

【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25年8月22日(2013.8.22)

【公開番号】特開2011-18338(P2011-18338A)

【公開日】平成23年1月27日(2011.1.27)

【年通号数】公開・登録公報2011-004

【出願番号】特願2010-156620(P2010-156620)

【国際特許分類】

G 06 T 7/40 (2006.01)

【F I】

G 06 T 7/40 100Z

【手続補正書】

【提出日】平成25年7月4日(2013.7.4)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

文書のページの画像上のマーキングをマーキング種類に従って分類するための方法であつて、

前記文書のページの画像を区分器に供給するステップと、

前記区分器により、連結成分を見つけ、横方向に近い罫線及び縦方向に近い罫線を削除し、いくつかの連結成分を更に分割することで断片を求めることにより、受け取られた前記画像を、同じマーキング種類である可能性が高いと識別される前景画素構造の複数の断片へと分割するステップと、

各画素が既知のマーキング種類に従ってラベル付けされているグランドトゥルース画像群により学習したものである分類器に対して前記複数の断片を供給するステップであつて、前記分類器はそれら各断片についてのカテゴリスコアを求めるステップと、

前記断片が前記分類器により分類されると、その断片に含まれるすべての画素に同じラベルを割り当てるステップと、

を含む方法。

【請求項2】

前記分割するステップは、形態学的演算により検出され除去される横及び縦の線を区分する第1段階と、サイズに関するテストに合格するまで連結成分を再帰的に分割する第2段階と、を含む、請求項1に記載の方法。

【請求項3】

前記分割するステップでの分割は、動的計画法に基づき、分割すべき断片を横切るコスト最小の分割経路を見つけることにより行われる、請求項2に記載の方法。

【請求項4】

前記分割経路のコストは、当該経路に沿った画素のコストの総和である、請求項3に記載の方法。

【請求項5】

前記断片は、単語よりも小さいか、又は、個々の文字又は数字よりも小さい、請求項1に記載の方法。

【請求項6】

予備的なカテゴリ割り当てが、近傍の断片群に対してなされたカテゴリ割り当ての統計

に基づいて更新される、請求項 1 に記載の方法。

【請求項 7】

前記分類器のパラメータは、多くの測定値の集まりの中から有益な特徴を選択しそれら選択した特徴に相対的な重みを割り当てる判別的AdaBoost学習により取得される、請求項1 に記載の方法。

【請求項 8】

前記分類は、印刷されたテキストの断片を検出するためのものであり、証拠の拠り所の1つは、断片の頂上点群と底点群との間に見られる整列度合い及びその断片の隣の断片群の整列度合いである、請求項 1 に記載の方法。

【請求項 9】

文書内の画像上のマーキングをマーキング種類に従って分類するためのシステムであって、

プロセッサ上で動作する共に前記画像を含んだ前記文書を受け取るように構成された区分器であって、連結成分を見つけ、横方向に近い罫線及び縦方向に近い罫線を削除し、いくつかの連結成分を更に分割することで断片を求めるこにより、受け取られた前記画像を、同じマーキング種類である可能性が高いと識別される前景画素構造の複数の断片へと分割する区分器と、

前記プロセッサ上で動作する共に前記断片群を受け取るように構成され、各断片についてのカategoriscoreを求める分類器と、

を備え、前記分類器は、各画素が既知のマーキング種類に従ってラベル付けされているグランドトゥルース画像群により学習したものであるとともに、前記分類器は、前記断片が前記分類器により分類されると、その断片に含まれるすべての画素に同じラベルを割り当てる、システム。

【請求項 10】

文書内の画像上のマーキングをマーキング種類に従って分類するための方法であって、前記画像を含んだ前記文書を区分器に供給するステップと、

前記区分器により、連結成分を見つけ、横方向に近い罫線及び縦方向に近い罫線を削除し、いくつかの連結成分を更に分割することで断片を求めるこにより、受け取られた前記画像を、同じマーキング種類である可能性が高いと識別される前景画素構造の複数の断片へと分割するステップであって、形態学的演算により検出され除去される横及び縦の線を区分する第1段階と、サイズに関するテストに合格するまで連結成分を再帰的に分割する第2段階と、を含むステップと、

前記断片群を2段階分類器に供給し、前記2段階分類器により各断片にカategoriscoreを付与するステップであって、前記2段階分類器は、各画素が既知のマーキング種類に従ってラベル付けされているグランドトゥルース画像群により学習したものであり、前記2段階分類器のうちの第1段階の分類器は、1カategoriあたり1つの一対全部分類器からなる組を含んでおり、+1から-1までの間のカategoriscoreからなる配列を生成し、前記2段階分類器のうちの第2段階の分類器は、前記第1段階の分類器からのカategoriscoreの配列を受け取ってすべてのマーキング種類のカategoriのスコアの通知を受けることで、精密化したスコアの配列を生成するところのステップと、

前記断片が前記分類器によりあるマーキング種類に分類されると、その断片に含まれるすべての画素に同じラベルを割り当てるステップと、

を含む方法。