

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第4142420号
(P4142420)

(45) 発行日 平成20年9月3日 (2008.9.3)

(24) 登録日 平成20年6月20日 (2008.6.20)

(51) Int.Cl.

F I

G O 6 T 7 / 0 0 (2 0 0 6 . 0 1)
G O 6 T 1 / 0 0 (2 0 0 6 . 0 1)
G O 6 T 1 / 2 0 (2 0 0 6 . 0 1)
G O 6 T 7 / 2 0 (2 0 0 6 . 0 1)

G O 6 T 7 / 0 0 3 0 0 B
G O 6 T 7 / 0 0 1 0 0 B
G O 6 T 7 / 0 0 2 5 0
G O 6 T 7 / 0 0 3 5 0 A
G O 6 T 1 / 0 0 3 4 0 A

請求項の数 2 (全 42 頁) 最終頁に続く

(21) 出願番号	特願2002-351817 (P2002-351817)	(73) 特許権者	500046438
(22) 出願日	平成14年12月3日 (2002.12.3)		マイクロソフト コーポレーション
(65) 公開番号	特開2003-216951 (P2003-216951A)		アメリカ合衆国 ワシントン州 9805
(43) 公開日	平成15年7月31日 (2003.7.31)		2-6399 レッドモンド ワン マイ
審査請求日	平成17年11月4日 (2005.11.4)		クロソフト ウェイ
(31) 優先権主張番号	10/006, 927	(74) 代理人	100077481
(32) 優先日	平成13年12月3日 (2001.12.3)		弁理士 谷 義一
(33) 優先権主張国	米国 (US)	(74) 代理人	100088915
			弁理士 阿部 和夫
		(72) 発明者	ルイ ヨン
			アメリカ合衆国 98074 ワシントン
			州 サマリッシュ 239 プレイス ノ
			ースイースト 2508

最終頁に続く

(54) 【発明の名称】 複数のキューを使用する複数の個人の自動検出および追跡のためのコンピュータ可読な記憶媒体

(57) 【特許請求の範囲】

【請求項 1】

内容のフレーム内で追跡されない顔に関する候補領域を検出する複数の命令を記憶した1つまたは複数のコンピュータ可読な記憶媒体であって、前記複数の命令が、1つまたは複数のプロセッサによって実行される時に、前記1つまたは複数のプロセッサに、

前記フレーム内に動きがあるかどうかを検出させ、

前記フレーム内に動きがある場合に、前記候補領域を識別するために動きベース初期化を実行させ、

前記フレーム内にオーディオがあるかどうかを検出させ、

前記フレーム内にオーディオがある場合に、前記候補領域を識別するためにオーディオベース初期化を実行させ、

前記フレーム内に動きがなく、かつ、前記フレーム内にオーディオがない場合に、前記候補領域を識別するために高速顔検出器を使用させ、

前記動きベース初期化を前記1つまたは複数のプロセッサに実行させる前記複数の命令は、前記1つまたは複数のプロセッサに、

前記フレームを横切る複数の線の複数の画素で動きがあるかどうかを判定させ、

前記複数の線のうちの複数の線の複数の線分についてフレーム差分の合計を生成させ、

前記複数の線のそれぞれについて、最大の合計を有する前記線分を選択させ、

前記選択された線分の最も滑らかな領域を識別させ、

前記最も滑らかな領域が、人間の上部に似ているかどうかを検査させ、

前記候補領域として、前記最も滑らかな領域のうちで人間の頭に似ている部分を抽出させる

命令を含むことを特徴とする1つまたは複数のコンピュータ可読な記憶媒体。

【請求項2】

前記動きがあるかどうかを前記1つまたは複数のプロセッサに判定させる前記命令は、
前記1つまたは複数のプロセッサに、

前記複数の画素のそれぞれについて、前記フレーム内の前記画素の輝度値と、1つまたは複数の他のフレーム内の対応する画素の輝度値との間の差が閾値を超えるかどうかを判定させる

命令を含むことを特徴とする請求項1に記載の1つまたは複数のコンピュータ可読な記憶媒体。

10

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、画像処理および/またはオーディオ処理、および/またはコンピュータビジョンに関し、詳細には、複数の個人の自動検出および追跡のためのコンピュータ可読な記憶媒体に関する。

【0002】

【従来の技術】

ビデオデータを分析するシステムが、ますます普及しつつある。ビデオ会議システムがそのようなシステムの例であり、これを用いると、会議参加者が地理的に異なる位置にいる可能性があるという事実にもかかわらず、視覚的に対話して会議を行うことができる。ビデオ会議の視覚面は、通常は、電話会議より魅力的であり、それと同時に、1人または複数の参加者が会議位置へ移動しなければならない生の会議に対して低コストの代替となる(かつ、通常は、より短時日の通知で行うことができる)。

20

【0003】

いくつかの現在のビデオ会議システムは、カメラを移動するために、自動化されたオーディオベースの検出技法および/またはプリセットを使用する(例えば、カメラのパンまたはチルト)。しかし、現在のビデオ会議システムには、多数の問題が存在する。そのような問題の1つが、オーディオベースの話者検出技法の精度が低くなる可能性があることである。さらに、ビデオ会議システムは、通常は、会議に何人の参加者がいるか(参加者が会議に加わるか去る時を含めて)、参加者がいる場所(座っているのか立っているのか)、または参加者が現在話しているかどうかを知らない。いくつかのシステムは、参加者情報(例えば、参加者の数および位置)を手動でプログラムすることができるが、これは、プログラムされる情報のユーザ入力を必要とし、これによって、参加者が室内で移動する能力ならびに参加者が会議に参加する能力が制限される傾向がある。

30

【0004】

本明細書に記載の複数の個人の自動的検出および追跡が、上記および他の問題の解決に役立つ。

【0005】

40

【非特許文献1】

P. Viola and M.J. Jones, "Robust real-time object detection", Technical Report Series, Compaq Cambridge Research Laboratory, CXRL 2001/01, Feb. 2001

【0006】

【非特許文献2】

S.Z. Li, Q.D. Fu, L. Gu, B. Scholkopf, Y.M. Cheng, H.J. Zhang., "Kernel Machine Based learning for Multi-View Face Detection and Pose Estimation," Proceedings of 8th IEEE International Conference on Computer Vision, Vancouver, Canada, July 9-12, 2001

【0007】

50

【非特許文献3】

J. Shi and C. Tomasi, "Good Features to Track," IEEE Conf. on Computer Vision and Pattern Recognition, pp. 593-600, 1994.

【0008】

【発明が解決しようとする課題】

本発明の目的は、複数の個人の自動的な検出および追跡のためのコンピュータ可読な記憶媒体を提供することである。

【0009】

【課題を解決するための手段】

一態様によれば、内容のフレーム(例えばオーディオおよび/またはビデオ)が、受け取られ、フレーム内の新しい顔領域の1つまたは複数の候補区域が識別される。その後、階層検証を使用して、人間の顔が候補区域にあるかどうかを検証し、階層検証によって人間の顔が候補区域にあることが検証される場合に、候補区域に顔が含まれることの指示を行う。区域の検証の後に、複数のキュー(cue)を使用して、フレームからフレームへと、内容内の検証された顔のそれぞれを追跡する。

【0010】

一態様によれば、この検出追跡フレームワークに、自動初期化モジュール、階層検証モジュール、および複数キュー追跡モジュールという3つのメインモジュールがある。内容のフレーム(例えばオーディオおよび/またはビデオ)が受け取られ、フレーム内の新しい顔(または他のオブジェクト)領域に関する1つまたは複数の候補区域が、自動初期化モジュールによって識別される。階層検証モジュールは、人間の顔が候補区域にあるかどうかを検証するのに使用され、階層検証モジュールが、人間の顔が候補区域にあると検証する場合に、候補区域に顔が含まれることの指示が行われる。区域を検証した後に、複数キュー追跡モジュールが、複数のキューを使用して、フレームからフレームへと、内容内の検証された顔のそれぞれを追跡する。追跡処理全体の中に、追跡される顔が、階層検証モジュールによって継続的に検証される。信頼性レベルが高い場合に、複数キュー追跡モジュールが、その顔を追跡し、信頼性レベルが低くなる場合には、その特定の顔の追跡を打ち切る。追跡モジュールおよび検証モジュールは、初期化モジュールがさらなる候補を供給するのを待つ。

【0011】

同一の符号が、この文書全体を通じて、類似する構成要素および/または特徴を参照するのに使用される。

【0012】

【発明の実施の形態】

複数の個人の自動検出および追跡を、本明細書で説明する。ビデオ内容および/またはオーディオ内容を分析して、内容のフレーム内で個人を自動的に検出する。検出された後に、これらの個人が、連続するフレームで自動的に追跡される。個人の追跡が失われた場合には、その個人が、もう一度自動的に検出され、個人の追跡が再開される。

【0013】

図1および2に、堅牢な自動化された検出および追跡を使用することができる例示的環境を示す。図1では、複数(n)個のビデオ会議システム102が、1つまたは複数の他のシステムそれぞれにオーディオ/ビデオ内容を通信することができ、会議参加者が、システム102のそれぞれに位置して、互いを見、聞くことができる。普通のパン/チルト/ズームカメラ、360°パノラマカメラ(例えば、機械的ではなくデジタルにパン/チルト/ズームすることができるもの)など、さまざまな異なるカメラシステムを、ビデオ会議システム102と共に使用することができる。そのような360°パノラマカメラシステムは、放物面鏡装置を指すカメラを使用し、そして、さまざまな較正技法を使用して、画像のひずみを除去して普通の画像にし、この画像から、カメラの周囲の360°全方向画像を作成することができる。そのような360°パノラマカメラシステムの例は、発明者Yong Rui、Anoop Gupta、Johnathan J. Cadiz、Ross G. Cutlerによる2001年6月14日に出願された「Automated Online Broad

10

20

30

40

50

dcasting System and Method Using an Omni-Directional Camera System for Viewing Meetings Over a Computer Network」という名称の同時係属の米国特許出願第09/681,843号に見られる。もう1つのそのような360°パノラマカメラシステムでは、複数のカメラ(それぞれが360°未満の視野を有する)を配置し、その結果、それらのカメラが一緒になって、約360°の視野を提供するようにする。

【0014】

ビデオ会議システム102のそれぞれに、対応するシステム102の複数の個人を堅牢に自動的に検出し、追跡する、追跡モジュール104が含まれる。この検出および追跡は、カメラのチルト/パン/ズーム、個人の強調(例えば、個人を指す矢印または個人を囲む円)など、さまざまな目的に使用することができる。

10

【0015】

ビデオ会議システム102を、さまざまな形のいずれかで一緒に結合することができる。例えば、1つまたは複数の電話回線(ISDNなどのデジタル回線を含む)を使用して、直接に、または中央装置もしくは拠点を通じてのいずれかで、複数のビデオ会議システム102と一緒に結合することができ、普通のデータネットワーク(例えば、インターネット、イントラネットなど)を使用して、複数のシステム102と一緒に結合することができる。

【0016】

図2では、追跡モジュール114を含むシステム112が、内容116を受け取る。内容116は、通常は、オーディオ/ビデオ内容であるが、代替として、他のタイプの内容(例えば、共有されるホワイトボードなど)を含めることができ、かつ/または、オーディオ内容またはビデオ内容を含めないものとしてすることができる。追跡モジュール114は、内容116を分析し、内容116内の個人の画像および/またはオーディオに基づいて複数の個人を堅牢に自動的に検出し、追跡する。内容116は、システム112のカメラおよびマイクロホン、内容が記録される記録媒体(例えば磁気テープ、光ディスクなど)、電話回線、またはネットワーク入力など、さまざまな形のいずれかでシステム112から使用可能にすることができる。

20

【0017】

図3に、堅牢な自動化された検出および追跡を使用する例示的システム130を示す。システム130は、例えば、図1のビデオ会議システム102または図2のシステム112のいずれかとしてすることができる。システム130には、検出および追跡モジュール132、通信モジュール134、オーディオ取込モジュール136、およびビデオ取込モジュール138が含まれる。ホワイトボード取込モジュールなど、さまざまな他のモジュール(図示せず)も、含めることができる。通信モジュール134は、図1のビデオ会議システム102または、分析される内容をそこから受け取ることができる他の装置など、システム130に関する他のシステムとの通信を管理する。通信モジュール134は、さまざまな普通のおよび/または独自のプロトコルをサポートすることができる。

30

【0018】

オーディオ取込モジュール136は、システム130の一部である1つまたは複数のマイクロホン(図示せず)を介するなど、システム130でのオーディオ内容の取込を管理する。さらなる処理(例えば、ビームフォーミング技法の使用)を行って、オーディオ品質を高めることもできる。オーディオ内容は、デジタルフォーマットに変換され(必要な場合に)、追跡のために検出および追跡モジュール132から使用可能にされる。ビデオ取込モジュール138は、システム130の一部である(例えば、固定されたカメラ、普通のパン/チルト/ズームカメラ、360°パノラマカメラを含めることができる)1つまたは複数のビデオ取込装置(例えば、アナログまたはデジタルのビデオカメラ(図示せず))などの、システム130でのビデオ内容の取込を管理する。ビデオ内容の取り込まれたフレームが、デジタルフォーマットに変換され(必要な場合に)、個人の検出および追跡のために検出および追跡モジュール132から使用可能にされる。オーディオ内容およびビデオ内容は、互いに相関され(例えば取込の時に)、したがって、内容の特定の部分(例えばフレーム)のどれについても、ビデオ内容とオーディオ内容の両方が既知になる。代替実施形態では、1つまたは複数のモジュール134、136、および138を含めないものとしてすることができる。例えば、システムを、

40

50

ビデオ取込モジュール138またはオーディオ取込モジュール136のいずれかを含まないものとしてすることができる。

【0019】

検出および追跡モジュール132に、自動初期化モジュール140、階層検証モジュール142、複数キュー追跡モジュール144、および顔/候補追跡リスト146が含まれる。検出および追跡モジュール132は、ビデオ内容のうちで、人間の顔を含むか潜在的に含む領域を自動的に検出し、さまざまなキューを使用して、検出された領域を追跡する。これらの領域を、本明細書ではオブジェクトとも称する。検出および追跡モジュール132は、顔または顔候補を含む複数の領域を検出し、これらの複数の領域を同時に追跡することができる。

【0020】

検出および追跡モジュール132は、フレームなどの、内容の部分を分析する。例えば、ビデオ内容は、通常は、毎秒複数のフレーム(例えば静止画像)として取り込まれる(通常は、毎秒15-60フレームであるが、他の速度を使用することができる)。これらのビデオフレームならびに対応するオーディオ内容(例えば、1秒の1/15から1/60のオーディオデータ)が、モジュール132による検出および追跡のフレームとして使用される。オーディオを記録する時に、オーディオは、通常は、ビデオより高い速度でサンプリングされる(例えば、ビデオについて毎秒15から60枚の画像を取り込むことができるのに対して、数千個のオーディオサンプルを取り込むことができる)。オーディオサンプルは、さまざまな異なる形で特定のビデオフレームに対応するものとしてすることができる。例えば、あるビデオフレームが取り込まれた時から次のビデオフレームが取り込まれる時までの範囲のオーディオサンプルを、そのビデオフレームに対応するオーディオフレームとすることができる。もう1つの例として、ビデオ取込フレームの時間に関してセンタリングされたオーディオサンプルを、そのビデオフレームに対応するオーディオフレームとすることができる(例えば、ビデオが、毎秒30フレームで取り込まれる場合に、オーディオフレームを、ビデオフレームが取り込まれる時の1/60秒前から、ビデオフレームが取り込まれた時の1/60秒後までの範囲とすることができる)。

【0021】

さらに、いくつかの状況で、ビデオ内容がない場合がある。これらの状況では、オーディオ内容のフレームを、サンプリングされたオーディオからさまざまな形で生成することができる。例えば、1秒の1/30ごとまたは1/60ごとのオーディオサンプルによって、オーディオ内容のフレームを構成することができる。

【0022】

いくつかの状況で、オーディオ内容に、ビデオ内容に直接に対応しないデータが含まれる場合がある。例えば、オーディオ内容が、ビデオ内容の中の人の音声ではなく、音楽のサウンドトラックである場合がある。これらの状況では、本明細書で説明する検出および追跡は、オーディオ内容なしでビデオ内容に頼ることができる。

【0023】

本明細書では、主に、ビデオ内容およびオーディオ内容の使用に関して説明するが、検出および追跡モジュール132は、その代わりに、ビデオ内容だけまたはオーディオ内容だけに基いて動作することができる。オーディオ内容がない状況では、下でオーディオ内容の処理に関して説明する処理が、実行されない。同様に、ビデオ内容がない状況では、下でビデオ内容の処理に関して説明する処理が、実行されない。

【0024】

顔/候補追跡リスト146は、人間の顔を含むか潜在的に含む検出された領域のそれぞれについて情報を維持する。顔を潜在的に含むが顔の存在が検証されていない領域を、候補領域と称する。示される例では、各領域が、中心座標148、境界ボックス150、追跡時間152、および最後の検証からの時間154によって記述される。顔または顔候補を含むビデオ内容の領域は、中心座標および境界ボックスによって定義される。中心座標148は、領域のほぼ中央を表し、境界ボックス150は、中心座標のまわりの長方形領域を表す。この長方形領域は、顔または顔候補を含む領域であり、検出および追跡モジュール132によって追跡

10

20

30

40

50

される。追跡時間152は、領域内の顔または顔候補が、どれほど長く追跡されたかを表し、最後の検証からの時間154は、領域内の顔または顔候補がどれほど前に検証されたか(下で詳細に説明するように、階層検証モジュール142によって)を表す。

【0025】

リスト146に示されている、各領域を記述する情報は、例示的なものに過ぎず、さまざまな他の情報を、その代わりに使用することができる。例えば、中心座標148を、含まないようにすることができる。もう1つの例として、長方形以外の、円、楕円、三角形、五角形、六角形、または自由形状などの領域形状を使用することができる。

【0026】

追跡リスト146には、顔および顔候補の両方が記録され、この2つは、さまざまな形で互いから区別することができる。例えば、2つのサブリスト(一方は顔を識別し、他方は顔候補を識別する)を維持することができ、あるいは、追加のフィールドを追加して、各フィールドに顔または顔候補のいずれかとしてラベルを付けることができ、あるいは、最後の検証からの時間154に固有とすることができる(例えば、この値が空白である場合に、それが、この領域がまだ顔を含むものとして検証されたことがなく、したがって、顔候補であることを意味する)。その代わりに、単一のリスト146ではなく、複数のリストを含めることができる(例えば、顔用の1つのリストと、顔候補用のもう1つのリスト)。

10

【0027】

動作中に、検出および追跡モジュール132が、フレームごとに内容を分析する。各フレームについて、モジュール132は、自動初期化モジュール140をアクティブにし、自動初期化モジュール140は、新しい顔領域の候補を検出する。そのような候補のそれぞれは、新しい顔(すなわち、現在は追跡されていない顔)を潜在的に含む、ビデオ内容の領域である。検出された後に、候補領域が、階層検証モジュール142に渡され、階層検証モジュール142は、候補領域に実際に顔が含まれるかどうかを検証する。階層検証モジュール142は、候補ごとに信頼性レベルを生成し、信頼性レベルが閾値を超える場合に、顔領域として候補を保存することを決定し、追跡リスト146に領域の記述を追加する。信頼性レベルが閾値を超えない場合には、階層検証モジュール142は、その候補を破棄する。

20

【0028】

複数キュー追跡モジュール144は、追跡リスト146内で識別される領域のそれぞれを追跡する。追跡モジュール144は、さまざまな視覚的キューを使用して、内容のフレームからフレームへと領域を追跡する。追跡される領域内の顔のそれぞれは、人の少なくとも一部の画像である。通常、人は、内容が生成されつつある間に、立ち上がる、座る、歩く、椅子に腰掛けて動くなど、動くことができる。内容の各フレームで顔検出を実行するのではなく、モジュール132は、顔(検出された後の)を含む領域をフレームからフレームへと追跡するが、これは、通常は計算的に顔検出より安価である。

30

【0029】

追跡されることに加えて、追跡リスト146からの顔を含む各領域は、階層検証モジュール142によって繰り返して再検証される。複数キュー追跡モジュール144または、代替として階層検証モジュール142が、領域をモジュール142によって再検証する必要があるかどうかを判定することができる。領域を、規則的な間隔または不規則な間隔で再検証することができる。領域を再検証する時に、階層検証モジュール142は、領域の新しい信頼性レベルを生成し、信頼性レベルを閾値と比較する。新しい信頼性レベルが閾値を超える場合に、その領域の最後の検証からの時間154が、リセットされ、領域が、追跡リスト146内に残される。しかし、新しい信頼性レベルが閾値を超えない場合には、領域が、追跡リスト146から削除される。

40

【0030】

複数キュー追跡モジュール144が、その追跡を失う状況が生じる可能性があることに留意されたい。階層検証モジュール142は、顔を含む領域の追跡が失われた(例えばその領域の信頼性レベルが低い)時を識別することによって、この状況を解決する。これによって、自動初期化モジュール140が、領域を再検出でき、再検出した領域の追跡を進行させられ

50

るようになる。

【 0 0 3 1 】

自動初期化

自動初期化モジュール140は、1つまたは複数の技法を使用して、新しい顔領域の候補を検出する。この技法には、動きベースの初期化、オーディオベースの音源位置、高速顔検出が含まれる。動きベース初期化モジュール156は、フレーム間差分(ビデオ内容の複数のフレームの間の相違)を使用して動きを検出し、動きが検出された領域に顔が含まれるかどうかを判定する。オーディオベース初期化モジュール158は、ビデオ内容に対応するオーディオ内容を分析し、音が受け取られた方向を検出し、その方向のビデオ内容の領域を検索して、音が受け取られた方向の領域に顔が含まれるかどうかを判定する。モジュール156および158の両方が、ビデオ内容の各フレームを分析するように動作する。その代わりに、モジュール156および158の一方が、顔を検出できない場合に限って、モジュール156および158の他方が、ビデオ内容の特定のフレームを操作することができる。

10

【 0 0 3 2 】

高速顔検出モジュール160は、ビデオ内容のフレームに動きまたは音声がない時に動作する。その代わりに、モジュール160が、フレーム内に動きおよび/または音声があるが、モジュール156およびモジュール158の両方が、顔を検出しない時(または、その代わりに、モジュール156または158が顔を検出するかどうかに関係なく)動作することができる。高速顔検出モジュール160は、高速顔検出器(fast face detector)を使用して、ビデオ内容のフレームを分析し、フレーム内の顔を検出する。信頼性消失領域検出モジュール162は、領域の再検証が、顔を含む領域の信頼性の消失をもたらしたことが自動初期化モジュール140に通知される時に動作する。領域が顔を含むことの信頼性が消失した場合であっても、顔がその領域の近くにある可能性が高い。信頼性消失領域検出モジュール162は、モジュール156、158、および160のそれぞれと通信して、モジュール156、158、および160に、この領域の周囲のビデオ内容の区域を分析させて、領域内の顔の検出を試みる。領域の周囲の区域の正確なサイズは、実装によって変化する可能性がある(例えば、一実施形態では、区域を、領域の上下に領域の高さの半分だけ延ばし、領域の左右に領域の幅の半分だけ延ばすことができる)。

20

【 0 0 3 3 】

図4は、新しい顔領域の候補を検出する例示的処理200を示す流れ図である。図4の処理は、図3の自動初期化モジュール140によって実行され、ソフトウェアで実行することができる。

30

【 0 0 3 4 】

最初に、オーディオ/ビデオ内容のフレームを受け取る(202)。内容のこのフレームは、さまざまな供給源のいずれかから受け取ることができる。例えば、内容のフレームを、図3のシステム130の1つまたは複数の取込装置によって取り込むことができ、あるいは、内容を、他所で取り込み、システム130に通信することができる(例えば、取外し可能記憶装置を介して、ネットワーク接続または電話回線接続を介してなど)。受け取られた後に、フレーム内の動きを検出する試みが、フレームの画素を、オーディオ/ビデオ内容の前のフレームの対応する画素と比較することによって行われる(動作204)。動きが検出される場合には、動きベース初期化を実行して、フレーム内の新しい顔領域の候補を識別する(動作206)。動きベース初期化を使用する新しい顔領域の候補のすべてが、動作206で識別された後に、フレーム内のオーディオを検出する試みを行う(動作208)。オーディオが検出される場合に、オーディオベース初期化を実行して、フレーム内の新しい顔領域の候補を識別する(動作210)。動きベース初期化および/またはオーディオベース初期化に基づく新しい顔領域の識別された候補のすべてを、顔検証のために階層検証モジュール142に渡す(動作212)。

40

【 0 0 3 5 】

動作204に戻って、フレーム内で動きが検出されない場合に、フレーム内でオーディオを検出する試みを行う(動作214)。オーディオが検出される場合には、オーディオベース初

50

期化を実行して、フレーム内の新しい顔領域の候補を識別し(動作210)、処理は動作212に進む。しかし、オーディオが検出されない場合には、高速顔検出器を使用して、新しい顔領域の候補を識別する(動作216)。高速顔検出に基づく新しい顔領域の識別された候補のすべてを、顔検証のために階層検証モジュール142に渡す(動作212)。

【0036】

フレームのうちで、動きまたはオーディオを検出する試みが行われる区域、または高速顔検出器が使用される区域は、状況に基づいて変更することができる。追跡リスト146に、顔または顔候補が含まれない状況では、フレームの区域は、フレーム全体である。追跡リスト146に1つまたは複数の顔または顔候補が含まれる状況では、フレームの区域に、現在追跡されていない(すなわち、追跡リスト146にリストされていない)すべての区域が含まれる。信頼性消失領域検出モジュール162が、特定の区域を分析することを要求する状況では、フレームの区域は、モジュール162によって識別される区域である。

10

【0037】

図3に戻って、動きベース初期化モジュール156は、フレーム内の画素を、前のフレームおよび/または後続フレーム内の対応する画素と比較することによって、ビデオ内容のフレームを分析し、フレームの間および/または中の各画素で動きがあるかどうかを検出する。移動する個人は、ビデオ内容の前景にあると思われ、モジュール156は、この移動する前景の形状を識別することを試みる。形状が、人の上体のシルエット(大きい肩の上の小さい頭)に似ている場合に、その形状が、顔候補として判定される。

20

【0038】

図5は、動きベース初期化を使用して新しい顔領域の候補を識別する例示的処理240を示す流れ図である。図5の処理は、図3の動きベース初期化モジュール156によって実行され、ソフトウェアで実行することができる。

【0039】

最初に、各画素で動きがあるかどうかの判定を行う(動作242)。この判定は、画素を、前のフレームの対応する画素と比較することによって、フレームの画素ごとに行われる。比較は、例えば、画素の輝度(例えばグレイレベル)または色値によって行うことができる。さまざまな従来のフィルタを、比較の前に画素に適用することもできる。ビデオ内容は、画素の普通の2次元(x,y)座標系を使用して見ることができる。あるフレームの特定の座標位置の画素は、同一の座標位置にあるもう1つのフレームの画素に対応する。分析されるフレームの区域内の各画素は、次式によって生成されるフレーム差分を有する。

30

【0040】

【数1】

$$D_t(x,y) = \begin{cases} 1, & I_t(x,y) - I_{t-1}(x,y) > d_{th} \\ 0, & \text{それ以外} \end{cases}$$

【0041】

ここで、 $D_t(x,y)$ は、フレームtの画像内の位置(x,y)の画素と、フレームt-1の画像内の位置(x,y)の画素の間のフレーム差分であり、 $I_t(x,y)$ は、フレームtの画像の位置(x,y)の画素であり、 $I_{t-1}(x,y)$ は、フレームt-1の画像の位置(x,y)の画素であり、 d_{th} は、画素が動き画素であるかどうかを決定する閾値である。 d_{th} の正確な値は、フレームがカラーまたはグレイスケールのどちらであるか、どのようなフィルタリングが行われたか(使用される場合)など、実施形態によって変更することができる。1特定の例では、画素が256レベルグレイスケールの場合に、20という値を、 d_{th} に使用することができる。

40

【0042】

代替案では、フレーム差分を、2つのフレームだけではなく、3つまたはそれ以上のフレームに基づいて生成することができる。一実施形態では、3つのフレーム(例えば、 I_{t-1} 、 I_t 、および I_{t+1})を使用して、動いている画素を検出する。 $I_t(x,y) - I_{t-1}(x,y)$ と $I_{t+1}(x,y) - I_t(x,y)$ の両方で大きいフレーム差分(例えば、 d_{th} より大きい)を有する画素だけが、動い

50

ている画素である。

【 0 0 4 3 】

フレーム差分を与えられて、分析されるフレーム区域内の画像の各水平線の各可能な線分のフレーム差分の合計を生成する(動作244)。分析されるフレーム区域内の画像に、複数の水平線が含まれる。画素の水平の行のそれぞれを、そのような線とすることができ、あるいは、画素の水平の行のn番目ごと(例えば、2番目、3番目、4番目、5番目など)を、そのような線とすることができる。各線の、線上の異なる起点および終点を有する多数の線分が存在する。可能な線分に沿ったフレーム差分の合計は、分析される区域内の最も可能性の高い前景線分を識別する試みに使用される。これを、さらに詳細に図6に示す。

【 0 0 4 4 】

10

図6に、ビデオ内容のフレームの例示的画像を示す。画像270は、顔または顔候補を含むものとして既に追跡されている2つの領域272および274と、新しい顔領域の候補について分析される残りの区域276を含むものとして図示されている。画像に個人278が含まれると仮定すると、2つの水平線280および282が、起点*i*および終点*j*で画像278と交わる。特定の線*li*上の点*i*と*j*の間のすべての画素が、前景でなければならず、2つの連続する水平線の間の境界が、滑らかさの制約も有しなければならない(類似する中心と類似する幅を有する傾向がある)。フレーム差分の合計は、起点*i*および終点*j*を有する水平線*li*の部分の識別するのに使用される。

【 0 0 4 5 】

各水平線について、水平線上の各可能な線分のフレーム差分の合計*S*は、次式によって生成される。

20

【 0 0 4 6 】

【数2】

$$S(i, j) = \sum_{x=i}^j D(x, y) \quad 0 < i < j < N, y \in [0, M]$$

【 0 0 4 7 】

ここで、*i*は線分の起点、*j*は線分の終点、*D*(*x*, *y*)は、線分に沿った位置*x*, *y*でのフレーム差分、*N*は、水平線の長さ、*M*は、水平線の数である。

30

【 0 0 4 8 】

すべての可能な*i*および*j*の合計を計算できる速度を高めるために、下記の処理を使用する。まず、0と*N*の間のすべての値*i*について、一切を含めて、下記を生成する。

$S(i, i) = D(i, y), i \in [0, N]$

次に、*k*=1から*k*=*N*まで、次式を計算する。

$S(i, i+k) = S(i, i+k-1) + S(i+k, i+k), i \in [0, N-k]$

【 0 0 4 9 】

図5に戻って、水平線上の各可能な線分のフレーム差分の合計を生成した後、水平線ごとに、最大の合計を有する線分を、その線上で最も可能性が高い前景線分として選択する(動作246)。最大の合計を有する線分が、実際に新しい顔領域の候補の一部であるかどうか、下で述べるように滑らかさの制約に依存する。最も可能性が高い線分のうちで最も滑らかな領域を判定する(動作248)。最も滑らかな領域は、すべての水平線にまたがって滑らかさの制約を検討することによって生成される。これは、下記のように達成される。処理は、*y*=0(一番上の水平線)かつ $E^0(i^{(0)}, j^{(0)}) = S(i^{(0)}, j^{(0)})$ から開始され、下記の再帰関数に従うことによって、*y*=*M*(一番下の水平線)まで伝搬する。

40

【 0 0 5 0 】

【数3】

$$E^0(i^{(y)}, j^{(y)}) = S(i^{(y)}, j^{(y)}) + \max_{i^{(y-1)}, j^{(y-1)} \in [0, N]} \left\{ E^0(i^{(y-1)}, j^{(y-1)}) + C \left(\begin{bmatrix} i^{(y-1)} \\ j^{(y-1)} \end{bmatrix}, \begin{bmatrix} i^{(y)} \\ j^{(y)} \end{bmatrix} \right) \right\}$$

【 0 0 5 1 】

ここで、 $i^{(y)}$ および $j^{(y)}$ は、 (y) 番目の水平線上の境界であり、 N は、画像の幅である。 $C(\dots)$ パラメータは、滑らかさエネルギー項である。 $C(\dots)$ パラメータは、連続する線の間の滑らかでない境界に大きいペナルティを与え、次式によって定義される。

【 0 0 5 2 】

【 数 4 】

10

$$C \left(\begin{bmatrix} i^{(y-1)} \\ j^{(y-1)} \end{bmatrix}, \begin{bmatrix} i^{(y)} \\ j^{(y)} \end{bmatrix} \right) = c_c \cdot \left| \frac{i^{(y)} + j^{(y)}}{2} - \frac{i^{(y-1)} + j^{(y-1)}}{2} \right| + c_w \cdot |(j^{(y)} - i^{(y)}) - (j^{(y-1)} - i^{(y-1)})|$$

【 0 0 5 3 】

ここで、 c_c は、線分中央の非滑らかさに関するペナルティ係数であり、 c_w は、線分幅の非滑らかさに関するペナルティ係数である。ペナルティ係数 c_c および c_w に異なる値を使用することができ、一例示的实施形態では、 c_c および c_w の値のそれぞれが、0.5 である。

最も滑らかな領域は、次式を判定することによって得ることができる。

【 0 0 5 4 】

20

【 数 5 】

$$\max_{i^{(M)}, j^{(M)} \in [0, N]} (E^0(i^{(M)}, j^{(M)}))$$

【 0 0 5 5 】

この最も滑らかな領域を与えられて、すべての水平線上の境界を見つけるバックトレースを実行することができる。

【 0 0 5 6 】

最も滑らかな領域を与えられて、その領域が人間の上部に似ているかどうかに関する判定を行う(動作250)。図示の例では、人間の上部に、大きい肩の上の小さい頭が含まれる。

30

したがって、最も滑らかな領域が、幅広い部分(肩)の上に位置するほぼ楕円形の部分(頭)を有するかどうかに関する検査を行う。一実施形態では、この検査が、まず、隣接する水平線上での幅の最大の変化を見つけることによって首の部分を検出することによって行われる。その後、首の上の領域(頭領域)が、下側領域(肩領域)より小さい平均幅を有するかどうかに関する検査を行う。頭領域の幅対高さの比が、約1:1.2であるかどうかの検査も行う。これらの検査のすべてが真である場合に、検出された領域が、人間の上部のシルエットに似ていると判定される。

【 0 0 5 7 】

領域が、人間の上部に似ている場合に、領域のうちで頭を含む(肩は含まない)部分が、抽出され(動作252)、新しい顔領域の候補として識別される(動作254)。この抽出された領域は、人間の頭のほぼ楕円形の領域または頭の周囲の区域(例えば、頭の付近の長方形の領域)である可能性がある。しかし、領域が人間の上部に似ていない場合には、フレームから、新しい顔領域の候補が検出されない(動作256)。

40

【 0 0 5 8 】

一実施形態では、動作254で新しい顔領域の候補が識別され、フレーム内に追加の領域がある(動作254で識別された候補または他の顔または顔候補を含まない)場合に、図5の処理が繰り返される。これによって、フレーム内で新しい顔領域の追加の候補を識別できるようになる。

【 0 0 5 9 】

図3および4に戻って、オーディオベース初期化モジュール158が、音声を受け取られた方

50

向を検出するために音源ロケータ(sound source locator)を使用することによって、オーディオ/ビデオ内容のフレームを分析する(図4の動作210)。モジュール158は、この音声が入りの会話であると仮定し、したがって、顔領域の候補を含む可能性があるビデオ内容の領域を示すと仮定する。音声を受け取られる方向は、さまざまな異なる形で判定することができる。一実施形態では、1つまたは複数のマイクロホンアレイが、音声を取り込み、1つまたは複数の音源位置測定アルゴリズムが、音声が入った方向を判定するのに使用される。周知の到着時間遅延(time-delay-of-arrival、TDOA)技法(例えば、一般化相互相関(generalized cross-correlation、GCC)手法)などのさまざまな異なる普通の音源位置測定アルゴリズムを使用することができる。

【0060】

ビデオ内容がない状況では、顔判定を、複数のマイクロホンの適当な配置によって達成することができる。3つまたはそれ以上のマイクロホンを使用し、そのうちの少なくとも2つが異なる水平面上にあり、少なくとも2つが異なる垂直平面に配置される場合に、音源の(x,y)座標を判定することができる。例えば、2つのマイクロホンを、垂直面内に配置し、2つのマイクロホンを、水平面内に配置することができる。さまざまな従来の音源位置測定アルゴリズムのどれであっても、個人の口であると仮定される音源の(x,y)位置の判定に使用することができる。この音源位置自体を、検出された顔領域として扱うことができ(話者の口が話者の顔の一部であるものとして)、その代わりに、位置を拡張し(例えば、2%または3%だけ増やす)、拡張された位置を検出された顔領域として使用することができる。

【0061】

音声を受け取られた方向に対応する画像の区域を与えられて、初期化モジュール158は、その区域を分析し、その区域の画像に皮膚色モデルをあてはめることを試みる。この試みが成功である場合に、皮膚色モデルがあてはめられた区域が、新しい顔領域の候補として識別される。一実施形態では、皮膚色モデルが、HSV(色相、彩度、色価)色空間モデルであり、多数の皮膚色トレーニングデータが、モデルのトレーニングに使用される。オーディオが、領域内に顔があることを既に示しているので、粗な検出処理(例えば皮膚色モデル)を使用して、顔を突き止めることができることに留意されたい。

【0062】

ビデオ内容が使用可能でない状況では、モジュール158は、皮膚色モデルを使用せずに(皮膚色モデルを適用できるビデオ内容がないので)、音源位置判定に頼る。

【0063】

高速顔検出モジュール160は、高速顔検出器を使用して、フレームの画像の区域について顔を検出する。検出モジュール160によって使用される高速顔検出器は、下で詳細に説明する階層検証モジュール142によって使用される顔検出器と異なるものとしてすることができる。計算と精度のトレードオフのために、モジュール160によって使用される顔検出器は、階層検証モジュール142によって使用される顔検出器より高速であるが、精度が低い。しかし、モジュール160および142は、同一の顔検出アルゴリズムに基づくが、モジュール160による検出の速度をモジュール142による検出の速度より高めるために、異なるパラメータまたは閾値を使用するものとしてすることができる。その代わりに、モジュール160および142を、2つの異なる顔検出アルゴリズムに基づくものとしてすることができる。検出モジュール160によって使用される検出器は、通常は、階層検証モジュール142によって使用される検出器より高速である。

【0064】

さまざまな顔検出アルゴリズムを、高速顔検出モジュール160の基礎として使用することができ、使用されるアルゴリズムの主な特性は、その速度である。高速顔検出モジュール160の目的は、必要な場合に精度を犠牲にして、すばやく顔を検出することである。顔検出は、正面のみとすることができ、あるいは、複数ビュー(正面検出に制限されない)とすることができる。そのようなアルゴリズムの例が、文献に記載されている(例えば、非特許文献1参照)。そのようなアルゴリズムのもう1つの例は、上記文献に記載されたものに

似るが、広い範囲の視野の角度をカバーする検出器から始めて、それぞれがより狭い範囲の視野の角度をカバーする複数の検出器のセットに進む、複数ステージの検出器を使用することが異なる。オブジェクトが、検出器のあるステージから別のステージに渡され、各検出器が、オブジェクトを、顔または顔以外のいずれかとして分類する。オブジェクトが、いずれかの検出器によって顔以外として分類されるとすぐに、そのオブジェクトが処理から捨てられ、検出器のすべてのステージを通過し、顔として分類されるオブジェクトだけが、顔として識別される。

【 0 0 6 5 】

したがって、動きベース初期化、オーディオベース音源位置、および高速検出技法の1つまたは複数を使用して、自動初期化モジュール140が、新しい顔領域の候補を検出する。これらの候補は、候補に実際に顔が含まれるかどうかに関する検証のために、階層検証モジュール142に渡される。すべてのフレームに新しい顔が含まれるのではなく、したがって、自動初期化モジュール140は、上で参照した技法のすべてを使用する場合であっても、フレーム内で新しい顔領域の候補を検出しない場合があることに留意されたい。

【 0 0 6 6 】

階層検証

図3の階層検証モジュール142は、自動初期化モジュール140によって識別された候補顔領域を検証する。さらに、検出および追跡モジュール132が、複数キュー追跡モジュール144が動作中にオブジェクトを見失う可能性を考慮に入れる。これは、オクルージョン(例えば、ビデオ取込装置と追跡される個人の間を別の参加者が歩く時など)または突然の照明の変化などのさまざまな理由から発生する可能性がある。階層検証モジュール142は、規則的または不規則な間隔で、追跡される各オブジェクトを再検証し、オブジェクトを顔から顔候補に適当にダウングレードする。間隔の長さは、追跡がどれほど正確であることが望まれるか(短い間隔によって精度が改善される傾向がある)、使用可能な計算能力の量(検証のタイプに依存して、追跡が再検証より少ない計算能力を要する場合がある)、および検証モジュールの計算的出費に基づいて、変更することができる。

【 0 0 6 7 】

一実施形態では、階層検証モジュール142が、オブジェクトを顔として検証し、オブジェクトを顔または顔以外のいずれかとして識別する。その代わりに、検証モジュール142が、異なる特徴(例えば、オーディオ、色ヒストグラム距離、境界の回りのエッジ検出結果、顔検出結果など)に基づく確率的検証結果を出力することもできる。それを行う際に、出力される確率的検証結果を、下で詳細に説明するパーティクルフィルタリング(particle-filtering)の重み付け方式と組み合わせることができる。

【 0 0 6 8 】

計算の考慮事項のゆえに、階層検証モジュール142では、複数レベル階層処理を使用して、オブジェクトが顔を含むことを検証する。検証処理は、必要であれば、荒い処理から微細な処理へ、高速だが正確でない検証から始めて、低速だが高精度の検証に進む。図示の例では、階層処理に2つのレベルが含まれる。代替案では、3つまたはそれ以上のレベルを、階層処理に含めることができる。

【 0 0 6 9 】

図3の階層検証モジュール142には、高速色ベース検証モジュール164と、複数ビュー顔検出モジュール166が含まれる。検証モジュール142は、オブジェクトが、通常は連続するフレームの間で大幅に色を変化させないと仮定する。色ベース検証モジュール164は、現在のフレームのオブジェクトの色ヒストグラムと、前のフレームのオブジェクトの推定された色ヒストグラムの間の類似性に基づいてオブジェクトを検証する。類似性が高い時に、追跡の消失が発生しなかったと仮定され、複数ビュー顔検出モジュール166を呼び出す必要はない。しかし、類似性が低い時には、追跡の消失が発生した可能性があり、したがって、オブジェクトを顔から顔候補にダウングレードし、複数ビュー顔検出モジュール166に渡す。複数ビュー顔検出モジュール166が、オブジェクトを顔として検証する場合には、オブジェクトが、顔候補から顔にアップグレードされる。しかし、検出モジュール166

10

20

30

40

50

が、オブジェクトを顔として検証しない場合には、オブジェクトが、追跡リスト146から削除される。

【 0 0 7 0 】

一実施形態では、色ベース検証モジュール164が、フレームごとにその検証を実行するが、複数ビュー顔検出モジュール166は、より低い頻度でその検証を実行する。例として、複数ビュー顔検出モジュール166は、数秒に1回その検証を実行する場合があるが、上で述べたさまざまな要因に基づいて、異なる間隔を使用することもできる。

【 0 0 7 1 】

図7は、階層検証を実行する例示的処理320を示す流れ図である。処理320は、図3の階層検証モジュール142によって実行され、ソフトウェアで実行することができる。

10

【 0 0 7 2 】

最初に、関心を持たれる区域の画像を得る(動作322)。関心を持たれる区域は、自動初期化モジュール140によって識別された候補領域、または再検証に関する領域とすることができる。階層検証モジュール142に、分析される区域の指示と共にフレーム全体を渡すか、その代わりに、フレームのうちで分析される区域を含む部分だけを渡すことができる。受け取られた後に、高速色ベース検証を使用して、顔が区域にあるかどうかを検証する(動作324)。

【 0 0 7 3 】

動作324の高速色ベース検証を、図8に関してさらに詳細に示す。図8の処理324は、図3の高速色ベース検証モジュール164によって実行され、ソフトウェアで実行することができる。最初に、現在のフレーム t のオブジェクトの色ヒストグラム($q_t(x)$)を生成する(動作362)。前のフレームのオブジェクトの推定色ヒストグラム($p_{t-1}(x)$)も生成される(動作364)。推定色ヒストグラム($p_{t-1}(x)$)は、次式に従って生成される。

20

$$p_{t-1}(x) = \alpha \cdot q_{t-1}(x) + (1 - \alpha) \cdot p_{t-2}(x)$$

【 0 0 7 4 】

ここで、 α は、重みを表し、 $q_{t-1}(x)$ は、前のフレーム $t-1$ のオブジェクトの色ヒストグラムであり、 $p_{t-2}(x)$ は、前のフレーム $t-1$ のオブジェクトについて生成される推定色ヒストグラムである。異なる実施形態でさまざまな範囲の α の値を使用することができ、正確な値は、ヒストリの信頼と、現在のフレームの信頼の間のトレードオフとして選択される(例えば、1つの例示的実施形態で、 α の値を、0.25から0.75までの範囲とすることができる)。オブジェクトの推定色ヒストグラム $p_{t-1}(x)$ は、したがって、各フレームのオブジェクトの色ヒストグラムに基づいて更新される。

30

【 0 0 7 5 】

その後、2つのヒストグラムの類似性を判定する(動作366)。2つのヒストグラム $q_t(x)$ および $p_{t-1}(x)$ の類似性の尺度を判定するために、周知のバタチャリア(Bhattacharyya)係数を、次のように使用する。

【 0 0 7 6 】

【 数 6 】

$$\rho(p_{t-1}(x), q_t(x)) = \int \sqrt{p_{t-1}(x) \cdot q_t(x)} dx$$

40

【 0 0 7 7 】

ここで、 ρ は、統計的仮説テストの分類誤差の確率を表し、誤差の確率が大きいほど、2つの分布が似ている。 ρ の値は、0から1までの範囲であり、1が、2つのヒストグラムが同一であることを意味し、0が、2つのヒストリが全く異なることを意味する。この類似性の尺度を、本明細書では信頼性レベルとも称する。その代わりに、K-L発散、ヒストグラム交叉など、他の周知の類似性尺度を使用することができる。

【 0 0 7 8 】

その後、2つのヒストグラムの間の類似性が、閾値量を超えるかどうかに関する検査を行う(動作368)。差が、閾値量より大きい場合には、顔が検証される(動作370)、すなわち、

50

オブジェクトが、顔を含むものとして検証される。しかし、差が、閾値量を超えない場合には、顔が検証されない(動作372)、すなわち、オブジェクトが、顔を含むものとして検証されない。異なる実施形態で、異なる閾値を使用することができる。1つの例示的实施形態では、閾値を、0.90から0.95の範囲とすることができ、1つの特定の实施形態では、0.94である。

【0079】

図7に戻って、処理は、顔が検証されたかどうかに基づいて進行する(動作326)。顔が検証される場合には、顔候補から顔にアップグレードされ(まだ顔でない場合に)(動作328)、階層検証処理が完了し(動作330)、この時にこの関心を持たれる区域に関してこれ以上の検証は実行されない。しかし、顔が検証されない場合には、顔を、顔から顔候補にダウングレードする(現在顔である場合)(動作332)。顔を含むオブジェクトが、図3の複数ビュー顔検出モジュール166に渡され、複数ビュー顔検出モジュール166が、複数ビュー顔検出を使用して、顔が区域内にあるかどうかを検証する(動作334)。

10

【0080】

複数ビュー顔検出モジュール166は、異なるポーズでまたは複数のビューからの人間の顔を検出する(すなわち、頭が、傾くか、画像取込装置から離れる方に回転されるなどの場合であっても顔を検出する)ことを試みる、1つまたは複数の検出処理を使用する。さまざまな顔検出技法のどれでも、複数ビュー顔検出モジュール166によって使用することができる。

【0081】

そのような複数ビュー顔検出処理が、カーネルマシンベースの処理(kernel machine based process)である(例えば、非特許文献2参照)。この検出処理の概要を下に示す。

20

【0082】

【数7】

$$I_p \in \mathbb{R}^N$$

【0083】

が、顔の、窓付きのグレイレベル画像または外見であるものとする。すべての左に回転された顔(91°と180°の間のビュー角度を有するもの)が、右回転に鏡像化され、その結果、すべてのビュー角度が、0°と90°の間になると仮定する。ポーズを、L個の離散値の組に量子化する(例えば、10個の等間隔の角度0°から90°について、L=10を選択し、0°が、右側面ビューに対応し、90°が、正面ビューに対応するものとする)。

30

【0084】

トレーニング顔画像の組が、学習のために提供されると仮定する。画像 I_p は、ビュー内の変化だけではなく、照明の変化にも影響される。トレーニングセットは、各顔画像が、できる限り真に近いビュー値を用いて手動でラベルを付けられ、その後、最も近いビュー値に従ってL個のグループの1つに割り当てられるという点で、ビューによってラベル付けされる。これによって、顔のビューサブ空間を学習するための、L個のビューによってラベル付けされた顔画像サブセットが作られる。顔以外の画像のもう1つのトレーニングセットも、顔検出のトレーニングに使用される。

40

【0085】

ここで、 l に従ってインデクシングされる $L+1$ 個のクラスがあり、 $l \in \{0, 1, \dots, L-1\}$ が、顔のL個のビューに対応し、 $l=L$ が、顔以外のクラスに対応する。顔検出およびポーズ推定という2つのタスクが、入力 I_p を $L+1$ 個のクラスの1つに分類することによって、共同で実行される。入力が、L個の顔クラスの1つに分類される場合に、顔が、検出され、対応するビューが、推定されるポーズになり、そうでない場合には、入力パターンが、顔以外のパターンとみなされる。

【0086】

カーネルマシンを使用する顔検出およびポーズ推定のための学習は、2ステージすなわち、カーネルプリンシパルコンポーネント分析(kernel principal component analysis、KP

50

CA) ビュー-サブ空間学習の1ステージと、カーネルサポートベクトルクラシファイヤ(kernel support vector classifier、KSVC)クラシファイヤトレーニング用の1ステージで実行される。ステージ1のトレーニングは、L個の顔ビューサブセットからL個のKPCAビュー-サブ空間を学習することを目指す。1組のカーネルプリンシパルコンポーネント(KPC)が、各ビューサブセットから学習される。最も重要なコンポーネント(例えば上位50個)が、ビュー-サブ空間を構成する基本ベクトルとして使用される。このステージでの学習は、それぞれがサポートベクトルおよび対応する係数の組によって決定されるL個のビュー-サブ空間をもたらす。各ビューチャンネル内のKPCAによって、入力画像空間から出力KPCA特徴空間(最も重要なコンポーネントのコンポーネント数と同一の次元を有する)への非線形写像が効率的に実行される。

10

【0087】

ステージ2では、顔検出に関して顔パターンと顔以外のパターンを区別するためにL個のKSVCをトレーニングすることを目指す。ここでは、顔以外のサブセットならびにL個のビュー顔サブセットからなるトレーニングセットを使用する。KSVCを、ビューごとにトレーニングして、対応するKPCAサブ空間内の特徴に基づくL+1個のクラス分類を実行する。対応するビューのKPCAサブ空間への射影を、特徴ベクトルとして使用する。周知のワンアゲinstザレスト(one-against-the-rest)法を、KSVCの複数クラス問題の解決に使用する。ステージ2では、L個のKSVCが得られる。

【0088】

テストステージでは、ビューIごとに、テストサンプルをKPCA特徴エクストラクタに提示して、そのビューの特徴ベクトルを得る。そのビューの対応するKSVCによって、入力に対するL+1個のクラスの応答としての出力ベクトル

20

【0089】

【数8】

$$y_l = (y_l^c \mid c = 0, \dots, L)$$

【0090】

が計算される。これは、L個のビューチャンネルのすべてについて行われ、その結果、L個のそのような出力ベクトル $\{y_l \mid l=0, \dots, L-1\}$ が作られる。値

30

【0091】

【数9】

$$y_l^c$$

【0092】

が、I番目のビューのKPCAサブ空間内の特徴に関して入力 I_p がクラスcに属するという判断の証拠である。最終的な分類の決定は、L個のビューチャンネルのすべてからの証拠を融合させることによって行われる。融合の1つの方法が、証拠の合計である。すなわち、クラス $c=0, \dots, L$ のそれぞれについて、次式を計算する。

40

【0093】

【数10】

$$y^c(I_p) = \sum_{l=0}^{L-1} y_l^c$$

【0094】

この計算によって、 I_p をクラスcに分類するための総合的な証拠が得られる。最終決断は、証拠を最大にすることによって行われ、 $c^* = \arg \max_c y^c(I_p)$ の場合に、 I_p が c^* に属する。

50

【 0 0 9 5 】

図7の説明を続けると、その後、処理は、顔が複数ビュー顔検出によって検証されたかどうかに基づいて進行する(動作336)。顔が検証される場合には、顔が、顔候補から顔にアップグレードされ(動作328)、階層検証処理が完了する(動作330)。しかし、顔が検証されない場合には、図3の追跡リスト146から候補を除去し(動作338)、階層検証処理が完了する(動作330)。

【 0 0 9 6 】

階層検証について分析すべきビデオ内容がない状況では、適当な時に、オーディオキューだけを検証に使用することができる。例えば、顔が追跡されている人が連続的に話している時、または周知のオーディオ話者ベース識別が実行される時(これによって、音源を個別の話者の声に結び付けられるようになり、特定の音源位置から来る声が、その音源位置から前に受け取られたものと同じの話者識別と一致するかどうかを判定することによって検証が実行される)に、オーディオキューだけを使用することができる。

【 0 0 9 7 】

複数キュー追跡

ビデオ内容のフレーム内で顔を検出した後に、その顔を、図3の複数キュー追跡モジュール144が、ビデオ内容の後続のフレームで追跡する。顔が追跡されている参加者が、動きまわる可能性があり、したがって、顔の位置が、ビデオ内容の異なるフレームで異なる可能性がある。さらに、参加者が、頭を回転する(例えば、ビデオ取込装置を直接見なくなるように)可能性があり、さまざまなオクルージョンが発生する可能性があり(例えば、参加者が顔の前で自分の手を動かす場合がある)、照明が変化する可能性がある。複数キュー追跡モジュール144は、フレームからフレームへ発生する可能性があるこれらのさまざまな変化を計上することを試みる。さらに、これらの変化のゆえに、いくつかのキューが、追跡するには信頼性がない状態になる場合がある。複数キュー追跡モジュール144は、フレームからフレームへ発生する可能性があるキュー信頼性の変化を計上することも試みる。

【 0 0 9 8 】

さまざまなキューが、顔を追跡する際に追跡モジュール144によって使用される。一実施形態では、これらの追跡キューに、顔の形状(楕円としてモデル化される)、動き、エッジ、前景色、および背景色が含まれる。その代わりに、これらのキューの1つまたは複数を使用しないか、オーディオキューなどの追加のキューを使用することができる。

【 0 0 9 9 】

複数キュー追跡モジュール144は、オーディオ内容が使用可能な時に、追跡を支援するのに(または、追跡の単独の基礎として)オーディオキューを使用することができる。オーディオベースの追跡は、音源位置処理に基づいて実行され、オーディオベース検出が上で述べた図3のオーディオベース初期化モジュール158によって実行されるのと同じの形で実行される。

【 0 1 0 0 】

図9は、複数キュー追跡を実行する例示的処理400を示す流れ図である。処理400は、図3の複数キュー追跡モジュール144によって実行され、ソフトウェアで実行することができる。

【 0 1 0 1 】

最初に、前のフレーム $t-1$ からの追跡結果とオブジェクトの動力学(周知のランジュバン過程によってモデル化され、下で詳細に説明する)に基づいて、オブジェクトが現在のフレーム t 内にあるかどうかに関する予測を行う(動作402)。観測を、オブジェクトの予測される輪郭の法線の組に沿って収集し(動作404)、観測尤度関数を、法線上のすべての画素について評価する(動作406)。フレーム $t-1$ からフレーム t への状態遷移確率を評価し(動作408)、所与の観測に関する最もよい輪郭を判定する(動作410)。最もよい楕円を、検出された輪郭に基づいてフレーム t の画像にあてはめ(動作412)、モデルを、次のフレーム $t+1$ に関する使用に適応させる(動作414)。

【 0 1 0 2 】

複数キュー追跡モジュール144には、図9の動作を実行するさまざまなモジュールが含まれる。図示の例では、追跡モジュール144に、観測尤度モジュール168、滑らかさ制約モジュール170、輪郭選択モジュール172、およびモデル適応モジュール174が含まれる。

【 0 1 0 3 】

複数キュー追跡モジュール144は、人間の頭の追跡に焦点を合わせるが、この頭は、楕円形の形状(約1:1.2)を有する。追跡される顔に関する人間の頭は、さまざまな追跡キューを有する楕円であるモデルによって表される。ビデオ内容のフレームの画像を分析する時に、モデルが、画像内のさまざまな位置と比較され、どの位置がモデルに最もよく一致するかに関する判定が行われる。モデルに最もよく一致する位置を、新しいフレーム内の顔として選択する。

10

【 0 1 0 4 】

図10に、このモデル化および比較を詳細に示す。図10では、実線の曲線422が、直前のフレーム $t-1$ からの追跡結果に基づく、特定のフレーム t での人間の頭の予測された輪郭を表す。破線の曲線424は、フレーム t での人間の頭の真の輪郭を表す。予測された輪郭422の複数(M 個)の法線426に沿って、測定値の組を収集する。点428($c(\varphi)$)は、 φ 番目の法線上の真の輪郭点である。点430($p_\varphi(N)$)は、 φ 番目の法線上の予測された輪郭点である。複数キュー追跡モジュール144は、予測された輪郭422上のできる限り多くの輪郭点を、真の輪郭線424上の輪郭点と同一にすることによって、真の輪郭424を突き止めることを試みる。

20

【 0 1 0 5 】

図3の観測尤度モジュール168は、次式のように線 φ 上の画素 λ の画像輝度を表す値 $\rho_\varphi(\lambda)$ を生成する。

$$\rho_\varphi(\lambda) = I(x_{\lambda\varphi}, y_{\lambda\varphi})$$

ここで、 φ は、1から M (法線426の総数)の範囲にわたり、 λ は、法線に沿って $-N$ から N までの範囲にわたり(各法線が $2N+1$ 画素を有する)、 $x_{\lambda\varphi}, y_{\lambda\varphi}$ は、 φ 番目の法線の画素 λ の対応する画像座標であり、 $I(x_{\lambda\varphi}, y_{\lambda\varphi})$ は、点 $(x_{\lambda\varphi}, y_{\lambda\varphi})$ での画像輝度である。

【 0 1 0 6 】

輪郭点を検出するために、異なるキュー(例えば、エッジの輝度、前景と背景の色モデル)および前の制約(例えば輪郭の滑らかさの制約)を、隠れマルコフモデル(HMM)を使用することによって統合することができる。隠れマルコフモデルは、当業者に周知であり、したがって、本明細書に記載の複数の個人の自動化された追跡に関する部分を除いて、これ以上説明しない。HMMの隠れ状態は、各法線上の真の輪郭点である($s = \{s_1, \dots, s_\varphi, \dots, s_M\}$ と表す)。HMMの観測 $O = \{O_1, \dots, O_\varphi, \dots, O_M\}$ が、各法線 φ に沿って収集される。HMMは、複数の状態(この例では $2N+1$ 個)、観測モデル $P(O_\varphi | s_\varphi)$ 、および遷移確率 $p(s_\varphi | s_{\varphi-1})$ によって指定される。

30

【 0 1 0 7 】

観測尤度モジュール168は、下記のように、複数キュー観測尤度関数の生成に移る。線 φ での観測(O_φ と表す)に、複数のキュー、例えば、線に沿った画素輝度(すなわち、 $\rho_\varphi(\lambda), \lambda \in [-N, N]$)およびエッジ輝度(すなわち、 z_φ)を含めることができる。エッジ検出結果 z_φ の観測尤度モデルは、周知のソベル(Sobel)エッジ検出器またはキャニイ(Canny)エッジ検出器などのさまざまな普通のエッジ検出処理のいずれかを使用して導出することができる。雑音および画像クラッタに起因して、各法線 φ に沿って複数のエッジがある可能性がある。値 J を使用して、検出されたエッジの数を表す($z_\varphi = (z_1, z_2, \dots, z_J)$)。 J 個の検出されたエッジのうちで、多くとも1つが、図10の真の輪郭線424である。したがって、 $J+1$ 個の仮説

40

$$H_0 = \{e_j = F : j = 1, \dots, J\}$$

$$H_1 = \{e_j = T, e_k = F : k = 1, \dots, J, k \neq j\}$$

を定義することができる。ここで、 $e_j = T$ は、 j 番目のエッジが、真の輪郭線に関連することを意味し、 $e_j = F$ は、 j 番目のエッジが、真の輪郭線に関連しないことを意味する。した

50

がって、仮説 H_0 は、エッジのどれもが、真の輪郭線に関連しないことを意味する。

【 0 1 0 8 】

画像クラッタが、空間密度 γ を有する線に沿った周知のポアソン(Poisson)過程であり、真の目標測定値が、標準偏差 σ_z の正規分布であると仮定すると、エッジ尤度モデルは、次式として得られる。

【 0 1 0 9 】

【 数 1 1 】

$$p(z_\phi | s_\phi = \lambda_\phi) \propto 1 + \frac{1}{\sqrt{2\pi}\sigma_z q \gamma} \sum_{m=1}^J \exp\left(-\frac{(z_m - \lambda_\phi)^2}{2\sigma_z^2}\right) \quad 10$$

【 0 1 1 0 】

ここで、 q は、仮説 H_0 の前の確率である。

【 0 1 1 1 】

エッジ尤度モデルの他に、例えば、混合色モデルなどの、前景および背景の領域特性に関する他のキューが、HMMフレームワークに統合される。 $p(v|FG)$ および $p(v|BG)$ が、それぞれ前景(FG)および背景(BG)の色分布を表すものとする。経験的確率 $P(BG|v)$ および $P(FG|v)$ は、次のように導出することができる。

【 0 1 1 2 】

【 数 1 2 】

$$\begin{aligned} P(BG|v) &= \frac{p(v|BG)}{p(v|BG) + p(v|FG)} \\ P(FG|v) &= \frac{p(v|FG)}{p(v|BG) + p(v|FG)} \end{aligned} \quad (1)$$

【 0 1 1 3 】

$s_\phi = \lambda_\phi$ が、線 ϕ 上の輪郭点である場合に、線分 $[-N, s_\phi]$ が、前景にあり、線分 $[s_\phi + 1, N]$ が、背景にある。エッジ尤度モデルと色の経験的確率を組み合わせると、下記のHMMに関する複数キュー観測尤度関数がもたらされる。

【 0 1 1 4 】

【 数 1 3 】

$$P(O_\phi | s_\phi) = p(z | s_\phi) \cdot \prod_{i=-N}^{s_\phi} P(BG | v = \rho_\phi(i)) \cdot \prod_{i=s_\phi+1}^N P(FG | v = \rho_\phi(i)) \quad (2)$$

【 0 1 1 5 】

オーディオキュー(例えば音源位置および特定の位置から来る音声の尤度に基づく)などの他のキューも、同様の形で統合することができる。分析されるビデオ内容がない状況では、オーディオキューだけが使用される。その代わりに、そのようなオーディオキューに加えてまたはその代わりに、オーディオを、下で詳細に説明するアンセンテッドパーティクルフィルタリング(unscented particle-filtering)を用いる提案関数として使用することができる。

【 0 1 1 6 】

HMMのもう1つの構成要素が、遷移確率であり、これによって、時刻 $t-1$ の状態が、どのように時刻 t の別の状態に遷移するかが決定される。図3の滑らかさ制約モジュール170が、推移確率を導出する。

【 0 1 1 7 】

10

20

30

40

50

滑らかな輪郭を得るために、推移確率を使用して、滑らかさ制約をエンコードし、粗さにペナルティを与える。図10を参照すると、法線426が、密(例えば、30法線程度)である時に、隣接する法線426の真の輪郭線424の点が、予測された輪郭線422(各法線上で0としてインデクシングされる)からの同一の変位を有する傾向があることが分かる。この相関を使用して、滑らかな輪郭を得るのを助ける。

【0118】

HMMでは、現在の状態 s_φ を与えられれば、現在の観測 o_φ は、前の状態 $s_{\varphi-1}$ および前の観測 $o_{\varphi-1}$ と独立である。さらに、マルコフ過程の特性のゆえに、 $p(s_\varphi | s_1, s_2, \dots, s_{\varphi-1}) = p(s_\varphi | s_{\varphi-1})$ が得られる。

【0119】

輪郭の滑らかさ制約を、状態遷移 $p(s_\varphi | s_{\varphi-1})$ によって、次のように取り込むことができる。

【0120】

【数14】

$$p(s_\varphi | s_{\varphi-1}) = c \cdot \exp(-(s_\varphi - s_{\varphi-1})^2 / \sigma_s^2) \quad (3)$$

【0121】

ここで、 c は正規化定数、 σ_s は、輪郭の滑らかさを規定する事前定義の定数である。この遷移確率では、隣接する線の間の輪郭点の突然の変化にペナルティが与えられ、したがって、滑らかな輪郭がもたらされる。最適の輪郭は、輪郭選択モジュール172によって得ることができる。

【0122】

上の計算(3)に基づいて滑らかさ制約モジュール170によって生成される遷移確率では、法線上の他の画素にかかわらずに輪郭点を検討する。代替案では、滑らかさ制約モジュール170が、JPDAF(joint probability data association filter)ベースの方法を使用して、輪郭滑らかさ制約だけでなく、法線上の複数の(例えばすべての)画素で観測された領域滑らかさ制約もエンコードする。図示の例では、動的計画法に基づくJPDAF過程を使用して、リアルタイム性能を向上させる。

【0123】

通常の条件の下では、人体の部分(例えば顔または頭)の画素輝度値が、その領域の内側で滑らかに変化する。したがって、人間追跡において、前景と背景が、滑らかな領域特性を有し、その結果、2つの隣接する線の測定値が類似することが、適当な仮定である。 s_φ および $s_{\varphi+1}$ が、それぞれ線 φ および線 $\varphi+1$ 上の輪郭点であるものとする。これらの2つの輪郭点によって、2つの線が前景線分と背景線分に分割される。領域滑らかさ仮定に基づく、 s_φ および $s_{\varphi+1}$ が、互いに近くなければならないだけでなく、2つの線上の他のすべての画素も、十分に一致しなければならない。領域滑らかさ制約を得るために、joint probability data association filterを使用して、線マッチングを行う。すなわち、これは、単一の点对単一の点のマッチングの問題ではなく、 $(2N+1)$ 個の点对 $(2N+1)$ 個の点のマッチングの問題である。線に沿ったすべての画素を一緒に検討することによって、より堅牢なマッチング結果を得ることができる。このJPDAF過程に基づく遷移確率は、したがって、通常はより正確である。 $D^F(i, j)$ および $D^B(i, j)$ が、それぞれ前景のマッチングディスタンス(線 φ 上の $[-N, i]$ と線 $\varphi+1$ 上の $[-N, j]$)および背景のマッチングディスタンス(線 φ 上の $[i+1, N]$ と線 $\varphi+1$ 上の $[j+1, N]$)であるものとする。遷移確率を、次のように定義して、計算(3)に関して上で説明した確率と置換することができる。

【0124】

【数15】

$$\log(p(s_2 | s_1)) = D^F(s_1, s_2) + D^B(s_1, s_2) + (s_2 - s_1)^2 / \sigma_s^2 \quad (4)$$

【 0 1 2 5 】

領域の滑らかさの概念は、図11に示された合成された画像によって示すことができる。2つの領域すなわち、背景クラッタを表す長方形領域460と、オブジェクトを表すほぼ円形の領域462が示されている。2つの隣接する法線464および466も示されている。点aおよびbは、線464上で検出されたエッジ点であり、点cおよびdは、線466上で検出されたエッジ点である。目標は、輪郭点が、この2つの線464および466のどこにあるかを見つけることである。2つの線464および466に沿った輝度の測定値を、図12に示す。測定482は、線464に沿った輝度を表し、測定484は、線466に沿った輝度を表す。測定482および484は、多少のひずみを除いて互いに類似する。輪郭の滑らかさ制約だけに基づいて、aからcまでの輪郭とbからcまでの輪郭は、 $|a-c|$ $|b-c|$ なので、ほぼ同一の量の滑らかさエネルギーを有する。しかし、領域滑らかさの前提も考慮する場合には、可能な輪郭は、adまたはbcであって、acまたはbdではない。輪郭候補adおよびbcを、すべての観測線に基づいてHMMによってさらに弁別することができる。

10

【 0 1 2 6 】

新しい遷移確率を得るために、状態のすべての可能な対の間のマッチング $((2N+1)^2)$ を計算する。図13に、マッチングディスタンスの計算を線図的に示す。線464および466を与えられて、マッチングディスタンスの計算は、次の再帰式で説明することができ、図13で見ることができる。

20

【 0 1 2 7 】

【 数 1 6 】

$$D^F(i, j) = \min \begin{cases} D^F(i-1, j) + d(\rho_1(i), \rho_2(j)) \\ D^F(i, j-1) + d(\rho_1(i), \rho_2(j)) \\ D^F(i-1, j-1) + d(\rho_1(i), \rho_2(j)) \end{cases}$$

30

【 0 1 2 8 】

ここで、 $d(\cdot, \cdot)$ は、2つの画素のマッチングのコストである。 $D^F(i, j)$ は、線464上の線分 $[-N, i]$ と線466上の線分 $[-N, j]$ の間の最良マッチングディスタンスである。 $D^F(0, j) = D^F(i, 0) = 0$ 、ただし $i, j \in [-N, N]$ から開始して、上の再帰を使用して、 $i = -N$ から N まで、および $j = -N$ から N までについてマッチングディスタンス $D^F(i, j)$ を得る。類似する処理を経て、 $D^B(i, j)$ を計算するが、この場合には、 $D^B(N, N) = 0$ から $D^B(-N, -N)$ までである。すべてのマッチングディスタンスを得た後に、状態遷移確率を計算でき、輪郭追跡を、下で詳細に説明する図3の輪郭選択モジュール172によって達成することができる。

【 0 1 2 9 】

観測シーケンス $O = \{O_\phi, \phi \in [1, M]\}$ および遷移確率 $a_{i, j} = p(s_{\phi+1} = j | s_\phi = i)$ を与えられて、輪郭選択モジュール172は、周知のビタビ(Viterbi)アルゴリズムを使用して、次のように最尤状態シーケンス s^* を見つけることによって、最良の輪郭を判定する。

40

【 0 1 3 0 】

【 数 1 7 】

$$s^* = \arg \max_s P(s | O) = \arg \max_s P(s, O)$$

【 0 1 3 1 】

値 $V(\phi, \lambda)$ は、次のように定義される。

【 0 1 3 2 】

50

【数 1 8】

$$V(\phi, \lambda) = \max_{s_{\phi-1}} P(O_{\phi-1}, s_{\phi-1}, s_{\phi} = \lambda)$$

【0 1 3 3】

マルコフの条件独立の前提を使用して、 $V(\phi, \lambda)$ を、次のように再帰的に計算することができる。

【0 1 3 4】

【数 1 9】

$$V(\phi, \lambda) = P(O_{\phi} | s_{\phi} = \lambda) \cdot \max_j P(s_{\phi} = \lambda | s_{\phi-1} = j) V(j, \phi-1)$$

10

$$j^*(\phi, \lambda) = P(O_{\phi} | s_{\phi} = \lambda) \cdot \arg \max_j P(s_{\phi} = \lambda | s_{\phi-1} = j) V(j, \phi-1)$$

【0 1 3 5】

ここで、初期化は、 $V(1, \lambda) = \max_{s_1} P(O_1 | s_1) P(s_1)$ であり、初期状態確率は、 $P(s_1) = 1/(2N+1)$ 、 $s_1 \in [-N, N]$ である。項 $j^*(\phi, \lambda)$ によって、線 ϕ での状態 λ からの「最良の前の状態」が記録される。したがって、シーケンスの終りに、 $\max_s P(O, s) = \max_{\lambda} V(M, \lambda)$ が得られる。最適状態シーケンス s^* は、 $s_M^* = \arg \max_{\lambda} V(M, \lambda)$ かつ $s_{\phi-1}^* = j^*(s_{\phi}^*, \phi)$ から始めて、 j^* をバックトラッキングすることによって得ることができる。

20

【0 1 3 6】

最良状態シーケンス $s^* = \{s_1^*, \dots, s_M^*\}$ を与えられて、線 ϕ 上の最良の輪郭点 s_{ϕ}^* の対応する画像座標を、 $[x_{\phi}, y_{\phi}]$ と表す。パラメトリック輪郭モデルとして楕円が使用されるので、輪郭点 $[x_{\phi}, y_{\phi}]$ のそれぞれについて、下記が成り立つ。

【0 1 3 7】

【数 2 0】

$$ax_{\phi}^2 + by_{\phi}^2 + cx_{\phi}y_{\phi} + dx_{\phi} + ey_{\phi} - 1 = 0$$

【0 1 3 8】

これらの式の行列表現が、

$$A^*f = b$$

であり、ここで、

【0 1 3 9】

【数 2 1】

$$A = \begin{bmatrix} x_1^2 & y_1^2 & x_1 y_1 & x_1 & y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_M^2 & y_M^2 & x_M y_M & x_M & y_M \end{bmatrix}$$

40

【0 1 4 0】

および $b = [1, 1, \dots, 1]^T$ である。最良一致楕円のパラメータ $f^* = [a, b, c, d, e]^T$ は、平均最小自乗(LMS)解によって得ることができる。

$$f^* = (A^T A)^{-1} A^T b \quad (5)$$

上の楕円表現 $f = [a, b, c, d, e]^T$ は、数学的に便利である。しかし、この5つのパラメータの明確な物理的解釈はない。追跡では、通常は次の5要素楕円表現を使用する。

$$\theta = [x, y, \alpha, \beta, \phi]$$

ここで、 (x, y) は、楕円の中心、 α および β は、楕円の長軸および短軸の長さ、 ϕ は、楕円の向きである。 f および θ が、同一の楕円の2つの表現なので、本明細書ではこれらを交

50

換可能に使用する。

【 0 1 4 1 】

動的環境では、追跡されるオブジェクトと背景の両方が、徐々に外見を変化させる可能性がある。したがって、モデル適応モジュール174が、観測尤度モデルを動的に適応させる。観測尤度モデルを適応させる方法の1つが、フレームt-1にピタビアルゴリズムによって返された輪郭を完全に信頼し、輪郭の内側と外側のすべての画素の平均をとって、フレームtの新しい前景/背景色モデルを得ることである。しかし、フレームt-1で誤差が発生した場合に、この手順は、モデルを誤った形で適応させる可能性がある。したがって、モデル適応モジュール174は、確率的な形で観測モデルをトレーニングする。

【 0 1 4 2 】

フレームt-1で得られた輪郭を完全に信頼するのではなく、前向き後向きアルゴリズムを使用することによって、観測モデルを更新する方法に関する決定を行う。「前向き確率分布」は、次のように定義される。

$$\alpha_{\phi}(s) = p(O_1, O_2, \dots, O_{\phi}, s_{\phi} = s)$$

これは、次のように再帰を使用して計算することができる。

【 0 1 4 3 】

【数 2 2】

$$\alpha_1(s) = p(s_1 = s)p(O_1 | s_1 = s)$$

$$\alpha_{\phi+1}(s) = \left[\sum_u \alpha_{\phi}(u) a_{u,s} \right] p(O_{\phi+1} | s_{\phi+1} = s)$$

【 0 1 4 4 】

同様に、「後向き確率分布」は、次のように定義される。

$$\beta_{\phi}(s) = p(O_{\phi+1}, O_{\phi+2}, \dots, O_M, s_{\phi} = s)$$

これは、次のように再帰を使用して計算することができる。

【 0 1 4 5 】

【数 2 3】

$$\beta_M(s) = 1$$

$$\beta_{\phi}(s) = \sum_u a_{s,u} p(O_{\phi+1} | s_{\phi+1} = u) \beta_{\phi+1}(u)$$

【 0 1 4 6 】

前向き確率および後向き確率を計算した後に、線φでの各状態の確率を、次のように計算することができる。

【 0 1 4 7 】

【数 2 4】

$$P(s_{\phi} = s | \mathbf{O}) = \frac{\alpha_{\phi}(s) \beta_{\phi}(s)}{\sum_u \alpha_{\phi}(u) \beta_{\phi}(u)}, \quad s \in [-N, N]$$

【 0 1 4 8 】

これは、測定線φでのsに輪郭点を有する確率を表す。

この確率に基づいて、画素 λ_{ϕ} が前景(または背景)内にある確率を、次のように法線に沿って $P(s_{\phi} = s | \mathbf{O})$ を積分することによって計算することができる。

【 0 1 4 9 】

【数 2 5】

$$P(\lambda_\phi \in BG) = 1 - P(\lambda_\phi \in FG) = \prod_{s=-N}^{\lambda_\phi} p(s_\phi = s | \mathbf{O})$$

【 0 1 5 0 】

この確率によって、観測モデルの適応中に異なる画素に重みを付ける堅牢な方法が与えられる。より信頼性のある形で分類された画素は、色モデルにより多く貢献し、信頼性の低い形で分類された画素は、色モデルへの貢献が少ない。

【 0 1 5 1 】

【 数 2 6 】

10

$$p(v | BG) = \frac{\sum_{s=-N}^N P(s \in BG) \cdot O_\phi(s)}{\sum_{s=-N}^N P(s \in BG)}$$

$$p(v | FG) = \frac{\sum_{s=-N}^N P(s \in FG) \cdot O_\phi(s)}{\sum_{s=-N}^N P(s \in FG)} \quad (6)$$

【 0 1 5 2 】

20

新たに適応されたモデルは、追跡中の変化する色分布を反映する。新たに適応されたモデルは、次のフレームの輪郭検索中に式(1)に挿入される。図示の例で、遷移確率は、通常は追跡処理中に相対的に安定したままになる傾向があるので、トレーニングされない。その代わりに、色分布のトレーニングに類似する形で遷移確率をトレーニングすることができる。

【 0 1 5 3 】

図9に戻って、複数キュー追跡処理400を、さらに、図14に関して見ることができる。図14に、時刻t-1のフレーム522から時刻tの次のフレーム524へのオブジェクトの追跡を示す。前のフレームt-1での追跡結果およびオブジェクトの動力学に基づいて、現在のフレームtでのオブジェクトの位置に関する予測を行う(動作402)。予測された輪郭の法線の組に沿って観測を収集する(動作404)。周知のランジュバン過程を使用して、人間の動きの動力学をモデル化する。

30

【 0 1 5 4 】

【 数 2 7 】

$$\begin{bmatrix} \theta_t \\ \dot{\theta}_t \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ 0 & a \end{bmatrix} \begin{bmatrix} \theta_{t-1} \\ \dot{\theta}_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix} m_t$$

【 0 1 5 5 】

ここで、 $\theta = [x, y, \alpha, \beta, \phi]$ が、パラメトリック楕円であり、

40

【 0 1 5 6 】

【 数 2 8 】

$$a = \exp(-\beta_\theta \tau), \quad b = \bar{v} \sqrt{1 - a^2}$$

【 0 1 5 7 】

である。 β_θ は、比率定数であり、mは、正規分布 $N(0, Q)$ から引き出される熱励起過程であり、 τ は、打ち切り時間ステップであり、

【 0 1 5 8 】

【 数 2 9 】

50

$\bar{v}$

【0159】

は、定常状態二乗平均平方根速度である。

【0160】

観測尤度関数を、エッジ検出および上の計算(2)を使用することによる線上の各画素の色値に基づいて、法線 ϕ 上のすべての画素について評価する(動作406) $p(0_\phi | s_\phi = \lambda_\phi)$, $\lambda_\phi \in [-N, N]$, $\phi \in [1, M]$

JPDFAFに基づく状態遷移確率も、上の計算(4)に示されているように評価する(動作408)。

【0161】

10

前に計算した観測尤度および遷移確率行列を用いて、所与の観測に関する最良輪郭を、ピタビアルゴリズムによって見つけ(動作410)、検出された輪郭に基づいて、最良楕円を、上の計算(6)を使用してあてはめる(動作412)。

【0162】

その後、前向き後向きアルゴリズムを使用して、法線上の各画素の(前景および背景への)ソフト分類を推定し、上の計算(6)に基づいて前景および背景の色モデルを更新する(動作414)。

図9の処理400を、ビデオ内容のフレームごとに繰り返す。

【0163】

複数キュー追跡モジュール144は、キュー信頼性およびキュー信頼性の変化を計上することを試みる。例えば、前景と背景の両方の特性をモデル化し(上の計算(1)参照)、モデルを上記の計算(2)で使用する、境界を検出する(例えば、前景と背景の色が似ている場合に、これは境界検出にあまり貢献せず、処理は、動きなどの、より識別力のある他のキューにより多く頼る)。背景および前景のモデルは、追跡中に適応もされるが、これは上の計算(6)によって表されている。

20

【0164】

上で述べた複数キュー追跡処理に対して、さまざまな修正を行うこともできる。1代替案によれば、追跡される顔の1つまたは複数の特徴点の組が維持され、新しいフレームのそれぞれが、特徴点のその組を突き止めるために分析される。特徴点の組を突き止めた後に、顔の位置を、点の突き止められた組に基づいて粗いレベルで推定することができ、この粗い推定が、上で述べたパラメトリック輪郭追跡処理の初期推測として使用される。言い換えると、新しいフレームを分析して、上で述べた予測された位置を信頼するのではなく、パラメトリック輪郭追跡処理の初期推測を突き止める。この修正は、連続するフレームの間のオブジェクトの動きが大きい(上で述べた予測された位置が、後続フレーム内の実際の輪郭位置に十分に近くない可能性があるのに十分に大きい)状況で、特に有用になる可能性がある。

30

【0165】

目の端、口の端、鼻孔など、さまざまな異なる特徴点を追跡することができる。オーディオの音源も、視覚的特徴に加えてまたはその代わりに、特徴点として追跡することができる。周知のルーカス-カナデ特徴トラッカ(例えば、非特許文献3参照)など、さまざまな異なる特徴追跡処理を使用することができる。

40

【0166】

上で述べた複数キュー追跡処理に対して行うことができるもう1つの修正は、確率的サンプリングを実行する時に、状態空間からではなく、特徴点(検出された輪郭点)からサンプリングすることである。例えば、複数の輪郭点を、検出された輪郭点のすべておよびサンプリングされた輪郭点にあてはめられるパラメトリック形状からサンプリングすることができる。

【0167】

複数キュー追跡処理に対して行うことができるもう1つの修正が、顔の複数の可能な位置の追跡、言い換えると、1つの仮説ではなく複数の仮説の追跡である。パーティクルフィ

50

ルタリング技法を使用して、複数の仮説を維持し、その結果、弱い仮説が即座に除去されないようにすることができる。そうではなく、弱い仮説が、維持され、よい選択であることを証明する時間を与えられる。次は、アンセンテッドパーティクルフィルタと称する、そのようなパーティクルフィルタ技法の1つである。

【 0 1 6 8 】

アンセンテッドカルマンフィルタ(unscented Kalman filter、UKF)を使用するアンセンテッドパーティクルフィルタ(UPF)が、複数キュー追跡モジュール144によって、複数の仮説の追跡に使用される。アンセンテッド変換(unscented transformation、UT)が、 $g()$ のテイラー級数展開の2次(Gaussian priorの3次)までの平均および共分散を計算するのに使用される。 n_x が、 x の次元、

10

【 0 1 6 9 】

【数 3 0】

$\bar{x}$

【 0 1 7 0 】

が、 x の平均、 P_x が、 x の共分散であるものとして、UTは、 $y=g(x)$ の平均および共分散を次のように計算する。

【 0 1 7 1 】

まず、 $2n_x+1$ 個のシグマ点 $S_i=\{X_i, W_i\}$ を、決定論的に生成する。

【 0 1 7 2 】

20

【数 3 1】

$$X_0 = \bar{x}$$

$$X_i = \bar{x} + (\sqrt{(n_x + \lambda)P_x})_i \quad i = 1, \dots, n_x$$

$$X_i = \bar{x} - (\sqrt{(n_x + \lambda)P_x})_i \quad i = n_x + 1, \dots, 2n_x$$

$$W_0^{(m)} = \lambda / (n_x + \lambda), \quad W_0^{(c)} = W_0^{(m)} + (1 - \alpha^2 + \beta)$$

$$W_i^{(m)} = W_i^{(c)} = 1 / (2 \cdot (n_x + \lambda)) \quad i = 1, \dots, 2n_x$$

30

$$\lambda = \alpha^2(n_x + \kappa) - n_x$$

(7)

【 0 1 7 3 】

ここで、 κ は、シグマ点と平均

【 0 1 7 4 】

【数 3 2】

$\bar{x}$

【 0 1 7 5 】

の間の距離を制御するスケーリングパラメータ、 α は、非線形関数 $g()$ から生じる高次の影響を制御する正のスケーリングパラメータ、 β は、0番目のシグマ点の重みづけを制御するパラメータ、

40

【 0 1 7 6 】

【数 3 3】

$$(\sqrt{(n_x + \lambda)P_x})_i$$

【 0 1 7 7 】

は、行列の平方根の i 番目の列である。一実施形態では、スカラーの場合に、 $\alpha=1$ 、 $\beta=0$ 、および $\kappa=2$ である。0番目のシグマ点の重みが、平均と共分散を計算する場合に異なるこ

50

とに留意されたい。

【 0 1 7 8 】

次に、シグマ点を、非線形変換を介して伝搬させる。

$$Y_i = g(X_i) \quad i=0, \dots, 2n_x \quad (8)$$

yの平均および共分散は、次のように計算される。

【 0 1 7 9 】

【 数 3 4 】

$$\bar{y} = \sum_{i=0}^{2n_x} W_i^{(m)} Y_i, \quad P_y = \sum_{i=0}^{2n_x} W_i^{(c)} (Y_i - \bar{y})(Y_i - \bar{y})^T \quad (9)$$

10

【 0 1 8 0 】

このyの平均および共分散は、テイラー級数展開の2次まで正確である。

アンセンテッドカルマンフィルタ(UKF)は、雑音成分

【 0 1 8 1 】

【 数 3 5 】

$$x_t^a = [x_t^T m_t^T n_t^T]^T$$

【 0 1 8 2 】

を含むように状態空間を拡張することによって、UTを使用して実施することができる。 $N_a = N_x + N_m + N_n$ が、拡張された状態空間の次元であるものとし、 N_m および N_n が、雑音 m_t および n_t の次元、 Q および R が、雑音 m_t および n_t の共分散であるものとして、UKFを、次のように要約することができる。

【 0 1 8 3 】

初期化

【 0 1 8 4 】

【 数 3 6 】

$$\bar{x}_0^a = [\bar{x}_0^T 0 0]^T, \quad P_0^a = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{bmatrix} \quad (10)$$

30

【 0 1 8 5 】

時刻のインスタンスtのそれぞれについて、下記を繰り返す。

a) 上の計算7の手順を使用して、シグマ点を計算する。

【 0 1 8 6 】

【 数 3 7 】

$$X_{t-1}^a = [\bar{x}_{t-1}^a \quad \bar{x}_{t-1}^a \pm \sqrt{(n_a + \lambda) P_{t-1}^a}] \quad (11)$$

40

【 0 1 8 7 】

b) 時刻更新

【 0 1 8 8 】

【 数 3 8 】

$$X_{t|t-1}^x = f(X_{t-1}^x, X_{t-1}^v), \quad \bar{x}_{t|t-1} = \sum_{i=0}^{2n_a} W_i^{(m)} X_{i,t|t-1}^x \quad (12)$$

$$Y_{t|t-1} = h(X_{t|t-1}^x, X_{t-1}^n), \quad \bar{y}_{t|t-1} = \sum_{i=0}^{2n_a} W_i^{(m)} Y_{i,t|t-1}^x \quad (13)$$

$$P_{t|t-1} = \sum_{i=0}^{2n_a} W_i^{(c)} \left[X_{i,t|t-1}^x - \bar{x}_{t|t-1} \right] \left[X_{i,t|t-1}^x - \bar{x}_{t|t-1} \right]^T \quad (14)$$

【 0 1 8 9 】

10

c) 測定値更新

【 0 1 9 0 】

【 数 3 9 】

$$P_{y_t, y_t} = \sum_{i=0}^{2n_a} W_i^{(c)} \left[Y_{i,t|t-1} - \bar{y}_{t|t-1} \right] \left[Y_{i,t|t-1} - \bar{y}_{t|t-1} \right]^T \quad (15)$$

$$P_{x_t, y_t} = \sum_{i=0}^{2n_a} W_i^{(c)} \left[X_{i,t|t-1}^x - \bar{x}_{t|t-1} \right] \left[Y_{i,t|t-1} - \bar{y}_{t|t-1} \right]^T \quad (16)$$

$$K_t = P_{x_t, y_t} P_{y_t, y_t}^{-1} \quad (17)$$

20

$$\bar{x}_t = \bar{x}_{t|t-1} + K_t (y_t - \bar{y}_{t|t-1}), \quad P_t = P_{t|t-1} - K_t P_{y_t, y_t} K_t^T \quad (18)$$

【 0 1 9 1 】

UKFを用いると、最も最近の観測を、状態推定に簡単に組み込むことができる(例えば、上の測定値更新c))が、UKFでは、状態分布が正規分布であることが前提にされる。その一方で、パーティクルフィルタでは、任意の分布をモデル化することができるが、提案分布に新しい観測 y_t を組み込むことが困難である。UKFは、パーティクルフィルタの提案分布を生成するのに使用され、ハイブリッドUPFがもたらされる。具体的に言うと、各パーティクルの提案分布は、次のようになる。

【 0 1 9 2 】

30

【 数 4 0 】

$$q(x_t^{(i)} | x_{0:t-1}^{(i)}, y_{1:t}) = N(\bar{x}_t^{(i)}, P_t^{(i)}), \quad i = 1, \dots, N \quad (19)$$

【 0 1 9 3 】

ここで、

【 0 1 9 4 】

【 数 4 1 】

$\bar{x}_t$

40

【 0 1 9 5 】

および P_t は、UKFを使用して(計算(10)-(18))計算される、 x の平均および共分散である。事後分布 $p(x_t | x_{t-1}, y_{0:t})$ を近似するのに正規分布の前提が非現実的である場合であっても、別個の

【 0 1 9 6 】

【 数 4 2 】

$\bar{x}_t$

【 0 1 9 7 】

および P_t を用いて個々のパーティクルを生成するより問題が少ないことに留意されたい。

50

さらに、UKFは、後の平均および共分散を2次まで近似するので、システムの非線形性が、よく保存される。UPF処理は、UKFステップおよび計算(19)を、汎用パーティクルフィルタアルゴリズムに組み込むことによって、簡単に得られる。

【 0 1 9 8 】

図15は、例示的なUPF処理550を示す流れ図である。図15の処理は、図3の複数キュー追跡モジュール144によって実行され、ソフトウェアで実行することができる。

【 0 1 9 9 】

最初に、パーティクル

【 0 2 0 0 】

【 数 4 3 】

10

$$x_i^{(i)}, i = 1, \dots, N$$

【 0 2 0 1 】

を、計算(11)-(18)を使用してUKFを用いて更新して、

【 0 2 0 2 】

【 数 4 4 】

$$\bar{x}_i^{(i)}$$

【 0 2 0 3 】

および

20

【 0 2 0 4 】

【 数 4 5 】

$$P_i^{(i)}$$

【 0 2 0 5 】

を得る(動作552)。次に、パーティクル

【 0 2 0 6 】

【 数 4 6 】

$$x_i^{(i)}, i = 1, \dots, N$$

30

【 0 2 0 7 】

を、提案分布

【 0 2 0 8 】

【 数 4 7 】

$$q(x_t^{(i)} | x_{0:t-1}^{(i)}, y_{1:t}) = N(\bar{x}_t^{(i)}, P_t^{(i)})$$

【 0 2 0 9 】

からサンプリングする(動作554)。その後、次の計算(20)を使用して、パーティクルの重みを計算する(動作556)。

40

【 0 2 1 0 】

【 数 4 8 】

$$\begin{aligned}
\tilde{w}_t^{(i)} &= \frac{p(y_{1:t}|x_{0:t}^{(i)})p(x_{0:t}^{(i)})}{q(x_{0:t-1}^{(i)}|y_{1:t-1})q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})} \\
&= \tilde{w}_{t-1}^{(i)} \frac{p(y_{1:t}|x_{0:t}^{(i)})p(x_{0:t}^{(i)})}{p(y_{1:t-1}|x_{0:t-1}^{(i)})p(x_{0:t-1}^{(i)})q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})} \\
&= \tilde{w}_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})}
\end{aligned} \tag{20}$$

10

【 0 2 1 1 】

その後、次の式(21)を使用して、重要性の重みを正規化する(動作558)。

【 0 2 1 2 】

【数 4 9 】

$$w_t(x_{0:t}^{(i)}) = \tilde{w}_t(x_{0:t}^{(i)}) / \sum_{i=1}^N \tilde{w}_t(x_{0:t}^{(i)}) \tag{21}$$

【 0 2 1 3 】

ここで、パーティクル $\{x_{0:t}^{(i)}, w_t(x_{0:t}^{(i)})\}$ が、既知の分布 q から引き出され、

【 0 2 1 4 】

20

【数 5 0 】

$$\tilde{w}_t(x_{0:t}^{(i)})$$

【 0 2 1 5 】

および

【 0 2 1 6 】

【数 5 1 】

$$w_t(x_{0:t}^{(i)})$$

【 0 2 1 7 】

30

は、正規化されていない重要性の重みおよび正規化された重要性の重みである。

有効パーティクルサイズ S を、次の計算(22)を使用して判定する(動作560)。

【 0 2 1 8 】

【数 5 2 】

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \frac{p(y_t|x_t^{(i)})p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, y_{1:t})} = \tilde{w}_{t-1}^{(i)} p(y_t|x_t^{(i)}) \tag{22}$$

【 0 2 1 9 】

$S < S_T$ の場合には、重みつきパーティクルを掛け(または抑制し)て、 N 個の等しい重みのパーティクルを生成する(動作562)。その後、次の計算(23)を使用して、 $g(\cdot)$ の期待値を計算する(動作564)。

40

【 0 2 2 0 】

【数 5 3 】

$$E_p(g(x_{0:t})) = \lim_{N \rightarrow \infty} \sum_{i=1}^N g(x_{0:t}^{(i)}) w_t(x_{0:t}^{(i)}) \tag{23}$$

【 0 2 2 1 】

x_t の条件付き平均を、 $g_t(x_t) = x_t$ を用いて計算することができ、 x_t の条件付き共分散を、

【 0 2 2 2 】

【数 5 4 】

50

$$g_i(x_i) = x_i x_i^T$$

【 0 2 2 3 】

を用いて計算することができる。

【 0 2 2 4 】

オーディオに基づく参加者の追跡への図15のUPF処理550の使用を、ここで説明する。2つのマイクロホンが、通常は、水平にパンする角度を推定するのに十分である。水平にパンする角度に基づく追跡を、ここで説明するが、話者の垂直にチルトする角度に基づいて追跡するのに、類似する動作を実行することができる。図16に、例示的な複数マイクロホン環境を示す。図16では、2つのマイクロホンが、位置AおよびBに配置され、音源が、位置Cに配置されると仮定する。音源の距離(すなわち|OC|)が、マイクロホン対のベースライン|AB|よりはるかに長い時に、パン角度 $\theta = \angle COX$ を、次のように推定することができる。

10

【 0 2 2 5 】

【数 5 5 】

$$\theta = \angle COX \approx \angle BAE = \arcsin \frac{|BE|}{|AB|} = \arcsin \frac{D \times v}{|AB|} \quad (24)$$

【 0 2 2 6 】

ここで、Dは、2つのマイクロホンの間の時間遅延、 $v=342$ m/sは、空气中を伝わる音の速度である。

20

【 0 2 2 7 】

追跡応用例でUPFフレームワークを使用するために、4つのエンティティを最初に確立する、すなわち、計算(12)で使用されるシステム動力学 $x_t = f(x_{t-1}, m_{t-1})$ 、計算(13)で使用されるシステム観測 $y_t = h(x_t, n_t)$ 、計算(22)で使用される尤度 $p(y_t | x_t)$ 、および計算(18)で使用されるイノベーション

【 0 2 2 8 】

【数 5 6 】

$$y_t - \bar{y}_{t|t-1}$$

30

【 0 2 2 9 】

である。この4つのエンティティを確立した後に、追跡は、図15のUPF処理550を使用して簡単に進む。

【 0 2 3 0 】

システム動力学モデル $x_t = f(x_{t-1}, m_{t-1})$ は、次のように決定される。

【 0 2 3 1 】

【数 5 7 】

$$x = [\theta, \dot{\theta}]^T$$

40

【 0 2 3 2 】

が、状態空間であるものとし、これらが、パン角度およびパン角度の速度であるものとする。話している人の動きの動力学をモデル化するために、周知のランジュバン過程 $d^2 \theta / dt^2 + \beta_\theta \cdot d\theta / dt = m$ を使用する。この離散形は次の通りである。

【 0 2 3 3 】

【数 5 8 】

$$\begin{bmatrix} \theta_t \\ \dot{\theta}_t \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ 0 & a \end{bmatrix} \begin{bmatrix} \theta_{t-1} \\ \dot{\theta}_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix} m_t$$

$$a = \exp(-\beta_\theta \tau), \quad b = \bar{v} \sqrt{1-a^2} \quad (25)$$

【 0 2 3 4 】

ここで、 β_θ は、速度定数、 m は、 $N(0, Q)$ から引き出される熱励起過程、 τ は、打切り時間ステップ、

【 0 2 3 5 】

10

【 数 5 9 】

$\bar{v}$

【 0 2 3 6 】

は、定常状態二乗平均平方根速度である。

【 0 2 3 7 】

この系の観測モデル $y_t = h(x_t, n_t)$ は、次のように決定される。系の観測 y_t は、時間遅延 D_t である。上の計算(24)に基づいて、観測は、

$$y_t = D_t = h(\theta_t, n_t) = |AB|v \sin \theta_t + n_t \quad (26)$$

によって状態に関係し、ここで、 n_t は、 $N(0, R)$ の正規分布に従う測定値雑音である。

20

【 0 2 3 8 】

尤度モデル $p(y_t | x_t)$ は、次のように決定される。 J が、GCCF(一般化相互相関関数)内のピークの数であるものとする。 J 個のピーク位置のうち、せいぜい1つが、真の音源からのものである。したがって、 $J+1$ 個の仮説を定義する。

$$H_0 = \{c_j = C : j = 1, \dots, J\}$$

$$H_j = \{c_j = T, c_k = C : k = 1, \dots, J, k \neq j\} \quad (27)$$

ここで、 $c_j = T$ は、 j 番目のピークが、真の音源に関連することを意味し、そうでない場合には $c_j = C$ である。したがって、仮説 H_0 は、ピークのどれもが真の源に関連しないことを意味する。したがって、組み合わされた尤度モデルは、次のようになる。

【 0 2 3 9 】

30

【 数 6 0 】

$$\begin{aligned} p(y_t | x_t) &= \pi_0 p(y_t | H_0) + \sum_{j=1}^J \pi_j p(y_t | H_j) \\ &= \pi_0 U + N_m \sum_{j=1}^J \pi_j N(D_j, \sigma_D) \\ \text{s.t. } \pi_0 + \sum_{j=1}^J \pi_j &= 1 \end{aligned} \quad (28)$$

【 0 2 4 0 】

ここで、 π_0 は、仮説 H_0 の前の確率であり、 $\pi_j, j=1, \dots, J$ は、 j 番目のピークの相対的な高さから得ることができ、 N_m は、正規化係数であり、 D_j は、 j 番目のピークに対応する時間遅延であり、 U は、一様分布を表し、 $N()$ は、正規分布を表す。

40

【 0 2 4 1 】

イノベーションモデル

【 0 2 4 2 】

【 数 6 1 】

$y_t - \bar{y}_{t-1}$

【 0 2 4 3 】

は、次のように決定される。尤度モデルと同様に、イノベーションモデルも、複数ピーク

50

の事実を考慮する必要がある。

【 0 2 4 4 】

【 数 6 2 】

$$y_t - \bar{y}_{t|t-1} = \sum_{j=1}^J \pi_j (D_j - \bar{y}_{t|t-1}) \quad (29)$$

【 0 2 4 5 】

ここで、

【 0 2 4 6 】

【 数 6 3 】

$$\bar{y}_{t|t-1}$$

10

【 0 2 4 7 】

は、UKF(上の計算(18)を参照)から得られる予測された測定値である。

【 0 2 4 8 】

視覚データに基づく参加者の追跡に図15のUPF処理550を使用することは、オーディオデータに基づく参加者の追跡に似ている。追跡応用例でUPFフレームワークを使用するために、まず、4つのエンティティを確立する、すなわち、システム動力学モデル $x_t = f(x_{t-1}, m_{t-1})$ 、システム観測モデル $y_t = h(x_t, n_t)$ 、尤度モデル $p(y_t | x_t)$ 、およびイノベーションモデル

20

【 0 2 4 9 】

【 数 6 4 】

$$y_t - \bar{y}_{t|t-1}$$

【 0 2 5 0 】

である。これらの4つのエンティティを確立した後に、追跡は、図15のUPF処理550を使用して簡単に進む。

【 0 2 5 1 】

システム動力学モデル $x_t = f(x_{t-1}, m_{t-1})$ は、次のように決定される。 (r, s) が、画像座標を表すものとする。輪郭ベースの追跡では、システム状態が、楕円の中心の位置、その水平速度、および垂直速度すなわち

30

【 0 2 5 2 】

【 数 6 5 】

$$x_t = [r_t, s_t, \dot{r}_t, \dot{s}_t]^T$$

【 0 2 5 3 】

である。可聴データのシステム動力学モデルと同様に、周知のランジュバン過程を採用して、人間の動きの動力学をモデル化する。

【 0 2 5 4 】

【 数 6 6 】

$$\begin{bmatrix} r_t \\ s_t \\ \dot{r}_t \\ \dot{s}_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & \tau & 0 \\ 0 & 1 & 0 & \tau \\ 0 & 0 & a_r & 0 \\ 0 & 0 & 0 & a_s \end{bmatrix} \begin{bmatrix} r_{t-1} \\ s_{t-1} \\ \dot{r}_{t-1} \\ \dot{s}_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ b_r \\ b_s \end{bmatrix} m_t \quad (30)$$

40

【 0 2 5 5 】

システム観測モデル $y_t = h(x_t, n_t)$ は、次のように決定される。楕円が、現在の状態位置 $(r_t$

50

, s_t)でセンタリングされる。K本の放射線が、楕円の中心から楕円の境界に交わるように生成される。楕円の中心は、ローカル座標系の原点として使用され、したがって、交差(u_k, v_k), $k=1, 2, \dots, K$ を、

【 0 2 5 6 】

【 数 6 7 】

$$u_k = \sqrt{\tan^2 \varphi_k / (1.44 \tan^2 \varphi_k + 1)}$$

$$v_k = \sqrt{1 / (1.44 \tan^2 \varphi_k + 1)}$$

(31)

10

【 0 2 5 7 】

として得ることができ、楕円の式と放射線の式を共同で解くことによって、

【 0 2 5 8 】

【 数 6 8 】

$$\begin{cases} \frac{u_k^2}{1} + \frac{v_k^2}{1.2^2} = 1 \\ u_k = v_k \tan(\varphi_k) \end{cases}$$

(32)

20

【 0 2 5 9 】

を得ることができる。

【 0 2 6 0 】

ローカル(u, v)座標を画像座標に逆変換すると、次の観測が得られる。

$$y_t = h(x_t, n_t)$$

$$= [(u_k + r_t, v_k + s_t)] + n_t, k=1, 2, \dots, K \quad (33)$$

ここで、 n_t は、 $N(0, R)$ の正規分布に従う測定値雑音である。観測モデルが、非常に非線形であることに留意されたい。

【 0 2 6 1 】

30

尤度モデル $p(y_t | x_t)$ は、次のように決定される。エッジ輝度を使用して、状態尤度をモデル化する。K本の放射線のそれぞれに沿って、周知のキャニイエッジ検出器を使用して、エッジ輝度を計算する。結果の関数は、可聴データに関する尤度モデルのGCCFに似た複数ピーク関数である。複数のピークは、この放射線に沿って複数のエッジ候補があることを意味する。ピークの数 J であるものとして、可聴データの尤度モデルで開発したものと同一の尤度モデルを使用して、放射線 k に沿ったエッジ尤度をモデル化することができる。

【 0 2 6 2 】

【 数 6 9 】

$$p^{(k)}(y_t | x_t) = \pi_{k0} p^{(k)}(y_t | H_0) + \sum_{j=1}^J \pi_{kj} p^{(k)}(y_t | H_j)$$

$$= \pi_{k0} U + N_m \sum_{j=1}^J \pi_{kj} N((u_k, v_k)_j, \sigma_{kj})$$

40

【 0 2 6 3 】

したがって、K本の放射線のすべてを考慮に入れた総合的な尤度は、

【 0 2 6 4 】

【 数 7 0 】

$$p(y_t|x_t) = \prod_{k=1}^K p^{(k)}(y_t|x_t)$$

【 0 2 6 5 】

になる。

【 0 2 6 6 】

イノベーションモデル

【 0 2 6 7 】

【 数 7 1 】

$$y_t - \bar{y}_{t-1}$$

10

【 0 2 6 8 】

は、次のように決定される。尤度モデルと同様に、イノベーションモデルも、複数ピークの事実を考慮する必要がある。

【 0 2 6 9 】

【 数 7 2 】

$$y^{(k)}_t - \bar{y}^{(k)}_{t-1} = \sum_{j=1}^J \pi_{kj} \left((u_k, v_k)_{t,j} - (u_k, v_k)_{t-1} \right)$$

【 0 2 7 0 】

20

ここで、 $k=1, 2, \dots, K$, π_{kj} は、放射線 k に沿った j 番目のピークの混合重みであり、対応するエッジ輝度から得ることができる。

【 0 2 7 1 】

全般的なコンピュータ環境

図17に、本明細書に記載の複数の個人の自動検出および追跡を実施するのに使用することができる全般的なコンピュータ環境600を示す。コンピュータ環境600は、コンピューティング環境の1例に過ぎず、コンピュータアーキテクチャおよびネットワークアーキテクチャの使用または機能性の範囲に関する制限を暗示することは意図されていない。コンピュータ環境600を、例示的なコンピュータ環境600に示された構成要素のいずれかまたはその組合せに関する依存性または要件を有するものと解釈してもならない。

30

【 0 2 7 2 】

コンピュータ環境600には、コンピュータ602の形の汎用コンピューティング装置が含まれる。コンピュータ602は、例えば、図1のシステム102、図2の112、図3のシステム130などとしてすることができる。コンピュータ602の構成要素には、1つまたは複数のプロセッサまたは処理ユニット604、システムメモリ606、および、プロセッサ604を含むさまざまなシステム構成要素をシステムメモリ606に結合するシステムバス608を含めることができるが、これに制限はされない。

【 0 2 7 3 】

システムバス608は、メモリバスまたはメモリコントローラ、周辺バス、accelerated graphics port、および、さまざまなバスアーキテクチャのいずれかを使用するプロセッサバスまたはローカルバスを含む、さまざまなタイプのバス構造のいずれかの1つまたは複数を表す。例として、そのようなアーキテクチャに、Industry Standard Architecture(ISA)バス、Micro Channel Architecture(MCA)バス、Enhanced ISA(EISA)バス、Video Electronics Standards Association(VESA)ローカルバス、および、メザニンバスとも称するPeripheral Component Interconnects(PCI)バスを含めることができる。

40

【 0 2 7 4 】

コンピュータ602に、通常は、さまざまなコンピュータ可読媒体が含まれる。そのような媒体は、コンピュータ602によってアクセス可能な入手可能な媒体とすることができ、これには、揮発性媒体と不揮発性媒体、取外し可能媒体と取外し不能媒体の両方が含まれる。

50

【 0 2 7 5 】

システムメモリ606には、ランダムアクセスメモリ(RAM)610などの揮発性メモリ、および/または読取専用メモリ(ROM)612などの不揮発性メモリの形のコンピュータ可読媒体が含まれる。起動中などに、コンピュータ602内の要素の間で情報を転送するのを助ける基本ルーチンを含む基本入出力システム(BIOS)614が、ROM612に保管される。RAM610には、通常は、プロセッサユニット604によって即座にアクセス可能および/または現在操作されているデータおよび/またはプログラムモジュールが含まれる。

【 0 2 7 6 】

コンピュータ602には、他の取外し可能/取外し不能、揮発性/不揮発性のコンピュータ記憶媒体も含めることができる。例として、図17に、取外し不能不揮発性磁気媒体(図示せず)から読み取り、これに書き込むハードディスクドライブ616、取外し可能不揮発性磁気ディスク620(例えば、「フロッピー(登録商標)ディスク」)から読み取り、これに書き込む磁気ディスクドライブ618、および、CD-ROM、DVD-ROM、または他の光学媒体などの取外し可能不揮発性光ディスク624から読み取り、かつ/または書き込む光ディスクドライブ622を示す。ハードディスクドライブ616、磁気ディスクドライブ618、および光ディスクドライブ622は、それぞれ、1つまたは複数のデータ媒体インターフェース626によってシステムバス608に接続される。その代わりに、ハードディスクドライブ616、磁気ディスクドライブ618、および光ディスクドライブ622を、1つまたは複数のインターフェース(図示せず)によってシステムバス608に接続することができる。

【 0 2 7 7 】

ディスクドライブおよびそれに関連するコンピュータ可読媒体は、コンピュータ可読命令、データ構造、プログラムモジュール、およびコンピュータ602用の他のデータの揮発性ストレージを提供する。この例には、ハードディスク616、取外し可能磁気ディスク620、および取外し可能光ディスク624が示されているが、磁気カセットまたは他の磁気記憶装置、フラッシュメモリカード、CD-ROM、デジタル多用途ディスク(DVD)または他の光記憶装置、ランダムアクセスメモリ(RAM)、読取専用メモリ(ROM)、電氣的消去可能プログラマブル読取専用メモリ(EEPROM)、および類似物など、コンピュータによってアクセス可能なデータを保管できる他のタイプのコンピュータ可読媒体も、例示的なコンピューティングシステムおよびコンピューティング環境の実施に使用することができることを理解されたい。

【 0 2 7 8 】

例えばオペレーティングシステム627、1つまたは複数のアプリケーションプログラム628、他のプログラムモジュール630、およびプログラムデータ632を含む、任意の数のプログラムモジュールを、ハードディスク616、磁気ディスク620、光ディスク624、ROM612、および/またはRAM610に保管することができる。そのようなオペレーティングシステム627、1つまたは複数のアプリケーションプログラム628、他のプログラムモジュール630、およびプログラムデータ632(またはその組合せ)のそれぞれによって、分散ファイルシステムをサポートする常駐構成要素のすべてまたは一部を実施することができる。

【 0 2 7 9 】

ユーザは、キーボード634およびポインティングデバイス636(例えば「マウス」)などの入力装置を介して、コンピュータ602にコマンドおよび情報を入力することができる。他の入力装置638(具体的には図示せず)に、マイクロホン、ジョイスティック、ゲームパッド、衛星放送用パラボラアンテナ、シリアルポート、スキャナ、および/または類似物を含めることができる。これらおよび他の入力装置は、システムバス608に結合される入出力インターフェース640を介して処理ユニット604に接続されるが、パラレルポート、ゲームポート、またはuniversal serial bus(USB)などの他のインターフェースおよびバス構造によって接続することができる。

【 0 2 8 0 】

モニタ642または他のタイプの表示装置を、ビデオアダプタ644などのインターフェースを介してシステムバス608に接続することもできる。モニタ642に加えて、他の出力周辺装置

10

20

30

40

50

に、入出力インターフェース640を介してコンピュータ602に接続することができる、スピーカ(図示せず)およびプリンタ646などの構成要素を含めることができる。

【0281】

コンピュータ602は、リモートコンピューティング装置648などの1つまたは複数のリモートコンピュータへの論理接続を使用して、ネットワーク化された環境で動作することができる。例として、リモートコンピューティング装置648を、パーソナルコンピュータ、ポータブルコンピュータ、サーバ、ルータ、ネットワークコンピュータ、ピアデバイス、または他の一般的なネットワークノード、および類似物とすることができる。リモートコンピューティング装置648は、コンピュータ602に関して本明細書で説明した要素および特徴の多くまたはすべてを含めることができるポータブルコンピュータとして図示されている。

10

【0282】

コンピュータ602とリモートコンピュータ648の間の論理接続は、ローカルエリアネットワーク(LAN)650および一般的な広域ネットワーク(WAN)652として図示されている。そのようなネットワーキング環境は、オフィス、会社全体のコンピュータネットワーク、イントラネット、およびインターネットでありふれたものである。

【0283】

LANネットワーキング環境で実施される時に、コンピュータ602は、ネットワークインターフェースまたはネットワークアダプタ654を介してローカルネットワーク650に接続される。WANネットワーキング環境で実施される時に、コンピュータ602に、通常は、広域ネットワーク652を介する通信を確立するための、モデム656または他の手段が含まれる。モデム656は、コンピュータ602の内蔵または外付けとすることができるが、入出力インターフェース640または他の適当な機構を介してシステムバス608に接続することができる。図示されたネットワーク接続は例示的であること、ならびにコンピュータ602および648の間で通信リンクを確立する他の手段を使用できることを理解されたい。

20

【0284】

コンピューティング環境600と共に図示されたものなどのネットワーク化された環境では、コンピュータ602に関して図示されたプログラムモジュールまたはその一部を、リモートメモリストレージデバイスに保管することができる。例として、リモートアプリケーションプログラム658が、リモートコンピュータ648の記憶装置に常駐する。例示のために、アプリケーションプログラムおよび、オペレーティングシステムなどの他の実行可能プログラム構成要素が、この図では別個のブロックとして図示されているが、そのようなプログラムおよび構成要素が、さまざまな時にコンピューティング装置602の異なるストレージ構成要素に常駐し、コンピュータのデータプロセッサによって実行されることを理解されたい。

30

【0285】

分散ファイルシステムの実施形態150を、プログラムモジュールなどの、1つまたは複数のコンピュータまたは他の装置によって実行される、コンピュータ実行可能命令の全般的な文脈で説明することができる。一般に、プログラムモジュールには、特定のタスクを実行するか特定の抽象データ型を実装する、ルーチン、プログラム、オブジェクト、コンポーネント、データ構造などが含まれる。通常、プログラムモジュールの機能性を、さまざまな実施形態で望みに応じて組み合わせるか分散することができる。

40

【0286】

暗号化されたファイルのファイルフォーマットの一実施形態を、何らかの形のコンピュータ可読媒体を介して保管または伝送することができる。コンピュータ可読媒体は、コンピュータによってアクセスできる、あらゆる使用可能な媒体とすることができる。制限ではなく例として、コンピュータ可読媒体に、「コンピュータ記憶媒体」および「通信媒体」を含めることができる。

【0287】

「コンピュータ記憶媒体」には、コンピュータ可読命令、データ構造、プログラムモジュ

50

ール、または他のデータなどの情報の記憶のためのあらゆる方法または技術で実施される、揮発性および不揮発性の、取外し可能および取外し不能の媒体が含まれる。コンピュータ記憶媒体には、RAM、ROM、EEPROM、フラッシュメモリまたは他のメモリ技術、CD-ROM、デジタル多用ディスク(DVD)または他の光学記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置または他の磁気記憶装置、または、所望の情報を保管するのに使用でき、コンピュータによってアクセスすることができる他の媒体が含まれるが、これに制限はされない。

【0288】

「通信媒体」は、通常は、搬送波または他のトランスポート機構など、変調されたデータ信号内でコンピュータ可読命令、データ構造、プログラムモジュール、または他のデータを実施する。通信媒体に、情報配送媒体も含まれる。用語「変調されたデータ信号」は、信号内に情報をエンコードする形でその特性の1つまたは複数を設定または変更される信号を意味する。制限ではなく例として、通信媒体に、有線ネットワークまたは直接配線接続などの有線媒体と、音響、RF、赤外線、および他の無線媒体などの無線媒体が含まれる。上記の任意の組合せも、コンピュータ可読媒体の範囲に含まれる。

【0289】

本明細書では、主に人間の顔に関して説明したが、本明細書に記載の人間の顔と同様に、他のオブジェクトを自動的に検出し、かつ/または追跡することができる。

【0290】

結論

上の説明では、構造的特徴および/または方法論的動作に固有の言語を使用した、請求項で定義される発明が、記載された特定の特徴または動作に制限されないことを理解されたい。そうではなく、特定の特徴および動作は、本発明の実施の例示的形態として開示される。

【図面の簡単な説明】

【図1】堅牢な自動化された識別および追跡を使用することができる例示的環境を示す図である。

【図2】堅牢な自動化された識別および追跡を使用することができるもう1つの例示的環境を示す図である。

【図3】堅牢な自動化された識別および追跡を使用する例示的システムを示す図である。

【図4】新しい顔領域の候補を検出する例示的処理を示す流れ図である。

【図5】動きベース初期化を使用して新しい顔領域の候補を識別する例示的処理を示す流れ図である。

【図6】ビデオ内容のフレームの例示的画像を示す図である。

【図7】階層検証を実行する例示的処理を示す流れ図である。

【図8】高速色ベース検証の例示的処理を示す図である。

【図9】複数キュー追跡を実行する例示的処理を示す流れ図である。

【図10】複数キュー追跡の例示的なモデル化および比較を追加的に詳細に示す図である。

【図11】領域滑らかさの概念を示す画像の図である。

【図12】図11の輝度の測定値を示す図である。

【図13】マッチングディスタンスの例示的な計算を線図的に示す図である。

【図14】あるフレームから次のフレームへのオブジェクトの例示的な追跡を示す図である。

【図15】例示的なアンセンテッドパーティクルフィルタ処理を示す流れ図である。

【図16】例示的な複数マイクロホン環境を示す図である。

【図17】例示的な一般的なコンピュータ環境を示す図である。

【符号の説明】

130 システム

132 検出および追跡モジュール

10

20

30

40

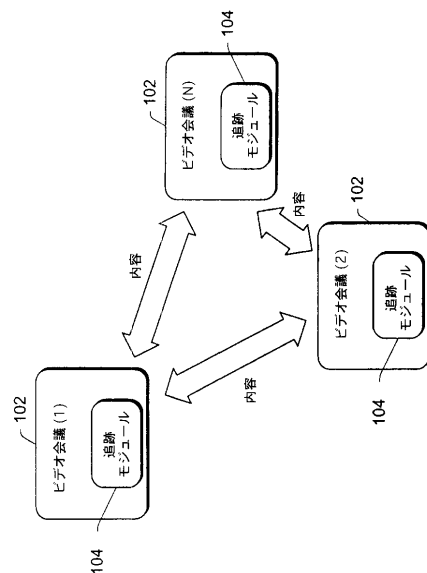
50

- 134 通信モジュール
- 136 オーディオ取込モジュール
- 138 ビデオ取込モジュール
- 140 自動初期化モジュール
- 142 階層検証モジュール
- 144 複数キュー追跡モジュール
- 146 顔/候補追跡リスト
- 148 中心座標
- 150 境界ボックス
- 152 追跡時間
- 154 最後の検証からの時間
- 156 動きベース初期化モジュール
- 158 オーディオベース初期化モジュール
- 160 高速顔検出モジュール
- 162 信頼性消失領域検出モジュール
- 164 高速色ベース検証モジュール
- 166 複数ビュー顔検出モジュール
- 168 観測尤度モジュール
- 170 滑らかさ制約モジュール
- 172 輪郭選択モジュール
- 174 モデル適応モジュール

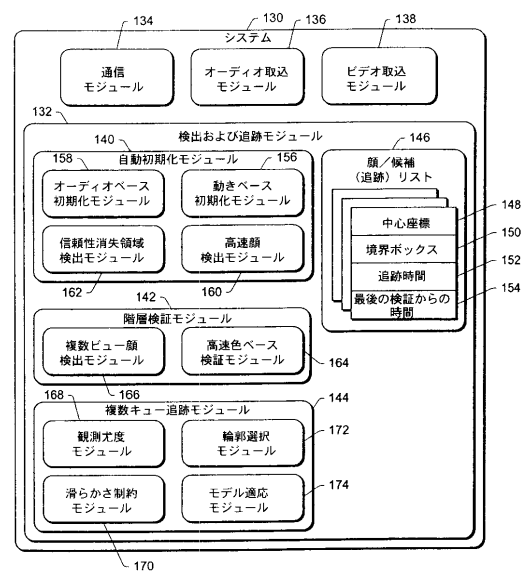
10

20

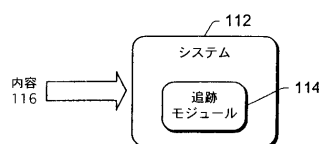
【図 1】



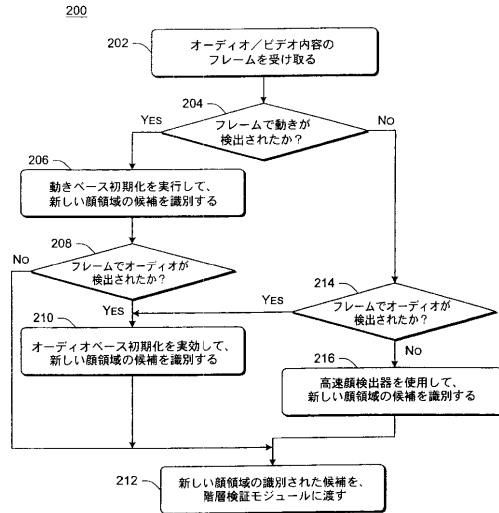
【図 3】



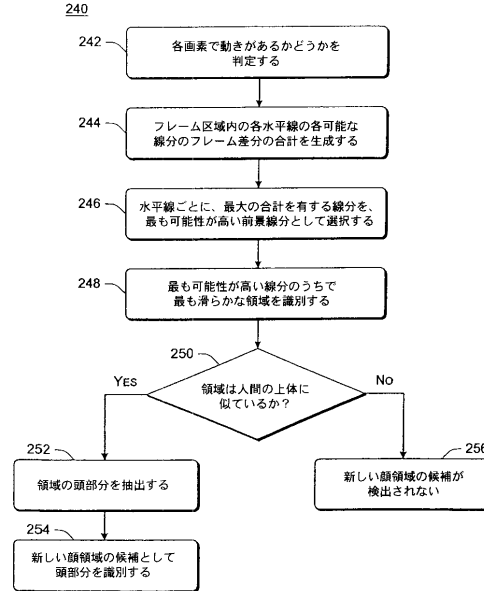
【図 2】



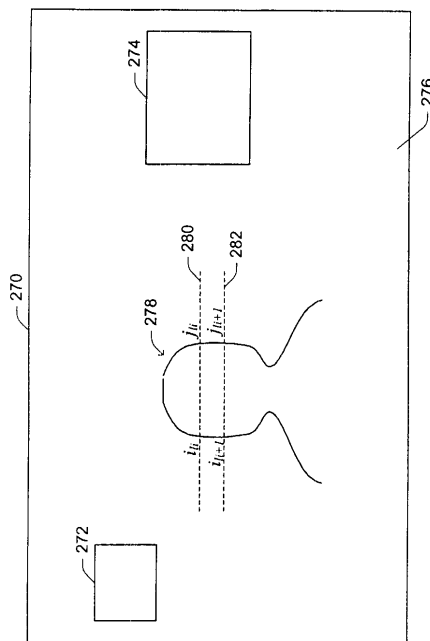
【図 4】



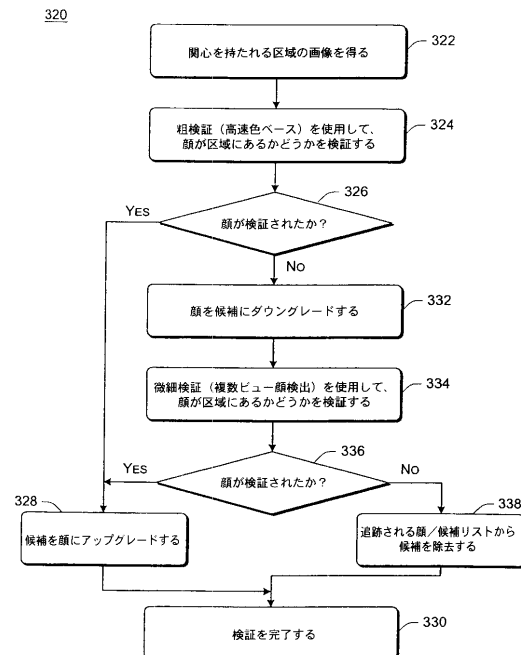
【図 5】



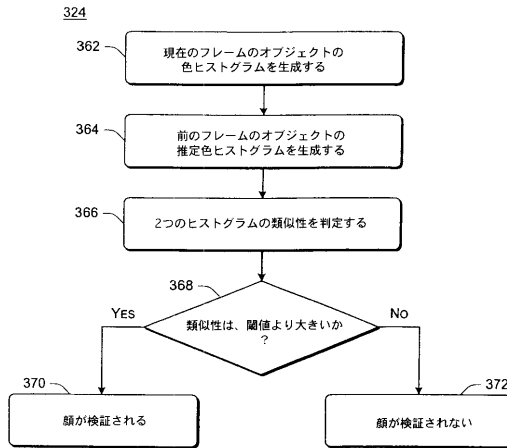
【図 6】



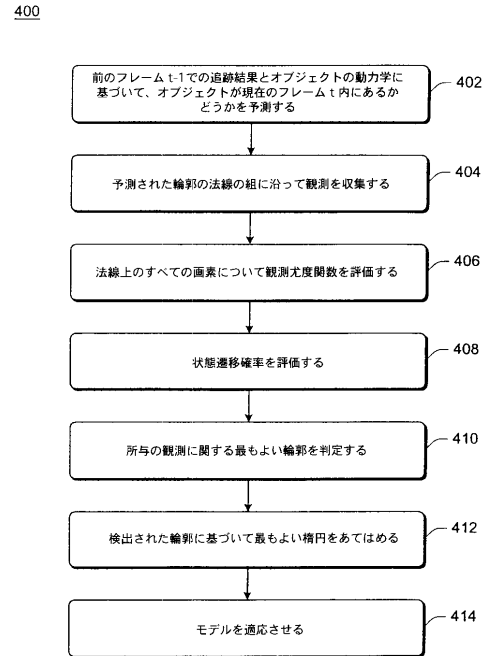
【図 7】



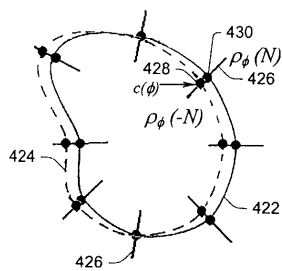
【図 8】



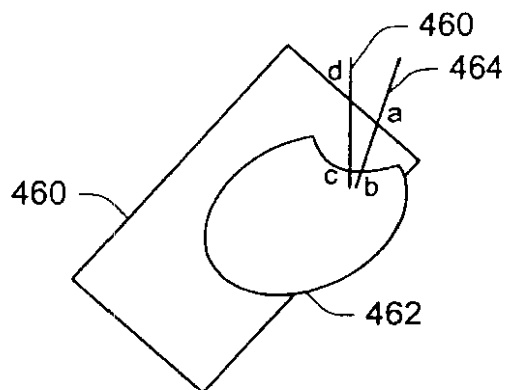
【図 9】



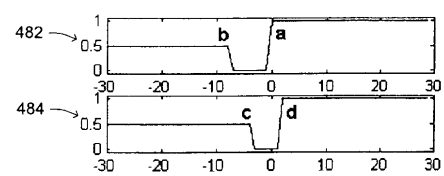
【図 10】



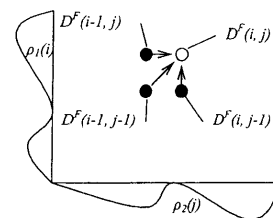
【図 11】



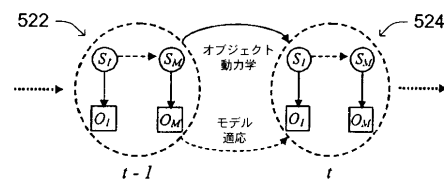
【図 12】



【図 13】

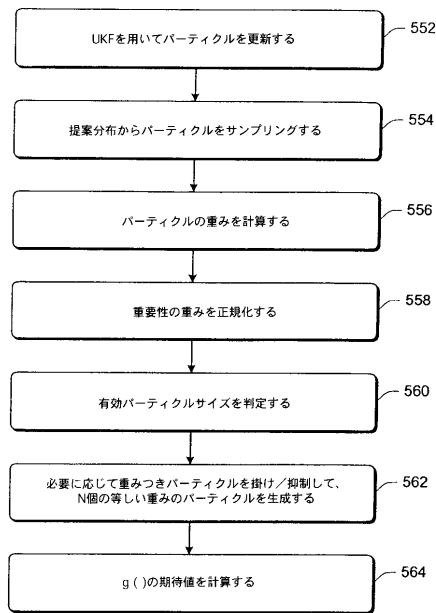


【図 14】

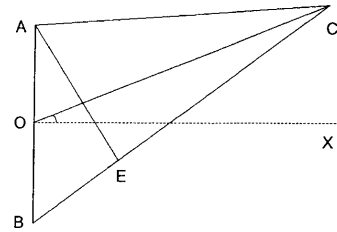


【図 15】

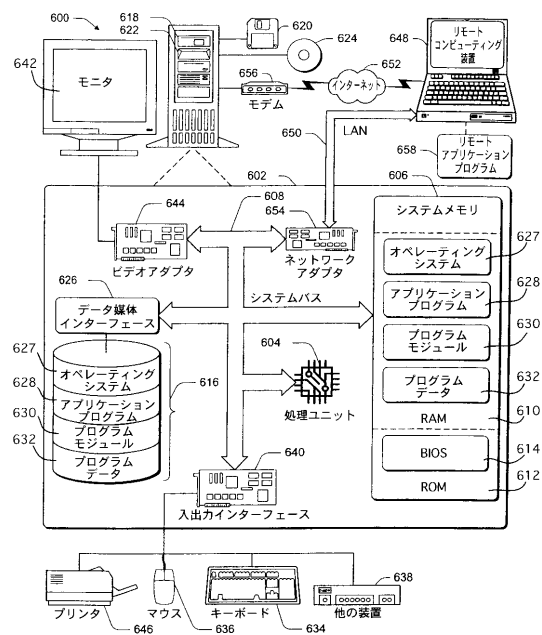
550



【図 16】



【図 17】



フロントページの続き

(51)Int.Cl.

F I

G 0 6 T 1/20 B

G 0 6 T 7/20 B

(72)発明者 チェン ユンキャン

アメリカ合衆国 6 1 8 0 1 イリノイ州 アーバナ ノース マッシューズ アベニュー 4 0 5
ピーアイ 2 3 2 3

審査官 松尾 俊介

(56)参考文献 国際公開第01/041064(WO,A1)

国際公開第00/018128(WO,A1)

特開2001-094857(JP,A)

特開2001-092974(JP,A)

特開平09-293185(JP,A)

特開2001-309225(JP,A)

米国特許第05805733(US,A)

特開平11-265455(JP,A)

Ingemar J.Cox,Sunita L.Hingorani,An Efficient Implementation of Reid's Multiple Hypothesis Tracking Algorithm and Its Evaluation for the Purpose of Visual Tracking,IEEE Transactions on Pattern Analysis and Machine Intelligence,米国,IEEE,1996年 2月28日,Vol.18,No.2, pp.138-150, <http://ieeexplore.ieee.org/iel1/34/10295/00481539.pdf>

(58)調査した分野(Int.Cl.,DB名)

G06T7/00~7/60

G06T1/00~1/20