

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7373367号

(P7373367)

(45)発行日 令和5年11月2日(2023.11.2)

(24)登録日 令和5年10月25日(2023.10.25)

(51)国際特許分類

F I

G 0 6 T 7/00 (2017.01)

G 0 6 T 7/00 3 5 0 C

G 0 6 T 7/49 (2017.01)

G 0 6 T 7/49

G 0 6 V 30/14 (2022.01)

G 0 6 V 30/14

請求項の数 9 (全21頁)

(21)出願番号	特願2019-209628(P2019-209628)	(73)特許権者	000004352
(22)出願日	令和1年11月20日(2019.11.20)		日本放送協会
(65)公開番号	特開2021-82056(P2021-82056A)		東京都渋谷区神南2丁目2番1号
(43)公開日	令和3年5月27日(2021.5.27)	(74)代理人	110001807
審査請求日	令和4年10月4日(2022.10.4)		弁理士法人磯野国際特許商標事務所
		(72)発明者	遠藤 伶
			東京都世田谷区砧一丁目10番11号
			日本放送協会放送技術研究所内
		(72)発明者	河合 吉彦
			東京都世田谷区砧一丁目10番11号
			日本放送協会放送技術研究所内
		(72)発明者	望月 貴裕
			東京都世田谷区砧一丁目10番11号
			日本放送協会放送技術研究所内
		審査官	長谷川 素直

最終頁に続く

(54)【発明の名称】 文字領域検出モデル学習装置およびそのプログラム、ならびに、文字領域検出装置およびそのプログラム

(57)【特許請求の範囲】

【請求項1】

画像内の文字領域を検出するために用いるニューラルネットワークのモデルを学習する文字領域検出モデル学習装置であって、

画像に含まれる単独文字の領域分布を示す文字マップおよび前記画像の特徴を示す特徴マップを生成する単独文字検出モデルを用いて、学習用画像から前記文字マップおよび前記特徴マップを生成する単独文字検出手段と、

前記学習用画像に含まれる単独文字の領域を示す正解データである領域座標から前記学習用画像に含まれる単独文字の領域分布を示す正解マップを生成する正解マップ生成手段と、

前記文字マップと前記正解マップとの誤差を算出する単独文字誤差算出手段と、

前記単独文字誤差算出手段で算出された誤差を小さくする方向に前記単独文字検出モデルのパラメータを更新する第1パラメータ更新手段と、

前記文字マップおよび前記特徴マップから前記単独文字のペアが同じ文字列に含まれるか否かを示すペア属性を算出するペア属性推定モデルを用いて、前記文字マップで特定される単独文字のペアのペア属性を算出するペア属性算出手段と、

前記学習用画像に含まれる文字列の領域を示す正解データである領域座標から前記単独文字のペアについての正解の属性を求め、前記ペア属性との誤差を算出するペア属性誤差算出手段と、

前記ペア属性誤差算出手段で算出される誤差を小さくする方向に前記ペア属性推定モデル

ルのパラメータを更新する第2パラメータ更新手段と、
を備えることを特徴とする文字領域検出モデル学習装置。

【請求項2】

前記ペア属性推定モデルは、グラフコンボリューションネットワークで構成され、
前記ペア属性算出手段は、

前記文字マップで特定される単独文字のペアをノードとし、ノード同士で同一の単独文字を持つノード間をエッジで接続したグラフ構造を生成するグラフ構造生成手段と、

前記ペア属性推定モデルを用いて、前記ノードに含まれる単独文字の位置における前記特徴マップの特徴量を前記ノードの特徴量として前記ノードのペア属性を算出するノード属性算出手段と、

を備えることを特徴とする請求項1に記載の文字領域検出モデル学習装置。

【請求項3】

前記第2パラメータ更新手段は、前記ペア属性推定モデルのパラメータとともに、前記単独文字検出モデルのパラメータを重畳して更新することを特徴とする請求項1または請求項2に記載の文字領域検出モデル学習装置。

【請求項4】

前記単独文字検出モデルは、

画像から予め定めたチャンネル数の特徴量を複数の畳み込み層を介して抽出するコンボリューションニューラルネットワークで構成された第1ネットワークと、

前記第1ネットワークで抽出された特徴量に対して、拡大と畳み込み層による畳み込みとを繰り返すことで、予め定めた大きさの前記特徴マップを生成するとともに、前記特徴マップを1チャンネルに畳み込んで前記文字マップを生成するコンボリューションニューラルネットワークで構成された第2ネットワークと、

を連結して構成していることを特徴とする請求項1から請求項3のいずれか一項に記載の文字領域検出モデル学習装置。

【請求項5】

コンピュータを、請求項1から請求項4のいずれか一項に記載の文字領域検出モデル学習装置として機能させるための文字領域検出モデル学習プログラム。

【請求項6】

画像内の文字領域を検出する文字領域検出装置であって、

画像に含まれる単独文字の領域分布を示す文字マップおよび前記画像の特徴を示す特徴マップを生成する学習済のニューラルネットワークで構成された単独文字検出モデルを用いて、入力された画像から前記文字マップおよび前記特徴マップを生成する単独文字検出手段と、

前記文字マップおよび前記特徴マップから前記単独文字のペアが同じ文字列に含まれるか否かを示すペア属性を算出する学習済のニューラルネットワークで構成されたペア属性推定モデルを用いて、前記文字マップで特定される単独文字のペアのペア属性を算出するペア属性算出手段と、

前記ペア属性で同じ文字列に含まれる単独文字の領域を統合した前記文字領域を算出する文字領域算出手段と、

を備えることを特徴とする文字領域検出装置。

【請求項7】

前記ペア属性推定モデルは、グラフコンボリューションネットワークで構成され、
前記ペア属性算出手段は、

前記文字マップで特定される単独文字のペアをノードとし、ノード同士で同一の単独文字を持つノード間をエッジで接続したグラフ構造を生成するグラフ構造生成手段と、

前記ペア属性推定モデルを用いて、前記ノードに含まれる単独文字の位置における前記特徴マップの特徴量を前記ノードの特徴量として前記ノードのペア属性を算出するノード属性算出手段と、

を備えることを特徴とする請求項6に記載の文字領域検出装置。

10

20

30

40

50

【請求項 8】

前記単独文字検出モデルは、

画像から予め定めたチャンネル数の特徴量を複数の畳み込み層を介して抽出するコンボリューションニューラルネットワークで構成された第 1 ネットワークと、

前記第 1 ネットワークで抽出された特徴量に対して、拡大と畳み込み層による畳み込みとを繰り返すことで、予め定めた大きさの前記特徴マップを生成するとともに、前記特徴マップを 1 チャンネルに畳み込んで前記文字マップを生成するコンボリューションニューラルネットワークで構成された第 2 ネットワークと、

を連結して構成していることを特徴とする請求項 6 または請求項 7 に記載の文字領域検出装置。

10

【請求項 9】

コンピュータを、請求項 6 から請求項 8 のいずれか一項に記載の文字領域検出装置として機能させるための文字領域検出プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像内の文字領域を検出するための文字領域検出モデルを学習する文字領域検出モデル学習装置およびそのプログラム、ならびに、文字領域検出モデルを用いて画像内の文字領域を検出する文字領域検出装置およびそのプログラムに関する。

【背景技術】

20

【0002】

従来、画像内の文字領域を検出する手法として、文字に正対した形で撮影された画像から、文字領域を検出する手法が一般的であった（例えば、特許文献 1 参照）。

しかし、このような手法を用いた場合、撮影条件を限定しない情景画像内では、文字領域が矩形形状ではないため、文字領域を検出することは困難であった。

そこで、近年では、機械学習技術（ニューラルネットワーク）を利用して、情景画像内に映った文字領域を検出する手法が種々提案されている。

【0003】

例えば、非特許文献 1 には、図 1 1 に示すように、文字を含む画像 I を入力した際に、文字列の領域を示す領域座標 O を文字の領域（R 1, R 2, …）ごとに出力するように学習されたニューラルネットワーク NN 1 を用いて、文字領域を検出する手法が開示されている。この手法は、1 文字以上の文字列の単位で文字領域を検出する。

30

【0004】

また、例えば、非特許文献 2 には、図 1 2 に示すように、文字を含む画像 I を入力した際に、1 文字（単独文字）ごとの領域分布を示す文字マップ M₁ と、文字間の領域分布を示す文字間マップ M₂ とを出力するように学習されたニューラルネットワーク NN 2 を用いて、文字領域を検出する手法が開示されている。この手法は、ニューラルネットワーク NN 2 を用いて、画像 I から、文字マップ M₁ と文字間マップ M₂ と生成し、それらを重ね合わせたマップ M₃ を生成する。そして、この手法は、マップ M₃ の文字・文字間の重複した領域（R 1, R 2, …）の領域座標 O を文字領域として検出する。

40

【先行技術文献】

【特許文献】

【0005】

【文献】特開 2003-256771 号公報

【非特許文献】

【0006】

【文献】Xiaobing Wang, Yingying Jiang, Zhenbo Luo, Cheng-Lin Liu, Hyunsoo Choi, Sungjin Kim, “Arbitrary Shape Scene Text Detection with Adaptive Text Region Representation”, In IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.6449-6458, 2019.

50

【文献】Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, Hwalsuk Lee, “Character Region Awareness for Text Detection”, In IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.9365-9374, 2019.

【発明の概要】

【発明が解決しようとする課題】

【0007】

非特許文献1に記載の手法（以下、従来手法1）では、画像内における検出対象の文字列が占める形状（アスペクト比）は文字数に応じて大きく変化する。そのため、従来手法1は、ニューラルネットワークの学習を十分に行うことが困難であり、例えば、顔認識等の領域形状が安定した物体の検出に比べ、高精度に文字列を検出することができないという問題がある。

10

【0008】

これに対し、非特許文献2に記載の手法（以下、従来手法2）は、ニューラルネットを用いて単独文字と文字間とを検出するため、検出対象となる領域の形状が比較的安定しており、従来手法1よりは文字列の検出精度を上げることができる。

しかし、従来手法2は、単独文字と文字間との統合を、単純なルールベースのアルゴリズムで行うため、例えば、狭い範囲に複数の文字列が密集している場合に、それらを1つの文字列として検出する等、文字列の密集の度合いによっては、正しく文字列を検出することができないという問題がある。

【0009】

20

本発明は、このような問題に鑑みてなされたものであり、画像内の文字領域を精度よく検出することが可能なモデルを学習する文字領域検出モデル学習装置およびそのプログラム、ならびに、文字領域検出装置およびそのプログラムを提供することを課題とする。

【課題を解決するための手段】

【0010】

前記課題を解決するため、本発明に係る文字領域検出モデル学習装置は、画像内の文字領域を検出するために用いるニューラルネットワークのモデルを学習する文字領域検出モデル学習装置であって、単独文字検出手段と、正解マップ生成手段と、単独文字誤差算出手段と、第1パラメータ更新手段と、ペア属性算出手段と、ペア属性誤差算出手段と、第2パラメータ更新手段と、を備える構成とした。

30

【0011】

かかる構成において、文字領域検出モデル学習装置は、単独文字検出手段によって、単独文字検出モデルを用いて、学習用画像から文字マップおよび特徴マップを生成する。単独文字検出モデルは、画像の特徴を示す特徴マップを生成するニューラルネットワークと、特徴マップから画像に含まれる単独文字の領域分布を示す文字マップを生成するニューラルネットワークとを連結して構成することができる。

また、文字領域検出モデル学習装置は、正解マップ生成手段によって、学習用画像に含まれる単独文字の領域を示す正解データである領域座標から学習用画像に含まれる単独文字の領域分布を示す正解マップを生成する。

そして、文字領域検出モデル学習装置は、単独文字誤差算出手段によって、文字マップと正解マップとの誤差を算出する。

40

そして、文字領域検出モデル学習装置は、第1パラメータ更新手段によって、単独文字誤差算出手段で算出された誤差を小さくする方向に単独文字検出モデルのパラメータを更新する。これによって、文字領域検出モデル学習装置は、単独文字の位置を検出するための単独文字検出モデルを学習することができる。

【0012】

また、文字領域検出モデル学習装置は、ペア属性算出手段によって、ペア属性推定モデルを用いて、文字マップで特定される単独文字のペアのペア属性を算出する。ペア属性推定モデルは、文字マップおよび特徴マップから単独文字のペアが同じ文字列に含まれるか否かを示すペア属性を算出するニューラルネットワークで構成することができる。

50

そして、文字領域検出モデル学習装置は、ペア属性誤差算出手段によって、学習用画像に含まれる文字列の領域を示す正解データである領域座標から単独文字のペアについての正解の属性を求め、ペア属性算出手段で算出されたペア属性との誤差を算出する。

そして、文字領域検出モデル学習装置は、第2パラメータ更新手段によって、ペア属性誤差算出手段で算出される誤差を小さくする方向にペア属性推定モデルのパラメータを更新する。これによって、文字領域検出モデル学習装置は、単独文字のペアが同じ文字列を構成する文字であるか否かを判定するためのペア属性推定モデルを学習することができる。

なお、文字領域検出モデル学習装置は、コンピュータを、前記した各手段として機能させるための文字領域検出モデル学習プログラムで動作させることができる。

【0013】

また、前記課題を解決するため、本発明に係る文字領域検出装置は、画像内の文字領域を検出する文字領域検出装置であって、単独文字検出手段と、ペア属性算出手段と、文字領域算出手段と、を備える構成とした。

【0014】

かかる構成において、文字領域検出装置は、単独文字検出手段によって、画像に含まれる単独文字の領域分布を示す文字マップおよび画像の特徴を示す特徴マップを生成する学習済のニューラルネットワークで構成された単独文字検出モデルを用いて、入力された画像から文字マップおよび特徴マップを生成する。

そして、文字領域検出装置は、ペア属性算出手段によって、文字マップおよび特徴マップから単独文字のペアが同じ文字列に含まれるか否かを示すペア属性を算出する学習済のニューラルネットワークで構成されたペア属性推定モデルを用いて、文字マップで特定される単独文字のペアのペア属性を算出する。

【0015】

そして、文字領域検出装置は、文字領域算出手段によって、ペア属性で同じ文字列に含まれる単独文字の領域を統合した文字領域を算出する。例えば、文字領域算出手段は、同じ文字列の単独文字の領域を含む外接矩形等によって文字領域を算出する。

これによって、文字領域検出装置は、画像内において、文字列として認識される文字領域を検出する。

なお、文字領域検出装置は、コンピュータを、前記した各手段として機能させるための文字領域検出プログラムで動作させることができる。

【発明の効果】

【0016】

本発明は、以下に示す優れた効果を奏するものである。

本発明は、ニューラルネットワークの学習によって、単独文字同士が同じ文字列に属するか否かを判定するため、従来のような単純なルールベースのアルゴリズムで文字列の判定を行う手法に比べて、柔軟に文字列の判定を行うことができる。

これによって、本発明は、従来手法に比べて、画像から精度よく文字領域を検出することができる。

【図面の簡単な説明】

【0017】

【図1】本発明の第1実施形態に係る文字領域検出モデル学習装置の構成を示すブロック図である。

【図2】単独文字検出モデルのニューラルネットワークの構成例を示すネットワーク図である。

【図3】正解マップ生成手段における正解マップを生成する手法を説明するための説明図である。

【図4】ペア属性算出手段のグラフ構造生成手段におけるグラフ構造を生成する手法を説明するための説明図である。

【図5】ペア属性算出手段のノード属性算出手段が算出する文字のペア属性を説明するための説明図である。

10

20

30

40

50

【図 6】 ペア属性推定モデルのニューラルネットワークの構成例を示すネットワーク図である。

【図 7】 特徴マップと文字位置の特徴量との関係を説明するための説明図である。

【図 8】 本発明の第 1 実施形態に係る文字領域検出モデル学習装置の動作を示すフローチャートである。

【図 9】 本発明の第 2 実施形態に係る文字領域検出装置の構成を示すブロック図である。

【図 10】 本発明の第 2 実施形態に係る文字領域検出装置の動作を示すフローチャートである。

【図 11】 従来の第 1 の文字領域検出手法の概要を示す概要図である。

【図 12】 従来の第 2 の文字領域検出手法の概要を示す概要図である。

【発明を実施するための形態】

【0018】

以下、本発明の実施形態について図面を参照して説明する。

〔文字領域検出モデル学習装置の構成〕

図 1 を参照して、本発明の第 1 実施形態に係る文字領域検出モデル学習装置 1 の構成について説明する。

【0019】

文字領域検出モデル学習装置 1 は、画像内の文字領域を検出するために用いるニューラルネットワークのモデル（文字領域検出モデル）を学習するものである。

文字領域検出モデル学習装置 1 は、学習用画像 I_L と学習用正解データ D_L とを対とした学習データを用いて学習を行う。

【0020】

学習用画像 I_L は、1 文字以上の文字列を 1 ヶ所以上含んだ画像である。ここでは、学習用画像 I_L を、チャンネル数 C 、高さ H 画素、幅 W 画素（ $C \times H \times W$ ）とする。例えば、学習用画像 I_L として、RGB のカラー画像を用いた場合、チャンネル数は“3”である。

【0021】

学習用正解データ D_L は、対となる学習用画像 I_L に含まれる単独文字領域座標データ D_1 と、文字列領域座標データ D_2 とで構成される。

単独文字領域座標データ D_1 は、学習用画像 I_L 内の 1 文字（単独文字）ごとの領域座標 $C_1, C_2, \dots, C_m$ （ m は画像内に含まれる文字数）である。この単独文字の領域座標は、単独文字を囲む 4 角形の 4 頂点の座標で構成される。また、単独文字を囲む 4 角形は、矩形である必要はなく、台形、平行四辺形、不等辺四辺形等、文字の変形状状に応じた形状であればよい。

【0022】

文字列領域座標データ D_2 は、学習用画像 I_L 内の文字列ごとの領域座標 $S_1, S_2, \dots, S_n$ （ n は画像内に含まれる文字列数）である。この文字列の領域座標は、文字列を構成する単独文字を 1 文字以上囲む多角形の各頂点の座標で構成される。また、文字列を囲む多角形は、文字列を含めば、台形、平行四辺形、不等辺四辺形等、どのような形状でも構わないが、単独文字の内包を簡易に判定するため、矩形形状の 4 角形が好ましい。

なお、文字列は、1 文字以上の文字のまとまりを示す。しかし、分かち書きで記述された英文の文章のように空白を挟んだ文章の場合、1 つの文章を、空白で区切った複数の文字列とするか、空白を含んだ 1 つの文字列とするかは、文字領域をどの単位で検出したいかによって、予め定めておけばよい。例えば、「I have a dog.」を、「I」、「have」、「a」および「dog.」の 4 つの文字列とするか、「I have a dog.」の 1 つの文字列とするかは、いずれか一方に予め定めて学習データを生成しておく。

【0023】

図 1 に示すように、文字領域検出モデル学習装置 1 は、単独文字検出手段 10 と、正解マップ生成手段 11 と、単独文字誤差算出手段 12 と、パラメータ更新手段 13 と、ペア属性算出手段 14 と、ペア属性誤差算出手段 15 と、パラメータ更新手段 16 と、モデル記憶手段 17 と、を備える。

10

20

30

40

50

【0024】

単独文字検出手段10は、画像に含まれる単独文字の領域分布を示す文字マップおよび画像の特徴を示す特徴マップを生成するニューラルネットワークで構成された単独文字検出モデル N_1 を用いて、学習用画像 I_L から文字マップおよび特徴マップを生成するものである。

【0025】

単独文字検出モデル N_1 は、画像に対して畳み込み演算を行うことで特徴量を生成する第1ネットワークと、特徴量に対して畳み込み演算を行い画像の大きさに対応した特徴マップを生成し、特徴マップに対して畳み込み演算を行うことで文字マップを生成する第2ネットワークとを連結したニューラルネットワークのモデルである。

10

【0026】

ここで、図2を参照（適宜図1参照）して、単独文字検出モデル N_1 の構成例について説明する。

図2に示すように、単独文字検出モデル N_1 は、第1ネットワーク N_{11} と第2ネットワーク N_{12} とを連結したニューラルネットワークとして構成することができる。

【0027】

第1ネットワーク N_{11} は、画像 I に対して複数の畳み込み層を介して特徴量 f を抽出するコンボリューションニューラルネットワークである。この第1ネットワーク N_{11} は、例えば、VGG (Visual Geometry Group) 等の既存のネットワークを用いることができる。なお、第1ネットワーク N_{11} は、VGG以外にも、ResNet (Residual Network)、Inception等、一般的な物体認識ネットワークの特徴抽出部分のネットワークを用いることができる。

20

【0028】

第2ネットワーク N_{12} は、第1ネットワーク N_{11} で抽出される特徴量 f に対して、拡大と畳み込み層による畳み込みとを繰り返すことで、予め定めた大きさの特徴マップ M_f を生成するとともに、畳み込み層を介して特徴マップ M_f から1チャンネルの文字マップ M_c を生成するネットワークである。

この第2ネットワーク N_{12} は、特徴量 f を拡大し、拡大した特徴量に同じ大きさの第1ネットワーク N_{11} で生成された中間特徴量を連結して畳み込みを行う処理を、特徴量が予め定めた大きさになるまで繰り返す。なお、畳み込みの際に、必ずしも中間特徴量を連結する必要はないが、特徴量の下層への畳み込みを行わないパスを設けることで、モデル学習時における勾配消失を防止することができるため好ましい。

30

【0029】

特徴マップ M_f は、特徴量 f を画像 I の画素に対応付けた情報である。特徴マップ M_f は、例えば、画像 I がチャンネル数“3”、高さ H 画素、幅 W 画素（ $3 \times H \times W$ ）で、特徴量 f のチャンネル数が“16”の場合、チャンネル数“16”、高さ H 画素、幅 W 画素（ $16 \times H \times W$ ）となる。

【0030】

文字マップ M_c は、画像 I に含まれる単独文字の領域分布を示す情報である。文字マップ M_c のチャンネル数は“1”で、高さおよび幅は、特徴マップ M_f と同じ H 画素および W 画素（ $1 \times H \times W$ ）である。この文字マップ M_c が後記する正解マップ生成手段11で生成される正解マップとなるように、単独文字検出モデル N_1 が学習されることになる。

40

図1に戻って、文字領域検出モデル学習装置1の構成について説明を続ける。

【0031】

単独文字検出手段10は、生成した文字マップを単独文字誤差算出手段12に出力する。また、単独文字検出手段10は、生成した特徴マップおよび文字マップをペア属性算出手段14に出力する。

【0032】

正解マップ生成手段11は、学習用正解データ D_L から、学習用画像 I_L に含まれる単独文字の領域分布を示す情報である正解マップを生成するものである。

50

ここでは、正解マップ生成手段 11 は、単独文字の中心と単独文字以外の領域とで異なる値を割り当て、単独文字の中心から単独文字領域の縁に近づくほど、単独文字以外の領域の値に近づくように値を割り当てることで、正解マップを生成する。

例えば、正解マップ生成手段 11 は、単独文字の中心の画素値を“1.0”（例えば、256階調における画素値“255”に相当）、単独文字以外の領域の画素値を“0.0”（例えば、256階調における画素値“0”に相当）とし、単独文字の中心から単独文字領域の縁に近づくほど“0.0”に近づくように値を割り当てる。

【0033】

ここで、図3を参照（適宜図1参照）して、正解マップの生成手法について説明する。なお、図3では、説明を簡略化するため、1文字のみ記載された学習用画像 I_L を例として説明するが、複数文字が記載されている場合でも同様である。

10

図3に示すように、学習用画像 I_L に単独文字“A”が存在し、学習用正解データ D の単独文字領域座標データ D_1 の1つの単独文字の領域座標（例えば、 C_1 ）として、4頂点 P_1 、 P_2 、 P_3 、 P_4 が設定されていたとする。

このとき、正解マップ生成手段 11 は、二次元ガウス分布を適用した正方形画像（例えば、 256×256 画素） G_d を生成する。ここでは、正方形画像 G_d の中心の画素値を“1.0”、画像端の画素値を“0.0”とする。

【0034】

そして、正解マップ生成手段 11 は、学習用画像 I_L と同じ大きさで全面に“0.0”の値を初期設定した正解マップ M_r の4頂点 P_1 、 P_2 、 P_3 、 P_4 と、正方形画像 G_d の4頂点とが一致するように、正方形画像 G_d を透視変換して、正解マップ M_r に書きする。

20

これによって、正解マップ生成手段 11 は、学習用正解データ D_L に含まれる単独文字の分布領域として、単独文字の中心位置と領域形状とを模式的に表した正解マップ M_r を生成することができる。

図1に戻って、文字領域検出モデル学習装置1の構成について説明を続ける。

【0035】

単独文字誤差算出手段 12 は、単独文字検出手段 10 で生成された文字マップと、正解マップ生成手段 11 で生成された正解マップとの誤差を算出するものである。

単独文字誤差算出手段 12 における誤差計算には、例えば、平均二乗誤差（MSE：Mean squared error）、バイナリ交差エントロピ（Binary cross-entropy）等、文字マップと正解マップとの各画素値の差が大きいほど、大きな値を誤差として算出する関数を用いることができる。

30

単独文字誤差算出手段 12 は、算出した文字マップと正解マップとの誤差をパラメータ更新手段 13 に出力する。

【0036】

パラメータ更新手段（第1パラメータ更新手段） 13 は、単独文字誤差算出手段 12 で算出された誤差を小さくするように、単独文字検出モデル N_1 のパラメータを更新するものである。

パラメータ更新手段 13 におけるパラメータの更新には、例えば、確率的勾配降下法（SGD：Stochastic Gradient Descent）、Adam（Adaptive moment estimation）等、一般的なニューラルネットワークの最適化手法を用いることができる。

40

パラメータ更新手段 13 は、確率的勾配降下法等によって、モデル記憶手段 17 に記憶されている単独文字検出モデル N_1 のパラメータを更新する。

【0037】

ペア属性算出手段 14 は、ニューラルネットワークで構成されたペア属性推定モデル N_2 を用いて、単独文字検出手段 10 で検出された単独文字の各ペアが、同じ文字列に属する文字か否かを示すペア属性を算出するものである。

ペア属性算出手段 14 は、グラフ構造生成手段 140 と、ノード属性算出手段 141 と、を備える。

50

【0038】

グラフ構造生成手段140は、単独文字検出手段10で生成された文字マップに基づいて、単独文字のペアをノードとし、ノード同士で同一の単独文字を持つノード間をエッジで接続したグラフ構造を生成するものである。

グラフ構造生成手段140は、文字マップにおいて局所値（ここでは、局所最大値）を持つ画素の位置を単独文字の位置とし、グラフ構造を生成する。ただし、グラフ構造生成手段140は、局所最大値のうち、予め定めた閾値（例えば、0.5）を超える画素を単独文字の位置とすることが好ましい。そして、グラフ構造生成手段140は、単独文字の位置に対応付けて、固有のラベルを付与する。

なお、グラフ構造生成手段140において、検出された単独文字が1文字以下の場合、ペア属性算出手段14は、ペア属性の算出を行わないこととする。

10

【0039】

ここで、図4を参照（適宜図1参照）して、グラフ構造生成手段140が生成するグラフ構造の例について説明する。

図4に示すように、文字マップMcに4つの単独文字が存在しているものとする。なお、図4中、「a」、「b」、「c」、「d」は、説明の都合上、単独文字の位置を識別するためのラベルとして記載したもので、実際に文字マップMc上に記述されているものではない。

【0040】

グラフ構造生成手段140は、単独文字のすべてのペアとなる「ab」、「ac」、「ad」、「ba」、「bc」、「bd」、「ca」、「cb」、「cd」、「da」、「db」、「dc」の12個のペアをそれぞれノードNとして設定する。

20

【0041】

なお、グラフ構造生成手段140は、これらすべてのノードを必ずしもすべて使用する必要はない。例えば、グラフ構造生成手段140は、ノードに含まれる単独文字同士の距離（画像上の距離）が離れていると判断される場合、そのノードを除外することとしてもよい。

具体的には、グラフ構造生成手段140は、単独文字ごとに、当該単独文字を含むノードのペア間の距離が短い方から順に順位付けし、予め定めた数n（例えば、n=5）を超えるノードを削除候補とする。そして、グラフ構造生成手段140は、ノードに含まれる両方の単独文字で、当該ノードが削除対象となったものを削除する。

30

【0042】

例えば、図4において、単独文字「b」に着目した場合、「b」を含むノードは、ペア間の距離が「ab」=「bd」<「bc」となる。ここで、予め定めた数nを“2”とした場合、グラフ構造生成手段140は、ノード「bc」を除外候補とする。同様に、単独文字「c」に着目した場合、「c」を含むノードは、ペア間の距離が「ac」=「cd」<「bc」となり、ノード「bc」が除外候補となる。

このように、単独文字「b」、「c」について、両方ともノード「bc」が除外候補となったため、グラフ構造生成手段140は、ノード「bc」を除外する。

なお、単独文字のペアにおいて、いずれか一方が除外候補となった場合に、そのペアのノードを削除することとしてもよい。

40

あるいは、グラフ構造生成手段140は、ノードに含まれる単独文字のペア間の距離が予め定めた閾値を上回る場合に、そのノードを除外することとしてもよい。

【0043】

また、グラフ構造生成手段140は、設定したそれぞれのノードNにおいて、「ab」、「ac」のように、同じ単独文字（ここでは、「a」）のラベルを共通に含むノードN間にエッジEを設定する。一方、グラフ構造生成手段140は、「ab」、「cd」のように、同じ単独文字を含まないノードN間にはエッジEを設定しないものとする。

これによって、グラフ構造生成手段140は、単独文字のペア（ラベル対）をノードN、ノードN同士で同一の単独文字を持つノード間をエッジEで接続したグラフ構造Gを生

50

成する。なお、図4のグラフ構造Gは、一部のノードおよびエッジを省略している。

図1に戻って、文字領域検出モデル学習装置1の構成について説明を続ける。

【0044】

グラフ構造生成手段140は、生成したグラフ構造を単独文字の位置とともにノード属性算出手段141に出力する。

【0045】

ノード属性算出手段141は、ニューラルネットワークで構成されたペア属性推定モデル N_2 を用いて、グラフ構造生成手段140で生成されたグラフ構造と、特徴マップとに基づいて、単独文字同士のペア属性を算出するものである。

このノード属性算出手段141は、ペア属性推定モデル N_2 を用いて、図5に示すように、グラフ構造GのノードNごとに、ノード属性としてペア属性を算出する。

ペア属性（ノード属性）は、単独文字が同じ文字列に属するペアである属性（例えば、属性値“0”）と、異なる文字列に属するペアである属性（例えば、属性値“1”）の2種類である。なお、図5では、「a」および「b」が同じ文字列に属し、「c」および「d」が同じ文字列に属している状態を示している。

【0046】

ここで、図6を参照（適宜図1参照）して、ペア属性推定モデル N_2 の構成例について説明する。

図6に示すように、ペア属性推定モデル N_2 は、グラフコンボリューションネットワーク（GCN：Graph Convolutional Networks）で構成される。なお、図6のペア属性推定モデル N_2 は、図4に例示したグラフ構造Gの「ab」のノードにエッジを接続するノードについて図示しているが、他のノードについても同様である。

ペア属性推定モデル N_2 は、エッジEで接続されたノードNに対応する2つの単独文字の特徴量を、ノード特徴量として入力し、順次畳み込み演算を行うことで、ノードNごとにペア属性を出力するネットワークである。

【0047】

単独文字の特徴量は、図7に示すように、チャンネル数“C”、高さH画素、幅W画素（ $C \times H \times W$ ）の特徴マップ M_f において、単独文字の位置に対応する1チャンネルごとの値をチャンネル数分合算した数値列である。

例えば、図6において、ペア属性推定モデル N_2 に入力する「ab」のノードNの場合、当該ノードに対応するノード特徴量は、「a」の特徴量 f_a と「b」の特徴量 f_b とを要素ごとに足し合わせた数値列とする。他のノードについても同様である。

ペア属性推定モデル N_2 は、出力として、ノードNごとに“0”～“1”の範囲の値となるペア属性を出力する。

【0048】

図1に戻って、文字領域検出モデル学習装置1の構成について説明を続ける。

ノード属性算出手段141は、ペア属性推定モデル N_2 を用いて算出したノード（ラベル対）ごとのペア属性を、2つの単独文字の位置とともに、ペア属性誤差算出手段15に出力する。

【0049】

ペア属性誤差算出手段15は、学習用正解データ D_L に基づいて、ペア属性算出手段14で算出されたペア属性の誤差を算出するものである。

ペア属性誤差算出手段15は、学習用正解データ D_L の文字列領域座標データ D_2 の領域座標 $S_1 \sim S_n$ を参照し、ペア属性算出手段14で算出されたペア属性に対応する2つ単独文字の位置が同じ領域に含まれるか否かを正解属性とし、ペア属性と正解属性との誤差を算出する。

【0050】

正解属性は、ペア属性と同様に2種類とし、ペア属性に対応する2つ単独文字の位置が同じ領域に含まれる場合、正解属性の値を“0”、同じ領域に含まれない場合、正解属性の値を“1”とする。正解属性が“0”の場合、2つ単独文字は同じ文字列に含まれ、正解属性

10

20

30

40

50

が“ 1 ”の場合、2つ単独文字は異なる文字列に含まれることになる。

ペア属性誤差算出手段15における誤差計算には、交差エントロピー (Cross-entropy) 等、算出したペア属性が正解属性と異なる場合に値が大きくなる関数を用いることができる。

ペア属性誤差算出手段15は、算出したノードごとのペア属性と正解属性との誤差を、パラメータ更新手段16に出力する。

【0051】

パラメータ更新手段 (第2パラメータ更新手段) 16は、ペア属性誤差算出手段15で算出された2つの単独文字のペア属性と正解属性との誤差を小さくするように、単独文字検出モデル N_1 およびペア属性推定モデル N_2 のパラメータを更新するものである。

パラメータ更新手段16におけるパラメータの更新には、例えば、確率的勾配降下法 (SGD)、Adam等、一般的なニューラルネットワークの最適化手法を用いることができる。

パラメータ更新手段16は、確率的勾配降下法等によって、モデル記憶手段17に記憶されている単独文字検出モデル N_1 およびペア属性推定モデル N_2 のパラメータを更新する。

【0052】

なお、単独文字検出モデル N_1 のパラメータは、パラメータ更新手段13において更新されるため、必ずしもパラメータ更新手段16において更新する必要はない。

しかし、パラメータ更新手段16において、単独文字検出モデル N_1 のパラメータを重畳して更新することで、文字列を精度よく検出するための単独文字の特徴を抽出することが可能になる。

【0053】

モデル記憶手段17は、画像内の文字領域を検出するためのニューラルネットワークで構成された文字領域検出モデルのパラメータを記憶するものである。このモデル記憶手段17は、半導体メモリ等の一般的な記憶媒体で構成することができる。

文字領域検出モデルは、単独文字検出モデル N_1 およびペア属性推定モデル N_2 で構成される。

単独文字検出モデル N_1 のパラメータは、単独文字検出手段10によって参照され、パラメータ更新手段13およびパラメータ更新手段16によって更新される。

ペア属性推定モデル N_2 のパラメータは、ペア属性算出手段14によって参照され、パラメータ更新手段16によって更新される。

【0054】

以上説明したように文字領域検出モデル学習装置1を構成することで、文字領域検出モデル学習装置1は、画像内の文字領域を検出するための文字領域検出モデル (単独文字検出モデル N_1 およびペア属性推定モデル N_2) を学習することができる。

【0055】

このように、文字領域検出モデル学習装置1は、文字列の判定をニューラルネットワークで学習することで、複数の文字列が狭い範囲に密集している場合等、複雑な状態で画像内に文字列が存在している場合でも、精度よく文字列の領域を判定することが可能なモデルを学習することができる。

なお、文字領域検出モデル学習装置1は、コンピュータを、前記した各手段として機能させるための文字領域検出モデル学習プログラムで動作させることができる。

【0056】

〔文字領域検出モデル学習装置の動作〕

次に、図8を参照 (構成については適宜図1参照) して、本発明の第1実施形態に係る文字領域検出モデル学習装置1の動作について説明する。

ステップS10において、単独文字検出手段10は、学習用画像 I_L を入力する。

ステップS11において、単独文字検出手段10は、モデル記憶手段17に記憶されている単独文字検出モデル N_1 を用いて、学習用画像 I_L に対応する画像特徴である特徴マ

10

20

30

40

50

ップと、学習用画像 I_L に対応する単独文字の領域分布を示す文字マップとを生成する。

ステップ S_{12} において、正解マップ生成手段 11 は、学習用正解データ D_L から、学習用画像 I_L 内の単独文字ごとの正解の領域を示す正解マップを生成する。

【0057】

ステップ S_{13} において、単独文字誤差算出手段 12 は、ステップ S_{11} で生成された文字マップと、ステップ S_{12} で生成された正解マップとの誤差を算出する。

ステップ S_{14} において、パラメータ更新手段 13 は、ステップ S_{13} で算出された誤差を小さくするように、単独文字検出モデル N_1 のパラメータを更新する。

【0058】

ステップ S_{15} において、ペア属性算出手段 14 のグラフ構造生成手段 140 は、文字マップにおいて局所最大値を持つ画素の位置を単独文字の位置として検出する。

10

ステップ S_{16} において、グラフ構造生成手段 140 は、単独文字の位置が2以上検出されたか否かを判定する。

ここで、単独文字の位置が2以上検出されなかった場合（ステップ S_{16} で No ）、ペア属性算出手段 14 は、ペア属性の算出を行わずに、ステップ S_{22} に動作を移す。

一方、単独文字の位置が2以上検出された場合（ステップ S_{16} で Yes ）、ステップ S_{17} において、ペア属性算出手段 14 のグラフ構造生成手段 140 は、ステップ S_{11} で生成された文字マップに基づいて、単独文字のペアをノードとし、ノード同士で同一の単独文字を持つノード間をエッジで接続したグラフ構造を生成する。

【0059】

20

ステップ S_{18} において、ペア属性算出手段 14 のノード属性算出手段 141 は、モデル記憶手段 17 に記憶されているペア属性推定モデル N_2 を用いて、ステップ S_{11} で生成された特徴マップと、ステップ S_{17} で生成されたグラフ構造とから、ノード属性として、単独文字同士のペア属性を算出する。

ステップ S_{19} において、ペア属性誤差算出手段 15 は、学習用正解データ D_L の文字列領域座標データ D_2 の領域座標 $S_1 \sim S_n$ を参照し、ステップ S_{18} で算出されたペア属性に対応する2つ単独文字の位置が同じ領域に含まれるか否かの属性を、正解属性として生成する。

【0060】

ステップ S_{20} において、ペア属性誤差算出手段 15 は、ステップ S_{19} で生成された正解属性と、ステップ S_{18} で算出されたペア属性との誤差を算出する。

30

ステップ S_{21} において、パラメータ更新手段 16 は、ステップ S_{20} で算出された誤差を小さくするように、単独文字検出モデル N_1 およびペア属性推定モデル N_2 のパラメータを更新する。

【0061】

ステップ S_{22} において、文字領域検出モデル学習装置 1 は、予め定めた学習の終了条件を満たしたか否かを判定する。ここで、学習の終了条件は、例えば、すべての学習データ（学習用画像 I_L 、学習用正解データ D_L ）による学習が終了した場合、パラメータ更新手段 13 、 16 におけるパラメータの更新が予め定めた閾値内に収束した場合等である。

ここで、まだ、終了条件に達していない場合（ステップ S_{22} で No ）、文字領域検出モデル学習装置 1 は、ステップ S_{10} に戻って動作を継続する。

40

一方、終了条件に達した場合（ステップ S_{22} で Yes ）、文字領域検出モデル学習装置 1 は、動作を終了する。

【0062】

〔文字領域検出装置の構成〕

次に、図9を参照して、本発明の第2実施形態に係る文字領域検出装置2の構成について説明する。

【0063】

文字領域検出装置2は、文字領域検出モデル学習装置1（図1）で学習された文字領域検出モデル（単独文字検出モデル N_1 およびペア属性推定モデル N_2 ）を用いて、画像内

50

の文字領域を検出するものである。

文字領域検出装置 2 は、単独文字検出手段 10 B と、ペア属性算出手段 14 B と、モデル記憶手段 17 B と、文字領域算出手段 18 と、を備える。

【0064】

単独文字検出手段 10 B は、画像に含まれる単独文字の領域分布を示す文字マップおよび画像の特徴を示す特徴マップを生成するニューラルネットワークで構成された学習済の単独文字検出モデル N_1 を用いて、画像 I から文字マップおよび特徴マップを生成するものである。

この単独文字検出手段 10 B は、入力する画像 I が文字領域を検出する対象の画像である点、文字マップの出力先が文字領域算出手段 18 である点を除いて、文字領域検出モデル学習装置 1（図 1）の単独文字検出手段 10 と同じ機能を有する。

【0065】

ペア属性算出手段 14 B は、ニューラルネットワークで構成された学習済のペア属性推定モデル N_2 を用いて、単独文字検出手段 10 B で検出された単独文字の各ペアが、同じ文字列に属する文字か否かを示すペア属性を算出するものである。

ペア属性算出手段 14 B は、グラフ構造生成手段 140 B と、ノード属性算出手段 141 B と、を備える。

【0066】

グラフ構造生成手段 140 B は、単独文字検出手段 10 B で生成された文字マップに基づいて、単独文字のペアをノードとし、ノード同士で同一の単独文字を持つノード間をエッジで接続したグラフ構造を生成するものである。

このグラフ構造生成手段 140 B は、基本的に文字領域検出モデル学習装置 1（図 1）のグラフ構造生成手段 140 と同じ機能を有する。

ただし、グラフ構造生成手段 140 B は、文字マップにおいて局所値（ここでは、局所最大値）を持つ画素の位置として、単独文字の位置が 1 つしか検出されなかった場合、グラフ構造の生成を行わず、文字領域算出手段 18 に単独文字の位置のみを通知することとする。なお、単独文字の位置が 1 つも検出されなかった場合、図示を省略した表示装置にその旨を表示することとしてもよい。

【0067】

ノード属性算出手段 141 B は、ニューラルネットワークで構成された学習済のペア属性推定モデル N_2 を用いて、グラフ構造生成手段 140 B で生成されたグラフ構造と、特徴マップとに基づいて、単独文字同士のペア属性を算出するものである。

このノード属性算出手段 141 B は、ペア属性の出力先が文字領域算出手段 18 である点を除いて、文字領域検出モデル学習装置 1（図 1）のノード属性算出手段 141 と同じ機能を有する。

【0068】

モデル記憶手段 17 B は、文字領域検出モデル学習装置 1（図 1）で学習された文字領域検出モデル（単独文字検出モデル N_1 およびペア属性推定モデル N_2 ）を記憶するものである。このモデル記憶手段 17 B は、半導体メモリ等の一般的な記憶媒体で構成することができる。

【0069】

文字領域算出手段 18 は、単独文字検出手段 10 B で生成された文字マップと、ペア属性算出手段 14 B で算出されたペア属性とに基づいて、同じ文字列に含まれる単独文字の領域を統合した文字領域を算出するものである。

文字領域算出手段 18 は、単独文字領域検出手段 180 と、文字領域統合手段 181 と、を備える。

【0070】

単独文字領域検出手段 180 は、単独文字の位置における単独文字の領域を検出するものである。ここでは、単独文字領域検出手段 180 は、ペア属性算出手段 14 B からペア属性とともに入力される単独文字の位置（ここでは、局所最大値の位置）における単独文

10

20

30

40

50

字の領域を検出する。なお、単独文字領域検出手段180は、ペア属性算出手段14Bから、単独文字の位置を1つだけ入力した場合、1つの単独文字の領域を検出する。

【0071】

具体的には、単独文字領域検出手段180は、単独文字検出手段10Bで生成された文字マップにおいて、単独文字の位置を既知の前景とし、ラベルを割り当てる。また、単独文字領域検出手段180は、単独文字以外の領域を示す値として設定されている画素値（ここでは、“0.0”）の領域を背景とする。そして、単独文字領域検出手段180は、前景および背景と設定した画素以外の画素が前景であるとの単独文字の領域に属するかを判定することで、単独文字の領域を検出する。

【0072】

このように、前景と背景とを分割する手法は、一般的な領域分割手法を用いればよく、例えば、Watershed（分水嶺）アルゴリズムを用いることができる。Watershedアルゴリズムは、画像の局所値（ここでは、局所最大値）に前景を設定し、画像の輝度勾配によって前景の輪郭を検出する手法である。

これによって、単独文字領域検出手段180は、単独文字ごとの領域を検出することができる。

単独文字領域検出手段180は、検出した単独文字ごとの領域を、単独文字を識別するラベルとともに、文字領域統合手段181に出力する。

【0073】

文字領域統合手段181は、単独文字領域検出手段180で検出された単独文字の領域を、同じ文字列を構成する領域に統合するものである。

文字領域統合手段181は、ペア属性算出手段14で算出されたペア属性に基づいて、同じ文字列に属する単独文字領域検出手段180で検出された単独文字の領域を統合する。

この文字領域統合手段181は、予め定めた閾値（例えば、0.5）よりも大きい値となるペア属性の単独文字を同じ文字列に属するものとする。

【0074】

文字領域統合手段181は、統合した領域を、画像Iに含まれる文字領域として外部に出力する。なお、文字領域統合手段181は、単独文字が1つのみの場合、当該単独文字の領域を1文字の文字列とみなして文字領域を外部に出力する。

この文字領域統合手段181において、外部に出力する文字領域の出力形式は特に限定されるものではない。例えば、同じ文字列に含まれるすべての単独文字の領域に外接する外接矩形の4つの頂点の座標（合計8つの数値）、外接矩形の中心座標（あるいは左上座標）、幅および高さ（合計4つの数値）等である。なお、回転を含んで外接矩形を設定する場合であれば、外接矩形の中心座標（あるいは左上座標）、幅、高さおよび回転角（合計5つの数値）等である。

もちろん、出力形式は、外接矩形に限定されず、最小外接円や多角形ポリゴンであってもよい。

【0075】

以上説明したように文字領域検出装置2を構成することで、文字領域検出装置2は、ニューラルネットワークである文字領域検出モデル（単独文字検出モデルN₁およびペア属性推定モデルN₂）を用いて、画像内の文字領域を検出することができる。

これによって、文字領域検出装置2は、複数の文字列が狭い範囲に密集している場合等、複雑な状態で画像内に文字列が存在している場合でも、精度よく文字列の領域を検出することができる。

なお、文字領域検出装置2は、コンピュータを、前記した各手段として機能させるための文字領域検出プログラムで動作させることができる。

【0076】

〔文字領域検出装置の動作〕

次に、図10を参照（構成については適宜図9参照）して、本発明の第2実施形態に係る文字領域検出装置2の動作について説明する。なお、モデル記憶手段17Bには、予め

10

20

30

40

50

文字領域検出モデル学習装置 1 (図 1) で学習された文字領域検出モデル (単独文字検出モデル N_1 およびペア属性推定モデル N_2) が記憶されているものとする。

【0077】

ステップ S 3 0 において、単独文字検出手段 1 0 B は、画像 I を入力する。

ステップ S 3 1 において、単独文字検出手段 1 0 B は、モデル記憶手段 1 7 B に記憶されている単独文字検出モデル N_1 を用いて、画像 I に対応する画像特徴である特徴マップと、画像 I に対応する単独文字の領域分布を示す文字マップとを生成する。

【0078】

ステップ S 3 2 において、ペア属性算出手段 1 4 B のグラフ構造生成手段 1 4 0 B は、文字マップにおいて局所最大値を持つ画素の位置を単独文字の位置として検出する。

ステップ S 3 3 において、グラフ構造生成手段 1 4 0 B は、単独文字の位置を検出したか否かを判定する。

ここで、単独文字の位置を検出できなかった場合 (ステップ S 3 3 で No)、文字領域検出装置 2 は、動作を終了する。

一方、単独文字の位置を検出できた場合 (ステップ S 3 3 で Yes)、ステップ S 3 4 において、グラフ構造生成手段 1 4 0 B は、単独文字の位置が 2 以上検出されたか否かを判定する。

ここで、単独文字の位置が 2 以上検出されなかった場合 (ステップ S 3 4 で No)、ペア属性算出手段 1 4 は、ペア属性の算出を行わずに、ステップ S 3 7 に動作を移す。

【0079】

一方、単独文字の位置が 2 以上検出された場合 (ステップ S 3 4 で Yes)、ステップ S 3 5 において、ペア属性算出手段 1 4 B のグラフ構造生成手段 1 4 0 B は、ステップ S 3 1 で生成された文字マップに基づいて、単独文字のペアをノードとし、ノード同士で同一の単独文字を持つノード間をエッジで接続したグラフ構造を生成する。

ステップ S 3 6 において、ペア属性算出手段 1 4 B のノード属性算出手段 1 4 1 B は、モデル記憶手段 1 7 B に記憶されているペア属性推定モデル N_2 を用いて、ステップ S 3 1 で生成された特徴マップと、ステップ S 3 5 で生成されたグラフ構造とから、ノード属性として、単独文字同士のペア属性を算出する。

【0080】

ステップ S 3 7 において、文字領域算出手段 1 8 の単独文字領域検出手段 1 8 0 は、Watershed アルゴリズム等によって、文字マップにおいて、ステップ S 3 2 で検出された単独文字の位置における単独文字の領域を検出する。

【0081】

ステップ S 3 8 において、文字領域算出手段 1 8 の文字領域統合手段 1 8 1 は、ステップ S 3 6 で算出されたペア属性に基づいて、同じ文字列に属するステップ S 3 7 で検出された単独文字の領域を文字領域として統合する。なお、単独文字が 1 文字の場合、文字領域統合手段 1 8 1 は、単独文字が 1 文字の領域を文字列の文字領域とする。

ステップ S 3 9 において、文字領域統合手段 1 8 1 は、文字領域を所定の出力形式に変換して外部に出力する。

以上の動作によって、文字領域検出装置 2 は、画像内に存在する文字列の領域を検出することができる。

【0082】

以上、本発明の実施形態について説明したが、本発明は、これらの実施形態に限定されるものではない。

〔変形例〕

ここでは、図 2 で説明した単独文字検出モデル N_1 は、入力する画像 I の大きさ ($H \times W$) と、出力する特徴マップ M_f および文字マップ M_c の大きさ ($H \times W$) を、同じ大きさとした。しかし、この大きさは、高さ W と幅 W との比が同じであれば、必ずしも同じ大きさである必要はない。

【0083】

10

20

30

40

50

例えば、特徴マップ M_f および文字マップ M_c の大きさを、 $1/2$ ($H/2 \times W/2$)、 $1/4$ ($H/4 \times W/4$) 等、予め定めた縮小比で縮小した大きさとしてもよい。

この場合、文字領域検出モデル学習装置 1 は、学習用正解データ D_L の単独文字領域座標データ D_1 や、文字列領域座標データ D_2 の領域座標の座標値を同じ縮小比で縮小して使用すればよい。

また、この場合、文字領域検出装置 2 は、文字領域算出手段 18 において、出力する文字領域の座標を、縮小比の逆数で拡大すればよい。

これによって、文字領域検出モデル学習装置 1 および文字領域検出装置 2 における計算処理負荷を軽減させることができる。ただし、この場合、小さい文字列の検出精度を劣化させることになるため、処理負荷と精度とのトレードオフによって、特徴マップ M_f および文字マップ M_c の大きさを予め定めればよい。

10

【0084】

また、ここでは、図 6 で説明したペア属性推定モデル N_2 は、ノードの特徴量として、図 7 に示す特徴マップ M_f から生成される 2 つの単独文字の特徴量を合算したものをを用いた。しかし、ノードの特徴量は、これに限定されるものではない。

例えば、単独文字の特徴量を合算したものではなく、連結したものをを用いてもよい。その場合、「a b」，「b a」のようにノードを構成する単独文字が同じであっても、連結する順序が異なるものは異なるノードとして扱う方が望ましい。ただし、ノードの数が 2 倍になるため、メモリ消費量の観点から合算を使い、「a b」，「b a」を同一のノードとして扱うことが好ましい。

20

【0085】

また、例えば、ノードの特徴量には、特徴マップ M_f から生成される特徴量に、さらに、ノードに属する単独文字のペア間の距離、角度特徴等の幾何学的特徴量を付加してもよい。

具体的には、ペアとなる 2 つの単独文字の位置座標を $P_1 = (x_1, y_1)$ 、 $P_2 = (x_2, y_2)$ とした場合、以下の式 (1) に示すペア間の距離 d を用いればよい。また、角度特徴として、以下の式 (2)、式 (3) に示す正弦値 $\sin \theta$ 、余弦値 $\cos \theta$ を用いればよい。

【0086】

【数 1】

$$d = \|P_2 - P_1\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad \cdots \text{式(1)}$$

$$\sin \theta = (y_2 - y_1)/d \quad \cdots \text{式(2)}$$

$$\cos \theta = (x_2 - x_1)/d \quad \cdots \text{式(3)}$$

30

【0087】

これによって、文字領域検出モデル学習装置 1 は、ペア属性推定モデル N_2 をさらに精度よく学習することができる。また、文字領域検出装置 2 は、ペア属性推定モデル N_2 を用いてさらに精度よく文字領域を検出することができる。

40

【0088】

また、ここでは、ペア属性推定モデル N_2 を、グラフコンボリューションネットワーク (GCN) で構成した。

しかし、ペア属性推定モデル N_2 は、例えば、線形結合構造で構成された他のニューラルネットワークで構成しても構わない。ただし、ペアの属性を検出する精度と、メモリの使用効率の観点から、ペア属性推定モデル N_2 は、GCN で構成することが好ましい。

【0089】

また、ここでは、ペア属性を、2 つの単独文字が同じ文字列に含まれるか否かを示す属性としたが、さらに、他の属性を追加してもよい。

50

例えば、２つの単独文字が、「同じ文字列に含まれ、かつ、隣り合う文字であるか否か」、「同じ文字列に含まれ、かつ、一方の単独文字が文字列の先頭に位置するか否か」等の単独文字の位置に関する属性を追加してもよい。

この場合、文字領域検出装置２は、文字領域算出手段１８において、文字領域を出力する際に、単独文字の位置関係を属性として併せて出力すればよい。

この位置関係の属性は、文字領域内の文字認識を行う場合の有用な情報として活用することができる。

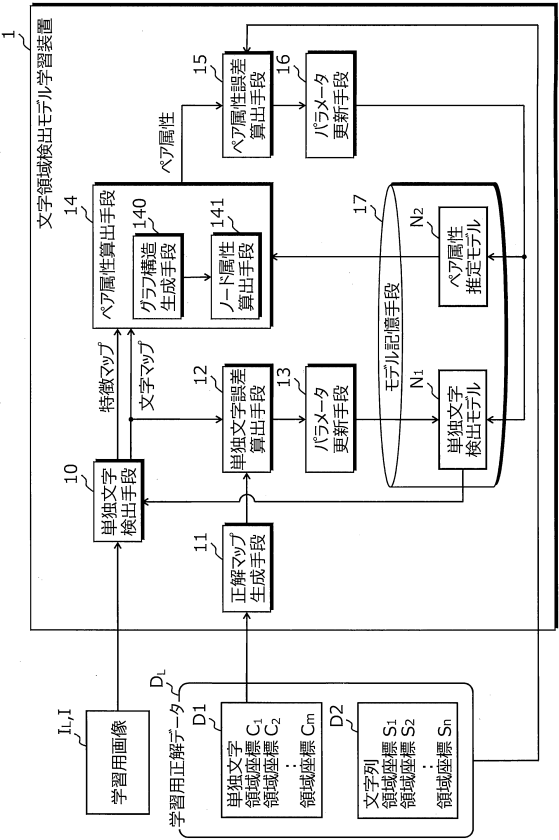
【符号の説明】

【００９０】

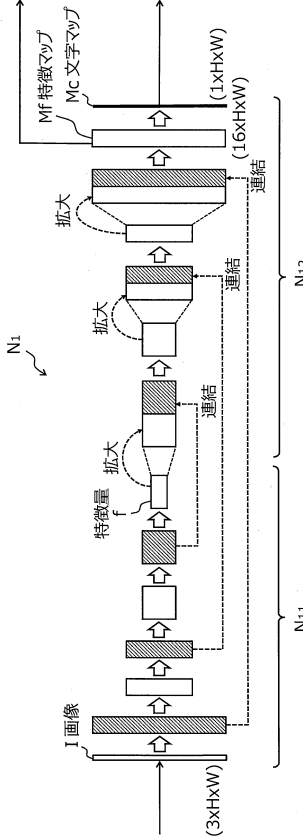
１	文字領域検出モデル学習装置	10
１０	単独文字検出手段	
１１	正解マップ生成手段	
１２	単独文字誤差算出手段	
１３	パラメータ更新手段（第１パラメータ更新手段）	
１４	ペア属性算出手段	
１４０	グラフ構造生成手段	
１４１	ノード属性算出手段	
１５	ペア属性誤差算出手段	
１６	パラメータ更新手段（第２パラメータ更新手段）	
１７	モデル記憶手段	20
２	文字領域検出装置	
１０Ｂ	単独文字検出手段	
１４Ｂ	ペア属性算出手段	
１４０Ｂ	グラフ構造生成手段	
１４１Ｂ	ノード属性算出手段	
１７Ｂ	モデル記憶手段	
１８	文字領域算出手段	
１８０	単独文字領域検出手段	
１８１	文字領域統合手段	
Ｎ１	単独文字検出モデル（文字領域検出モデル）	30
Ｎ１１	第１ネットワーク	
Ｎ２２	第２ネットワーク	
Ｎ２	ペア属性推定モデル（文字領域検出モデル）	
Ｍｆ	特徴マップ	
Ｍｃ	文字マップ	

【図面】

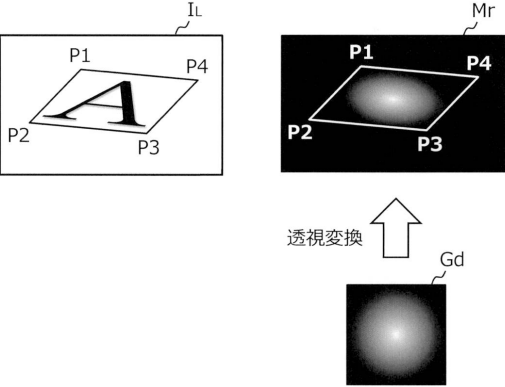
【図 1】



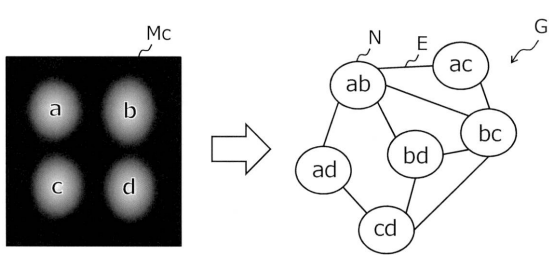
【図 2】



【図 3】



【図 4】



10

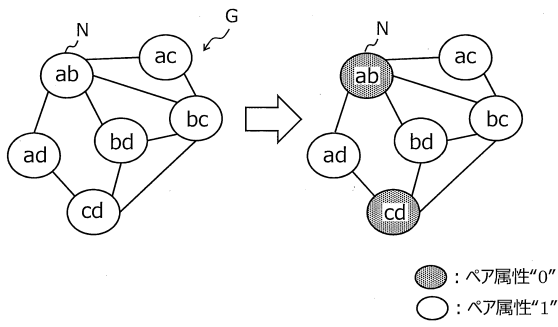
20

30

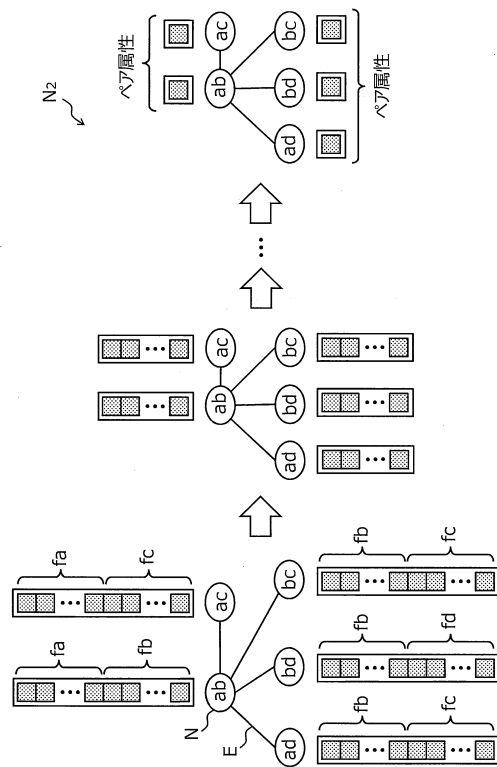
40

50

【図 5】



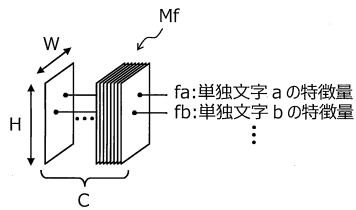
【図 6】



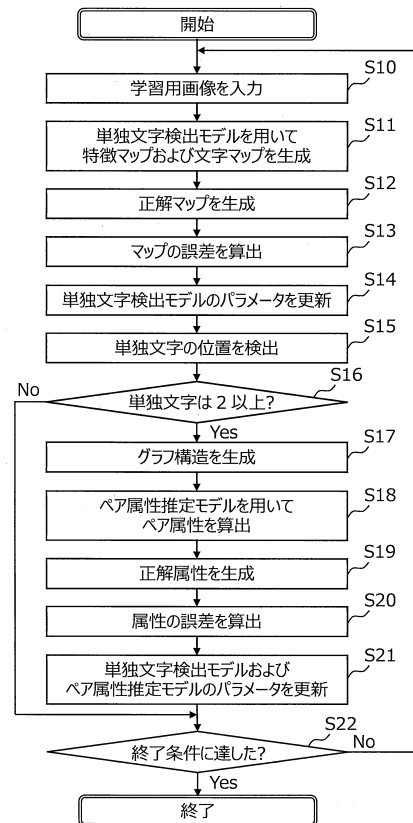
10

20

【図 7】



【図 8】

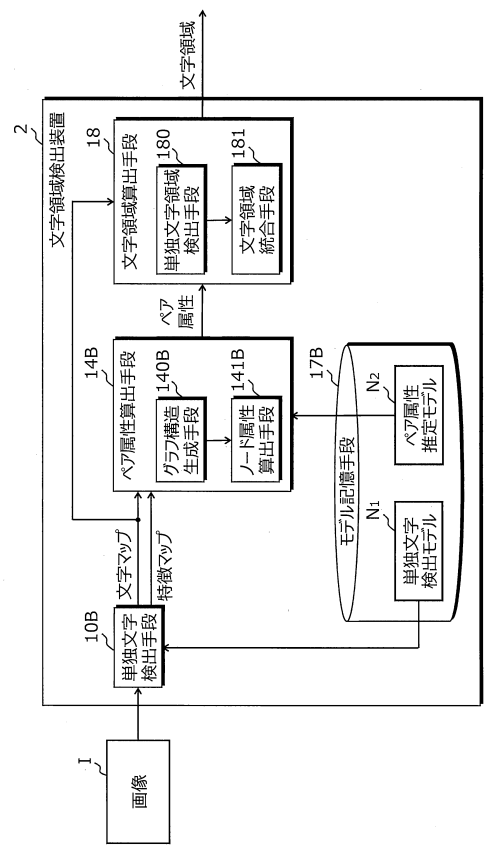


30

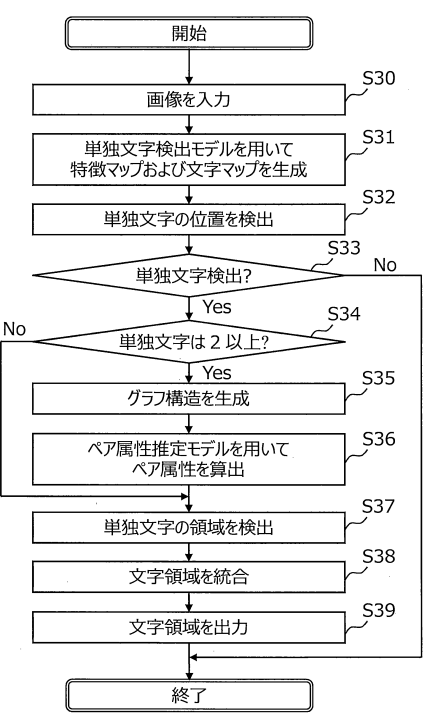
40

50

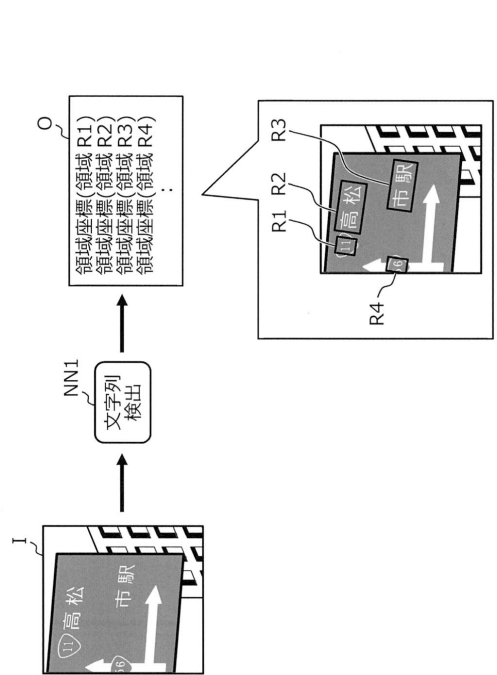
【図 9】



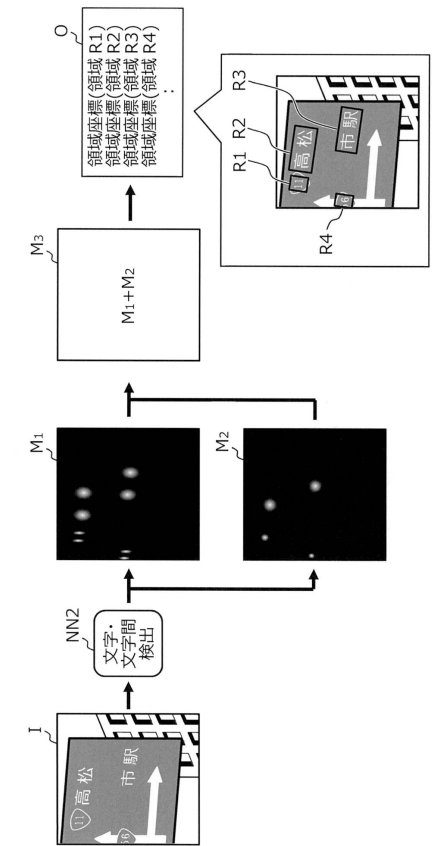
【図 10】



【図 11】



【図 12】



10

20

30

40

50

フロントページの続き

- (56)参考文献 特開 2 0 0 1 - 1 9 5 5 4 2 (J P , A)
国際公開第 2 0 1 9 / 1 7 4 4 0 5 (W O , A 1)
- (58)調査した分野 (Int.Cl., D B 名)
- G 0 6 T 7 / 0 0
G 0 6 V 3 0 / 0 0