

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5796107号
(P5796107)

(45) 発行日 平成27年10月21日(2015.10.21)

(24) 登録日 平成27年8月21日(2015.8.21)

(51) Int.Cl. F I
G06T 7/40 (2006.01) G06T 7/40 I00A

請求項の数 23 外国語出願 (全 29 頁)

(21) 出願番号	特願2014-103652 (P2014-103652)	(73) 特許権者	000001007 キヤノン株式会社 東京都大田区下丸子3丁目30番2号
(22) 出願日	平成26年5月19日 (2014.5.19)	(74) 代理人	100076428 弁理士 大塚 康德
(65) 公開番号	特開2014-229314 (P2014-229314A)	(74) 代理人	100112508 弁理士 高柳 司郎
(43) 公開日	平成26年12月8日 (2014.12.8)	(74) 代理人	100115071 弁理士 大塚 康弘
審査請求日	平成26年7月16日 (2014.7.16)	(74) 代理人	100116894 弁理士 木村 秀二
(31) 優先権主張番号	201310196300.7	(74) 代理人	100130409 弁理士 下山 治
(32) 優先日	平成25年5月24日 (2013.5.24)	(74) 代理人	100134175 弁理士 永川 行光
(33) 優先権主張国	中国 (CN)		
(31) 優先権主張番号	201310196315.3		
(32) 優先日	平成25年5月24日 (2013.5.24)		
(33) 優先権主張国	中国 (CN)		

最終頁に続く

(54) 【発明の名称】 テキスト検出の方法及び装置

(57) 【特許請求の範囲】

【請求項1】

少なくとも1つの連結成分を含む画像においてテキスト領域を検出するテキスト検出方法であって、

前記画像からテキスト背景領域を検出する検出工程と、

前記少なくとも1つの連結成分をフィルタリングしてテキスト候補となる連結成分を残すフィルタリング工程と、

前記検出工程において検出された前記テキスト背景領域に基づいて、前記テキスト候補となる連結成分をグループ化して少なくとも1つの連結成分グループを形成し、且つ前記少なくとも1つの連結成分グループに基づいて少なくとも1つのテキスト領域を生成するグループ化工程と、

を備え、

前記フィルタリング工程において、前記テキスト背景領域の境界内にはない連結成分に対するフィルタリングは、前記テキスト背景領域の境界内の連結成分に対するフィルタリングより厳しいことを特徴とするテキスト検出方法。

【請求項2】

前記テキスト背景領域は前記画像中のテキストの周囲領域であって、規則的な境界線、及び、均一な色又はグレースケールを有することを特徴とする請求項1に記載の方法。

【請求項3】

前記検出工程は、

前記少なくとも1つの連結成分の特徴、
 前記少なくとも1つの連結成分の境界内に配置された連結成分であり、前記少なくとも1つの連結成分に対して高いコントラストを有する前記少なくとも1つの連結成分中のメンバ連結成分の統計、及び

前記少なくとも1つの連結成分と他のテキスト背景領域との間の関係、
 のうちの少なくとも1つに基づいて前記少なくとも1つの連結成分から前記テキスト背景領域を選択する選択工程を含むことを特徴とする請求項1又は2記載の方法。

【請求項4】

前記少なくとも1つの連結成分の前記特徴は、
 前記連結成分の色又はグレースケールの均一性、
 前記連結成分のサイズ、
 前記連結成分の形状、
 前記連結成分の境界線の規則性、
 前記画像中の前記連結成分の位置、
 前記連結成分の平均グレースケール値、及び
 前記連結成分のグレースケール値分布

のうちの少なくとも1つを含むことを特徴とする請求項3に記載の方法。

【請求項5】

前記メンバ連結成分の前記統計は、
 前記連結成分中の前記メンバ連結成分の数、
 第1の事前定義済みの閾値より高いテキスト信頼度を有する前記メンバ連結成分のシード連結成分の数、
 前記連結成分中の前記メンバ連結成分の平均テキスト信頼度、及び
 前記連結成分中の前記メンバ連結成分の総面積と前記連結成分の面積との比
 のうちの少なくとも1つを含むことを特徴とする請求項3に記載の方法。

【請求項6】

前記少なくとも1つの連結成分と他のテキスト背景領域との間の前記関係に基づいて前記少なくとも1つの連結成分から前記テキスト背景領域を選択する選択工程は、
 前に判定されたあらゆるテキスト背景領域におけるメンバ連結成分ではなく、且つ前記前に判定されたテキスト背景領域と同一のメンバ連結成分を有さない前記連結成分に
 応答して連結成分をテキスト背景領域として判定する判定工程を含むことを特徴とする請求項3に記載の方法。

【請求項7】

前記フィルタリング工程において、あらゆるテキスト背景領域の境界内の連結成分及びあらゆるテキスト背景領域の境界内にない連結成分は、

前記連結成分のサイズ、
 前記連結成分の形状、
 前記連結成分のバウンディングボックスのアスペクト比、
 前記連結成分とそのバウンディングボックスの面積比、
 周長と前記連結成分の面積との比、及び
 前記連結成分のテクスチャ特徴、

のうちの少なくとも1つに基づいてフィルタリングされることを特徴とする請求項1に記載の方法。

【請求項8】

あらゆるテキスト背景領域内にない前記連結成分は、
 ストローク幅の統計、及び
 境界線画素数と連結成分画素数との比、

のうちの少なくとも1つに更に基づいてフィルタリングされることを特徴とする請求項7に記載の方法。

【請求項9】

10

20

30

40

50

少なくとも1つの連結成分を含む画像においてテキスト領域を検出するテキスト検出方法であって、

前記画像からテキスト背景領域を検出する検出工程と、

前記少なくとも1つの連結成分をフィルタリングしてテキスト候補となる連結成分を残すフィルタリング工程と、

前記検出工程において検出された前記テキスト背景領域に基づいて、前記テキスト候補となる連結成分をグループ化して少なくとも1つの連結成分グループを形成し、且つ前記少なくとも1つの連結成分グループに基づいて少なくとも1つのテキスト領域を生成するグループ化工程と、

を備え、

前記グループ化工程は、

a) 前記候補となるテキスト連結成分をそれぞれのテキスト背景領域に割り当てること、及びテキスト背景領域に割り当て不可能な前記候補となるテキスト連結成分を外側領域に割り当てる工程と、

b) 各テキスト背景領域及び前記外側領域の連結成分をそれぞれグループ化して連結成分グループを形成する工程と、

c) 各テキスト背景領域及び前記外側領域間の種々の領域からの連結成分グループを組み合わせる前記少なくとも1つのテキスト領域を生成する工程と

を更に含むことを特徴とするテキスト検出方法。

【請求項10】

前記工程 b) は、

b1) 各テキスト背景領域及び前記外側領域の連結成分を暗い領域の明るい連結成分の集合及び明るい領域の暗い連結成分の集合に分割する工程と、

b2) それぞれ前記明るい連結成分の集合及び前記暗い連結成分の集合内に連結成分グループを生成する工程と、

b3) 空間関係及び外観の類似性のうちの少なくとも1つに基づいて、前記明るい連結成分の集合と前記暗い連結成分の集合とを組み合わせる工程と

を更に含むことを特徴とする請求項9に記載の方法。

【請求項11】

前記工程 c) において、

行方向の整合性、

前記連結成分の平均サイズの類似性、

前記連結成分の平均ストローク幅の類似性、及び

前記連結成分の平均的な色又はグレースケールの類似性、

のうちの少なくとも1つに基づいて種々の領域からの連結成分グループを組み合わせることを特徴とする請求項9に記載の方法。

【請求項12】

前記工程 b2) において、前記連結成分グループは連結成分クラスタリングにより生成され、

当該連結成分クラスタリングは、以下の制約、

ある特定の方向に従う連結成分の中心の位置合わせ、

連結成分のサイズの類似性、

連結成分の形状の類似性、

連結成分の色又はグレースケールの類似性、

連結成分のストローク幅の類似性、及び

連結成分間の距離

のうちの少なくとも1つを使用することを特徴とする請求項10に記載の方法。

【請求項13】

前記工程 b2) は、

ハフ変換により連結成分グループを生成する工程と、

10

20

30

40

50

以下の制約の少なくとも1つを用いて、生成された連結成分グループに含まれる連結成分をフィルタリングする工程とを更に有する

前記制約は、

ある特定の方向に従う連結成分の中心の位置合わせ、
連結成分のサイズの類似性、
連結成分の形状の類似性、
連結成分の色又はグレースケールの類似性、
連結成分のストローク幅の類似性、及び
連結成分間の距離

である、

ことを特徴とする請求項10に記載の方法。

【請求項14】

前記工程c)において前記組み合わせることに対する規則は、前記工程b)において前記グループ化することに対する規則より厳しいことを特徴とする請求項9に記載の方法。

【請求項15】

前記フィルタリング工程は、

前記テキスト背景領域以外の前記少なくとも1つの連結成分の各々のテキスト信頼度を算出する工程と、

ここで前記計算において、あらゆるテキスト背景領域の前記境界内の連結成分は他の連結成分より重視される、

第2の事定義済みの閾値より高いテキスト信頼度を有する連結成分を前記テキスト候補連結成分として判定する工程と

を含むことを特徴とする請求項1に記載の方法。

【請求項16】

前記フィルタリング工程は、

前記テキスト背景領域以外の前記少なくとも1つの連結成分毎に、前記連結成分があらゆるテキスト背景領域の前記境界内に配置されることに応答して、前記連結成分を第1の連結成分として識別するか、あるいは前記連結成分を第2の連結成分として識別する工程と、

前記連結成分が前記テキスト候補連結成分であるかを判定するために、前記第1の連結成分及び前記第2の連結成分の各々に対して第1のフィルタリング工程を実行する工程と、

前記第2の連結成分が前記テキスト候補連結成分であるかを更に判定するために、前記第1のフィルタリング工程により前記テキスト候補連結成分として判定される前記第2の連結成分の各々に対して第2のフィルタリング工程を実行する工程と

を含むことを特徴とする請求項1に記載の方法。

【請求項17】

前記第1の連結成分及び前記第2の連結成分の各々に対して前記第1のフィルタリング工程を実行する工程は、

前記連結成分が前記テキスト候補連結成分であるかを判定するために、前記連結成分の1つ以上の第1の特徴に基づいて前記第1のフィルタリング工程を実行する工程を含むことを特徴とする請求項16に記載の方法。

【請求項18】

前記第1のフィルタリング工程により前記テキスト候補連結成分として判定される前記第2の連結成分の各々に対して前記第2のフィルタリング工程を実行する工程は、

前記第2の連結成分が前記テキスト候補連結成分であるかを更に判定するために、前記第2の連結成分の1つ以上の第2の特徴に基づいて前記第2のフィルタリング工程を実行する工程を含むことを特徴とする請求項16に記載の方法。

【請求項19】

前記テキストチャ特徴は、

10

20

30

40

50

ローカルバイナリパターン、
エッジ方向ヒストグラム、及び
勾配のヒストグラム

のうちの少なくとも1つを含むことを特徴とする請求項7に記載の方法。

【請求項20】

前記第1のフィルタリング工程において、前記第1の特徴は、カスケード規則として使用されるか、あるいは訓練分類器に入力される特徴ベクトルとして組み合わされることを特徴とする請求項17に記載の方法。

【請求項21】

前記第2のフィルタリング工程において、前記第2の特徴は、カスケード規則として使用されるか、あるいは訓練分類器に入力される特徴ベクトルとして組み合わされることを特徴とする請求項18に記載の方法。

【請求項22】

少なくとも1つの連結成分を含む画像においてテキスト領域を検出するテキスト検出装置であって、

前記画像からテキスト背景領域を検出する検出手段と、

前記少なくとも1つの連結成分をフィルタリングしてテキスト候補となる連結成分を残すフィルタリング手段と、

前記検出手段において検出された前記テキスト背景領域に基づいて、前記テキスト候補となる連結成分をグループ化して少なくとも1つの連結成分グループを形成し、且つ前記少なくとも1つの連結成分グループに基づいて少なくとも1つのテキスト領域を生成するグループ化手段と、

を備え、

前記フィルタリング手段において、前記テキスト背景領域の境界内にはない連結成分に対するフィルタリングは、前記テキスト背景領域の境界内の連結成分に対するフィルタリングより厳しいことを特徴とするテキスト検出装置。

【請求項23】

少なくとも1つの連結成分を含む画像においてテキスト領域を検出するテキスト検出装置であって、

前記画像からテキスト背景領域を検出する検出手段と、

前記少なくとも1つの連結成分をフィルタリングしてテキスト候補となる連結成分を残すフィルタリング手段と、

前記検出手段において検出された前記テキスト背景領域に基づいて、前記テキスト候補となる連結成分をグループ化して少なくとも1つの連結成分グループを形成し、且つ前記少なくとも1つの連結成分グループに基づいて少なくとも1つのテキスト領域を生成するグループ化手段と、

を備え、

前記グループ化手段は、

a) 前記候補となるテキスト連結成分をそれぞれのテキスト背景領域に割り当てること、及びテキスト背景領域に割り当て不可能な前記候補となるテキスト連結成分を外側領域に割り当てる手段と、

b) 各テキスト背景領域及び前記外側領域の連結成分をそれぞれグループ化して連結成分グループを形成する手段と、

c) 各テキスト背景領域及び前記外側領域間の種々の領域からの連結成分グループを組み合わせて前記少なくとも1つのテキスト領域を生成する手段と

を更に含むことを特徴とするテキスト検出装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、テキスト検出に関し、特に、テキスト検出の方法及び装置、並びにテキスト

10

20

30

40

50

情報抽出の方法及びシステムに関する。

【背景技術】

【0002】

画像、特に自然画像においてテキストを検出することは、視覚障害者及び外国人のためのコンピュータ化支援、画像及び映像の自動検索、並びに都市環境でのロボットナビゲーション等の多数の画像認識の適応例にとって極めて重要である。

【0003】

それにもかかわらず、自然シーンでのテキスト検出は難しい問題である。印刷されたページ、ファックス及び名刺の走査とは対照的に、主な課題は、フォント、サイズ、スキュー角、斜体及び傾斜による歪み等のテキストの多様性にある。不均一な照明及び反射、劣悪な照明条件、並びに複雑な背景等の環境要因により、より複雑化している。

【0004】

関連文献において、自然シーンにおいてテキスト領域を検出するテキスト検出方法は、通常、図1に示されたフローチャートに従う。図1の方法100はブロック110から開始し、画像から成分を生成する。ここで、成分は、同様の色又はグレースケール、あるいはストローク幅を有する画素グループである連結成分(CC)等であってもよい。

【0005】

次にブロック120において、種々の特徴が各成分から抽出され、その特徴に基づいて非テキスト成分がフィルタリングされる結果、候補となるテキスト成分が確保される。

【0006】

次にブロック130において、確保された候補となるテキスト成分は、テキスト行又は単語を定型化するために共にグループ化され、テキスト行又は単語のバウンディングボックス(テキストを含む矩形等の最小多角形)としてテキスト領域を出力する。

【0007】

従来技術の一般的な問題は、従来技術が、エッジ、隅、ストローク、色及びテクスチャ等のテキスト領域の特徴のみによってテキストを検出しようとすることである。しかし、殆どの場合に有用となりうるテキスト周囲のコンテキスト情報は無視される。その結果、従来技術では、複雑な非テキスト領域による誤検出及び自然シーン中のテキストの広範囲の多様性による非検出が発生する。従って、テキスト領域の特徴によってのみ、直接テキストを検出しようとする際には厳しいトレードオフがある。

【0008】

従って、テキスト領域の周囲の背景情報を利用する画像でのテキスト検出のために改善された方法が必要である。

【0009】

Yasuhiro Kunishige、Feng Yaokai、Seiichi Uchida、International Conference on Document Analysis and Recognition (ICDAR)により出版された文献「Scenery character detection with environmental context」、1049~1053ページ、2011年において、コンテキスト情報の概念を使用するテキスト検出方法が提案される。特に、文献は、対象成分の周囲に10画素の余白を追加することで形成される対象成分の拡張領域からコンテキスト特徴を抽出する。更に方法は、6個のシーン成分カテゴリ、すなわち「空」、「緑」、「看板」、「地面」、「建物」及び「その他」のうちの1つに成分を分類する。この方法の1つの問題は、「空」、「緑」等の一般的なカテゴリがテキスト検出には不適切であり、且つ「看板」が例えばロゴ、タグ、スクロール、ポスター等の全ての種類のテキスト背景領域を範囲に含めないことである。この方法の別の問題は、コンテキスト情報が、シーンの変動に適應できない固定のサイズの余白領域から抽出されることである。更に別の問題は、情報がCCのグループ化において有益である一方で、この方法が空間におけるCCの関係、例えばどのCCがある看板に属しているかを取得できないことである。

10

20

30

40

50

【発明の概要】**【発明が解決しようとする課題】****【0010】**

従って、上述の問題の少なくとも1つに対処する必要がある。

【0011】

本発明者は、自然シーン中の殆どのテキストは、容易に認識されるように、テキストに対して高いコントラストを有する相対的に同様の背景領域上に印刷されていることに着目した。これは、テキスト検出にとって有用だろう。

【0012】

それにより、新規のテキスト検出の方法及び装置は、画像、特に自然シーン画像でのテキスト検出の性能を向上させるために本発明において提案される。テキストを取り囲むこの一般的な背景領域を規定するために、新しい概念であるテキスト背景領域(TBR)が本発明に導入される。自然シーン画像において、TBRは、通常、看板、ロゴ、タグ、スクロール及びポスター等として存在するが、これらの形態に限定されない。テキスト領域を直接見つけるのではなく、最初にTBRを見つけて、TBR内の成分及び全てのTBRの外、すなわち外側領域(OR)の成分として成分を分類することにより、テキストを探索する。テキストは、TBRの外よりTBR内に現れる可能性がより高いと仮定される。また、1つのテキスト行/単語は、2つの領域(2つのTBR又はTBR及びOR)を交差することはめったにない。その仮定に基づいて、TBR情報は、連結成分のフィルタリング及び/又は連結成分のグループ化において使用されうる。

【課題を解決するための手段】**【0013】**

本発明の第1の態様によると、少なくとも1つの連結成分(CC)を含む画像においてテキスト領域を検出する方法が提供される。方法は、画像から少なくとも1つのテキスト背景領域(TBR)を検出するTBR検出ステップと、少なくとも1つのCCをフィルタリングして少なくとも1つの候補となるテキストCCを確保するCCフィルタリングステップと、TBR検出ステップにおいて検出されたTBRに基づいて少なくとも1つの候補となるテキストCCをグループ化して少なくとも1つのCCグループを形成し、且つ少なくとも1つのCCグループに基づいて少なくとも1つのテキスト領域を生成するCCグループ化ステップとを備える。

【0014】

本発明の第2の態様によると、少なくとも1つの連結成分(CC)を含む画像においてテキスト領域を検出するテキスト検出装置が提供される。装置は、画像からテキスト背景領域(TBR)を検出するように構成されたTBR検出ユニットと、少なくとも1つのCCをフィルタリングして少なくとも1つの候補となるテキストCCを確保するように構成されたCCフィルタリングユニットと、TBR検出ユニットにおいて検出されたTBRに基づいて少なくとも1つの候補となるテキストCCをグループ化して少なくとも1つのCCグループを形成し、且つ少なくとも1つのCCグループに基づいて少なくとも1つのテキスト領域を生成するように構成されたCCグループ化ユニットとを備える。

【0015】

本発明の第3の態様によると、テキスト情報抽出方法が提供される。方法は、本発明の第1の態様に係るテキスト検出方法を使用して入力画像又は入力映像からテキスト領域を検出するステップと、検出されたテキスト領域からテキストを抽出するステップと、抽出されたテキストを認識してテキスト情報を取得するステップとを備える。

【0016】

本発明の第4の態様によると、テキスト情報抽出システムが提供される。システムは、入力画像又は入力映像からテキスト領域を検出するように構成された本発明の第2の態様に係るテキスト検出装置と、検出されたテキスト領域からテキストを抽出するように構成された抽出装置と、抽出されたテキストを認識してテキスト情報を取得するように構成された認識装置とを備える。

10

20

30

40

50

【 0 0 1 7 】

これらの特徴を利用することにより、本発明に係る方法、装置及びシステムは、画像中のテキストの場所を迅速に且つ／あるいは高精度に示すことができ、その結果、テキスト検出の性能が向上する。

【 0 0 1 8 】

図面を参照して、以下の説明から本発明の更なる特徴及び利点が明らかになるだろう。

【図面の簡単な説明】

【 0 0 1 9 】

本明細書に組み込まれ且つその一部を構成する添付の図面は、本発明の実施形態を例示し、説明と共に本発明の原理を説明するのに役立つ。

【図 1】図 1 は、画像においてテキスト領域を検出する従来技術の方法を示すフローチャートである。

【図 2】図 2 は、本発明の実施形態を実現できるコンピュータシステムの例示的なハードウェア構成を示すブロック図である。

【図 3】図 3 は、本発明の一実施形態に係る画像においてテキスト領域を検出するテキスト検出方法を示すフローチャートである。

【図 4】図 4 は、本発明の一実施形態に係る C C と他の T B R との関係に基づいて C C が T B R であるかを判定する例示的な処理を示す図である。

【図 5 A】図 5 A は、本発明の一実施形態に係る図 3 の C C フィルタリングステップを示すフローチャートである。

【図 5 B】図 5 B は、本発明の別の実施形態に係る図 3 の C C フィルタリングステップを示すフローチャートである。

【図 5 C】図 5 C は、本発明の一実施形態に係る訓練分類器を使用する図 3 の C C フィルタリングステップにおける例示的な処理を示す図である。

【図 6】図 6 は、図 3 の C C グループ化ステップを実行することを示す例示的なフローチャートである。

【図 7】図 7 は、図 6 のグループ化ステップを実行することを示す例示的なフローチャートである。

【図 8 A】、

【図 8 B】、

【図 8 C】、

【図 8 D】、

【図 8 E】、

【図 8 F】図 8 A ~ 図 8 F は、本発明に係るテキスト検出方法を使用してそれぞれの処理結果を示す例示的な画像を示す図である。

【図 9】図 9 は、本発明の一実施形態に係る画像においてテキスト領域を検出するテキスト検出装置を示すブロック図である。

【図 10 A】図 10 A は、本発明の一実施形態に係る図 9 の C C フィルタリングユニットを示すブロック図である。

【図 10 B】図 10 B は、本発明の別の実施形態に係る図 9 の C C フィルタリングユニットを示すブロック図である。

【図 11 A】図 11 A は、本発明の一実施形態に係る図 9 の C C グループ化ユニットを示すブロック図である。

【図 11 B】図 11 B は、本発明の一実施形態に係る図 11 A のグループ化ユニットを示すブロック図である。

【図 12】図 12 は、本発明の一実施形態に係るテキスト情報抽出方法を示すフローチャートである。

【図 13】図 13 は、本発明の一実施形態に係るテキスト情報抽出システムを示すブロック図である。

【発明を実施するための形態】

10

20

30

40

50

【 0 0 2 0 】

以下、図面を参照して、本発明の実施形態を詳細に説明する。本明細書に組み込まれ且つその一部を構成する添付の図面は、本発明の実施形態を例示し、説明と共に本発明の原理を説明するのに役立つ。

【 0 0 2 1 】

尚、同様の図中符号及びアルファベットは図中同様の項目を示すため、項目は、1つの図面において規定されれば、後続の図面に対して説明されることを要しない。

【 0 0 2 2 】

また、本発明において、「第1の」及び「第2の」等の用語は、要素又はステップを区別するためだけに使用され、時間順序、存在又は重要性を示すことを意図しない。

10

【 0 0 2 3 】

図2は、本発明の実施形態を実現できるコンピュータシステム1000のハードウェア構成を示すブロック図である。

【 0 0 2 4 】

図2に示されるように、コンピュータシステムはコンピュータ1110を備える。例えばコンピュータ1110は、デジタルカメラ又はスマートフォンであってよい。コンピュータ1110は、処理ユニット1120と、システムメモリ1130と、取り外し不可能な不揮発性メモリインタフェース1140と、取り外し可能な不揮発性メモリインタフェース1150と、ユーザ入力インタフェース1160と、ネットワークインタフェース1170と、ビデオインタフェース1190と、出力周辺インタフェース1195とを備え、それらはシステムバス1121を介して接続される。

20

【 0 0 2 5 】

システムメモリ1130は、ROM(読み出し専用メモリ)1131及びRAM(ランダムアクセスメモリ)1132を備える。BIOS(基本入出力システム)1133はROM1131に常駐する。オペレーティングシステム1134、アプリケーションプログラム1135、他のプログラムモジュール1136及びいくつかのプログラムデータ1137は、RAM1132に常駐する。

【 0 0 2 6 】

ハードディスク等の取り外し不可の不揮発性メモリ1141は、取り外し不可の不揮発性メモリインタフェース1140に接続される。取り外し不可の不揮発性メモリ1141は、例えばオペレーティングシステム1144、アプリケーションプログラム1145、他のプログラムモジュール1146及びいくつかのプログラムデータ1147を格納できる。

30

【 0 0 2 7 】

フラッシュドライブ1151及びCD-ROMドライブ1155等の1つ以上の取り外し可能な不揮発性メモリドライブは、取り外し可能な不揮発性メモリインタフェース1150に接続される。例えば、SDカード等のフラッシュメモリ1152は、フラッシュドライブ1151に挿入可能であり、CD(コンパクトディスク)1156は、CD-ROMドライブ1155に挿入可能である。処理される画像は、不揮発性メモリに格納可能である。

40

【 0 0 2 8 】

マイク1161及びキーボード1162等の入力装置は、ユーザ入力インタフェース1160に接続される。

【 0 0 2 9 】

コンピュータ1110は、ネットワークインタフェース1170によりリモートコンピュータ1180に接続可能である。例えばネットワークインタフェース1170は、ローカルエリアネットワーク1171を介してリモートコンピュータ1180に接続可能である。あるいは、ネットワークインタフェース1170は、モデム(変調器-復調器)1172に接続可能であり、モデム1172は、ワイドエリアネットワーク1173を介してリモートコンピュータ1180に接続される。

50

【 0 0 3 0 】

リモートコンピュータ 1 1 8 0 は、リモートアプリケーションプログラム 1 1 8 5 を格納するハードディスク等のメモリ 1 1 8 1 を含めることができる。

【 0 0 3 1 】

ビデオインタフェース 1 1 9 0 は、本発明の実施形態に係る 1 つ以上の処理結果を表示するために使用されてもよいモニタ 1 1 9 1 に接続される。

【 0 0 3 2 】

出力周辺インタフェース 1 1 9 5 は、プリンタ 1 1 9 6 及びスピーカ 1 1 9 7 に接続される。

【 0 0 3 3 】

図 2 に示されたコンピュータシステムは、例示にすぎず、本発明、その適応例又は使用を限定することを全く意図しない。

【 0 0 3 4 】

図 2 に示されたコンピュータシステムは、スタンドアロンコンピュータ又は装置の処理システムとして、場合によっては 1 つ以上の不要な構成要素が除去された状態あるいは 1 つ以上の更なる構成要素が追加された状態で、実施形態のいずれかに対して実現されてもよい。

【 0 0 3 5 】

図 3 は、本発明の一実施形態に係る画像においてテキスト領域を検出するテキスト検出方法 3 0 0 を示すフローチャートである。本発明に係るテキスト検出方法を使用して処理結果を示す図 8 A ~ 図 8 F を更に参照して説明する。

【 0 0 3 6 】

一実施形態によると、画像は少なくとも 1 つの連結成分 (C C : Connected Component) を含む。 C C は、同様の色又はグレースケール値を含む画素クラスタ (画素群) である。 1 つのクラスタにおける画素は、空間において 4 近傍又は 8 近傍で接続可能である。 C C は、例えば色クラスタリング、適応 2 値化及び形態素処理等により、 T B R 検出ステップの前に画像から生成されうる。実施形態の 1 つにおいて、内容が参考として本明細書に取り入れられる「 Robust wide baseline stereo from maximally stable extremal regions」、 J . M a t a s、 O . C h u m、 M . U r b a n 及び T . P a j d l a、 Proc . o f B r i t i s h M a c h i n e V i s i o n C o n f e r e n c e、 3 8 4 ~ 3 9 6 ページ、 2 0 0 2 年において説明された M S E R 方法に基づいて、 C C はグレースケール画像から生成されうる。一実施形態において、暗い背景上の明るいテキスト及び明るい背景上の暗いテキストの双方を検出するために、 C C の生成は、一方は元の画像用であり且つ他方は倒像用である 2 つのチャンネルに適用可能である。しかし、これは必ずしも要求されない。

【 0 0 3 7 】

一例として、図 8 A 及び図 8 B は、それぞれ、グレースケール画像及び画像中の生成された C C を示す。図 8 B において、黒線のボックスの各々は C C を示す。すなわち、黒線のボックスは、 C C のバウンディングボックス (C C を含む最小の四角形) である。

【 0 0 3 8 】

ブロック 3 1 0 において、テキスト背景領域 (T B R : Text Background Region) 検出ステップは、入力画像から少なくとも 1 つの T B R を検出するために実行される。

【 0 0 3 9 】

一実施形態によると、 T B R 検出ステップは、画像に含まれた C C に基づいて実行されうる。 T B R は、特殊な特性を含むような C C であってよい。一実施形態によると、 T B R は、規則的な境界線及び均一な色又はグレースケールを有する画像中のテキストの周囲領域であってよい。自然シーン画像において、 T B R は、通常、看板、ロゴ、タグ、スクロール、ポスター等として存在するが、これらの形態に限定されない。

【 0 0 4 0 】

一実施形態によると、TBRは、画像に含まれたCCから選択されうる。各CCは、それがTBRであるかを判定するためにチェックされうる。特性の3つの態様は、CCの特徴、CC中のメンバCCの統計的特徴、及びCCと他のTBRとの関係の判定において考慮されうる。これらの態様は、TBRの検出のために個々に又はあらゆる組合せで使用されうる。ここで、現在のCC中のメンバCCは、現在のCCの境界内に配置されたCCであり、現在のCCに対して高いコントラストを有する。メンバCCは、現在のCCの反対側のチャンネルから抽出されうる。例えば、暗いCCのメンバCCは暗いCCの領域内の明るいCCであってよく、明るいCCのメンバCCは明るいCCの領域内の暗いCCであってよい。

【0041】

図8Cは、白線のボックスで示される、2つのTBRが検出された画像を示す。

【0042】

[CCの特徴]

CCは、それがTBRであるかを判定するため、その特徴に基づいてチェックされうる。CCの特徴は、例えば、CCの色又はグレースケールの均一性、CCのサイズ、CCの形状、CCの境界線の規則性、画像中のCCの位置、CCの平均グレースケール値及びCCのグレースケール値分布のうちの少なくとも1つを含みうる。

【0043】

尚、TBRのサイズは、通常、相対的に大きい。従って、一実施形態によると、全てのCCはサイズによりソート可能であり、TBRは、上位n個の最も大きなCCから選択可能である。

【0044】

尚、TBRは、通常、画像の余白領域ではなく、画像中の顕著な位置に配置される。従って、別の実施形態によると、余白領域に配置されたCCは、非TBR領域として除外されうる。例えば余白領域は、1/m像幅等の指定された幅又は1/m像高を有する画像の外側ループ領域として規定されうる。

【0045】

更に別の実施形態によると、TBRが、通常、規則的な境界線を有するため、CCの境界線の規則性は、TBRの判定において考慮されうる。境界線の規則性は、CCの密度(バウンディングボックスにおけるCCの占有率)、境界線率(境界線画素とCC画素の量との比)及び境界線の対称性(4つの四分円における密度差により評価されうる4つの四分円における境界線の類似性)により測定されうる。

【0046】

更に別の実施形態によると、TBRがフラッシュにより形成されるべきではないため、いくつかの特徴は、TBRをフラッシュから区別するために使用される。尚、フラッシュは、通常、中央において平均よりも非常に高いグレースケール値を含む高い平均グレースケール値を有する。従って、CCの平均グレースケール値及びグレースケール値分布は、TBRをフラッシュから区別するために使用されうる。

【0047】

[CC中のメンバCCの統計]

CCは、それがTBRであるかを判定するために、CC中のメンバCCの統計に基づいてチェックされうる。ここで、現在のCC中のメンバCCは、現在のCCの境界内に配置されたCCであり、現在のCCに対して高いコントラストを有する。メンバCCの領域は、完全に現在のCCの境界内にある。メンバCCは、現在のCCの反対側のチャンネルから抽出されうる。例えば、暗いCCのメンバCCは暗いCCの領域内の明るいCCであってよく、明るいCCのメンバCCは明るいCCの領域内の暗いCCであってよい。

【0048】

メンバCCの統計は、例えば、CC中のメンバCCの数、メンバCCのシードCCの数、CC中のメンバCCの平均テキスト信頼度及びCC中のメンバCCの総面積とCCの面積との比のうちの少なくとも1つを含みうる。

10

20

30

40

50

【 0 0 4 9 】

説明のために、CC中のメンバCCの数は閾値より多いことが好ましい。

【 0 0 5 0 】

メンバCCの間のシードCCの数は、閾値より多いことが好ましい。ここで、シードCCは、テキスト成分である可能性が非常に高いCCを示す。例えばシードCCは、事前定義済みの閾値より高いテキスト信頼度を有するCCであってよい。シードCCを選択するために、特徴の集合は、CCのテキスト信頼度を算出するためにCCから抽出されうる。

【 0 0 5 1 】

一例として、図8Dは、白線のボックスにおけるいくつかのシードCCを示す。

【 0 0 5 2 】

シードCCを選択するために一般的に使用される特徴は、CCのサイズ、CCの幅/高さの比、CCの密度(すなわち、バウンディングボックス内のCC画素の占有率)、CCのストローク幅の統計的特徴及びCCの領域から抽出されたテクスチャ特徴を含みうる。一実施形態において、特徴は、テキスト信頼度を調整する際に規則として使用されうる。別の実施形態において、テキスト分類器は、テキストCC及び非テキストCCの双方を含む訓練集合に基づいて習得されうる。分類器は、CCの特徴を入力として使用し、CCのテキスト信頼度値を出力する。

【 0 0 5 3 】

[CCと他のTBRとの関係]

CCは、それがTBRであるかを判定するために、他のTBRとの関係に基づいてチェックされうる。一実施形態によると、少なくとも1つのCCと他のTBRとの関係に基づいてCCからTBRを選択することは、前に判定されたあらゆるTBRにおけるメンバCCではなく、且つ前に判定されたTBRと同一のメンバCCを有さないCCに回答して、CCをTBRとして判定することを含みうる。換言すると、TBRは、通常、互いに重複しないかあるいは互いを含まない。

【 0 0 5 4 】

一例として、図4は、CCと他のTBRとの関係に基づいてCCがTBRであるかを判定する処理400を示す。

【 0 0 5 5 】

ブロック410において、現在のCCが前に判定されたTBRのメンバCCであるかを判定する。現在のCCは、前に判定されたTBRのメンバCCである場合にはTBRではないものとして識別される。現在のCCが前に判定されたTBRのメンバCCでない場合、処理400はブロック420に進み、現在のCCが前に判定されたTBRと同一のメンバCCを有するかを判定する。現在のCCは、前に判定されたTBRと同一のメンバCCを全く有さない場合にがTBRとして識別される。現在のCCが前に判定されたTBRと同一のメンバCCを有する場合、処理400はブロック430に進み、現在のCC及び前に判定されたTBRのうちどちらがよりTBRのようであるかを判定する。現在のCCがよりTBRのようである場合、前に判定されたTBRはTBRの集合から除去され(ブロック440)、現在のCCはTBRとして識別される。前に判定されたTBRがよりTBRのようである場合、現在のCCは、TBRではないものとして識別される。

【 0 0 5 6 】

ブロック430においてどちらがよりTBRのようであるかを判定する種々の方法があってよい。例えば、上述の基準、例えば境界線の規則性、CCの密度及びメンバCCの平均テキスト信頼度等のうちの1つ以上が使用されうる。

【 0 0 5 7 】

再度、図3を参照する。画像からTBRが検出されているため、処理300はブロック320に進む。ブロック320において、CCフィルタリングステップは、少なくとも1つのCCをフィルタリングして少なくとも1つの候補となるテキストCCを確保するために実行される。

【 0 0 5 8 】

10

20

30

40

50

特に及び好ましくは、CCからTBRを選択した後、残りのCCは、候補となるテキストCC及び非テキストCCを含む非TBR CCである。このステップの後、候補となるテキストCCが確保されるが、非テキストCCは除去される。

【0059】

一実施形態によると、TBR情報は、CCをフィルタリングするために使用される。TBRの境界内のCC及びあらゆるTBRの境界内でないCCの双方は、同一の規則に基づいてフィルタリングされることが好ましい。例えば、TBRの境界内のCC及びあらゆるTBRの境界内でないCCは、CCのサイズ、CCの形状、CCのバウンディングボックスのアスペクト比、CCとそのバウンディングボックスとの面積比、周長とCCとの面積の比及びCCのテクスチャ特徴のうちの少なくとも1つに基づいてフィルタリングされる。

10

【0060】

好ましくは、CCをフィルタリングする場合、CCのサイズは、経験値に基づいて所定の範囲にあるものとして選択されうる。テキストCCのアスペクト比が、通常、高すぎないため、CCのバウンディングボックスのアスペクト比は、所定の閾値より低いものとして選択されうる。CCとそのバウンディングボックスとの面積比は、低すぎるべきではなく、経験値に従って所定の閾値より高いものとして選択されうる。CCの面積比に対する周長は、所定の閾値より短くなりうる。ノイズCCの面積比に対する周長が、通常、相対的に長いため、これはノイズCCを除去するためである。CCから抽出されたウェーブレット、Gabor、LBP等のテクスチャ特徴は、テキスト信頼度を算出するために使用されうるため、CCをフィルタリングする際に組み込まれうる。

20

【0061】

尚、テキストは、TBRの外よりTBR内に現れる可能性がより高い。従って、別の実施形態によると、TBR情報は、フィルタリングの効率及び精度を更に向上させるために、CCをフィルタリングする際に使用されうる。

【0062】

特に、例えばCCフィルタリングステップにおいて、あらゆるTBRの境界内でないCCに対するフィルタリングは、TBRの境界内のCCに対するフィルタリングより厳しくてよい。別の例の場合、あらゆるTBRの境界内でないCCは、TBRの境界内のCCより多くの規則によりフィルタリングされうる。

30

【0063】

これは、あらゆるTBRの境界内でないものとして判定されるCCが算出された相対的に低いテキスト信頼度を有する結果、ノイズを被りやすいためである。従って、微フィルタリングは、非テキストCCを除去するために実行されうる。

【0064】

あらゆるTBRの境界内でないCCは、ストローク幅の統計及びノ又はCCの境界線画素数とCCの画素数との比に更に基づいてフィルタリングされることが好ましい。例えばストローク幅の統計は、ストローク幅の分散と平均ストローク幅との比を含みうる。

【0065】

図8Dは、フィルタリングの結果を示す。確保されたテキスト候補CCは、黒線又は白線で描画されたバウンディングボックスにより示される。ここで、白線のバウンディングボックスはシードCCを示し、黒線のバウンディングボックスは非シードCCを示す。図8Bと比較して、CCの一部、特にTBRの外側に配置されたCCは除去される。

40

【0066】

CCのフィルタリングを実行する例示的な方法を示すために、以下において2つの実施形態を説明する。

【0067】

図5Aは、本発明の一実施形態に係る図3のCCフィルタリングステップを示すフローチャートである。

【0068】

50

図5 Aにおいて、非T B R C Cの各々に対するテキスト信頼度は、C CがあらゆるT B Rの境界内に配置されるかに基づいて算出される。

【0069】

ブロック510において、各非T B R C Cのテキスト信頼度は、T B R情報に基づいて算出される。計算において、C C I RはC C O Rより重視される。

【0070】

ブロック520において、C Cのテキスト信頼度が事前定義済みの閾値Tより高いかを判定する。テキスト信頼度が閾値より高い場合、C Cはテキスト候補C Cとして判定される。テキスト信頼度が閾値より低い場合、C Cは非テキストC Cとして判定される。

【0071】

本実施形態の特定の一例は、以下の通り提供される。現在のC Cのテキスト信頼度は、ベイズの定理により規定されうる。

$$P(A|B) = P(B|A)P(A)/P(B)$$

式中、 $P(A)$ は事前確率であり、 $P(B|A)$ は条件付き確率であり、 $P(A|B)$ は事後確率である。

【0072】

C Cのフィルタリングの特定の例を考慮すると、Aは、ある特定のC Cのラベル(テキスト又は非テキスト)を示す確率変数である。 $P(A)$ は、現在のC Cのテキスト存在の事前確率を表す。 $P(A)$ はT B Rにより判定されうる。C C I Rは、C C O Rより高い $P(A)$ を与えられうる。

【0073】

$P(B|A)$ は、テキスト存在の条件付き確率である。 $P(B|A)$ は、テキスト領域が何に見えるかを説明する。従って、値は、テキスト領域自体から抽出されたテキスト特徴に基づいて算出される。 $P(B)$ は、現在のC Cの存在確率である。C Cが固定される場合、 $P(B)$ は一定値である。

【0074】

$P(A|B)$ は、現在のC Cのテキスト信頼度である。 $P(A|B)$ は、C C自体のテキスト特徴及びC Cに関連したT B R情報の双方による影響を受ける。事前定義済みの閾値より高い $P(A|B)$ 値を有するC Cは、テキスト候補C Cとして確保される。

【0075】

この例において、C C I Rに対する $P(A)/P(B)$ は1として設定可能であり、C C O Rに対する $P(A)/P(B)$ は、(0, 1)の範囲の値として設定可能である。

【0076】

図5 Bは、本発明の別の実施形態に係るC Cのフィルタリングを示すフローチャートである。

【0077】

図5 Bにおいて、フィルタリングは、T B R情報に基づいていくつかの段階、例えば2つの段階で非T B R C Cに対して実行される。例えば2つの段階は、粗フィルタリング及び微細フィルタリングを含む。全ての非T B R C Cは粗フィルタリングにかけられるが、あらゆるT B Rの境界外のC Cのみが微細フィルタリングにかけられてもよい。単純な特徴は粗フィルタリングの際に使用可能であり、より複雑な特徴は微細フィルタリングの際に使用可能である。従って、あらゆるT B Rの境界内にはないC Cに対するフィルタリングは、T B Rの境界内のC Cに対するフィルタリングより厳しく行われる。従って、微細フィルタリングにかけられるC Cの量は減少し、方法の効率は向上する。

【0078】

ブロック530において、非T B R C Cは、C C I R、すなわちあらゆるT B Rの境界内のC Cと、C C O R、すなわちあらゆるT B Rの境界内にはないC Cとの2つのグループに分離される。

【0079】

ブロック540において、粗フィルタリング等の第1のフィルタリングステップは、全

10

20

30

40

50

ての非 T B R C C に対して実行される。特に、非 T B R C C の各々は、候補となるテキスト C C 又は非テキスト C C として判定される。

【 0 0 8 0 】

第 1 のフィルタリングステップは、C C が候補となるテキスト C C であるかを判定するために、非 T B R C C の各々の 1 つ以上の第 1 の特徴に基づいて実行されうる。第 1 の特徴は、C C から抽出され、C C のサイズ、C C の形状、C C のバウンディングボックスのアスペクト比、C C の密度 (C C とそのバウンディングボックスの面積比)、C C の面積比に対する周長及び C C のテクスチャ特徴を含むがそれらに限定されない相対的に単純な特徴であってよい。一例として、テクスチャ特徴は、ローカルバイナリパターン、エッジ方向ヒストグラム及び勾配のヒストグラムを含みうるが、それらに限定されない。

10

【 0 0 8 1 】

第 1 の特徴は、カスケード規則として使用されうるか、あるいは訓練分類器に入力される特徴ベクトルとして組み合わせられうる。カスケード規則の閾値又は分類器は、テキストサンプル及び非テキストサンプルの双方から習得されうる。カスケード規則がフィルタリングの際に使用される場合、各入力 C C は事前定義済みの規則によりチェック可能であり、規則の少なくとも 1 つを満たさない C C は除去される。

【 0 0 8 2 】

ブロック 5 5 0 において、微細フィルタリング等の第 2 のフィルタリングステップは、候補となるテキスト C C O R が候補となるテキスト C C であるかを更に判定するために、候補となるテキスト C C O R、すなわち第 1 のフィルタリングステップにより候補となるテキスト C C として判定される C C O R の各々に対して実行される。C C O R が、ブロック 5 4 0 において候補となるテキスト C C として判定されたとしてもノイズを被りやすいため、第 2 のフィルタリングは非テキスト C C を更に除去してよい。

20

【 0 0 8 3 】

第 2 のフィルタリングステップにおいて、ブロック 5 4 0 において使用された特徴に対してより厳しい条件が採用されてよく、且つ / あるいはフィルタリングのためにいくつかの他の特徴が使用されてよい。他の特徴は、ストローク幅の統計 (例えば、ストローク幅の分散と平均ストローク幅との比) 及び / 又は境界線画素数と C C 画素数との比) を含みうる。

【 0 0 8 4 】

30

第 1 のフィルタリングステップと同様に、第 2 のフィルタリングに対する特徴は、カスケード規則として使用されうるか、あるいは訓練分類器に入力される特徴ベクトルとして組み合わせられうる。カスケード規則の閾値又は分類器は、テキストサンプル及び非テキストサンプルの双方から習得されうる。カスケード規則がフィルタリングの際に使用される場合、各入力 C C は事前定義済みの規則によりチェック可能であり、規則の少なくとも 1 つを満たさない C C は除去される。

【 0 0 8 5 】

図 5 C は、訓練分類器を使用する C C のフィルタリングを示すフローチャートである。図 5 C の方法は、第 1 のフィルタリングステップ及び第 2 のフィルタリングステップの双方に適用可能である。図 5 C において、テキストサンプル及び非テキストサンプルを含む訓練サンプルは、分類器を訓練するために使用される。C C は、テキスト候補 C C を取得するために訓練分類器により分類される。訓練及び分類のために抽出された特徴は、第 1 のフィルタリング及び第 2 のフィルタリングに関連して上述した特徴である。本発明の主題を不必要に不明確にすることを回避するために、フローチャートに関する更なる詳細については説明しない。

40

【 0 0 8 6 】

再度、図 3 を参照する。候補となるテキスト C C を取得した後、処理 3 0 0 はブロック 3 3 0 に進む。ブロック 3 3 0 において、C C グループ化ステップは、T B R 検出ステップにおいて検出された T B R に基づいて少なくとも 1 つの候補となるテキスト C C をグループ化して少なくとも 1 つの C C グループを形成し、且つ少なくとも 1 つの C C グループ

50

に基づいて少なくとも1つのテキスト領域を生成するために実行される。

【0087】

CCグループ化ステップ330を実行する例示的なフローチャートを図6に示す。図6に示されるように、CCグループ化ステップ330は、ステップ610～630を含みうる。

【0088】

CCグループ化ステップに対する入力は、候補となるテキストCCである。

【0089】

ステップ610において、候補となるテキストCCは、それぞれのテキスト背景領域に割り当てられる。TBRに割り当て不可能な候補となるテキストCCは、外側領域に割り

10

【0090】

ステップ610は、図5Bのブロック530に示されるようなステップに類似する。従って、ステップ610についての説明は省略する。尚、ステップ310でTBRが検出されない場合、全てのCCは外側領域にある。候補となるテキストCCを割り当てた後、ステップ620に進む。

【0091】

ステップ620において、各TBR及び外側領域のCCは、それぞれ、グループ化されてCCグループを形成する。

【0092】

20

このステップにおいて、1つの領域のCCは、空間関係及び外観の類似性に基づいてグループ化される。図6のグループ化ステップ620を実行する例示的なフローチャートを図7に示す。

【0093】

図7に示されるように、グループ化ステップ620は、ステップ6201～6203を含みうる。

【0094】

ステップ6201において、各TBR及び外側領域のCCは、暗い領域の明るいCCの集合及び明るい領域の暗いCCの集合に分割される。

【0095】

30

ステップ6202において、CCグループは、それぞれ、明るいCCの集合及び暗いCCの集合内に生成される。

【0096】

好ましい一実施形態によると、CCグループはCCクラスタリングにより生成されうる。CCクラスタリングは、ある特定の方向に従うCCの中心の位置合わせ、CCのサイズの類似性、CCの形状の類似性、CCの色又はグレースケールの類似性、CCのストローク幅の類似性及びCC間の距離の制約のうちの一つ以上を使用する。

【0097】

別の実施形態によると、CCグループは、最初にハフ変換により生成される。ハフ変換の結果、1つの行上に中心があるCCは共にグループ化される。次に、生成されたCCグループに含まれたCCは、上記の制約を使用してフィルタリングされる。

40

【0098】

ステップ6203において、明るいCC及び暗いCCが、それぞれ、ステップ6202でグループ化された後、明るいCCの集合及び暗いCCの集合は、空間関係及び/又は外観の類似性に基づいて組み合わせられる。

【0099】

ステップ6203でCCグループを組み合わせるために一般的に使用される特徴は、例えば、2つのある特定のグループのバウンディングボックスの重複率、サイズの類似性(サイズの差は、2つのグループの高さの最大値より低いことが好ましい)及び行方向の類似性(方向の差は、30度より小さいことが好ましい)を含む。上記の特徴の一つ又はあ

50

らゆる組合せが実際に使用されてもよい。

【0100】

ステップ6203の完了後、各TBR及び外側領域のCCグループがそれぞれ形成されている。図6のステップ630に進む。

【0101】

ステップ630において、各TBR及び外側領域間の種々の領域からのCCグループは、組み合わされてテキスト領域を生成する。

【0102】

このステップにおいて、種々の領域からのCCグループは、行方向の整合性、CCの平均サイズの類似性、CCの平均ストローク幅の類似性及びCCの平均的な色又はグレースケールの類似性のうちの少なくとも1つに基づいて組み合わされる。

10

【0103】

種々の領域からのCCグループを組み合わせる規則は、1つの領域からのCCグループをグループ化する規則より厳しくてよいことが好ましい。種々の領域からのCCグループが互いに重複しないため、バウンディングボックスの重複率は使用されなくてもよい。

【0104】

CCグループ化ステップ330は、ステップ630が完了する時に完了する。

【0105】

図8Eは、CCグループ化ステップ330の結果を示す。候補となるテキストCCは、同一の文字列に属するCCが白線を使用して接続されるテキスト行/単語にグループ化されていることが図8Eからわかるだろう。

20

【0106】

テキスト検出方法300は、CCグループ化ステップ330が完了する時に終了する。

【0107】

図8Fは、テキスト検出方法300の結果を示す。テキスト行/単語のバウンディングボックスに基づくテキスト領域は、入力画像から検出されていることが分かりうる。最後に検出されたテキスト領域は、白線のボックスで示される。

【0108】

次に、本発明の一実施形態に係る画像においてテキスト領域を検出するテキスト検出装置900のブロック図を示す図9を参照する。装置900は、図3～図7を参照して説明した方法を実現するために使用されうる。簡潔にするために、図3～図7を参照して説明したものに類似するいくつかの詳細をここでは省略する。しかし、これらの詳細も装置900に適用可能であってもよいことが理解されるだろう。

30

【0109】

図9に示されるような一実施形態によると、テキスト検出装置900は、テキスト背景領域(TBR)検出ユニット910と、CCフィルタリングユニット920と、CCグループ化ユニット930とを備える。

【0110】

テキスト背景領域(TBR)検出ユニット910は、画像からTBRを検出するように構成されうる。

40

【0111】

CCフィルタリングユニット920は、少なくとも1つのCCをフィルタリングして少なくとも1つの候補となるテキストCCを確保するように構成されうる。

【0112】

CCグループ化ユニット930は、TBR検出ユニットにおいて検出されたTBRに基づいて少なくとも1つの候補となるテキストCCをグループ化して少なくとも1つのCCグループを形成し、且つ少なくとも1つのCCグループに基づいて少なくとも1つのテキスト領域を生成するように構成されうる。

【0113】

本明細書において、TBRは、画像中のテキストの周囲領域として規定されてよく、規

50

則的な境界線及び均一な色又はグレースケールを有する。

【0114】

一実施形態によると、TBR検出ユニット910は、少なくとも1つのCCの特徴、少なくとも1つのCCの境界内に配置されたCCであり、少なくとも1つのCCに対して高いコントラストを有する少なくとも1つのCC中のメンバCCの統計、及び少なくとも1つのCCと他のTBRとの間の関係のうちの少なくとも1つに基づいて、少なくとも1つのCCからTBRを選択するように構成されうる。

【0115】

例えば、少なくとも1つのCCの特徴は、CCの色又はグレースケールの均一性、CCのサイズ、CCの形状、CCの境界線の規則性、画像中のCCの位置、CCの平均グレースケール値及びCCのグレースケール値分布のうちの少なくとも1つを含みうる。

10

【0116】

例えば、メンバCCの統計は、CC中のメンバCCの数、第1の事前定義済みの閾値より高いテキスト信頼度を有するメンバCCのシードCCの数、CC中のメンバCCの平均テキスト信頼度及びCC中のメンバCCの総面積とCCの面積との比のうちの少なくとも1つを含みうる。

【0117】

例えば、少なくとも1つのCCと他のTBRとの間の関係に基づいて少なくとも1つのCCからTBRを選択することは、前に判定されたあらゆるTBRにおけるメンバCCではなく、且つ前に判定されたTBRと同一のメンバCCを有さないCCにตอบสนองして、CCをTBRとして判定することを含みうる。これは、TBR検出ユニット910により図4に示されたようなフローチャートを実行することで実現されうる。

20

【0118】

一実施形態によると、CCフィルタリングユニット920において、あらゆるTBRの境界内にないCCに対するフィルタリングは、あらゆるTBRの境界内のCCに対するフィルタリングより厳しくてよい。

【0119】

一実施形態によると、CCフィルタリングユニット920は、以下の条件、すなわちCCのサイズ、CCの形状、CCのバウンディングボックスのアスペクト比、CCとそのバウンディングボックスの面積比、周長とCCの面積との比及びCCのテクスチャ特徴のうちの少なくとも1つに基づいて、あらゆるTBRの境界内のCC及びあらゆるTBRの境界内にないCCをフィルタリングするように構成される。CCフィルタリングユニット920は、更に以下の条件、すなわちストローク幅の統計及びCCの境界線画素数とCCの画素数との比の少なくとも一方に基づいて、あらゆるTBR内にないCCをフィルタリングするように構成される。

30

【0120】

図10Aは、本発明の一実施形態に係る図9のCCフィルタリングユニットを示すブロック図である。

【0121】

図10Aに示されるように、一実施形態によると、CCフィルタリングユニット920は、TBR以外の少なくとも1つのCCの各々のテキスト信頼度を算出するように構成された算出ユニット1010であり、計算において、あらゆるTBRの境界内のCCが他のCCより重視される算出ユニット1010と、事前定義済みの閾値より高いテキスト信頼度を有するCCをテキスト候補CCとして判定するように構成された判定ユニット1020とを備える。

40

【0122】

図10Bは、本発明の別の実施形態に係る図9のCCフィルタリングユニットを示すブロック図である。

【0123】

図10Bに示されるように、別の実施形態によると、CCフィルタリングユニット9

50

20は、TBR以外の少なくとも1つのCC毎に、CCがあらゆるTBRの境界内に配置されることに応答して、CCを第1のCCとして識別するか、あるいはCCを第2のCCとして識別するように構成された識別ユニット1102と、CCがテキスト候補CCであるかを判定するために、第1のCC及び第2のCCの各々に対して第1のフィルタリングステップを実行するように構成された第1のフィルタリングユニット1104と、CCがテキスト候補CCであるかを更に判定するために、第1のフィルタリングステップによりテキスト候補CCとして判定される第2のCCの各々に対して第2のフィルタリングステップを実行するように構成された第2のフィルタリングユニット1106とを備える。

【0124】

一実施形態によると、第1のフィルタリングユニット1104は、CCがテキスト候補CCであるかを判定するように、CCの1つ以上の第1の特徴に基づいて第1のフィルタリングステップを実行するように更に構成されうる。

10

【0125】

一実施形態によると、第2のフィルタリングユニット1106は、CCがテキスト候補CCであるかを更に判定するように、CCの1つ以上の第2の特徴に基づいて第2のフィルタリングステップを実行するように更に構成されうる。

【0126】

図11Aは、本発明の一実施形態に係る図9のCCグループ化ユニット930を示すブロック図である。

【0127】

一実施形態によると、CCグループ化ユニット930は、割り当てユニット9301と、グループ化ユニット9302と、第1の組み合わせユニット9303とを更に備えうる。

20

【0128】

割り当てユニット9301は、候補となるテキストCCをそれぞれのテキスト背景領域に割り当て、且つTBRに割り当て不可能な候補となるテキストCCを外側領域に割り当てるように構成されうる。

【0129】

グループ化ユニット9302は、各TBR及び外側領域のCCをそれぞれグループ化してCCグループを形成するように構成されうる。

30

【0130】

第1の組み合わせユニット9303は、各TBR及び外側領域間の種々の領域からのCCグループを組み合わせる前記少なくとも1つのテキスト領域を生成するように構成されうる。

【0131】

一実施形態によると、第1の組み合わせユニット9303は、以下の条件、すなわち行方向の整合性、CCグループのグループバウンディングボックスの重複率、CCの平均サイズの類似性、CCの平均ストローク幅の類似性及びCCの平均的な色又はグレースケールの類似性のうちの少なくとも1つに基づいて、種々の領域からのCCグループを組み合わせるように構成されうる。

40

【0132】

図11Bは、本発明の一実施形態に係る図11Aのグループ化ユニット9302を示すブロック図である。

【0133】

一実施形態によると、グループ化ユニット9302は、分割ユニット9302-1と、生成ユニット9302-2と、第2の組み合わせユニット9302-3とを更に備えうる。

【0134】

分割ユニット9302-1は、各TBR及び外側領域のCCを暗い領域の明るいCCの集合及び明るい領域の暗いCCの集合に分割するように構成されうる。

50

【 0 1 3 5 】

生成ユニット 9 3 0 2 - 2 は、それぞれ明るい C C の集合及び暗い C C の集合内に C C グループを生成するように構成されうる。

【 0 1 3 6 】

第 2 の組み合わせユニット 9 3 0 2 - 3 は、空間関係及び外観の類似性のうちの少なくとも 1 つに基づいて、明るい C C の集合と暗い C C の集合とを組み合わせるように構成されうる。

【 0 1 3 7 】

一実施形態によると、生成ユニット 9 3 0 2 - 2 において、C C グループは C C クラスタリングにより生成されうる。C C クラスタリングは、以下の制約、すなわちある特定の方向に従う C C の中心の位置合わせ、C C のサイズの類似性、C C の形状の類似性、C C の色又はグレースケールの類似性、C C のストローク幅の類似性及び C C 間の距離のうちの少なくとも 1 つを使用してよい。

10

【 0 1 3 8 】

一実施形態によると、生成ユニット 9 3 0 2 - 2 は、ハフ変換により C C グループを生成し、且つ以下の制約、すなわちある特定の方向に従う C C の中心の位置合わせ、C C のサイズの類似性、C C の形状の類似性、C C の色又はグレースケールの類似性、C C のストローク幅の類似性及び C C 間の距離のうちの少なくとも 1 つを使用して生成された C C グループに含まれた C C をフィルタリングするように更に構成されうる。

【 0 1 3 9 】

一実施形態によると、第 1 の組み合わせユニット 9 3 0 3 において組み合わせることに対する規則は、グループ化ユニット 9 3 0 2 においてグループ化することに対する規則より厳しくてよい。

20

【 0 1 4 0 】

本発明に係るテキスト検出の方法及び装置は、種々の適応例を有する。例えばそれは、カメラが取り込んだ画像又は映像からテキスト情報を自動的に抽出する際に使用されうる。

【 0 1 4 1 】

図 1 2 は、本発明の一実施形態に係るテキスト情報抽出方法を示す。

【 0 1 4 2 】

図 1 2 に示されるように、ブロック 1 2 1 0 において、入力画像又は入力映像からのテキスト領域は、図 3 ~ 図 7 を参照して説明したテキスト検出方法に係るテキスト検出方法を使用して検出される。

30

【 0 1 4 3 】

ブロック 1 2 2 0 において、テキストは、検出されたテキスト領域から抽出されうる。選択的に、ブロック 1 2 4 0 において示されるように、入力映像中のテキストは、入力映像からテキスト領域を検出する際に追跡されうる。

【 0 1 4 4 】

ブロック 1 2 3 0 において、テキスト認識は、抽出されたテキストに対して実行されてテキスト情報を取得しうる。

40

【 0 1 4 5 】

次に、本発明の一実施形態に係るテキスト情報抽出システム 1 3 0 0 のブロック図を示す図 1 3 を参照する。システム 1 3 0 0 は、図 1 2 を参照して説明した方法を実現するために使用されうる。

【 0 1 4 6 】

図 1 3 に示されるように、システム 1 3 0 0 は、テキスト検出装置 1 3 1 0 と、抽出装置 1 3 2 0 と、認識装置 1 3 3 0 とを備える。

【 0 1 4 7 】

テキスト検出装置 1 3 1 0 は、入力画像又は入力映像からテキスト領域を検出するように構成され、且つ図 9 に関連して説明した装置 9 1 0 と同一であってよい。

50

【 0 1 4 8 】

抽出装置 1 3 2 0 は、検出されたテキスト領域からテキストを抽出するように構成される。

【 0 1 4 9 】

認識装置 1 3 3 0 は、抽出されたテキストを認識してテキスト情報を取得するように構成される。

【 0 1 5 0 】

選択的に、システム 1 3 0 0 は追跡装置 1 3 4 0 を更に備える。追跡装置 1 3 4 0 は、テキスト検出装置 1 3 1 0 が入力映像からテキスト領域を検出するように構成される際に入力映像中のテキストを追跡するように構成される。

10

【 0 1 5 1 】

図 9 ~ 図 1 1 及び図 1 3 に関連して上述したユニット及び装置は、種々のステップを実現する例示的なモジュール及び / 又は好ましいモジュールであることが理解されるだろう。モジュールは、ハードウェアユニット (例えば、プロセッサ又は特定用途向け集積回路等) 及び / 又はソフトウェアモジュール (例えば、コンピュータプログラム) であってよい。種々のステップを実現するモジュールは、完全に上述されていない。しかし、ある特定の処理を実行するステップがある場合、同一の処理を実現する対応する機能モジュール又は機能ユニット (ハードウェア及び / 又はソフトウェアにより実現された) があってもよい。上述及び後述のステップとこれらのステップに対応するユニットとの全ての組合せが構成する技術的解決法が完全で且つ適用可能である限り、それらによる技術的解決方法は本発明の開示内容に含まれる。

20

【 0 1 5 2 】

また、種々のユニットにより構成された上述の装置及びシステムは、機能モジュールとしてコンピュータ等のハードウェアデバイスに組み込まれる。当然、コンピュータは、これらの機能モジュールに加えて、他のハードウェアコンポーネント又はソフトウェアコンポーネントを有する。

【 0 1 5 3 】

本発明の方法、装置及びシステムは、多くの方法で実行可能である。例えば、本発明の方法及び装置は、ソフトウェア、ハードウェア、ファームウェア又はそれらのあらゆる組合せにより実行可能である。方法のステップの上述の順序は例示することのみを意図し、特に指示のない限り、本発明の方法のステップは特に上述された順序に限定されない。それに加えて、いくつかの実施形態において、本発明は、本発明に係る方法を実現する機械可読命令を含む記録媒体に記録されたプログラムとしても実施されてもよい。従って、本発明は、本発明に係る方法を実現するプログラムを格納する記録媒体も範囲に含む。

30

【 0 1 5 4 】

例を用いて本発明のいくつかの特定の実施形態を詳細に実証したが、上述の例は、本発明の範囲を限定することではなく、例示することのみを意図することが当業者により理解されるべきである。上述の実施形態は、本発明の範囲及び趣旨から逸脱することなく変更可能であることが当業者により理解されるべきである。本発明の範囲は、添付の特許請求の範囲により規定される。

40

【図1】

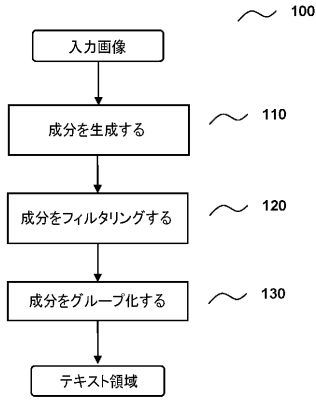


Fig. 1

【図2】

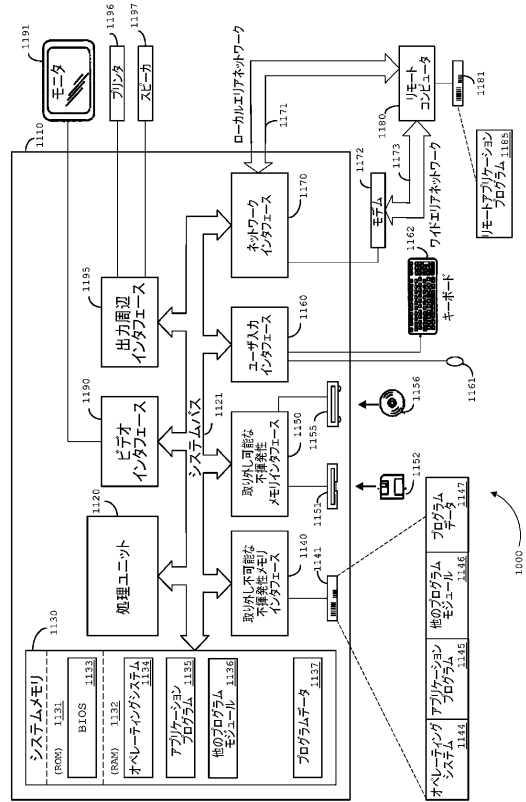


Fig. 2

【図3】

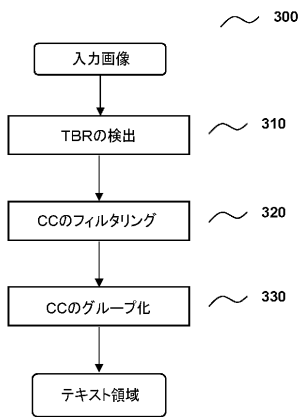


Fig. 3

【図4】

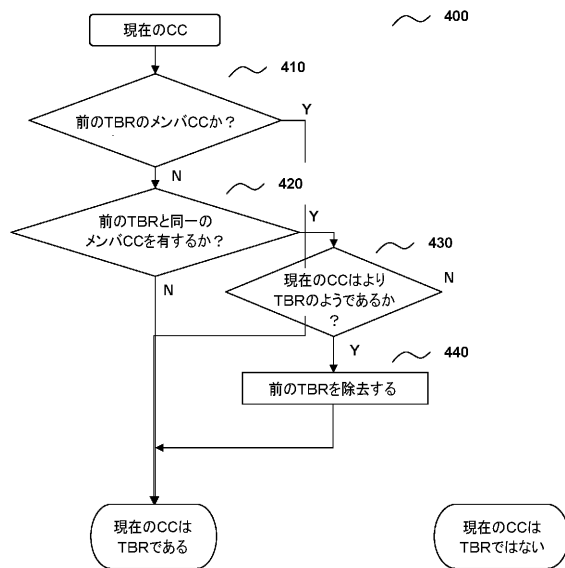


Fig. 4

【図5A】

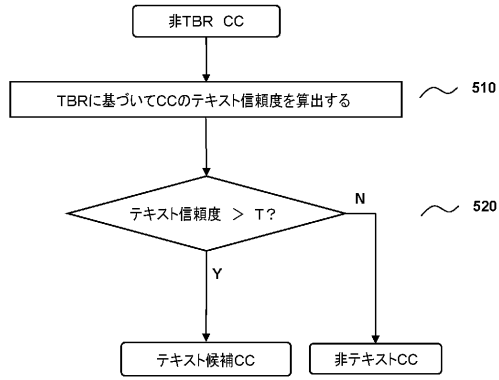


Fig. 5A

【図5B】

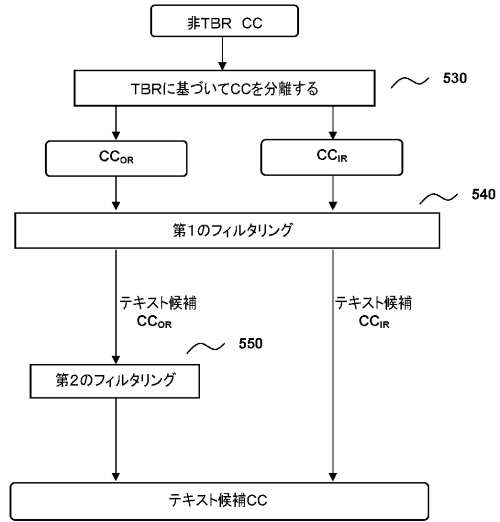


Fig. 5B

【図5C】

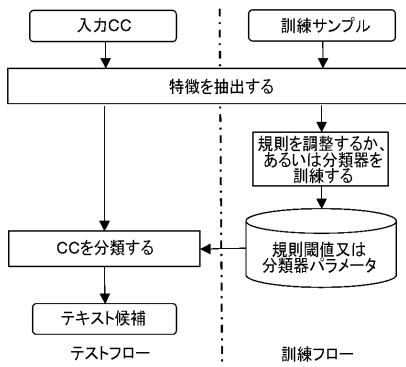


Fig. 5C

【図6】

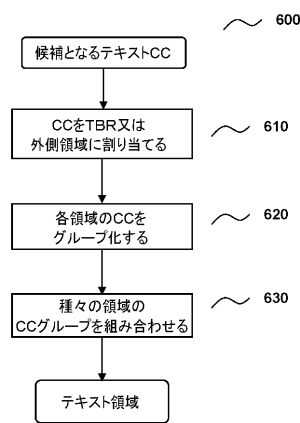


Fig. 6

【図7】

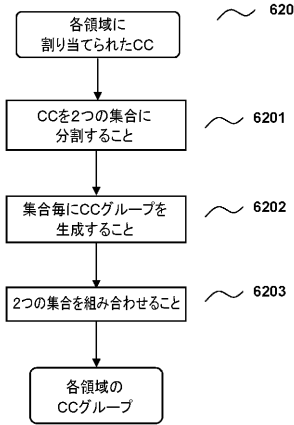


Fig. 7

【図9】

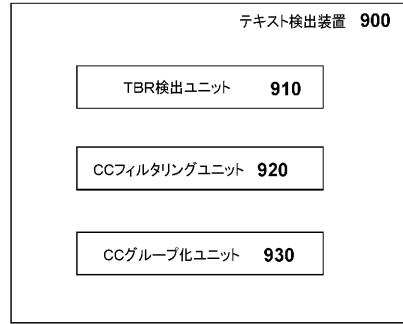


Fig. 9

【図10A】

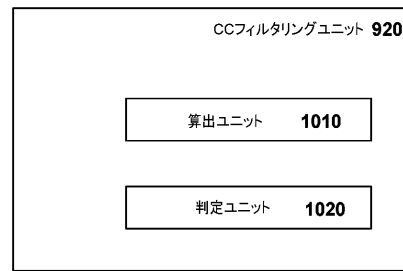


Fig. 10A

【図10B】

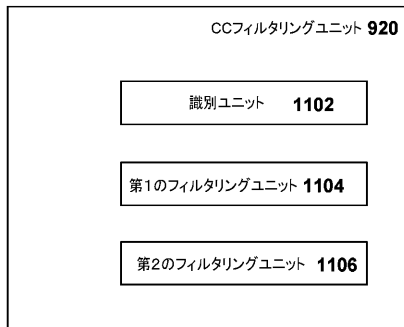


Fig. 10B

【図11B】

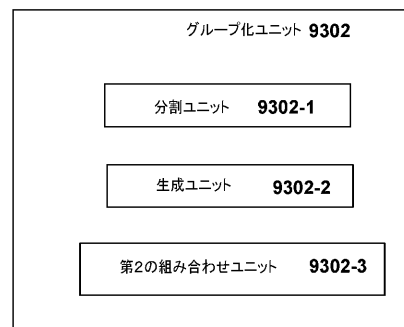


Fig. 11B

【図11A】

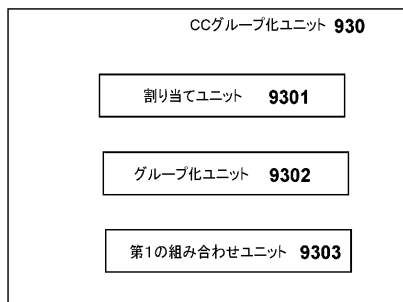


Fig. 11A

【図12】

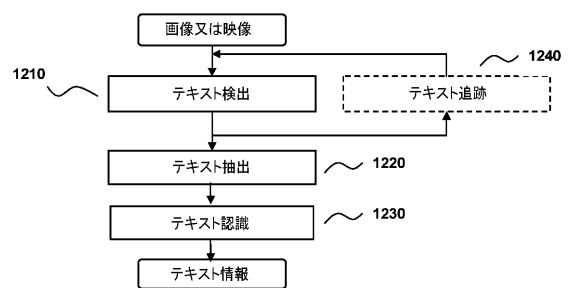


Fig. 12

【 図 1 3 】

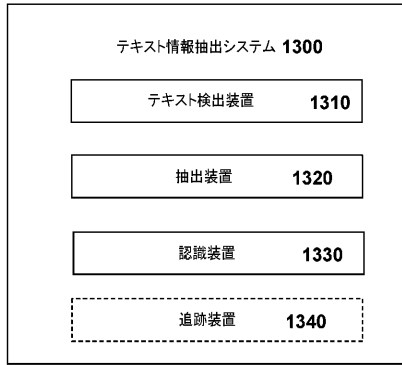


Fig. 13

【 図 8 A 】



Fig. 8A

【 8 B 】



Fig. 8B

【 8 C 】



Fig. 8C

【 8 D 】

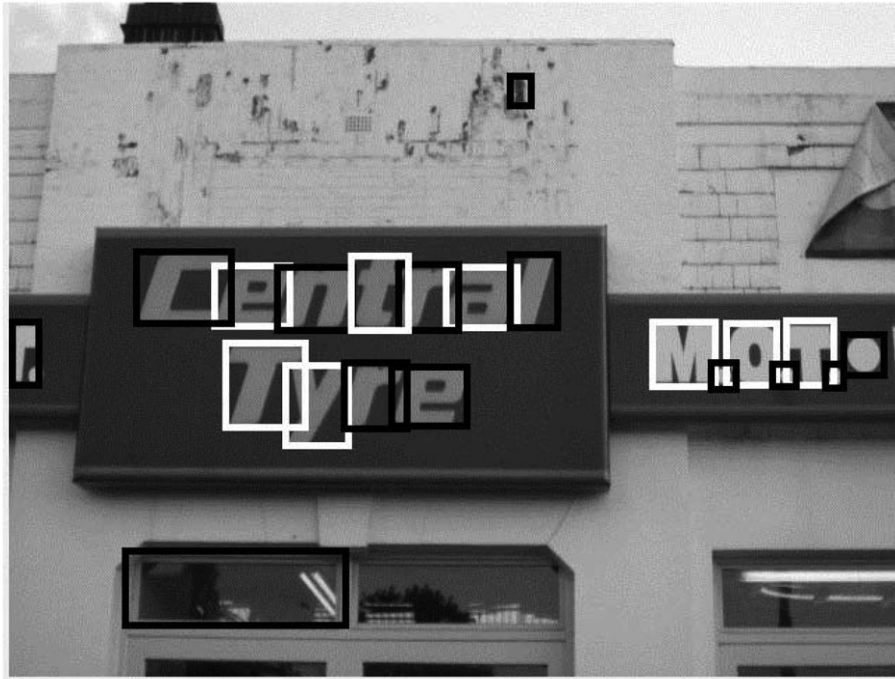


Fig. 8D

【 8 E 】



Fig. 8E

【 8 F 】



Fig. 8F

フロントページの続き

(72)発明者 マー ウェンフォア

中華人民共和国 ベイジン 100080, ハイディアン ディストリクト, ベイスホワン
ロード, ナンバー9, イング ビルディング, エフ12エー キヤノン インフォメーショ
ン テクノロジー(ベイジン)コーポレーション リミテッド内

(72)発明者 ツァオハイ ルオ

中華人民共和国 ベイジン 100080, ハイディアン ディストリクト, ベイスホワン
ロード, ナンバー9, イング ビルディング, エフ12エー キヤノン インフォメーショ
ン テクノロジー(ベイジン)コーポレーション リミテッド内

審査官 佐田 宏史

(56)参考文献 特開平11-238098(JP, A)

特開平07-093476(JP, A)

特開2009-123206(JP, A)

特開2004-127203(JP, A)

特開平11-239363(JP, A)

特開2010-067223(JP, A)

特開2010-277227(JP, A)

特開2011-139366(JP, A)

特開2012-222581(JP, A)

米国特許出願公開第2011/0069885(US, A1)

北田 英樹、若原 徹, “エッジ成分の空間的配置に着目した情景画像からの文字列抽出”, 情
報処理学会 シンポジウム 画像の認識・理解シンポジウム(MIRU) 2011, 日本, 情
報処理学会, 2011年 7月20日, pp.1502-1507

(58)調査した分野(Int.Cl., DB名)

G06T 7/00-7/60