

(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2006-165944
(P2006-165944A)

(43) 公開日 平成18年6月22日(2006.6.22)

(51) Int.CI.

H04L 9/32 (2006.01)
G06K 9/00 (2006.01)

F 1

H04L 9/00
G06K 9/00

テーマコード(参考)

5 B 0 6 4
5 J 1 0 4

審査請求 未請求 請求項の数 18 O L (全 21 頁)

(21) 出願番号
(22) 出願日特願2004-353597 (P2004-353597)
平成16年12月7日 (2004.12.7)

(71) 出願人 000005108
株式会社日立製作所
東京都千代田区丸の内一丁目6番6号

(74) 代理人 110000062
特許業務法人第一国際特許事務所

(72) 発明者 伊藤 信治
神奈川県川崎市麻生区王禅寺1099番地
株式会社日立製作所システム開発研究所
内

(72) 発明者 本多 義則
神奈川県川崎市麻生区王禅寺1099番地
株式会社日立製作所システム開発研究所
内

最終頁に続く

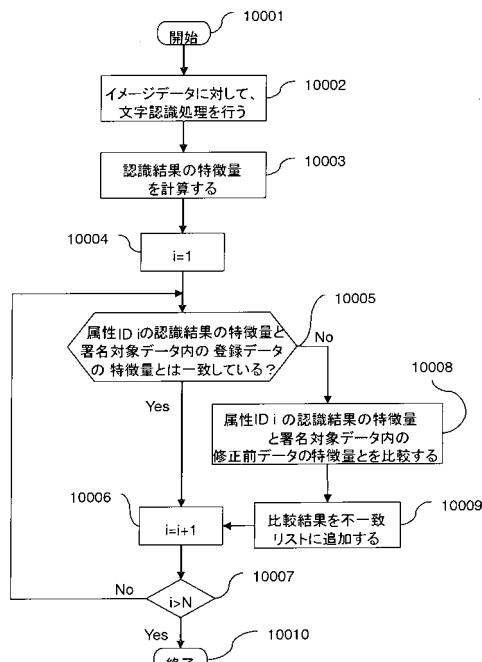
(54) 【発明の名称】イメージデータへの登録方法及び装置、登録プログラム及びそれを記録した記録媒体、並びにイメージデータの検証方法及び装置、検証プログラム及びそれを記録した記録媒体

(57) 【要約】

【課題】テキスト情報の記載されたイメージデータの改ざん場所を特定する。

【解決手段】テキスト情報の記載されたイメージデータに対して文字認識処理を行い、その文字認識結果とイメージデータに対して電子署名を付与する。電子署名は文字認識結果とイメージデータのそれぞれの特徴量を結合したデータに対して付与する。改ざん場所の特定は、電子署名の検証とイメージデータに文字認識処理を行い、更に、文字認識の結果の特徴量を計算し、保存されている文字認識結果の特徴量と比較することで、両特徴量が不一致となる箇所を改ざん箇所として特定する。

【選択図】 図10



【特許請求の範囲】**【請求項 1】**

1 項目以上のテキスト部分を含むイメージデータを取得するイメージデータ取得ステップと、

前記イメージデータ取得ステップで取得された前記イメージデータから前記テキスト部分のテキスト情報を認識するテキスト情報認識ステップと、

前記イメージデータ及び前記テキスト情報認識ステップで認識された1項目以上のテキスト情報のそれぞれについて特徴量を計算する特徴量計算ステップと、

前記特徴量計算ステップで計算された前記特徴量に対して電子署名を付与する署名付与ステップと、

前記イメージデータ、前記特徴量及び前記署名付与ステップで付与された前記電子署名を記録する記録ステップとを有することから成るイメージデータの登録方法。

【請求項 2】

前記認識したテキスト情報を修正するテキスト情報修正ステップを更に有し、

前記特徴量計算ステップにおいて、前記テキスト情報修正ステップで修正された前記テキスト情報の特徴量を更に計算し、

前記署名付与ステップにおいて、修正された前記テキスト情報について前記特徴量計算ステップで計算された前記特徴量に対して更に電子署名を付与し、

前記記録ステップにおいて、前記修正されたテキスト情報の前記特徴量とその電子署名を更に記録することから成る請求項1に記載のイメージデータの登録方法。

【請求項 3】

前記特徴量計算ステップにおいて、前記イメージデータ若しくは認識又は修正された前記テキスト情報のサイズに応じて、前記特徴量の計算方法が異なることから成る請求項1又は2に記載のイメージデータの登録方法。

【請求項 4】

前記特徴量の計算において、前記イメージデータ若しくは認識又は修正された前記テキスト情報のサイズが、予め決められた値よりも小さい場合には前記イメージデータ又は前記テキスト情報のサイズを前記特徴量とし、前記予め決められた値以上である場合には前記イメージデータ又は前記テキスト情報のサイズが入力される一方向性関数の出力値を前記特徴量とすることから成る請求項3に記載のイメージデータの登録方法。

【請求項 5】

請求項1乃至請求項4の何れかに記載の前記各ステップをコンピュータに実行させることから成るイメージデータへの登録プログラム。

【請求項 6】

請求項5に記載の前記登録プログラムを記録してあることから成るコンピュータでの読み取りが可能な記録媒体。

【請求項 7】

1 項目以上のテキスト部分が含まれているイメージデータを取得するイメージデータ取得部と、

前記イメージデータ取得部で取得された前記イメージデータ内の前記テキスト部分の前記テキスト情報を認識するテキスト情報認識部と、

前記イメージデータ及び前記テキスト情報認識部で認識された1項目以上のテキスト情報のそれぞれについて特徴量を計算する特徴量計算部と、

前記特徴量計算部で計算された前記特徴量に対して電子署名を付与する署名付与部と、

前記イメージデータ、前記特徴量及び前記署名付与部で付与された前記電子署名を記録する記録部とを有することから成るイメージデータの登録装置。

【請求項 8】

前記テキスト情報認識部で認識された前記テキスト情報を修正するテキスト情報修正部を更に有し、

前記特徴量計算部は、前記テキスト情報修正部で修正された前記テキスト情報の特徴量

10

20

30

40

50

を更に計算し、

前記署名付与部は、修正された前記テキスト情報の前記特徴量に対して更に電子署名を付与し、

前記記録部は、修正された前記テキスト情報の前記特徴量とその電子署名を更に記録することから成る請求項7に記載のイメージデータの登録装置。

【請求項9】

前記特徴量計算部は、認識又は修正された前記テキスト情報のサイズに応じて、前記特徴量の計算方法を異にすることから成る請求項7又は8に記載のイメージデータの登録装置。

【請求項10】

前記特徴量計算部は、前記イメージデータ若しくは認識又は修正された前記テキスト情報のサイズが、予め決められた値よりも小さい場合には前記イメージデータ又は前記テキスト情報のサイズを前記特徴量とし、前記予め決められた値以上である場合には前記イメージデータ又は前記テキスト情報のサイズが入力される一方向性関数の出力値を前記特徴量とすることから成る請求項9に記載のイメージデータの登録装置。

【請求項11】

1項目以上のテキスト部分が含まれているイメージデータと、前記イメージデータ内の前記テキスト部分についてのテキスト情報の特徴量と、当該特徴量に対して付与された電子署名とを取得する取得ステップと、

前記取得ステップで取得された前記電子署名を検証する署名検証ステップと、

前記イメージデータ内の前記テキスト情報を認識するテキスト情報認識ステップと、

前記認識したテキスト情報の特徴量を計算する特徴量計算ステップと、

前記取得ステップで取得された前記テキスト情報の前記特徴量と前記テキスト情報認識ステップで認識された前記テキスト情報の前記特徴量とを比較する比較ステップと、

前記比較ステップでの比較において前記取得ステップで取得された前記特徴量と前記テキスト情報認識ステップで認識された前記特徴量とが不一致であるとされた前記テキスト情報が前記改ざん場所であると特定する改ざん場所特定ステップとを有することから成るイメージデータの検証方法。

【請求項12】

前記イメージデータ内の前記テキスト情報の配置を定義したフォーマット定義情報を読み込むフォーマット定義情報読み込みステップを更に有し、

前記比較ステップにおいて、前記テキスト情報認識ステップで認識された前記テキスト情報の前記特徴量は、前記フォーマット定義情報読み込みステップで読み込まれた前記フォーマット定義情報から取得された前記イメージデータ内の前記テキスト情報についての前記特徴量と比較されることから成る請求項11に記載のイメージデータの検証方法。

【請求項13】

前記比較ステップにおいて、前記取得ステップで取得された前記イメージデータ内の前記テキスト情報の前記特徴量と前記テキスト情報認識ステップで認識された前記テキスト情報の前記特徴量とが不一致となるときに、前記テキスト情報の場所を前記イメージデータ上で表示する改ざん場所表示ステップを更に有することから成る請求項11又は12に記載のイメージデータの検証方法。

【請求項14】

請求項11乃至請求項13の何れかに記載の前記各ステップをコンピュータに実行させることから成るイメージデータの検証プログラム。

【請求項15】

請求項14に記載の前記検証プログラムを記録してあることから成るコンピュータでの読み取りが可能な記録媒体。

【請求項16】

1項目以上のテキスト部分が含まれているイメージデータと、前記イメージデータ内の前記テキスト部分についてのテキスト情報の特徴量と、当該特徴量に対して付与された電

10

20

30

40

50

子署名とを取得する取得部と、

前記取得部で取得された前記電子署名を検証する署名検証部と、

前記イメージデータ内の前記テキスト情報を認識するテキスト情報認識部と、

前記認識したテキスト情報の特徴量を計算する特徴量計算部と、

前記取得部で取得された前記テキスト情報の前記特徴量と前記テキスト情報認識部で認識された前記テキスト情報の前記特徴量とを比較する比較部と、

前記比較部での比較において前記取得部で取得された前記特徴量と前記テキスト情報認識部で認識された前記特徴量とが不一致であるとされた前記テキスト情報を前記改ざん場所として特定する改ざん場所特定部とを有することから成るイメージデータの検証装置。

【請求項 1 7】

前記イメージデータ内の前記テキスト情報の配置を定義したフォーマット定義情報を読み込むフォーマット定義情報読み込み部を更に有し、

前記比較部において、前記テキスト情報認識部で認識された前記テキスト情報の前記特徴量は、前記フォーマット定義情報読み込み部で読み込まれた前記フォーマット定義情報から取得された前記イメージデータ内の前記テキスト情報についての前記特徴量であることから成る請求項 1 6 に記載のイメージデータの検証装置。

【請求項 1 8】

前記比較部において、前記取得部で取得された前記イメージデータ内の前記テキスト情報の前記特徴量と前記テキスト情報認識部で認識された前記テキスト情報の前記特徴量とが不一致とされた前記テキスト情報の場所を前記イメージデータ上で表示する改ざん場所表示部とを更に有することから成る請求項 1 6 又は 1 7 に記載のイメージデータの検証装置。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、イメージデータに対する登録方法及び装置、登録プログラム及びそれを記録した記録媒体、並びに登録されたイメージデータが改ざんされた場合にその改ざん場所を特定可能なイメージデータの検証方法及び装置、検証プログラム及びそれを記録した記録媒体に関する。

【背景技術】

【0 0 0 2】

情報技術の発展に伴い、従来、紙で保存が義務付けられていた書類をスキャニング装置を用いて電子化しイメージデータとして保存するというニーズが高まっている。しかし、電子化されたイメージデータは、紙文書と比較して、痕跡を残さずに改ざん可能であるという問題がある。

【0 0 0 3】

この問題に対して、特許文献 1 には、イメージデータに対して電子署名を付与することで、改ざんを検知可能とする手法として電子データ化方法及び装置が開示されている。

【0 0 0 4】

また、特許文献 2 では、電子透かし技術を用いることでイメージデータの改変場所を特定する手法が開示されている。

【0 0 0 5】

しかし、紙文書の場合であれば、修正の跡や筆跡やインクの違いなどから改ざんされた場所を特定できるが、特許文献 1 で開示されている手法では、紙文書を電子化して得られるイメージデータの改ざんを検知することは可能であっても、改ざんされている場所まで検知できない。

【0 0 0 6】

また、特許文献 2 で開示されている手法により、ブロック単位（例、8 X 8 ピクセル）で改ざんされた場所を検知可能であるが、テキスト情報の記載されたイメージデータに対して項目単位（例、紙文書が例えば、領収書のような場合において、名前、日付、金額な

10

20

30

40

50

どの項目)で改ざんを検知することはできない。

【特許文献1】特開2002-230202号公報(段落[0026]~[0031]、図1)

【特許文献2】特許3342677「コンテンツデータ鑑定装置」(CA2307534
"CONTENT DATA JUDGING APPARATUS")

【発明の開示】

【発明が解決しようとする課題】

【0007】

本発明では、イメージデータが改ざんされたことを後で検証可能にする電子署名を伴う登録と、そうした電子署名が付与されたイメージデータから項目単位で改ざん場所の特定を可能にする検証を得る点で解決すべき課題がある。 10

【0008】

この発明の目的は、紙文書を電子化して得られるイメージデータについて改ざんがあつたことのみならず、改ざんが行われた場所をも特定することが可能となるイメージデータに対する登録方法及び装置、登録プログラム及びそれを記録した記録媒体、並びにイメージデータの改ざん場所が特定可能な検証方法及び装置、こうした改ざん場所を特定可能な検証プログラム及びそれを記録した記録媒体を提供することである。

【課題を解決するための手段】

【0009】

上記課題を解決するため、本発明によるイメージデータの登録方法は、1項目以上のテキスト部分を含むイメージデータを取得するイメージデータ取得ステップと、前記イメージデータ取得ステップで取得された前記イメージデータから前記テキスト部分のテキスト情報を認識するテキスト情報認識ステップと、前記イメージデータ及び前記テキスト情報認識ステップで認識された1項目以上のテキスト情報のそれぞれについて特徴量を計算する特徴量計算ステップと、前記特徴量計算ステップで計算された前記特徴量に対して電子署名を付与する署名付与ステップと、前記イメージデータ、前記特徴量及び前記署名付与ステップで付与された前記電子署名を記録する記録ステップとを有することから成っている。 20

【0010】

このイメージデータの登録方法によれば、イメージデータ取得ステップにおいて1項目以上のテキスト部分を含むイメージデータが取得され、テキスト情報認識ステップにおいては、上記取得されたイメージデータ内のテキスト部分のテキスト情報が認識される。特徴量計算ステップにおいてイメージデータの特徴量と、テキスト情報認識ステップで認識された1項目以上のテキスト情報について特徴量が計算される。署名付与ステップにおいては、特徴量計算ステップで計算されたイメージデータの特徴量と、テキスト情報について特徴量に対して電子署名が付与される。電子署名は、例えば、イメージデータの管理者のみが知る秘密鍵による暗号化で行われる。記録ステップにおいては、イメージデータ、特徴量及び署名付与ステップで付与された電子署名が記録され、イメージデータの検証に用いることができる。 30

【0011】

このイメージデータの登録方法において、前記認識したテキスト情報を修正するテキスト情報修正ステップを更に有し、前記特徴量計算ステップにおいて、前記テキスト情報修正ステップで修正された前記テキスト情報の特徴量を更に計算し、前記署名付与ステップにおいて、修正された前記テキスト情報について前記特徴量計算ステップで計算された前記特徴量に対して更に電子署名を付与し、前記記録ステップにおいて、前記修正されたテキスト情報の前記特徴量とその電子署名を更に記録することができる。認識したテキスト情報が誤っている場合には、テキスト情報修正ステップにおいてテキスト情報を修正する。特徴量計算ステップにおいてはテキスト情報修正ステップで修正したテキスト情報の特徴量を計算し、署名付与ステップにおいては、修正したテキスト情報の特徴量に対して更に電子署名を付与し、記録ステップにおいて、修正されたテキスト情報の前記特徴量とそ 40
50

の電子署名を更に記録する。

【0012】

上記イメージデータの登録方法において、前記特徴量計算ステップでは、前記イメージデータ若しくは認識又は修正された前記テキスト情報のサイズに応じて、前記特徴量の計算方法を異ならすことが好ましい。データのサイズは、データの内容に応じて非常に変化するので、サイズに応じて計算方法を変更し、同じ程度のサイズの情報量に対して特徴量を計算する方が好ましい。

【0013】

特徴量の計算方法を異ならす上記イメージデータの登録方法において、前記特徴量の計算では、前記イメージデータ若しくは認識又は修正された前記テキスト情報のサイズが、予め決められた値よりも小さい場合には前記イメージデータ又は前記テキスト情報のサイズを前記特徴量とし、前記予め決められた値以上である場合には前記イメージデータ又は前記テキスト情報のサイズが入力される一方向性関数の出力値を前記特徴量とすることができる。テキスト情報のサイズが予め決められた値よりも小さい場合には、そのサイズを特徴量とするが、予め決められた値よりも大きい場合にはテキスト情報のサイズが入力される一方向性関数の出力値を特徴量とする。一般に、イメージデータのサイズはテキスト情報のサイズと比較して非常に大きい。したがって、特徴量の計算方法として、データのサイズが予め決められた値よりも大きい場合には一方向性関数を用いてサイズを出力値として下げる方が好ましい。この場合、一方向性関数としては、例えば、ハッシュ関数が挙げられる。

10

20

【0014】

この発明によるイメージデータの登録プログラムは、上記のイメージデータの登録方法における各ステップをコンピュータに実行させるプログラムである。また、この発明による記録媒体は、上記登録プログラムを記録してあり、コンピュータでの読み取りが可能な記録媒体である。

【0015】

この発明によるイメージデータの登録装置は、1項目以上のテキスト部分が含まれているイメージデータを取得するイメージデータ取得部と、前記イメージデータ取得部で取得された前記イメージデータ内の前記テキスト部分の前記テキスト情報を認識するテキスト情報認識部と、前記イメージデータ及び前記テキスト情報認識部で認識された1項目以上のテキスト情報のそれについて特徴量を計算する特徴量計算部と、前記特徴量計算部で計算された前記特徴量に対して電子署名を付与する署名付与部と、前記イメージデータ、前記特徴量及び前記署名付与部で付与された前記電子署名を記録する記録部とを有することから成る。

30

【0016】

このイメージデータの登録装置によれば、イメージデータ取得部は1項目以上のテキスト部分が含まれているイメージデータを取得し、テキスト情報認識部はイメージデータ取得部で取得されたイメージデータ内のテキスト部分についてテキスト情報を認識し、特徴量計算部はイメージデータの特徴量とテキスト情報認識部が認識したテキスト情報の特徴量とを計算する。署名付与部は、特徴量計算部が計算したイメージデータの特徴量とテキスト情報の特徴量とに対して電子署名を付与する。電子署名は、例えば、イメージデータの管理者のみが知る秘密鍵による暗号化で行われる。記録部は、イメージデータの検証に用いるため、イメージデータ、特徴量及び署名付与部が付与した電子署名を記録する。

40

【0017】

このイメージデータの登録装置は、前記テキスト情報認識部で認識された前記テキスト情報を修正するテキスト情報修正部を更に有し、前記特徴量計算部は、前記テキスト情報修正部で修正された前記テキスト情報の特徴量を更に計算し、前記署名付与部は、修正された前記テキスト情報の前記特徴量に対して更に電子署名を付与し、前記記録部は、修正された前記テキスト情報の前記特徴量とその電子署名を更に記録することができる。即ち、テキスト情報認識部が認識したテキスト情報が誤っている場合には、テキスト情報修正

50

部において人手にてテキスト情報を修正可能とする。修正されたテキスト情報について、特徴量の計算、電子署名及び記録が行われる。詳細には、特徴量計算部はテキスト情報修正部で修正したテキスト情報の特徴量を計算し、署名付与部は修正したテキスト情報の特徴量に対して更に電子署名を付与し、記録部は修正されたテキスト情報の前記特徴量とその電子署名を更に記録する。

【0018】

このイメージデータの登録装置において、前記特徴量計算部は、認識又は修正された前記テキスト情報のサイズに応じて、前記特徴量の計算方法を異にすることができる。データのサイズは、データの内容に応じて非常に変化するので、サイズに応じて計算方法を変更し、同じ程度のサイズの情報量に対して特徴量を計算する方が好ましい。

10

【0019】

このイメージデータの登録装置において、前記特徴量計算部は、前記イメージデータ若しくは認識又は修正された前記テキスト情報のサイズが、予め決められた値よりも小さい場合には前記イメージデータ又は前記テキスト情報のサイズを前記特徴量とし、前記予め決められた値以上である場合には前記イメージデータ又は前記テキスト情報のサイズが入力される一方向性関数の出力値を前記特徴量とすることができる。一般に、イメージデータのサイズはテキスト情報のサイズと比較して非常に大きい。したがって、特徴量の計算方法として、データのサイズが予め決められた値よりも大きい場合には一方向性関数を用いてサイズを出力値として下げる方が好ましい。この場合、一方向性関数としては、例えば、ハッシュ関数が挙げられる。

20

【0020】

この発明によるイメージデータの検証方法は、1項目以上のテキスト部分が含まれているイメージデータと、前記イメージデータ内の前記テキスト部分についてのテキスト情報の特徴量と、当該特徴量に対して付与された電子署名とを取得する取得ステップと、前記取得ステップで取得された前記電子署名を検証する署名検証ステップと、前記イメージデータ内の前記テキスト情報を認識するテキスト情報認識ステップと、前記認識したテキスト情報の特徴量を計算する特徴量計算ステップと、前記取得ステップで取得された前記テキスト情報の前記特徴量と前記テキスト情報認識ステップで認識された前記テキスト情報の前記特徴量とを比較する比較ステップと、前記比較ステップでの比較において前記取得ステップで取得された前記特徴量と前記テキスト情報認識ステップで認識された前記特徴量とが不一致であるとされた前記テキスト情報が前記改ざん場所であると特定する改ざん場所特定ステップとを有することから成っている。

30

【0021】

このイメージデータの検証方法によれば、取得ステップにおいて、1項目以上のテキスト部分が含まれているイメージデータと、前記イメージデータ内の前記テキスト部分についてのテキスト情報の特徴量と、当該特徴量に対して付与された電子署名とが取得される。署名検証ステップにおいては、取得ステップで取得された電子署名が検証される。電子署名が有効なものであると判断されると、テキスト情報認識ステップにおいては、イメージデータ内のテキスト部分についてそのテキスト情報が認識される。特徴量計算ステップにおいては、認識されたテキスト情報の特徴量が計算される。そして、比較ステップにおいては、取得ステップで取得されたテキスト情報の特徴量とテキスト情報認識ステップで認識されたテキスト情報の特徴量とが比較される。比較ステップでの比較において、取得ステップで取得された特徴量は正当な特徴量であるが、テキスト情報認識ステップで認識された特徴量と不一致である場合には、改ざん場所特定ステップはイメージデータが改ざんされたためと推察され、その不一致となった特徴量を持つテキスト情報が改ざん場所であると特定する。

40

【0022】

このイメージデータの検証方法において、前記イメージデータ内の前記テキスト情報の配置を定義したフォーマット定義情報を読み込むフォーマット定義情報読み込みステップを更に有し、前記比較ステップにおいて、前記テキスト情報認識ステップで認識された前記

50

テキスト情報の前記特徴量は、前記フォーマット定義情報読み込みステップで読み込まれた前記フォーマット定義情報から取得された前記イメージデータ内の前記テキスト情報についての前記特徴量と比較することができる。イメージデータが、所定の書式を持つ書類のイメージデータである場合には、書式が定まっていないためにイメージデータ内の領域指定から開始して文字認識を行う場合に比べて、フォーマット定義情報を読み込むことでテキスト情報の認識までの動作を迅速化することができる。

【0023】

上記イメージデータの検証方法において、前記比較ステップでは、前記取得ステップで取得された前記イメージデータ内の前記テキスト情報の前記特徴量と前記テキスト情報認識ステップで認識された前記テキスト情報の前記特徴量とが不一致となるときに、改ざん場所表示ステップでは、前記テキスト情報の場所を前記イメージデータ上で表示することができる。両特徴量が不一致となるテキスト情報の場所をイメージデータ上で表示することで、イメージデータの利用者は、容易にイメージデータの改ざん場所を認識することができる。

【0024】

この発明によるイメージデータの検証プログラムは、上記のイメージデータの検証方法における各ステップをコンピュータに実行させるプログラムである。また、この発明による記録媒体は、上記検証プログラムを記録しており、コンピュータでの読み取りが可能な記録媒体である。

【0025】

この発明によるイメージデータの検証装置は、1項目以上のテキスト部分が記載されているイメージデータと、前記イメージデータ内の前記テキスト部分についてのテキスト情報の特徴量と当該特徴量に対して付与された電子署名とを取得する取得部と、前記取得部で取得された前記電子署名を検証する署名検証部と、前記イメージデータ内の前記テキスト情報を認識するテキスト情報認識部と、前記認識したテキスト情報の特徴量を計算する特徴量計算部と、前記取得部で取得された前記テキスト情報の前記特徴量と前記テキスト情報認識部で認識された前記テキスト情報の前記特徴量とを比較する比較部とを有し、前記比較部での比較において前記取得部で取得された前記特徴量と前記テキスト情報認識部で認識された前記特徴量とが不一致であるとされた前記テキスト情報を前記改ざん場所として特定する改ざん場所特定部とを有することから成っている。

【0026】

このイメージデータの検証装置によれば、取得部は、1項目以上のテキスト部分が記載されているイメージデータと、前記イメージデータ内の前記テキスト部分についてのテキスト情報の特徴量と当該特徴量に対して付与された電子署名とを取得する。署名検証部は、取得部で取得された電子署名を検証する。電子署名が有効であると検証されると、テキスト情報認識部はイメージデータ内のテキスト部分についてテキスト情報を認識する。特徴量計算部は認識したテキスト情報の特徴量を計算し、比較部は、取得部で取得されたテキスト情報の特徴量（正当な特徴量）と、テキスト情報認識部で認識されたテキスト情報の特徴量とを比較する。比較部での比較において、取得部で取得された特徴量とテキスト情報認識部で認識された特徴量とが不一致であるとされると、改ざん場所特定部は、その特徴量が不一致であるテキスト情報がイメージデータにおいて改ざんされた場所であると特定する。

【0027】

このイメージデータの検証装置において、前記イメージデータ内の前記テキスト情報の配置を定義したフォーマット定義情報を読み込むフォーマット定義情報読み込み部を更に有し、前記比較部において、前記テキスト情報認識部で認識された前記テキスト情報の前記特徴量は、前記フォーマット定義情報読み込み部で読み込まれた前記フォーマット定義情報から取得された前記イメージデータ内の前記テキスト情報についての前記特徴量であるとすることができます。イメージデータが、所定の書式を持つ書類のイメージデータである場合には、書式が定まっていないためにイメージデータ内の領域指定から開始して文字認識

10

20

30

40

50

を行う場合に比べて、フォーマット定義情報を読み込むことでテキスト情報の認識までの動作を迅速化することができる。

【0028】

このイメージデータの検証装置において、前記取得部で取得された前記イメージデータ内の前記テキスト情報の前記特徴量と前記テキスト情報認識部で認識された前記テキスト情報の前記特徴量とが前記比較部で不一致とされた前記テキスト情報の場所を前記イメージデータ上で表示する改ざん場所表示部を有することができる。両特徴量が不一致となるテキスト情報の場所をイメージデータ上で表示することで、イメージデータの利用者は、容易にイメージデータの改ざん場所を認識することができる。

【0029】

このように、テキスト部分を含むイメージデータに対してテキスト情報認識（文字認識）処理が行われ、テキスト情報認識結果を含むデータに対して電子署名が付与され、イメージデータ、文字認識結果とイメージデータのそれぞれの特徴量、及び電子署名が保存される。電子署名に対して逆の作用として組み合わされる検証においては、電子署名の検証を行い、電子署名が有効と判定されれば、イメージデータに対して文字認識処理を行い、その文字認識の結果の特徴量を計算し、保存されている文字認識結果の特徴量と比較することで、不一致となる特徴量が存在すればその特徴量を持つテキスト情報に対応したテキスト部分が改ざん箇所として特定される。これにより、従来目視での確認が必要であった改ざん場所の特定作業を自動化することができ、紙書類を電子化した場合における監査業務に必要な時間を大幅に削減できる。

【発明の効果】

【0030】

本発明による、イメージデータの登録方法及び装置、登録プログラム及びそれを記録した記録媒体、並びに署名が付与されたイメージデータが改ざんされた場合にその改ざん場所を特定可能なイメージデータの検証方法及び装置、検証プログラム及びそれを記録した記録媒体においては、イメージデータの改ざんを検知できるだけでなく、改ざんされたテキスト部分の場所を特定することができる。

【発明を実施するための最良の形態】

【0031】

本発明によるイメージデータの登録方法及びイメージデータの改ざん場所特定可能な検証方法の第1の実施例について、図面を参照しつつ説明する。本実施例では、従来のイメージデータのみに対する電子署名方法ではなく、イメージデータと、イメージデータに対して行った文字認識結果との両方に対して電子署名を適用することで、紙書類のイメージデータの改ざん場所を特定することができる方法について説明する。

【0032】

図1は、本実施例におけるイメージデータの改ざん場所を特定可能な検証方法を実行する情報処理システムの概略を示す構成例である。本システムを構成するコンポーネントは、スキヤニング装置1002、登録装置1001、データ保管装置1003、検証装置1004、ネットワーク1005である。本実施例においては、機能毎に装置1001、1003及び1004を分離し、それぞれの装置をネットワーク1005で繋いでいるが、1台の装置に複数の装置の機能を備えてよい。

【0033】

スキヤニング装置1002は、紙の書類をイメージデータに変換する装置である。紙の書類としては、申込書や契約書などの業務によってフォーマットが一定である紙の定形書類である場合や、領収書や請求書などの発行元によりフォーマットの異なる紙の非定形書類がある。

【0034】

登録装置1001、データ保管装置1003、検証装置1004は、それぞれ、CPU、メモリ、記憶装置、表示装置、キーボードを備えたコンピュータである。登録装置1001は、イメージデータ取得1006、文字認識1007、署名生成1008、データ登

10

20

30

40

50

録 1009 の各機能を備える。文字認識 1007 では、スキャニング装置 1002 又は記憶装置からイメージデータを読み込み、イメージデータ内に含まれるテキスト部分について文字認識処理、即ち、テキスト情報認識処理を行う。更に、署名生成 1008 においては、イメージデータ及び文字認識結果に対して署名が付与され、データ保管装置 1003 にイメージデータ、文字認識結果及び署名を登録する。

【0035】

データ保管装置 1003 は、データ受信 1010、署名検証 1011、データ保管 1012、及びデータ提供 1013 の各機能を備える。データ保管装置 1003 は、データ受信 1010 において登録装置 1001 から署名を含むデータを受け取り、署名検証 1011 において署名検証を行う。署名が有効な場合に、データ保管 1012 において記憶装置にデータ（イメージデータ、文字認識結果、署名データ）を保管する。署名が有効でない場合には、データを保管せずにエラーを登録装置 1001 に返信する。なお、文字認識結果はイメージデータの検索用に利用できるが、必要ない場合には保管しなくてもよい。

【0036】

検証装置 1004 は、検証対象データ取得 1014、署名検証 1011、改ざん場所特定 1015、改ざん場所表示 1016 の各機能を備える。検証装置 1004 は、検証対象データ取得 1014 においてデータ保管装置 1003 から検証対象データ（イメージデータ、文字認識結果、署名データ）を取得し、署名検証 1011 において検証対象データのうちの署名について署名検証を行う。検証が無効である場合には、改ざん場所特定 1015 において改ざん場所検知の処理を行う。更に、改ざん場所表示 1016 において、視覚的に判るようにイメージデータの改ざん場所を表示する。

【0037】

図 2 は、登録装置 1001 における文字認識（テキスト情報認識）1007 の処理フローである。本フローが、本発明におけるテキスト情報認識ステップに関連する。

ステップ 2001：開始。

ステップ 2002：イメージデータ取得 1006 で取得したイメージデータが定形書類である場合にはステップ 2003 へ進む。そうでない場合（非定形書類である場合）にはステップ 2005 へ進む。

ステップ 2003：フォーマット定義情報を読み込む。フォーマット定義情報とは、以降のステップにおける文字認識を行う場所（範囲）とその場所に記載されている内容の識別情報及びデータ形式（例：数字、文字など）を定義している情報である。

ステップ 2004：フォーマット定義情報に従い、指定された箇所の文字認識を行う。

ステップ 2005：イメージデータ全体に対して文字認識を行う。文字認識処理は行単位で行い、それぞれの行の認識結果を 1 つの要素として扱う。

ステップ 2006：文字認識結果を表示する。正しく文字を認識できているかをユーザに確認してもらうために行う。なお、誤認識の可能性がない場合には、このステップは省略してもよい。

ステップ 2007：文字認識結果に誤りがある場合にデータを修正する（ユーザ操作による）。

ステップ 2008：文字認識結果データを作成する。

ステップ 2009：終了。

【0038】

図 3 は、ステップ 2003 におけるフォーマット定義情報 3001 のフォーマットの一例を示している。フォーマット定義情報 3001 は、帳票 ID 3002、属性 ID 3003、データ形式 3004 及び読取範囲 3005 より構成される。帳票 ID 3002 は、書類のフォーマットを識別する情報である。なお、帳票 ID 3002 における 0 × 0 0 0 0 は、書類のフォーマットが非定形であることを示す。属性 ID 3003 は、文字認識を行う情報の識別子である。書類には、名前、日付、金額などが記載されており、属性 ID 3003 はそれらを識別するための情報である。データ形式 3004 は、文字列、数字等の各属性の型情報である。読取範囲 3005 は、各属性 ID 3003 の情報が記載されてい

10

20

30

40

50

るイメージデータ上の範囲を表す情報であり、4つの値は左から、始点のX座標、始点のY座標、終点のX座標、終点のY座標を表している。読み取範囲3005は、イメージデータ上において、始点と終点を結ぶ線を対角線とする長方形の内部に対応している。

【0039】

図4は、ステップ2008における文字認識結果データ4001のフォーマットの一例を示す図である。文字認識結果データ4001は、属性ID3003、登録データ4002及び修正前データ4003より構成される。属性ID3003は、定形書類の場合、認識されたデータの内容の識別子であり、それぞれの属性ID3003は、例えば、日付、名前、金額などに関連付けられている。非定形書類の場合は、属性ID3003は行番号とされる。登録データ4002は、文字認識の結果であり、ユーザが修正した場合には修正したデータとなる。修正前データ4003は、ユーザが修正する前のデータであり、ユーザが修正しなかった場合には空欄(データなし)となる。

【0040】

図5は、登録装置1001における署名生成1008での処理フローである。

ステップ5004：イメージデータ5001及び文字認識1007の出力である文字認識結果データ4001の各属性の登録データ5002と修正前データ5003に対してそれぞれ特徴量5005、5006、5007を計算する。ここで、各属性の修正前データ5003の特徴量5007はデータが存在する場合(ユーザが文字認識結果を修正した場合)のみ計算する。特徴量の計算の具体的な内容については、図7を参照して後述する。

ステップ5008：イメージデータの特徴量5005と文字認識結果の各属性ID3003と特徴量5006、5007とを結合したデータに対して、署名値を計算する。署名値は、結合データに対して秘密鍵を適用することによって暗号化された値である。

ステップ5009：特徴量5005、5006、5007、属性ID3003及び署名値から署名データ5010を作成する。

当フローの5007までが特徴量計算ステップに相当し、それ以後が本発明における署名付与ステップに相当する。

【0041】

図6は、署名データ5010のフォーマットを示している。署名データ5010は、公開鍵証明書6001、署名対象データ6002及び署名値6003より構成される。公開鍵証明書6001は、認証局(公的な性格を帯びるものから、民間の団体まである)が発行している公開鍵の証明書であり、上記秘密鍵に対応し且つ署名データ5010を検証するための公開鍵を含み、且つ公開鍵の持ち主の証明を果たしている。署名対象データ6002は、イメージデータの特徴量5005と、各属性ID3003(属性ID1～属性IDN)についてのそれぞれの認識結果6004とを含んでいる。各属性ID3003の認識結果6004は、各属性ID_i(i=1～N)とその登録データ4002の特徴量5006を含む。更に、ユーザが文字認識結果を修正した場合には、各属性ID3003の認識結果6004には修正前データ4003の特徴量5007が含まれる。その他、時刻情報など任意のデータを署名対象データ6002に含めてよい。ユーザによる文字認識結果の修正は、本発明におけるテキスト情報修正ステップに対応している。

【0042】

また、署名データ5010に含まれるデータには、図5に示したデータ以外を含めてもよい。さらに、公開鍵証明書6001、署名対象データ6002、署名値6003は関連付けがされていれば、分離された状態で保管してもよい。署名データ5010のフォーマットとしては、例えば、署名の標準フォーマットとして利用されているRFC3369“Cryptographic Message Syntax”で定義されているSig ned Data構造体を利用することができます。

【0043】

図7は、ステップ5004における特徴量の計算方法の詳細を示すフローチャートである。

ステップ7001：開始。

10

20

30

40

50

ステップ7002：入力データ(*in*)のデータサイズ(*s*)を計算する。

ステップ7003：データサイズ(*s*)が予め決められた値である閾値(*L*)より小さい場合にはステップ7004へ進む。そうでない場合には、ステップ7005へ進む。即ち、イメージデータ若しくは認識又は修正されたテキスト情報のサイズ(入力データのサイズ)に応じて、特徴量の計算方法が異なっている。

ステップ7004：特徴量を入力データ(*in*)とする。即ち、入力データのサイズが閾値(*L*)より小さい場合には、その入力データのサイズが特徴量とされる。

ステップ7005：特徴量 = $h(in)$ として特徴量を計算する。関数 $h()$ としては、一例として、一方向性関数としてのハッシュ関数を利用してもよい。

ステップ7006：終了。

10

【0044】

図8はデータ登録1009において、登録装置1001がデータ保管装置1003に送信するデータ8001のフォーマットを示している。送信データ8001は帳票ID3002、イメージデータ5001、文字認識結果データ4001、署名データ5010より構成される。その他、登録者を識別する情報や登録日時等、図8に示したデータ以外のデータを送信データ8001に含めてよい。データ保管装置1003は、登録装置1001が送信したデータを受信し、署名の検証を行い、署名が有効である場合に受信したデータを保管する。データ登録1009は、本発明における記録ステップに対応している。

【0045】

図9は検証装置1004における署名検証1011の処理手順を示している。

20

ステップ9001：開始。

ステップ9002：署名データ5010の中に含まれる公開鍵証明書6001を検証する。

ステップ9003：公開鍵証明書6001が有効である場合にはステップ9004へ進む。そうでない場合にはステップ9010へ進む。

ステップ9004：公開鍵証明書6001から公開鍵を取り出す。

ステップ9005：イメージデータ5001(改ざんの可能性がある)の特徴量を計算し、署名対象データ6002に含まれる特徴量5005と比較する。

ステップ9006：ステップ9005において特徴量が一致する場合にはステップ9007へ進む。そうでない場合にはステップ9010へ進む。

30

ステップ9007：公開鍵、署名対象データ6002、署名値6003から署名を検証する。即ち、取り出した公開鍵で署名値6003を解読することによって、元のイメージデータの特徴量5005と文字認識結果の各属性ID3003と特徴量5006、5007とを結合したデータが復号化される。復号化された特徴量5005、文字認識結果の各属性ID3003及びそれらの特徴量5006、5007と、署名対象データ6002に含まれているイメージデータ5001の特徴量5005、各属性ID3003及びそれらの特徴量5006、5007とを比較することで、署名が有効か否かを検証することができる(一致していれば、署名は有効と判断される)。

ステップ9008：署名が有効である場合にはステップ9009へ進む。そうでない場合にはステップ9010へ進む。

40

ステップ9009：(正常)終了。

ステップ9010：ステップ9003において公開鍵証明書6001が有効でない場合及びステップ9006において特徴量が一致しない場合には、エラー終了とする。

【0046】

データ保管装置1003の署名検証1011において、エラー終了(ステップ9008)の場合には、受信したデータを破棄し、登録装置1001にエラーである旨を返送する。正常終了の場合には、受信したデータをデータ保管1012の処理によりデータ保管装置1003内に記録する。また、データ保管装置1003は、データ提供機能1013の機能を備える。データ提供機能1013は、データ保管装置1003に保管したデータを他の計算機からの要求に対して、データを送信する機能である。これにより、他の計算機

50

(例、検証装置 1004) は、データ保管装置 1003 に保管されているデータを参照できる。データ提供機能 1013 は、ユーザ認証機能やアクセス制御機能を備えることで、参照可能なユーザを制限してもよい。

【0047】

検証装置 1004 は、検証対象データ取得 1014 を利用して、データ保管装置 1003 から検証対象データ(イメージデータ 5001、文字認識結果データ 4001、署名データ 5010)を取得する。検証装置 1004 は、取得したデータから署名検証機能 1011、改ざん場所特定 1015 を実行する。署名検証 1011 は、図 9 に示すフローチャートと同じデータベース順で実行される。署名検証 1011においてエラーとなった場合には、改ざん場所特定機能 1015 が実行される。更に、改ざん場所特定 1015において、改ざんリストに 1 つ以上の値が存在する場合には、その結果をもとに改ざん場所表示 1016 を実行し、改ざん場所表示を行う。10

【0048】

図 10 は、改ざん場所特定 1015 の処理手順を示している。

ステップ 10001：開始。

ステップ 10002：イメージデータ 5001 に対して、文字認識処理を行う。なお、ここでの文字認識処理は、図 2 に示す文字認識処理の処理手順のうち、ステップ 2006(文字認識結果の表示)、2007(ユーザーによる認識結果の修正)の処理を省略した処理である。更に、ステップ 2008 で作成する文字認識結果データ 4001 は全ての修正前データ 4004 が空欄となっている。20

ステップ 10003：認識結果の特徴量を計算する。

ステップ 10004：内部変数 $i = 0$ に設定する。

ステップ 10005：属性 ID i の認識結果の特徴量と署名対象データ 6002 内の登録データの特徴量 5006 が一致しているか確認する。一致している場合にはステップ 10006 へ進む。そうでない場合にはステップ 10008 へ進む。

ステップ 10006：内部変数 $i = i + 1$ とする。

ステップ 10007： $i > N$ の場合にはステップ 10010 へ進む。そうでない場合にはステップ 10005 へ戻る。ここで、 N は認識結果の数であり、文字認識結果データ 4001 の属性 ID の数と一致する。

ステップ 10008：属性 ID i の認識結果の特徴量と署名対象データ 6002 内の修正前データの特徴量 5007 を比較する。30

ステップ 10009：比較結果を不一致リストに追加する。なお、修正前データの特徴量 5007 が存在しない場合には不一致とする。

ステップ 10010：終了。

【0049】

図 11 は、改ざん場所特定 1015 で作成する不一致リスト 11001 のフォーマットの一例を示している。不一致リスト 11001 は、属性 ID 3003、認識結果 11002、比較結果 11003 を含む。認識結果 11002 は、ステップ 10002 の文字認識結果である。比較結果 11003 は、修正前データの特徴量 5007 と認識結果 11002 の特徴量を比較して一致する場合には「1」、一致しない場合には「0」を指定する。なお、修正前データがない場合には「0」となる。40

【0050】

図 12 は、改ざん場所表示 1016 の処理手順の一例を示すフローチャートである。

ステップ 12001：開始。

ステップ 12002：フォーマット定義情報 3001 を読み込む。

ステップ 12003：不一致リスト 11001 からデータを 1 つ取り出す。なお、データは、図 11 に示す不一致リスト 11001 の行単位で取り出す。取り出したデータは不一致リスト 11001 から削除する。

ステップ 12004：フォーマット定義情報 3001 と属性 ID から当該属性の情報が記載されているイメージデータ上の位置を特定し、マークする。マークは、例えば、当該50

箇所を枠で囲うなど、視覚的にわかるようにする。また、比較結果 11003 の違いにより、色を分けても良い。比較結果 11003 が「1」の場合には文字認識処理が誤っている可能性があり、ユーザに確認を求めるようとする。また、「0」の場合には改ざんされている可能性が非常に高いので、改ざんされている旨を画面に表示する。

ステップ 12005：不一致リスト 11001 にデータが残っている場合にはステップ 12003 へ進む。そうでない場合にはステップ 12006 へ進む。

ステップ 12006：終了。

【0051】

以上の手順により、イメージデータに対して文字認識処理を施して電子署名を付与し、検証時に再度文字認識処理を行って結果を比較することでイメージデータ上の改ざん場所を特定することができる。改ざん場所を特定の機能を付加するために、データ量が従来と比べて増加するが、その増加量はイメージデータ量と比較して僅かであるので、特段の問題はない。

【0052】

また、本実施例を応用すれば、改ざん場所を特定できるだけでなく、データを復元することもできる。各属性（氏名、日付、金額など）はランダムな値ではなく、「氏名」であれば、過去に登録した「氏名」と一致する可能性もあり、「日付」「金額」などはある程度範囲を絞ることができる。「氏名」であれば、データベースに登録されている全「氏名」を取り出し、特徴量をそれぞれ計算して比較し、署名データ 5010 内の特徴量と一致すれば、それが正しいデータとなる。また、「日付」「金額」についても同様であり、範囲を絞ってそれぞれの特徴量を計算して比較することでデータを復元することができる。

【0053】

図 13 には、図 1 に示すイメージデータの登録・保管・検証システムにおける登録装置の概略の一例が示されている。イメージデータの登録装置 1 は、1 項目以上のテキスト部分が含まれているイメージデータを取得するイメージデータ取得部 2 と、イメージデータ取得部 2 で取得されたイメージデータ内のテキスト部分についてのテキスト情報を認識するテキスト情報認識部 3 と、イメージデータ取得部 2 で取得されたイメージデータ及びテキスト情報認識部 3 で認識されたテキスト情報のそれぞれについて特徴量を計算する特徴量計算部 4 と、特徴量計算部 4 で計算された特徴量に対して電子署名を付与する署名付与部 5 とを有している。テキスト情報認識部 3 で認識されたテキスト情報に誤りがあれば、テキスト情報修正部 6 で修正され、修正されたテキスト情報が特徴量計算部 4 に送られる。イメージデータ取得部 2 で取得されたイメージデータ、特徴量計算部 4 で計算された特徴量、及び署名付与部 5 で付与された電子署名は、記録部 7 で記録される。

【0054】

図 14 には、図 1 に示すイメージデータの登録・保管・検証システムにおける検証装置の概略の一例が示されている。図 14 に示すイメージデータの検証装置 10 は、1 項目以上のテキスト部分が含まれているイメージデータと、そのイメージデータ内のテキスト部分についてのテキスト情報の特徴量と、当該特徴量に対して付与された電子署名とを取得する取得部 11 を備えている。検証装置 10 は、また、取得部 11 で取得された電子署名を検証する署名検証部 12 と、署名検証部 12 で電子署名が有効とされた場合に、取得部 11 で取得されているイメージデータ内のテキスト部分についてのテキスト情報を認識するテキスト情報認識部 13 と、テキスト情報認識部 13 で認識したテキスト情報の特徴量を計算する特徴量計算部 14 とを更に有している。検証装置 10 は、更に、取得部 13 で取得されているテキスト情報の特徴量と、テキスト情報認識部 13 で認識されたテキスト情報の特徴量とを比較する比較部 15 を備えている。紙書類が定型用紙に記載されているものである場合には、その定型用紙のフォーマット定義情報読み込み部 16 からフォーマット定義情報が読み込まれて、比較部 15 に特徴量の情報が与えられる。検証装置 10 は、更にまた、比較部 15 での比較において取得部 11 で取得されたテキスト情報の特徴量と、対応するテキスト部分についてテキスト情報認識部 13 で認識されたテキスト情報の特徴量とが不一致であるとき、そのテキスト情報を改ざん場所として特定する改ざん場所特

10

20

30

40

50

定部16を有し、更に、改ざん場所特定部16で特定された改ざん場所をイメージデータ上で表示させる改ざん場所表示部17を備えている。

【0055】

上記したようなイメージデータ、イメージデータ内のテキスト情報の特徴量及び当該特徴量に対して付与された電子署名とを取得する取得ステップと、電子署名を検証する署名検証ステップと、テキスト情報を認識するテキスト情報認識ステップと、テキスト情報の特徴量を計算する特徴量計算ステップと、両特徴量を比較する比較ステップと、両特徴量が不一致であることに応じた改ざん場所特定ステップとを有するイメージデータの検証方法については、各ステップをコンピュータに実行させることができるイメージデータの検証プログラムとして具体化することができる。10

【0056】

こうしたプログラムは、CD-ROMやDVD-ROMのような記録媒体に記録させた状態で市場に流通させることができる。また、プログラムデータとして、インターネット等の公衆回線を通じてダウンロード可能な状態にして、ユーザが購入することができるようにもよい。上記イメージデータの検証方法のプログラムには、イメージデータ内のテキスト情報の配置を定義したフォーマット定義情報を読み込むフォーマット定義情報読み込みステップを有するもの、また、両特徴量が不一致となるときにテキスト情報の場所をイメージデータ上で表示する改ざん場所表示ステップを更に有するもの等が含まれることは明らかである。

【図面の簡単な説明】

20

【0057】

【図1】本発明の実施例の構成を示す図。

【図2】文字認識処理のフローチャートを示す図。

【図3】文字認識処理におけるフォーマット定義情報を示す図。

【図4】文字認識結果データを示す図。

【図5】署名生成処理のフローチャートを示す図。

【図6】署名データのフォーマットを示す図。

【図7】特徴量の計算方法のフローチャートを示す図。

【図8】登録データのフォーマットを示す図。

【図9】署名検証処理のフローチャートを示す図。

30

【図10】改ざん場所特定のフローチャートを示す図。

【図11】不一致リストのフォーマットを示す図。

【図12】改ざん場所表示のフローチャートを示す図。

【図13】図1に示すイメージデータの登録・保管・検証システムにおける登録装置の概略の一例を示す図。

【図14】図1に示すイメージデータの登録・保管・検証システムにおける検証装置の概略の一例を示す図。

【符号の説明】

【0058】

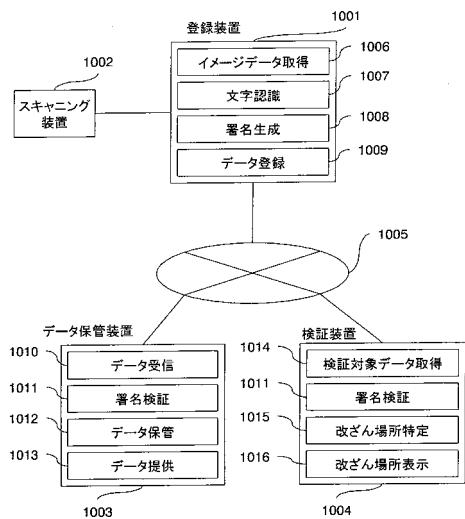
40

- 1 イメージデータの登録装置
- 2 イメージデータ取得部
- 3 テキスト情報認識部
- 4 特徴量計算部
- 5 署名付与部
- 6 テキスト情報修正部
- 7 記録部
- 10 イメージデータの検証装置
- 11 取得部
- 12 署名検証部
- 13 テキスト情報認識部

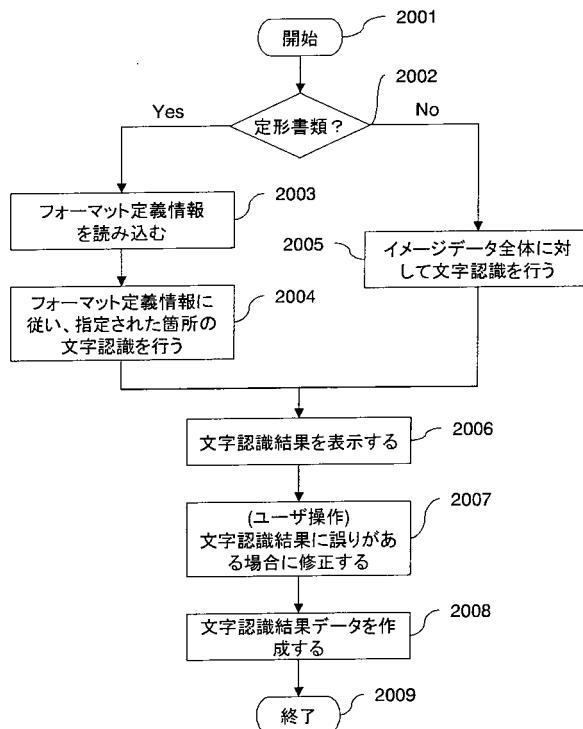
50

1 4	特徴量計算部	
1 5	比較部	
1 6	フォーマット定義情報読み込み部	
1 7	改ざん場所特定部	
1 8	改ざん場所表示部	
1 0 0 1	登録装置	
1 0 0 2	スキヤニング装置	
1 0 0 3	データ保管装置	
1 0 0 4	検証装置	
1 0 0 5	ネットワーク	10
1 0 0 6	イメージデータ取得	
1 0 0 7	文字認識	
1 0 0 8	署名生成	
1 0 0 9	データ登録	
1 0 1 0	データ受信	
1 0 1 1	署名検証	
1 0 1 2	データ保管	
1 0 1 3	データ提供	
1 0 1 4	検証対象データ取得	
1 0 1 5	改ざん場所特定	20
1 0 1 6	改ざん場所表示	
3 0 0 1	フォーマット定義情報	
3 0 0 2	帳票ID	
3 0 0 3	属性ID	
3 0 0 4	データ形式	
3 0 0 5	読み取り範囲	
4 0 0 1	文字認識結果データ	
4 0 0 2	登録データ	
4 0 0 3	修正前データ	
5 0 0 1	イメージデータ	30
5 0 0 5	イメージデータの特徴量	
5 0 0 6	各属性の登録データの特徴量	
5 0 0 7	各属性の修正前データの特徴量	
5 0 1 0	署名データ	
6 0 0 1	公開鍵証明書	
6 0 0 2	署名対象データ	
6 0 0 3	署名値	
8 0 0 1	登録装置からデータ保管装置への送信データ	
1 1 0 0 1	不一致リスト	
1 1 0 0 2	認識結果	40
1 1 0 0 3	比較結果	

【図1】



【図2】



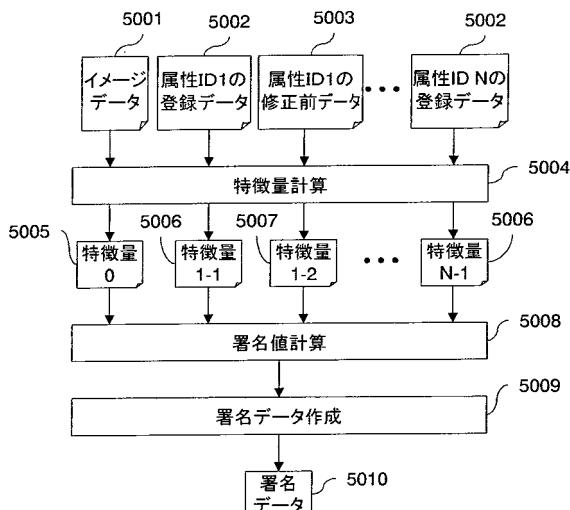
【図3】

3001	3002	3003	3004	3005
帳票ID	属性ID	データ形式	読み取範囲	
0x0000	0	—	(0,0,0,0)	
0x00a1	1	文字列	(x11,y11,x21,y21)	
	2	数字	(x12,y12,x22,y22)	
	
	N	文字列	(x1N,y1N,x2N,y2N)	
0x032a	1	数字	(x11,y11,x21,y21)	
	
	

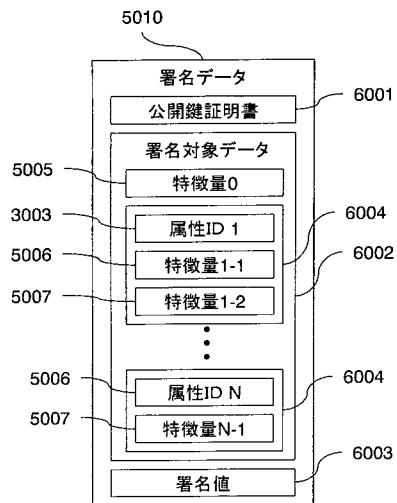
【図4】

4001	3003	4002	4003
属性ID	登録データ	修正前データ	
1	山田太郎	山田大郎	
2	1000	—	
...	
N	2004/7/1	—	

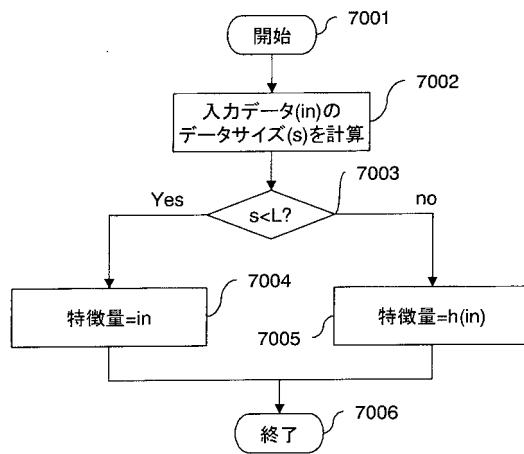
【図5】



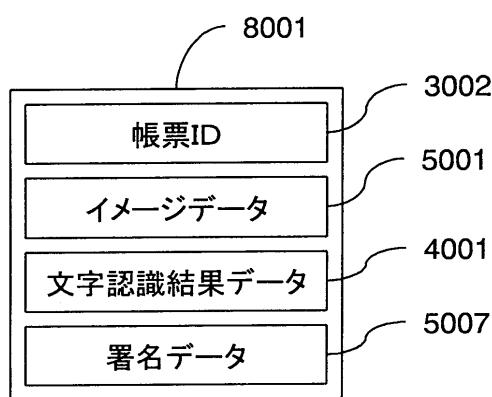
【図6】



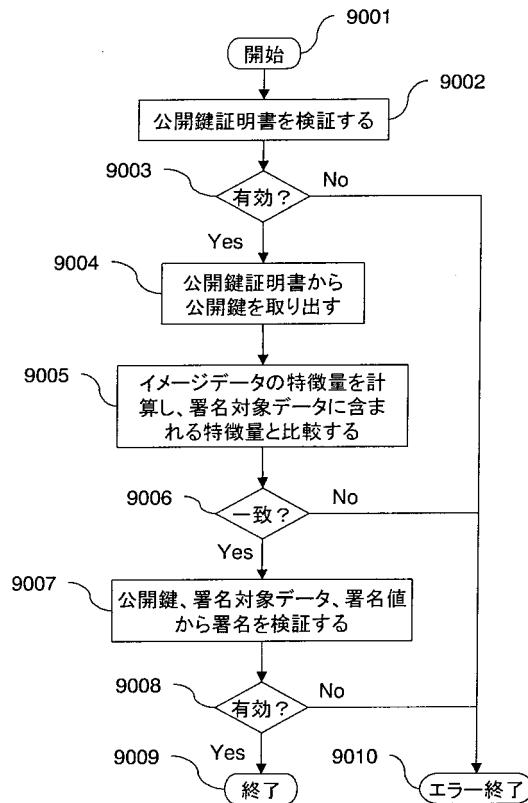
【図7】



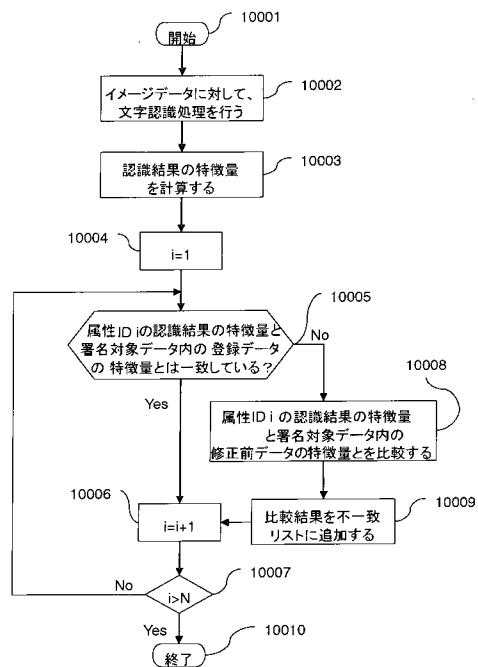
【図8】



【図9】



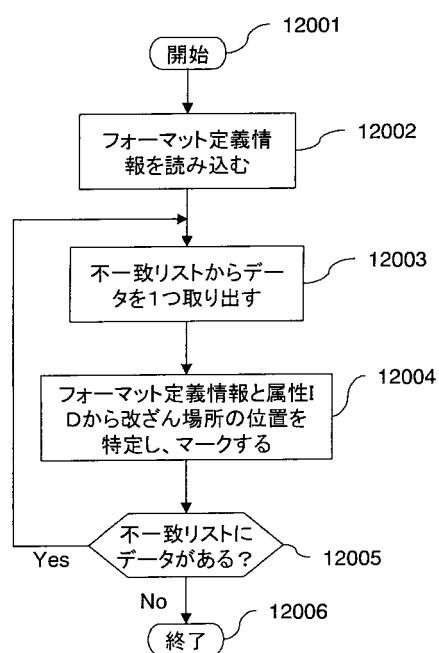
【図10】



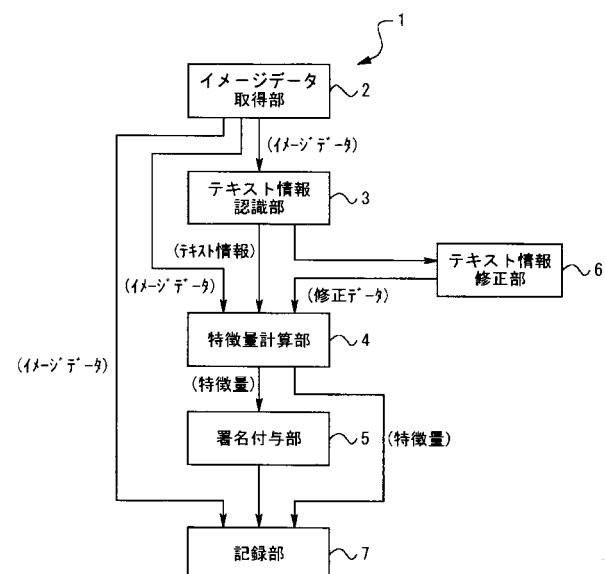
【図11】

属性ID	認識結果	比較結果
1	山田大郎	1
...

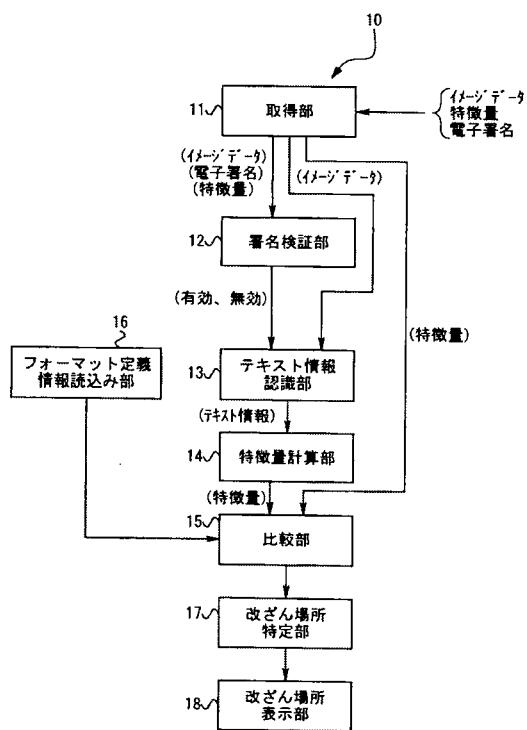
【図12】



【図13】



【図14】



フロントページの続き

(72)発明者 谷川 嘉伸
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72)発明者 酒井 瑞洋
神奈川県川崎市鹿島田890番地 株式会社日立製作所金融第二事業部内

F ターム(参考) 5B064 AA01 AA07 AB02 BA01
5J104 AA08 AA09 LA03 LA05 LA06 PA14