

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】平成19年1月18日(2007.1.18)

【公開番号】特開2002-170079(P2002-170079A)

【公開日】平成14年6月14日(2002.6.14)

【出願番号】特願2000-367675(P2000-367675)

【国際特許分類】

G 06 K 9/20 (2006.01)

G 06 Q 10/00 (2006.01)

【F I】

G 06 K 9/20 3 4 0 C

G 06 F 19/00 3 0 0 C

【手続補正書】

【提出日】平成18年11月27日(2006.11.27)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】文書画像の特徴量に基づいて、文書書式識別のための文書書式データを作成する作成手段と、

前記文書書式データを保存する保存手段と、

文書書式を識別すべき文書の画像について前記作成手段を用いて文書書式データを得て、これを前記保存手段で保存した文書書式データと比較し、相似関係があるか否かを判断する判断手段と、

前記判断手段にて相似関係があると判断された場合に、前記識別すべき文書と、前記保存手段で保存した文書との相似状態を表す相似情報を抽出する相似情報抽出手段と、

前記相似情報抽出手段にて抽出した相似情報及び前記文書書式データに基づいて文書書式の類似度を計算し、前記識別すべき文書の文書書式を識別する識別手段と

を備えることを特徴とする文書書式識別装置。

【請求項2】前記識別手段は、前記相似情報に基づいて、前記類似度の計算に用いる文書書式データを補正し、類似度の計算を行うことを特徴とする請求項1記載の文書書式識別装置。

【請求項3】前記相似情報は、前記識別すべき文書と前記保存手段で保存した文書との変倍率を含み、前記識別手段は、該変倍率を類似度の計算に用いる文書書式データに乗算することを特徴とする請求項2記載の文書書式識別装置。

【請求項4】前記相似情報は、前記識別すべき文書と前記保存手段で保存した文書との原点ずれ量を含み、前記識別手段は、該原点ずれ量を類似度の計算に用いる文書書式データに加算することを特徴とする請求項3記載の文書書式識別装置。

【請求項5】前記相似情報は、前記変倍率に基づいて算出したペナルティを含み、前記識別手段は、該ペナルティを類似度の計算に加算することを特徴とする請求項2記載の文書書式識別装置。

【請求項6】前記相似情報抽出手段は、

前記識別すべき文書を構成する構成要素と前記保存手段で保存した文書を構成する構成要素とが互いに対応するように配列する配列手段と、

前記配列手段によって配列した前記構成要素の個数が互いに等しく、かつ予め定められた数以上の場合に、前記各構成要素の各文書上の特定位置座標に基づいて相関係数を算出

し、該相関係数によって得られる回帰直線に基づいて文書画像の変倍率をもとめる変倍率算出手段と、

前記変倍率算出手段によりもとめた変倍率から類似度の計算に加算するペナルティを算出するペナルティ算出手段と

を備えることを特徴とする請求項2記載の文書書式識別装置。

【請求項7】前記相似情報抽出手段は、前記変倍率算出手段によりもとめた変倍率の適正を判定する判定手段を更に備え、

前記判定手段は、前記判定手段にて変倍率が不適正であると判定した場合、前記相似情報を用いずに前記類似度計算を行うことを特徴とする請求項6記載の文書書式識別装置。

【請求項8】前記相似情報抽出手段は、

前記変倍率算出手段によって算出された前記相関係数に基づく前記回帰直線のY軸切片を前記各文書上の原点位置のずれ量とするずれ量算出手段を更に備えることを特徴とする請求項6記載の文書書式識別装置。

【請求項9】文書書式識別装置における文書書式識別方法であって、

作成手段が、文書画像の特徴量に基づいて、文書書式識別のための文書書式データを作成する作成工程と、

保存手段が、前記文書書式データを保存する保存工程と、

判断手段が、文書書式を識別すべき文書の画像について前記作成工程を用いて文書書式データを得て、これを前記保存工程で保存した文書書式データと比較し、相似関係があるか否かを判断する判断工程と、

相似情報抽出手段が、前記判断工程にて相似と判断された場合に、前記識別すべき文書と、前記保存工程で保存した文書との相似状態を表す相似情報を抽出する相似情報抽出工程と、

識別手段が、前記相似情報抽出工程にて抽出した相似情報及び前記文書書式データに基づいて文書書式の類似度を計算し、前記識別すべき文書の文書書式を識別する識別工程とを備えることを特徴とする文書書式識別方法。

【請求項10】前記識別工程は、前記相似情報に基づいて、前記類似度の計算に用いる文書書式データを補正し、類似度の計算を行うことを特徴とする請求項9記載の文書書式識別方法。

【請求項11】前記相似情報は、前記識別すべき文書と前記保存工程で保存した文書との変倍率を含み、前記識別工程は、該変倍率を類似度の計算に用いる文書書式データに乗算することを特徴とする請求項10記載の文書書式識別方法。

【請求項12】前記相似情報は、前記識別すべき文書と前記保存工程で保存した文書との原点ずれ量を含み、前記識別工程は、該原点ずれ量を類似度の計算に用いる文書書式データに加算することを特徴とする請求項11記載の文書書式識別方法。

【請求項13】前記相似情報は、前記変倍率に基づいて算出したペナルティを含み、前記識別工程は、該ペナルティを類似度の計算に加算することを特徴とする請求項10記載の文書書式識別方法。

【請求項14】前記相似情報抽出工程は、

配列手段が、前記識別すべき文書を構成する構成要素と前記保存工程で保存した文書を構成する構成要素とが互いに対応するように配列する配列工程と、

変倍率算出手段が、前記配列工程によって配列した前記構成要素の個数が互いに等しく、かつ予め定められた数以上の場合に、前記各構成要素の各文書上の特定位置座標に基づいて相関係数を算出し、該相関係数によって得られる回帰直線に基づいて文書画像の変倍率をもとめる変倍率算出手段と、

ペナルティ算出手段が、前記変倍率算出手工程によりもとめた変倍率から類似度の計算に加算するペナルティを算出するペナルティ算出手工程と

を備えることを特徴とする請求項10記載の文書書式識別方法。

【請求項15】前記相似情報抽出工程は、判定手段が、前記変倍率算出手工程によりもとめた変倍率の適正を判定する判定工程を更に備え、

前記判別工程は、前記判定工程にて変倍率が不適正であると判定した場合、前記相似情報用いずに前記類似度計算を行うことを特徴とする請求項14記載の文書書式識別方法。

【請求項16】 前記相似情報抽出工程は、

ずれ量算出手段が、前記変倍率算出工程によって算出された前記相関係数に基づく前記回帰直線のY軸切片を前記各文書上の原点位置のずれ量とするずれ量算出工程を更に備えることを特徴とする請求項14記載の文書書式識別方法。

【請求項17】 請求項9乃至16のいずれか1つに記載の文書書式識別方法をコンピュータによって実行させるための制御プログラムを格納する記憶媒体。