

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第5355625号  
(P5355625)

(45) 発行日 平成25年11月27日(2013.11.27)

(24) 登録日 平成25年9月6日(2013.9.6)

(51) Int.Cl.

G06K 9/20 (2006.01)

F 1

G06K 9/20 340 J

請求項の数 24 (全 22 頁)

(21) 出願番号	特願2011-129862 (P2011-129862)
(22) 出願日	平成23年6月10日(2011.6.10)
(65) 公開番号	特開2012-3756 (P2012-3756A)
(43) 公開日	平成24年1月5日(2012.1.5)
審査請求日	平成25年5月30日(2013.5.30)
(31) 優先権主張番号	12/814,448
(32) 優先日	平成22年6月12日(2010.6.12)
(33) 優先権主張国	米国(US)

早期審査対象出願

(73) 特許権者	511072895 キング・アブドゥラジズ・シティ・フォー・サイエンス・アンド・テクノロジー(ケイ・エイ・シィ・エス・ティ) KING ABDULAZIZ CITY FOR SCIENCE AND TECHNOLOGY (KACST) サウジアラビア、11442 リヤド、ピイ・オウ・ボックス・6086、ザ・ナショナル・センター・フォー・テクノロジー・ディベロップメント
(74) 代理人	110001195 特許業務法人深見特許事務所

最終頁に続く

(54) 【発明の名称】光学式文字認識用に画像を前処理するための方法およびシステム

## (57) 【特許請求の範囲】

## 【請求項 1】

光学式文字認識(OCR:Optical Character Recognition)用に画像を前処理する方法であつて、画像は複数の欄を含み、複数の欄のうちの各欄は、アラビア文字および非文字項目の少なくとも一方を含み、前記方法は、

複数の構成要素の中のアラビア文字および非文字項目の少なくとも一方に関連付けられる複数の構成要素を定めることを備え、構成要素は一連の接続されたピクセルを含み、前記方法は、さらに、

前記複数の構成要素に関連付けられる行高さおよび欄間を計算することと、

行高さおよび欄間の少なくとも一方に基づき、前記複数の構成要素のうちの少なくとも 1 つの構成要素を、前記複数の欄のうちのある欄に関連付けることと、

前記複数の欄のうちの各欄について第 1 の組の特性パラメータを計算することと、

前記第 1 の組の特性パラメータに基づき、前記複数の欄のうちの各欄の中の複数の構成要素を合成して、少なくとも 1 つのアラビア副単語および少なくとも 1 つのアラビア単語の少なくとも一方を形成することとを備え、

前記第 1 の組の特性パラメータは、各欄に関連付けられる行高さ、各欄に関連付けられる語間、各欄に関連付けられる行間、各構成要素に対応するピクセルの数、各構成要素の幅、各構成要素の高さ、各構成要素の座標、各構成要素の密度、および各構成要素のアスペクト比の少なくとも 1 つであり、

前記各欄に関連付けられる行間を計算することは、

10

20

各欄に関連付けられる複数の構成要素の複数の横並び突出部のヒストグラムを作成することを含み、前記複数の横並び突出部のうちのある横並び突出部は、ラスタスキャンの各掃引に対応して複数の構成要素に関連付けられるピクセルの数を示し、さらに

2つの連続する最大横並び突出部間の平均距離を計算することと、

前記平均距離に基づき行間を計算することとを含む、方法。

#### 【請求項 2】

前記画像は、濃淡画像およびカラー画像の少なくとも一方を2進画像に変換することによって得られる、請求項1に記載の方法。

#### 【請求項 3】

前記画像は、ごま塩雜音をふるい落とすことによって得られる、請求項1に記載の方法 10。

#### 【請求項 4】

前記画像は、変形ハフ変換を用いて歪みを修正することによって得られ、前記変形ハフ変換はアラビア文字に適合される、請求項1に記載の方法。

#### 【請求項 5】

前記複数の構成要素を定めることは、

前記画像に対してラスタスキャンを行なうこと、

前記ラスタスキャンの少なくとも1回の掃引に対応する複数の構成要素のうちの少なくとも1つに関連付けられる複数のピクセルを特定すること、および

複数のピクセル間の相互接続に基づき前記複数のピクセルを統合して、少なくとも1組の接続されたピクセルを形成することを含む、請求項1に記載の方法。 20

#### 【請求項 6】

ピクセルは当該ピクセルの8個の隣接するピクセルの少なくとも1つと相互接続される、請求項5に記載の方法。

#### 【請求項 7】

前記行高さを計算することは、

前記複数の構成要素の各自の高さに対応する高さのヒストグラムを作成すること、

高さのヒストグラムから頻出高さを特定すること、および

頻出高さに基づき行高さを計算することを含む、請求項1に記載の方法。 30

#### 【請求項 8】

前記欄間は、行高さに基づき計算される、請求項7に記載の方法。

#### 【請求項 9】

各欄に関連付けられる語間を計算することは、

各欄に関連付けられた複数の構成要素の連続する構成要素間の間隔のヒストグラムを作成することと、

ヒストグラムから頻出間隔を特定することとを含み、前記頻出間隔は行高さによって定められるしきい値範囲内にあり、さらに

前記頻出間隔に基づき語間を計算することを含む、請求項1に記載の方法。

#### 【請求項 10】

前記連続する構成要素は、縦方向に重なる構成要素および所定の距離離れている構成要素の少なくとも一方を含み、前記縦方向に重なる構成要素は、縦軸に沿って少なくとも1つの座標を共有する、請求項9に記載の方法。 40

#### 【請求項 11】

前記複数の構成要素を合成することは、

前記少なくとも1つの副単語および少なくとも1つ単語の少なくとも一方を形成するために、語間に基づき、各欄に関連付けられる連続する構成要素を結合することと、

前記第1の組の特性パラメータに基づき、アラビア文字に関連付けられる複数の構成要素から、非文字項目に関連付けられる複数の構成要素のうちの少なくとも1つの構成要素をふるい落とすこととを含む、請求項9に記載の方法。

#### 【請求項 12】

ある欄に関連付けられる少なくとも 1 つの座標に基づき、複数の欄を分類することをさらに備え、前記少なくとも 1 つの座標は、画像における欄の位置に関連付けられる、請求項 11 に記載の方法。

#### 【請求項 13】

前記方法はさらに、各欄に関連付けられる、各副単語および各単語の少なくとも一方に関連付けられる第 2 の組の特性パラメータを計算することを備え、前記第 2 の組の特性パラメータは、各副単語および各単語の少なくとも一方に関連付けられる行高さ、各副単語および各単語の少なくとも一方に関連付けられる語間、ならびに各副単語および各単語の少なくとも一方に関連付けられる行間の 1 つであり、さらに

第 2 の組の特性パラメータに基づき少なくとも 2 つの副単語をグループ化して、少なくとも 1 つの副単語および少なくとも 1 つの単語の一方を形成することを備える、請求項 1 に記載の方法。 10

#### 【請求項 14】

前記方法は、前記少なくとも 1 つの副単語および前記少なくとも 1 つの単語を、各副単語および各単語の少なくとも一方に関連付けられる行高さ、ならびに各副単語および各単語の少なくとも一方に関連付けられる行間の少なくとも一方に基づき、少なくとも 1 本の横行に分割することをさらに備える、請求項 13 に記載の方法。

#### 【請求項 15】

光学式文字認識（O C R）用に画像を前処理するためのシステムであって、画像は複数の欄を含み、複数の欄の各欄は、アラビア文字および非文字項目の少なくとも一方を含み、前記システムは、

メモリと、

前記メモリに結合されるプロセッサとを備え、前記プロセッサは、

複数の欄の中のアラビア文字および非文字項目の少なくとも一方に関連付けられる複数の構成要素を定め、構成要素は一連の接続されたピクセルを含み、前記プロセッサは、さらに、

前記複数の構成要素に関連付けられる行高さおよび欄間を計算し、

行高さおよび欄間に基づき、複数の構成要素のうちの少なくとも 1 つの構成要素を、複数の欄のうちのある欄に関連付け、

前記複数の欄のうちの各欄について第 1 の組の特性パラメータを計算し、

第 1 の組の特性パラメータに基づき、複数の欄のうちの各欄の中の複数の構成要素を合成して、少なくとも 1 つのアラビア副単語および少なくとも 1 つのアラビア単語の少なくとも一方を形成することを行ない、

前記第 1 の組の特性パラメータは、各欄に関連付けられる行高さ、各欄に関連付けられる語間、各欄に関連付けられる行間、各構成要素に対応するピクセルの数、各構成要素の幅、各構成要素の高さ、各構成要素の座標、各構成要素の密度、および各構成要素のアスペクト比の少なくとも 1 つであり、

各欄に関連付けられる行間を計算するために、前記プロセッサは、

各欄に関連付けられる前記複数の構成要素の中の複数の横並び突出部のヒストグラムを作成し、前記複数の横並び突出部のうちのある横並び突出部は、ラスタスキャンの各掃引に対応して、前記複数の構成要素に関連付けられるピクセルの数を示し、さらに、前記プロセッサは、

2 つの連続する最大横並び突出部間の平均距離を計算し、

前記平均距離に基づき行間を計算する、システム。

#### 【請求項 16】

前記プロセッサは、

濃淡画像およびカラー画像の少なくとも一方を 2 進画像に変換すること、

ごま塩雜音をふるい落とすこと、および

変形ハフ変換を用いて歪みを修正すること、のうちの少なくとも一つを行なう、請求項 15 に記載のシステム。 50

**【請求項 17】**

複数の構成要素を定めるために、前記プロセッサは、  
 画像に対してラスタスキャンを行ない、  
 ラスタスキャンの少なくとも 1 回の掃引に対応して前記複数の構成要素の少なくとも 1  
 つの構成要素に関連付けられる複数のピクセルを特定し、  
 複数のピクセル間の相互接続に基づき、前記複数のピクセルを統合して少なくとも 1 組  
 の接続されたピクセルを形成する、請求項 15 に記載のシステム。

**【請求項 18】**

前記行高さを計算するために、前記プロセッサは、  
 前記複数の構成要素の各々の高さに対応する高さのヒストグラムを作成し、  
 前記高さのヒストグラムから頻出高さを特定し、  
 前記頻出高さに基づき行高さを計算する、請求項 15 に記載のシステム。 10

**【請求項 19】**

前記プロセッサは、行高さに基づき欄間を計算する、請求項 18 に記載のシステム

**【請求項 20】**

各欄に関連付けられる語間を計算するために、前記プロセッサは、  
 各欄に関連付けられる前記複数の構成要素のうちの連続する構成要素間の間隔のヒスト  
 グラムを生成し、  
 前記ヒストグラムから頻出間隔を特定し、前記頻出間隔は行高さによって定められるし  
 きい値範囲内にあり、さらに、前記プロセッサは、 20

前記頻出間隔に基づき語間を計算する、請求項 15 に記載のシステム。

**【請求項 21】**

前記プロセッサは、  
 語間に基づき各欄に関連付けられる連続する構成要素を結合して、少なくとも副単語お  
 よび少なくとも 1 つの単語の少なくとも一方を形成し、  
 前記第 1 の組の特性パラメータに基づき、アラビア文字に関連付けられる複数の構成要  
 素から非文字項目に関連付けられる前記複数の構成要素のうちの少なくとも 1 つの構成要  
 素をふるい落とす、請求項 20 に記載のシステム。

**【請求項 22】**

前記プロセッサは、ある欄に関連付けられる少なくとも 1 つの座標に基づき、複数の欄  
 を分類し、前記少なくとも一つの座標は画像における欄の位置に関連付けられる、請求項  
 21 に記載のシステム。 30

**【請求項 23】**

前記プロセッサは、  
 各欄に関連付けられる、各副単語および各単語の少なくとも一方に関連付けられる第 2  
 の組の特性パラメータを計算し、第 2 の組の特性パラメータは、各副単語および各単語の  
 少なくとも一方に関連付けられる行高さ、各副単語および各単語の少なくとも一方に関連  
 付けられる語間、ならびに各副単語および各単語の少なくとも一方に関連付けられる行間  
 のうちの 1 つであり、さらに、前記プロセッサは、

第 2 の組の特性パラメータに基づき少なくとも 2 つの副単語をグループ化して、少なく  
 とも 1 つの副単語および少なくとも 1 つの単語の一方を形成する、請求項 15 に記載のシ  
 ステム。 40

**【請求項 24】**

前記プロセッサは、少なくとも 1 つの副単語および少なくとも 1 つの単語を、各副単語  
 および各単語の少なくとも一方に関連付けられる行高さ、ならびに各副単語および各単語  
 の少なくとも一方に関連付けられる行間の少なくとも一方に基づき、少なくとも 1 つの横  
 行に分割する、請求項 23 に記載のシステム。

**【発明の詳細な説明】****【技術分野】****【0001】**

## 発明の分野

本発明は一般に光学式文字認識(OCR:Optical Character Recognition)用に画像を前処理することに関し、画像はアラビア文字および／または非文字項目を含む。より具体的には、本発明は複数の欄を含む画像の前処理に関し、各欄はアラビア文字および／または非文字項目を含む。

### 【背景技術】

#### 【0002】

##### 発明の背景

スキャンされた文章を編集可能および検索可能な文章に自動的に変換するには、正確かつ安定した光学式文字認識（OCR）システムを使用する必要がある。英語の文章に対するOCRシステムは、さまざまな理由により、高いレベルの精度に達している。その主な理由の1つは、OCRシステムへの入力として、英語の文章を分離された文字に前処理する機能にある。英語の文章の各文字は、印刷された英語の文章が繋がっていない性質により、分離することができる。しかし、スキャンされた繋がっている文字はOCRシステムへの課題であり、ピッチが変化している場合にその精度を落とす。

10

### 【発明の概要】

#### 【発明が解決しようとする課題】

#### 【0003】

アラビア語のスキャンされた文章は一連の繋がっている文字を含むので、文字に分割するのが難しい。アラビア文字での別の問題は、文字や後に続く母音の発音を示すために、多くの点やアクセント記号が文字の上下に入っていることである。これにより、英語向けに設計された前処理技術が正確にアラビア文字を処理することに用いられるのを妨げることとなる。

20

#### 【0004】

アラビア文字の別の特徴は、アラビア語の文章は母音を示すアクセント記号を付けてもまたは付けなくても書くことができる点にある。さらに、英語の文章は大文字または小文字の表示を有するが、多くのアラビア語の文字は、その位置が単語の初め、単語の間、もしくは単語の終わりにあるのか、または単独の単語としてあるのかに応じて、3つまたは4つの形を含む。したがって、アクセント記号および単語内での文字の位置により、アラビア文字にはさまざまな組合せが可能であるので、現在のOCR前処理システムでアラビア文字を前処理することは不的確となる。

30

#### 【0005】

さらに、アラビア文字および非文字項目の欄を複数有する画像では、各欄に関連付けられるアラビア文字はフォントのサイズ、スタイル、色などが変わり得る。フォントサイズが変わることにより、隣接する欄は行が揃わず、正確に分割できないかもしれない。

### 【課題を解決するための手段】

#### 【0006】

したがって、アラビア文字および／または非文字項目を含む複数の欄を有する画像を前処理するための方法およびシステムが必要である。

#### 【0007】

40

添付の図面であって、同じ参照符号は、それぞれの図面において同じまたは機能的に類似した要素を指し、以下の詳細な説明とともに明細書の中に組込まれてその一部をなす図面は、さまざまな実施例を示し、かつ本発明に従うさまざまな原理および利点を説明する役割を果たす。

### 【図面の簡単な説明】

#### 【0008】

【図1】本発明のさまざまな実施例に従い、アラビア文字および／または非文字項目を有する複数の欄を含む画像の一例を示す図である。

【図2】本発明の一実施例に従い、画像に関連付けられる歪みを修正する際の画像の一例を示す図である。

50

【図3】本発明の一実施例に従い、2進画像に変換された画像の一例を示す図である。

【図4】本発明の一実施例に従い、光学式文字認識のために画像を前処理する方法のフロー図である。

【図5】本発明の一実施例に従い、複数の構成要素を定める方法のフロー図である。

【図6】本発明の一実施例に従い、行の高さを計算する方法のフロー図である。

【図7】本発明の一実施例に従い、複数の欄のうちのある欄に関連付けられる1つ以上の構成要素を有する画像を示す図である。

【図8】本発明の一実施例に従い、語間を計算する方法のフロー図である。

【図9】本発明の一実施例に従い、行間を計算する方法のフロー図である。

【図10】本発明の一実施例に従い、1つ以上の副単語および／または1つ以上の単語を形成するために、複数の構成要素を合成する方法を示す図である。 10

【図11】本発明の一実施例に従い、非文字項目が削除された画像の一例を示す図である。

【図12】本発明の一実施例に従い、行高さおよび行間に応じて、1つ以上の単語および1つ以上の副単語を1つ以上の横行に分割する一例を示す図である。

【図13】本発明のさまざまな実施例に従い、アラビア文字および／または非文字項目を含む複数の欄を含む画像を前処理するためのシステムのブロック図である。

#### 【発明を実施するための形態】

##### 【0009】

図面の要素は簡潔におよび明瞭にするために示されており、必ずしも尺度通りに描かれていなければならないことは、当業者なら理解するであろう。たとえば、図面の一部の要素の寸法は、本発明の実施例をわかりやすくするために、他の要素に対して拡大されて示されているかもしれない。 20

##### 【0010】

###### 発明の詳細な説明

本発明に従う実施例を詳細に説明する前に、実施例は主に光学式文字認識（OCR）用にアラビア文字および非文字項目を含む画像を前処理するための方法およびシステムに関する方法の工程および装置の構成要素の組合せに基づいていることに注意しなければならない。画像は複数の欄を含み、各欄はアラビア文字および／または非文字項目を含む。したがって、装置の構成要素および方法の工程は、図面において適する場合は従来の記号によって示され、ここでの記載が当業者にとって容易に明らかとなる詳細でもって開示を曖昧にしないよう、本発明の実施例の理解に関連する具体的な詳細のみが示されている。 30

##### 【0011】

本明細書では、第1および第2、上および下などのような相關的用語は、あるエンティティまたは動作を別のエンティティまたは動作と区別するためにのみ用いられており、これらのエンティティまたは動作間において実際にこのような関係または順序を必ずしも必要としないまたは意味しない。「含む」、「有する」またはその他のこののような用語の変形は、限定されない含有を網羅するために意図されており、一連の要素を含むプロセス、方法、物品または装置は、これらの要素のみを含むのではなく、プロセス、方法、物品もしくは装置に明記されていない要素、または固有の他の要素をも含み得る。「...を含む」の用語が付いている要素は、それ以外の制限がなければ、その要素を含むプロセス、方法、物品または装置において付加的同一要素の存在を排除するものではない。 40

##### 【0012】

ここに記載される発明の実施例は、OCR用に画像を前処理する方法の機能の一部、大部分、またはすべてを実施するために、特定の非トランザクション・クライアント回路と併せて、1つ以上の従来のトランザクション・クライアントと、その1つ以上のトランザクション・クライアントを制御する固有の記憶されているプログラム命令とを含み得ることは理解されるであろう。画像は複数の欄を含み、各欄はアラビア文字および／または非文字項目を含む。非トランザクション・クライアント回路は、無線受信装置、無線送信装置、信号ドライバ、クロック回路、電源回路、およびユーザ入力装置を含むことができる。 50

が、これらに限定されない。したがって、これらの機能は、O C R用に、アラビア文字および非文字項目を含む画像を前処理する方法の工程として解釈することができる。代替的に、機能の一部またはすべては、プログラム命令が記憶されていない状態マシンによって、または1つ以上の特定用途向け集積回路(A S I C)において実施することができ、各機能または特定の機能の一部の組合せは、カスタム論理として実施される。これら2つのアプローチを組合せて用いることもできる。これらの機能の方法および手段がここに記載される。さらに、当業者なら、ここに開示されている概念および原理により、たとえば利用可能な時間、現行の技術および経済的な点を考慮して動機付けられる著しい努力および多くの設計的選択事項があったとしても、最小限の実験でもってこのようなソフトウェア命令、プログラムおよびI Cを容易に生成できると考えられる。

10

#### 【0013】

一般に、さまざまな実施例に従い、本発明は、O C R用に画像を前処理するための方法およびシステムを提供する。画像は複数の欄を含み、各欄はアラビア文字および/または非文字項目を含む。本方法は、複数の欄の中のアラビア文字および/または非文字項目に関連付けられる複数の構成要素を定めることを含む。ここで、構成要素は1組の繋がっているピクセルを含む。つぎに、複数の構成要素に関連付けられる行高さおよび欄間が計算される。その後、複数の構成要素のうちの1つ以上の構成要素は、行の高さおよび/または欄間に基づき、複数の欄のうちのある欄に関連付けられる。さらに、複数の欄のうちの各欄に関連付けられる第1の組の特性パラメータが計算される。各欄に関連付けられる複数の構成要素は、第1の組の特性パラメータに基づいて合成されて、1つ以上の副単語および/または1つ以上の単語を形成する。

20

#### 【0014】

図1は、本発明のさまざまな実施例に従い、複数の欄を含む画像の一例を示し、複数の欄の各欄はアラビア文字および/または非文字項目を含む。画像は濃淡画像またはカラー画像のいずれかであり得る。さらに、画像はごま塩雑音を含み、歪んでいるかもしれない。O C R用に画像を前処理する前に、画像に関連付けられるごま塩雑音および歪みは取除かれる。さらに、画像は濃淡画像またはカラー画像から2進画像に変換される。

#### 【0015】

画像に関連付けられる歪みは、画像に関連付けられる基線を定めて、基線の配列に基づき画像を正しい位置に置くことによって修正される。基線の配列は、変形ハフ変換によって定められ、横並び突出部は複数の方向で定められる。横並び突出部は、画像の前景に関連付けられるピクセルの数を示す。アラビア語の近似単語長さに対応する妥当なランレンジングスが考慮されて、最も高いピクセル密度を有する方向が決定される。最も高いピクセル密度の方向が、基線の配列と一致すると考えられる。その後、画像は基線の配列に基づき正しい位置に置かれる。図2は、画像に関連付けられる歪みを修正する際の画像の一例を示す。

30

#### 【0016】

歪みを修正する際、画像に関連付けられるごま塩雑音が取除かれる。ごま塩雑音は、任意に起こる白および黒ピクセルを表わし、暗い背景上の白い点として、および明るい背景上の黒い点を含み得る。一実施例において、ごま塩雑音はメディアンフィルタおよび/または多数フィルタを用いることによって除去することができる。当業者にとって、ごま塩雑音は当該技術分野におけるノイズ除去技術を用いることによっても除去できることは明らかであろう。

40

#### 【0017】

その後、画像は濃淡画像またはカラー画像から2進画像に変換される。たとえば画像が濃淡画像の場合、0から255の各ピクセル値を0のピクセル値または1のピクセル値に変換することにより、画像は2進画像に変換される。ある実施例において、ピクセル値0は背景値を表わし、ピクセル値1は前景値を表わす。代替的に、ピクセル値0は前景値を表わし、ピクセル値1は背景値を表わしてもよい。ピクセル値0は白ピクセルに関連付けられ、ピクセル値1は黒ピクセルに関連付けられる。

50

**【0018】**

ある画素のピクセル値を変換する前に、濃淡画像にしきい値が定められ、しきい値より上のピクセル値はピクセル値1に変換され、しきい値より下のピクセル値はピクセル値0に変換される。一実施例において、しきい値は濃淡画像のピクセル値のヒストグラムを作成することによって計算される。ヒストグラムは、各ピクセル値の頻度を表わす。このヒストグラムを作成する際、連続するピクセル値の頻度を加算して、その連続するピクセル値を、連続するピクセル値の結合された頻度を有する単一のピクセル値に置き換えることにより、平滑化されたヒストグラムを生成することができる。考慮される連続するピクセル値の数は、予め定めることができる。後で、平滑化されたヒストグラムの2つの最も顕著なピークが選択され、この2つの顕著なピーク間の最小の谷が定められる。最も低い谷の中で最も低い頻度を有するピクセル値が、しきい値として選択される。図3は例示的に2進に変換された画像を示す。

10

**【0019】**

別の例であって、画像がカラー画像の場合、カラー画像はまず濃淡画像に変換され、次に上記のように2進画像に変換される。一実施例において、カラー画像を濃淡画像に変換するために、全国テレビジョン方式委員会(NTSC)のデフォルト値を用いることができる。

**【0020】**

画像を2進画像に変換する際、ピクセル値0およびピクセル値1の発生数が数えられる。より低いカウントの2進値は前景値であると考えられ、より高いカウントを有する2進値は背景値であると考えられる。

20

**【0021】**

図4を参照すると、本発明の一実施例に従い、光学式文字認識用に画像を前処理する方法のフロー図が示される。画像は複数の欄を含み、各欄はアラビア文字および/または非文字項目を含む。前記のように、画像の2進画像への変換、ごま塩雜音の除去、および画像に関連付けられる歪みの修正のいずれか1つ以上を行なうことにより、画像が得られる。画像を前処理するために、画像の中のアラビア文字および/または非文字項目に関連付けられる複数の構成要素がステップ402で決定される。ここで、構成要素は1組の繋がっているピクセルを含む。構成要素は、文字が他の文字に繋がらない場合、アラビア文字の1文字を表わす。したがって、複数の文字が他の文字に繋がる場合、繋がっている文字は1つの構成要素であると考えられる。複数の構成要素を決定する方法は、図5と併せてさらに説明される。

30

**【0022】**

複数の構成要素を決定する際、複数の構成要素に関連付けられる行高さおよび欄間がステップ404で計算される。複数の構成要素に関連付けられる行の高さは、複数の構成要素の各構成要素の高さに対応する高さのヒストグラムを作成することによって計算される。行高さおよび/または欄間は、画像の複数の構成要素のすべての構成要素の平均値に基づき計算される。たとえば、行高さは、複数の構成要素のすべての構成要素に対して平均化された頻出高さである。行高さを計算する方法は、図6と併せてさらに詳しく説明される。欄間は、行高さの関数として動的に計算される。ステップ406において、複数の構成要素のうちの1つ以上の構成要素は、図7で例示的に示されるように、行高さおよび/または欄間に基づき、複数の欄のうちのある欄に関連付けられる。すなわち、複数の構成要素は、図7の702、704および706で例示的に示されるように、複数の構成要素に関連付けられる行高さおよび欄間に基づき、複数の欄に分離される。たとえば、2つの横方向に連続する構成要素間の間隔が、欄間よりも小さければ、その構成要素は同じ欄のものであると考えられ、それに応じて分離される。

40

**【0023】**

複数の構成要素がある欄に関連付けられると、ステップ408において、第1の組の特性パラメータが各欄について計算される。ある実施例において、第1の組の特性パラメータは、各欄に関連付けられる行高さ、各欄に関連付けられる語間、各欄に関連付けられる

50

行間、各構成要素のピクセルの数、各構成要素の幅、各構成要素の高さ、各構成要素の座標、各構成要素の密度、および各構成要素のアスペクト比を含む。行高さ、語間、および行間を計算する方法は、それぞれ図6、図8、および図9と併せて説明される。

#### 【0024】

その後、ステップ410において、各欄に関連付けられた複数の構成要素は、第1の組の特性パラメータに基づき合成されて、1つ以上の副単語および／または1つ以上の単語を形成する。複数の構成要素を合成する方法は、図10と併せてさらに説明される。

#### 【0025】

図5は、本発明の一実施例に従い、複数の構成要素を定める方法のフロー図を示す。ステップ502において、ラスタスキャンが画像に対して行なわれる。ラスタスキャンでは複数の掃引を行ない、複数の構成要素に対応する各ピクセル列に対して1回掃引される。10 ラスタスキャンの1回以上の掃引の実行により、画像の前景に関連付けられる1つ以上のピクセルがステップ504において特定される。画像の前景は、複数の構成要素に対応する。その後、ステップ506において、複数のピクセル間の相互接続に基づき、その複数のピクセルは統合されて、1つ以上の組の接続ピクセルを形成する。一実施例において、複数のピクセルは、8個の隣接ピクセルと1つ以上繋がっている場合に相互接続していると考えられる。こうして、アラビア文字の連続する文字は、連続する文字に関連付けられる1つ以上のピクセルが互いに相互接続されている場合に、单一の構成要素を形成する。

#### 【0026】

たとえば、ラスタスキャンの現行の掃引で特定されたピクセルは、当該ピクセルが前回の掃引で特定されたピクセルと繋がる場合には、そのピクセルと統合される。現行の掃引で特定されたピクセルが、前回の掃引で特定された複数のピクセルと繋がる場合、当該ピクセルはその複数のピクセルと統合される。別の例では、現行の掃引で特定された複数のピクセルが繋がっている場合、その複数のピクセルは統合される。同様に、ラスタスキャンの後続の掃引で特定される1つ以上のピクセルは、その1つ以上のピクセルが互いに繋がる場合にも統合される。統合されたピクセルは、1組の繋がっているピクセルを形成し、複数の構成要素のうちのある構成要素に対応付けられる。20

#### 【0027】

接続するピクセルの組を決定する際に、各構成要素のピクセルの数、各構成要素の幅、各構成要素の高さ、各構成要素の座標、各構成要素の密度、および各構成要素のアスペクト比のいずれか1つ以上は、各構成要素に関連付けられる接続ピクセルをトラッキングすることにより計算される。30

#### 【0028】

図6を参照すると、本発明の一実施例に従い、行の高さを計算する方法のフロー図が示される。本方法はステップ602において、複数の構成要素の各々の高さに対応する高さのヒストグラムを作成することを含む。ヒストグラムは、複数の構成要素の各々の高さの頻出を表わす。ヒストグラムを作成する際、連続する高さ値を加算して、その連続する高さ値を、連続する高さ値の結合された頻度を有する单一の高さ値と置き換えることにより、平滑化されたヒストグラムを生成することができる。考慮される連続する高さ値の数は予め定められてもよい。たとえば、連続する高さ値の数が3個であると定められたのなら、高さが20ピクセルの頻度は、高さが19ピクセルの頻度プラス高さが20ピクセルの頻度プラス高さが21ピクセルの頻度となる。40

#### 【0029】

平滑化されたヒストグラムが得られると、頻出高さがステップ604で特定される。頻出高さを特定するために、アラビア文字に対応するアクセント記号や句読点の小さな構成要素の高さは除外される。これは、しきい値高さを設定し、頻出高さを特定するのに、しきい値高さより大きい高さを有する構成要素のみを考慮することによって行なわれる。頻出高さは、画像が複数の文字サイズを有する場合には、画像の主要文字サイズを表わす。

#### 【0030】

頻出高さを特定する際、行高さはステップ606において頻出高さに基づき計算される50

。行高さは、頻出高さおよび乗率の積として計算される。乗率は頻出高さに依存する。行高さは、1つ以上の単語および/または1つ以上の副単語を、アラビア文字の1つ以上の横行に分割するのに用いることができる。さらに、行高さは、図8と併せて説明したように、語間を計算するために用いられる。

#### 【0031】

図8は、本発明の一実施例に従い、語間を計算する方法のフロー図を示す。本方法は、ステップ802において、複数の構成要素のうちの連続する構成要素間の間隔のヒストグラムを作成することを含む。一実施例において、縦方向に重なり、かつ他の構成要素によって分けられていない2つの構成要素はすべて連続する構成要素であると考えられる。2つの構成要素は、縦軸に沿って1つ以上の共通の座標を共有する場合、縦方向に重なる。すなわち、連続する構成要素は、アラビア文字1行に属する。代替的に、2つの構成要素が縦方向に重ならない場合、2つの構成要素は予め定められた距離で分けられている場合に、連続する構成要素であると考えられる。

10

#### 【0032】

連続する構成要素間の間隔のヒストグラムを作成する際、平滑化されたヒストグラムは、連続する間隔値を加算することにより生成できる。連続する間隔値は、連続する間隔値の結合させられた頻度を有する单一の間隔値と置き換えられる。たとえば、10ピクセルの間隔値の頻度は、9ピクセルの間隔値の頻度と、10ピクセルの間隔値の頻度と、11ピクセルの間隔値の頻度との合計と置き換えられる。

20

#### 【0033】

ステップ804において、平滑化されたヒストグラムから頻出間隔が特定される。頻出間隔は、行高さによって定められるしきい値範囲内から特定される。たとえば、5分の1の行高さと半分の行高さとの間にある頻出間隔値を対象とすることができる。ステップ806において、語間は頻出間隔に基づき計算される。語間は、アラビア文字の2つの連続するワードの間の間隔である。

30

#### 【0034】

図9は、本発明の一実施例に従い、行間を計算する方法のフロー図を示す。ステップ902において、前景に対応する複数の構成要素の複数の横並び突出部のヒストグラムが作成される。横並び突出部は、ラスタスキャンの掃引に対応する複数の構成要素に関連付けられるピクセルの数を示す。たとえば、ラスタスキャンの掃引が、複数の構成要素に関連付けられる15個のピクセルを特定すると、その掃引に対するピクセル列の横並び突出部は15である。

#### 【0035】

その後、ステップ904において、2つの連続する最大横並び突出部間の平均距離が計算される。最大横並び突出部は、最も高い密度の領域を表わす。その後、ステップ906において、行間は、平均距離に基づき計算される。

#### 【0036】

第1の組の特性パラメータを計算する際、複数の構成要素は合成されて、図10と併せて説明されるように、1つ以上の副単語および/または1つ以上の単語を形成する。

40

#### 【0037】

図10は、本発明の一実施例に従い、1つ以上の副単語および/または1つ以上の単語を形成するために、複数の構成要素を合成する方法を示す。ステップ1002において、連続する構成要素間の間隔が、語間の係数未満である場合に結合される。語間に加えて、連続する構成要素の座標も、連続する構成要素が結合される場合に定めることができる。連続する構成要素の語間および/または座標に基づいて連続する構成要素を結合することは、アラビア文字のある単語または副単語に対応する異なる構成要素の結合を引起す。

#### 【0038】

たとえば、アクセント記号に関連付けられる構成要素は、構成要素の語間および位置に基づき、属する単語と結合される。一つの単語は1つ以上の構成要素を含み得る。構成要素の位置は、構成要素の座標によって定められる。ある構成要素に関連付けられる第1の

50

組の特性パラメータが、アラビア文字の句読点またはアクセント記号と類似しており、かつアラビア語の文字に対応する構成要素に対して適切に隣接している場合、その構成要素は文字とともにグループ化されて単語または副単語を形成する。さもなければ、構成要素はノイズであると考えられ、除去される。

#### 【0039】

アラビア文字に関連付けられる構成要素を結合することに加えて、非文字項目に関連付けられる構成要素も、語間に基づき結合される。非文字項目に関連付けられる構成要素は結合されて、1つ以上のより大きい構成要素を形成する。

#### 【0040】

ステップ1004において、非文字項目に関連付けられる構成要素は、第1の組の特性パラメータに基づき、アラビア文字に関連付けられる構成要素からふるい落とされる。たとえば、大きい高さ、大きい幅、および低い密度を有する構成要素は取除かれる。これらの構成要素は、ある欄の周りまたは他の非文字項目の周りの枠またはボーダーに対応し得る。同様に、大きい高さ、小さい幅、および高い密度を有する構成要素は縦線として認識され、除去される。横線は小さい高さ、大きい幅、および高い密度を有するものとして認識される。

10

#### 【0041】

同様に、他の非文字項目も1つ以上のフィルタに基づき除去される。この1つ以上のフィルタは、画像の共通に起こる構成要素の長さ、構成要素の幅、構成要素のアスペクト比、構成要素の密度、および構成要素の合計数を用いて、非文字項目をアラビア文字からふるい落とす。2つ以上の欄にわたる非文字項目も、非文字項目に関連付けられる構成要素の寸法を、アラビア文字に関連付けられる構成要素の最もよく起こる寸法と比較することによって除去される。図11は、非文字項目が除去された画像を例示的に示す。

20

#### 【0042】

非文字項目をアラビア文字からふるい落とした後、1つ以上の単語および1つ以上の副単語の第2の組の特性パラメータが計算される。第2の組の特性パラメータは、各副単語および各単語の少なくとも一方に関連付けられる行高さ、各副単語および各単語の少なくとも一方に関連付けられる語間、ならびに各副単語および各単語の少なくとも一方に関連付けられる行間を含む。第2の組の特性パラメータは、1つ以上の副単語および/または1つ以上の単語を形成するために、複数の構成要素を結合するプロセスの精度をさらに上げるために、計算される。第2の組の特性パラメータに基づき、1つ以上の副単語をグループ化して、1つ以上の副単語および/または1つ以上の単語を形成する。

30

#### 【0043】

その後、縦方向に重なり、かつ複数の欄のうちのある欄に関連付けられる1つ以上の副単語および1つ以上の単語は、分割されてアラビア文字の横行を形成する。一実施例において、1つ以上の副単語および1つ以上の単語は、行高さおよび/または行間に基づいても分割されてもよい。たとえば、互いに重なる縦方向の構成要素を1つ以上有するので2本の横行が一緒に分割されると、その2本の横行は、行高さおよび/または行間に基づき分けられる。図12は行高さおよび行間に依存して、1つ以上の単語および1つ以上の副単語を1つ以上の横行に分割する例を示す。

40

#### 【0044】

こうして、OCR用に画像を前処理する方法が開示される。画像は複数の欄を含み、各欄はアラビア文字および/または非文字項目を有する。この方法は、アラビア文字および非文字項目に関連付けられる複数の構成要素を定めることを含む。複数の構成要素のうちのある構成要素は、アラビア語の1つ以上の文字または1つ以上の非文字項目を表わす。複数の文字が相互接続されるのなら、構成要素は2つ以上の文字を表わす。

#### 【0045】

複数の構成要素を決定する際、複数の構成要素に関連付けられる行高さおよび欄間が計算される。行高さおよび欄間は、すべての欄にわたるすべての構成要素の平均値を表わす。複数の構成要素は、平均行高さおよび平均欄間に基づき、1つ以上の欄に分離される。

50

後で、各欄の複数の構成要素に関連付けられる第1の組の特性パラメータが計算される。各欄に関連付けられる複数の構成要素は、後で第1の組の特性パラメータに基づき合成されて、1つ以上の副単語および/または1つ以上の単語を形成する。

#### 【0046】

ここに開示されている方法は、繋がっている文字を含み、かつ複数の欄を含むアラビア文字を正確に前処理して分割することを可能にする。本方法は、構成要素がノイズであるのかアラビア文字の一部であるのかを判断する場合に、アラビア文字に関連付けられるアクセント記号および句読点を考慮する。さらに、本方法は画像が複数の欄を含むか否かを特定し、それらを分離する。

#### 【0047】

図13は本発明の一実施例に従い、光学式文字認識(OCR)のために、アラビア文字および/または非文字項目を含む画像を前処理するためのシステム1300のブロック図を示す。画像は複数の欄を含み、各欄はアラビア文字および/または非文字項目を含む。画像は濃淡画像およびカラー画像のどちらかであり得る。さらに、画像はごま塩雜音を含み、歪んでいるかもしれない。図13に示されるように、システム1300は、メモリ1302と、メモリ1302に結合されるプロセッサ1304とを含む。OCR用に画像を前処理する前に、プロセッサ1304は、図1と併せて説明したように、変形ハフ変換を用いて画像に関連付けられる歪みを除去する。その後、プロセッサ1304はごま塩雜音を除去し、濃淡画像またはカラー画像を2進画像に変換する。一実施例において、ごま塩雜音は、メディアンフィルタおよび/または多数フィルタを用いて除去され得る。ここでは、画像を前処理するために、プロセッサ1304はアラビア文字および/または非文字項目に関連付けられる複数の構成要素を定める。構成要素は接続されたピクセルの組を含む。構成要素は、文字が他の文字と繋がらない場合、アラビア語文字の1つの文字を表わす。したがって、複数の文字が他の文字と繋がる場合、繋がっている文字は1つの構成要素であると考えられる。

#### 【0048】

一実施例において、複数の構成要素を定めるために、プロセッサ1304は画像に対してラスタスキャンを行なう。ラスタスキャンでは複数の掃引を行ない、複数の構成要素に対応する各ピクセル列に対して1回掃引される。ラスタスキャンの1回以上の掃引の実行により、画像の前景に関連付けられる1つ以上のピクセルが特定される。画像の前景は、複数の構成要素に対応する。その後、プロセッサ1304は、複数のピクセル間の相互接続に基づき、複数のピクセルを統合して、1つ以上の組の接続ピクセルを形成する。統合されたピクセルは1組の接続ピクセルを形成し、複数の構成要素のうちのある構成要素に関連付けられる。

#### 【0049】

こうしてプロセッサ1304によって定められた複数の構成要素は、メモリ1402に記憶することができ、プロセッサ1304によって用いられて、複数の構成要素に関連付けられる行高さおよび欄間が計算される。行高さおよび欄間を用いて、複数の構成要素のうちの1つ以上の構成要素を、複数の欄のある欄に関連付ける。すなわち、複数の構成要素が行高さおよび/または欄間を満たすのなら、複数の構成要素はプロセッサ1304によって複数の欄に分離される。たとえば、2つの縦にまたは横に連続する構成要素間の間隔が欄間よりも小さければ、それらの構成要素間は同じ欄のもであると考えられて、分離される。その後、複数の構成要素に関連付けられる第1の組の特性パラメータが計算される。ある実施例において、第1の組の特性パラメータは、各欄に関連付けられる行高さ、各欄に関連付けられる語間、各欄に関連付けられる行間、各構成要素のピクセルの数、各構成要素の幅、各構成要素の高さ、各構成要素の座標、各構成要素の密度、および各構成要素のアスペクト比を含む。その後、プロセッサ1304は第1の組の特性パラメータに基づき、複数の構成要素を合成する。合成された構成要素は、1つ以上の副単語および/または1つ以上の単語を形成する。

#### 【0050】

10

20

30

40

50

一実施例において、プロセッサ 1304 は、複数の構成要素の各々の高さに対応する高さのヒストグラムを作成することによって行高さを計算する。ヒストグラムから頻出高さがプロセッサ 1304 によって特定される。その後、プロセッサ 1304 は頻出高さおよび乗率の積として、行高さを計算する。乗率は頻出高さに依存する。行高さを用いて、1 つ以上の単語および/または 1 つ以上の副単語を、アラビア文字の 1 つ以上の横行に分割することができる。さらに、プロセッサ 1304 は行高さを用いて語間を計算する。

#### 【0051】

次に、プロセッサ 1304 は、複数の構成要素の連続する構成要素間の間隔のヒストグラムを作成することにより、語間を計算する。プロセッサ 1304 は、ヒストグラムから頻出間隔を特定する。頻出間隔はしきい値範囲内から特定され、そのしきい値範囲は行高さに基づいている。その後、語間はプロセッサ 1304 によって頻出間隔に基づき計算される。語間は、アラビア文字の 2 つの連続するワードの間の間隔である。10

#### 【0052】

プロセッサ 1304 は、複数の構成要素の複数の横並び突出部のヒストグラムを作成することによって、行間を計算するよう構成されている。横並び突出部は、ラスタスキャンの各掃引に対応する複数の構成要素に関連付けられるピクセルの数を示す。次に、2 つの連続する横並び突出部間の平均距離は、プロセッサ 1304 によって計算される。その後、プロセッサ 1304 は平均距離に基づき、行間を計算する。

#### 【0053】

さらに、プロセッサ 1304 は、各構成要素のピクセル数、各構成要素の幅、各構成要素の高さ、各構成要素の座標、各構成要素の密度、および各構成要素のアスペクト比を定める。20

#### 【0054】

前述のように、プロセッサ 1304 は、第 1 の組の特性パラメータに基づき、各欄に関連付けられる複数の構成要素を合成する。これを行なうため、プロセッサ 1304 は、構成要素間の間隔が、その欄に関連付けられる語間の係数よりも小さい場合に、連続する構成要素を結合する。各欄に関連付けられる語間に加えて、連続構成要素が結合されるのなら、連続する構成要素の座標も定めることができる。さらに、プロセッサ 1304 は、図 10 と併せて説明されたように、第 1 の組の特性パラメータに基づき、アラビア文字に関連付けられる構成要素から、非文字項目に関連付けられる構成要素をふるい落とす。非文字項目をふるい落とすことは、1 つ以上の副単語および/または 1 つ以上の単語をもたらす。30

#### 【0055】

プロセッサ 1304 は、1 つ以上の副単語および/または 1 つ以上の単語に関連付けられる第 2 の組の特性パラメータを計算するよう構成されている。第 2 の組の特性パラメータは、各副単語および/または各単語に関連付けられる行高さ、各副単語および/または各単語に関連付けられる語間、ならびに各副単語および/または各単語に関連付けられる行間を含む。次に、2 つ以上の副単語は、第 2 の組の特性パラメータに基づきプロセッサ 1304 によってグループ化されて、1 つ以上の副単語および/または 1 つ以上の単語を形成する。すなわち、2 つ以上の副単語は第 2 の組の特性パラメータに基づきグループ化されて、完全な単語またはより大きい副単語を形成する。40

#### 【0056】

1 つ以上の副単語および 1 つ以上の単語を形成する際、プロセッサ 1304 は、縦方向に重なりかつ複数の欄のある欄に関連付けられる 1 つ以上の副単語および 1 つ以上の単語を分割して、アラビア文字の横行を形成する。一実施例において、1 つ以上の副単語および 1 つ以上の単語は、行高さおよび/または行間にに基づき、プロセッサ 1304 によって分割されてもよい。

#### 【0057】

本発明の多様な実施例は、OCR 用に画像を前処理するための方法およびシステムを提供する。画像は複数の欄を含み、各欄はアラビア文字および/または非文字項目を含む。50

本発明は、アラビア文字を、O C R システムによって正確に処理することができる分割された行の副単語および単語に分割する。本方法は、構成要素がノイズであるのかアラビア文字の一部であるのかを判断する場合に、アラビア文字に関連付けられるアクセント記号および句読点を考慮に入れる。

#### 【 0 0 5 8 】

当業者は、ここに記載される利点および他の利点は一例であって、本発明のさまざまな実施例の利点すべてを含むことは意図されていないと認識するであろう。

#### 【 0 0 5 9 】

以上で、本発明の具体的実施例が説明された。しかし、当業者なら、さまざまな変形および変更が、添付の請求項に記載されている本発明の範囲から逸脱することなく行なうことができると理解するであろう。したがって、明細書本文および図面は限定するのではなく例示するものであり、変形はすべて本発明の範囲内に含まれることが意図される。利益、利点、問題の解決、および利益、利点または解決を引起すまたは顕著にする要素は、クレームのいずれかまたはすべてにおける重大な、必要な、または必須の特徴もしくは要素であると考えるべきではない。本発明は、本願の係属中になされた補正を含む添付の請求項およびこれら請求項の均等物すべてによってのみ規定される。10

#### 【 符号の説明 】

#### 【 0 0 6 0 】

1 3 0 0 システム

1 3 0 2 メモリ

1 3 0 4 プロセッサ

4 0 2 アラビア文字および / または非文字項目に関連付けられる複数の構成要素を定める20

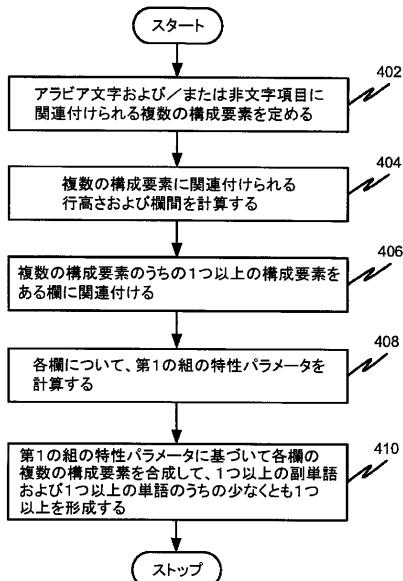
4 0 4 複数の構成要素に関連付けられる行高さおよび欄間を計算する

4 0 6 複数の構成要素のうちの 1 つ以上の構成要素がある欄に関連付ける

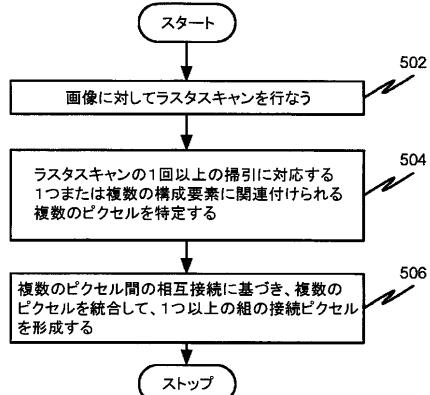
4 0 8 各欄について、第 1 の組の特性パラメータを計算する

4 1 0 第 1 の組の特性パラメータに基づいて各欄の複数の構成要素を合成して、1 つ以上の副単語および 1 つ以上の単語のうちの少なくとも 1 つ以上を形成する

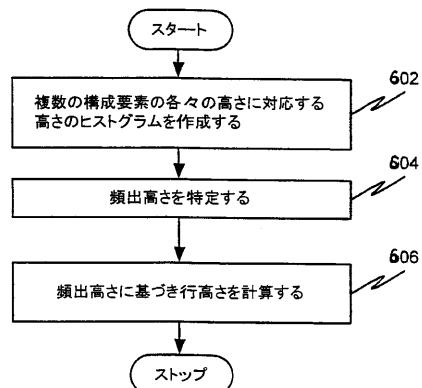
【図4】



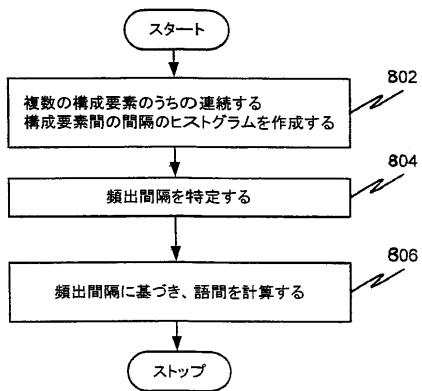
【図5】



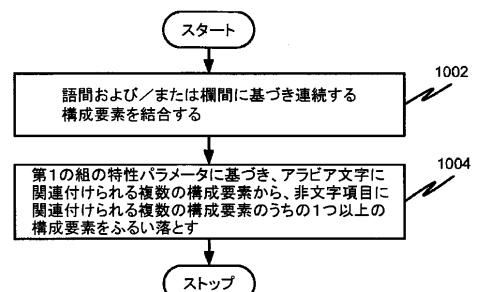
【図6】



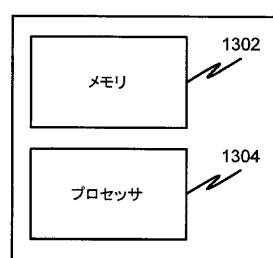
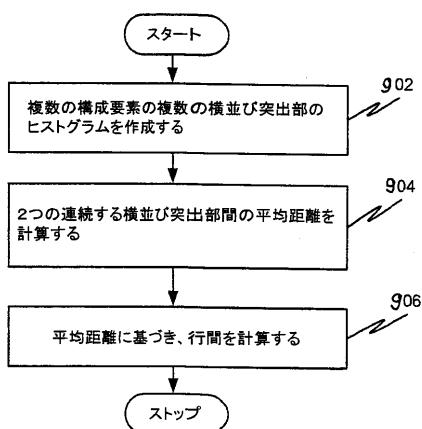
【図8】



【図10】



【図9】



【図13】

1300

【図1】



【図2】



( 义 3 )

( 义 7 )



[ ﺖ ١ ١ ]

مدينة الملك عبد العزيز  
لعلوم والتكنولوجيا  
KACST

بسم الله الرحمن الرحيم

# أخبار المدينة

نشرة إعلامية شهرية تصدر عن مدينة الملك عبد العزيز للعلوم والتكنولوجيا

العدد ١١٥ - ربيع الآخر ١٤٣٢ هـ - أبريل ٢٠١٣

وزارته للتربية  
المعاهد والمدارس  
الابتدائية والمتوسطة  
المدارس الاعدادية والثانوية

إنفاذية سعد بن  
أميريحة لتطوير  
الاتصالات وخدمات  
الإنترنت والهاتف  
المحمول

مكتب العلاج  
والتنمية المتخصصة  
مكتب عهد ناسا  
لعلوم الفضاء والرائد  
ضباء كلية أنهار

معالي وزير الثقافة والإعلام  
وزير مدينة «العلوم  
والتكنولوجيا»

## التأسيس لعلاقة استراتيجية بين المدينة و جامعة الملك عبد الله للعلوم والتكنولوجيا

قام معالي وزير الثقافة والإعلام - محمد العزيز

بن معين الدين جوهر برزيارة قصر المدينة يوم السبت ١٧/٣/١٤٣٢هـ، وكان به استثنائه معالي رئيس المدينة د. محمد بن إبراهيم السويلي ونائبه، رئيس لجنة دعم البحث العلمي د. عبد الله بن الحسين الرشيد وعدد من المسؤولين بالجامعة.

فيما، عمالق رئيس الجامعة، ووزير الثقافة والإسلام والوزير الأعلى للطريق والمطرق إلى المدينة وأهميتها العلمية والتكنولوجية، وعمالي ورؤوف مصطفى ووزير الإسلام على العهد التي تدعاها مدينة الملك عبد العزيز للعلوم والتكنولوجيا في مجال البحث العلمي وتحقيق أشغاله على مستوى المملكة فضلاً عن بعض التحولات التي تشهدها بفضل من مسارات الدعم والتنمية من خلال إصلاح وتأهيل متخصصون أثروا ما يدعوه المدينه بـ «هذا المكان».

افتتح سمو نائب رئيس مدينة الملك عبد العزيز للعلوم والتكنولوجيا ملتقى الباحثون، رئيس اللجنة العلمية المنظمة للملتقى، الشهري السادس للعلوم والتكنولوجيا والابتكار، الأستاذ، تركي بن سعيد، وذلك في قاعة المؤتمرات بمقر جامعة الملك عبد الله للعلوم والتكنولوجيا.

**د. تركي بن سعood : دعم ٣٠٠ مليون ريال بحثي استراتيجي بتكلفة تجاوزت ٦٤٥ مليون ريال خلال عام واحد**

أكد سمو نائب رئيس مدينة الملك عبد العزيز للعلوم والتكنولوجيا ملتقى الباحثون، رئيس اللجنة العلمية المنظمة للملتقى، الشهري السادس للعلوم والتكنولوجيا والابتكار، الأستاذ، تركي بن سعيد، أن سموه، أن الخطة الخمسية الأولى للعلوم والتكنولوجيا والابتكار حققت خلال العام المالي الماضي ١٤٣١/١٤٣٢هـ (٩٠٠٠) إنجازات كبيرة ونكلة نوعية غير مسبوقة، على مستوى المملكة والتكنولوجيا عامة، في دعم وتمويل منظومة البحث والتطوير والابتكار في المملكة ... (البيان ص:٦)

【 义 1 2 】



---

フロントページの続き

(72)発明者 フセイン・ハリド・アル-オマリ

ヨルダン、11953 アンマン、ピイ・オウ・ボックス・2750

(72)発明者 モハメド・スレイマン・ホルシード

サウジアラビア、11443 リヤド、ビルディング・17、ロック・エフ、オフィス・415  
、ピイ・オウ・ボックス・6086、キング・アブドゥルアジズ・シティ・フォー・サイエンス・  
アンド・テクノロジー

審査官 広 島 明芳

(56)参考文献 特表2009-545807(JP,A)

特開2005-346707(JP,A)

特開平09-218923(JP,A)

Jawad H AlKhateeb,外3名, Component-based Segmentation of Words from Handwritten Arabic Text , International Journal of Computer Systems Science and Engineering 5:1 2009 , 2009年, p.54-58

(58)調査した分野(Int.Cl., DB名)

G 06 K 9 / 20