

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号  
特許第5357331号  
(P5357331)

(45) 発行日 平成25年12月4日 (2013. 12. 4)

(24) 登録日 平成25年9月6日 (2013. 9. 6)

(51) Int. Cl.

F I

G O 6 T 7/00 (2006. 01)

G O 6 T 7/40 (2006. 01)

G O 6 T 7/00 1 5 0

G O 6 T 7/40 Z

請求項の数 26 (全 20 頁)

(21) 出願番号	特願2012-514018 (P2012-514018)	(73) 特許権者	000005326
(86) (22) 出願日	平成22年5月28日 (2010. 5. 28)		本田技研工業株式会社
(65) 公表番号	特表2012-529110 (P2012-529110A)		東京都港区南青山二丁目1番1号
(43) 公表日	平成24年11月15日 (2012. 11. 15)	(74) 代理人	100064414
(86) 国際出願番号	PCT/US2010/036656		弁理士 磯野 道造
(87) 国際公開番号	W02010/141369	(74) 代理人	100111545
(87) 国際公開日	平成22年12月9日 (2010. 12. 9)		弁理士 多田 悦夫
審査請求日	平成25年5月28日 (2013. 5. 28)	(72) 発明者	ランガナサン, アナンス
(31) 優先権主張番号	61/217, 930		アメリカ合衆国、オハイオ州 43067
(32) 優先日	平成21年6月4日 (2009. 6. 4)		-9705、レイモンド、21001 ス
(33) 優先権主張国	米国 (US)		テート ルート 739、ホンダ パテン
(31) 優先権主張番号	12/789, 292		ツ アンド テクノロジーズ ノース ア
(32) 優先日	平成22年5月27日 (2010. 5. 27)		メリカ、エルエルシー気付
(33) 優先権主張国	米国 (US)		
早期審査対象出願		審査官	新井 則和
			最終頁に続く

(54) 【発明の名称】 無作為多項ロジットを用いる意味場面区画

(57) 【特許請求の範囲】

【請求項1】

無作為多項ロジット ( R M L ) を使用して場面区画をするコンピュータを用いる方法であって、

前記コンピュータは、

複数の画像と、前記複数の画像のそれぞれについての複数のデジタル代表値と、を含む訓練画像集合であって、前記複数の画像のそれぞれの被写体が意味ラベルを有する訓練画像集合を受け取り、

前記訓練画像集合内の前記複数の画像のそれぞれに対応し、それぞれが複数の画素からなるテクストン画像であって、該テクストン画像の画素値のそれぞれが、前記訓練画像集合に含まれる、テクストン画像が対応する画像の画素値の代表値によって置換されるテクストン画像を生成し、

前記複数のテクストン画像から、1又は複数のテクスチャレイアウト特徴を選択し、

前記選択されたテクスチャレイアウト特徴に基づいて、前記 R M L 分類子の複数の多項ロジスティック回帰モデルを学習し、

前記訓練画像集合内の被写体の意味ラベルに基づいて、前記複数の多項ロジスティック回帰モデルの性能を評価し、

前記1又は複数のテクスチャレイアウト特徴を選択することは、

前記 R M L 分類子によって現在使用されているテクスチャレイアウト特徴の統計的有意性に基づいて、前記現在使用されているテクスチャレイアウト特徴を無作為に選択された

新たなテクスチャレイアウト特徴と交換することを含むこと、  
を特徴とする方法。

【請求項 2】

前記訓練画像情報内の前記画像に対応する前記複数のテクスツォン画像を生成することは、

前記訓練画像集合内の画像間のコントラストのばらつき及び輝度のばらつきを除去し、  
フィルタバンクを用いて前記訓練画像集合の前記画像を多重化し、  
前記多重化された画像をクラスタリングすることを含むこと、  
を特徴とする請求項 1 に記載の方法。

【請求項 3】

前記フィルタバンクは、

異なる縮尺におけるガウシアン、異なる縮尺におけるガウシアンの微分係数、及び、異なる縮尺におけるガウシアンのラプラシアンからなる 17 次元フィルタバンクであること、  
を特徴とする請求項 2 に記載の方法。

【請求項 4】

前記多重化された画像をクラスタリングすることは、

階層的 k 平均クラスタリング方法を使用して前記多重化された画像をクラスタリングすることを含むこと、  
を特徴とする請求項 2 に記載の方法。

【請求項 5】

前記多重化された画像の 17 次元ベクトルは、

前記ベクトルが属するクラスタの識別子によって表されること、  
を特徴とする請求項 4 に記載の方法。

【請求項 6】

前記特徴の統計的有意性は、

前記特徴に関連付けられた係数値と前記係数値の標準偏差とを比較することによって決定されること、  
を特徴とする請求項 1 に記載の方法。

【請求項 7】

前記テクスツォン画像から選択された前記テクスチャレイアウト特徴は、

前記テクスツォン画像の長方形領域及びテクスツォン語を含むこと、  
を特徴とする請求項 1 に記載の方法。

【請求項 8】

前記コンピュータは、

前記長方形領域の内側にある画素のうち画素値が前記テクスツォン語の値に等しい画素の数が、前記長方形領域の内側にある画素の総数に占める百分率を算出すること、  
を特徴とする請求項 7 に記載の方法。

【請求項 9】

前記テクスツォン語の値は、

前記テクスツォン語のクラスタ識別子に等しいこと、  
を特徴とする請求項 8 に記載の方法。

【請求項 10】

前記 RML 分類子の複数の多項ロジスティック回帰モデルを学習することは、

前記 RML 分類子の複数の多項ロジスティック回帰モデルごとに、

選択されたテクスチャレイアウト特徴のテクスツォン語の確率分布を推定し、

前記テクスツォン語の前記推定された確率分布の誤差を最小化し、

前記テクスツォン語の前記推定された確率分布を最適化することを含むこと、

を特徴とする請求項 1 に記載の方法。

【請求項 11】

10

20

30

40

50

前記 R M L 分類子の複数の多項ロジスティック回帰モデルを学習することは、  
 前記複数の多項ロジスティック回帰モデルからの、テキスト語の確率分布を平均し、  
 前記テキスト語ごとに最終出力ラベルを生成することを含むこと、  
 を特徴とする請求項 10 に記載の方法。

【請求項 12】

前記コンピュータは、  
 前記多項ロジスティック回帰モデルの性能が性能閾値より低い場合は、前記 R M L 分類子の多項ロジスティック回帰モデルを再学習すること、  
 を特徴とする請求項 1 に記載の方法。

【請求項 13】

前記コンピュータは、  
 場面区画の対象となる入力画像に対して前記 R M L 分類子を適用すること、  
 を特徴とする請求項 1 に記載の方法。

【請求項 14】

無作為多項ロジット ( R M L ) 分類子を使用して場面区画をするコンピュータシステムであって、

前記コンピュータシステムは、  
 画像テキスト化モジュール、特徴選択モジュール及び R M L 分類子を有し、  
 前記テキスト化モジュールは、  
複数の画像と、前記複数の画像のそれぞれについての複数のデジタル代表値と、を含む訓練画像集合であって、前記複数の画像のそれぞれの被写体が意味ラベルを有する訓練画像集合を受け取り、

前記訓練画像集合内の前記複数の画像のそれぞれに対応し、それぞれが複数の画素からなるテキスト画像であって、該テキスト画像の画素値のそれぞれが、前記訓練画像集合に含まれる、テキスト画像が対応する画像の画素値の代表値によって置換されるテキスト画像を生成し、

前記特徴選択モジュールは、  
 前記複数のテキスト画像から、1 又は複数のテキストチャレイアウト特徴を選択し、  
 前記 R M L 分類子は、  
 前記選択されたテキストチャレイアウト特徴に基づいて、前記 R M L 分類子の複数の多項ロジスティック回帰モデルを学習し、  
 前記訓練画像集合内の前記被写体の前記意味ラベルに基づいて、前記複数の多項ロジスティック回帰モデルの性能を評価し、

前記 1 又は複数のテキストチャレイアウト特徴を選択することは、  
 前記 R M L 分類子によって現在使用されているテキストチャレイアウト特徴の統計的有意性に基づいて、前記現在使用されているテキストチャレイアウト特徴を無作為に選択された新たなテキストチャレイアウト特徴と交換することを含むこと、

を特徴とするコンピュータシステム。

【請求項 15】

前記テキスト化モジュールは、  
 前記訓練画像集合内の画像間のコントラストのばらつき及び輝度のばらつきを除去し、  
 フィルタバンクを用いて前記訓練画像集合を多重化し、  
 前記多重化された画像をクラスタリングすること、  
 を特徴とする請求項 14 に記載のコンピュータシステム。

【請求項 16】

前記フィルタバンクは、  
 異なる縮尺におけるガウシアン、異なる縮尺におけるガウシアンの微分係数、及び、異なる縮尺におけるガウシアンのラプラシアンからなる 17 次元フィルタバンクであること、  
 を特徴とする請求項 15 に記載のコンピュータシステム。

## 【請求項 17】

前記多重化された画像の 17 次元ベクトルは、  
前記ベクトルが属するクラスタの識別子によって表されること、  
を特徴とする請求項 16 に記載のコンピュータシステム。

## 【請求項 18】

前記テキスト画像から選択された前記テキストチャレイアウト特徴は、  
前記テキスト画像の長方形領域及びテキスト語を含むこと、  
を特徴とする請求項 14 に記載のコンピュータシステム。

## 【請求項 19】

前記特徴選択モジュールは、  
前記長方形領域の内側にある画素のうち画素値が前記テキスト語の値に等しい画素の数が、前記長方形領域の内側にある画素の総数に占める百分率を算出すること、  
を特徴とする請求項 18 に記載のコンピュータシステム。

10

## 【請求項 20】

前記 RML 分類子は、  
前記 RML 分類子の複数の多項ロジスティック回帰モデルのそれぞれを設定し、  
選択されたテキストチャレイアウト特徴のテキスト語の確率分布を推定し、  
前記テキスト語の前記推定された確率分布の誤差を最小化し、  
前記テキスト語の前記推定された確率分布を最適化すること、  
を特徴とする請求項 14 に記載のコンピュータシステム。

20

## 【請求項 21】

前記 RML 分類子は、  
前記複数の多項ロジスティック回帰モデルからの、テキスト語の確率分布を平均し、  
前記テキスト語ごとに最終出力ラベルを生成すること、  
を特徴とする請求項 20 に記載のコンピュータシステム。

## 【請求項 22】

前記 RML 分類子は、  
前記多項ロジスティック回帰モデルの性能が性能閾値より低い場合は、前記 RML 分類子の多項ロジスティック回帰モデルを再学習すること、  
を特徴とする請求項 14 に記載のコンピュータシステム。

30

## 【請求項 23】

前記 RML 分類子は、  
場面区画の対象となる入力画像に対して前記 RML 分類子を適用すること、  
を特徴とする請求項 14 に記載のコンピュータシステム。

## 【請求項 24】

無作為多項ロジット (RML) 分類子を使用して場面区画をするコンピュータを機能させるためのコンピュータプログラムであって、

前記コンピュータプログラムは、  
前記コンピュータに対して、

複数の画像と、前記複数の画像のそれぞれについての複数のデジタル代表値と、を含む訓練画像集合であって、前記複数の画像のそれぞれの被写体が意味ラベルを有する訓練画像集合を受け取り、

40

前記訓練画像集合内の前記複数の画像のそれぞれに対応し、それぞれが複数の画素からなるテキスト画像であって、該テキスト画像の画素値のそれぞれが、前記訓練画像集合に含まれる、テキスト画像が対応する画像の画素値の代表値によって置換されるテキスト画像を生成し、

前記複数のテキスト画像から、1 又は複数のテキストチャレイアウト特徴を選択し、  
前記選択されたテキストチャレイアウト特徴に基づいて、前記 RML 分類子の複数の多項ロジスティック回帰モデルを学習し、

前記訓練画像集合内の前記被写体の前記意味ラベルに基づいて、前記複数の多項ロジス

50

ティック回帰モデルの性能を評価する処理を実行させ、

前記 1 又は複数のテクスチャレイアウト特徴を選択することは、

前記 R M L 分類子によって現在使用されているテクスチャレイアウト特徴の統計的有意性に基づいて、前記現在使用されているテクスチャレイアウト特徴を無作為に選択された新たなテクスチャレイアウト特徴と交換することを含むこと、

を特徴とするコンピュータプログラム。

【請求項 2 5】

前記コンピュータプログラムは、

前記コンピュータに対して、

前記 R M L 分類子の複数の多項ロジスティック回帰モデルを学習させる際に、

前記 R M L 分類子の複数の多項ロジスティック回帰モデルごとに、

選択されたテクスチャレイアウト特徴のテキストン語の確率分布を推定し、

前記テキストン語の前記推定された確率分布の誤差を最小化し、

前記テキストン語の前記推定された確率分布を最適化する処理を実行させること、

を特徴とする請求項 2 4 に記載のコンピュータプログラム。

【請求項 2 6】

前記コンピュータプログラムは、

前記コンピュータに対して、

場面区画の対象となる入力画像に対して前記 R M L 分類子を適用する処理を実行させること、

を特徴とする請求項 2 4 に記載のコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、一般的には、画像処理システムにおける場面区画に関する。具体的には、多項ロジスティック回帰モデルを使用する意味場面区画に関する。

【背景技術】

【0002】

高度道路交通システムのようなコンピュータ画像分析において、共通タスクとは、取得された画像内の街路場面を分類することである。当該共通タスクは、道、他の車両又は歩行者を検出し、高度道路交通システムのユーザが潜在的に危険な状態にいと警告を発することが多い。取得された画像内で関心被写体を検出するには、一般的に、画像に関心領域に区分し、及び/又は、さらに関心領域に関心被写体に区分することが必要になる。

【0003】

場面区画は、盛んに研究されている分野であり、ロボットシステム及び自動車システムでの適用例のように、現実世界の問題に対する広範囲の適応例を有する。1つの従来型場面区画方法は、特徴のコードブック、すなわちテキストン画像 (texton image) のような離散表現を使用する。特徴のコードブックは、画像の空間的關係を用いて、又は画像の空間的關係を用いずに、全体画像又は画像の特定の領域をモデル化する。入力画像のテキストン (texton) は、離散化されたテクスチャ語である。このテクスチャ語は、入力画像をフィルタバンクに通し、フィルタバンクからの出力をクラスタリングすることによって知られる。当該方法の問題は、当該方法が専ら画像レベルの場面区画向けであるということである。そこで、画像レベルの特徴及び統計データが不充分である画像中において、特に小さいサイズの被写体を検出し、その位置を決定することが課題となる。

【0004】

他の従来型場面区画方法は、入力画像のテクスチャレイアウト特徴を使用し、テキストンに影響を及ぼす特徴選択を促進する。当該従来型場面区画方法の一例は、テキストン生成及びテキストン分類の両者を目的とする、意味テクスチャ林 (semantic texture forest) を使用する。このような特徴の数は非常に大きいので、当該方法において使用される場面区画エンジンの訓練は、非常に低速であり、訓練データ集合のサイズが大きくなるに

10

20

30

40

50

つれて、そして、訓練データ集合内の被写体分類のバリエーションが増加するにつれて、このような場面区画の性能は低下する。

【図面の簡単な説明】

【0005】

【図1】本発明の一実施形態に係る、無作為多項ロジット分類子を使用する、意味場面区画のためのコンピュータシステムの図である。

【図2】本発明の一実施形態に係る、無作為多項ロジット分類子の多項ロジスティック回帰モデルを学習するフローチャートである。

【図3A】本発明の一実施形態に係る、意味場面区画の白黒図である。

【図3B】本発明の一実施形態に係る、意味場面区画の図3Aの白黒図に対応するカラー図である。

10

【図4】本発明の一実施形態に係る、画像テキスト化モジュールのブロック図である。

【図5A】本発明の一実施形態に係る、特徴選択の白黒図である。

【図5B】本発明の一実施形態に係る、特徴選択の図5Aの白黒図に対応するカラー図である。

【図6】本発明の一実施形態に係る、入力画像の意味場面区画のための、学習済の多項ロジスティック回帰モデルを適用するフローチャートである。

【発明を実施するための形態】

【0006】

図を参照しつつ、本発明の実施形態を説明する。同じ参照番号は、同一の構成要素又は機能的に類似する構成要素を示す。

20

【0007】

本明細書において、“一実施形態”又は“ある実施形態”と言う場合は、本発明の少なくとも1つの実施形態に、その実施形態に関連して記述される1つの特徴又は構造が含まれていることを意味する。本明細書のあちこちに“ある実施形態では”という語が出現しても、必ずしも同一の実施形態を指しているわけではない。

【0008】

後記する詳細説明のいくつかの部分は、アルゴリズム用語や、コンピュータメモリ内のデータビット作業を示す象徴的な表現による。これらのアルゴリズム的な説明や表現は、情報処理分野の当業者が、自らの業績の要旨を、同分野の他の当業者に最も効率的に伝えるために用いる手段である。アルゴリズムとは、ここでは、そして一般的にも、ある所望の結果に至る複数のステップ（命令）の首尾一貫したシーケンスのことを言う。ステップとは、物理量に対する物理的操作を要求するステップのことである。通常、必ずしも必要条件ではないが、それらの数値は、記憶され、送信され、合成され、比較されかつ操作され得る、電子的、磁氣的又は光学的信号の形を取る。これらの信号のことを、ビット、値、要素、シンボル、文字、語又は番号等と呼ぶことが主として用語の共通化の理由から便宜である。さらに、物理量に対する物理的操作を要求するステップの配列のうちいくつかのものは、一般性を失うことなく、モジュール又はコードデバイスと呼ぶことが便宜である。

30

【0009】

しかしながら、このような用語の全ては適当な物理量と関連付けられており、これらの物理量に付された単なる便宜的なラベルに過ぎない。後記において特段の説明がない限り、明細書本文全体を通じて、“処理”、“計算”、“算出”、“決定”又は“表示”等の用語を用いた説明は、（電子的な）物理量としてのデータを、コンピュータシステムのメモリ、レジスタ又は他の情報ストレージのなかで操作しかつ変形するコンピュータシステムや同様の電子的コンピューティングデバイスの動作や処理のことを言う。

40

【0010】

本発明のいくつかの側面は、アルゴリズムの形になったプロセスステップや命令を含む。本発明のプロセスステップや命令は、ソフトウェア、ファームウェア又はハードウェアによって実装され、ソフトウェアで実施される場合は、ダウンロードされることが可能で

50

あり、多様なオペレーティングシステムが用いる別のプラットフォームから操作されることも可能である。

【 0 0 1 1 】

本発明は、操作を実行する装置にも関する。この装置は、所与の目的を達成する専用装置であってもよいし、コンピュータに記憶されたコンピュータプログラムによって動作する汎用コンピュータであってもよい。このようなコンピュータプログラムは、コンピュータが読取り可能な媒体に記憶され得る。その媒体とは、コンピュータシステムバスに接続可能な、フロッピーディスク（登録商標）、光ディスク、CD-ROM、電磁光学的ディスク、ROM、RAM、EPROM、EEPROM、電磁的又は光学的カード、ASIC又は電子的命令を記憶し得るあらゆる媒体を含むが、これらに限定されない。さらに、本明細書で言うコンピュータは、単体のプロセッサを含んでもよいし、コンピュータ性能を増加させるために複数のプロセッサを用いた装置であってもよい。

10

【 0 0 1 2 】

ここでいうアルゴリズムやディスプレイは、もともと特定のコンピュータや他の装置に關係している訳ではない。ここで教示される内容に従って、プログラムとともに多様な一般目的システムが使用され得る。又は、必要とされる方法ステップを実行するためにさらに特化した装置を作成することも便宜である。これらのシステムのそれぞれについてどのような構成が必要となるかは、後記する明細書本文から明らかになる。さらには、本発明は特定のプログラム言語を参照して記載されるものではない。本発明の教示を実施するために、多様なプログラム言語が使用され得る。後記において特定の言語に言及した場合、それは本発明の実施可能性及びベストモードを開示するためである。

20

【 0 0 1 3 】

さらに、本明細書で使用される言語は、主として意味の取りやすさ及び教示目的から選択されているのであって、本発明の主旨を限定するために選択されているのではない。従って、本発明の開示は、特許請求の範囲にて記述されている本発明の範囲を例示するものであるが限定するものではない。

【 0 0 1 4 】

本発明の実施形態は、場面区画ごとに、無作為多項ロジット（RML）分類子を学習し、場面区画ごとに、入力画像に対し学習済のRML分類子を適用する。RML分類子を使用する場面区画システムは、事前にラベル付けされた訓練画像集合を通じてRML分類子を学習し、無作為に選択されたテクスチャレイアウト特徴によって、RML分類子が場면을区画する性能を増加させる。場面区画システムは、訓練画像集合内の画像間のコントラストのばらつき及び輝度のばらつきを除去し、訓練画像集合内のすべての画像を多重化し、クラスタリングすることによって、訓練画像を事前処理する。場面区画システムは、特徴に関連付けられた統計的に有意な測定値に基づいて、RML分類子によって使用される特徴を、無作為に選択されたテクスチャレイアウト特徴に置換する。RML分類子の多項ロジスティック回帰モデルのそれぞれは、選択されたテクスチャレイアウト特徴のテキスト語の確率分布を推定する。場面区画システムは、RML分類子の多項ロジスティック回帰モデルのそれぞれの確率分布を平均することによって、テキスト語ごとに、最終ラベルを生成する。学習済のRML分類子は、場面区画ごとに入力画像に対して適用され得る。本明細書に開示する場面区画システムは、ロボットシステム及び自動車システムでの適用例のような、現実世界の問題に対する広範囲の適応例も有する。

30

40

【 0 0 1 5 】

開示されるシステム（及び方法）の一実施形態は、無作為多項ロジット（RML）分類子を学習し、場面区画ごとにRML分類子を適用することを含む。システムは、画像テキスト化モジュール、特徴選択モジュール及びRML分類子を含む。画像テキスト化モジュールは、画像の被写体が事前にラベル付けされている訓練画像集合を受け取る。画像テキスト化モジュールはさらに、訓練画像集合から、対応するテキスト画像を生成する。特徴選択モジュールは、テキスト画像から、1又は複数のテクスチャレイアウト特徴を無作為に選択する。RML分類子は、複数の多項ロジスティック回帰モデルを有する

50

。RML分類子は、選択されたテクスチャレイアウト特徴を使用して、多項ロジスティック回帰モデルのそれぞれを学習する。RML分類子はさらに、場面区画ごとに、入力画像に対して、学習済の多項ロジスティック回帰モデルを適用する。

(システム概要)

【0016】

図1は、本発明の一実施形態に係る無作為多項ロジット(RML)分類子126を使用する、意味場面区画のためのコンピュータシステム100を示す。コンピュータシステム100は、メモリ120、入力制御装置130、出力制御装置140及びプロセッサ150を有する。ある実施形態においては、コンピュータシステム100は、訓練画像集合110Aを受け取り、RML分類子を学習する。他の実施形態においては、コンピュータシステム100は、学習済のRML分類子126によって、場面区画ごとに入力画像110Bを受け取る。訓練画像集合110Aは、複数の事前にラベル付けされた画像を有する。入力画像110Bは、コンピュータシステム100によって場面区画ごとに受け取られた、任意の静止画像又はビデオフレームである。訓練画像集合110Aを使用してRML分類子126を学習することは、図2を参照しさらに後記される。入力画像110Bに対してRML分類子126を適用することは、図6を参照しさらに後記される。

【0017】

ある実施形態においては、訓練画像集合110Aは、移動している自動車にカメラを取り付けて路上のオートバイを撮像することによって入手されたビデオシーケンスを有する。ビデオ画像は、移動している自動車のサイドバックミラー内でドライバーが視認するものと同じである。ビデオ画像は、撮像されるオートバイのタイプ、視点、ノイズの量及び照明の質の点で異なり得る。例えば、ある訓練画像集合110Aは、全体で約5800個のフレームを有する6個のビデオシーケンスから選択された63個のフレームを含む。選択されたフレームは、オートバイ、道、空及びその他のような、それらのフレームが表す被写体のタイプに応じてラベル付けされる。事前にラベル付けされた訓練画像集合110Aは、コンピュータシステム100によって受け取られ、RML分類子126を学習する。RML分類子126の学習結果は、RML分類子126の性能を評価するために訓練画像集合110Aのラベルと比較される。

【0018】

メモリ120は、データ、及び/又は、プロセッサ150によって実行され得る命令を格納する。命令は、ここに記述される任意の及び/又はすべての技術を実行するためのコードを含む。メモリ120は、DRAM装置、静的ランダムアクセスメモリ(SRAM)、フラッシュ(登録商標)RAM(非揮発性ストレージ)、これらの組合せ、又は、当業界で知られている他の記憶装置であってもよい。ある実施形態においては、メモリ120は、画像テクストン化モジュール122、特徴選択モジュール124、RML分類子126、及び、データストア128を有する。

【0019】

画像テクストン化モジュール122は、入力画像から対応するテクストン画像を生成するためのコンピュータ実行可能な命令を有する。入力画像から生成されたテクストン画像は、複数の画素を有する画像である。テクストン画像内のそれぞれの画素値は、入力画像内の対応する画素値の代表値である。具体的には、入力画像の画素値のそれぞれは、代表値、例えば、クラスタ識別子に置換される。この代表値は、事前処理後の入力画像の画素値に対応する。例えば、入力画像は、フィルタバンクを用いて多重化され、入力画像の画素ごとに、17次元ベクトルとなる。フィルタリング後の入力画像の17次元ベクトルがクラスタリングされると、17次元ベクトルのそれぞれは、当該ベクトルが属するクラスタの識別子(例えば、クラスタ番号)によって表される。入力画像の画素値のそれぞれが、対応するクラスタ識別子によって置換されると、テクストン画像が入手される。テクストン画像のそれぞれの画素は、テクスチャクラスタとも呼ばれるテクストン語である。画像テクストン化モジュール122は、図4を参照してさらに後記される。

【0020】



特徴選択モジュール 124 は、図 5 A 及び図 5 B を参照して後記される無作為テキストチャレイアウト特徴選択を実行するための、コンピュータ実行可能な命令を有する。テキストチャレイアウト特徴は、テキスト画像のうちの選択された部分、例えば、テキスト画像のうちの選択された長方形領域である。選択されたテキストチャレイアウト特徴ごとに、特徴選択モジュール 124 は、その画素値が所定のテキストチャクラスタのクラスタ識別子に等しい画素の百分率を算出する。図 2、図 3 A 及び図 3 B を参照して後記される、RML 分類子 126 の複数の多項ロジスティック回帰モデルを学習し適用するためのコンピュータ実行可能な命令を、RML 分類子 126 は有する。本発明のある実施形態においては、データストア 128 は、RML 分類子 126 の学習結果及び / 又は意味区画結果を格納する。

10

#### 【0021】

図 2 は、本発明の一実施形態に係る、無作為多項ロジット分類子 126 の多項ロジスティック回帰モデルを学習するフローチャートである。まず、コンピュータシステム 100 は、複数の画像を含む訓練画像集合を受け取る (ステップ 210)。ここでは、それぞれの画像は、オートバイ、道、空及びその他のような、画像内の被写体のタイプに応じてラベル付けされている。画像テキスト化モジュール 122 は、訓練画像集合内の画像から、対応するテキスト画像を生成する (ステップ 212)。特徴選択モジュール 124 は、テキスト画像からテキストチャレイアウト特徴を無作為に選択し、選択されたテキストチャレイアウト特徴を RML 分類子 126 に送信する (ステップ 214)。RML 分類子 126 の多項ロジスティック回帰モデルのそれぞれは、テキスト画像の選択された特徴内の所与のテキスト語の確率分布をモデル化する (ステップ 216)。RML 分類子 126 は、多項ロジスティック回帰モデルの性能を評価する (ステップ 218)。多項ロジスティック回帰モデルの性能が満足できるものである場合は、RML 分類子 126 は、学習処理を終了する。それ以外の場合は、RML 分類子 126 は、テキスト画像からテキストチャレイアウト特徴を新たに選択したうえで、学習処理を繰り返す (ステップ 220)。

20

#### 【0022】

図 3 A は、本発明の一実施形態に係る、意味場面区画の白黒図である。まず、訓練画像集合 302 は、図 2 で記述された多項ロジスティック回帰モデルを学習するために、図 1 のコンピュータシステム 100 に受け取られる。説明を単純化するために、図 3 A 及び図 3 B においては、本来複数の画像を含む訓練画像集合は、1つの画像 302 として表される。同様に、対応するテキスト画像も1つのテキスト画像 306 として表され、ラベル付けされた出力 (すなわち、訓練画像集合の画像に対応する場面区画された画像) も1つのラベル付けされた画像 310 として表される。訓練画像集合は、事前に処理され、多重化され、クラスタリングされ (304)、訓練画像集合の画像に対応するテキスト画像 306 を生成する。特徴選択モジュール 124 は、複数のテキストチャレイアウト特徴 308 (例えば、選択されたテキストチャレイアウト特徴を表す長方形) を無作為に選択し、これらを RML 分類子 126 に送信し、RML 分類子 126 の多項ロジスティック回帰モデルを学習する。図 3 A においては、RML 分類子 126 は、全部で N 個の多項ロジスティック回帰モデルを有する。ここで、N は、設計パラメータである。多項ロジスティック回帰モデルのそれぞれは、選択されたテキストチャレイアウト特徴内の所与のテキスト語の確率分布を算出し、入力画像 302 のラベル付けされた出力 310 を生成する。出力のラベルは、すべての入力画像の場面区画の結果を示す。図 3 B は、本発明の一実施形態に係る、意味場面区画の図 3 A の白黒図に対応するカラー図である。

30

40

(RML 画像テキスト化)

#### 【0023】

図 4 は、本発明の一実施形態に係る、画像テキスト化モジュール 122 のブロック図である。図 4 の実施形態においては、画像テキスト化モジュール 122 は、画像事前処理モジュール 410、画像多重化モジュール 420 及び画像クラスタリングモジュール 430 を含む。画像テキスト化モジュール 122 は、訓練画像集合を事前処理し、事前処理された訓練画像集合を多重化し、クラスタリングすることによって、訓練画像集合から

50

対応するテキスト画像を生成する。

【 0 0 2 4 】

一般に、複数の画像からなる画像集合は、訓練画像集合内の画像間のコントラストのばらつき及び輝度のばらつきを含む。コントラストのばらつき及び輝度のばらつきは、テキスト算出に対して悪影響を与え得る。画像事前処理モジュールジュール 4 1 0 は、テキスト算出の前に、画像集合内のコントラストのばらつき及び輝度のばらつきを除去する。具体的には、画像事前処理モジュールジュール 4 1 0 は、処理後の画像が平坦化されたヒストグラムとなりゼロ平均単位標準偏差を有するように、画像集合内の画像を処理する。

【 0 0 2 5 】

画像多重化モジュール 4 2 0 は、フィルタバンクを用いて、事前処理された訓練画像集合を多重化する。ある実施形態においては、画像多重化モジュール 4 2 0 は、17次元のフィルタバンクを使用する。フィルタバンクは、縮尺  $k$ 、 $2k$  及び  $4k$  ではガウシアンからなり、縮尺  $2k$  及び  $4k$  では  $x$  軸及び  $y$  軸方向のガウシアンの微分係数からなり、縮尺  $k$ 、 $2k$ 、 $4k$  及び  $8k$  ではガウシアンのラプラシアンからなり、 $k$  は設計パラメータである。ガウシアンは、CIE L a b 色空間のすべての3チャンネルについて算出される。当該フィルタ以外は、輝度チャンネルに対してのみ適用される。

【 0 0 2 6 】

画像クラスタリングモジュール 4 3 0 は、多重化された訓練画像集合をクラスタリングする。ある実施形態においては、画像クラスタリングモジュール 4 3 0 は、多重化された画像集合から生成された17次元ベクトルを、階層的  $k$  平均クラスタリング方法を使用してクラスタリングする。多重化された後の訓練画像集合の画像の17次元ベクトルのそれぞれは、そのベクトルが属するクラスタの識別子（例えば、クラスタ番号）によって表される。画像の画素値のそれぞれが、対応するクラスタ識別子によって置換されると、テキスト画像が入手される。テキスト画像のそれぞれの画素は、テキスト語である。テキスト語は、テキストクラスタとも呼ばれる。画像クラスタリングモジュール 4 3 0 は、三角不等式を使用し、 $k$  平均クラスタリングを加速し得る。当業者に知られている他のクラスタリング方法は、本発明の実施形態に対し容易に利用可能である。

（無作為多項ロジット（RML）分類子）

【 0 0 2 7 】

1つのRML分類子は、 $N$ 個の多項ロジスティック回帰モデルからなる。これらの回帰モデルのそれぞれは、以下の式（1）のように、入力ベクトル  $x$  が与えられると、ラベル  $y$  の確率分布をモデル化する。

【 0 0 2 8 】

【数 1】

$$\pi_{il} = p_i(y = l | x, \beta_i) = \begin{cases} \exp\left(\beta_{i0} + \sum_{f=1:M} \beta_{if} \phi_f(x)\right) / Z, & l = 1 : L-1 \\ 1/Z, & l = L \end{cases} \quad (1)$$

【 0 0 2 9 】

ここで、 $i$  及び  $l$  は、それぞれ、モデル集合及びラベル集合に対するインデクスである。 $Z$  は、確率分布の和を単位元にする正規化定数である。 $(\cdot)$  は、ベクトル  $x$  を入力とする特徴関数である。 $\phi_{i-1}$  は、被写体カテゴリ  $l$  ごとに検出関数を定義する、長さ  $(L-1)$  の係数のベクトルである。これらのベクトルのそれぞれを並べると、多項ロジスティック回帰モデルごとに、すべての係数を示す  $(L-1) \times (M+1)$  の行列  $\phi_i$  を得る。

【 0 0 3 0 】

RML分類子 1 2 6 を学習するということは、訓練画像集合から、係数  $\beta_i$  を学習するということである。訓練画像集合から、より小さな  $N$  個の集合が繰り返しサンプリングされ

10

20

30

40

50

る。当該小さな集合を用いて、RML分類子126の個々の多項ロジスティック回帰モデルが学習される。1つのモデルあたりM個の特徴というように、個々のモデルごとの特徴も無作為に選択される。ここで、Mは通常小さな数（例えば、10から20まで）である。使用される特徴の数を増加させると出力分散が増大するので、当該通常の数を超えて特徴の数が増加すると性能は低下する。RMLの最終的な出力ラベル分布は、以下の式（2）のように、個々のモデルの出力を平均することによって算出される。

【0031】

【数2】

$$\hat{\pi}_l = \sum_{i=1:N} \pi_{li} \quad (2)$$

10

【0032】

個々の回帰モデルごとの係数は、すべての訓練データに亘って、式（1）において記述される分散の誤差を最小化する最尤フレームワーク内で学習される。対数尤度を考慮すると、最大化されるべき関数は式（3）となる（モデル数のインデックスは省略されている）。

【0033】

【数3】

$$L(\beta|\{x,y\}) = \sum_{\{x,y\}} \log \pi_y \quad (3)$$

【0034】

20

ここで、 $\{x, y\}$ は、訓練画像データであり、 $y$ は、式（1）で定義される確率分布である。

【0035】

式（3）の対数尤度は、当業者に知られている勾配降下法又は二次法によって最適化され得る。例えば、ある特定の係数についての式（3）の勾配は、式（4）で与えられる。

【0036】

【数4】

$$\frac{\partial L}{\partial \beta_{lf}} = \sum_{\{x,y\}} \phi_f(x) (I(l=y) - \pi_{li}) \quad (4)$$

30

【0037】

ここで、 $I(\cdot)$ は、引数が真である場合に単位元を出力する指示関数である。（RMLテクスチャレイアウト特徴選択）

【0038】

式（1）から式（4）までに記述されるRML分類子126の性能は、テクスチャレイアウト特徴選択を通じて改良することが可能である。テクスチャレイアウト特徴は、テクストン画像の選択された部分、例えば、テクストン画像の選択された長方形領域である。RML特徴選択は、RML分類子によって現在使用されている特徴の統計的有意性に基づいて、現在使用されている特徴を無作為に選択された特徴と交換することによって実行される。そして、選択された特徴は、RML分類子によって使用され、その複数の多項ロジスティック回帰モデルを学習する。ある多項ロジスティック回帰モデルで使用されるすべての特徴が統計的に有意であるとき、無作為化された特徴選択によってモデルが改良される。

40

【0039】

ある実施形態においては、特徴選択モジュール124は、選択された特徴の統計的有意性を決定するために、縮尺の如何に依らず単一の検定を使用する。特徴に対応する係数の列がすべて極端に小さな値である場合は、その特徴は、式（1）に記述される回帰モデルに貢献しない。特徴の貢献を決定するための縮尺の如何に依らない検定は、 $t$ の値をそれらの標準偏差と比較することにより、 $t$ の値の統計的な有意性を確認することである。

$|\beta_{1f}| \leq 2\sigma_{1f}$  かつ  $\forall l \in [1 : L-1]$  であり、 $\sigma_{1f}$  が対応する標準偏差を表す場合、特徴  $\phi_f$  はモデルから消去され、それに替わって他の特徴が無作為に選択される。

その後、回帰モデルは、現在の係数を初期値として、最適化のために再学習される。消去された特徴は統計的に有意ではないので、他の特徴についての係数値は一般的には有意に変化することなく、再学習は効率的に進む。

【 0 0 4 0 】

係数推定の標準偏差は、以下の式 ( 5 ) のような、対数尤度関数のヘッシアンから算出され得る。

【 0 0 4 1 】

【 数 5 】

$$\frac{\partial^2 L}{\partial \beta_{chl}^2} = \sum_{\{x,y\}} -\phi_h(x)\phi_f(x)\pi_c(I(c=l) - \pi_l) \quad (5)$$

【 0 0 4 2 】

ここで、 $c$  及び  $l$ 、並びに、 $h$  及び  $f$  は、それぞれ、ラベル集合及び特徴集合に対するインデクスである。ヘッシアンの逆元は、そこから標準偏差が入手され得る、の共分散行列である。多項ロジスティック回帰モデル内のすべての特徴が統計的に有意であるとき、モデルは、最尤法に基づく無作為化された特徴選択によって改良される。 $L$  をモデルの対数尤度とすると、量  $-2 \log L$  は、カイ二乗統計量に従い、最良適合モデルに対応するとき最小値となる。したがって、1つの特徴のみが異なる2つのモデルの場合、小さな統計量を有するモデルが保持される。

【 0 0 4 3 】

テクスチャレイアウト特徴選択をさらに図示するために、以下に RML 特徴選択の一例の疑似コードを示す。

【 0 0 4 4 】

【 表 1 】

---

**Algorithm 1 RML Feature Selection**

---

**Input:** Current multinomial logistic regression model  $ML = \{\phi_{1 \dots L}, \beta\}$ , coefficient standard deviations

$\sigma$ , rounds of feature selection  $S$

$LL \leftarrow \log\text{-likelihood of } ML$

**for**  $i = 1$  to  $S$  **do**

$B \leftarrow \text{set of feature } \phi_{1f} \text{ s.t. } |\beta_{1f}| \leq 2\sigma_{1f} \forall f \in 1 : L-1$

**if**  $B$  is empty **then**

        Swap a randomly selected feature from current model for a randomly selected new feature  $\phi'$

**else**

        Swap a randomly selected feature from  $B$  for a randomly selected new feature  $\phi'$

**end if**

$ML' \leftarrow \text{new model learnt using maximum likelihood}$

$LL' \leftarrow \log\text{-likelihood of } ML'$

**if**  $LL' > LL$  **then**

$ML \leftarrow ML', LL \leftarrow LL'$

**end if**

**end for**

**return**  $ML$

---

【 0 0 4 5 】

ある実施形態においては、テクスチャレイアウト特徴は、長方形  $r$  及びテキスト用語  $t$

10

20

30

40

50

からなる。テキスト語は、前記では、テクスチャクラスとも呼ばれている。画素  $p$  ごとに、特徴選択モジュール 124 は、長方形  $r$  の内側にあるテキスト語  $t$  の割合を算出する。ここで、長方形  $r$  は、画素  $p$  が原点となる座標系内に入るように平行移動されている。例えば、選択されたテクスチャレイアウト特徴ごとに、特徴選択モジュール 124 は、画素値がテキスト語  $t$  のクラス識別子に等しい画素が、長方形  $r$  の内側にある百分率を算出する。テクスチャレイアウト特徴が画像内の局所的テクスチャ関係、例えば、ボートが通常水面上にあるという関係を取得することがわかる。さらに、当該関係は、複数のテクスチャレイアウト特徴値の線形結合として表されており、画素に基づく場面ラベリングを実行するのに充分である。

【0046】

10

テクスチャレイアウト特徴に基づいて RML 分類子を学習するために、特徴選択モジュール 124 は、まず、特徴候補の総数が  $N_r \times N_t$  となるように、 $N_r$  個の長方形領域を無作為に事前選択する。ここで、 $N_t$  は、学習処理において使用されているコードブック内のテキスト語の数である。その後、RML 分類子内の多項回帰モデルごとに、 $M < N_r \times N_t$  個の特徴の集合が無作為に選択され、式 (1) で与えられた分布を生成する。

【0047】

RML は、画素にラベル付けしたデータを使用して、教師付で学習される。画素及びその画素ラベルが評価された特徴値は、訓練インスタンスの構成要素となる。当該訓練データの無作為に選択された部分集合は、個々の回帰モデルを学習するのに使用される。実行時間に亘って、回帰モデル内の特徴は、入力画像について評価され、回帰モデルを通して、式 (2) のような出力ラベルを得る。

20

【0048】

図 5 A を参照する。図 5 A は、本発明の一実施形態に係る、特徴選択の白黒図である。画像 510 a は、RML 分類子学習処理において使用される訓練画像集合の画像の 1 つである。画像 510 b は、画像 510 a に対応するテキスト画像である。RML 分類子の個々の回帰モデルについて式 (1) で定義される係数 (512) は、最尤フレームワーク内で学習される。図 5 A の例では、合計 4 個のテクスチャレイアウト特徴 (520 a、530 a、540 a、550 a) が無作為に選択されている。選択されたテクスチャレイアウト特徴のそれぞれは、テキスト画像 510 b の一部である 1 つの長方形及び 1 つのテキスト語 (例えば、520 b、530 b、540 b、550 b) からなる。選択されたテクスチャレイアウト特徴ごとに、特徴選択モジュール 124 は、テクスチャレイアウト特徴の長方形の内側にある画素であって、その画素値が選択された特徴に関連付けられたテキスト語の値に等しい画素の百分率を算出する。テキスト語がテキスト化後のクラス識別子によって表されている場合は、特徴選択モジュール 124 は、選択された特徴長方形内にあり、画素値がクラス識別子に等しい画素の数の、長方形の内側にある画素の総数に対する比率を算出する。

30

【0049】

図 5 B は、本発明の一実施形態に係る、特徴選択の図 5 A の白黒図に対応するカラー図である。色付けされた長方形のそれぞれは、長方形ウィンドウを介し、さらに、テクスチャレイアウト特徴に関連付けられたテキスト語 (図では色が語を示す) を介してテクスチャレイアウト特徴を表している。画素値が画素ラベルを示していないとき、場面内のその画素についての特徴は、統計的に有意ではない。統計的に有意ではないのは、特徴の形状が大きすぎる又は小さすぎるからである。図 5 B の例では、学習されている多項回帰モデルにおいて係数  $\beta_4$  が乗じられる 4 つ目の特徴は、統計的に有意ではなく、特徴選択を通じて消去される。

40

(場面区画への RML 分類子の適用)

【0050】

前記した学習済の RML 分類子 126 は、従来型の場面区画方法と比較して正確性及び効率が改良されたテクスチャベースの場面区画のために使用され得る。図 6 は、本発明の一実施形態に係る、意味場面区画の対象となる入力画像に対して RML 分類子 126 を適

50

用するフローチャートである。学習済のRML分類子126を有するコンピュータシステム100は、まず、場面区画の対象となる入力画像を受け取る(ステップ610)。画像テクストン化モジュール122は、RML分類子学習処理の間に学習されたクラスタを使用して、対応するテクストン画像を生成する(ステップ612)。RML分類子126は、入力画像に対して学習済の多項ロジスティック回帰モデルを適用する(ステップ614)。例えば、RML分類子126は、RML分類子126の多項ロジスティック回帰モデルのそれぞれが式(1)を使用してモデル化した画素ラベルの確率分布を算出する。RML分類子126は、入力画像ごとに、区画ラベルによって表される意味画像区画を生成する(ステップ616)。例えば、RML分類子126は、RML分類子126の個々の多項ロジスティック回帰モデルの出力を、式(2)を使用して平均することによって、入力画像の最終出力ラベル分布を算出する。最後に、コンピュータシステム100は、画像の意味画像区画を出力する(ステップ618)。

10

#### 【0051】

高度道路交通システム及び他の適用例において使用される路上場面を分類するために、RML分類子は、事前にラベル付けされた画像を有する訓練画像集合を介して学習される。RML分類子は、複数の多項ロジスティック回帰モデルからなる。多項ロジスティック回帰モデルのそれぞれは、訓練画像集合から無作為に選択された特徴の部分集合について作動し、選択された特徴に対応する画素のラベルについての確率分布を出力する。最尤フレームワークを使用すると、多項ロジスティック回帰モデルは、大きな特徴空間内においても改良された性能で作動できる。学習済のRML分類子は、場面区画の対象となる入力画像に対して効率的に使用され得る。開示されるRMLベースの場面区画装置は、ロボットシステム及び自動車システムにおける適用のように、現実世界の問題に広範囲に適用できる。

20

#### 【0052】

本発明の特定の実施形態と応用例をここに記載したが、付属の特許請求の範囲に定義される本発明の技術的精神と範囲から逸脱しない限りにおいて、本発明はここに記載される厳密な構成および要素にのみ限定されることなく、本発明の方法および装置の構成、動作およびその詳細について、各種の変更、修正および変形を加えることが可能であることを理解すべきである。

#### 【符号の説明】

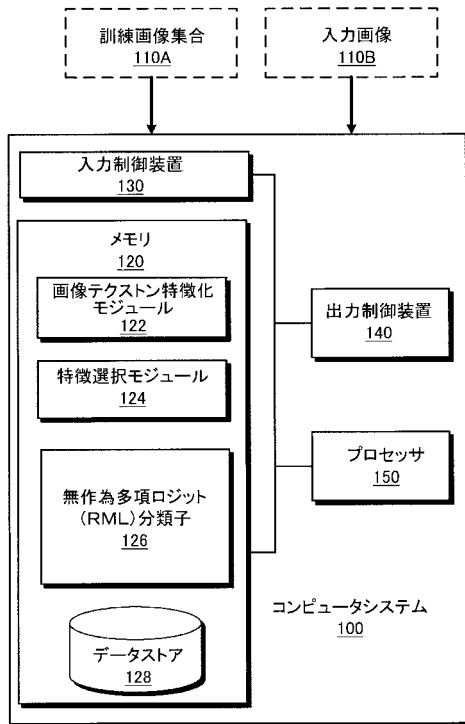
30

#### 【0053】

- 100 コンピュータシステム
- 110A 訓練画像集合
- 110B 入力画像
- 120 メモリ
- 122 画像テクストン化モジュール
- 124 特徴選択モジュール
- 126 無作為多項ロジット(RML)分類子
- 128 データストア
- 140 出力制御装置
- 150 プロセッサ

40

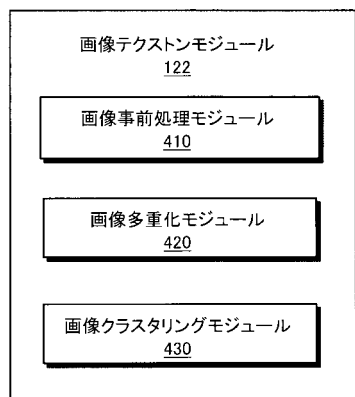
【図 1】



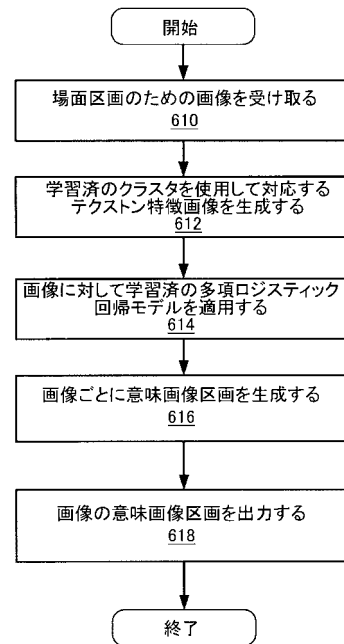
【図 2】



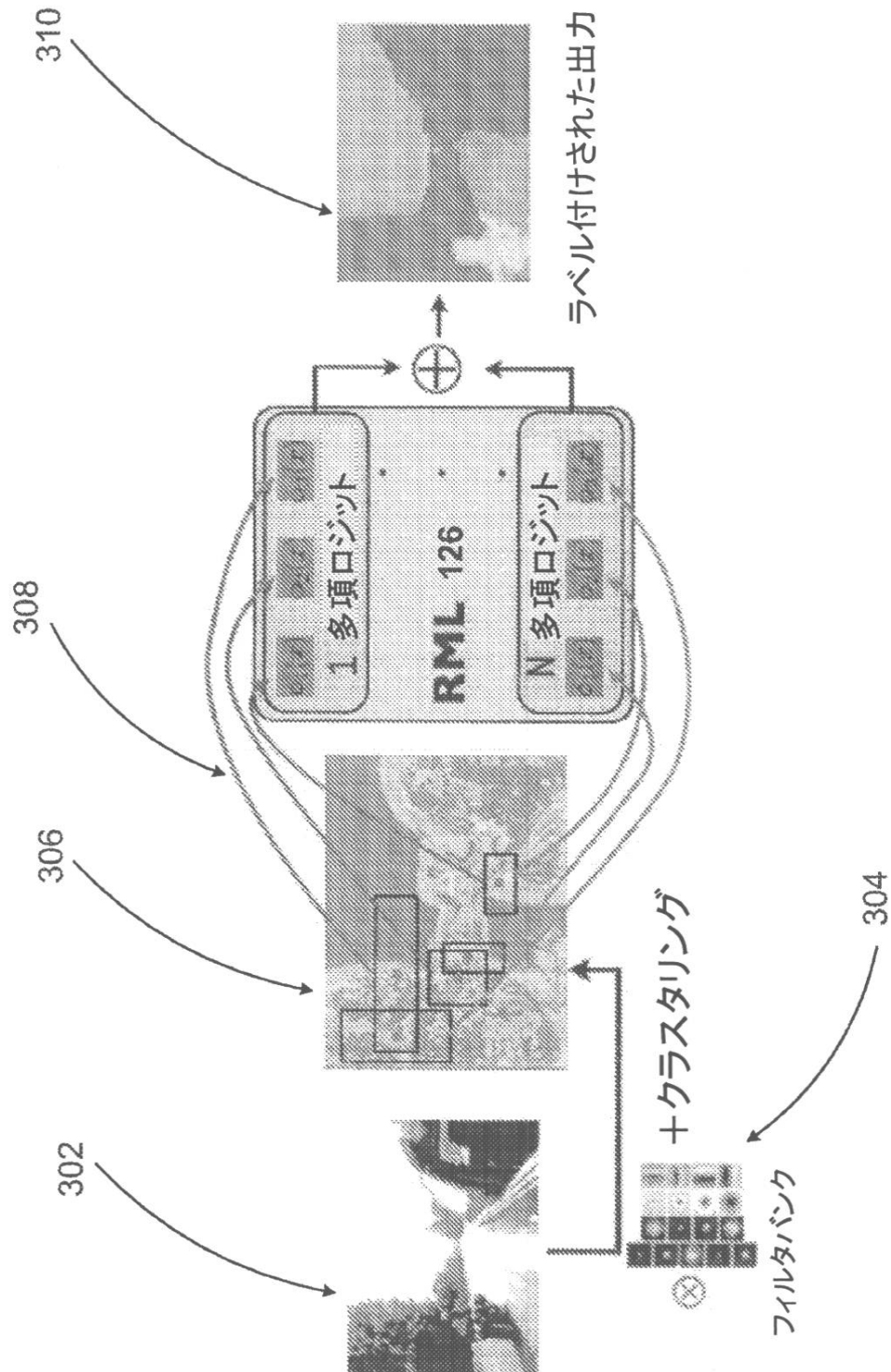
【図 4】



【図 6】

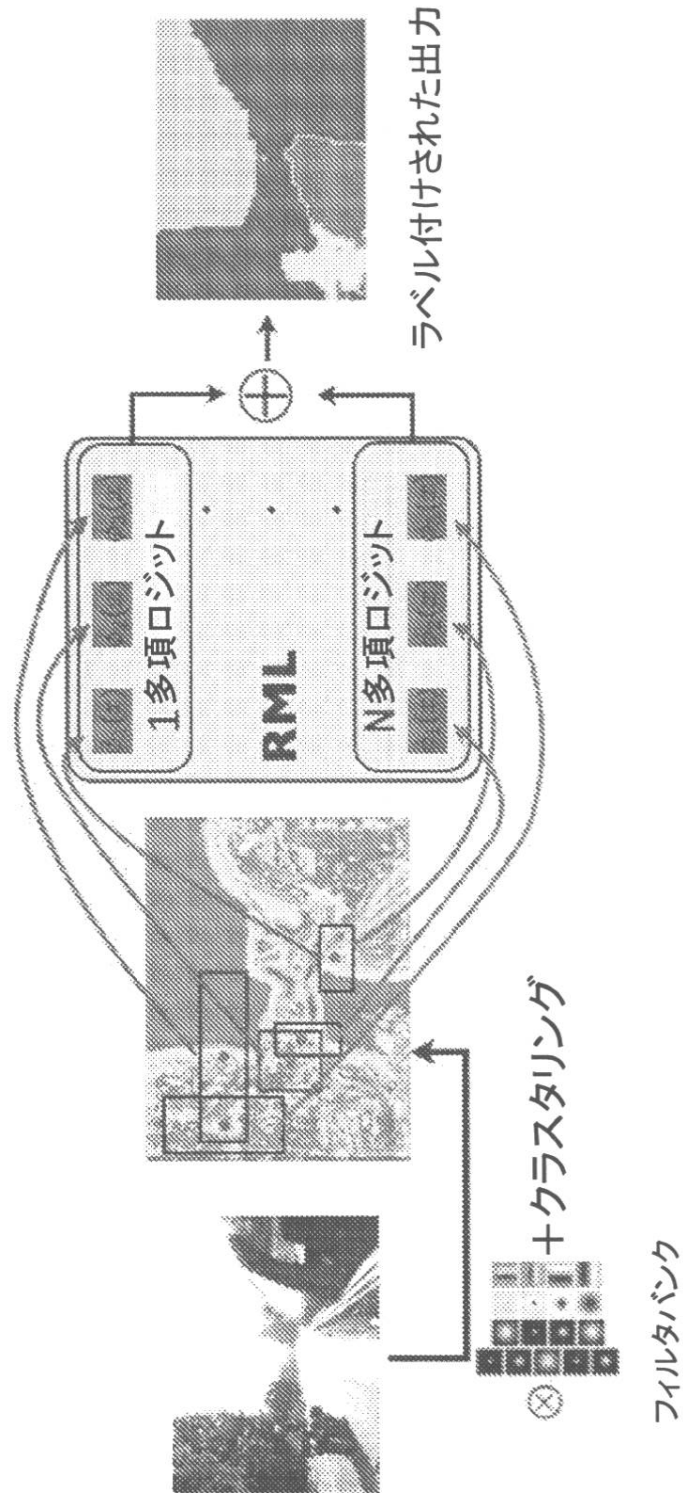


【図 3 A】

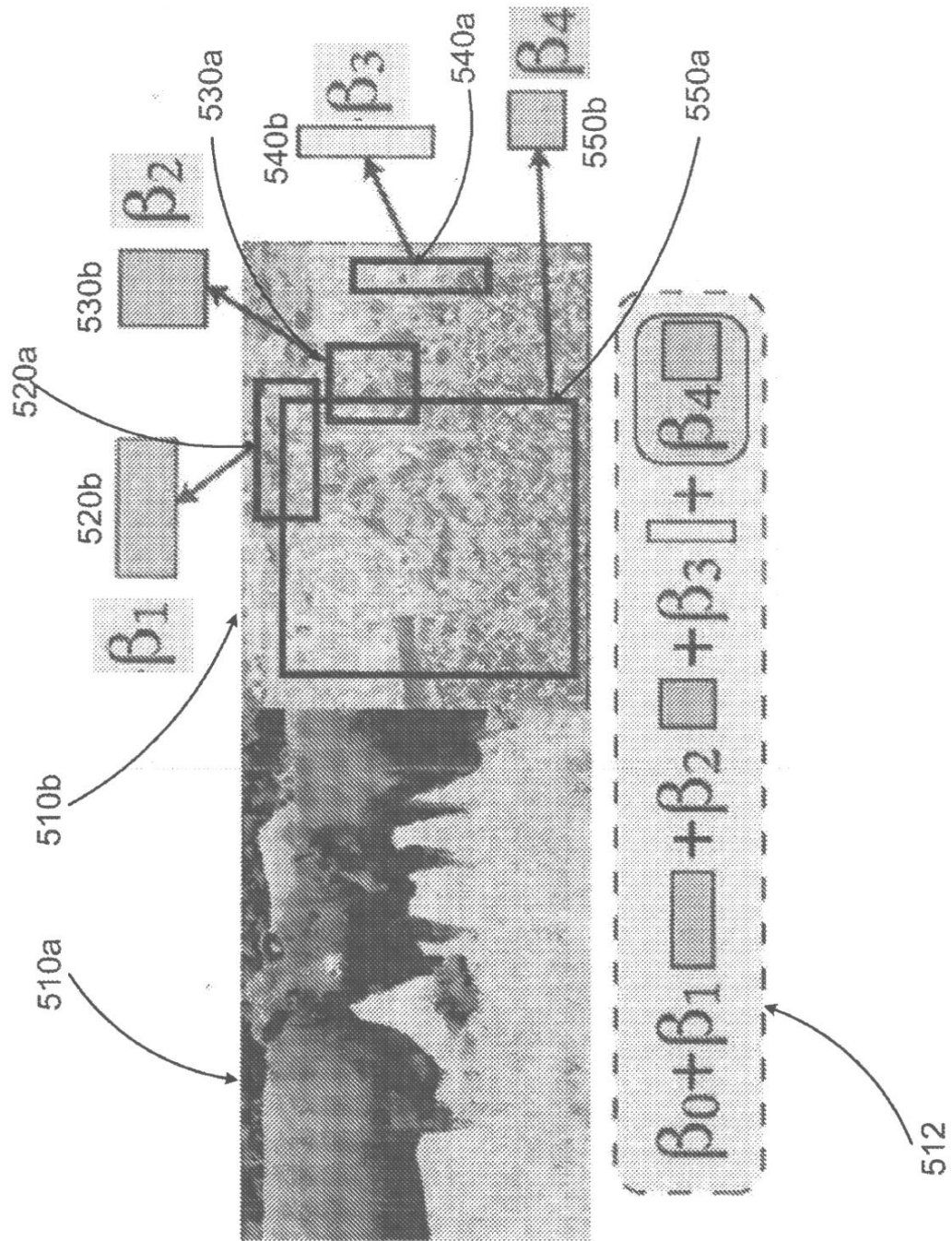




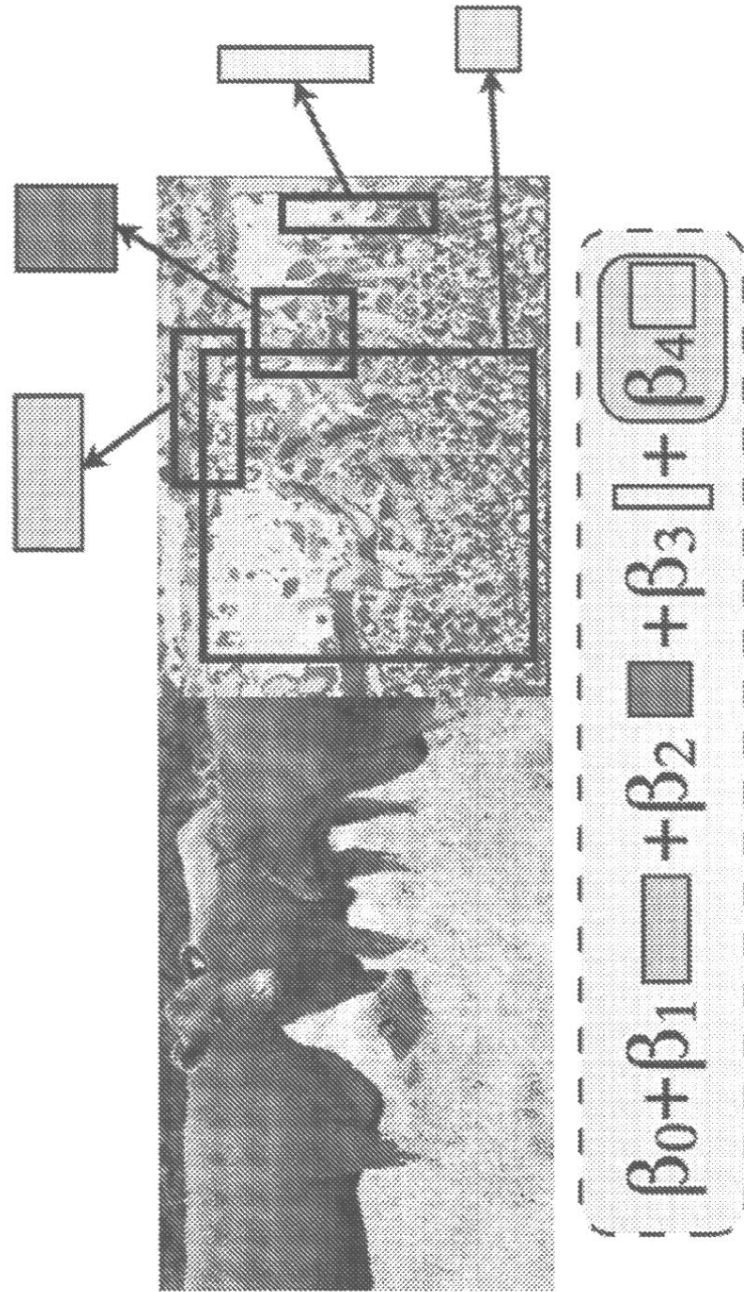
【図 3 B】



【図 5 A】



【 図 5 B 】



---

フロントページの続き

- (56)参考文献 米国特許出願公開第2008/0075361(US,A1)  
米国特許出願公開第2004/0252870(US,A1)  
米国特許出願公開第2008/0027917(US,A1)

- (58)調査した分野(Int.Cl., DB名)  
G06T 1/00 - 7/60