

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6188976号

(P6188976)

(45) 発行日 平成29年8月30日 (2017. 8. 30)

(24) 登録日 平成29年8月10日 (2017. 8. 10)

(51) Int. Cl.		F I	
G06T	7/44	(2017.01)	G06T 7/44
G06T	7/00	(2017.01)	G06T 7/00 350B
G06K	9/20	(2006.01)	G06K 9/20 340J
			G06K 9/20 340L

請求項の数 20 (全 17 頁)

(21) 出願番号	特願2017-18957 (P2017-18957)	(73) 特許権者	517038176
(22) 出願日	平成29年2月3日 (2017. 2. 3)		株式会社ストラッドビジョン
(65) 公開番号	特開2017-138989 (P2017-138989A)		大韓民国 37673 ギョンサンブクト
(43) 公開日	平成29年8月10日 (2017. 8. 10)		ポハンシ ナムグ チョンアムロ 67
審査請求日	平成29年2月6日 (2017. 2. 6)		コンプレックス 4, 4427
(31) 優先権主張番号	15/014, 441	(74) 代理人	100121728
(32) 優先日	平成28年2月3日 (2016. 2. 3)		弁理士 井関 勝守
(33) 優先権主張国	米国 (US)	(74) 代理人	100165803
			弁理士 金子 修平
早期審査対象出願		(72) 発明者	チョ ホジン
			大韓民国 37673 ギョンサンブクト
			ポハンシ ナムグ チョンアムロ 77
			、106
		審査官	佐田 宏史
			最終頁に続く

(54) 【発明の名称】 画像に含まれるテキストを検出する方法、装置及びコンピュータ読み取り可能な記録媒体

(57) 【特許請求の範囲】

【請求項 1】

画像に含まれるテキストを検出する方法であって、

(a) 入力画像が取得されると、装置が、前記入力画像に含まれているピクセルのフィーチャー値を参照にして、前記入力画像の少なくとも1つのテキスト候補を検出したり、検出するようにサポートするステップ、

(b) 前記装置が、前記検出されたテキスト候補が含まれる領域内で選択された少なくとも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、(i) 前記検出されたテキストの候補を第1閾値と比較して強いテキスト(strong text)または強くないテキスト(non-strong text)に分類したり、分類するようにサポートして、(ii) 前記強くないテキスト(non-strong text)に分類されたテキスト候補を第2閾値と比較して、弱いテキスト(weak text)またはノンテキスト(non-text)に分類したり、分類するようにサポートしているステップ、及び

(c) 前記装置が、前記強いテキスト(strong text)と前記弱いテキスト(weak text)の関連情報を参照して前記弱いテキスト(weak text)を前記強いテキスト(strong text)に分類したり、分類するようにサポートするかどうかを判断するステップを含み、

前記(c)ステップは、

前記装置が、前記強いテキスト(strong text)と前記弱いテキスト(weak text)の関連情報を参照して前記弱いテキスト(weak text)を前記強いテキスト(strong text)またはノンテキスト(non-text)に分類したり、分類するようにサポートし、

前記（a）ステップで、

前記装置は、前記入力画像から所定の領域に含まれるすべてのピクセルのフィーチャー値が前記所定の領域の外側の境界領域のフィーチャー値よりも大きいまたは小さいとき、前記所定の領域を1つのテキスト候補として検出したり、検出するようにサポートし、

前記装置は、前記所定の領域のピクセル輝度レベルを変更しながら、前記少なくとも一つのテキスト候補を検出したり、検出するようにサポートし、

前記検出された少なくとも一つのテキスト候補から、（i）前記所定の領域のピクセル輝度レベルが相対的に大きい下位レベルのテキスト候補と（ii）前記下位レベルのテキスト候補の領域を含み、前記所定の領域のピクセル輝度レベルが比較的小さい上位レベルのテキスト候補のうち、前記所定の領域のピクセル輝度レベルが最大のテキスト候補を、それぞれ子ノードと親ノードに割り当て、木構造を構成したり、構成するようにサポートすることを特徴とする、テキストを検出する方法。

10

【請求項2】

前記検出されたテキストの候補が複数のとき、前記（a）ステップの後に、（a1）前記装置が、前記検出された複数のテキスト候補が含まれる領域が所定の割合以上重なり合うテキスト候補を含む少なくとも一つのテキスト候補セットを識別し、識別された各テキストの候補セットに対して所定の条件を満たしているテキストの候補を選択したり、選択するようにサポートするステップをさらに含み、

前記（b）ステップは、

前記所定の条件を満たしているテキスト候補が含まれている領域内で選択された少なくとも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、（i）前記所定の条件を満たしているテキスト候補を前記第1閾値と比較して、前記強いテキスト(strong text)または前記強くないテキスト(non-strong text)に分類したり、分類するようにサポートして、（ii）前記強くないテキスト(non-strong text)に分類されたテキストの候補を前記第2しきい値と比較して、前記弱いテキスト(weak text)または前記ノンテキスト(non-text)に分類したり、分類するようにサポートすることを特徴とする、請求項1に記載のテキストを検出する方法。

20

【請求項3】

前記木構造から特定のノードに対して、前記特定のノードに対応するテキスト候補が含まれる領域と、前記特定のノードに対応するテキスト候補の領域を含み、前記所定の領域のピクセル輝度レベルが比較的小さい、祖先ノードに対応するテキスト候補が含まれる領域が所定割合以上に重なる場合が所定回数以上である場合、

30

前記装置は、自分の特定の祖先ノードに対しテキスト候補が含まれる領域の変化量が比較的最も小さなノードに対応するテキスト候補を選択したり、選択するようにサポートして、

前記（b）ステップは、

前記選択されたテキストの候補が含まれる領域内で選択された少なくとも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、（i）前記所定の条件を満たしているテキストの候補を前記第1閾値と比較して、前記強いテキスト(strong text)または前記強くないテキスト(non-strong text)に分類したり、分類するようにサポートして、（ii）前記強くないテキスト(non-strong text)に分類されたテキスト候補を前記第2しきい値と比較して、前記弱いテキスト(weak text)または前記ノンテキスト(non-text)に分類したり、分類するようにサポートすることを特徴とする、請求項1に記載のテキストを検出する方法。

40

【請求項4】

（d）前記装置が、前記第1閾値と比較して分類された強いテキスト(strong text)及び前記弱いテキスト(weak text)から編入された強いテキスト(strong text)をグルーピングしたり、グルーピングするようサポートするステップをさらに含むことを特徴とする、請求項1に記載のテキストを検出する方法。

【請求項5】

50

前記装置は、前記第3強いテキスト(strong text)を前記第1強いテキスト(strong text)及び前記第2強いテキスト(strong text)と一緒にグルーピングしたり、グルーピングするようサポートすることを特徴とする、請求項4に記載のテキストを検出する方法。

前記 (b) ステップは、

10

前記 (b) ステップで、

20

前記 (b) ステップで、

前記装置は、AdaBoostアルゴリズムでトレーニングされた分類機を用いて前記検出されたテキスト候補を分類したり、分類するようにサポートすることを特徴とする、請求項1に記載のテキストを検出する方法。

前記(c)ステップで、

前記の関連情報は、前記強いテキスト(strong text)と前記弱いテキスト(weak text)の位置情報、サイズ情報、カラー情報、ストローク幅情報の少なくとも一部を含むことを特徴とする、請求項1に記載のテキストを検出する方法。

30

前記検出されたテキスト候補が含まれる領域は、

エキストラ マージン(extra margin)を最小限に抑えながら前記検出されたテキストの候補を含むバウンディングボックスの領域であることを特徴とする、請求項1に記載のテキストを検出する方法。

画像に含まれるテキストを検出する装置において、

入力画像を取得する通信部、及び

40

50

t)に分類したり、分類するようにサポートするかどうかを判断するプロセッサを含み、
前記 (i i i) で、

前記プロセッサは、前記強いテキスト(strong text)と前記弱いテキスト(weak text)の
関連情報を参照して前記弱いテキスト(weak text)を前記強いテキスト(strong text)または
ノンテキスト(non-text)に分類したり、分類するようにサポートし、

前記 (i) で、

前記プロセッサは、前記入力画像から所定の領域に含まれるすべてのピクセルのフィー
チャー値が前記所定の領域の外側の境界領域のフィーチャー値より大きい小さいとき、
前記所定の領域を1つのテキスト候補として検出したり、検出するようにサポートし、

前記プロセッサは、前記所定の領域のピクセル輝度レベルを変更しながら、前記少なく
とも一つのテキスト候補を検出したり、検出するようにサポートし、

前記検出された少なくとも一つのテキスト候補から、(i) 前記所定の領域のピクセル
輝度レベルが相対的に大きい下位レベルのテキスト候補と (i i) 前記下位レベルのテキ
スト候補の領域を含み、前記所定の領域のピクセル輝度レベルが比較的小さい上位レ
ベルのテキスト候補のうち、前記所定の領域のピクセル輝度レベルが最大のテキスト候
補を、それぞれ子ノードと親ノードに割り当て、木構造を構成したり、構成するよう
にサポートする

ことを特徴とする、テキストを検出する装置。

【請求項 1 2】

前記検出されたテキストの候補が複数のとき、前記 (i) の後に、前記プロセッサが、
前記検出された複数のテキスト候補が含まれる領域が所定割合以上に重なり合うテキ
スト候補を含む少なくとも一つのテキスト候補セットを識別し、識別された各テキ
スト候補セットに対して所定の条件を満たしているテキストの候補を選択したり、選
択するようにサポートして、

前記所定の条件を満たしているテキスト候補が含まれる領域内で選択された少なく
とも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、(i) 前
記所定の条件を満たしているテキストの候補を前記第1閾値と比較して、前記強い
テキスト(strong text)または前記強くないテキスト(non-strong text)に分類
したり、分類するようにサポートして、(i i) 前記強くないテキスト(non-strong
text)に分類されたテキストの候補を前記第2閾値と比較して、前記弱いテキ
スト(weak text)または前記ノンテキスト(non-text)に分類したり、分類する
ようにサポートすることを特徴とする、請求項 1 1 に記載のテキストを検出する
装置。

【請求項 1 3】

前記木構造から特定のノードに対して、前記特定のノードに対応するテキスト候補
が含まれる領域と、前記特定のノードに対応するテキスト候補の領域を含み、前記
所定の領域のピクセル輝度レベルが比較的小さい、祖先ノードに対応するテキ
スト候補が含まれる領域が所定割合以上に重なる場合が所定回数以上である場合、

前記プロセッサが、自分の特定の祖先ノードに対しテキスト候補が含まれる領域
の変化量が比較的最も小さなノードに対応するテキスト候補を選択したり、選択
するようにサポートして、

前記選択されたテキストの候補が含まれる領域内で選択された少なくとも一つのピ
クセルのフィーチャー値またはこれを変換した値を参照して、(i) 前記所定の条件
を満たすテキスト候補を前記第1閾値と比較して、前記強いテキスト(strong
text)または前記強くないテキスト(non-strong text)に分類したり、分類する
ようにサポートして、(i i) 前記強くないテキスト(non-strong text)に分類
されたテキストの候補を前記第2閾値と比較して、前記弱いテキスト(weak
text)または前記ノンテキスト(non-text)に分類したり、分類するよう
にサポートすることを特徴とする、請求項 1 1 に記載のテキストを検出する
装置。

【請求項 1 4】

前記プロセッサが、前記 (i i i) の後に (i v) 前記第1閾値と比較して分類された

10

20

30

40

50

強いテキスト(strong text)及び前記弱いテキスト(weak text)から編入された強いテキスト(strong text)をグルーピングしたり、グルーピングするようにサポートしていることを特徴とする、請求項 1 1 に記載のテキストを検出する装置。

【請求項 1 5】

第 1 強いテキスト(strong text)及び第 1 弱いテキスト(weak text)から編入された第 2 強いテキスト(strong text)がグルーピングされて前記第 2 強いテキスト(strong text)と第 2 弱いテキスト(weak text)の関連情報を参照して前記第 2 弱いテキスト(weak text)が第 3 強いテキスト(strong text)に分類されると、

前記プロセッサは、前記第 3 強いテキスト(strong text)を前記第 1 強いテキスト(strong text)及び前記第 2 強いテキスト(strong text)と一緒にグルーピングしたり、グルーピングするようにサポートすることを特徴とする、請求項 1 4 に記載のテキストを検出する装置。

10

【請求項 1 6】

前記 (i i) で、

前記第 1 閾値と前記第 2 閾値の間に少なくとも一つの間閾値があるとき、(i i - 1) 前記強くないテキスト(non-strong text)に分類されたテキストの候補を、前記の間閾値と比較して、前記強いテキスト(strong text)または次順位弱いテキスト(weak text)に分類したり、分類するようにサポートして、(i i - 2) 前記次順位弱いテキスト(weak text)を前記第 2 閾値と比較して、前記弱いテキスト(weak text)または前記ノンテキスト(non-text)に分類したり、分類するようにサポートすることを特徴とする、請求項 1 1 に記載のテキストを検出する装置。

20

【請求項 1 7】

前記 (i i) で、

前記プロセッサは、前記検出されたテキスト候補が含まれる領域内で選択されたピクセルのフィーチャー値をMLBP(Mean Local Binary Pattern)に変換した値を参照にして、前記分類の作業を実行したり、実行するようにサポートしていることを特徴とする、請求項 1 1 に記載のテキストを検出する装置。

【請求項 1 8】

前記 (i i) で、

前記プロセッサは、AdaBoostアルゴリズムでトレーニングされた分類機を用いて前記検出されたテキスト候補を分類したり、分類するようにサポートすることを特徴とする、請求項 1 1 に記載のテキストを検出する装置。

30

【請求項 1 9】

前記 (i i i) で、

前記の関連情報は、前記強いテキスト(strong text)と前記弱いテキスト(weak text)の位置情報、サイズ情報、カラー情報、ストローク幅情報の少なくとも一部を含むことを特徴とする、請求項 1 1 に記載のテキストを検出する装置。

【請求項 2 0】

前記検出されたテキスト候補が含まれる領域は、

エキストラ マージン(extra margin)を最小限に抑えながら前記検出されたテキストの候補を含むバウンディングボックスの領域であることを特徴とする、請求項 1 1 に記載のテキストを検出する装置。

40

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、画像に含まれるテキストを検出する方法、装置、及びコンピュータ読み取り可能な記録媒体に関するもので、より詳細には、入力画像が取得されると、前記入力画像に含まれているピクセルのフィーチャー値を参照にして、前記入力画像から少なくとも 1 つのテキスト候補を検出したり、検出するようにサポートするプロセス、前記検出された

50

テキスト候補が含まれる領域内で選択された少なくとも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、(i) 前記検出されたテキスト候補を第1閾値と比較して強いテキスト (strong text) または強くないテキスト (non-strong text) に分類したり、分類するようにサポートして、(i i) 前記 強くないテキスト (non-strong text) に分類されたテキスト候補を第2閾値と比較して弱いテキスト (weak text) またはノンテキスト (non-text) に分類したり、分類するようにサポートするプロセス、及び前記強いテキスト (strong text) と前記弱いテキスト (weak text) の関連情報を参照して前記 弱いテキスト (weak text) を前記強いテキスト (strong text) に分類したり、分類するようにサポートするかどうかを判断するプロセスを含む方法、装置及びコンピュータ読み取り可能な記録媒体に関するものである。

10

【背景技術】

【0002】

画像内のテキストは、有意義な情報を提供する場合が多いので、シーンのテキスト (scenetext) を検出し、認識することは画像及びビデオの検索 (image and video retrieval)、多言語翻訳、automotiveassistanceなど、さまざまなコンピュータビジョンアプリケーションで重要と考えられてきた。

【0003】

シーンテキスト検出アルゴリズム (scene text detection algorithm) は、画像上でのテキスト (文字) を検出するためのアルゴリズムで、テキストの候補を抽出する方法に応じて大きくスライディングウィンドウ (sliding window) 方式と連結成分分析 (connected component analysis、connected component labellingとも呼ばれる) 方式とに分割することができる。

20

【0004】

スライディングウィンドウ方式は、ウィンドウをイメージの全ての場所から複数のスケールでシフトさせ、シーンの画像のテキストを検出する技術である。この技術は、入力画像に対して徹底的に検索が行われるため、どれだけ多くのテキスト領域が検出されるかを示す要精検率 (recallrate) が高いという長所がある。一方、徹底したウィンドウスキャンにより、多くの計算量を避けることができず、多くのテキスト候補が大量の偽陽性 (false positive) を引き起こし得る。したがって、リアルタイムアプリケーションには不適合である。スライディングウィンドウ (sliding window) 方式は、X.Chen and AL Yuille .Detecting and reading text in natural scenes. In Proc. CVPR 2004、pages 366-373、2004などで紹介されたことがある。

30

【0005】

このように、スライディングウィンドウ方式が多く計算量を要するため、近年では連結成分分析方式が多く使用される傾向にある。連結成分分析方式は、入力画像から同様のテキスト属性を共有するピクセルの集合であるテキスト候補を抽出し、ノンテキスト (non-text) 候補を抑制 (suppression) するために、テキストの候補を分析 (精製) する作業を実行する。SWT (stroke width transform) とMSER (maximally stable extremal region) は、連結成分分析の代表的な技術である。この方法は、シーンのテキスト検出で最先端の性能を提供する。連結成分分析方式はB. Epshtein、E. Ofek、andY.Wexler. Detecting text in natural scenes with ストローク幅 (strokewidth) transform. In Proc. CVPR2010、pages 2963-2970、2010などで紹介されたことがある。

40

【0006】

しかし、連結成分分析方式でテキスト候補を分析 (精製) するために使用される一般的な制約 (constraints) は、複数の トゥルーテキスト (truetext) を検出するのに制限的に評価され、結果的に低い要精検率を示す限界がある。

【0007】

したがって、画像内のテキスト検出時の最適な性能を出しながらも、高い要精検率を持つことができるテキスト検出技術が要求されている。

【発明の概要】

50

【発明が解決しようとする課題】**【0008】**

本発明は、上述した問題点をすべて解決することをその目的とする。

【0009】

また、本発明は、最適なパフォーマンスを出しながらも、高い要精検率を持つことができるテキスト検出技術を提供することを他の目的とする。

【0010】

また、本発明は、多重閾値を利用して、テキスト候補を強いテキスト、弱いテキスト、ノンテキストに分類することにより、高い要精検率を持つことができるテキスト検出技術を提供することを他の目的とする。

10

【0011】

また、本発明は、テキストの候補が強いテキストと弱いテキストに分類された後、強いテキストと弱いテキストの位置情報、サイズ情報、カラー情報、ストローク幅情報の少なくとも一部を参照にして弱いテキストを強いテキストに編入させることで、高い要精検率を持つことができるテキスト検出技術を提供することを他の目的とする。

【課題を解決するための手段】**【0012】**

前記目的を達成するための本発明の代表的な構成は以下の通りである。

【0013】

本発明の一態様によれば、画像に含まれているテキストを検出する方法であって、（a）入力画像が取得されると、装置は、前記入力画像に含まれているピクセルのフィーチャー値を参照にして、前記入力画像から少なくとも1つのテキスト候補を検出したり、検出するようにサポートしているステップ、（b）前記装置が前記検出されたテキスト候補が含まれる領域内で選択された少なくとも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、（i）前記検出されたテキストの候補を第1閾値と比較して強いテキストまたは強くないテキスト（non-strong text）に分類したり、分類するようにサポートして、（i i）前記強くないテキストに分類されたテキスト候補を第2閾値と比較して弱いテキストまたはノンテキストに分類したり、分類するようにサポートしているステップ、及び（c）前記装置が、前記強いテキストと前記弱いテキストの関連情報を参照して前記弱いテキストを前記強いテキストに分類したり、分類するようにサポートするかどうかを判断するステップを含む方法が提供される。

20

30

【0014】

本発明の他の態様によれば、画像に含まれているテキストを検出する装置であって、画像を取得する通信部、及び前記画像が取得されると、（i）前記入力画像に含まれているピクセルのフィーチャー値を参照して、前記入力画像の少なくとも1つのテキスト候補を検出したり、検出するようにサポートして、（i i）前記検出されたテキスト候補が含まれる領域内で選択された少なくとも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、前記検出されたテキストの候補を第1閾値と比較して強いテキストまたは強くないテキストに分類したり、分類するようにサポートして、前記強くないテキストに分類されたテキスト候補を第2閾値と比較して弱いテキストまたはノンテキストに分類したり、分類するようにサポートしており、（i i i）前記強いテキストと前記弱いテキストの関連情報を参照して前記弱いテキストを前記強いテキストに分類したり、分類するようにサポートするかどうかを判断するプロセッサを含む装置が提供される。

40

【0015】

この他にも、本発明を実施するための他の方法、装置、及び前記の方法を実行するためのコンピュータプログラムを記録するためのコンピュータ読み取り可能な記録媒体がさらに提供される。

【発明の効果】**【0016】**

本発明によれば、次のような効果がある。

50

【0017】

本発明は、最適なパフォーマンスを出しながらも、高い要精検率を持つことができるテキスト検出技術を提供する効果がある。

【0018】

また、本発明は、多重閾値を利用して、テキスト候補を強いテキスト、弱いテキスト、ノンテキストに分類することで、テキスト検出時の高い要精検率を達成することができる効果がある。

【0019】

また、本発明は、テキストの候補が強いテキストと弱いテキストに分類された後、強いテキストと弱いテキストの位置情報、サイズ情報、カラー情報、ストローク幅情報などの関連情報を参照にして弱いテキストを強いテキストに編入させることで、テキスト検出時の高い要精検率を達成することができる効果がある。

10

【図面の簡単な説明】

【0020】

【図1】図1は、本発明の一実施形態に係る画像に含まれるテキストを検出する装置の構成を示すブロック図である。

【図2】図2は、本発明の一実施形態に係る装置は、画像に含まれているテキストを検出する過程を示すフローチャートである。

【図3】図3は、本発明の一実施形態に係るER木構造の例を示す図である。

【図4】図4は、MLBP変換の例を示す図である。

20

【図5】図5は、本発明に係るテキスト検出手法を使用して例示画像からテキストを検出する場合は、各過程での中間結果を示す図である。

【図6】図6は、他のテキスト検出技術と本発明の性能を比較実験した結果を示す図である。

【図7】図7は、他のテキスト検出技術と本発明の性能を比較実験した結果を示す図である。

【図8】図8は、他のテキスト検出技術と本発明の性能を比較実験した結果を示す図である。

【発明を実施するための形態】

【0021】

30

後述する本発明の詳細な説明は、本発明が実施されることができると特定の実施例を例示として示す添付図面を参照する。これらの実施例は、当業者が本発明を実施することができる様十分に詳細に説明されている。本発明の様々な実施例は、互いに異なるが、相互に排他的である必要はないことが理解されるべきである。例えば、ここに記載されている特定の形状、構造、及び特性は、一実施形態に関連して、本発明の精神と範囲を逸脱なく、他の実施例で実装することができる。また、それぞれの開示された実施例内の個々の構成要素の位置または配置は、本発明の精神と範囲を逸脱しなく変更されることがあることが理解されるべきである。したがって、後述する詳細な説明は、限定的な意味として取りたいものではなく、本発明の範囲は、適切に説明された場合、その請求項が主張するように均等なすべての範囲と添付された請求項によってのみ限定される。図面で同様の参照符号は、いくつかの側面に亘って同一または類似の機能を指す。

40

【0022】

本発明は、以下の基準(criteria)が考慮されたテキスト検出技術を提案する。本発明者は、下記の基準を満たしている、本発明に係る検出機をキャニーテキストデテクター(Canny Text Detector)と命名する。各基準を満たすための、本発明の構成については、当構成に対する説明で詳しく述べられる。

【0023】

1.リコール(recall)：できるだけ多くのテキスト領域が検出されなければならない。

2.プリシジョン(precision)：検出結果はノンテキスト(non-text)領域を含んでいて

50

はならない。

3.ユニークネス(uniqueness)：検出されたテキストは、一度だけ表示(mark)しなければならない。

4.コンパクトネス(compactness)：検出された領域（バウンディングボックス）は、エキストラマージン(extra margin)を最小限に抑えるように、テキスト領域を含むべきである。

【0024】

以下では、前記の基準を満たしている本発明の様々な実施形態に係る画像に含まれるテキストを検出する装置及び方法について詳細に説明する。

【0025】

図1は、本発明の一実施形態に係る画像に含まれるテキストを検出する装置の構成を示したブロック図であり、図2は、本発明の一実施形態に係る装置が、画像に含まれているテキストを検出する過程を示すフローチャートである。

【0026】

図1を参照すると、本発明の一実施形態に係る画像に含まれるテキストを検出する装置(100)は、通信部(110)、プロセッサ(120)と格納部(130)を含んでいる。

【0027】

通信部(110)は、外部装置とデータを送受信したり、内部の構成要素の間でデータを送受信する構成である。特に、通信部(110)は、テキストを検出するための入力画像を取得することができる(S210)。

【0028】

プロセッサ(120)は、各種データの演算を実行する構成である。まず、プロセッサ(120)は、前記入力画像が取得されると、前記入力画像に含まれているピクセルのフィーチャー値を参照にして、前記入力画像の少なくとも1つのテキスト候補を検出したり、検出するようにサポートすることができる(S220)。

【0029】

このとき、プロセッサ(120)は、入力画像から所定の領域に含まれるすべてのピクセルのフィーチャー値が前記所定の領域の外側の境界領域のフィーチャー値よりも大きいとき、前記所定の領域を1つのテキスト候補として検出したり、検出するようにサポートすることができる。これは、次の数式で表現することができる。

【0030】

(数式1)

$$R_t = \{x | I(x) > I(y) \quad \forall x \in R_t, \forall y \in B(R_t)\}$$

ここで、xとyは、シングルチャンネル(single channel)画像Iのピクセルのインデックスであり、tはテキスト候補領域を抽出するために使用されている閾値(threshold)であり、B(Rt)は、Rtの境界領域のピクセルの集合である。もちろん、逆の場合であって、入力画像から所定の領域に含まれるすべてのピクセルのフィーチャー値が前記所定の領域の外側の境界領域のフィーチャー値よりも小さいとき、前記所定の領域を1つのテキスト候補として検出したり、検出するようにサポートすることができる。

【0031】

一方、次のように、ERs(Extremal Regions)を利用して、ER木構造を構成することもできる。つまり、プロセッサ(120)は、前記所定の領域のピクセル輝度(intensity)レベルを変更しながら、前記少なくとも一つのテキスト候補を検出したり、検出するようにサポートすることができ、前記検出された少なくとも一つのテキスト候補から、(i)前記所定の領域のピクセル輝度レベルが相対的に大きい下位レベルのテキスト候補と(i)前記下位レベルのテキスト候補の領域を含み、前記所定の領域のピクセル輝度レベル

10

20

30

40

50

が比較的小さい上位レベルのテキスト候補のうち前記所定の領域のピクセル輝度レベルが最も大きいテキスト候補を、それぞれ子ノードと親ノードに割り当て、木構造を構成したり、構成するようにサポートすることができる。

【0032】

図3は、本発明の一実施形態に係るER木構造の例を示す図である。

【0033】

図3に図示された例では、輝度レベルが136であるテキスト候補と、前記輝度レベルが136であるテキスト候補の領域を含み、輝度レベルが133であるテキスト候補がそれぞれER木構造の子ノードと親ノードに割り当てられていることを知ることができる。

【0034】

次に、プロセッサ(120)は、多重閾値分類を実行したり、実行するようにサポートすることができる(S240)。つまり、プロセッサ(120)は、検出されたテキスト候補が含まれる領域内で選択された少なくとも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、前記検出されたテキストの候補を第1閾値と比較して強いテキストまたは強くないテキストに分類したり、分類するようにサポートして、前記強くないテキストに分類されたテキスト候補を第2閾値と比較して弱いテキストまたはノンテキストに分類したり、分類するようにサポートすることができる。例えば、プロセッサ(120)は、AdaBoostアルゴリズムでトレーニングされた分類機を用いて前記検出されたテキスト候補を分類したり、分類するようにサポートすることができる。

【0035】

ここで、前記検出されたテキスト候補が含まれる領域は、エキストラ マージン(extra margin)を最小限に抑えながら前記検出されたテキストの候補を含むバウンディングボックスの領域であることがある。

【0036】

また、強いテキストは検出されたテキストの候補が実際にテキストである可能性が相対的に高いテキスト候補を意味する。一方、弱いテキストは検出されたテキストの候補が実際にテキストである可能性が強いテキストより相対的に低いテキスト候補を言い、ノンテキストは、検出されたテキストの候補が実際にテキストである可能性が非常に低いテキスト候補を言う。

【0037】

また、第1閾値は、高閾値であることができ、第2閾値は、低閾値であることがある。異なる閾値は、異なる精度を満足させることができるが、例えば、第1閾値は、99%の精度を、第2閾値は90%の精度を満足させることができる。このように、第2閾値は、第1閾値に比べて低い正確性が示されるが、比較的高い要精検率を示すことができる。

【0038】

また、プロセッサ(120)は、検出されたテキスト候補が含まれる領域内で選択されたピクセルのフィーチャー値をMLBP(Mean Local Binary Pattern)に変換した値を参照して前記分類の作業を実行したり、実行するようにサポートすることができる。

【0039】

図4は、MLBP変換の例を示す図である。

【0040】

図4を参照すると、検出されたテキスト候補が含まれる領域内で選択されたピクセルを囲む8つの隣接するピクセル(3×3パッチ内の隣接するピクセル)の平均値が先に計算される。そして、中央の選択されたピクセルを除いた残りの8つの隣接するピクセルと前記計算された平均値が比較され、特定の隣接するピクセルの値が平均値よりも大きい場合は、隣接するピクセルに1が割り当てられ、小さい場合は0が割り当てられる。そして、左上の隣接するピクセルから時計回りに、新たに割り当てられた値が読まれ、8つのビットからなる2進数でエンコードされる。図4に示された例では、MLBP値は2進数00110101₍₂₎であることがわかる。しかし、本発明で使用するすることができるフィーチャー(feature)がMLBPに限定されるものではなく、様々な他のフィーチャーが使用さ

10

20

30

40

50

れることがあるとするものである。

【0041】

一方、テキスト候補が二ステップを経て二つの閾値と比較されて分類されることもあるが、（これをdoublethreshold classificationとすることにする。）二ステップ以上を経て、分類が行われることもある。すなわち、前記第1閾値と前記第2閾値の間に少なくとも一つの間閾値があることがあり、この場合、プロセッサ（120）は、（ $i i - 1$ ）「non-strong text」に分類されたテキスト候補を前記中間閾値と比較して、前記強いテキストまたは次順位弱いテキストに分類したり、分類するようにサポートして、（ $i i - 2$ ）前記次順位弱いテキストを前記第2閾値と比較して、前記弱いテキストまたは前記ノンテキストで分類したり、分類するようにサポートすることができる。

10

【0042】

一方、プロセッサ（120）は、多重閾値分類を実行する前に、「Non-MaximumSuppression」を実行したり、実行するようにサポートすることもできる（S230）。

【0043】

つまり、検出されたテキストの候補が複数のとき、プロセッサ（120）は、前記検出された複数のテキスト候補が含まれる領域が所定割合以上にお互い重なり合うテキスト候補を含む少なくとも一つのテキスト候補セットを識別し、識別された各テキストの候補セットに対して所定の条件を満たしているテキストの候補を選択したり、選択するようにサポートすることができる。この場合、プロセッサ（120）は、前記所定の条件を満たしているテキスト候補が含まれている領域内で選択された少なくとも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、（ i ）前記所定の条件を満たしているテキストの候補を前記第1閾値と比較して、前記強いテキストまたは前記強くないテキストに分類したり、分類するようにサポートして、（ $i i$ ）前記強くないテキストに分類されたテキストの候補を前記第2閾値と比較して、前記弱いテキストまたは前記ノンテキストに分類したり、分類するようにサポートすることができる。

20

【0044】

前記のように、テキスト候補が含まれる領域が相互に重なる場合は、MSE R s (Maximally Stable Extremal Region)とER sでしばしば現れる。ER木構造は、ユニークネス(uniqueness)基準を保障するために繰り返されるER (node)を抑制し、マキシマムスタビリティを持つ1つのERだけを選択することができる。

30

【0045】

ER木構造を構成した場合「Non-MaximumSuppression」過程を具体的に説明すると、（a）前述したER木構造内の特定のノードに対して、前記特定のノードに対応するテキスト候補が含まれる領域と、前記特定のノードに対応するテキスト候補領域を含み、前記所定の領域のピクセル輝度レベルが比較的小さい、祖先ノードに対応するテキスト候補が含まれる領域が所定割合以上に重なる場合が所定回数以上である場合、（b）プロセッサ（120）は、自分の特定の祖先ノードに対しテキスト候補が含まれる領域の変化量が相対的に最も小さなノードに対応するテキスト候補を選択したり、選択するようにサポートすることができる。

【0046】

（a）の判断は、下記の式に使用されることができる。
（数式2）

40

$$O(R_{t-k}, R_t) = \frac{|R_t|}{|R_{t-k}|}$$

ここで、 R_{t-k} はER木構造で R_t の親ノードを指す。 $|R|$ はRのバウンディングボックス領域（テキスト候補が含まれる領域）を指す。

【0047】

50

例えば、 $0(R_{t-k}, R_t) > 0.7$ である場合が5回以上の場合 (b) が実行されることがある。(b)は、スタビリティが最大のERが選択されていることとして表現されることがありますが、スタビリティは次の式で表すことができる。

(数式3)

$$S(R_t) = \frac{(|R_{t-k}| - |R_t|)}{|R_t|}$$

例えば、数式2で $k=2$ に設定することができる。この場合、スタビリティは、特定のノードと特定のノードの祖父母ノードに対し考慮することができる。

10

【0048】

ちなみに、スタビリティが同じERが二つ以上ある場合、ERの領域（バウンディングボックスの領域）が最小であるものが選択されることがある。これにより、ユニークネスとコンパクトネス基準を満たすことができる。

【0049】

また、 R_t で重なる回数が3より大きく、 R_t が「non-maximum」制約によって除去されなければ R_t は、テキスト候補として考慮されることがある。しかし、テキスト候補が過度に大きい小さいアスペクト比を持つ場合、そのようなテキスト候補は削除されることがある。

20

【0050】

前記の場合、プロセッサ(120)は、前記選択されたテキストの候補が含まれる領域内で選択された、少なくとも一つのピクセルのフィーチャー値またはこれを変換した値を参照して、(i)前記所定の条件を満たしているテキスト候補を前記第1閾値と比較して、前記強いテキストまたは前記強くないテキストに分類したり、分類するようにサポートして、(ii)前記強くないテキストに分類されたテキストの候補を前記第2閾値と比較して、前記弱いテキストまたは前記ノンテキストに分類したり、分類するようにサポートすることができる。

【0051】

再び、図2を参照すると、プロセッサ(120)は、ヒステリシスを考慮して、テキストトラッキングを実行したり、実行するようにサポートすることができる(S250)。つまり、プロセッサ(120)は、前記のように分類された強いテキストと弱いテキストの関連情報を参照して前記弱いテキストを前記強いテキストに分類したり、分類するようにサポートするかどうかを判断することができる。より具体的には、プロセッサ(120)は、前記強いテキストと前記弱いテキストの関連情報を参照して前記弱いテキストを前記強いテキストまたはノンテキストに分類したり、分類するようサポートすることができる。

30

【0052】

前述したように強いテキストに分類されたテキストの候補は、実際にtextである可能性が高い。しかし、弱いテキストに分類されたテキストの候補の場合にも、実際にテキストである可能性がある。したがって、本発明は、弱いテキストの場合強いテキストと特徴が類似した場合強いテキストに編入させる。つまり、強いテキストの隣で弱いテキストに分類されたテキストの候補を探して、このようなテキストの候補が強いテキストと同様の特徴を持つ場合弱いテキストを強いテキストに変換する。これにより、高い要精検率を達成することができる。

40

【0053】

強いテキストと弱いテキストが類似の特徴があることを判断するために考慮される強いテキストと弱いテキストの関連情報は、前記強いテキストと前記弱いテキストの位置情報、サイズ情報、カラー情報、ストローク幅情報の少なくとも一部を含むことすることができる。

50

【0054】

次に、プロセッサ（120）は、前記第1閾値と比較して分類された強いテキスト及び前記弱いテキストから編入された強いテキストをグルーピングしたり、グルーピングするようにサポートすることができる（S260）。

【0055】

また、第1強いテキストとこれに隣接した第1弱いテキストから編入された第2強いテキストがグルーピングされて前記第2強いテキストとこれに隣接した第2弱いテキストの関連情報を参照して前記第2弱いテキストが第3強いテキストに分類されると、プロセッサ（120）は、前記第3強いテキストを前記第1強いテキスト及び前記第2強いテキストと一緒にグルーピングしたり、グルーピングするようにサポートすることができる。このように、本発明は、テキストグルーピング時に出発点となった強いテキストから一つのテキストずつ拡張しながらグルーピングを行うことができるようになる。

10

【0056】

また、出力にコンパクトなバウンディングボックスを提供するために、「minimum-area encasingrectangle」が計算されることができる。キャラクターピクセルの2D座標を使用して2Dイメージ空間でグルーピングされたテキストを取り囲む最小の長方形を計算することができる。

【0057】

図5は、本発明に係るテキスト検出手法を使用して例示画像からテキストを検出する場合は、各過程での中間結果を示す図である。

20

【0058】

図5の（a）の入力画像が取得された場合、（b）は、「non-maximum suppression」が行われて選択されたERを示す。（c）は、テキスト候補が高い閾値を利用して分類された結果を示すが、比較的少ない数の強いテキストが分類されたことを知ることができる。（d）は、低い閾値を利用して、追加の分類がなされ、多くのテキスト候補が弱いテキストに分類された結果を示す。（e）は、テキストトラッキングの結果を示す。その結果、多くの弱いテキストが強いテキストに編入される。（f）は、最終的にテキストグルーピングが行われた結果を示す。

【0059】

一方、プロセッサ（120）は、前述した通信部（110）と、後述する格納部（130）との間のデータの流れを制御する機能を実行する。つまり、プロセッサ（120）は、入力データからテキストを検出する装置（100）の各構成要素間のデータの流れを制御することにより、通信部（110）、格納部（130）でそれぞれ固有の機能を実行するように制御する。

30

【0060】

プロセッサ（120）は、MPU(Micro Processing Unit)やCPU(Central Processing Unit)、キャッシュメモリ(Cache Memory)、データバス(Data Bus)などのハードウェア構成を含むことができる。また、オペレーティングシステム、特定の目的を実行するアプリケーションのソフトウェア構成をさらに含むこともある。

【0061】

格納部（130）は、各種データを格納する構成である。特に、格納部（130）は、テキスト検出の各ステップで算出された中間結果のデータを格納することができる。また、オペレーティングシステム、特定の目的を実行するアプリケーションのソフトウェアを格納することができる。

40

【0062】

以下では、前述したテキスト検出方法及び装置を実装して、本発明者が実験した結果を簡単に説明する。

【0063】

図6～8は、他のテキスト検出技術と本発明の性能を比較実験した結果を示す図である。

50

【0064】

図6は、ICDAR 2011のデータセットのキャラクターレベルのリコール(recall)についてM.-C. Sung、B. Jun、H. Cho、and D.Kim.Scene text detection with robust character candidate extraction method. In Proc.ICDAR 2015、pages 426-430、2015で提案された最新のテキスト候補の検出方法と本発明を比較した結果を示すテーブルである

【0065】

図6に示されている「Sung et al.」と比較して、本発明は、ER木構造で「non-maximum suppression」を利用して、テキスト候補の数を減らし、initial ERがSung et al.の1/3程度であることが分かることができる。本発明の最終的なlocalization結果は、関連性が落ちるテキスト候補を90%以上削減しながら、かなりの要精検率を達成したと評価することができる。

10

【0066】

図7は、ICDAR 2013のデータセットに対して、本発明と他の技術のプリシジョン、リコール、ハーモニックミーン(harmonic mean、調和平均)を評価した結果を示すテーブルである。

【0067】

図7に示すように、X.-C. Yin、X. Yin、K.Huang、and H.-W. Hao. Robust text detection in natural scene images. IEEE Trans. Pattern Analysis Machine Intelligence、36(5):970-983、May 2014で提案されたアルゴリズムと比較しても、本発明のハーモニックミーン(調和平均)などが優れていることができる。

20

【0068】

図8は、Y.-F. Pan、X. Hou、and C.-L. Liu. A hybrid approach to detect and localize texts in natural scene images. IEEE Trans. Image Processing、20(3):800-813、2011年から提案された多言語データセットに対して実験した結果を示すテーブルである。

【0069】

図8に示すように、本発明は、他のアルゴリズムと比較して改善されたリコール、プリシジョン、ハーモニックミーン(調和平均)を示すことが分かる。

【0070】

一方、これまでに説明した本発明に係る実施例は、様々なコンピュータの構成要素を介して実行することができるプログラム命令の形で実装されて、コンピュータ読み取り可能な記録媒体に記録することができる。前記コンピュータ読み取り可能な記録媒体は、プログラム命令、データファイル、データ構造などを単独で、または組み合わせて含むことができる。前記コンピュータ読み取り可能な記録媒体に記録されているプログラムの命令は、本発明のために特別に設計されて構成されたものか、コンピュータソフトウェア分野の当業者に公知されて使用可能なものであることもできる。コンピュータ読み取り可能な記録媒体の例には、ハードディスク、フロッピーディスク、及び磁気テープのような磁気媒体、CD-ROM、DVDなどの光記録媒体、プティカルディスク(floptical disk)のような自己-光媒体(magneto-optical media)、及びROM、RAM、フラッシュメモリなどのようなプログラム命令を格納して実行するように特別に構成されたハードウェア装置が含まれる。プログラム命令の例には、コンパイラによって作られるような機械語コードだけでなく、インタプリタなどを使用して、コンピュータによって実行されることができる高級言語コードも含まれる。前記のハードウェア装置は、本発明に係る処理を実行するために1つ以上のソフトウェアモジュールとして動作するように構成することができ、その逆も同様である。

30

40

【0071】

以上で、本発明の具体的な構成要素などの特定の事項と限定された実施例及び図面によって説明されたが、これは本発明のより全般的な理解を助けるために提供されたものであり、本発明が前記実施例に限定されるものではなく、本発明が属する技術分野で通常の知識を有する者であれば、このような記載から多様な修正及び変形を図ることができる。

【0072】

50

したがって、本発明の思想は、前記説明した実施例に限定されて決まってはならず、後述する特許請求の範囲だけでなく、この特許請求の範囲と均等にはまたは等価的に変形されたすべてのものは、本発明の思想の範疇に属するものである。

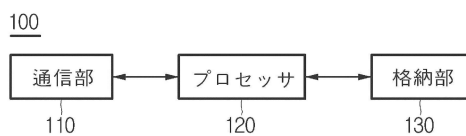
【符号の説明】

【0073】

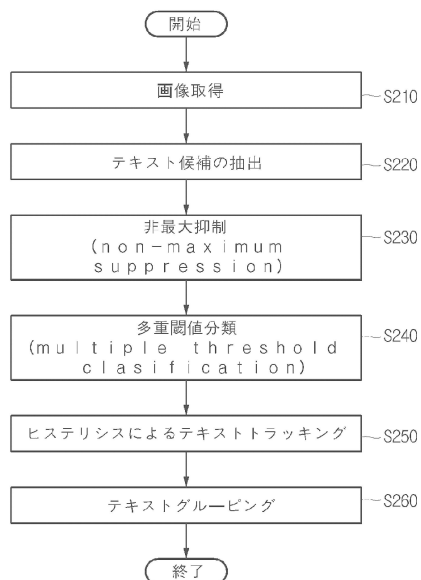
- 100：イメージに含まれているテキストを検出する装置
- 110：通信部
- 120：プロセッサ
- 130：格納部

10

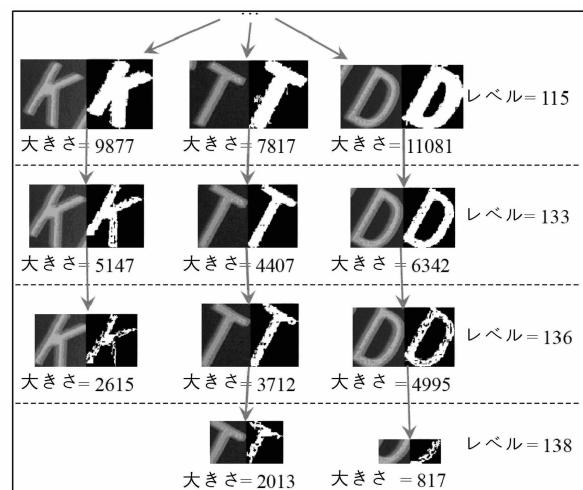
【図1】



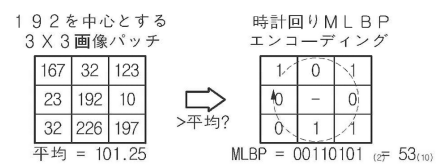
【図2】



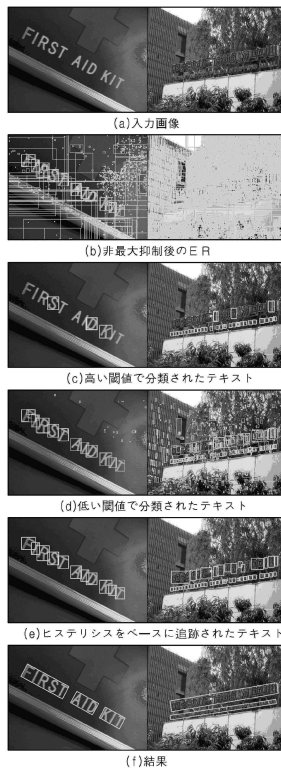
【図3】



【図4】



【図 5】



【図 6】

方法	候補の数	リコール (%)
全てのER	6,051,331	96.6
MSERs	39,762	53.9
SUNG et al. initial ER リファイされたER	1,729,833 93,357	89.6 87.7
我々の方法 NMS後のER 最終の文字	629,932 8,121	95.1 87.4

【図 7】

方法	リコール	プリシジョン	調和平均
我々の方法	78.45	86.26	82.17
SUNG et al.	74.23	88.65	80.80
TIAN et al.	75.89	85.15	80.25
YIN et al.	66.45	88.47	75.89
NEUMANN AND MATAS	64.84	87.51	74.49
YIN et al.	65.11	83.98	73.35
BAI et al.	68.24	78.89	73.18
SHI et al.	62.85	84.70	72.16

【図 8】

方法	リコール	プリシジョン	調和平均
PAN et al.	65.9	64.5	65.5
ベースライン	67.2	78.6	72.4
YIN et al.	68.5	82.6	74.6
TIAN et al.	78.4	84.7	81.4
我々の方法	93.5	93.1	93.3

フロントページの続き

- (56)参考文献 特開2014-229314 (JP, A)
特表2014-530424 (JP, A)
特開2014-228953 (JP, A)
特開2000-298725 (JP, A)
特開2001-092921 (JP, A)
宮本 一正、外4名, “複数特徴量を用いた低品質定型文字の一認識手法”, 電子情報通信学会論文誌, 日本, 社団法人電子情報通信学会, 1999年 4月25日, Vol.J82-D-II, No.4, pp .771-779

- (58)調査した分野(Int.Cl., DB名)
G06T 1/00, 7/00-7/90
G06K 9/20, 9/62