択された、一連の手書きの文字や数字(あるいはもつと一般的には、記号)で構成されているという知識をうまく活用したものとはなっていない。このような知識を利用できないと、計算速度や精度、あるいはその両方に対してマイナスの作用を及ぼしかねない。

30

【先行技術文献】

【非特許文献】

【0009】

【非特許文献1】Sakoe et al., "Dynamic programming algorithm optimization for spoken word recognition", IEEE Transactions on Acoustics, Speech and Signal processing, pp. 159-65 (1978)

40

【非特許文献2】Jebara et al., "Spectral clustering and embedding with hidden Markov models", in the 2007 European Conference on Machine Learning (ECML 2007)

【発明の概要】

【発明が解決しようとする課題】

【0010】

ここでワードスポットティングは説明の例として用いられているが、以下で述べることは

50

、（1）オブジェクトを順序づけられたベクトル系列でモデル化し、（2）その順序化されたベクトル系列の距離指標あるいは他の定量的比較指標を計算する、ことにより、より一般的にオブジェクトの定量的比較をおこなうことに関する。このような定量的比較は、前述のオンライン及びオフラインの手書き認識への応用はもとより、音声認識（この場合のオブジェクトは話された単語やフレーズを表す音声系列である）、画像コンテンツスポットティング（この場合のオブジェクトは対象オブジェクトを表す画像部分である）、生物配列処理（この場合のオブジェクトはDNAの部分配列、タンパク質の部分配列、などである）、などへの応用も可能である。

【課題を解決するための手段】

【0011】

10

ここで、説明のための例として挙げる1つの例示的実施形態において、記憶媒体が、モデル基準を有する参照用の半連續隠れマルコフモデル（S C - H M M）を備えた入力ベクトル系列と、参照用の重みパラメータの順序化された系列との比較を実行し得る命令を格納する。その比較には以下のステップを含む比較方法を利用する。入力ベクトルの重みパラメータの順序化された系列を生成するために、入力ベクトル系列をモデル基準を有するS C - H M Mを利用してモデル化するステップと、定量的な比較指標を生成するために、入力ベクトル重みパラメータの順序化された系列と、参照用重みパラメータの順序化された系列とを比較するステップと、を含む比較方法を利用する。

【0012】

20

ここで、説明のために挙げる別の例示的実施形態において、開示されるオブジェクトの比較方法が以下のステップを含む。第1のオブジェクトの第1の順序化されたベクトル系列表示を生成するステップと、第2のオブジェクトの第2の順序化されたベクトル系列表示を生成するステップと、汎用基準を用いた半連續隠れマルコフモデルを利用して第1の順序化ベクトル系列表示をモデル化して生成したモデルパラメータの第1の順序化系列によって第1のオブジェクトを表示するステップと、汎用基準を用いた半連續隠れマルコフモデルを利用して第2の順序化ベクトル系列表示をモデル化して生成したモデルパラメータの第2の順序化系列によって第1のオブジェクトを表示するステップと、第1の順序化系列と第2の順序化系列とを比較して定量的比較指標を生成するステップと、である。

【0013】

30

ここで、説明のための更に別の例示的実施形態において、以下のステップを含む文書化処理システムが開示される。複数の単語あるいは句の画像を生成するために、文書ページの少なくとも1つの画像を分割するステップと、複数の単語あるいは句の画像の入力用順序化ベクトル系列表示を生成するステップと、固定された平均および分散パラメータを有するガウス分布成分から構成されるガウス混合モデル（G M M）基準を利用する半連續隠れマルコフモデルを用いて、入力用順序化ベクトル系列表示をモデル化して対応する重みパラメータ系列を生成するステップと、その重みパラメータ系列を対象モデルの1つまたは複数の語あるいは句と比較することに基づいて文書にラベル付けするステップと、である。

【図面の簡単な説明】

40

【0014】

【図1】ワードスポットティングエンジンの概略図である。

【図2】図1のワードスポットティングエンジンに用いる汎用基準を生成するためのシステム例の概略図である。

【図3】図1のワードスポットティングエンジンの変形版であり、対象モデルの語あるいは句がワードスポットティング操作の前に生成され格納される。

【図4】図3のワードスポットティングエンジンを利用した文書ラベル付けシステムの概略図である。

【図5】コンピュータで生成したテキストの対象語“M a d a m e”をコンピュータで生成した10個の異なるフォントで示した、10個の画像のトレーニングセットである。

50

【発明を実施するための形態】

【0015】

図1を参照すると、例示のワードスポットティングエンジン10が手書きテキスト画像12を処理して、手書きテキスト12と、対象の語あるいは句の1つまたは複数の手書きサンプル画像14との類似性を定量的に評価する。手書き文書を光学的にスキャンすることにより、あるいは、ユーザがペン、スタイラスやその他の適当な手入力用具を用いて手入力できるタッチセンサスクリーンを備えた、例えばタブレット型パーソナルコンピュータ（“タブレットPC”）やパーソナルデータアシスタント（PDA）を用いて手書きのコンテンツを電子的に直接入力することにより、画像12、14が適切に生成される。画像12、14の生成には所望により、文書のページ画像あるいは他の大きな画像の一部分として、テキスト画像12、14を分離あるいは識別するための分割が行われてもよい。画像12、14は、任意の適当な画像フォーマット、例えばJPEG、TIFF、GIF、PNGなどで格納することができる。画像12、14は所望により、ゆがみ、傾斜および寸法が正規化され、これらの観点からの画像12、14の比較のし易さを高めることもできる。

【0016】

ワードスポットティングエンジン10が特徴抽出エンジン16を含み、手書きのテキスト画像12を処理して、対応する順序化された特徴ベクトル系列22を生成する。適切な手法においては、1つのウィンドウを左から右へスライドさせ、それぞれのウィンドウに対してひと組の特徴（即ち特徴ベクトル）を計算することによって、順序化された特徴ベクトル系列22が得られる。適切な特徴のタイプは、例えば以下のようない文文献で述べられている。Martí、ほか：“Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system”，Int. J. of Pattern Recognition and Artificial Intelligence，vol. 15 pp. 65-90 (2001)。

Vinciarelli ほか：“Offline recognition of unconstrained handwritten texts using HMMs and statistical language models”，IEEE Transactions on Pattern Analysis and Machine Intelligence，vol. 26 no. 6，pp. 709-20 (2004)。

Rodriguez ほか：“Local gradient histogram features for word spotting in unconstrained handwritten documents”，2008 International Conference on Frontiers in Handwriting Recognition (ICFHR 2008)。

同様に、ワードスポットティングエンジン10は、特徴抽出エンジン16'を含み、これは、対象の語あるいは句の手書きサンプルの1つまたは複数の画像14を処理して、対応する1つまたは複数の参照用の順序化された特徴ベクトル系列24を生成する。特徴抽出エンジン16'により遂行される変換は、特徴抽出エンジン16により遂行される変換と作用上は同一である。従って、別々の、ただし機能的には等価な特徴抽出エンジン16、16'を利用することも考えられるが、実際に2つのエンジン16、16'が1つの同一エンジンであってもよい。

【0017】

一方における順序化された特徴ベクトル系列22と、他方での1つまたは複数の参照用の順序化された特徴ベクトル系列24とが、画像12、14の比較のための適切なベースを提供する。実際に、单一の画像を参照する既存のワードスポットティングエンジンにおいては、動的時間伸縮法（DTW）の距離が順序化系列22、24に直接適用されて、定量

的な比較指標が計算される。しかしながらこの手法は複数の参照画像へは適用できない。そして、さらにこのような手法は、手書きの語や句のそれぞれが、アルファベットやその他の先駆的に知られた記号の組から選択された、一連の手書きの文字や数字（あるいはもつと一般的には、記号）で構成されているという先駆的な知識をうまく活用したものとはなっていない。ここでは、このような先駆的な知識を活用して計算速度を向上させ、精度を向上させ、または速度と精度の両方を向上させることが望まれる。

【0018】

この目的に対して、順序化された特徴ベクトル系列22を、半連續隠れマルコフモデル（S C - H M M）用モデル化エンジン26を用いてモデル化する。このモデル化エンジン26は、テキスト画像12、14を構成することが予想される手書きの文字や数字、もしくはより一般的には記号、の組に関する先駆的な知識を組み込んだ汎用基準30を使用する。ここで説明する例示的な実施形態においては、汎用基準30はガウス混合モデル（G M M）であり、ガウス分布成分が固定された平均値および分散パラメータ（ここで、"分散パラメータ"は広義に使用されており、分散、標準偏差や他のガウス分布成分の"拡がり"を示すパラメータを表す）を有し、これが先駆的情報を表す。ただし、調整可能な荷重パラメータも有しており、それはS C - H M Mモデル化エンジン26で調整される。順序化された特徴ベクトル系列22を以下では、長さTの系列Xとして表し、その成分の特徴ベクトルを x_i , $i = 1, \dots, T$ と表す。すなわち、系列は次のように表される。 $X = \{x_1, x_2, \dots, x_T\}$ S C - H M Mモデル化においては、それぞれの時刻tにおいてシステムは隠れた状態であると仮定され、この状態をここでは離散的潜在変数 q_t で表す。S C - H M M状態数32はモデル化のパラメータであり、ここではNと表す。こうして、1からNまでラベル付けされたN個の状態がある。例示的なS C - H M Mモデル化の実施形態の汎用基準30は、K個のガウス分布成分を持った汎用G M Mである。それぞれの状態iは3つのタイプのパラメータで記述される。すなわち、初期占有確率 $\pi_i = P(q_t = i)$ 、遷移確率 $a_{ij} = P(q_t = j | q_{t-1} = i)$ 、出力確率 $P(x_i | q_t = i)$ である。ここでの実施例では、左から右へのH M Mを仮定し、ジャンプはないものとする。そのようなH M Mは以下の性質をもつ： $a_{ij} = 0$ ($j \neq i$ 及び $j = i + 1$) 出力確率は例示の汎用G M M基準30で記述されるものと仮定する。この実施例においては、分散パラメータは対角共変行列であると仮定する。それは、計算コストが軽減され、かつ任意の分布は対角共変行列を有するガウス分布の結合により任意の精度で近似できるからである。

【0019】

S C - H M M状態数N32を順序化された特徴ベクトル22の系列長Tで計測することは有効である。ここで用いる表記では、 $N = \cdot T$ であり、 \cdot は(0, 1)の範囲内のスケール因子である。パラメータ \cdot はここでは「圧縮因子」とも呼ばれる。それはS C - H M Mが直観的には、順序化された特徴ベクトル22のT観測中の情報を $N = \cdot T$ 状態に圧縮するからである。

【0020】

S C - H M Mモデル化において、出力確率のすべてのガウス分布成分は、ここで汎用基準30と称する共通セットのK個のガウス分布に属するように強制される。言い換えれば、K個のガウス分布の平均および分散パラメータは、S C - H M Mモデル化の間、調整されない。 p_k が、平均ベクトル μ_k 及び共変行列 Σ_k を有する汎用G M M30のk番目のガウス分布成分を表すとする。出力確率は以下のように表される。

【数1】

$$p(x_t | q_t = i) = \sum_{k=1}^K \omega_{ik} p_k(x_t) \quad (1)$$

従ってS C - H M Mパラメータは、平均 μ_k 及び分散パラメータ Σ_k を含む、系列に依存しないすなわち共通の、つまり固定すなわち静的なパラメータと、遷移確率 a_{ij} 及びガウス

10

20

30

40

50

分布成分重みパラメータ ω_{ik} を含む、系列に依存するパラメータと、に分離することができる。固定すなわち静的な平均 p_k 及び分散パラメータ σ_k^2 は、手書きスタイルに関する先駆的な情報を S C - H M M モデル化に注入する。

【0021】

S C - H M M モデル化エンジン 2 6 は、N = T 状態 3 2 を有する S C - H M M と、K 個のガウス分布成分を持つ汎用基準 3 0 とを利用して、順序化された特徴ベクトル 2 2 をモデル化する。適切なモデル化手法においては、G M M 基準 3 0 の静的な平均 μ_k および分散パラメータ σ_k^2 は再び固定され、遷移確率 a_{ij} 及び重みパラメータ ω_{ik} が適切に調整されて、次の再評価公式を用いる期待値最大化 (E M) 法を利用して、順序化された特徴ベクトル 2 2 をモデル化する。

10

【数 2】

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_{ij}}{\sum_{t=1}^{T-1} \gamma_i(t)} \quad (2)$$

【数 3】

$$\hat{\omega}_{ik} = \frac{\sum_{t=1}^T \gamma_{ik}(t)}{\sum_{t=1}^T \sum_{n=1}^N \gamma_{in}(t)} \quad (3)$$

20

ここで、 $\gamma_i(t)$ は x_i が状態 i によって生成される確率、 $\gamma_{ik}(t)$ は x_i が状態 i と混合成分 k によって生成される確率、そして $\gamma_{ij}(t)$ は x_i が状態 i によって生成され、 x_{i+1} が状態 j によって生成される確率、を示す。これらの後天的因子は、フォワード・バックワードアルゴリズムで適切に計算される。これについては、例えば、Rabiner による、"A tutorial on hidden Markov models and selected applications in speech recognition", Proc. of the IEEE vol. 77, p. 257-86 (1989) を参照されたい。

30

【0022】

S C - H M M モデル化の出力には、適合された遷移確率 a_{ij} 及び重みパラメータ ω_{ik} が含まれる。しかし、ほとんどの H M M モデル化の適用において、適合重みパラメータ ω_{ik} が遷移確率 a_{ij} に比べて実質的により識別力がある。従って例示の実施形態においては、S C - H M M モデル化エンジン 2 6 の出力は、順序化された重みパラメータ系列 3 6 、すなわち、適合重みパラメータの順序化された系列 ω_{ik} である。ほかのよく検討された実施形態においては、順序化された適合遷移確率系列 a_{ij} は、適合重みパラメータ ω_{ik} の使用に追加して、あるいはその代わりに含まれるように考慮されている。

【0023】

同様に、ワードスボッティングエンジン 1 0 が S C - H M M モデル化エンジン 2 6' を含み、N = T 状態 3 2 を有する S C - H M M と、K 個のガウス分布成分を持った汎用基準 3 0 とを利用して、1 つまたは複数の参照用の順序化された特徴ベクトル系列 2 4 をモデル化する。S C - H M M エンジン 2 6' により遂行されるモデル化は、S C - H M M エンジン 2 6 により遂行されるモデル化と同一である。従って、機能的には等価な別々の特徴抽出エンジン 2 6 、2 6' を利用することも考えられるが、実際に 2 つのエンジン 2 6 、2 6' が 1 つの同一エンジンであってもよい。1 つまたは複数の参照用画像 1 4 が、対象としている語や句の手書きサンプル画像の訓練用のセットなどのよう、参照画像を複数含む場合には、S C - H M M モデル化する際に調整パラメータの a_{ij} と ω_{ik} を 1 つの参照用サンプル画像に対して調整するのではなく、複数の訓練用のサンプル画像の参照用のセットのそれぞれに対して適切に調整の訓練を行う。いずれの場合でも、S C - H M M モ

40

50

デル化エンジン 26'の出力は、適合重みパラメータ α_{ik} の順序化された系列によって定義された、対象モデルの語あるいは句 38 である。遷移パラメータ a_{ij} は、所望により追加もしくは代替のモデル成分として含まれる。

【0024】

手書きのテキスト画像 12 および参照用の 1 つのテキスト画像もしくは複数の画像 14 に対する表示 36、38 が、GMM 基準 30 の先駆的情報を、共通の固定された、すなわち静的な平均 μ_k 及び分散パラメータ σ_k を介して組み込むことは有利なことである。更に、表示 36、38 は簡潔になっている。適合重みパラメータの順序化された系列 α_{ik} は、インデックス $i = 1, \dots, N = \cdot T$ および $k = 1, \dots, K$ の範囲を動くので、重みパラメータの総数は $N \times K$ である。ここで、 N は SC-HMM の状態数であり、 K は GMM 基準 30 のガウス分布成分の数である。例示された動的時間伸縮法 (DTW) の距離コンパレータ 40 などのような、順序化系列コンパレータは、表示 36、38 に適切に作用して、定量的比較指標 42 を生成する。図に示した DTW コンパレータの代わりに、実質的に任意の他のタイプのコンパレータを利用することができる。好ましくは、コンパレータが順序化系列 36 と 38 の間の可能な整合を考察する。そして、ベイズの誤差確率や、カルバック・ライブラー (KL) の発散、バタチャリヤの類似性、などの状態（即ち、重みベクトル）間の類似性の局所指標を求めるさまざまな方法を利用することができます。コンパレータは、最適経路あるいは全経路の合計を考察する。DTW 距離は最適な経路を考察する。それを以下で説明する。その他の距離あるいは他の比較アルゴリズムもまた検討する。

10

【0025】

DTW はベクトル系列間の伸縮性のある距離である。長さがそれぞれ T_x 、 T_y である 2 つのベクトル系列 X および Y について、DTW がこの系列間の整合のあらゆる可能性を考察する。ここで整合とは、ある条件を満足するようなベクトル間の一連の対応を言う。それぞれの整合に対して、ベクトルとベクトルの距離の総和が求められ、これらの距離の最小値として DTW 距離が定義される。これは言い換えれば、最適整合に沿った距離であり、ワーピング経路とも呼ばれる。可能なすべての整合を直接評価するのは、計算コストがかさむ。従って、DTW のほとんどの実装においては、2 次時間で距離を計算するために動的計画法アルゴリズムが用いられる。これは、プレフィックス $\{x_1, \dots, x_m\}$ と $\{y_1, \dots, y_n\}$ との間の部分距離 DTW(m, n) を考慮に入れる。ここで $m = 1, \dots, T_x$ 、 $n = 1, \dots, T_y$ である。部分距離 DTW(m, n) は次式で適切に与えられる。

20

【数 4】

$$DTW(m, n) = \min \left\{ \begin{array}{l} DTW(m-1, n) \\ DTW(m-1, n-1) \\ DTW(m, n-1) \end{array} \right\} + d(m, n) \quad (4)$$

ここで、 $d(m, n)$ は、例えばユークリッド空間距離として表されるような、 x_m と y_n との間のベクトルとベクトルとの距離である。DTW 距離をワーピング経路の長さで割ると、一般に性能の向上がみられる。距離関数 $d(_, _)$ は所望により類似の指標に置き換えることができる。ただし式(4)の $\min \{\dots\}$ を $\max \{\dots\}$ へ変える必要が生じる。式(4)を用いて、行列 DTW(m, n) を 1 列づつ埋めていくことができるので、アルゴリズムのコストは $O(T_x T_y D)$ のオーダである。ここで D は特徴ベクトルの次元である。

40

【0026】

DTW 距離を状態系列へ展開するには、ベクトルとベクトルの距離 $d(_, _)$ を状態と状態の距離に置き換えれば十分である。対象としている、系列に依存する状態パラメータは、順序化された混合重み系列 36、38 である。従って、2 つの状態間の距離は、2 つの混合重みベクトル間の距離として定義される。バタチャリヤ類似性指標では、これは

50

GMM間の真のバタチャリヤ類似性の近似に対応することが示される。以下においては、
【数5】

$$f = \sum_{i=1}^M \alpha_i f_i, \quad g = \sum_{j=1}^N \beta_j g_j$$

は、それぞれSC-HMMモデル化エンジン26、と26'による2つのGMM出力を表しており、 $\alpha = (\alpha_1 \dots \alpha_M)$ はSC-HMMモデル化エンジン26によるGMM出力の重みベクトル α_i であり、 $\beta = (\beta_1 \dots \beta_N)$ はSC-HMMモデル化エンジン26'によるGMM出力の重みベクトル β_j である。確率積の核(PPK)は以下のように適切に定義される。

【数6】

$$K_{ppk}^\rho(f, g) = \int_x (f(x)g(x))^\rho dx \quad (5)$$

これはJebbaraによる"Probability product kernels", Journal of Machine Learning Research vol. 5 pp. 819-44 (2004)を参照されたい。バタチャリヤ類似性は特別の場合であり、 $B(f, g) = K_{ppk}^{1/2}(f, g)$ である。 f および g がガウス混合モデルの場合には、バタチャリヤ類似性 B は閉じた形の式では表されないが、 B を次に示す上界で近似することによって近似できる。

【数7】

$$B(f, g) \leq \sum_{i=1}^M \sum_{j=1}^N (\alpha_i \beta_j)^{1/2} B(f_i, g_j) \quad (6)$$

値 $B(f_i, g_j)$ が、ガウス分布成分のペア間のバタチャリヤ類似性に対応し、これには閉じた形の式が存在する。例えば、Jebbaraによる"Probability product kernels", Journal of Machine Learning Research vol. 5 pp. 819-44 (2004)を参照されたい。SC-HMMの場合には、GMM出力確率は、同じガウス分布成分の組、すなわち、 $M = N$, $f_i = g_j$ に対して定義され、値 $B(f_i, g_j)$ が事前に計算される。そのような場合には、2つの状態の類似性は、対応する2つの重みベクトル間の類似性そのものとなる。

【0027】

しかしながら、計算コストはガウス分布成分の数の2次式のままであり、典型的な応用例では2~3から数百以上のオーダになるものもある。計算コストを下げることが望まれる場合には、所望により、次のような界への追加の近似が行われる。ここで、ガウス分布成分は明確に分離していることが仮定される。すなわち、 $i \neq j$ の場合には $B(f_i, g_j) = 0$ であることが仮定される。この近似は一般に特徴空間の次元が増えるにつれて妥当性を増す。定義により、 $B(f_i, f_i) = 1$ である。これにより以下の近似が成り立つ。

【数8】

$$B(f, g) \approx B(\alpha, \beta) = \sum_{i=1}^M (\alpha_i, \beta_i)^{1/2} \quad (7)$$

これは、重みベクトル α と β の間の離散バタチャリヤ距離である。重みベクトルの平方根が格納される場合には、この量は内積として計算するのに効率的である。

【0028】

図2を参照すると、汎用GMMモデル30の生成の適切な手法が示されている。GMMモデル30は、手書きスタイルに関する先駆的情報を具現化することが望ましい。この目的のために、手書きスタイルデータベース50が提供され、そこには様々な手書きのサンプル画像が含まれている。手書きスタイルデータベース50は、数及び範囲において十分

10

20

30

40

50

な手書きサンプルを含んでおり、ワードスポットエンジン 10 が出会うと予想される範囲の手書きスタイルを表示できる。十分な多様性を与えるために、対象とする地域に住んでいる人、および / または出会う書き手の教育レベルとして想定される全範囲の人、などからサンプルが集められる。その反対の極端な場合としては、単一の人のみによって書かれた文書にワードスポットティングを適用して処理することが考えられる（例えば、沢山の手紙を書いた歴史上の人物の広範囲にわたる著作を分析するような場合）。このようなものへの適用には、手書きスタイルデータベースは適切には、対象とするたった一人の書き手による手書きサンプル画像を含む。手書きスタイルデータベース 50 は一般的に対象とする語や句に限られるわけではない。それよりもむしろ幅広い手書きサンプルを提供する、広範な語や句を含み、ワードスポットティングエンジン 10 が出会う可能性のある幅広い手書きスタイルの形を提供できるようにする。

【 0 0 2 9 】

手書きスタイルデータベース 50 のサンプル画像は、特徴抽出エンジン 16" で処理される。これは機能的には特徴抽出エンジン 16 と同等のもので（実際、所望により 1 つの同一エンジンとして具現化されてもよい）、手書きのサンプル画像のそれぞれに対して特徴ベクトル系列を生成する。手書きスタイルモデル化エンジン 52 は適切には、汎用基準 30 として働く汎用 GMM を、特徴抽出エンジン 16" によって抽出された特徴ベクトル系列を利用して訓練する。この訓練では、サンプル画像から抽出された特徴ベクトルの系列順序は無視されて、特徴ベクトル系列はその代わりに（一般的に大きな）特徴ベクトルの無秩序な組と見なされる。特徴ベクトル抽出時に、テキスト画像に沿ってスライド窓が動くと、スライド窓が文字や、文字の一部、あるいは文字と文字とをつなぐ部分などのような典型的な手書き一次情報をサンプリングする。重みパラメータ κ 、平均パラメータ μ_k 、分散パラメータ κ を含むすべての GMM パラメータを無秩序な特徴ベクトルの組に合わせることにより、ガウス分布成分の "語彙" が生成される。汎用 GMM 30 を訓練する適切なアルゴリズムは、期待値最大化 (EM) アルゴリズムである。これは例えば、Billmes の "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models" , Technical Report TR-97-021 , Int. Computer Science Institute, (1998) に記述されている。訓練の後、調整された平均パラメータ μ_k 及び分散パラメータ κ は、後続の SC - HMM モデル化エンジン 26、26' によるモデル化において、先駆的情報を担う固定された静的パラメータとして扱われる。

【 0 0 3 0 】

前に述べた手書きテキストオブジェクト 12 は单一のオブジェクトであるが、対象とする語や句の 1 つまたは複数の手書きサンプル画像は、单一のオブジェクトであってもよいし、複数のオブジェクトであってもよい。また、オブジェクト 12 は複数のオブジェクトから構成されていてもよく、その場合には図 1 のシステムは、2 組の手書きテキスト画像などの、2 組のオブジェクトを比較するのに利用することが可能である。

【 0 0 3 1 】

図 3 を参照すると、変形型ワードスポットティングエンジン 10' が示されている。このワードスポットティングエンジンは、対象モデルの語あるいは句 38 を生成するための部品 16'、26' がワードスポットティングエンジン 10' から省略され、その代わり、対象モデルの語あるいは句 38 はワードスポットティングエンジン 10' への入力となっていることを除けば、図 1 のワードスポットティングエンジン 10 と同じである。対象語あるいは句の 1 つまたは複数の手書きサンプル画像 14 から、対象とするモデルの語あるいは句 38 を生成する処理が、省略した部品 16'、26' で構成される適切なモデル準備システム（表示せず）によって先ず実行され、その結果として得られる対象モデルの語あるいは句 38 が、後からワードスポットティングエンジン 10' によって検索、利用されることに備えて格納される、ということがこの変形型ワードスポットティングエンジン 10' で示されて

10

20

30

30

40

50

いる。この手法によれば、定量的比較をするために対象の語あるいは句 1 2 の新しい入力がある度に、対象モデルの語あるいは句 3 8 を再計算する必要がないので、適用対象によつては処理効率が向上する。

【 0 0 3 2 】

図 4 を参照すると、文書のラベル付けおよびアーカイブシステムが示されている。このシステムは、変形型ワードス포ッティングエンジン 1 0 'と、アーカイブ中の文書をラベル付けするための対象語あるいは句のモデルを含んだ対象モデル語あるいは句のデータベース 6 0 を利用する。文書をトピック別にラベル付けしてアーカイブする場合、例えば地球温暖化というトピックに関する文書を識別するための語句としては、例として「オゾン」、「地球温暖化」、「極氷冠」、「温室」等が含まれる。このような対象とするそれぞれの語あるいは句に対して、部品 1 6 '、2 6 '（図 1 参照）が適用されて、対応する対象モデルの語あるいは句が生成され、データベース 6 0 に格納される。10

【 0 0 3 3 】

一般的な処理の実例においては、文書は紙に書かれた形で受取られ、光学式のスキャナ 6 2 を用いて文書ページの画像が生成される。光学式スキャナ 6 2 でスキャンされた文書は通常手書きの文書である。しかし、本開示のワードス포ッティング技術は、タイプされた語、あるいはコンピュータで生成されたフォントで印刷された語などにもス포ッティングを適用できるように考えられている。ここで用いているように、「手書き文書」という用語あるいは類似の術語は、対象とする、あるいは対象となる可能性のある手書きコンテンツを有する文書を意味する。従って、手書き文書は、大量の、あるいは殆どが、コンピュータで生成した、もしくはタイプで書かれたテキストを含んでもよい。例えば、コンピュータで生成した書類に手書きで記入したものは、ほとんどがコンピュータ型のフォントで書かれたコンピュータ生成テキストで構成される。しかしながら、そのような書類も手書きで記入されると手書き文書と見なされる。それは対象とするコンテンツは手書きの記入部分であるからである。20

【 0 0 3 4 】

文書ページの画像は文書ページ画像分割プロセッサ 6 4 によって処理され、手書きの語あるいは句を表すテキスト画像（あるいはむしろ文書ページ画像の部分画像）を分離ないしは識別する。所望によっては、分割プロセッサ 6 4 は抽出されたテキスト画像の、ゆがみ、傾斜、寸法についての標準化を行い、これらに関する一様性も向上させる。所望によつては、プレフィルタ 6 6 が選択基準に基づいて事前のフィルタリングを行い、ある条件に適合しないテキスト画像を検討から除外する。例えば、プレフィルタ 6 6 は所望によつては、語あるいは句の標準化された寸法が、対象の語あるいは句の最短のものよりも小さいテキスト画像を除外する。そのようなプレフィルタをかけることによって、例えば 1 文字あるいは 2 文字の語をすべて除外することができる。プレフィルタでは所望によって更に、テキスト画像を含む四角形の縦、横のアスペクト比などの基準を検討することもできる。そしてこのような基準を用いて、アスペクト比が対象の語あるいは句のいずれのアスペクト比にも合致しないテキスト画像を除去することができる。30

【 0 0 3 5 】

プレフィルタ 6 6 を通過する分割されたテキスト画像は、ルーピング処理 6 8 により順番に処理される候補テキスト画像と見なされ、ワードス포ッタ 1 0 'に入力された手書きテキスト画像 1 2 として、候補テキスト画像のそれぞれが次々に処理される。更に、対象モデル語あるいは句のデータベース 6 0 に格納されたモデルはルーピング処理 7 0 によって順次処理され、ワードス포ッタ 1 0 'に入力された対象モデルの語あるいは句 1 4 として、それぞれのモデルが順番に処理される。このように、候補テキスト画像のそれぞれは、対象モデルの語あるいは句のそれぞれと比較され、候補テキスト画像が対象の語あるいは句のいずれかとマッチするかどうかを評価するための定量的比較指標の組が生成される。文書ラベラ 7 4 がこの情報を用いて文書をラベル化する。例えば、地球温暖化というトピックに関する対象の語あるいは句が、その文書に閾値の数よりも多く含まれる場合に、文書ラベラ 7 4 はその文書を「地球温暖化」という主題タグでラベル付けする。光学的に4050

スキャンされた文書ページは適切に連結されて p d f ファイル、もしくはその文書を格納するその他の連結ファイルとされ、次にそれが、文書ラベラ 7 4 によりその文書に付けられた任意のラベルを含むメタデータと共に、文書データベース 7 6 に格納される。更にそれに追加して、もしくはその代わりに、文書ラベル化の別の使い方をすることができる。例えば代替実施形態の 1 つとして、ラベラ 7 4 によって、対象の 1 つまたは複数のトピックに関係あることが表示された文書のみが文書データベース 7 6 に格納される。別の代替実施形態では、ラベラ 7 4 が対象の 1 つまたは複数のトピックに関係あるとする文書が人によるレビュー（例えばEメールを介して）に送られ、人が更に検討しレビューする。

【 0 0 3 6 】

既に述べたように例示的な S C - H M M モデル化エンジン 2 6 は、 S C - H M M の状態数 N 3 2 を利用して S C - H M M のモデル化を実行し、順序化された特徴ベクトル系列 2 2 は系列の長さ T を有し、そして、 $N = \cdot T$ の表記を導入した。ここで \cdot は (0 , 1) の範囲内のスケール因子であり、直観的には S C - H M M が順序化された特徴ベクトル系列 2 2 の T 個の観測に含まれる情報を $N = \cdot T$ の状態に圧縮するので、この因子をここでは"圧縮因子"とも呼ぶ。一般的に、 $\cdot = 1$ の場合には $N = T$ であり、これは非圧縮の S C - H M M モデル化プロセスと考えられる。また一方で、 $\cdot < 1$ の場合は状態数 N が系列長 T よりも小さくなり、 S C - H M M モデル化プロセスは圧縮を実行すると考えられ、計算スピードが向上することの代償として検索精度（ワードスポットティング応用の）が幾分減少することになる。次にこの圧縮因子 \cdot の影響を更に検討する。

【 0 0 3 7 】

D T W 指標のコストは、 $O(T_x T_y D)$ のオーダであり、ここで D は特徴ベクトルの次元である。図 1 のシステムあるいは図 3 のシステムにより実装される類似性指標のコストは、 $0 \sim 1$ に対して $O(\cdot^2 T_x T_y K)$ のオーダであることが示される。ここで、 K は汎用 G M M 基準 3 0 のガウス分布成分の数であることを思い起こされたい。これより K = D の場合は、 D T W 距離指標の場合に比べて計算コストが $1 / \cdot^2$ だけ縮減されることになる。

【 0 0 3 8 】

図 1 のシステムに実質的に一致する、定量的比較指標を用いたワードスポットティングシステムが発明者によって構築された。そしてその性能を D T W 距離指標そのものを用いるワードスポットティングと比較した。実際に構築されたワードスポットティングシステムを用いて行われたテストで、圧縮因子 \cdot を減少させれば一般的に性能が下がり、処理速度が上がる事が確認された。従って、ワードスポットティングシステム（あるいは、開示した定量的順序化ベクトル系列比較技術のその他のアプリケーション）の性能は、圧縮因子 \cdot を加減することにより調節が可能である。一般的に、圧縮因子 \cdot を高くすれば精度が向上し、逆に圧縮因子 \cdot を下げればスピードが向上する。ワードスポットティングテストの結果、圧縮因子 \cdot が 0 . 1 ~ 0 . 2 の範囲（これは計算コストの増加 $1 / \cdot^2$ が D T W 指標を単独で使用する場合の 2 5 ~ 1 0 0 倍の範囲に対応）にあれば、 D T W 指標単独操作によるワードスポットティングに比べて性能が同等、もしくは若干すぐれていることが示された。

【 0 0 3 9 】

これらの計算コストの値は、順序化系列コンパレータ 4 0 に対するものであって、 S C - H M M モデル化エンジン 2 6 、 2 6 ' により実行される処理の計算時間は含まれていない。しかし図 3 に示すように、一般的なワードスポットティングのやり方では、対象モデルの語あるいは句 3 8 は適切に事前計算される。従って、 S C - H M M モデル化エンジン 2 6 ' の計算コストはその影響を受けない。さらに、それぞれの手書きテキスト画像 1 2 が複数の対象モデルの語あるいは句 3 8 と比較されるとすると（例えば図 4 に示すように）、対応する重みパラメータの順序化系列 3 6 を生成するための S C - H M M モデル化エンジン 2 6 の処理は、手書きテキスト画像 1 2 のそれぞれに対して 1 回のみでよく、その後は順序化系列コンパレータ 4 0 によって適切に再使用されて対象モデルの語あるいは句 3 8 のそれぞれとの比較が行われる。対象モデルの語あるいは句の数が大量である場合には、これはワードスポットティング応用の場合にはしばしば起きることであるが、順序化系列表

10

20

30

40

50

コンパレータ 40 の計算コストが大部分を占めることになる。

【0040】

図 1、3 のワードスポットエンジン 10、10' は、図 2 の手書きスタイルモデル化エンジン 52 および図 4 の文書ラベル付けシステムと共に、さまざまな方法での実現が可能である。図示した実施形態においては、コンピュータ C が、これらのシステムの操作を実行するように、適切にプログラムされたデジタルプロセッサを含んでいる。また、インターネットベースのサービスシステム、パーソナルデータアシスタント (PDA) あるいは自動書式受信システム等の、その他の適切にプログラムされたデジタルプロセッサ、あるいはそれを含むコントローラあるいはデバイスも利用することが可能である。これに加えて、あるいはこれに代わって、記憶媒体がそのようなプロセッサやコントローラで実行できる命令を適切に格納して、本開示の図 1 および 3 のワードスポットティングエンジン 10、10' や、図 2 の手書きスタイルモデル化エンジン 52、図 4 の文書ラベル付けシステムの 1 つまたは複数を実行する。記憶媒体としては例えば、スタンドアローン部品の、ないしはコンピュータやその他のデジタルプロセッシングデバイスに組み込まれたハードディスクやその他の磁気記憶媒体、または、スタンドアローン部品の、ないしはコンピュータやその他のデジタルプロセッシングデバイスに組み込まれた光ディスクやその他の光記憶媒体、または、スタンドアローン部品ないしはコンピュータやその他のデジタルプロセッシングデバイスに組み込まれたフラッシュメモリやその他の静電気型記憶媒体、などがある。
。

【0041】

汎用モデル 30 を介して手書きに関する先駆的な情報を使用する、本開示の手法の別の利点は、コンピュータで生成したフォントを利用して生成した参照用のサンプルテキスト画像を利用して、手書きテキスト画像にワードスポットティングを実行することが可能なことである。すなわち、これらの実施形態では手書きテキスト画像 12 は手書きのままであるが、対象の語あるいは句の手書きサンプルの 1 つまたは複数の画像 14 は、対象の語あるいは句をタイプした、あるいはコンピュータで生成したサンプルの 1 つまたは複数の画像で置き換えることができる。タイプされたテキストサンプルの利点は、タイポグラフィックフォントで表示することによって任意のクエリーストリングに対するクエリーをオンラインで自動生成できることである。しかし、この手法はこれまで実際的ではなかった。それはタイプされたテキスト形状が概して手書きのものを表現できなかつたからであり、タイプされたテキストの実質的な一様性が、手書き画像の実質的な可変性をうまく表現できなかつたからである。
。

【0042】

しかし、図 1 あるいは図 3 の類似性指標を利用すると、汎用基準 30 に通常含まれていて、例えば図 2 に示す手書きスタイルデータベース 50 をを利用して訓練される、手書きに関する先駆的な情報によりこの不足を解消できる。従って、S C - H M M モデル化エンジン 26' が、タイプされたテキストあるいはコンピュータで生成されたフォントで印刷されたテキストとして表示された、対象の語あるいは句のサンプル画像を処理する場合、汎用 G M M 基準 30 の固定の、すなわち静的な平均パラメータ μ_k 及び分散パラメータ σ_k が、対象の語あるいは句のタイプされた、あるいはコンピュータで生成されたサンプルの S C - H M M モデルの中に手書きスタイルに関する先駆的な情報を入れ込む。
。

【0043】

図 1 に示された手書きサンプル 14 の代わりにタイプされたあるいはコンピュータで生成されたテキストサンプルを使用する図 1 のシステムの能力を実証するために、図 1 の定量比較指標を使って実際に構築されたワードスポットティングシステムを、入力 14 としてタイプされたあるいはコンピュータで生成されたテキストサンプルを用いて使用してみた。単一のコンピュータ生成フォントを用いて印刷されたサンプルを使用した場合、性能は良くないことがわかった。しかし、典型的な手書きとの類似性を有するさまざまなコンピュータ生成フォントを使って印刷されたサンプルを用いて入力 14 を構成すると、大幅に改良された性能が得られた。
。

10

20

30

40

50

【0044】

図5を参照すると、例えば、対象の語“Madame”が、手書きを模したデザインの10種のコンピュータ生成フォントで印刷されている。図5の10種のフォントは、(上から下へ、また左から右へ) French Script、Kunstler Script、Papyrus、Lucida handwriting、Rage Italic、Lucida Calligraphy、Harlow Solid、Freestyle Script、Comic Sans、Viner Handである。また、性能は特徴ベクトルの選択にも敏感であることがわかった。最もよい結果は局所勾配ヒストグラム(LGH)の特徴を利用して得られた。これについては、Rodriguezほかの“Local gradient histogram features for word spotting in unconstrained handwritten documents”，in 2008 International Conference on Frontiers in Handwriting Recognition (ICFHR 2008)を参照されたい。¹⁰ 実験は、LGH特徴と、図5に示した対象の語“Madame”用の訓練セットとを用いて行われ、(幅／アスペクト比などに基づく)刈り込みは行わず、図1の定量的比較指標を用いて行った。またDTW指標のみを用いても行った。図1の定量的比較指標を用いた性能は、DTW指標のみを用いた場合に比べてはるかに優れていた。ある性能指標によれば2倍の改善であった。

【0045】

いかなる特定の操作理論の制限を受けることなく、図1の定量的比較指標を用いた性能の大幅な改良は、汎用GMM基準30によって対象の語に入れ込まれ、多様な手書きサンプルを用いて訓練された、典型的な手書きの可変性に関する先駆的情報によるものと考えられる。²⁰

【0046】

例示の実施形態は、ここで例示の実施例として用いられたワードスポットティングに関するものである。しかし、開示した定量的比較指標は、一般的に、ベクトルの順序化系列により表すことのできるほとんど任意のタイプの対象に対する定量的比較に適用できる。対象は例えば、テキスト画像(例示の実施形態のような)とか、記録された音声や生体的序列、等であってもよい。³⁰

【0047】

例えば、対象のタイプが音声のコンテンツである場合、特徴抽出エンジン16は、適切には、例えば音声コンテンツに沿って時間的にスライドする窓を適用し、それぞれの時間窓内の音声コンテンツを表す特徴ベクトルを抽出することにより、順序化ベクトル系列を生成するように構成される。この音声認識応用においては、1つまたは複数の対象の語あるいは句のサンプル14は、適切には、1人または複数の人によって話された対象の語あるいは句を録音した1つまたは複数のサンプルである。汎用基準30は、手書きスタイルデータベース50を、代表的な人数の人によって発声された広範な録音音声を含む音声スタイルデータベースで置き換えることにより適切に生成される。

【0048】

別の適用例は画像コンテンツのスポットティングである。ここで特徴抽出エンジン16は、例えば、グリッドやその他の選択パターンあるいは選択アルゴリズムに従ってキーとなるパッチ(つまり画像の小区分)を選択し、それぞれのキーパッチから特徴ベクトルを定義する画像特徴を抽出することによって、順序化ベクトル系列を生成するように適切に構成されている。この画像コンテンツスポットティング応用においては、1つまたは複数の対象の語あるいは句のサンプル14は、適切には、対象の主題の1つまたは複数の画像である(例えば、対象の主題が“動物”であれば、サンプル画像は、適切には、猫、犬、ウサギ、等の画像である)。汎用基準30は、適切には、手書きスタイルデータベース50を、画像コンテンツスポットティングシステムによる処理が想定される広範な画像コンテンツを表す画像が含まれた、画像データベースで置き換えることにより生成される。⁴⁰

10

20

30

40

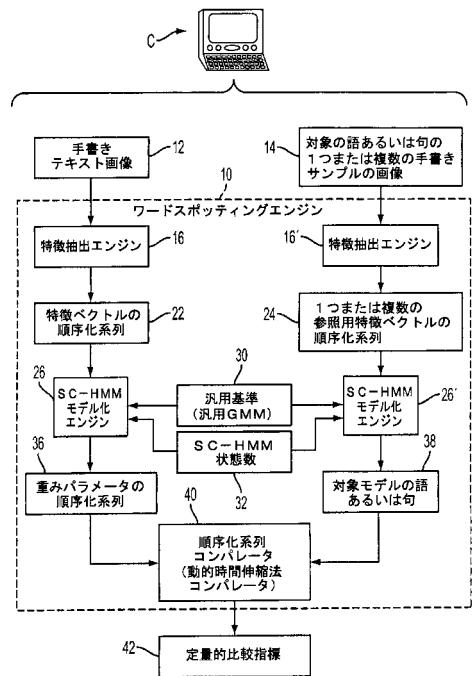
50

【符号の説明】

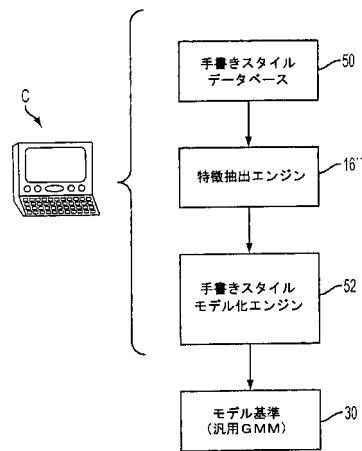
【0049】

10	ワードスパッティングエンジン	
10'	ワードスパッティングエンジン	
12	手書きテキスト画像	
14	対象の語あるいは句の1つまたは複数の手書きサンプルの画像	
16	特徴抽出エンジン	10
16'	特徴抽出エンジン	
16"	特徴抽出エンジン	
22	特徴ベクトルの順序化系列	
24	1つまたは複数の参照用特徴ベクトルの順序化系列	
26	SC - HMM モデル化エンジン	
26'	SC - HMM モデル化エンジン	
30	汎用基準(汎用GMM)	
32	SC - HMM 状態数	
36	重みパラメータの順序化系列	
38	対象モデルの語あるいは句	
40	順序化系列コンパレータ(動的時間伸縮法コンパレータ)	
42	定量的比較指標	
50	手書きスタイルデータベース	20
52	手書きスタイルモデル化エンジン	
60	対象モデルの語あるいは句のデータベース	
62	光学スキャナ	
64	文書ページ画像分割プロセッサ	
66	フィルタ基準による画像除去用プレフィルタ	
68	語あるいは句の候補画像ルーピング処理	
74	文書ラベラ	
76	文書データベース	
70	対象ルーピング処理の語あるいは句	
C	コンピュータ	30

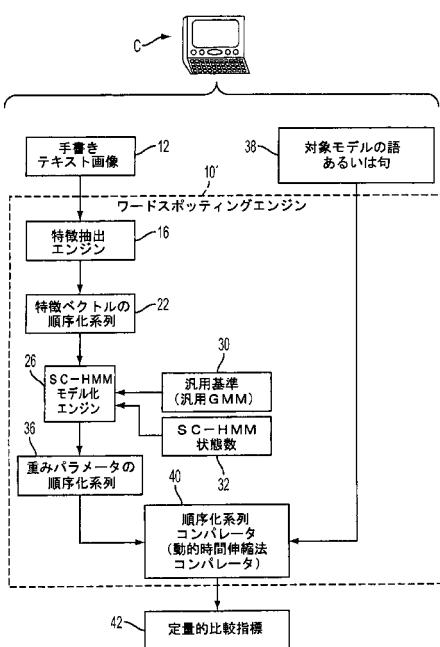
【図1】



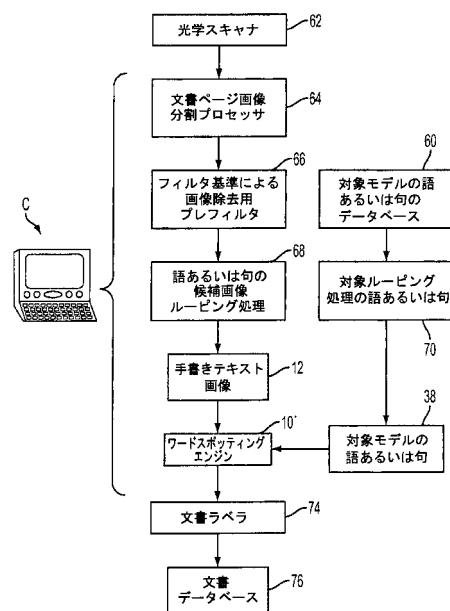
【図2】



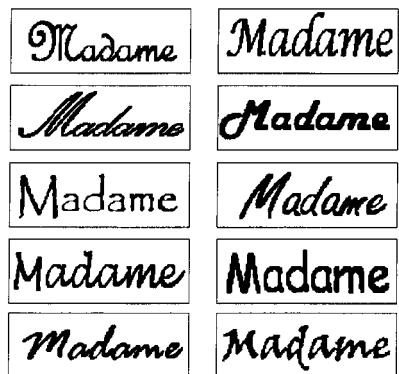
【図3】



【図4】



【図5】



フロントページの続き

(72)発明者 フローラン ペロンナン
フランス 38420 ドメーヌ リュ ジャン ジョール 107

審査官 新井 則和

(56)参考文献 特開2004-294916(JP,A)
Jebara et al., "Spectral clustering and embedding with hidden Markov models", the 2007 European Conference on Machine Learning (ECCML 2007), 2007年

(58)調査した分野(Int.Cl., DB名)

G06K 9/00 - 9/82
G06T 1/00 - 7/60