

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号  
特許第5461523号  
(P5461523)

(45) 発行日 平成26年4月2日 (2014.4.2)

(24) 登録日 平成26年1月24日 (2014.1.24)

(51) Int.Cl.

G O 6 F 17/30 (2006.01)

F I

G O 6 F 17/30 2 1 O D

G O 6 F 17/30 2 2 O B

G O 6 F 17/30 1 7 O B

G O 6 F 17/30 3 1 O Z

請求項の数 20 (全 16 頁)

(21) 出願番号	特願2011-504015 (P2011-504015)	(73) 特許権者	503260918
(86) (22) 出願日	平成21年4月10日 (2009.4.10)		アップル インコーポレイテッド
(65) 公表番号	特表2011-517818 (P2011-517818A)		アメリカ合衆国 9 5 0 1 4 カリフォル
(43) 公表日	平成23年6月16日 (2011.6.16)		ニア州 クパチーノ インフィニット ル
(86) 国際出願番号	PCT/US2009/002270		ープ 1
(87) 国際公開番号	W02009/131629	(74) 代理人	100092093
(87) 国際公開日	平成21年10月29日 (2009.10.29)		弁理士 辻居 幸一
審査請求日	平成24年4月2日 (2012.4.2)	(74) 代理人	100082005
(31) 優先権主張番号	12/102,395		弁理士 熊倉 禎男
(32) 優先日	平成20年4月14日 (2008.4.14)	(74) 代理人	100067013
(33) 優先権主張国	米国 (US)		弁理士 大塚 文昭
		(74) 代理人	100086771
			弁理士 西島 孝喜
		(74) 代理人	100122563
			弁理士 越柴 絵里

最終頁に続く

(54) 【発明の名称】 撮影位置シーケンス情報を用いた画像分類

(57) 【特許請求の範囲】

【請求項 1】

データ処理システムにより実行され、時間的に関係のある画像群の分類を容易にする方法であって、

前記時間的に関係のある画像群から、撮影位置シーケンスを特定するステップと、  
前記撮影位置シーケンスに集合的に関連付けられている撮影位置シーケンス情報を、予め決められた撮影位置シーケンスの特徴の複数のセットの各々であって、予め決められた分類に関連付けられているセットの各々と比較するステップと、

少なくとも前記比較ステップからの結果に基づいて、前記時間的に関係のある画像群に関連付けられる、特定された分類を特定するステップと、

前記特定された分類をプロセッサアクセス可能メモリシステムの中に保存するステップと、

を含むことを特徴とする方法。

【請求項 2】

前記撮影位置シーケンスは、少なくとも前記群の中の画像の少なくともいくつかに関連付けられている G P S 情報に基づいて特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 3】

前記時間的に関係のある画像群に集合的に関連付けられている画像コンテンツ情報を、予め決められた画像コンテンツの特徴の複数のセットの各々であって、予め決められた分

類に関連付けられているセットの各々と比較するステップをさらに含み、

前記特定するステップは、少なくとも前記2つの比較ステップからの結果に基づいて、前記特定された分類を特定することを特徴とする請求項1に記載の方法。

【請求項4】

前記特定された分類は、前記予め決められた分類の1つから特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5】

撮影位置シーケンス情報は、前記撮影位置シーケンスのトレースを含むことを特徴とする請求項1に記載の方法。

【請求項6】

前記トレースは、空間および時間領域が含むことを特徴とする請求項5に記載の方法。

【請求項7】

前記トレースの特徴を判断するステップをさらに含み、前記特徴は前記トレースによって表される空間的範囲か、前記トレースによって表される移動速度に関係し、前記比較ステップには、前記特徴を、前記予め決められた撮影位置シーケンスの特徴のセットのうちの少なくとも1つにおいて表される、対応する数値または数値範囲と比較することを含む請求項5に記載の方法。

【請求項8】

前記トレースの特徴を判断するステップをさらに含み、前記特徴はエントロピー、重心までの距離、分散、歪度、尖度、空間範囲、時区画、長さまたは速度に関係し、前記比較ステップには、前記特徴を、前記予め決められた撮影位置シーケンスの特徴のセットのうちの少なくとも1つにおいて表される、対応する数値または数値範囲と比較することを含む請求項5に記載の方法。

【請求項9】

前記予め決められた分類は、ハイキングと市内観光が含むことを特徴とする請求項1に記載の方法。

【請求項10】

前記予め決められた分類は、結婚式、野球、裏庭での活動、浜辺での活動、ドライブ旅行、ピクニック、クリスマス行事、誕生会をさら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】

データ処理システムに、時間的に関係のある画像群の分類を容易にするための方法を実行させる命令を記憶したプロセッサアクセス可能メモリを含むシステムであって、前記命令は、プロセッサに対し、

前記時間的に関係のある画像群から、撮影位置シーケンスを特定させること、

前記撮影位置シーケンスに集合的に関連付けられている撮影位置シーケンス情報を、予め決められた撮影位置シーケンスの特徴の複数のセットの各々であって、予め決められた分類に関連付けられているセットの各々と比較させること、

少なくとも前記比較ステップからの結果に基づいて、前記時間的に関係のある画像群に関連付けられる、特定された分類を特定させること、

前記特定された分類をプロセッサアクセス可能メモリシステムの中に保存させること、  
を実行させることを特徴とするシステム。

【請求項12】

前記撮影位置シーケンス情報は、撮影位置シーケンスのトレースを含むことを特徴とする請求項11に記載のプロセッサアクセス可能メモリを含むシステム。

【請求項13】

前記トレースには、空間および時間領域が含まれることを特徴とする請求項12に記載のプロセッサアクセス可能メモリを含むシステム。

【請求項14】

前記プロセッサに前記トレースの特徴を判断することを実行させる命令をさらに含み、

10

20

30

40

50

前記特徴は前記トレースによって表される空間的範囲か、前記トレースによって表される移動速度に関係し、前記比較は、前記特徴を、前記予め決められた撮影位置シーケンスの特徴のセットのうちの少なくとも1つにおいて表される、対応する数値または数値範囲と比較することを含む請求項12に記載のプロセッサアクセス可能メモリを含むシステム。

【請求項15】

前記プロセッサに前記トレースの特徴を判断することを実行させる命令をさらに含み、  
前記特徴はエントロピー、重心までの距離、分散、歪度、尖度、空間範囲、時区画、長さまたは速度に関係し、前記比較は、前記特徴を、前記予め決められた撮影位置シーケンスの特徴のセットのうちの少なくとも1つにおいて表される、対応する数値または数値範囲と比較することを含む請求項12に記載のプロセッサアクセス可能メモリを含むシステム。

10

【請求項16】

データ処理システムと、  
前記データ処理システムに通信可能に接続され、前記データ処理システムに、時間的に関係のある画像群の分類を容易にする方法を実行させる命令を記憶したメモリシステムと、を含むシステムであって、前記命令は、  
前記時間的に関係のある画像群から、撮影位置シーケンスを特定させること、  
前記撮影位置シーケンスに集合的に関連付けられている撮影位置シーケンス情報を、予め決められた撮影位置シーケンスの特徴の複数のセットの各々であって、予め決められた分類に関連付けられているセットの各々と比較させること、  
少なくとも前記比較からの結果に基づいて、前記時間的に関係のある画像群に関連付けられる、特定された分類を特定させること、  
前記特定された分類をプロセッサアクセス可能メモリシステムの中に保存させること、  
を実行させることを特徴とするシステム。

20

【請求項17】

前記撮影位置シーケンス情報は、前記撮影位置シーケンスのトレースを含むことを特徴とする請求項16に記載のシステム。

【請求項18】

前記トレースは、空間および時間領域を含むことを特徴とする請求項17に記載のシステム。

30

【請求項19】

前記トレースの特徴を判断する命令をさらに含み、前記特徴は前記トレースによって表される空間的範囲か、前記トレースによって表される移動速度に関係し、前記比較は、前記特徴を、前記予め決められた撮影位置シーケンスの特徴のセットのうちの少なくとも1つにおいて表される、対応する数値または数値範囲と比較することを含む請求項17に記載のシステム。

【請求項20】

前記トレースの特徴を判断する命令をさらに含み、前記特徴はエントロピー、重心までの距離、分散、歪度、尖度、空間範囲、時区画、長さまたは速度に関係し、前記比較は、前記特徴を、前記予め決められた撮影位置シーケンスの特徴のセットのうちの少なくとも1つにおいて表される、対応する数値または数値範囲と比較することを含む請求項17に記載のシステム。

40

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、人工知能と画像理解、特に、時間的に関係のある画像集合の分類に関する。

【背景技術】

【0002】

一般の人々による写真収集は広く普及している。このような写真の集合から意味論的に重要な情報を抽出することは、機械学習とコンピュータビジョンの分野において、盛んに

50

研究が進められている領域である。オブジェクト認識、すなわち顔、車、草、水、空その他、特定の種類のオブジェクトの検出に関する問題に焦点を当てた研究結果が数多く報告されている。そのほとんどは、画像の中から得られる低レベルの視覚的特徴（色、テクスチャ、線等）の利用に頼るものである。近年では、シーン検出や活動認識といった、意味論的により複雑な情報の抽出に対する関心がますます高まっている。たとえば、これまでのシステムでは、シーンやオブジェクトの視覚的分類を通じたイベント認識が試みられており、一例として、L. - J. リとL. フェイフェイ “What, where and who classfying events by scene and object recognition”, in Proc. IEEE Intl. Conf. on Computer Vision, 2007を参照されたい。このシステムは、ボロ 10  
、ボート競技、boche等、多数の独特のスポーツイベントの認識においては一定の成功を収めたと報告されているが、これは、この種のイベントの写真からは特異な視覚的特徴が観察できることによる。

#### 【0003】

上記の先行技術のいずれにおいても、従来通りの画像のクラスタリングと分類は、個々の画像に基づいて、たとえば色やエッジヒストグラムのような画像に基づく特徴、すなわち視覚的特徴集合（bag of visual features）だけを使って実行される（S. ラゼブニック、C. シュミット、J. ポンス “Beyond bags of features: spatial pyramid matching for recognizing natural scene categories”, in 20  
Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2006参照）。

#### 【0004】

しかしながら、同じイベントに属する複数の画像は相互に時間的および空間的相関関係を有するため、画像同士は独立していないことが多い。より詳しくは、個人的な画像集合には、画像の特徴以外に、コンテキスト情報が豊富にあり、意味論的理解のためには、通常、このようなコンテキスト情報が画像特徴を補うことになる。

#### 【先行技術文献】

#### 【非特許文献】

#### 【0005】

【非特許文献1】 “What, where and who classfying events by scene and object recognition”, L. - J. リとL. フェイフェイ, in Proc. IEEE Intl. Conf. on Computer Vision, 2007 30

【非特許文献2】 “Beyond bags of features: spatial pyramid matching for recognizing natural scene categories”, S. ラゼブニック、C. シュミット、J. ポンス, in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2006 40

#### 【発明の概要】

#### 【発明が解決しようとする課題】

#### 【0006】

したがって、画像間の関係を考慮した、改良された画像分類技術が求められている。

#### 【課題を解決するための手段】

#### 【0007】

本発明の各種の実施形態による、時間的に関係のある画像群の分類を容易にするシステムと方法によって、上記の問題に取り組み、当業界における技術的解決を実現する。本発明のある実施形態において、撮影位置シーケンスは、時間的に関係のある画像群から特定される。撮影位置シーケンス情報は、撮影位置シーケンスに集合的に関連付けられており、予め決められた撮影位置シーケンス特徴の複数のセットの各々と比較される。各セット 50

は、予め決められた分類に関連付けられている。時間的に関係のある画像群に関連付けられると特定される分類は、少なくとも比較ステップの結果に基づいて特定され、特定された分類は、プロセッサからアクセス可能なメモリシステムの中に保存される。

【発明の効果】

【0008】

本発明の利点の1つは、関連する撮影位置情報を有する個人的画像集合を使ってイベントを認識することである。

【0009】

上記の実施形態のほかにも、図面を参照し、以下の詳細な説明を読むことによって他の実施形態が明らかとなるであろう。

【図面の簡単な説明】

【0010】

【図1A】本発明の実施形態を実現するためのフローチャートである。

【図1B】本発明の実施形態を実現するためのフローチャートである。

【図2】空間的範囲と移動速度に関する異なるイベントの特徴を示す表である。

【図3】市内観光のイベントに関する空間と時間のトレース例を示す図である。

【図4】ハイキングのイベントに関する空間と時間のトレース例を示す図である。

【図5】分類表の例と、複数の方法により得られる手がかりからの認識結果を組み合わせるために用いられる信頼度を計算するための手順を示す図である。

【図6】本発明のある実施形態による、写真集合および関連する撮影位置トレース(trace)を用いたイベント認識システムを示す図である。

【発明を実施するための形態】

【0011】

集合的推論は、画像分類に有益なツールとなりうる。たとえば、特定の写真に関して、それが仕事かレジャーかを区別することは難しいかもしれないが、同じイベントに属する他の写真を見れば、その予測はもっとつきやすくなる可能性がある。本発明の実施形態において、画像分類は、関連付けられた撮影位置情報を有する、時間的に関係のある画像集合のコンテキストの中で実行される。「画像分類」という語句は、本明細書において、時間的に関係のある画像群を、予め決められた活動のセットの中の1つに分類するプロセスを指す。さらに、「イベント認識」と「活動認識」という語句は、本明細書において、画像分類の例である。またさらに、「画像」、「写真」、「フォト」という単語は、本明細書においては互換的に使用される。

【0012】

本発明は、本明細書に記載の実施形態を組み合わせたものも含む。「特定の実施形態」その他への言及は、本発明の少なくとも1つの実施形態の中にある特徴を指す。「ある実施形態」または「特定の実施形態」その他に別々に言及する場合は、必ずしも同じ実施形態(複数の場合もある)を指すとは限らず、これらの実施形態は、特にその旨が明記されている、または当業者にとって容易に明らかでないかぎり、相互に排他的ではない。「方法」または「複数の方法」その他への言及において単数形および/または複数形が使用されても、限定するものではない。

【0013】

さらに、別段の明確な断りがなく、文脈により必要でないかぎり、「または」という単語は、本明細書において、非排他的な意味で使用されることに注意すべきである。

【0014】

デジタルカメラにGPS(全地球測位システム)センサが搭載されていれば、撮影された個々の画像に関して、画像撮影の緯度、経度、時刻を収集することができる。さらに、撮影された個々の画像に関して、当業界で周知の技術を用いて、次のような情報を取得することができる。たとえば色やエッジヒストグラム等の低レベルの画像特徴、意味論的なオブジェクトとシーンの認識、オブジェクト検出器からの出力(たとえば顔)と画像分類器(屋内/屋外検出器、浜辺検出器等)からの出力、および同じ時間、同じ場所で撮影さ

10

20

30

40

50

れた画像（たとえば、同じイベントまたは同じ画像フォルダ内の画像）から得られるその他のコンテキスト情報。

【 0 0 1 5 】

図 6 は、本発明のある実施形態によるイベント認識のためのシステム 1 0 0 を示している。システム 1 0 0 は、データ処理システム 1 1 0、周辺システム 1 2 0、ユーザインタフェースシステム 1 3 0、プロセッサアクセス可能メモリシステム 1 4 0 を備える。プロセッサアクセス可能メモリシステム 1 4 0、周辺システム 1 2 0、ユーザインタフェースシステム 1 3 0 は、データ処理システム 1 1 0 に通信可能に接続される。

【 0 0 1 6 】

データ処理システム 1 1 0 は、図 1 および図 1 A のプロセス例を含む本発明の各種の実施形態のプロセスを実行する、1 つまたは複数のデータ処理デバイスを備える。「データ処理デバイス」または「データプロセッサ」という語句は、あらゆるデータ処理デバイスを含むものとされ、これにはたとえば、C P U（中央処理ユニット）、デスクトップコンピュータ、ラップトップコンピュータ、メインフレームコンピュータ、個人用携帯情報端末、B l a c k b e r r y（商標）、デジタルカメラ、携帯電話または、データを処理し、データを管理し、データを取り扱うその他のデバイスまたはその構成要素が含まれ、電気、磁気、光学、生物学的構成要素またはその他のいずれで実装されているかを問わない。

【 0 0 1 7 】

プロセッサアクセス可能メモリシステム 1 4 0 は、情報を保存するように構成された 1 つまたは複数のプロセッサアクセス可能メモリを含み、この情報とは、本明細書で説明する図 1 と図 1 A のプロセス例をはじめとする本発明の各種実施形態のプロセスを実行するのに必要な情報を含む。プロセッサアクセス可能メモリシステム 1 4 0 は、複数のコンピュータおよび/またはデバイスを介してデータ処理システム 1 1 0 に通信可能に接続された複数のプロセッサアクセス可能メモリを含む、分散型のプロセッサアクセス可能メモリシステムであってもよい。その一方で、プロセッサアクセス可能メモリシステム 1 4 0 は、分散型のプロセッサアクセス可能メモリである必要はなく、したがって、単独のデータプロセッサまたはデバイスの中に配置された 1 つまたは複数のプロセッサアクセス可能メモリであってもよい。

【 0 0 1 8 】

「プロセッサアクセス可能メモリ」という語句は、揮発性、不揮発性の電子、磁気、光またはその他のあらゆるプロセッサアクセス可能データ保存デバイスを含むものとし、たとえば、これらに限定されないが、レジスタ、フロッピー（登録商標）ディスク、ハードディスク、C o m p a c t D i s c、D V D、フラッシュメモリ、R O M、R A M であってよい。

【 0 0 1 9 】

「通信可能に接続される」という語句は、無線か有線かを問わず、デバイス、データプロセッサまたはプログラム間の、データが通信されてもよいあらゆる種類の接続を含むものとする。さらに、「通信可能に接続される」という語句は、単独のデータプロセッサ内のデバイスまたはプログラム間の接続、異なるデータプロセッサの中に配置されたデバイスまたはプログラム間の接続、およびデータプロセッサの中には配置されていないデバイス間の接続を含むものとする。この点に関して、プロセッサアクセス可能メモリシステム 1 4 0 はデータ処理システム 1 1 0 とは別に示されているが、当業者であれば、プロセッサアクセス可能メモリシステム 1 4 0 は、全体的または部分的にデータ処理システム 1 1 0 の中に格納されていてもよいことがわかるであろう。さらにこの点に関して、周辺システム 1 2 0 とユーザインタフェースシステム 1 3 0 はデータ処理システム 1 1 0 とは別に示されているが、当業者であれば、これらのシステムの一つまたは両方が、全体的または部分的にデータ処理システム 1 1 0 の中に格納されていてもよいことがわかるであろう。

【 0 0 2 0 】

周辺システム 1 2 0 は、デジタル画像をデータ処理システム 1 1 0 に供給するように構成された 1 つまたは複数のデバイスを含んでいてもよい。たとえば、周辺システム 1 2 0

10

20

30

40

50

は、デジタルビデオカメラ、携帯電話、一般的なデジタルカメラまたはその他のデータプロセッサを含んでいてもよい。データ処理システム110は、周辺システム120の中のあるデバイスからデジタルコンテンツ記録(digital content records)を受け取ると、そのデジタルコンテンツ記録をプロセッサアクセス可能メモリ140の中に保存してもよい。

#### 【0021】

ユーザインタフェースシステム130は、マウス、キーボード、別のコンピュータまたは、データ処理システム110へのデータの入力元となるあらゆるデバイスまたはデバイスの組み合わせを含んでいてもよい。この点に関して、周辺システム120はユーザインタフェースデバイス130とは別に示されているが、周辺システム120はユーザインタフェースシステム130の一部として含まれていてもよい。

10

#### 【0022】

ユーザインタフェースシステム130はまた、表示デバイス、プロセッサアクセス可能メモリまたは、データ処理システム110からのデータの出力先となるあらゆるデバイスまたはデバイスの組み合わせであってもよい。この点に関して、ユーザインタフェースシステム130がプロセッサアクセス可能メモリを内蔵している場合、図6ではユーザインタフェースシステム130とプロセッサアクセス可能メモリシステム140が別々に示されていても、これらのシステムはプロセッサアクセス可能メモリシステム130の一部であってもよい。

#### 【0023】

20

本発明は、画像分類、たとえば、人々が写真やビデオを撮影する一般的なイベントの自動認識等を容易にする。目標は、個人的な写真集合(一部はビデオを含んでいてもよい)から代表的なイベントを認識することであり、各イベントは、特定の環境で行われ、そのイベント中に撮影される以下の画像集合によって記録される具体的な人間の活動に対応する。

$E_i = \{I_j : j = 1, 2, \dots, n\}$

式中、 $I_j$ は1枚の画像(またはビデオ)である。本発明の好ましい実施形態では、頻繁に発生し、合理的に示差的な視覚的特徴を有する10種類のイベントまたは分類のリストを、 $C = \{\text{裏庭での活動、浜辺での活動、野球、誕生会、クリスマス行事、市内観光、ハイキング、ドライブ旅行、スキー、結婚式}\}$ と定義する。これら以外のイベントを定義し、含めることもでき、当業者であれば、本発明が上記のイベントに限定されないことがわかるであろう。リストCは、屋外と屋内の両方のイベントを含む。一般に、イベント認識はシーン認識より困難で、複雑であるが、それは、イベント認識にはより高レベルの意味論が関わり(L. J. リとL. フェイフェイ、"What, where and who classifying events by scene and object recognition" in Proc. IEEE Intl. Conf. on Computer Vision, 2007参照)、そのため、視覚的コンテンツがインスタンスごとに大きく異なる場合があり、分類器の性能を下げる原因となるからである。

30

#### 【0024】

40

本発明の実施形態で解決しようとする分類の問題は、次のように説明することができる。

入力：

視覚的イベントのデータベースは、ある分類器を設計するための教師データDとして入力され、 $D = \{E_i, 1(E_i)\}$ であり、この式中、 $E_i = \{T_{i,j}, I_{i,j} : j = 1, 2, \dots, |E_i|\}$ は、ある視覚的イベントを構成する画像セットを含む個々の画像フォルダ(群)を指し、 $|E_i|$ は、 $E_i$ の大きさを指す。また、 $T_{i,j} = \{x_{i,j}, y_{i,j}, t_{i,j}\}$ は、その写真の2次元位置( $x, y$ )と時間tのGPS記録であり、 $I_{i,j}$ は対応する画像である。集合的に、 $T_{i,j}$ のすべての記録がトレース $T_i$ を形成する。各画像フォルダに関して、Cの中の視覚的イベントクラス $1(E_i)$ の1つで

50

固有のラベルを付することができ、この場合、 $C = \{\text{裏庭での活動、浜辺での活動、野球、誕生会、クリスマス行事、市内観光、ハイキング、ドライブ旅行、スキー、結婚式}\}$ である。画像  $I_i, j$  のすべてが、所属する群  $E_i$  と同じラベルを共有する。教師データセット  $D$  には、 $C$  の中の各クラスについて十分なインスタンスが含まれているべきである。

出力：

新しい画像フォルダ  $E_i = \{T_j, j, I_i, j : j = 1, 2, \dots, |E_i|\}$  と学習された分類器に与えられるラベル  $1(E_i)$ 、たとえば  $1(E_i) = \text{ハイキング}$

【0025】

いずれのパターン分類の問題においても、高次元の特徴ベクトルを、事前に決定されたクラスのセットの1つに分類する必要がある。特徴間に冗長性があり、学習サンプルの数が限定されていることから、当初の高次元空間において1つの良好な分類器（最も可能性が高いのは、非線形分類器）を発見するのは困難であることが多い。これに対して、低次元特徴空間の中で多くの弱分類器を発見するほうが容易であり、この場合、発見された部分空間で構築された各分類器は、弱い分類ルールとしての役割を果たす。本発明のある実施形態では、有効な方法として、低次元空間で発見されたこれらの弱分類器を、多クラスブースティングによって組み合わせ、最終的な分類器を作る。ブースティングは当業界において周知であり、同様の他の手順を用いて最終的な分類器を作っても、本発明の範囲から逸脱しない。

【0026】

図1に、本発明のある実施形態による、時間的に関係のある画像群を分類する方法が示されている。まず、ステップ10で、撮影位置情報を有する、時間的に関係のある画像群を取得する。この画像群（本明細書では画像集合と呼ぶこともある）は、当業界で公知の技術により、時間的および位置的に関係のある写真群にすでに選別されているか、あるいは選別される。さらに、この画像集合には、GPS受信センサを備えるデジタルカメラを使って、ジオタグが付けられていてもよい。あるいは、ジオタグが付された写真を、独立したGPS受信機と（たとえば、Bluetoothリンクを通じて）通信可能なデジタルカメラを使って取得することもできる。あるいはまた、撮影後に写真に手作業で座標タグを付けるか、写真をまとめてGPSロギングデバイスと同期化させる方法もある。特定のイベントに対応するジオタグ付きの写真集合に関して、ステップ20で、はじめにGPS座標とタイムスタンプ情報を抽出し、空間と時間のスパースなトレース(sparse trace)を形成する。このトレースは、撮影位置のシーケンスと考えられる。位置情報を伴う写真が、時間的に関係のある画像群全体についてのトレースを形成するのに十分な量だけあれば、各写真が位置情報に関連付けられていることは必須でない点に注意されたい。

【0027】

次にステップ30で、トレースの特徴を計算し、トレースの形状または構造を特徴付ける。ステップ40で、このようなトレースの特徴を使って、その集合についての対応するイベントまたは分類を推測する。推測は、撮影位置シーケンスに集合的に関連付けられた情報を、予想される撮影位置シーケンス特徴の複数のセットの各々と比較することによって行われ、各セットは、あるイベントの分類に関連付けられている。

【0028】

ステップ20 - 40で位置シーケンストレースのような撮影位置シーケンス情報を使用することに加え、集合を分類するために、その集合の中の画像コンテンツ情報も任意で使用してよく、これがステップ50 - 70に示されている。特に、ステップ50で画像コンテンツ、すなわち画素情報を個々の写真から抽出してもよく、次のステップ60で、画像コンテンツから視覚的特徴を計算する。ステップ70で、同じイベントの中の個々の写真すべてにより集合的に表される視覚的情報から、イベントまたは分類を推測することができる。ここでも、推測は、時間的に関係のある画像群に集合的に関連付けられる画像コンテンツ情報を、予想される画像コンテンツ特徴の複数のセットの各々と比較することによって行われ、各セットは、あるイベントの分類に関連付けられている。平行する2通りの



推測経路があるため、ステップ80では、トレース特徴と視覚的特徴から別々に得られた推測結果を組み合わせることにより、共同的にイベントを推測できる。

【0029】

最後に、ステップ90で、推測されたイベント情報を使って、大きなデータベースの中で特定のイベントの写真を検索するか、他の用途のために写真集合に注記を付けることができる。

【0030】

図1Aが示す別の方法では、ステップ85において、トレース特徴と視覚的特徴の組み合わせ、すなわち複合情報から共同的にイベントを推測し、トレース特徴と視覚的特徴から別々に推測する中間のステップが省かれている。

【0031】

いくつかの例により、本発明の性能が、視覚的情報だけを使用する場合よりどのように優れているかを示すことができる。ある人々のグループがドライブ旅行に出て、ある主要都市を通ったとする。また、彼らはその都市と路上で写真を撮影したとする。視覚的情報だけを使用した場合、視覚的分類器はおそらく、このイベントにおける多数の写真を「市内観光」として正しく分類するかもしれないが、イベント全体としては誤った特徴づけを行う可能性がある。しかしながら、撮影位置シーケンスがこれらの写真に関連付けられていれば、写真群全体が、(都市を通過しているが)ドライブ旅行として正しく分類される可能性が高くなる。情報を融合した結果、このイベントはドライブ旅行であると最終的に判断される(本発明で、ドライブ旅行のイベントに都市の写真が含まれる可能性があることが学習される点に注意されたい)。

【0032】

別の例として、ある人々のグループが森にハイキングに行ったとする。裏庭の写真にも木々が含まれている可能性があるため、視覚的分類器は、これを裏庭でのイベントであると誤って判断するかもしれない。しかしながら、ここでも、その写真に撮影位置シーケンスが関連付けられていれば、本発明の実施形態が、イベント全体を裏庭のイベントではなくハイキングであると正しく判断する可能性が高くなる。

【0033】

視覚的特徴しか適用できなかった従来の画像および視覚的イベント分類の問題とは異なり、本発明の実施形態は、(1)GPS特徴と(2)視覚的特徴という2種類の特徴を利用する。特に、GPS特徴は、同じイベントに関係のある画像集合に対して定義されるため、個々の1枚の画像だけでなく、画像集合全体の特徴である。本発明の文脈内では、写真集合は、フォルダとも呼ばれる。

【0034】

図1に関して述べたスパースなトレースは、各種のイベントを差別化する上で非常に有益となりうる。図2にまとめるように、単純に空間的範囲と移動速度だけによって、イベントを、それぞれ前述のようなラベルに関連付けられるクラスにまとめることができる。本発明の実施形態は、写真が、当業界で周知の技術により、時間と位置において関係のある写真群に事前に選別されていることを前提としてもよい。市内観光、ハイキング、ドライブ旅行等のイベントは大きな空間的範囲にわたり、これに対して裏庭、浜辺、野球、ピクニック、結婚式のイベントは、空間的に制約されている傾向がある。ドライブ旅行であれば、人々はかなり高速で移動するが、市内観光、ハイキング、裏庭、浜辺、野球、ピクニック、結婚式での移動ははるかに遅い。範囲や速度よりもっと高度な特徴を設計して、トレースの形状や構造を特徴付けることができる。

【0035】

記録されたGPS情報に基づき、各画像フォルダは、空間-時間トレースで表され、 $T_i = \{ (x_{ij}, y_{ij}, T_{ij}) \mid j = 1, \dots, |T_i| \}$  となり、式中、 $|T_i|$  はフォルダの大きさ、すなわち画像の数に等しい。空間-時間領域における点のシーケンスは、そのもとになる画像撮影中の撮影者の移動を示すものとなる。このシーケンス単独で、空間-時間の中のその点において撮影された、対応する画像とは無関係に、イベ

10

20

30

40

50

ント認識作業のための情報となる。図3と図4は、それぞれ、数回の市内観光イベントとハイキングイベントに関するGPSトレースを示している。これらのトレースは、分類器を指定するための教師データDに使用してもよく、上述の分類問題の「入力」の部分とされる。注目すべき点は、これら2種類のイベントのトレースの形状特徴は相互に大きく異なるが、それと同時に、同じ種類のトレースは同様に見えることである。どちらのイベントも、空間と時間において比較的広い範囲にわたるが、市内観光のトレースのほうがトレースの形状の規則性が高く、これはもともになる都市の街区と街路の構造による。比較すると、ハイキングのトレースは、ジグザグの動きとすることもできるが、よりランダムで任意に見える。言い換えると、同じ種類のイベントに関する異なる移動インスタンスの中に構造的な類似性がある一方で、異なる種類のイベントに対応する移動の間には構造的な相違がある。

10

#### 【0036】

上記のような空間-時間トレースを考慮し、本発明の好ましい実施形態では、ステップ50において、次の13種類の特徴がトレースの形状または構造を表すものとして抽出され、それらの異なる重要性は学習によって決定され、一般性を損なうことなく[0, 0, 1, 0]の範囲内とされる。

エントロピー（特徴2つ）：時間および空間領域

重心までの距離（特徴1つ）

分散（特徴2つ）：時間および空間領域

歪度（特徴2つ）：時間および空間領域

20

尖度（特徴2つ）：時間と空間領域

空間的範囲（特徴1つ）と時区間（特徴1つ）

トレース長さ（特徴1つ）と平均速度（特徴1つ）

#### 【0037】

たとえば、以下に、上記の空間的範囲と空間的分散を定義する。(1)トレース $T_i$ の空間的範囲 $SR(T_i)$ は、視覚的イベントの活動の空間的範囲を説明する：

$$SR(T) = (\max\{x_j | x_j \in T\} - \min\{x_j | x_j \in T\}) \times (\max\{y_j | y_j \in T\} - \min\{y_j | y_j \in T\}) \quad (1)$$

(2)トレースの空間分散 $SV(T_i)$ は、活動の空間的分散を説明する：

#### 【数1】

30

$$SV(T) = \frac{1}{2} [VAR(\{x_j | x_j \in T\}) + VAR(\{y_j | y_j \in T\})]; \quad (2)$$

#### 【0038】

他の種類の特徴は低レベルの視覚的特徴であり、これらはステップ60において画像から計算される。視覚的ボキャブラリは、たとえばシーン認識のための視覚認識システムを構築するための有効な方法であることがわかっている(S.ラゼブニック、C.シュミット、J.ボンス、“Beyond bags of features: spatial pyramid matching for recognizing natural scene categories” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2006参照)。画像は、固定されたグリッドによって分けられ、画像パッチの順序付けされないセットとして表される。このような画像パッチについての適当な記述文が計算され、値域にまとめられて、「視覚的ボキャブラリ」が形成される。本発明の実施形態において、これと同じ方法を採用し、各画像グリッドを特徴付けるための色とテクスチャの両方の特徴を考慮するように拡張してもよい。

40

#### 【0039】

本発明のある実施形態において、色の特徴を抽出するために、1つの画像グリッドをさらに $2 \times 2$ の等しい大きさのサブグリッドに分ける。次に、各サブグリッドについて、R, G, Bの平均値を抽出して、 $4 \times 3 = 12$ の特徴ベクトルを形成し、これが4つのサブグリッドの色情報を特徴付ける。テクスチャ特徴を抽出するために、各サブグリッドにお

50

いて、方向に関する8つの値域を有する $2 \times 2$ のアレイからなるヒストグラムを計算する。したがって、 $4 \times 8 = 32$ 次元のSIFT記述子を用いて、各画像グリッド内の構造を特徴付けるが、これはラズベニックの考え方と似ている。本発明のある実施形態において、画像が200k画素より大きければ、これをまず200k画素の大きさに変える。次に、 $16 \times 16$ の画像グリッドサイズを設定し、重複するサンプリング間隔を $8 \times 8$ とする。一般に、1つの画像からそのようなグリッドが117できる。

【0040】

画像グリッドからすべての未加工の画像特徴を抽出した後に、本発明の実施形態において、教師データベースの中のすべての画像グリッドを、 $k$ -meansクラスタリングを通じてクラスタリングすることにより、別の色およびテクスチャのボキャブラリを構築する。どちらのボキャブラリも、500の大きさに設定される。あるイベントの中の全グリッドを累積させる（画像集合）ことによって、各イベントについて2つの正規化されたヒストグラム、すなわち、それぞれ色およびテクスチャのボキャブラリの単語分布に対応する $h_c$ と $h_t$ が得られる。 $h_c$ と $h_t$ を連結すると、その結果として、正規化された単語ヒストグラム、

【数2】

$$\sum_{i=1}^{1000} h_i(E) = 1.$$

となる。ヒストグラムの中の各値域は、対応する単語の発生頻度を示す。したがって、特徴ベクトル $f$ の応答は、以下のようになる。

【数3】

$$f_i(E) = \begin{cases} 1 & \text{if } h_i(E) > \theta_i \\ 0 & \text{else,} \end{cases} \quad (3)$$

式中、 $0 < \theta_i < 1$ は量子化閾値である。簡略化された指数関数を考えると、 $f_i$ は、そのイベント中に $i$ 番目の単語が十分頻繁に発生する場合には正の応答を示し、それ以外の場合の応答は $null$ となる。視覚的ボキャブラリには $c$ と $t$ の2種類があり、 $c$ は色のプリミティブ、 $t$ はテクスチャのプリミティブである。完全なボキャブラリを $c$ と $t$ として、色およびテクスチャの両方の情報を表すように、視覚的特長を計算することができる。

【0041】

低レベルの視覚的特徴のほかに、オブジェクトおよびシーン検出から得られる意味論的な視覚的特徴も考えることができる。各画像につき、シーンカテゴリのセット、

$S = \{\text{屋外、人工、浜辺、夕日、山、...}\}$

と、オブジェクトカテゴリのセット、

$O = \{\text{草、青い空、砂、雪、...}\}$

を定義することができる。

【0042】

上記のようなシーンとオブジェクトに関するバイナリSVM分類器のセットを構築することはよく知られており、たとえばラズベニックを参照されたい。決定論的判断を下すのではなく、本発明の好ましい実施形態では、ある画像がある特定のオブジェクトかシーンを含んでいるか否かを示すために、 $[0, 1]$ からの範囲の確率論的スコアを保持する。各画像には、複数のシーンまたはオブジェクトラベルが付けられてもよいことに注意されたい。したがって、各画像 $I$ に関して、結果は各数値が $[0, 1]$ の範囲の $k$ 次元の実数値特徴ベクトルとなり、 $k$ はシーンとオブジェクトのカテゴリの総和に等しく、各数値はそれぞれのSVM分類器からの検出スコアに対応する。

【0043】

視覚的特徴が計算されると、指数損失を用いた段階的加法モデリング(SAMME)と呼ばれる多クラスAdaBoost法を使って、教師データセット $D$ の中の画像を用いる視覚的分類器と、教師セット $D$ の中の撮影位置シーケンストレースを用いる撮影位置シー

10

20

30

40

50

ケンストレース分類器を構築することができる。SAMMEの詳細は、(J. Zhu (Zhu), S. El Alami, H. Zhou (Zhou), T. Hastie "Multi-class AdaBoost", Technical Report 2005, Stanford University) に記載されている。SAMMEは、Kバイナリ分類器ではなく、1つのKクラス分類器を構築することによって、Kクラス問題を直接取り扱う。SAMMEを用いる解決法はベイズの分類法則と一致しており、したがって、誤分類のエラーを最小限にするのに最適である。

【0044】

人気のある融合方式(たとえば、ブースティング)は、多くの弱分類器を線形に組み合わせ、強分類器を作っている。最終的な決定は、

【数4】

$$\hat{H}(\mathcal{D}_q) = \sum_{i=1}^L w_i h_i(\mathcal{D}_q), \quad (4)$$

となる。式中、 $w_i$  は最適化が必要な重み、 $h_i(\mathcal{D}_q) \in \{0, 1\}$  は、弱分類器  $h_i(\mathcal{D}_q)$  からの応答を示す。

【0045】

トレースの特徴、低レベルの視覚的特徴および意味論的な視覚的特徴の各々に関する個々の分類器が構築されたところで、本発明は、平行して行われるこれら分類の結果を、情報融合によって組み合わせる。ナイーブな融合方式の場合、たとえば、8つのイベントクラスに関して、以下を計算することができる。

GPS出力:  $h_g(X) = [4.2 \quad 0.0 \quad 0.5 \quad 4.9 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.7]$

視覚的出力:  $h_v(X) = [1.2 \quad 0.7 \quad 4.1 \quad 1.7 \quad 3.1 \quad 0.0 \quad 2.0 \quad 0.0]$

- 出力の総和:  $h(X) = h_g(X) + h_v(X) \quad (5)$

$h(X) = [5.4 \quad 0.7 \quad 5.6 \quad 6.6 \quad 3.1 \quad 0.0 \quad 2.0 \quad 0.7]$

この場合、予想されるラベル =  $\arg \max h(X) = 4$  となる。

【0046】

本発明の好ましい実施形態において、各クラスの信頼性は、ベイジアンネットワークを使った確率論的融合と同種の、確率に基づく融合を通じて得られ、すなわち、

$h(X) = W_g \times h_g(X) + W_v \times h_v(X) \quad (6)$

となる。

【0047】

重みまたは信頼度は、対応する分類モード(トレース、視覚、意味論)の分類表に基づいて計算することができる。分類表は当業界においてよく知られており、これは、各クラスの(各行の)サンプルが、考えられるクラスの(各列の)各々にどのように分類されるかを表している。したがって、良い分類器が有する分類表では、ゼロ以外の数値が表の対角線に沿って集中しているはずである。図5に示される、あるモードの8クラス分類問題の分類表の例を用いると、このモードの第一のクラスの重みは、次のように計算される。

$W_g(1) = 5 / (5 + 2 + 4 + 0 + 1 + 2 + 0 + 1) = 1 / 3 \quad (7)$

式中、分母は第一列の総和である。簡単に言えば、このモードのクラス1に分類された15のサンプルのうち、実際にクラス1に属するのは5つであるため、信頼度は  $1 / 3$  となる。

【0048】

本発明は、特定の好ましい実施形態を特に参照して詳細に説明したが、本発明の精神と範囲の中で変形版や改変形を実現できることが理解されるであろう。

【符号の説明】

【0049】

10

20

30

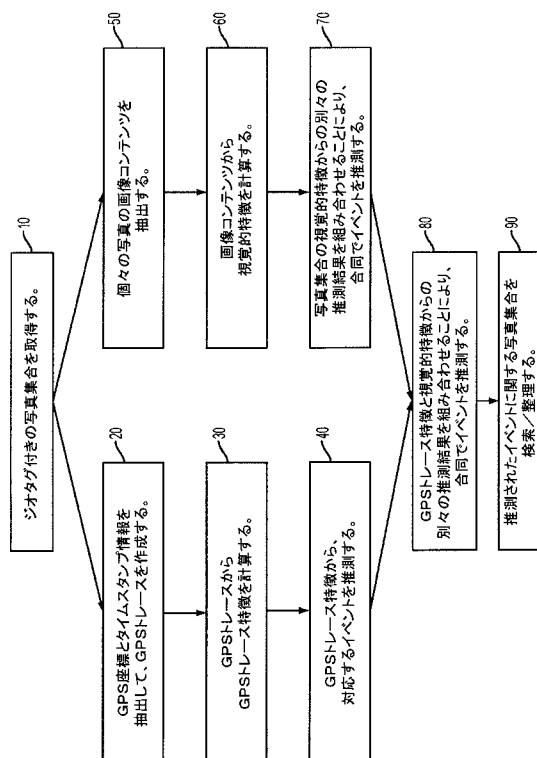
40

50

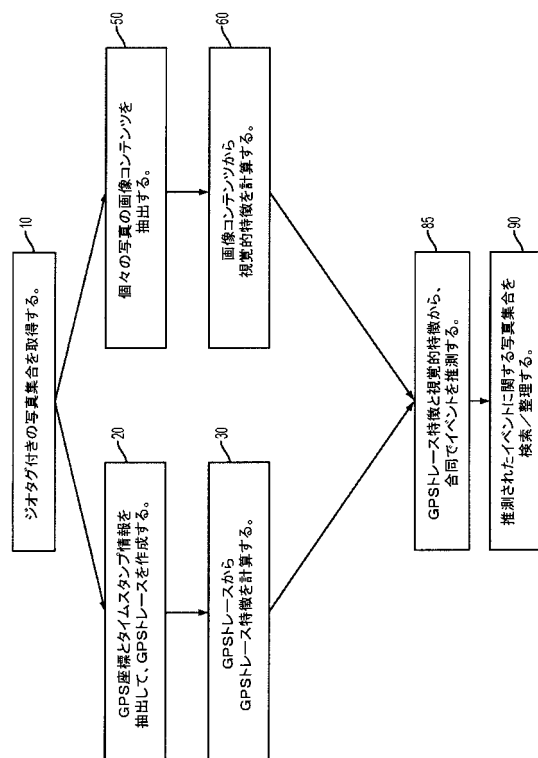
- 10 ステップ：ジオタグ付きの写真集合を取得する。
- 20 ステップ：ジオタグ付きの写真からのGPS座標とそれらのタイムスタンプ情報を抽出して、トレースを作成する。
- 30 ステップ：トレースからトレース特徴を計算する。
- 40 ステップ：トレース特徴から、対応するイベントを推測する。
- 50 ステップ：個々の写真の画像コンテンツを抽出する。
- 60 ステップ：画像コンテンツから視覚的特徴を計算する。
- 70 ステップ：写真集合の視覚的特徴から、対応するイベントを推測する。
- 80 ステップ：トレース特徴と視覚的特徴から別々に推測された結果を組み合わせることにより、合同でイベントを推測する。
- 85 ステップ：トレース特徴と視覚的特徴から、合同でイベントを推測する。
- 90 ステップ：推測されたイベントに関する写真集合を検索／整理する。

10

【図1A】



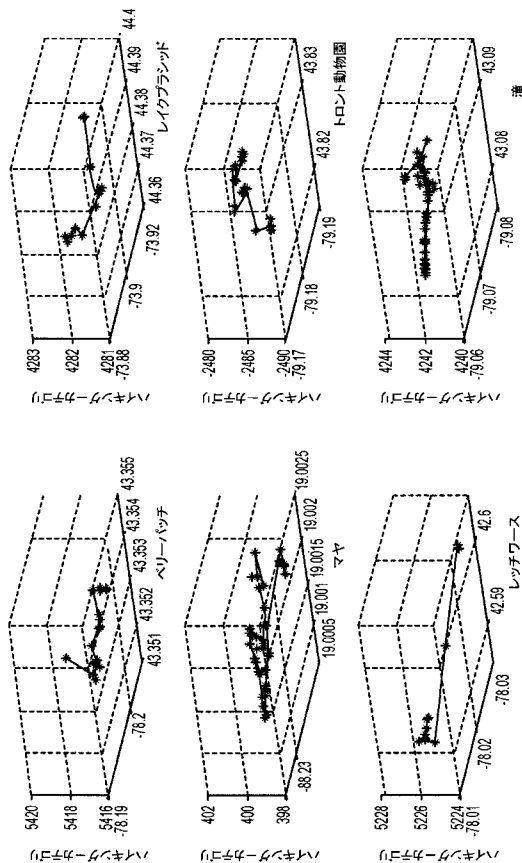
【図1B】



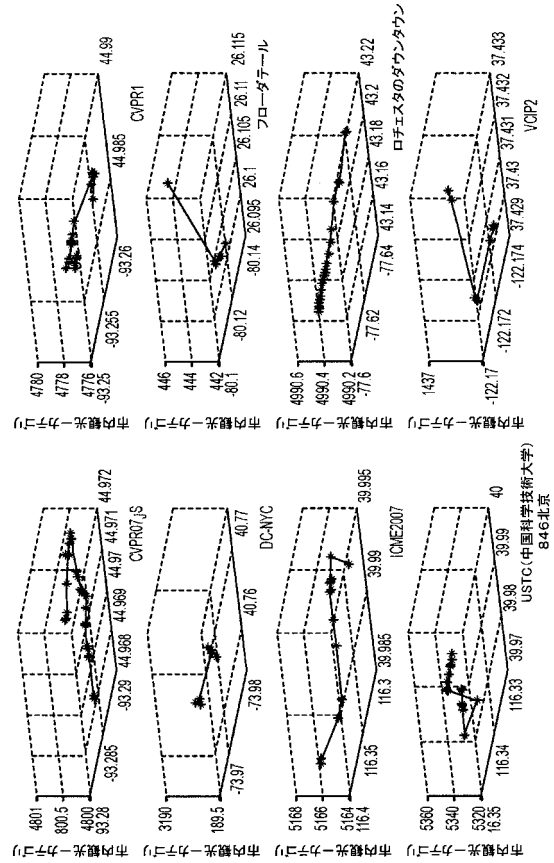
【図2】

空間範囲が大きいイベント	空間範囲が小さいイベント
市内観光、ハイキング、ドライブ旅行	裏庭、浜辺、野球、ピクニック、結婚式
移動速度が速いイベント	移動速度が低から中程度のイベント
ドライブ旅行	市内観光、ハイキング、裏庭、浜辺、野球、ピクニック、結婚式

【図4】



【図3】



【図5】

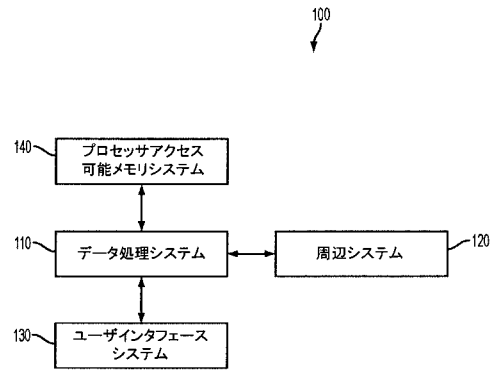
GPS特徴の分類表

1	0	2	0	0	1	0	0
2	2	0	1	0	1	0	1
4	1	0	0	0	0	0	0
0	0	0	7	1	0	0	0
1	0	0	1	5	0	0	0
2	2	1	0	0	0	0	0
0	0	0	0	0	0	8	0
1	2	0	0	0	0	0	0

$$W_d(1) = 5(5+2+4+0+2+0+1) = 13$$

USTC(中国科学技術大学)  
846北京

【図 6】



---

フロントページの続き

(72)発明者 ルオ ジエボー

アメリカ合衆国 ニューヨーク ロチェスター ステート ストリート 343

(72)発明者 ユエン ジュインソン

アメリカ合衆国 イリノイ エヴァンズトン リッジ アヴェニュー 2215 アpartment  
3ビー

審査官 早川 学

(56)参考文献 特開2005-174060(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/30