

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2006-508601
(P2006-508601A)

(43) 公表日 平成18年3月9日(2006.3.9)

(51) Int.C1.	F 1	テーマコード (参考)
HO4N 5/232 (2006.01)	HO4N 5/232	Z 5B057
G06T 1/00 (2006.01)	G06T 1/00	34OA 5C053
HO4N 5/225 (2006.01)	HO4N 5/225	C 5C065
HO4N 5/91 (2006.01)	HO4N 5/225	F 5C122
HO4N 9/04 (2006.01)	HO4N 5/91	Z

審査請求 未請求 予備審査請求 未請求 (全 54 頁) 最終頁に続く

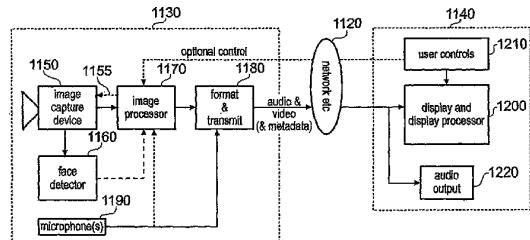
(21) 出願番号	特願2004-556510 (P2004-556510)	(71) 出願人	593081408 ソニー・ユナイテッド・キングダム・リミテッド Sony United Kingdom Limited
(86) (22) 出願日	平成15年11月28日 (2003.11.28)	(74) 代理人	100067736 弁理士 小池 晃
(85) 翻訳文提出日	平成17年8月1日 (2005.8.1)	(74) 代理人	100086335 弁理士 田村 榮一
(86) 國際出願番号	PCT/GB2003/005224	(74) 代理人	100096677 弁理士 伊賀 誠司
(87) 國際公開番号	W02004/051981		
(87) 國際公開日	平成16年6月17日 (2004.6.17)		
(31) 優先権主張番号	0227915.6		
(32) 優先日	平成14年11月29日 (2002.11.29)		
(33) 優先権主張國	英國 (GB)		
(81) 指定國	EP (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), JP, US		

最終頁に続く

(54) 【発明の名称】ビデオカメラ

(57) 【要約】

ビデオカメラ装置は、ビデオマテリアルを捕捉する撮像デバイスと、捕捉されたビデオマテリアルにおいて人間の顔を検出し、捕捉されたビデオマテリアルにおいて検出された顔の出現を特定する顔データを生成する顔検出器と、捕捉されたビデオマテリアルを表すデータを伝送及び/又は記録するデータ取扱媒体と、捕捉されたビデオマテリアル内の顔の検出に基づいて、データ取扱媒体によって送信又は保存されるデータを生成するプロセッサなどを備える。



【特許請求の範囲】**【請求項 1】**

ビデオマテリアルを捕捉する撮像デバイスと、
上記捕捉されたビデオマテリアルにおいて人間の顔を検出し、該捕捉されたビデオマテリアルにおいて検出された顔の出現を特定する顔データを生成する顔検出器と、

上記捕捉されたビデオマテリアルを表すデータを伝送及び／又は記録するデータ取扱媒体と、

上記捕捉されたビデオマテリアル内の顔の検出に基づいて、上記データ取扱媒体によって送信又は保存されるデータを生成するプロセッサとを備えるビデオカメラ装置。

【請求項 2】

上記データ取扱媒体は、

上記捕捉されたビデオマテリアルを保存するための記録媒体と、

上記顔検出器によって生成された顔データを含む、上記捕捉されたビデオマテリアルに関連するメタデータを保存するメタデータ保存手段とを備えることを特徴とする請求項 1 記載のビデオカメラ装置。

【請求項 3】

上記メタデータ保存手段は、上記捕捉されたビデオマテリアルと同じ記録媒体にメタデータを保存するよう構成されていることを特徴とする請求項 2 記載のビデオカメラ装置。

【請求項 4】

上記メタデータ保存手段は、当該ビデオカメラ装置に接続可能なリムーバブルストレージデバイスを備えることを特徴とする請求項 2 記載のビデオカメラ装置。

【請求項 5】

上記メタデータ保存手段は、当該ビデオカメラ装置の外部のストレージデバイスと、当該ビデオカメラ装置及び該ストレージデバイスとの間の無線リンクとを備えることを特徴とする請求項 2 記載のビデオカメラ装置。

【請求項 6】

当該ビデオカメラ装置は、フォーカス、ズーム及び／又は絞りの設定を調整可能なレンズを備え、上記メタデータは、該レンズのフォーカス、ズーム及び／又は絞りの設定に関する情報の少なくとも 1 つを含んでいることを特徴とする請求項 2 乃至 5 いずれか 1 項記載のビデオカメラ装置。

【請求項 7】

上記顔検出器は、上記レンズのフォーカス、ズーム及び／又は絞りの設定、及び／又は連続するビデオショットの開始点及び終了点に応じて動作することを特徴とする請求項 6 記載のビデオカメラ装置。

【請求項 8】

上記顔検出器は、上記捕捉されたビデオマテリアルの各フィールド又はフレーム内に人間の顔が存在している確率を検出し、

上記メタデータ保存手段は、上記捕捉されたビデオマテリアルの各連続するシーケンスから、該連続するシーケンスにおいて最も高い確率に関連付けられた、少なくとも 1 つの顔の表現を保存することを特徴とする請求項 2 乃至 7 いずれか 1 項記載のビデオカメラ装置。

【請求項 9】

当該ビデオカメラ装置は、単一の機器であることを特徴とする請求項 1 乃至 8 いずれか 1 項記載のビデオカメラ装置。

【請求項 10】

上記データ取扱媒体は、上記捕捉されたビデオマテリアルに関連する捕捉されたオーディオマテリアルを表すデータを保存する及び／又は伝送することを特徴とする請求項 1 乃至 9 いずれか 1 項記載のビデオカメラ装置。

【請求項 11】

音声検出器を備え、上記顔検出器は、捕捉されたオーディオマテリアルにおける音声の

10

20

30

40

50

検出に応じて動作することを特徴とする請求項 10 記載のビデオカメラ装置。

【請求項 12】

2つ以上のマイクロホンに接続され、上記プロセッサ及び／又は顔検出器は、該マイクロホンからのオーディオ信号に応じて動作し、現在の話者の顔を特定することを特徴とする請求項 10 又は 11 記載のビデオカメラ装置。

【請求項 13】

上記顔検出器による顔検出に応じて、上記データ取扱媒体によって保存及び／又は伝送される少なくとも幾つかの捕捉された画像のサブセットを導出するロジックを備える請求項 10 乃至 12 いずれか 1 項記載のビデオカメラ装置。

【請求項 14】

上記サブセットは、少なくとも、上記顔検出器によって検出された各顔を含むクロッピングされた画像を含むことを特徴とする請求項 13 記載のビデオカメラ装置。

【請求項 15】

上記サブセットは、捕捉された画像に関して、それぞれが 1 つの検出された顔を表し、該捕捉された画像において検出された顔の数に等しい複数のクロッピングされた画像を含むことを特徴とする請求項 13 記載のビデオカメラ装置。

【請求項 16】

上記クロッピングされた画像のそれぞれの表示特性を選択するためのユーザ操作子を備える請求項 15 記載のビデオカメラ装置。

【請求項 17】

上記データ取扱媒体は、伝送媒体であり、上記ユーザ操作子は、該伝送媒体の遠隔ノードに設けられていることを特徴とする請求項 16 記載のビデオカメラ装置。

【請求項 18】

上記サブセットは、捕捉された画像に関して、単一の検出された顔を表すクロッピングされた画像を含むことを特徴とする請求項 13 記載のビデオカメラ装置。

【請求項 19】

画像の一部において顔が検出されたか否かに応じて、該画像の一部に適用されるデータ圧縮率を変更するロジックを備える請求項 10 乃至 12 いずれか 1 項記載のビデオカメラ装置。

【請求項 20】

顔が検出されない捕捉された画像の部分に対しては、より高いデータ圧縮率を適用することを特徴とする請求項 19 記載のビデオカメラ装置。

【請求項 21】

上記顔検出器による顔検出に応じて、上記顔データに基づいて、上記撮像デバイスのレンズのズーム及び／又は向きを制御するロジックを備える請求項 10 乃至 12 いずれか 1 項記載のビデオカメラ装置。

【請求項 22】

それぞれが表示装置に接続された 2 つ以上の請求項 10 乃至 21 いずれか 1 項記載のビデオカメラ装置を備え、上記データ取扱媒体は、該 2 つ以上のビデオカメラ装置をリンクする伝送媒体であるビデオ会議システム。

【請求項 23】

請求項 10 乃至 21 いずれか 1 項記載のビデオカメラ装置を備える安全監視システム。

【請求項 24】

ビデオマテリアルを捕捉する撮像デバイスと、記録媒体とを備えるビデオカメラ装置の動作方法において、

上記捕捉されたビデオマテリアルにおいて人間の顔を検出し、該捕捉されたビデオマテリアルにおいて検出された顔の出現を特定する顔データを生成するステップと、

上記捕捉されたビデオマテリアル内の顔の検出に基づいて、送信又は保存のために、上記捕捉されたビデオマテリアルを表すデータを生成するステップとを有する動作方法。

【請求項 25】

10

20

30

40

50

請求項 2 4 記載の動作方法を実行するプログラムコードを含むコンピュータソフトウェア。

【請求項 2 6】

請求項 2 5 記載のプログラムコードを提供する提供媒体。

【請求項 2 7】

当該提供媒体は、記録媒体であることを特徴とする請求項 2 6 記載の提供媒体。

【請求項 2 8】

当該提供媒体は、伝送媒体であることを特徴とする請求項 2 6 記載の提供媒体。

【発明の詳細な説明】

【技術分野】

10

【0 0 0 1】

本発明は、ビデオカメラに関する。

【背景技術】

【0 0 0 2】

所謂カムコーダを含むビデオカメラ／レコーダ装置は、民生用及びプロフェッショナル用の用途の両方で広く用いられている。ビデオカメラ／レコーダ装置は、通常、オーディオ及びビデオマテリアルをテープ状記録媒体に記録するが、例えば、光ディスク又は磁気ディスク等の他の記録媒体も提案されている。

【0 0 0 3】

近年、プロフェッショナル用のカムコーダにおいて、カムコーダが捕捉したオーディオ及びビデオマテリアルと共に、幾つかの所謂「メタデータ」（追加的なデータ）を記録する技術も提案されている。メタデータは、オーディオ及びビデオ情報と共に、テープ状記録媒体に記録してもよく、例えば、フラッシュメモリカード等の独立した記録媒体に記録してもよく、無線リンクを介して外部データベースに伝送してもよい。いずれの場合も、メタデータの主な目的は、ユーザが後にマテリアルを最大限に活用できるよう、ユーザを補助することである。

【0 0 0 4】

幾つかのメタデータは、人間であるオペレータによって（例えば、キーボードを用いて）作成され、例えば、ファイリングの位置、出演者／司会者、日時、製作スタッフ、カメラの種類、カメラマン等がこのクリップを「良好なショット（good shot）」とみなしているか否か等を含む情報を定義する。更に、カムコーダ及び関連する装置によって、他のクラスのメタデータを自動的に生成してもよく、このようなメタデータとしては、例えば、カメラレンズのフォーカス、ズーム及び絞りの設定、地理的な位置（グローバルポジショニングシステム（Global Positioning System：G P S）受信機を用いる）、カメラのメンテナンス予定等がある。

【0 0 0 5】

これらの自動的に生成されたメタデータのクラスは、ユーザが後に、記録された大量のビデオクリップのグループから、特定のビデオクリップの位置を発見する際にある程度有用であるが、人間であるオペレータによって作成された第1のメタデータのクラスの方が、より有用性が高い。例えば、ユーザが後に特定の有名人を含むクリップを検索しようとすることはよくあるが、富士フィルム社のレンズ（Fuji lens）を用いて絞り f 1 . 8 で撮影されたクリップを検索する必要が生じることは希である。このように人間が作成するメタデータは、有用性が高いが、マテリアルを捕捉したとき又は後に、人間が必要な全てのデータを入力することは、非常に煩雑（したがって、高価）である。

【発明の開示】

【課題を解決するための手段】

【0 0 0 6】

本発明に係るビデオカメラ装置は、ビデオマテリアルを捕捉する撮像デバイスと、捕捉されたビデオマテリアルにおいて人間の顔を検出し、捕捉されたビデオマテリアルにおいて検出された顔の出現を特定する顔データを生成する顔検出器と、捕捉されたビデオマテ

20

30

40

50

リアルを表すデータを伝送及び／又は記録するデータ取扱媒体と、捕捉されたビデオマテリアル内の顔の検出に基づいて、データ取扱媒体によって送信又は保存されるデータを生成するプロセッサとを備える。

【0007】

本発明は、少なくとも幾つかの実施形態において、(ビデオマテリアルのコンテンツを素早く検索することを望む後のユーザにとって)人間が作成したメタデータ以上の高い有用性があるマシンによって生成されたメタデータの新たなクラス、すなわち、顔画像を提供し、捕捉されたビデオマテリアルとともに保存することにより上述の課題を解決する。他の実施例において、顔検出に応じて、例えば、帯域幅を減少させ又は利用可能な伝送帯域幅を有効に活用するように、カメラ装置から遠隔ノードに伝送される画像信号の性質を変更してもよい。顔の出現は、単に画像(例えば、フィールド又はフレーム)の指標として取り扱ってもよいが、この情報は、画像内における位置に関する情報を含んでいることが望ましい。カメラ装置は、必要条件ではないが、好ましくは、カムコーダとしても知られる単一の構成を有しているとよい。

10

【0008】

本発明のこの他の様々な側面及び特徴は添付の特許請求の範囲に記載されている。

【発明を実施するための最良の形態】

【0009】

以下、顔検出に関する実施例を説明するが、本発明は、必要なオブジェクトを表すトレーニング画像によるトレーニングによって例えば自動車等、他の如何なるオブジェクトの検出にも同様に適用できる。

20

【0010】

図1は、顔検出システム及び／又は非線形編集システムとして用いる汎用コンピュータ装置のブロック図である。コンピュータ装置は、処理ユニット10を備え、処理ユニット10は、中央処理装置(CPU)20と、ランダムアクセスメモリ(RAM)のようなメモリ30と、ディスクドライブ40のような不揮発性記憶装置と、他の通常の構成要素とを備える。コンピュータ装置は、ローカルエリアネットワーク又はインターネット(あるいは両方)のようなネットワーク50に接続している。また、コンピュータシステムは、キーボード60と、マウス又は他のユーザ入力デバイス70と、表示画面80とを備える。当業者には、汎用コンピュータ装置がここで記載していない多くの他の従来の部品を含むことは、明らかである。

30

【0011】

図2は、顔検出に用いるビデオカメラレコーダ(カムコーダ)のブロック図である。カムコーダ100は、画像を電荷結合素子(CCD)からなる画像捕捉素子120上に合焦点するレンズ110を備える。電子的な形式で得られる画像は、テープカセットのような記録媒体140に記録するために画像処理回路130によって処理される。また、画像捕捉素子120によって捕捉された画像は、アイピース160を介して見られるユーザ表示画面150上に表示される。

40

【0012】

画像と関連している音を捕捉するために、1つ以上のマイクロホンが用いられる。これらのマイクロホンは、フレキシブルケーブルによってカムコーダ100に接続され、あるいはカムコーダ100の本体に搭載されるという意味で、外部マイクロホンであるとも言える。1台以上のマイクロホンからのアナログオーディオ信号は、記録媒体140に記録するための適切なオーディオ信号を生成するために、オーディオ処理回路170によって処理される。

40

【0013】

なお、ビデオ及びオーディオ信号は、デジタル形式又はアナログ形式のいずれか、あるいは両方の形式で記録媒体140に記録することができる。したがって、画像処理回路130及びオーディオ処理回路170は、アナログ／デジタル変換器を備えていてもよい。

カムコーダ100のユーザは、レンズ110に電気的制御信号200を送るようにレン

50

ズ制御回路 190 に作用するユーザ制御 180 によって、レンズ 110 の性能における画角を制御することができる。一般的に、フォーカス及びズームのような属性はこのように制御されるが、レンズの絞り又は他の属性は、ユーザによって操作される。

【0014】

更に、2個のユーザ操作子を説明する。記録媒体 140 への記録を開始し、中止するために押しボタン 210 が設けられている。例えば、押しボタン 210 を1回押したときに記録を開始し、もう1回押したときに記録を中止することができる。あるいは、押した状態を維持することにより記録を行い、又はある時間、例えば5秒間押すことにより記録を開始するようにしてもよい。これらの如何なる構成においても、始めと終わりがある各「撮影 (shot)」(連続した記録期間)に対するカムコーダ 100 の記録操作の確認は、技術的に非常に簡単である。10

【0015】

図 2 に示す「素晴らしい撮影マーカ (good shot marker: 以下、GSM という。)」20 は、ユーザによって操作され、これにより、ビデオ及びオーディオマテリアルに関連した「メタデータ」(関連データ)が記録媒体 140 に格納される。この特別の撮影は、ある観点で「素晴らしい (good)」と操作者によって主観的にみなされたこと(例えば、俳優が特によく演じた、ニュースリポータが各言葉を正しく発音した等)を示している。

【0016】

メタデータは、記録媒体 140 上の予備領域(例えば「ユーザデータ」領域)に、用いられている特定のフォーマット及び規格に依存して、記録される。あるいは、メタデータはリムーバブルメモリスティック(登録商標)のメモリ(図示せず)のような別個の記録媒体に格納することができ、あるいはメタデータは、例えば無線リンク(図示せず)によって通信する外部データベース(図示せず)に格納することもできる。メタデータには、GSM の情報だけでなく、撮影条件 (shot boundaries)、レンズの属性、ユーザ(例えばキーボード(図示せず))による文字情報入力、全地球測位システム受信機(図示せず)からの地理的位置情報等が含まれてもよい。20

【0017】

以上、メタデータを記録可能なカムコーダについて説明した。次に、このようなカムコーダに顔検出を適用する方法について説明する。

【0018】

カムコーダ 100 は、顔検出器構成 230 を備える。適切な構成のより詳細は、後に説明するが、顔検出器 230 は、画像処理回路 130 から画像が供給され、このような画像が1つ以上の顔を含むか否かを検出、又は検出することを試みる。顔検出器 230 は、顔検出データを、「yes / no」フラグの形式で、あるいは、各検出された顔内の目の位置のような顔の画像座標を含むより詳細な形式で出力することができる。この情報は、メタデータの他の形として処理し、上述したフォーマットとは異なるフォーマットで格納することができる。30

【0019】

後述するように、顔検出は、検出処理における他の種類のメタデータを用いることにより、助けられる。例えば、顔検出器 230 は、レンズ 110 の現在のフォーカス及びズーミング設定を示すレンズ制御回路 190 からの制御信号が供給される。これらは、画像のフォアグラウンドで表示されるあらゆる顔の予想される画像サイズの初期の表示を与えることによって、顔検出器 2130 を補佐することができる。なお、この観点では、フォーカス及びズーミングの設定は、カムコーダ 100 と撮影されている個人との予想される距離、更にはレンズ 110 の倍率を表している。これらの2つの属性からの顔の大きさの平均に基づいて、得られる画像データ内における顔の予想される大きさ(画素)を算出することができる。40

【0020】

従来の(既知の)音声検出器 240 は、オーディオ処理回路 170 からオーディオ情報が供給され、このようなオーディオ情報内の音声の存在を検出する。音声の存在は、対応50

する画像に顔がある可能性を、音声を検出しないときに比して、より高い指標 (indicator) で示すことができる。

【0021】

最終的に、撮影境界 (shot boundaries) 及びユーザによって最も有益であるとみなされるそれらの撮影を示す GSM 情報 220 及び撮影情報 (制御 210 から) は、顔検出器 230 に供給される。

【0022】

勿論、カムコーダがアナログ記録技術に基づく場合、画像及びオーディオ情報を処理するために、更なるアナログ / デジタル変換器 (以下、A / D 変換器という。) が必要とされる。

10

【0023】

この実施例では、2段階の顔検出技術を用いる。図3は、トレーニング段階を具体的に説明する図であり、図4は、検出段階を具体的に説明する図である。

【0024】

以前に提案された顔検出方法 (以下に示す参照4及び5を参照) と異なり、この方法は、全体としてではなく顔の一部のモデリングに基づいている。顔の一部は、顔の特徴 (所謂「選択サンプリング (selective sampling)」) の推定位置上の中心のブロック、又は顔の通常間隔でサンプリング (所謂「標準サンプリング (regular sampling)」) されたブロックである。ここでは、主に、経験的検定で良い結果が得られた標準サンプリングについて説明する。

20

【0025】

トレーニング段階では、解析処理を、顔を含むことが知られている一組の画像に、及び (オプションとして) 顔を含まないことが知られている画像 (「顔でない画像 (nonface images)」) の別のセットに適用する。解析処理は、検定画像を後に (検出段階で) 比較することができる顔及び顔でない特徴の数学的モデルを構築する。

【0026】

したがって、数学的モデル (図3のトレーニング処理 310) を構築するための基本的な手順は次の通りである。

1. 同じ目位置を有するように正規化された顔の画像のセット 300 の各顔を、小さいブロックに一様にサンプリングする。
2. 後に説明する各ブロックの属性を算出する。
3. 属性を、異なる値の処理しやすい数に量子化する。
4. 次に、量子化属性を、そのブロック位置に関して 1 つの量子化値を生成するために組み合わせる。
5. そして、1 つの量子化値を、エントリとしてヒストグラム、例えば図5に示すヒストグラムに記録する。全てのトレーニング画像の全てのブロック位置に関する累積されたヒストグラム情報 320 は、顔の特徴の数学的モデルの基礎を形成する。

30

【0027】

上述のステップを多数の検定顔画像について繰り返すことによって、1 つのそのようなヒストグラムを、各可能なブロック位置に対して作成する。テストデータについては、更に後述する付録 A で説明する。そこで、 8×8 ブロックの配列を用いる方式では、64 個のヒストグラムを準備する。処理の後半部において、検定する量子化属性を、ヒストグラムのデータと比較する。データをモデル化するために全部のヒストグラムを用いるという事実は、例えばガウス分布又は他の分布を後にパラメータ化するか否かと仮定する必要はないことを意味する。データ記憶空間 (必要ならば) を節約するために、同じヒストグラムが異なるブロック位置に対して再生利用できるように、類似しているヒストグラムを併合することができる。

40

【0028】

検出段階で、検定画像 350 を顔検出器 340 で処理するために、検定画像 340 内の連続したウィンドウを、以下のように処理する。

50

6. ウィンドウを、一連のブロックのように一様にサンプリングし、そして、各ブロックに関する属性を算出して、上述のステップ1～4のように量子化する。

7. 各ブロック位置の量子化属性値の対応する「確率(probability)」を、対応するヒストグラムから調べる。すなわち、各ブロック位置のそれぞれの量子化属性を生成し、そのブロック位置に関して予め生成されたヒストグラムと比較する。ヒストグラムが「確率」データを高める方法については後述する。

8. 得られる全ての確率を互いに乗算して、ウィンドウを「顔」又は「顔でない」に分類するために、閾値と比較する最終の確率を形成する。「顔」又は「顔でない」の検出結果は絶対検出よりもむしろ確率ベースの方法であることは、言うまでもない。顔を含んでいない画像を間違って「顔」として検出(所謂誤検出(false positive))してしまうことがある。また、顔を含んでいる画像を間違って「顔でない」として検出(所謂見逃し検出(false negative))してしまうこともある。あらゆる顔検出システムの目標は、誤検出の割合及び見逃し検出の割合を減らすことであるが、現在の技術では、これらの割合をゼロに減らすことは、不可能ではないとしても困難である。

【0029】

上述のように、トレーニング段階において、一組の「顔でない」画像は、「顔でない」ヒストグラムの対応するセットを生成するために用いることができる。そして、顔の検出を達成するために、顔でないヒストグラムから生成される「確率」を、個々の閾値と比較し、検定ウィンドウが顔を含むためには、確率が閾値以下でなければならない。代わりに、顔でない確率に対する顔確率の比を、閾値と比較することができる。

【0030】

元のトレーニングセットを例えば位置、方向、大きさ、アスペクト比、背景の風景、照明の明るさ及び周波数成分(frequency content)の変化等の「合成変化(synthetic variations)」330で処理することによって、特別な(extra)トレーニングデータを生成することができる。

【0031】

ここで、属性及びそれらの量子化の導き方について説明する。従来の技術において、属性は、ウィンドウに表示された画像内に存在する異なる種類のブロックを表すコアブロック(又は固有ベクトル)である所謂固有ブロックに関して評価される。まず、固有ブロックの生成について、図6を参照して説明する。

【0032】

固有ブロックの生成

この実施例の属性は、所謂固有ブロックに基づいている。固有ブロックを、トレーニングセットのブロックの有効な具象的な才能(good representational ability)を有するように設計した。したがって、固有ブロックは、トレーニングセットからのブロックの大きなセットに対して重要な構成要素の解析を実行することによって生成された。この処理を、図6に示すとともに、付録Bにおいてより詳細に説明する。

【0033】

システムのトレーニング

実験を、トレーニングブロックの2の異なるセットによって行った。

【0034】

固有ブロックセットI

まず最初に、トレーニングセットの25個の顔画像から得られる一組のブロックを用いた。 16×16 ブロックを、重ならないように、 16 画素毎にサンプリングした。このサンプリングを、図6に示す。図6から明らかのように、 16×16 ブロックは、個々の 64×64 トレーニング画像から生成される。これにより、全体的に合計400個のトレーニングブロックが生成される。

【0035】

これらのトレーニングブロックから生成される第1の10個の固有ブロックを、図7に示す。

10

20

30

40

50

【0036】

固有ブロックセットII

固有ブロックの第2セットを、トレーニングブロックのより大きなセットから生成した。これらのブロックは、トレーニングセット内の500個の顔画像から得られた。この場合、 16×16 ブロックを、8画素が重なるように、8画素毎にサンプリングした。各 64×64 個のトレーニング画像から、49個のブロックが生成され、これを合計24,500個のトレーニングブロックに適用した。

【0037】

これらのトレーニングブロックから生成される第1の12の固有ブロックを、図8に示す。

10

【0038】

実験に基づく結果により、固有ブロックセットIIが固有ブロックセットIより僅かに有効な結果を与えることが分かった。

【0039】

このことは、固有ブロックセットIIが、顔画像から取られたトレーニングブロックのより大きなセットから算出され、顔の変化をより良く表すと認められるからである。しかしながら、性能の向上は、大きくない。

【0040】

ヒストグラムの作成

64×64 顔画像内の各サンプリングされたブロック位置に対してヒストグラムを作成した。ヒストグラムの数は、ブロック間隔に依存する。例えば、16画素のブロック間隔の場合、16の可能なブロック位置があり、したがって、16個のヒストグラムが用いられる。

20

【0041】

図9は、単一のブロック位置を表すヒストグラムを作成するための処理を示す。ヒストグラムを、M個の顔画像の大きなトレーニングセット400を用いて作成する。各顔画像に対する処理は、以下の処理を含んでいる。

- ・顔画像の位置(i, j)から関連したブロックを抽出する処理410。
- ・ブロックの固有ブロックベースの属性を算出して、これらの属性から関連したビン数420を判定する処理。
- ・ヒストグラム430内の関連したビン数を増加させる処理。

30

【0042】

属性の出現度数の分布の有効な表現を与えるヒストグラムを作成するために、この処理は、トレーニングセット内のM個の画像のそれぞれに対して繰り返される。理想的には、Mの値は、非常に大きく、例えば数千である。この処理は、一組の元の顔と各元の顔の数百個の合成変化とからなるトレーニングセットを用いることにより、容易に達成することができる。

【0043】

ヒストグラムビン数の生成

ヒストグラムビン数は、図10で示すように、以下の処理を用いて、所定のブロックから生成される。 64×64 ウィンドウ又は顔画像から 16×16 ブロック440を抽出する。ブロックは、一組の「固有ブロック重み(eigenblock weight)」を生成するためにA固有ブロックのセット450上で推定される。これらの固有ブロック重みは、この実行で用いられる「属性」である。固有ブロック重みは、-1 ~ +1 の範囲を有する。この処理については、付録Bでより詳細に説明する。各重みを、一組の量子化属性470($w_i, i = 1, \dots, A$)を生成するために、レベルLの固定量に量子化する。量子化重みを、以下のように、単一の値に結合する。

40

【0044】

【数1】

$$h = w_1 \cdot L^{A-1} + w_2 \cdot L^{A-2} + w_3 \cdot L^{A-3} + \dots + w_{A-1} \cdot L^1 + w_A \cdot L^0$$

【0045】

ここで、生成される値 h は、ヒストグラム bin 数 480 である。なお、ヒストグラムの bin の全数は、 L^A で与えられる。

【0046】

10 ピン「成分 (contents)」、すなわちその bin 数を増加する属性のセットの出現度数は、トレーニング画像の数 M で割ると、確率値であるとみなすことができる。しかしながら、確率を閾値と比較するので、実際には M による除算をする必要がなく、この値を、計算から除外する。したがって、以下の説明では、ピン「成分」は正確な意味では出現度数であるが、ピン「成分」を確率値 (probability values) と称し、確率値であるかのように取り扱う。

【0047】

上述の処理は、トレーニング段階と検出段階の両方で用いられる。

【0048】

顔検出段階

20 顔検出処理は、検定画像を 64×64 の移動ウィンドウ (moving 64x64 window) によってサンプリングして、各ウィンドウ位置における顔確率を算出する処理を含む。

【0049】

図 11 に、顔確率の計算を示す。ウィンドウ内の各ブロック位置において、前の章で説明したように、ブロックの bin 数 490 を算出する。ブロックの位置に対する適切なヒストグラム 500 を用いて、各 bin 数を調べ、その bin 数の確率 510 を決定する。そして、これらの確率の対数 (log) の和 520 を全てのブロックに亘って算出し、顔確率値 P_{face} (あるいは対数尤度値 (log likelihood value)) を生成する。

【0050】

この処理は、全体の検定画像のために確率「マップ (map)」を生成する。換言すれば、確率値は、画像の全体に亘る各可能なウィンドウ中心位置に関して導かれる。そして、これらの確率値の全ての長方 (あるいは他) 形状の配列への結合は、その画像に対応した確率「マップ」であると認められる。

【0051】

そして、このマップを反転し、これにより、顔を検出する処理は、反転マップ内の最小値を見つけ出すことを含む。所謂距離ベースの技術 (distance-based technique) を用いる。この技術は、次のように要約することができる。反転確率マップ内の最も小さい値を有するマップ (画素) 位置を選択する。この値が閾値 (TD) より大きいときは、それ以上は顔を選択しない。これは、終了基準 (termination criterion) である。一方、選択された中心画素位置に対応している顔の大きさのブロックを無効にし (すなわち、後続の計算から除外し)、終了基準に達するまで、候補顔位置検出処理を画像の残りの部分に対して繰り返す。

【0052】

顔でない方法

顔でないモデルは、顔を含まない画像の属性の確率分布を示すヒストグラムの更なるセットを含む。このヒストグラムは、トレーニング画像が顔の代わりに顔でない画像を含む以外は、顔モデルと正確に同じ方法で生成される。

【0053】

検出を通じて、顔モデルと顔でないモデルを用いた 2 つの対数確率値を計算する。そして、これらを、単に顔確率から顔でない確率を減算ことによって結合する。

【0054】

10

20

30

40

50

【数2】

$$P_{combined} = P_{face} - P_{nonface}$$

【0055】

そして、 $P_{combined}$ を、(極性反転前の)確率マップを生成するために用いる。

【0056】

なお、 P_{face} から $P_{nonface}$ を減算する理由は、それらが対数確率値であるからである。

【0057】

ヒストグラムの具体例

図12a～12fは、上述したトレーニング処理によって生成されるヒストグラムの具体例を示す図である。

【0058】

図12a、12b、12cは、顔画像のトレーニングセットから導かれ、図12d、12e、12fは、顔を含まない画像のトレーニングセットから導かれたものである。詳しくは、以下の表に示す。

【0059】

【表1】

10

20

	顔ヒストグラム	非顔ヒストグラム
ヒストグラム全体	図12a	図12d
$h=約1500$ の主ピークにズーム	図12b	図12e
$h=約1570$ の領域に更にズーム	図12c	図12f

【0060】

ピークが顔ヒストグラムと顔でないヒストグラムとでは異なる位置にあることが明らかに見られる。

【0061】

多尺度(multiscale)顔検出

検定画像における異なる大きさの顔を検出するために、検定画像を係数の範囲(range of factors)によって拡大縮尺し(scale)、距離(distance、すなわち確率)マップを各尺度(scale)に対して生成する。図13a～13cに、画像及びこれらの対応する距離マップを、3つの異なる尺度で示す。この手法は、最も小さい尺度(図13a)で大きな(中心の)対象に対して最良の応答(最高の確率又は最小の距離)を示し、大きな尺度でより小さい対象(主人物の左側)に対してより良い応答を示す。(マップ上のより暗い色は、反転マップにおけるより低い値、すなわちそこが顔であるというより高い確率を示す)。全ての尺度に対して最良の応答を示す位置を最初に検出ことによって、候補顔位置を異なる尺度に亘って抽出する。すなわち、最も高い確率(最も短い距離)を、全ての尺度で全ての確率マップ中で確定する。この候補位置は、顔としてラベルが付けられる第1の位置である。そして、その顔位置の中心に置かれるウィンドウは、各尺度の確率マップから削除される。削除されるウィンドウの大きさは、確率マップの尺度に比例する。

30

40

【0062】

この拡大縮小削除処理(scaled blanking out process)の具体例を、図13a～図13cに示す。特に、全てのマップに亘って最高の確率は、最大尺度マップ(図13c)の左側で検出される。図13cにおいて、顔の推定された大きさに対応する領域530を削

50

除する。より小さいマップにおいて、対応する縮小された領域 532、534 を削除する。

【0063】

重複検出を避けるために、マップの検定ウィンドウより大きな領域を削除する。特に、そのような重複検出を避けるのに適した領域の大きさは、検定ウィンドウの幅 / 長さに、それぞれの半分を加えた大きさである。

【0064】

次に良い応答を探索し、対応するウィンドウを次々に削除することによって、新たな顔を検出する。

【0065】

10
処理される尺度間で許容される間隔は、大きさの変化に対する処理の感度 (sensitivity) に影響される。尺度に対して不变であるという (scale invariance) この予備的な研究において、多くの場合、ある尺度で良い応答を示した顔が同様に隣の尺度で良い応答を示すように、この処理が大きさの変化に極端に敏感でないことが分かった。

【0066】

上述の説明では、画像における顔の大きさが検出処理の初めでは分かっていないときの顔検出について言及している。多尺度顔検出 (multiple scale face detection) の別の態様は、検出処理を確証するために、異なる尺度での 2 つ以上の並列検出に用いることがある。例えば、検出される顔が部分的に覆い隠され、あるいは人が帽子等をかぶっている場合に、この方法は、利点を有する。

【0067】

図 13d ~ 図 13g は、この処理を示す図である。トレーニング段階の間、装置を、「フルフェース (full face)」ヒストグラムデータを生成するために、検定顔全体を囲むウィンドウ (図 13d、上述したようにそれぞれのブロックに分割されている) 上でトレーニングし、更に、「ズームイン (zoomed in)」ヒストグラムデータ生成するために、検定顔の中心領域だけが含まれるように拡大した尺度のウィンドウ (図 13e) 上でトレーニングする。これにより、2 セットのヒストグラムデータを生成する。1 つのセットは、図 13d の「フルフェース」ウィンドウに関し、他のセットは、図 13e の「中心顔領域 (central face area)」ウィンドウに関する。

【0068】

30
検出段階の間、あらゆる与えられた検定ウィンドウ 536 において、検定ウィンドウが顔の予想される大きさの全部を囲み (図 13f)、また、他の検定ウィンドウがその予想される大きさで顔の主要な領域を含む (図 13g) ように、ウィンドウを、検定画像の 2 の異なる尺度に対して適用する。これらは、それぞれ、上述したように処理し、ウィンドウの種類に適したヒストグラムデータの各セットと比較する。各並列処理からの対数確率を加えた後、閾値と比較する。

【0069】

これらの多尺度顔検出の態様を一緒にすることにより、特に、格納しなければならないデータ量を著しく節約することができる。

【0070】

40
特に、これらの実施例において、図 13a ~ 図 13c の配置に対する複数の尺度は、等比数列的に配列されている。この具体例においては、等比数列の項として並べられた各尺度における隣り合う尺度は、次項が前項の⁴ 2 倍となるように配列されている。そして、図 13d ~ 図 13g によって説明した並列検出のために、より大きな尺度、中央領域、検出は、この数列において 3 段階高い尺度、すなわち「フルフェース」尺度よりも $2^{3/4}$ 倍大きな尺度で、数列において 3 段階高い尺度に関する属性データを用いて実行する。したがって、複数の尺度範囲の両端は別として、等比数列 (geometric progression) は、図 13d ~ 図 13g の並列検出が、順番で 3 段階高い他の多尺度に関して生成される属性データを用いて、常に実行できることを意味している。

【0071】

10

20

30

40

50

2つの処理（多尺度検出と並列尺度検出）は、様々な方法で組み合わせることができる。例えば、図13a～図13cの多尺度検出処理を最初に適用し、それから、図13d～図13gの並列尺度検出処理を、多尺度検出処理の間に識別された領域（及び尺度）で適用することができる。しかしながら、属性データの便利で効率的な使用は、以下のようにして達成することができる。

- ・各尺度で検定ウィンドウに関する属性を導く（図13a～図13cに示すように）。
- ・それらの属性を「フルフェース」ヒストグラムデータと比較して、「フルフェース」の距離マップのセットを生成する。
- ・属性を「ズームイン」ヒストグラムデータと比較して、「ズームイン」の距離マップのセットを生成する。
- ・各尺度nについて、尺度n+3の「ズームイン」距離マップを、尺度nの「フルフェース」距離マップに組み合わせる。
- ・図13a～図13cによって上述したように、組み合せた距離マップから顔位置を導く。

【0072】

更に、例えば正面、部分的に上、下、左、右等を見た様々なポーズを検出するために、並列検定を、実行することができる。ここで、ヒストグラムデータのそれぞれのセットが必要とされるとともに、結果が、「最大」関数を用いて好ましくは組み合せられ、すなわち、最も高い確率を与えるポーズは閾値処理に送られ、他は破棄される。

【0073】

顔追跡

顔追跡アルゴリズムについて説明する。追跡アルゴリズムは、画像シーケンスにおいて顔検出性能を向上させることを意図している。

【0074】

追跡アルゴリズムの初期の目標は、画像シーケンスの全てのフレームにおける全ての顔を検出することである。しかしながら、時々、シーケンス内の顔が検出できないことが認められる。これらの環境で、追跡アルゴリズムは、見逃した顔検出全体で補間するように補佐することができる。

【0075】

最終的に、顔追跡の目標は、画像シーケンスにおいて同じシーンに属しているフレームの各セットから有効なメタデータを出力できることである。このメタデータには、以下のものが含まれる。

- ・顔の数。
- ・各顔の「顔写真(Mugshot)」（個人の顔の画像を表す口語的な言葉、警察にファイルされている写真を照会する用語からきている）。
- ・各顔が最初に出現するフレーム番号。
- ・各顔が最後に出現するフレーム番号。
- ・各顔の識別（前のシーンで見られた顔に一致するか、顔のデータベースに一致したもの）-顔の識別には、顔の認識も必要とされる。

【0076】

追跡アルゴリズムは、顔検出アルゴリズムの結果を用い、画像シーケンスの各フレーム上で、その開始位置として独立して実行される。顔検出アルゴリズムは時々顔を逃がす（検出しない）こともあるので、見逃した顔を内挿する（interpolating）方法は有効である。このために、顔の次の位置を予測するためにカルマンフィルタ（Kalman filter）を用い、顔追跡を助けるために、肌色マッチングアルゴリズム（skin colour matching algorithm）を用いた。更に、顔検出アルゴリズムが頻繁に誤った採用の原因となるので、また、これらを排除する方法は有効である。

【0077】

このアルゴリズムを、図14に示す。

【0078】

10

20

30

40

50

このアルゴリズムについては以下に詳細に説明するが、要約すると、入力ビデオデータ545（画像シーケンスを表す）がこの出願で説明される種類の検出器540及び肌色マッチング検出器550に供給される。顔検出器540は、各画像内で1つ以上の顔を検出することを試みる。顔が検出されると、カルマンフィルタ560が起動され、その顔の位置を追跡する。カルマンフィルタ560は、画像シーケンスにおける次の画像内で同じ顔の予測される位置を生成する。目の位置比較器570、580は、顔検出器540が次の画像内のその位置（あるいは、その位置からある閾値距離の範囲内）で顔を検出したかを、検出する。顔が検出された場合、その検出された顔位置は、カルマンフィルタを更新するために用いられ、処理が続けられる。

【0079】

10

顔が予測された位置で、あるいは近くで検出されない場合、肌色マッチング回路550を用いる。肌色マッチング回路550は、厳密でない顔検出技術であり、その検出の閾値は顔検出器540よりも低く設定され、顔検出器540がその位置で顔があると検出することができないときでさえ、顔を検出する（顔があるとみなす）ことができる。肌色マッチング回路550によって「顔」が検出されると、その位置がカルマンフィルタ560に更新された位置として供給され、処理が続けられる。

【0080】

顔検出器450又は肌色マッチング回路550によって一致が検出されないとときは、カルマンフィルタを更新するために予測された位置を用いる。

【0081】

20

これらの結果の全ては、判定基準（下記参照）に対する対象である。したがって、例えば、1つの正しい検出に基づきシーケンスを通して追跡される顔、及び予測の残り又は肌色検出の残りは、破棄する。

【0082】

独立したカルマンフィルタは、追跡アルゴリズムにおいて各顔を追跡するために用いられる。

【0083】

顔を追跡するためにカルマンフィルタを用いるためには、顔を表す状態モデルを、生成しなければならない。状態モデルにおいて、各顔の位置は、左右の目の座標を含む4次元のベクトルによって表し、左右の目の座標は、ウィンドウの中央位置に対する所定の関係によって順番に導かれる。そして、以下の尺度を用いる。

30

【0084】

【数3】

$$p(k) = \begin{bmatrix} FirstEyeX \\ FirstEyeY \\ SecondEyeX \\ SecondEyeY \end{bmatrix}$$

40

【0085】

ここで、kはフレーム番号である。

【0086】

顔の現在の状態は、12次元のベクトルにおける、その位置、速度及び加速度によって表される。

【0087】

【数4】

$$\hat{z}(k) = \begin{bmatrix} p(k) \\ \vdots \\ \overset{\bullet}{p}(k) \\ \vdots \\ \ddot{p}(k) \end{bmatrix}$$

【0088】

検出された最初の顔

10

追跡アルゴリズムは、顔の存在を示す顔検出結果とともにフレームが供給されるまで、何もしない。

【0089】

そして、カルマンフィルタ560は、このフレームで検出された各顔毎に初期化される。その状態は、顔の位置によって、及び速度及び加速度をゼロとして初期化される。

【0090】

【数5】

$$\hat{z}_a(k) = \begin{bmatrix} p(k) \\ 0 \\ 0 \end{bmatrix}$$

20

【0091】

また、カルマンフィルタ560は、他の属性状態、すなわちモデル誤差共分散(state model error covariance)Qと、観測誤差共分散(observation error covariance)Rとが割り当てられる。また、カルマンフィルタの誤差共分散Pも、初期化される。これらのパラメータについては、後により詳細に説明する。次のフレーム及び全ての後続のフレームの初めに、カルマンフィルタの予測処理が実行される。

30

【0092】

カルマンフィルタの予測処理

各既存のカルマンフィルタに対して、顔の次の位置が、以下の標準カルマンフィルタ予測方程式を用いて予測される。カルマンフィルタは、前状態(フレームk-1)、及びフィルタの現状態(フレームk)を推定する他の内部及び外部変数を用いる。

状態予測式:

【0093】

【数6】

$$\hat{z}_b(k) = \Phi(k, k-1)\hat{z}_a(k-1)$$

40

【0094】

共分散予測方程式:

【0095】

【数7】

$$P_b(k) = \Phi(k, k-1)P_a(k-1)\Phi(k, k-1)^T + Q(k)$$

【0096】

50

ここで、 $Z_b \wedge (k)$ （「 $Z_b \wedge$ 」は、ハット Z_b を表すもとのする。以下同様）は、フレーム k に対するフィルタを更新する前の状態を示し、 $Z_a \wedge (k-1)$ はフレーム $k-1$ に対するフィルタを更新した後の状態（あるいは、それが新規のフィルタであるときは、初期化状態）を示し、 $(k, k-1)$ は状態推移行列である。様々な状態推移行列を、後述するように試した。同様に、 $P_b(k)$ は、フレーム k に対するフィルタを更新する前のフィルタの誤差共分散を示し、 $P_a(k-1)$ は、前フレームに対するフィルタを更新した後のフィルタの誤差共分散（あるいは、それが新規のフィルタであるときは、初期化された値）を示す。 $P_b(k)$ は、その精度をモデル化するフィルタの内部変数として考えることができる。

【0097】

$Q(k)$ は、状態モデルの誤差共分散である。 $Q(k)$ の高い値は、フィルタの状態の予測値（すなわち顔の位置）が高いレベルの誤差を有するとみなされることを意味する。このパラメータを調整することによって、フィルタの動作を変更することができ、顔検出に対して潜在的に向上させることができる。

【0098】

状態推移行列

状態推移行列 $(k, k-1)$ は、次の状態の予測をどのようにして行うかを決定する。運動方程式を用い、 $(k, k-1)$ に対して次の行列を導くことができる。

【0099】

【数8】

10

20

$$\Phi(k, k-1) = \begin{bmatrix} I_4 & I_4\Delta t & \frac{1}{2}I_4(\Delta t)^2 \\ O_4 & I_4 & I_4\Delta t \\ O_4 & O_4 & I_4 \end{bmatrix}$$

【0100】

ここで、 O_4 は 4×4 零行列であり、 I_4 は 4×4 単位行列である。 t は、簡単に1（すなわち、 t の単位はフレーム周期である）に設定することができる。

【0101】

この状態推移行列は、位置、速度及び加速度をモデル化する。しかしながら、予測状態を補正するために顔検出が利用できなかったときには、加速度を使用すると、顔の予測が画像の端の方に加速される傾向があることが、分かった。したがって、加速度を用いないより単純な状態推移行列が好ましい。

【0102】

【数9】

30

40

$$\Phi(k, k-1) = \begin{bmatrix} I_4 & I_4\Delta t & O_4 \\ O_4 & I_4 & O_4 \\ O_4 & O_4 & O_4 \end{bmatrix}$$

【0103】

各カルマンフィルタの予測された目位置 $Z_b \wedge (k)$ を、現フレームにおける全ての顔検出結果（存在する場合）と比較する。目位置間の距離が所定の閾値以下であるときは、顔検出は、カルマンフィルタによってモデル化される顔と同じ顔に属するとみなすことができる。そして、顔検出結果を、顔の現在の状態の観測 $y(k)$ として扱う。

50

【 0 1 0 4 】

【 数 1 0 】

$$y(k) = \begin{bmatrix} p(k) \\ 0 \\ 0 \end{bmatrix}$$

【 0 1 0 5 】

10

ここで、 $p(k)$ は顔検出結果の目の位置である。この観測は、予測を補正するのを助けるために、カルマンフィルタ更新過程の間、用いられる。

【 0 1 0 6 】

肌色マッチング

肌色マッチングは、顔検出結果と良く一致する顔に対しては用いられない。肌色マッチングは、カルマンフィルタによって位置が予測された顔に対して実行されるだけであり、現フレーム内の顔検出結果に一致しない、したがってカルマンフィルタの更新に役立つ観測データがない顔に対しては実行されない。

【 0 1 0 7 】

20

第1の手法において、各顔に対して、顔の前の位置を中心とした楕円状の領域が前フレームから抽出される。このような領域の具体例は、図16に示すように、顔ウィンドウ610内の領域600である。ガウス分布モデルに基づいて、 C_r 値及び C_b 値の平均値及び共分散の推定値を生成するために、この領域の色差データを用いる色モデルを、シードする。

【 0 1 0 8 】

そして、現フレームの予測された顔位置の周辺の領域を探索し、楕円状の領域で再び平均した色モデルと最も一致する位置を選択する。色一致が所定の類似基準を満たすときは、この位置を、前の章で説明した顔検出結果と同様の方法で、顔の現在の状態の観測 $y(k)$ として用いる。

【 0 1 0 9 】

30

図15a及び図15bは、探索領域の生成を具体的に説明する図である。特に、図15aは、次の画像630内の顔の予測された位置620を具体的に説明するための図である。肌色マッチング回路540において、次の画像で予測された位置620を囲んでいる探索領域640が、顔を検出するために探索される。

【 0 1 1 0 】

色の一一致が類似基準を満たさないときは、現フレームに対する信頼性が高い観測データは利用できない。代わりに、予測状態 Z_b^k を観測データとして用いる。

【 0 1 1 1 】

【 数 1 1 】

40

$$y(k) = \hat{z}_b(k)$$

【 0 1 1 2 】

上述した肌色マッチング方法は、単純なガウス分布の肌色モデルを用いる。モデルは、前フレームで顔の中央に置かれる楕円状の領域の上でシードされて、現フレームで最も一致している楕円状の領域を検出したものである。しかしながら、潜在的により有効な性能を得るために、更なる2つの方法、すなわち色ヒストグラム法 (colour histogram method) 及び色マスク法 (colour mask method) を開発した。これらを次に説明する。

【 0 1 1 3 】

50

色ヒストグラム法

この方法においては、追跡された顔で色の分布をモデル化するガウス分布を用いる代わりに、色ヒストグラムを用いる。

【0114】

前フレームの各追跡された顔に対して、顔の周りの正方形のウィンドウ内のCr値及びCb値のヒストグラムを計算する。ヒストグラムを計算するために、各画素に関するCr値とCb値を、まず、単一の値に組み合わせる。次に、全部のウィンドウ内のこれらの値の出現度数を評価するヒストグラムを計算する。組み合わせるCr値とCb値の数が大きい(256×256の可能な組合せ)ので、ヒストグラムを算出する前に、値を量子化する。

10

【0115】

前フレームの追跡された顔に対するヒストグラムを算出し、このヒストグラムを現フレームに対して用いて、最も類似した色分布を有する画像の領域を検出することによって、顔の最も有望な新たな位置を推定することを試みる。図15a、図15bに示すように、この推定は、現フレームの探索領域内のウィンドウ位置の範囲に対して正確に同じ方法でヒストグラムを算出することによって、行う。この探索領域は、予測された顔位置の周辺の所定の領域を含む。そして、前フレームの追跡された顔に関する元のヒストグラムと現フレームの各ヒストグラム間の二乗平均誤差(MSE)を算出することによって、これらのヒストグラムを比較する。現フレームの顔の推定される位置は、最小のMSEの位置によって与えられる。

20

【0116】

この色ヒストグラム法のアルゴリズムは、様々に変更することができ、以下のものが含まれる。

- ・2チャンネル(Cr、Cb)の代わりに3チャンネル(Y、Cr、Cb)を用いる。
- ・量子化レベルの数を変える。
- ・ウィンドウをブロックに分割して、各ブロックのヒストグラムを算出する。この方法では、色ヒストグラム法は、位置的に依存するようになる。この方法では、ヒストグラムの各対間のMSEを合計する。
- ・ウィンドウを分割するブロックの数を変える。
- ・実際に使用するブロックを変える。例えば、単に部分的に顔の画素を含む外側のブロックを省略する。

30

【0117】

これらの技術の実証的な試験で用いられる試験データに対して、最良の結果は次の条件を用いて達成された。しかしながら、他の条件により、異なる試験データを用いて、同様に良好な又はより良い結果が得られることもある。

- ・3チャンネル(Y、Cr及びCb)。
- ・各チャンネルに対して8つの量子化レベル(すなわち、ヒストグラムは $8 \times 8 \times 8 = 512$ ビン(bins)を含む)。

ウィンドウを16ブロックに分割する。

40

16個の全てのブロックを使用する。

【0118】

色マスク法

この方法は、上述した第1の方法に基づいている。色マスク法は、顔の画素分布を記述するのに、ガウス分布肌色モデルを用いる。

【0119】

上述した第1の方法における顔の中央に置かれた楕円状の領域を、色一致顔に用いる。この領域は、モデルを劣化する可能性がある背景画素の数を減らし、又は最小にすると認められる。

【0120】

存在色マスクモデルにおいて、類似した楕円状の領域を、例えばガウス分布モデルのパ

50

ラメータを設定するために R G B 又は Y C r C b の平均値及び共分散を適用することによって、前フレームで元の追跡された顔に色モデルをシードするために、更に用いる（あるいは、代わりに、ガウス分布モデルのようなデフォルトの色モデルを、以下に示すように、用いることができる）。しかしながら、現フレームにおいて最良の一一致を探索するときは、それは用いられない。代わりに、マスク領域を、前フレームからの元の顔のウィンドウにおける画素の分布に基づいて算出する。マスクは、最も色モデルと一致するウィンドウ内の 50 % の画素を検出することによって算出される。具体例を、図 17 a ~ 図 17 c に示す。詳細には、図 17 a は、検定における初期のウィンドウを具体的に示している。図 17 b は、色モデルをシードするために使用される橢円状のウィンドウを具体的に示している。図 17 c は、最も色モデルと一致する 50 % の画素によって定義されるマスクを具体的に示している。

10

【0121】

現フレームにおいて顔の位置を推定するために、予測された顔位置の周りの探索領域を探査し（前と同じ）、色モデルからの「距離」を各画素に対して算出する。「距離」は、その次元の分散によって各次元で正規化された平均値からの差を表している。得られる距離画像の具体例を図 18 に示す。この距離マップの各位置に対して（あるいは、計算時間を減らすためにサンプリングされた位置の少ないセットに対して）、距離画像の画素は、マスクの形をした領域上で平均されている。そして、最低の平均距離を有する位置を、このフレームにおける顔の位置の最良の推定値として、選択する。

20

【0122】

したがって、この方法は、橢円状の領域の代わりに、マスクの形をした領域を距離画像において用いるという点で、元の方法と異なる。これにより、色マッチング法は、色と形状の両方の情報を使用することができる。

【0123】

2つの変形例を提案し、技術の実証的な試験において実行した。

(a) ガウス分布肌色モデルは、前フレームにおいて追跡された顔の中央に置かれた橢円状の領域の C r 及び C b の平均値及び共分散を用いてシードされる。

(b) デフォルトのガウス分布肌色モデルは、前フレームのマスクを算出するために、及び現フレームにおいて距離画像を算出するための両方に、用いられる。

30

【0124】

ここで、ガウス分布肌色モデルの使用について説明する。肌色クラスのガウス分布モデルは、Y C b C r 色空間の色差成分を用いて作成される。肌色クラスへの検定画素の類似は、そして、評価ができる。この方法は、したがって各画素（独立して固有顔に拠点を置く方法の）に関する、肌色尤度推定値を提供する。

【0125】

検定画素の C b C r 値のベクトルを w とする。肌色クラス S に w が属している確率は、二次元のガウス分布によってモデル化される：

【0126】

【数 12】

40

$$p(w | S) = \frac{\exp\left[-\frac{1}{2}(w - \mu_s)\Sigma_s^{-1}(w - \mu_s)\right]}{2\pi|\Sigma_s|^{\frac{1}{2}}}$$

【0127】

ここで、分布の平均値 μ_s 及び共分散行列 Σ_s は、肌色値のトレーニングセットから予め推定される。肌色検出は、単独で用いた場合、効果的な顔検出法とはならない。これは、画像内には、肌色に類似する多くの領域が存在する可能性があり、これらの領域は、例えば顔以外の体の一部等、必ずしも顔ではないことが多いからである。しかしながら、この

50

肌色検出を顔追跡システムについて説明した手法と組み合わせて用いることにより、固有顔法の性能を向上させることができる。顔検出された目位置及びカルマンフィルタによる観測として色マッチングした目位置のいずれを認めるか、又はいずれの観測も認めないかに関する決定を保存する。この決定に関する情報は、各カルマンフィルタによってモデル化された顔に関する継続中の有効性を評価するために後に用いられる。

【0128】

カルマンフィルタ更新ステップ

更新ステップは、状態予測及び観測データに基づいて、現在のフレームに対するフィルタの出力を適切なものにするために用いられる。更新ステップでは、予測された状態と観測された状態との間の誤差に基づいて、フィルタの内部変数も更新される。

10

【0129】

更新ステップでは、以下の式が用いられる。

【0130】

カルマン利得式

【0131】

【数13】

$$K(k) = P_b(k)H^T(k)(H(k)P_b(k)H^T(k) + R(k))^{-1}$$

【0132】

20

状態更新式

【0133】

【数14】

$$\hat{z}_a(k) = \hat{z}_b(k) + K(k)[y(k) - H(k)\hat{z}_b(k)]$$

【0134】

共分散更新式

【0135】

【数15】

30

$$P_a(k) = P_b(k) - K(k)H(k)P_b(k)$$

【0136】

ここで、 $K(k)$ は、カルマンフィルタの他の内部変数であるカルマン利得を表す。カルマン利得は、観測された状態に基づいて、予測された状態 $y(k)$ をどれ程調整するかを決定するために用いられる。

【0137】

$H(k)$ は、観測行列を表す。これは、状態のどの部分が観測可能であるかを決定する。ここでは、顔の位置のみが観測可能であり、速度又は加速度は観測可能でないため、 $H(k)$ として以下の行列を用いる。

40

【0138】

【数16】

$$H(k) = \begin{bmatrix} I_4 & O_4 & O_4 \\ O_4 & O_4 & O_4 \\ O_4 & O_4 & O_4 \end{bmatrix}$$

【0139】

R (k) は、観測データの誤差共分散を表す。Q (k) と同様に、R (k) の値が大きいことは、フィルタ状態の観測された値（すなわち、顔検出結果又は色マッチング）の誤差が大きいことを意味する。このパラメータを調整することにより、フィルタの振る舞いを変更し、顔検出の精度を潜在的に向上させることができる。本願発明者らによる実験では、Q (k) に対して大きな値のR (k) は、適当であることが見出された（これは、予測された顔位置が観測より信頼性が高いものとして取り扱われていることを意味する）。なお、これらのパラメータはフレーム毎に変更してもよい。したがって、観測が顔検出結果（より信頼できる）に基づいているか、色マッチング（信頼性が低い）に基づいているかに基づいて、将来検討すべき領域について、R (k) と Q (k) の相対的な値を調整してもよい。

【0140】

各カルマンフィルタについて、更新状態 Z_a^k を顔の位置に関する最終決定として用いる。このデータはファイルに出力し、保存する。

【0141】

マッチングされた顔検出結果は、新たな顔として取り扱う。これらの新たな顔については、新たなカルマンフィルタを初期化する。以下の顔は削除される。

- ・ピクチャのエッジから外れた顔及び / 又は
- ・これらの顔をサポートする継続中の証拠がない顔（顔検出結果又は色マッチングより、カルマンフィルタ予測に基づく観測の割合が高い場合）

これらの顔については、関連するカルマンフィルタを除去し、ファイルにはデータを出力しない。

【0142】

この手法からの任意の変形例として、顔がピクチャから外れたことが検出された場合、その顔がピクチャから外れる前のフレームまでの追跡結果を保存し、有効な顔追跡結果として取り扱ってもよい（この結果が追跡結果に適用される他の判断基準に適合していることを前提とする）。これらの結果は、次のような更なる変数を用いることによって定式化及び構築することができる。

【0143】

`prediction_acceptance_ratio_threshold` (予測許容率閾値)

所定の顔の追跡の間に、許容されたカルマン予測顔位置の割合が、この閾値を超えた場合、追跡された顔は拒否される。この閾値は、現在 0.8 に設定している。

【0144】

`detection_acceptance_ratio_threshold` (検出許容率閾値)

全てのフレームに亘る最後のパスの間に、所定の顔について、許容された顔検出割合がこの閾値を下回った場合、追跡された顔は拒否される。この閾値は、現在 0.08 に設定している。

【0145】

`min_frames` (最少フレーム)

全てのフレームに亘る最後のパスの間に、所定の顔について、発生回数が `min_frames` より少ない場合、その顔は拒否される。このような場合が生じる可能性があるのは、シーケンスの最後の近くのみである。`min_frames` は、現在 5 に設定している。

10

20

30

40

50

【 0 1 4 6 】

`final_prediction_acceptance_ratio_threshold`及び`min_frames2`(最終予測許容率閾値及び最少フレーム2)

全てのフレームに亘る最後のパスの間に、追跡されている所定の顔の発生回数が`min_frames2`であり、許容されたカルマン予測顔位置の割合が`final_prediction_acceptance_ratio_threshold`を超えた場合、その顔は拒否される。このような場合が生じる可能性があるのは、シーケンスの最後の近くのみである。`final_prediction_acceptance_ratio_threshold`は、現在0.5に設定しており、`min_frames2`は、現在10に設定している。

【 0 1 4 7 】

`min_eye_spacing`(最小目間隔)

10

更に、追跡された顔の目の間隔が所定の最小距離を下回る場合、その顔を除外する。このような状況は、カルマンフィルタが目の間隔が狭くなっていると誤って判断し、例えば顔検出結果等、この判断を訂正する他の証拠がない場合に生じる。これを訂正しなければ、目の感覚は最終的に0になってしまう可能性がある。任意の変形例として、目の間隔に関する最小値又は下限を強制的に設定し、検出された目間隔が最小目間隔より小さい場合、検出処理はその目間隔を有する顔の探索を継続し、これより小さい目間隔を有する顔の探索は行わないようにしてもよい。

【 0 1 4 8 】

なお、追跡処理は、必ずしもビデオシーケンスを時間的に順方向に追跡する必要はない。画像データにアクセス可能であれば(すなわち、処理が実時間ではなく、又は画像データが時間的に継続する用途のためにバッファリングされている場合)、追跡処理を時間的に逆方向に行うことでもできる。又は、第1の顔が検出された場合(多くの場合ビデオシーケンスの途中で検出される)、追跡処理は、時間的に順方向及び逆方向の両方について開始してもよい。更なる任意の処理として、追跡処理は、ビデオシーケンス全体に亘って、時間的に順方向及び逆方向の両方について実行し、これらの追跡の結果を組み合わせて(例えば)許容基準に適合する追跡された顔が、追跡が実行されたいずれの方向についても有効な結果として含ませてもよい。

20

【 0 1 4 9 】

顔追跡のための重複規則(Overlap Rules for Face Tracking)

30

顔が追跡される場合、顔追跡が重複する可能性がある。重複が発生した場合、少なくとも幾つかのアプリケーションでは、追跡の1つを削除する必要がある。重複が発生した場合、どの顔追跡を持続するかを決定するために、一組の規則を用いる。

【 0 1 5 0 】

顔追跡においては、次のような3つの可能な追跡の種類がある。

D：顔検出 - 現在の顔の位置は、新たな顔検出によって確認される。

S：肌色追跡 - 顔検出は行われていないが、適切な肌色追跡が行われている。

P：予測 - 適切な顔検出も肌色追跡も行われておらず、カルマンフィルタからの予測された顔位置が使用されている。

【 0 1 5 1 】

以下に示す表は、2つの顔追跡が互いに重複した場合の優先順を定義している。

40

【 0 1 5 2 】

【表2】

顔1 顔2	D	S	P
D	最大顔サイズ	D	D
S	D	最大顔サイズ	S
P	D	S	最大顔サイズ

10

【0153】

すなわち、両方の追跡が同じ種類である場合、顔のサイズが最も大きい追跡を維持する。この他の場合、顔検出による追跡は、肌色追跡又は予測追跡より高い優先順位を有する。肌色追跡は、予測追跡より高い優先順位を有する。

【0154】

上述の追跡法において、顔追跡は、それまでの追跡に一致しない顔が検出される毎に開始される。ここで、多くの誤検出が行われ、既存の規則の1つ（例えば、予測許容率閾値（prediction_acceptance_ratio_threshold）に関連する規則）によって最終的に拒絶される前に幾つかのフレームに亘って誤った追跡がなされることがある。

【0155】

また、追跡を拒絶するための既存の規則例えば、予測許容率閾値（prediction_acceptance_ratio_threshold）検出許容率閾値（detection_acceptance_ratio_threshold）等の変数に関する規則は、対象となる人が比較的長時間、頭を横に向けることによってバイアスされる。実際には、このような場合、顔の追跡を継続することが望ましい。

【0156】

以下、これらの問題の解決策について説明する。

【0157】

まず、この手法の第1の側面では、誤検出によって誤った追跡が開始されることを防ぐ。ここでも、内部的には、顔追跡は、既存の追跡に一致しないあらゆる顔検出毎に開始される。但し、顔追跡は、それはアルゴリズムから出力されない。この追跡を維持するためには、追跡の最初の f 個のフレームが顔検出（すなわち、タイプD）である必要がある。最初の f 個のフレームの全てがタイプDである場合、追跡は、維持され、顔の位置は、 f 個のフレーム分経過後に、アルゴリズムから出力される。

【0158】

最初の n 個のフレームの全てがタイプDではない場合、顔追跡は終了し、この追跡に関する顔位置は出力されない。

【0159】

f は、通常、2、3又は5に設定する。

【0160】

この手法の第2の側面では、低い検出許容率（detection_acceptance_ratio）によって追跡を終了することなく、プロファイル内の顔を長時間に亘って追跡することができる。これを実現するために、 $\pm 30^\circ$ の固有ブロックによって顔が照合された場合、予測許容率閾値（variables prediction_acceptance_ratio_threshold）及び検出許容率閾値（detection_acceptance_ratio_threshold）に関する検査を用いないこととする。これに代えて、顔追跡を維持するために、以下の評価基準を含める。

【0161】

顔追跡を維持するために n 個のフレーム毎に、 g 個の連続した顔検出を必要とする。

【0162】

20

30

40

50

g は、多くの場合、 f と同様の値、例えば 1 ~ 5 フレームに設定され、 n は、例えば 10 秒間（フレームレートに応じて 250 ~ 300 フレームに相当する）等、カメラから離れた人を追跡できることを望む最大のフレームの数に対応している。

【 0 1 6 3 】

また、この手法は、予測許容率閾値 (prediction_acceptance_ratio_threshold) 及び検出許容率閾値 (detection_acceptance_ratio_threshold) を用いた規則と組み合わせてもよい。これに代えて、回転毎に、例えば、追跡の最初からではなく、最後の 30 のフレームのみについて予測許容率閾値と検出許容率閾値を適用してもよい。

【 0 1 6 4 】

顔追跡を終了するための他の基準として、所謂「誤色閾値 (bad colour threshold)」を上回った場合がある。この検査では、検出された顔の位置は、（顔の認識の種類が顔検出、肌色又はカルマン予測のいずれであっても）、肌色を検査することによって検証される。ここでは、肌色からの距離が誤色閾値 (bad_colour_threshold) を超える場合、顔追跡を終了する。

【 0 1 6 5 】

この手法では、顔の肌色は、肌色追跡の間のみ確認される。これは、予測された顔位置を用いた場合は、非肌色の誤検出が追跡され、又は顔追跡が非肌色の位置に迷い込む可能性があることを意味する。

【 0 1 6 6 】

この問題は、顔の認識の種類が如何なるものであっても（顔検出、肌色又はカルマン予測のいずれであっても）、肌色を検査することにより解決される。ここでは、肌色からの距離（誤差）が誤色閾値 (bad_colour_threshold) を超える場合、顔追跡を終了する。

【 0 1 6 7 】

肌色追跡の間に算出された各画素の肌色からの距離を用いることにより、この処理を効率的に実現することができる。顔領域上（どの肌色追跡方式を用いるかに応じて、マスクの形をした領域、楕円状の領域、顔ウインドウの全体のいずれの領域であってもよい。）で平均されたこの測定値が所定の閾値を超える場合、顔追跡を終了する。

【 0 1 6 8 】

顔追跡を終了するための更なる基準として、変化が非常に低い又は非常に高い場合に顔追跡を終了してもよい。この手法については、図 22 a ~ 図 22 c に関する説明の後に述べる。

【 0 1 6 9 】

図 14 に図式的に示す追跡システムは、更なる 3 つの特徴を有している。

【 0 1 7 0 】

ショット境界データ (Shot boundary data) 560 (検査中の画像シーケンスに関連付けられるメタデータ又は図 2 のカメラ内で生成されたメタデータ) は、画像シーケンス内の各隣接する「ショット」の境界 (limits) を定義する。カルマンフィルタは、ショット境界においてリセットされ、その後のショットに対しては予測を行わないようになる。このような予測は無意味であるためである。

【 0 1 7 1 】

ユーザメタデータ 542 及びカメラ設定メタデータ 544 は、顔検出器 540 に入力される。また、これらは、非追跡システムにおいても用いることができる。カメラ設定メタデータの具体例は、上述した通りである。ユーザメタデータは、例えば以下の情報を含むことができる。

- ・ 番組の種類（例えば、ニュース、インタビュー、ドラマ）。
- ・ 例えば「遠くからのショット」、「カメラが接近する」（特定の種類のカメラショットに基づいて、顔のサイズのサブレンジが予測される）、各ショットに何人の人が含まれるか（これによっても、顔のサイズのサブレンジが予測される）等、ショットの詳細に関する脚本情報。

10

20

30

40

50

・スポーツ関連情報。スポーツは、多くの場合、固定されたカメラ位置から標準的なビュー及びショットを用いて撮像される。これらのメタデータにおいて特定することによって、顔のサイズのサブレンジを導き出すことができる。

【0172】

番組の種類は、画像又は画像シーケンス内に存在すると予想される顔の種類に関連する。例えば、ニュース番組では、画像シーケンスの大部分において、画面の（例えば）10%を占める単一の顔が登場することが予想される。スケールが異なる顔の検出は、このデータに応じて、このサイズに近い顔の可能性を高めるように重み付けすることができる。これに代えて、又はこれに加えて、検索範囲を削減し、全ての可能なスケールを探索するのではなく、所定のスケールのサブセットのみを探索するようにしてもよい。これにより、顔検出処理の処理負担を軽減することができる。したがって、ソフトウェアに基づくシステムでは、ソフトウェアをより速やかに実行でき及び／又はよりパワーが小さいプロセッサで実行できる。また、ハードウェアに基づくシステム（例えば、特定用途向け集積回路（application-specific integrated circuit：ASIC）を含む。）では、ハードウェアに対する要求を軽減できる。

【0173】

上述した他の種類のユーザメタデータを同様に適用してもよい。「予想される顔のサイズ」のサブレンジは、例えば、メモリ30内に格納されたルックアップテーブルに登録してもよい。

【0174】

例えばレンズ110のフォーカス及びズームに関するカメラメタデータについて、画像の前景に存在する可能性がある任意の顔の予想される画像サイズに関する初期的な指標を与えることにより、顔検出器を補助することができる。これに関して、フォーカス及びズームの設定は、カムコーダ100と、撮像されている個人との間の予想される距離、及びレンズ110の拡大率を定義する。これら2つの属性から、平均顔サイズに基づき、結果として得られる画像データにおける顔の予想されるサイズ（画素数）を算出することができ、これに基づいて探索すべきサイズのサブレンジを決定し、又は予想される顔サイズの重み付けを行うことができる。

【0175】

この構成は、ビデオ会議又は所謂デジタル看板（デジタルサイネージ環境：digital signage environment）における使用に適している。

【0176】

ビデオ会議システムでは、ユーザは、ビデオマテリアルを「個々の話者」、「2人の話者のグループ」、「3人の話者のグループ」等に分類でき、この分類に基づいて、顔検出器は、予想される顔のサイズを導出し、探し、画像内の1つ以上の顔を強調表示することができる。

【0177】

デジタルサイネージ環境では、ビデオ画面上に広告マテリアルを表示できる。顔検出は、広告マテリアルを見る人々の顔を検出するために使用される。

【0178】

追跡アルゴリズムの利点

顔追跡法は、以下のような3つの主な利点を有する。

- ・顔検出結果が得られないフレームにおいて、カルマンフィルタリング及び肌色追跡を用いることにより、見逃された顔を埋める（fill in）ことができる。これにより、画像シーケンス間に亘って、真の許容率を高めることができる。

- ・顔を連續的に追跡することにより、顔のリンクを提供できる。アルゴリズムは、将来のフレームにおいて検出された顔が同じ個人の顔であるか、他の個人の顔であるかを自動的に知ることができる。したがって、このアルゴリズムから、シーン内の顔の数やこれらの顔が存在するフレームに関する情報を含むシーンメタデータを容易に作成することができ、各顔の代表的な顔写真を作成することもできる。

10

20

30

40

50

・顔の誤検出は、画像間で連続することは希であるため、顔の誤検出率を低くすることができる。

【0179】

図19a～図19cは、ビデオシーケンスに適用される顔追跡を説明する図である。

【0180】

具体的には、図19aは、連続するビデオ画像（例えば、フィールド又はフレーム）810から構成されるビデオシーン800を図式的に示している。

【0181】

この具体例では、画像810は、1又は複数の顔を含んでいる。詳しくは、このシーン内の全ての画像810は、画像810の図式的表現内における左上に示す顔Aを含んでいる。更に、一部の画像810は、画像810の図式的表現内における右下に示す顔Bを含んでいる。

10

【0182】

この図19aに示すシーンに顔追跡処理を適用したとする。顔Aは、当然、シーン全体に亘って追跡される。1つの画像820においては、直接検出によつては顔は追跡されていないが、上述した色マッチング法及びカルマンフィルタリング法により、「見逃された(missing)」画像820の前後の両側について、検出が連続していることを示唆する。

図19bは、検出された、各画像内に顔が存在する可能性を示している。図19bに示すように、この可能性は、画像830において最も高く、したがつて、顔Aを含むと判定された画像の部分840を顔Aに関する「ピクチャスタンプ」として用いる。ピクチャスタンプについては、後により詳細に説明する。

20

【0183】

同様に、顔Bは、異なるレベルの信頼度で検出され、特に画像850は、顔Bが存在する可能性が最も高いと判定されている。したがつて、顔Bを含むと判定された対応する画像の一部（部分860）をこのシーンにおける顔Bのピクチャスタンプとして使用する。（勿論、これに代えて、画像のより広い部分又は画像全体をピクチャスタンプとして用いてもよい）。

【0184】

追跡された各顔について、単一の代表的な顔ピクチャスタンプが必要である。純粹に顔確率に基づいて顔ピクチャスタンプを出力した場合、必ずしも、最良の品質のピクチャスタンプが保証されるわけではない。最高の品質のピクチャスタンプを得るためにには、ピクチャスタンプと同じ解像度例えば、 64×64 の画素で検出される顔が選択されるようなバイアス又は誘導を行うとよい。

30

【0185】

最高の品質のピクチャスタンプを得るためにには、以下のスキームを適用できる。

(1) (色追跡／カルマン追跡ではなく) 検出された顔を用いる。

(2) 顔検出の間、確率が高かった、すなわち、少なくとも閾値確率を超えた顔を用いる。

(3) 可能な限り 64×64 画素に近い顔を用い、これにより、再スケーリングによるアーチファクトを低減し、画質を改善する。

40

(4) (可能であれば) 追跡の早い段階での顔すなわち、追跡シーケンスにおける所定の初期の部分（例えば、追跡シーケンスの最初から10%又は20のフレーム等）の顔を用いないこの期間は、顔がまだ遠くにある（すなわち、小さい）又は不鮮明である可能性が高いためである。

【0186】

これを実現するための幾つかの規則を以下に示す。

【0187】

顔検出毎に以下の処理を行う。

【0188】

まず、メトリック $M = \text{face_probability} * \text{size_weight}$

50

n_g を算出する。ここで、 $\text{size_weighting} = \text{MIN}((\text{face_size} / 64) ^ x, (64 / \text{face_size}) ^ x)$ であり、 $x = 0.25$ である。次に、M が最も大きい顔ピクチャスタンプを選ぶ。

【0189】

これにより、各顔サイズの顔確率について、以下のような重み付けがなされる。

【0190】

【表3】

顔サイズ (face_size)	サイズ重み (size_weighting)	
16	0.71	10
19	0.74	
23	0.77	
27	0.81	
32	0.84	
38	0.88	20
45	0.92	
54	0.96	
64	1.00	
76	0.96	
91	0.92	
108	0.88	30
128	0.84	
152	0.81	
181	0.77	
215	0.74	
256	0.71	
304	0.68	40
362	0.65	
431	0.62	
512	0.59	

【0191】

実際には、この処理は、ルックアップテーブルを用いて行うことができる。

【0192】

重み関数の影響を小さくするために、0.25乗より小さい乗数、例えば $x = 0.2$ 又 50

は 0 . 1 を用いてもよい。

【 0 1 9 3 】

この重み付け法は、顔追跡全体に適用してもよく、最初の N 個のフレームのみに適用してもよい（これらの N 個フレームからのサイズが小さい顔の集合に対してのみ重み付けを適用してもよい）。N は、例えば、最初の 1 ~ 2 秒（25 ~ 50 フレーム）に対応する数とすることができます。

【 0 1 9 4 】

更に、±30° 以上の角度で検出された顔に対し、正面で検出された顔を優先させてよい（又は他の如何なるポーズを優先してもよい）。

【 0 1 9 5 】

非線形編集装置の表示画面を図 20 に示す。

【 0 1 9 6 】

非線型編集装置は、広く用いられているものであり、例えば図 1 に示すような汎用コンピュータ装置上で実行されるソフトウェアプログラムとして実現される場合も多い。これらの編集システムにより、ビデオ、オーディオ及びその他のマテリアルを編集し、各媒体アイテム（例えば、ビデオショット）が捕捉された順序から独立したメディア作品を制作することができる。

【 0 1 9 7 】

図 20 に示す表示画面には、ビデオクリップを見る能够性 900 と、後に説明する一組のクリップアイコン 910 と、タイムライン 920 とが表示され、タイムライン 920 は、編集されたビデオショット 930 の表現が含まれ、各ショットは、任意項目として、そのショットのコンテンツを示すピクチャスタンプ 940 を含んでいる。

【 0 1 9 8 】

あるレベルにおいて、図 19 a ~ 図 19 c を用いて説明したように導き出されたピクチャスタンプを各編集されたショットのピクチャスタンプ 940 として用い、これにより、捕捉された元のショットより短くなる場合もある編集された長さのショット内において、最も高い確率値を有する顔検出を表すピクチャスタンプをタイムラインに挿入し、ショットからの代表画像を示すことができる。この確率値を、基本的な顔検出閾値より高く設定することもできる閾値と比較し、信頼性が高い顔検出のみに基づいてピクチャスタンプを作成するようにしてよい。編集されたショット内で 2 以上の顔が検出された場合、最も確率が高い顔を表示するようにしてよい、タイムライン上に 2 以上の顔のピクチャスタンプを表示するようにしてよい。

【 0 1 9 9 】

非線型編集装置のタイムラインは、多くの場合、スケーリング機能を有し、表示画面の全幅に対応する長さは、出力されるメディア作品について、様々な異なる時間を表すことができる。したがって、例えば、隣接する 2 つのショットの特定の境界をフレームの精度で編集する場合、タイムラインのスケールを「拡大」し、表示画面が、出力メディア作品について比較的短い時間を表すようにすることができる。一方、出力メディア作品の概観を視覚化するといった目的では、タイムラインのスケールを縮小し、表示画面の幅に亘って、より長い時間を表示できるようにするとよい。このように、タイムラインのスケールの拡大又は縮小に応じて、出力メディア作品を構成する編集された各ショットを表示するために用いることができる表示領域を広くしたり、狭くしたりしてもよい。

【 0 2 0 0 】

拡張されたタイムラインのスケールでは、出力メディア作品を構成する編集された各ショットについて、1 つのピクチャスタンプ（図 19 a ~ 図 19 c に関連して説明したように導き出された）を表示するため以上に十分な余地がある。しかしながら、タイムラインのスケールを縮小すると、ピクチャスタンプを表示できなくなる場合がある。このような場合、ショットを「シーケンス」としてグループ化し、各シーケンスが、フェーズピクチャスタンプ（phase picture stamp）を表示するのに十分大きな表示サイズを有するよう

10

20

30

40

50

にする。シーケンス内には、最も確率が高い顔のピクチャスタンプを選択し、表示する。シーケンス内に顔が検出されない場合、任意の画像をタイムライン上に表示してもよく、画像を表示しなくてもよい。

【0201】

図20は、更に、2個の「顔タイムライン(face timelines)」925、935を示している。この顔タイムライン925、935は、「メイン」のタイムライン920と同じスケールで表示される。各顔タイムライン925、935は、追跡された単一の顔に対応し、この追跡された顔を含む編集されたシーケンスの部分を示している。ここで、複数の顔が同じ個人のものであることがユーザには判るが、これらの顔が追跡アルゴリズムでは関連付けられていないといった場合がある。この場合、ユーザは、顔タイムライン925、935上の関連する部分を選択し(例えば、複数のアイテムを選択するための標準的なウィンドウズ(登録商標)の選択機能を用いて行う)、画面上の「リンク」ボタン(図示せず)をクリックすることにより、これらの顔を関連付けることができる。これにより、顔タイムライン925、935は、検出された顔のグループの関連付けを反映し、1つのより長く追跡された顔として表示する。図21a及び図21bは、クリップアイコン910'、910"の2つの表示例を示している。クリップアイコン910'、910"は、図20に示す表示画面に表示され、これによりユーザは、個別のクリップを選択して、タイムラインに含め、及びこれらのクリップの開始位置及び終了位置(イン点及びアウト点)を編集することができる。このように、各クリップアイコン910'、910"は、装置内に保存されている各クリップの全体を代表している。

10

20

30

40

50

【0202】

図21aにおいて、クリップアイコン910'は、単一の顔ピクチャスタンプ912と、テキストラベル領域914とによって表現されている。テキストラベル領域914は、例えば、クリップの位置及び長さを定義するタイムコード情報を含んでいる。また、図21bに示すように、マルチパートクリップアイコンを用いて、2以上の顔ピクチャスタンプ916をクリップアイコン910"に表示してもよい。

【0203】

クリップアイコン910の他の表示例として、「顔の要約(face summary)」を提供し、検出された全ての顔をクリップアイコン910の組として、それらの顔が登場する順序(ソーススマテリアルにおける順序であっても、編集された出力シーケンスにおける順序であってもよい)で表示してもよい。ここでも、同じ個人の顔でありながら、追跡アルゴリズムでは関連付けられていない複数の顔について、ユーザがこれらの顔が同一の顔であると主観的に判断する場合がある。この場合、ユーザは、関連する顔クリップアイコン910を選択し(例えば、複数のアイテムを選択するための標準的なウィンドウズ(登録商標)の選択機能を用いて行う)、画面上の「リンク」ボタン(図示せず)をクリックすることにより、これらの顔を関連付けることができる。これにより、顔タイムラインは、検出された顔のグループの関連付けを反映し、1つのより長く追跡された顔として表示する。

【0204】

更なる可能性として、クリップアイコン910によりハイパーリンクを提供してもよく、これによりユーザがアイコン910の1つをクリックすると、ビューワ領域900において、対応するクリップが再生されるようにしてよい。

【0205】

例えば、監視又は閉回路テレビジョン(closed circuit television:以下、CCTVという。)システムにおいても同様の技術を用いることができる。顔が追跡された場合は必ず、又は少なくとも所定のフレーム数の間、顔が追跡された場合は必ずその顔が追跡された連続したビデオの部分に関してクリップアイコン910と同様のアイコンが生成される。このアイコンは、図20に示すクリップアイコンにアイコンと同様に表示される。アイコンをクリックすると、(ビューワ領域900と同様のウィンドウにおいて)その特定の顔が追跡されたビデオの部分が再生される。なお、この手法によって複数の異なる顔を追跡してもよく、ビデオの対応する部分は、重複していても、完全に一致していてもよい。

。

【0206】

図22a～図22cは、傾斜前処理法（gradient pre-processing technique）を図式的に説明する図である。

【0207】

画素の変化が殆どない画像ウィンドウは、固有顔又は固有ブロックに基づく顔検出構成によって、顔として検出される傾向がある。したがって、前処理ステップにおいて、顔検出処理の対象から、画素の変化が殆どない領域を除外するとよい。多数の尺度を有する方式（上述参照）では、前処理ステップを各尺度において実行してもよい。

【0208】

基本的な処理として、画像全体に亘る可能なウィンドウ位置について、「傾斜検査（gradient test）」を適用する。各ウィンドウ位置について所定の画素位置、例えばウィンドウ位置の中心の又は中心に近い画素に対して、このウィンドウに適用された検査の結果に応じてフラグを立て又はラベルを付す。この検査により、このウィンドウにおいては画素の変化が殆どないと判定された場合、このウィンドウ位置は、顔検出処理から除外する。

【0209】

第1のステップを図22aに示す。図22aは、画像における任意のウィンドウ位置を示している。上述のように、前処理は、可能な各ウィンドウ位置に対して繰り返し行う。傾斜前処理は、ウィンドウ全体に対して行ってもよいが、本願発明者らは、図22aに示すように、前処理を検査ウィンドウ1010の中心領域1000に適用した方が良好な結果が得られることを見出した。

【0210】

図22bに示すように、傾斜に基づく測定値は、ウィンドウ（又は図22aに示すようなウィンドウの中心領域）から得られ、この値は、ウィンドウから得られた水平方向及び垂直方向の両方における隣接する画素1011間の絶対距離の平均値である。各ウィンドウの中止領域には、この傾斜に基づく測定値のラベルを付し、これにより画像の傾斜「マップ」を作成する。次に、作成された傾斜マップを傾斜閾値と比較する。そして、傾斜に基づく測定値が傾斜閾値を下回る全てのウィンドウ位置をその画像に関する顔検出処理から除外する。

【0211】

また、画素の変化又は平均画素値からの画素の絶対平均偏差等の他の傾斜に基づく測定を用いてもよい。

【0212】

傾斜に基づく測定は、画素の輝度値に関して行うことが好ましいが、カラー画像の他の画像成分について行ってもよいことは言うまでもない。

【0213】

図22cは、例示的な画像から導出された傾斜マップの具体例を示している。

【0214】

ここで、傾斜が緩い領域1070（ハッティングで示す）は、顔検出から除外し、より傾斜が急峻な領域1080のみを顔検出に用いる。

【0215】

上述の具体例は、顔検出（トレーニング及び検出段階を含む）、及びカメラレコーダ及び編集装置におけるこの顔検出の可能な用途に関する具体例を説明するものである。なお、この技術は、この他の多数の用途で利用することができ、例えば、以下に限定されるものではないが、安全監視システム、包括的なメディア処理（例えば、ビデオテープレコーダのコントローラ）、ビデオ会議システム等に利用することもできる。

【0216】

他の実施例においては、高い画素差分を有するウィンドウ位置にフラグ又はラベルを付してもよく、及び顔検出処理から除外してもよい。「高い」画素差分とは、図29bを用

10

20

30

40

50

いて上述した測定値が上側の閾値を超えていることを意味する。

【0217】

したがって、傾斜マップは、上述のように、作成される。傾斜測定値が上述した（第1の）傾斜閾値より低い位置は、傾斜閾値が上側の閾値より高いあらゆる位置と同様に、顔検出処理から除外される。

【0218】

上述のように、「下側の閾値」処理は、好ましくは検査ウィンドウ1010の中心部分1000に適用される。「上側の閾値」処理も同様に適用できる。これは、各ウィンドウ位置に関して、単一の傾斜測定値のみを導出すればよいことを意味する。これに代えて下側の閾値検査に関して全体のウィンドウが使用されている場合、上側の閾値検査に関しても同様に全体のウィンドウを用いることができる。ここでも、各ウィンドウ位置毎に単一の傾斜測定値だけを導出すればよい。勿論、（例えば）検査ウィンドウ1010の中心部分1000を用いて下側の閾値検査のための傾斜測定値を導出し、上側の閾値検査に関しては、完全な検査ウィンドウを用いる等、2つの異なる構成を用いてもよい。

【0219】

上述した顔追跡を終了するための更なる基準として、その変化又は傾斜測定値が非常に低く又は非常に高くなつた場合に顔追跡を終了してもよい。

【0220】

この手法では、追跡された顔位置は、関心領域マップからの偏差によって検証される。追跡の次の繰り返しのためには、検出された尺度における顔のサイズの領域のマップのみを各顔毎に保存する。

【0221】

上述した傾斜前処理を行つた場合でも、肌色追跡された顔又はカルマン予測された顔が画像内の変化が小さい又は大きい（顔らしくない）領域に移動する可能性はある。したがって、傾斜前処理において、既存の顔追跡の周りの領域に関する変化値（又は、傾斜値）を保存する。

【0222】

顔の次の位置に関する最終的な決定が下された場合、（顔検出、肌色追跡又はカルマン予測のうちのいずれの認識を用いてよい。）その位置は、関心領域マップ内に保存されている変化（又は傾斜）値を用いて検証される。その位置が非常に高い又は非常に低い変化値（又は傾斜値）を有すると判定された場合その位置は顔らしくないと判定され、顔追跡は終了する。これにより、顔追跡が画像内の低い（又は高い）変化値を有する背景領域に逸れることが防がれる。

【0223】

これに代えて傾斜前処理が使用されていない場合であつても、新たな顔位置の変化値を新たに計算してもよい。いずれの場合も、用いられる変化測定は、従来の変化値であつてもよく、隣接する画素の差分の合計（傾斜値）であつてもよく、又は他の如何なる種類の変化値であつてもよい。

【0224】

図23は、ビデオ会議システムの構成を示している。2つのビデオ会議ステーション1100、1110は、例えば、インターネット、ローカルエリア又はワイドエリアネットワーク、電話回線、高ビットレート専用回線、ISDN回線等であるネットワーク接続1120を介して接続される。各ビデオ会議ステーション1100、1110は、基本的には、カメラ及び関連する送信装置1130と、表示及び関連する受信装置1140とを備える。テレビ会議システムの参加者は、カメラに撮像され、各ステーションに表示され、参加者の声は、各ステーションにおける1つ以上のマイクロホン（図23には示していない）に入力される。オーディオ及びビデオ情報は、ネットワーク1120を介して他方のステーションの受信機1140に伝送される。この他方のステーションにおいて、カメラによって撮像された画像が表示され、参加者の声は、スピーカ等の装置から再生される。

【0225】

10

20

30

40

50

なお、ここでは、説明を簡潔にするために、2つのステーションを示しているが、2つ以上のステーションがテレビ会議システムに参加してもよい。

【0226】

図24は、1つのカメラ／送信装置1130を1つの表示／受信装置1140に接続する1つのチャンネルを示している。

【0227】

カメラ／送信装置1130は、ビデオカメラ1150と、上述した技術を用いた顔検出器1160と、画像プロセッサ1170と、データフォーマッタ及び送信機1180とを備える。マイクロホン1190は、参加者の声を検出する。

【0228】

オーディオ、ビデオ、及び（オプションとして）メタデータ信号は、フォーマッタ及び送信機1180からネットワーク接続1120を介して表示／受信装置1140に送信される。また、ネットワーク接続1120を介して表示／受信装置1140から制御信号を受信してもよい。

【0229】

表示／受信装置は、例えば、表示画面及び関連する電子回路を含む表示及び表示プロセッサ1200と、ユーザ操作子1210と、例えばデジタル・アナログ（変換器）、増幅器及びスピーカを含むオーディオの出力構成1220とを備える。

【0230】

括的に言えば、顔検出器1160は、カメラ1150によって撮像された画像内の顔を検出（及び任意の機能として追跡）する。顔検出は、制御信号として画像プロセッサ1170に供給される。画像プロセッサは、以下に説明するように、様々な異なる手法で動作させることができるが、基本的には、画像プロセッサ1170は、カメラ1150によって撮像された画像をネットワーク1120を介して送信する前に処理する。この処理の主な目的は、ネットワーク接続1120の帯域幅又はビットレートを有効に活用することである。ここで、殆どの商業用途において、テレビ会議システムに適するネットワーク接続1120のコストは、ビットレートの要求に伴って高くなる。フォーマッタ及び送信機1180は、画像プロセッサ1170からの画像と、マイクロホン1190からの（例えば、アナログ・デジタル変換器（ADC）を介して、変換された）オーディオ信号と、オプションとして、画像プロセッサ1170によって行われた処理の性質を定義するメタデータとを結合する。

【0231】

以下、ビデオ会議システムの様々な動作モードについて説明する。

【0232】

図25は、更なるビデオ会議システムの構成を示す図である。ここで、顔検出器1160、画像プロセッサ1170、フォーマッタ及び送信機1180、表示及び表示プロセッサ1200の処理機能は、プログラミング可能なパーソナルコンピュータ1230によって実現される。表示画面（1200の一部）に表示されている画面は、図31を用いて後に説明する顔検出を用いたビデオ会議の1つの可能なモードを示しており、このモードでは、顔を含んでいる画像部分のみが一方の場所から他方の場所に送信され、この他方の場所において、タイル形式又はモザイク形式で表示される。この動作モードについては、後に説明する。

【0233】

図26は、図23～図25に示すシステムにおける動作のフローチャートである。図26、図28、図31、図33及び図34のフローチャートでは、カメラ／送信側（1130）によって実行される動作と表示／受信側（1140）で行われる動作とを区別して示している。

【0234】

まず、図26のステップ1300において、カメラ1150が画像を撮像する。ステップ1310において、顔検出器1160は、撮像された画像内の顔を検出する。理想的に

は、（上述のように）顔追跡は、顔検出における誤った中断を回避すること及びビデオ会議セッション中に特定の人物の顔を同様に取り扱うために用いられる。

【0235】

ステップ1320において、画像プロセッサ1170は、顔検出情報に応じて、撮像された画像をクロッピングする。このクロッピングは、以下のようにして行われる。

- ・まず、顔検出器1160によって検出された最も左上の顔を特定する。
- ・その顔の最も左上の極値(extreme)を検出する。これにより、クロッピングされる画像の左上角が定義される。
- ・最も右下の顔と、その顔の最も右下の極値について同様の処理を繰り返し、これにより、クロッピングされる画像の右下角が定義される。
- ・これらの2つの座標に基づいて、画像を長方形状に切り取る。

【0236】

そして、ステップ1320において、データフォーマッタ及び送信機1180によって、このクロッピングされた画像を送信する。この具体例では、追加的なメタデータを送信する必要はない。このように画像をクロッピングすることにより、フル画像画像を伝送する場合に比べて、ビットレートを低減することができ、又は同じビットレートを用いた場合の伝送品質を改善することができる。

【0237】

ステップ1130において、受信機側では、クロッピングされた画像がフルスクリーン画面で表示される。

【0238】

また、ユーザ操作子1210を用いて、画像をクロッピングするモードと、画像をクロッピングしないモードとの間で画像プロセッサ1170を切り換えるようにしてもよい。これにより、参加者は、受信端末において、相手側の部屋全体を表示させるか、画像において顔に関連する部分のみを表示させるかを任意に選択することができる。

【0239】

画像をクロッピングするための他の手法として、以下のようないくつかの処理を行ってもよい。

- ・最も左及び最も右の顔を特定する。
- ・ショットのアスペクト比を維持したまま、ピクチャの上半分の位置に顔が表示されるようにする。

【0240】

クロッピングの変形例として、カメラをズームし、これにより、送信された画像において、検出された顔をより大きく強調するようにしてもよい。この手法は、例えば、最終的な画像に関するビットレート削減法と組み合わせて行ってもよい。これを実現するために、画像プロセッサは、（図24の点線1155によって示すように）カメラの方向の制御（パン／チルト）及びレンズズームを制御することができる。

【0241】

図27a及び図27bは、図26のフローチャートに関連する画像の具体例を示している。詳しくは、図27aは、カメラ1150によって撮像されたフルスクリーン画像を示し、図27bは、その画像のズームされたバージョンを示している。

【0242】

図28は、図23～図25に示すシステムの他の動作モードを示すフローチャートである。ステップ1300は、図26のステップ1300と同じである。

【0243】

ステップ1340において、撮像された画像における各顔が特定され、例えば、表示すべき顔の周りにボックスを描画することによって強調される。また、各顔には、例えば、任意のラベルa、b、c…によってラベルが付される。ここで、顔追跡は、ラベルが付された顔が後に混同されることを回避するために特に有用である。ラベルが付された画像は、フォーマットされ、受信機に送信され、ステップ1350において、表示される。ステップ1360において、ユーザは、例えば、その顔に関連するラベルをタイプするこ

10

20

30

40

50

とによって表示するべき顔を選択する。この選択に基づく情報は、制御データとして画像プロセッサ 1170 に渡され、画像プロセッサ 1170 は、ステップ 1370 において、この要求された顔を分離する。この要求された顔は、受信機に送信される。この要求された顔は、ステップ 1380 において、表示される。ユーザは、ステップ 1360 において、異なる顔を選択し、現在表示されている顔をこれに置き換えることができる。このような構成においては、選択画像は、表示すべき顔を選択するためだけに用いられるので、より低いビットレートで伝送でき、したがって、帯域幅を削減することができる。これに代えて、上述の場合と同様、選択された個々の顔をより高いビットレートで伝送し、より高い画質を実現してもよい。

【0244】

10

図 29 は、図 28 のフローチャートに関連する画像の具体例を示している。ここでは、3つの顔が特定され、それぞれ a、b、c のラベルが付されている。ユーザは、これらの3つの文字の1つをユーザ操作子 1210 を用いてタイプすることによって、これらの顔の1つを選択でき、フルスクリーン表示させることができる。この処理は、上述と同様、メイン画像をクロッピングすることによって、又はカメラを顔にズームさせることによって実現することができる。図 30 は、受信機における選択のためのメニューとして、各顔の所謂サムネイル画像を表示する変形例を示している。

【0245】

図 31 は、図 23～図 25 に示すシステムの更なる動作モードを示すフローチャートである。ステップ 1300 とステップ 1310 は、図 26 のステップ 1300 とステップ 1310 と同様である。

20

【0246】

ステップ 1400 において、画像プロセッサ 1170 フォーマッタ及び送信機 1180 は、協働して、撮像された顔に関連するサムネイル画像のみを送信する。これらのサムネイル画像は、ステップ 1410 において、受信端末において、顔のメニュー又はモザイクとして表示される。オプションとして、ユーザは、ステップ 1420 において、1つの顔のみを選択し、拡大して表示させてもよい。ここで、他の顔を同じ画面上により小さいフォーマットで表示してもよく、又は拡大された画面が使用されている間、他の顔を表示しないようにしてもよい。すなわち、この構成と図 28 に示す構成との違いは、全ての顔に関するサムネイル画像が受信機に送信され、受信機側で、サムネイルをどのように表示するかに関する選択が行われるという点である。

30

【0247】

図 32 は、図 31 のフローチャートに関連する画像の具体例を示している。なお、初期画面では、3つのサムネイル 1430 が表示されるが、図 32 に示している段階では、参加者 c の顔が選択され、表示画面の左側に拡大して表示されている。但し、他の参加者に関するサムネイルも消されることなく残され、これにより、ユーザは、次に拡大して表示する画像を任意に選択することができる。

【0248】

40

なお、少なくともメイン画像をクロッピングするシステムにおいて、これらの実施例に示すサムネイル画像は、システムに存在する何らかの処理遅延を含んではいるが、「生の」サムネイル画像である。すなわち、撮像される参加者の画像が時間的に変化すれば、サムネイル画像もこれに応じて変化する。カメラズームを用いるシステムにおいては、サムネイルは、静止画であってもよく、あるいは、第 2 のカメラを用いてより広いアングルのシーンを撮像してもよい。

【0249】

図 33 は、更なる動作モードを示すフローチャートである。ステップ 1300 とステップ 1310 は、図 26 のステップ 1300 とステップ 1310 と同様である。

【0250】

50

ステップ 1440 において、有効なマイクロホンに最も近いことが検出された顔に関するサムネイル顔画像が送信される。勿論、この処理は、1つ以上のマイクロホンに基づ

いて行ってもよく、及び予め行われた選択又はどの参加者がどのマイクロホンの近くに座っているかに関するメタデータに基づいて行ってもよい。これは、それぞれのビデオ会議ステーションにおいて、簡単なメニュー形式の入力項目をユーザが選択することによって予め設定してもよい。有効なマイクロホンとは、例えば、ある期間（例えば、1秒間）に亘って平均化された最大の振幅のオーディオ信号を有するマイクロホンである。ここで、低域通過フィルタを用いて、例えば、咳や物を落としたときの雑音、又は2人の参加者が同時に発声した場合等に有効なマイクロホンが頻繁に交代してしまうことを回避することもできる。

【0251】

ステップ1450において、送信された顔が表示される。ステップ1460においては 10
、現在の有効なマイクロホンが随時検出される。

【0252】

この検出は、例えば、単一の有効なマイクロホンの検出であってもよく、又はこれに代えて、単純な三角測量法を用いて、複数のマイクロホンに基づいて、話者の位置を検出してもよい。

【0253】

図34は、更に他の動作モードを示すフローチャートであり、ここでもステップ130
0とステップ1310は、図26のステップ1300とステップ1310と同様である。

【0254】

ステップ1470において、撮像された画像における各顔及びその周囲の部分は、高い
解像度で、送信され、背景（撮像された画像の他の部分）は、低い解像度で、送信される。
これにより、ビットレートを有効に低減でき、又は顔及びその周辺の画像を強調する
ことができる。オプションとして、各顔の位置を定義するメタデータを送信してもよく、又
は、受信機において、異なる画像の部分の解像度に基づいて、顔の位置を判定してもよい
。

【0255】

ステップ1480では、受信端末において、画像が表示され、ユーザによる選択のため
に各顔には、オプションとしてラベルが付される。ステップ1490において、図32を
用いて説明した構成と同様に、ユーザが選択した顔が拡大されて表される。

【0256】

図23～図34では、ビデオ会議システムに関連させて本発明を説明したが、本発明は
、例えば安全監視（CCTV）システムにも同様に適用できる。この場合、通常、帰路チ
ャンネルは必要ではなく、図24に示す構成におけるカメラ／送信側構成は、CCTVカ
メラ側に設けられ、受信／表示構成は、監視設備側に設けられ、これらは、ビデオ会議に
ついて説明した構成と同様の手法で実現される。

【0257】

また、上述した本発明の具体例の少なくとも一部は、ソフトウェアにより制御されるデ
ータ処理装置によって実現できる。例えば、上述した1又は複数の要素は、ソフトウェア
により制御される汎用データ処理装置、又は特定用途向け集積回路やフィールドプログラ
マブルゲートアレー等の専用プログラムにより制御されるデータ処理装置によって実現し
てもよい。このようなソフトウェア又はプログラム制御を提供するコンピュータプログラ
ム、及びこのようなコンピュータプログラムを保存、伝送又は提供する保存、伝送又は伝
送媒体も本発明の側面とみなすことができる。

【0258】

参考文献のリスト及び付録を以下に示す。なお、疑いの余地をなくすために言及すれば
、これらのリスト及び付録は、本願の一部を構成する。これらの文献は、引用により本願
に援用されるものとする。

【0259】

参考文献

1. エイチ・シュナイダーマン (H. Schneiderman) 及びティー・カナデ (T. Kanade) 著 50

、「顔及び車に適用される3Dオブジェクト検出のための統計モデル（A statistical model for 3D object detection applied to faces and cars）」、コンピュータビジョン及びパターン検出に関するIEEEコンファレンス2000（IEEE Conference on Computer Vision and Pattern Detection, 2000）

2. エイチ・シュナイダーマン（H. Schneiderman）及びティー・カナデ（T. Kanade）著、「オブジェクト検出のための局的外観及び空間的関係の確率的モデリング（Probabilistic modelling of local appearance and spatial relationships for object detection）」、コンピュータビジョン及びパターン検出に関するIEEEコンファレンス1998（IEEE Conference on Computer Vision and Pattern Detection, 1998）

3. エイチ・シュナイダーマン（H. Schneiderman）著、「顔及び車に適用される3Dオブジェクト検出のための統計的手法」、カーネギーメロン大学（Carnegie Mellon University）ロボティクス研究科（Robotics Institute）博士論文、2000年 10

4. イー・ヘルマス（E. Hjelmas）及びビー・ケー・ロウ（B.K. Low）著、「顔検出：概観（Face Detection: A Survey）」、コンピュータビジョン及び画像理解（Computer Vision and Image Understanding）第83号、pp. 236～274、2001年

5. エム・エイチ・ヤング（M.-H.Yang）、ディー・クレイグマン（D.Kriegman）及びエヌ・アフジャ（N.Ahuja）著、「画像における顔の検出：概観（Detecting Faces in Images: A Survey）」、パターン分析及びマシン知能に関するIEEEトランザクション（IEEE Trans. on Pattern Analysis and Machine Intelligence）、第24巻1号pp. 34～58、2002年1月 20

付録A：トレーニング顔セット

1個のデータベースは、屋内の背景の前に立つ数千個の題材（subject）の画像を格納している。上述の手法の実験に用いられる他のトレーニングデータベースは、正面から左右までの様々な範囲の角度から見た人間の頭の8ビットのグレースケール画像を一万個以上格納している。当業者にとっては明らかであるが、これらとは異なる様々なトレーニングセットを用いてもよく、任意として、ローカルの母集団の顔の特徴を反映するようプロファイルされたトレーニングセットを用いてもよい。

【0260】

付録B：固有ブロック

顔検出及び認識のための固有顔法（図4及び図5参照）において、各 $m \times n$ 顔画像は、それぞれ長さ $m n$ のベクトルによって表現されるように記録される。これにより、各画像は、 $m n$ 次元空間内の点とみなすことができる。画像の組は、この大きな空間内の点の集合にマッピングされる。 30

【0261】

顔画像は、全体的な構成が類似しており、この $m n$ 次元画像空間においてランダムには分布せず、すなわち、顔画像は比較的低い次元のサブ空間によって記述できる。主成分分析（principal component analysis：以下、PCAという。）を用いて、画像空間内における顔画像の分布を最もよく説明するベクトルを見出すことができる。PCAでは、元の顔画像に対応する共分散行列の主固有ベクトル（principal eigenvectors）を判定する。これらのベクトルは、多くの場合顔空間とも呼ばれる、顔画像のサブ空間を定義する。各ベクトルは、 $m \times n$ 画像を表し、元の顔画像の一次結合である。これらのベクトルは、元の顔画像に対応する共分散行列の固有ベクトルであるため、及びこれらは顔に似た外観を有するため、これらのベクトルは、固有顔とも呼ばれる【4】。 40

【0262】

未知の画像が存在する場合、この画像は顔空間に射影される。これにより、この画像は、固有顔の重み付けされた和として表現される。

【0263】

この具体例では、近接相関法（closely related approach）を用いて、顔画像のブロックに関連する所謂「固有ブロック（eigenblock）」又は固有ベクトルを生成及び適用する。ブロックのグリッドは、（トレーニングセット内の）顔画像又は（検出段階の間に）検 50

査ウィンドウに適用され、固有顔処理に非常に類似した固有ベクトルに基づく処理を各ブロック位置に適用する（又は、データ処理の負荷を軽減する他の具体例においては、この処理は、まず、ブロック位置のグループに対して行い、全てのブロック位置において使用される固有ブロックの1つの組を生成する）。当業者には明らかであるが、中央のブロック等の幾つかのブロックは、画像の鼻の部分を表している場合が多く、顔が存在するか否かを判定するためのより重要な部分となる。

【0264】

固有ブロックの算出

固有ブロックは、以下のような手順で算出される。

(1) 画像のトレーニングセットを用いる。これらを、それぞれ $m \times n$ のサイズを有する画像ブロックに分割する。各ブロック位置について、それぞれが各画像におけるそのブロック位置から抽出された、次のような画像ブロックの組が得られる。10

【0265】

【数17】

$$\left\{ I_o^t \right\}_{t=1}^{N_T}$$

【0266】

(2) 正規化されたブロックのトレーニングセット20

【0267】

【数18】

$$\left\{ I^t \right\}_{t=1}^{N_T}$$

【0268】

を以下のようにして算出する。

【0269】

元のトレーニングセットからの各画像ブロック I_o^t を 0 と、1 の L2 ノルムとの平均値を有するように正規化し、各正規化された画像ブロック I^t を生成する。30

【0270】

各画像ブロック I_o^t , $t = 1, \dots, N_T$:

【0271】

【数19】

$$I' = \frac{I_o^t - \text{mean}_I}{\|I_o^t - \text{mean}_I\|}$$

ここで、40

$$\text{mean}_I = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n I_o^t[i, j]$$

及び

$$\|I_o^t - \text{mean}_I\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n (I_o^t[i, j] - \text{mean}_I)^2}$$

【0272】50

(すなわち、($I_0^t - \text{mean}_I I_0^t$) の L₂ ノルム)

(3) ベクトルのトレーニングセット

【 0 2 7 3 】

【 数 2 0 】

$$\{x^t\}_{t=1}^{N_T}$$

【 0 2 7 4 】

を各画像ブロック I_t の画素要素を辞書順に並べ替えることによって生成する。すなわち 10
、各 $m \times n$ 画像ブロック I_t を長さ $N = mn$ のベクトル x^t に並べ替える。

(4) 偏差ベクトルの組

【 0 2 7 5 】

【 数 2 1 】

$$D = \{x^t\}_{t=1}^{N_T}$$

【 0 2 7 6 】

を算出する。D は、N 個の列と N_T 個の行とを有する。

(5) 共分散行列 を算出する。

【 0 2 7 7 】

【 数 2 2 】

$$\Sigma = DD^T$$

【 0 2 7 8 】

は、 $N \times N$ のサイズを有する対称行列である。

(7) 以下の式を解くことにより、全ての固有ベクトルからなる組 P 及び共分散行列 の
固有値 i , $i = 1, \dots, N$ を求める。 30

【 0 2 7 9 】

【 数 2 3 】

$$\Lambda = P^T \Sigma P$$

【 0 2 8 0 】

ここで、P は、その対角に沿って（大きさの順に）固有値 i を有する $N \times N$ 対角行列で
あり、P は、それぞれ長さ N を有する N 個の固有ベクトルの組を含む $N \times N$ 行列である。
この分解は、カルーネン・レーベ変換 (Karhunen-Loeve Transform: KLT) としても知
られている。 40

【 0 2 8 1 】

固有ベクトルは、顔画像のブロック間の変化を互いに特徴付ける特徴の組であるとみな
すことができる。固有ベクトルは直交基底を形成し、これにより画像ブロックを表現する
ことができ、すなわち、原理的には、如何なる画像も固有ベクトルの重み付けされた和に
よって、誤差を生じることなく表現することができる。

【 0 2 8 2 】

画像空間内のデータ点の数（トレーニング画像の数）が空間の次元より少ない ($N_T < N$) 場合、意味がある固有ベクトルの数は N_T 個のみとなる。残りの固有ベクトルは、0
の固有値に関連付けられた値を有する。そこで、多くの場合、 $N_T < N$ であるため、 $i > N_T$
である全ての固有値は 0 となる。 50

【0283】

更に、トレーニングセット内の画像ブロックは、全体の構成において類似する（全てが顔の画像に由来する）ため、残りの固有ベクトルのうちの幾つかのみが、画像ブロック間の相異を強く特徴付ける。これらは、関連する固有値が最も大きい固有ベクトルである。関連する固有値が小さいこの他の固有ベクトルは、このような大きな相異を特徴付けることなく、したがって、顔を検出し、又は顔間の区別を行う目的では有用ではない。

【0284】

したがって、PCAにおいては、 $M < N_T$ として、固有値が大きいM個の主固有ベクトルのみを考慮し、すなわち、部分的LTTを実行する。すなわち、PCAでは、最も大きい固有値に対応するLTT基底の、より低次元のサブ空間を抽出する。

10

【0285】

主成分は顔画像間の大きな相異を記述するため、外観上、これらの主成分は顔ブロックの部分に類似し、ここでは、これらを固有ブロックと呼ぶ。なお、固有ベクトルという用語も同様に用いる。

【0286】

固有ブロックを用いた顔検出

未知の画像の顔への類似性、すなわち顔類似性（faceness）は、その画像が顔空間によってどれ程適切に表現できるかによって測定される。この処理は、トレーニング処理において用いたものと同じブロックのグリッドを用いて、ブロック毎に行う。

20

【0287】

この処理の第1段階では、画像を顔空間に射影する。

【0288】

画像の顔空間への射影

画像を顔空間に射影する前に、トレーニングセットに対して行った前処理と略々同様の前処理をこの画像に対して行う。

30

【0289】

【数24】

$$I = \frac{I_o - \text{mean}_I}{\|I_o - \text{mean}_I\|}$$

ここで、

$$\text{mean}_I = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n I_o[i, j]$$

及び

$$\|I_o - \text{mean}_I\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n (I_o[i, j] - \text{mean}_I)^2}$$

40

【0290】

（すなわち、 $(I_o - \text{mean}_I)$ のL2ノルム）

（3）画像の画素要素を辞書順に並べ替えることによって、偏差ベクトルを生成する。画像を長さ $N = m n$ の偏差ベクトル x に並べ替える。

【0291】

これらの前処理の後、以下のような単純な処理により、偏差ベクトル x を顔空間に射影

50

する。

(4) 顔空間への射影は、偏差ベクトル x をその固有ブロック成分に変換する処理を含む。この処理は、M 個の主固有ベクトル（固有ブロック） P_i , $i = 1, \dots, M$ による単純な乗算を含む。各重み y_i は、以下のようにして求めることができる。

【0292】

【数25】

$$y_i = P_i^T x$$

10

【0293】

ここで P_i は、 i 番目の固有ベクトルを表す。

【0294】

これらの重み y_i , $i = 1, \dots, M$ は、入力顔ブロックの表現における各固有ブロックの貢献を表す。

【0295】

外観が類似するブロックは、類似する重みの組を有し、外観が相異するブロックは、相異なる重みの組を有する。すなわち、ここでは、重みは、顔検出において顔ブロックを分類する特徴ベクトルとして用いている。

【図面の簡単な説明】

20

【0296】

【図1】顔検出装置及び／又は非線型編集装置として用いられる汎用コンピュータシステムの構成を示す図である。

【図2】顔検出に用いるビデオカメラ・レコーダ（カムコーダ）の内部構成を示す図である。

【図3】トレーニング処理を説明する図である。

【図4】検出処理を説明する図である。

【図5】特徴ヒストグラムを示すグラフ図である。

【図6】固有ブロックを生成するためのサンプリング処理を説明する図である。

【図7】固有ブロックの組を示す図である。

30

【図8】固有ブロックの組を示す図である。

【図9】ブロック位置を表すヒストグラム表現を構築する処理を説明する図である。

【図10】ヒストグラムビン番号の生成を説明する図である。

【図11】顔確率の算出を説明する図である。

【図12a】上述の手法を用いて生成されたヒストグラムの具体例を示すグラフ図である。

【図12b】上述の手法を用いて生成されたヒストグラムの具体例を示すグラフ図である。

【図12c】上述の手法を用いて生成されたヒストグラムの具体例を示すグラフ図である。

40

【図12d】上述の手法を用いて生成されたヒストグラムの具体例を示すグラフ図である。

【図12e】上述の手法を用いて生成されたヒストグラムの具体例を示すグラフ図である。

【図12f】上述の手法を用いて生成されたヒストグラムの具体例を示すグラフ図である。

【図13】図13a～13gは、所謂多尺度顔検出を説明する図である。

【図14】顔追跡アルゴリズムを説明する図である。

【図15】図15a及び15bは、肌色検出に使用される探索領域の抽出を説明する図である。

50

【図16】肌色検出に適用されるマスクを示す図である。

【図17】図17a～17cは、図16に示すマスクの用途を説明する図である。

【図18】距離マップを示す図である。

【図19】図19a～19cは、ビデオシーケンスに適用される顔追跡を説明する図である。

【図20】非線形編集装置の表示画面を示す図である。

【図21】図21a及び21bは、クリップアイコンの表示例を示す図である。

【図22】図22a～22cは、傾斜前処理法を図式的に説明する図である。

【図23】ビデオ会議システムの構成を示す図である。

【図24】ビデオ会議システムの構成をより詳細に示す図である。 10

【図25】ビデオ会議システムの構成をより詳細に示す図である。

【図26】図23～図25に示すシステムにおけるある動作モードのフローチャートである。

【図27】図27a及び27bは、図26のフローチャートに関連する画像の具体例を示す図である。

【図28】図23～図25に示すシステムにおける他の動作モードのフローチャートである。 20

【図29】図28のフローチャートに関連する画像の具体例を示す図である。

【図30】図28のフローチャートに関連する画像の具体例を示す図である。

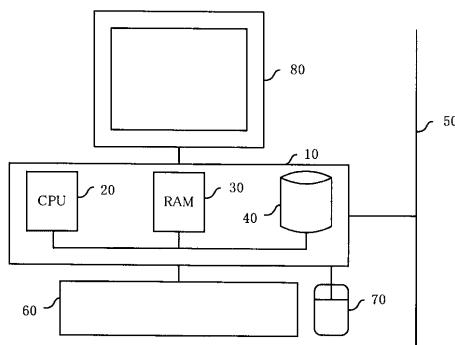
【図31】図23～図25に示すシステムにおける他の動作モードのフローチャートである。 20

【図32】図31のフローチャートに関連する画像の具体例を示す図である。

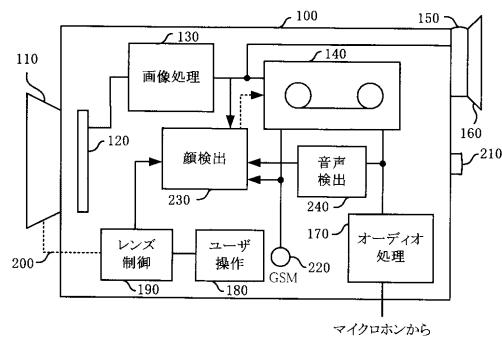
【図33】図23～図25に示すシステムにおける他の動作モードのフローチャートである。

【図34】図23～図25に示すシステムにおける他の動作モードのフローチャートである。

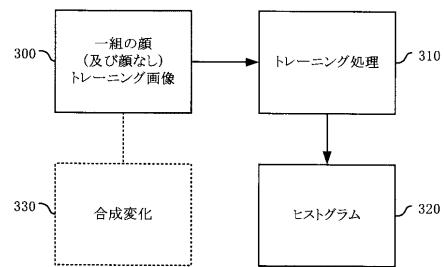
【図1】



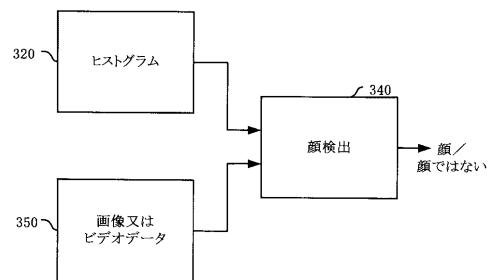
【図2】



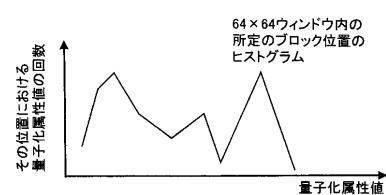
【図3】



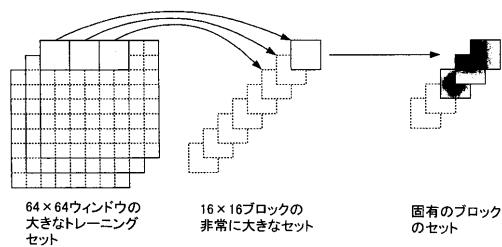
【図4】



【図5】



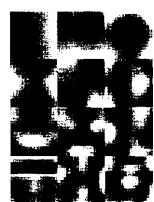
【図6】



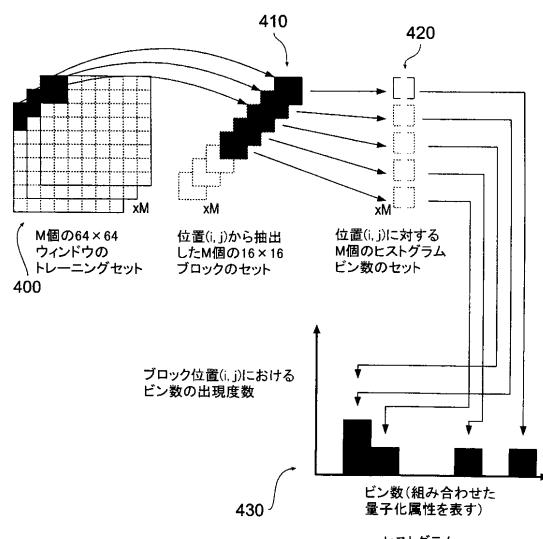
【図7】



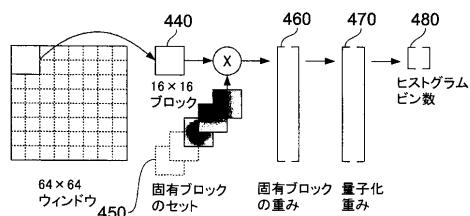
【図8】



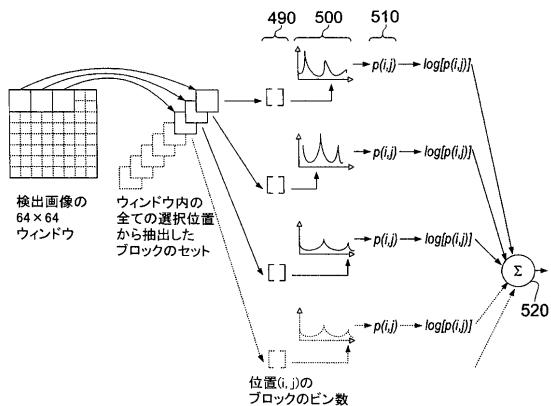
【図9】



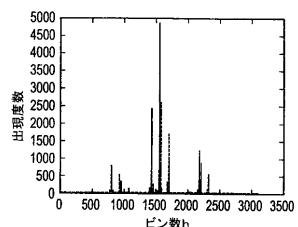
【図10】



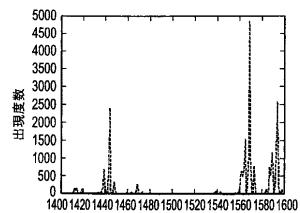
【図11】



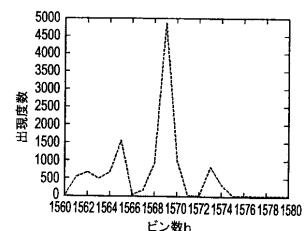
【図12 a】



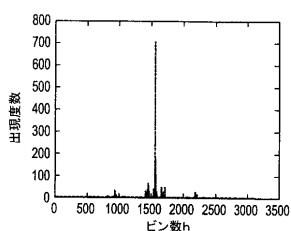
【図12 b】



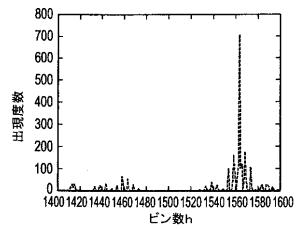
【図12 c】



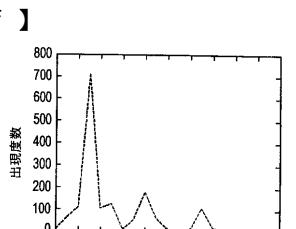
【図12 d】



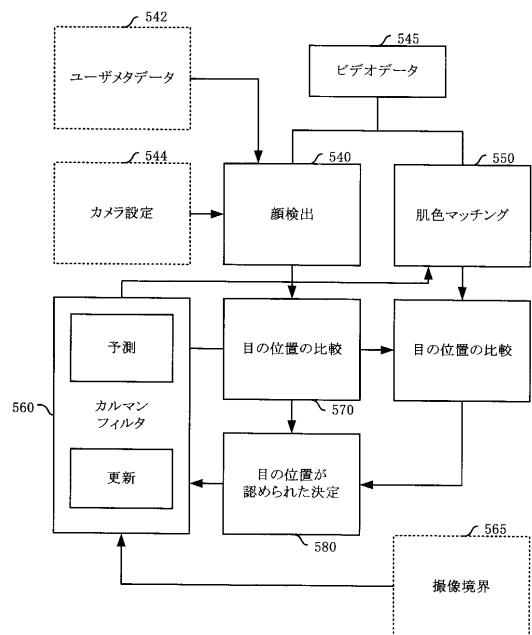
【図12 e】



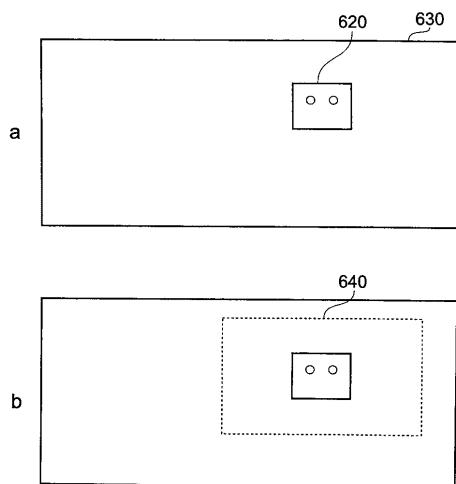
【図12 f】



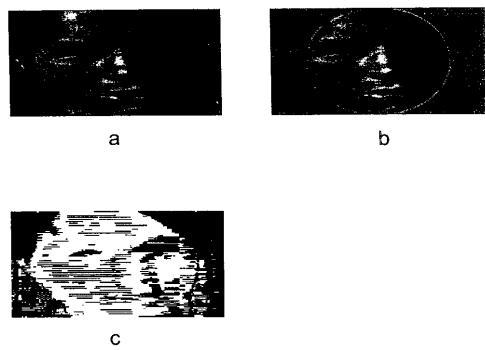
【図14】



【図15】



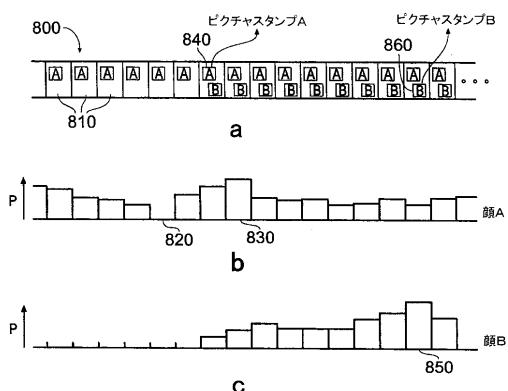
【図17】



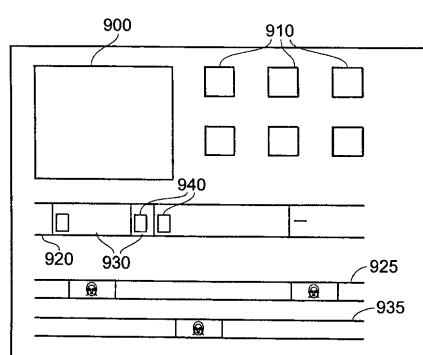
【図18】



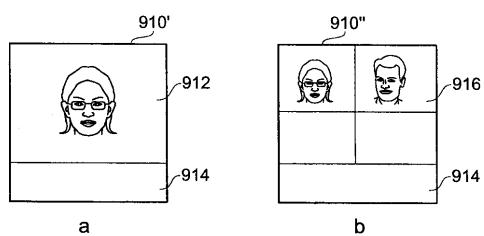
【図19】



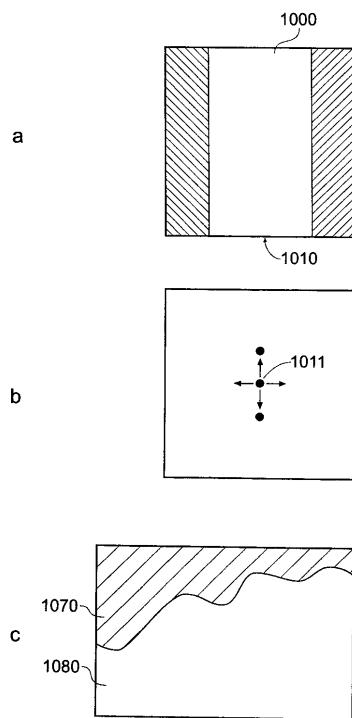
【図20】



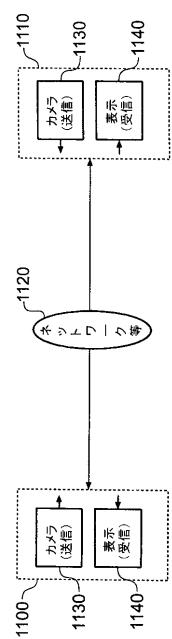
【図21】



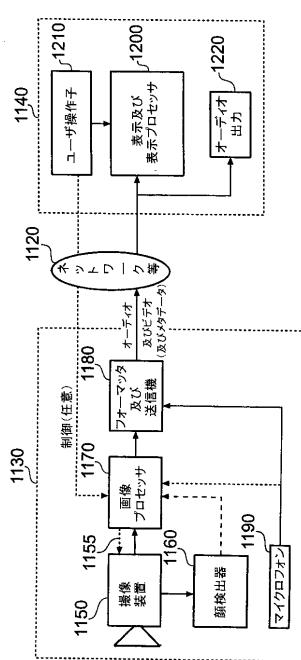
【図22】



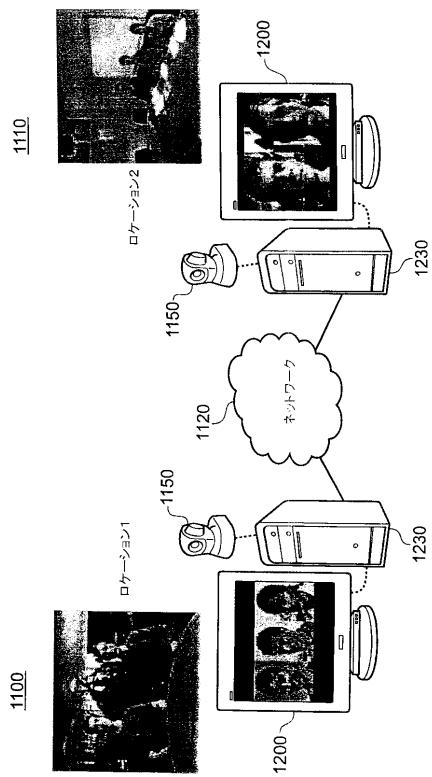
【図23】



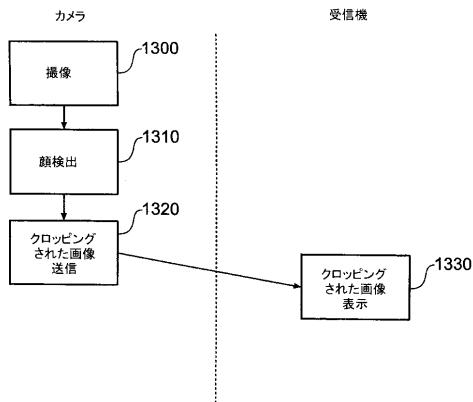
【図24】



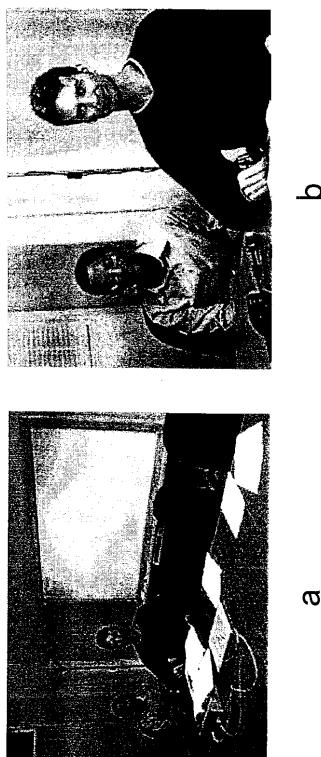
【図25】



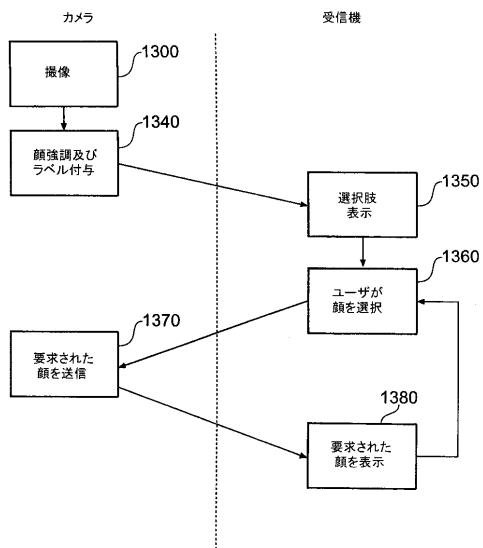
【図26】



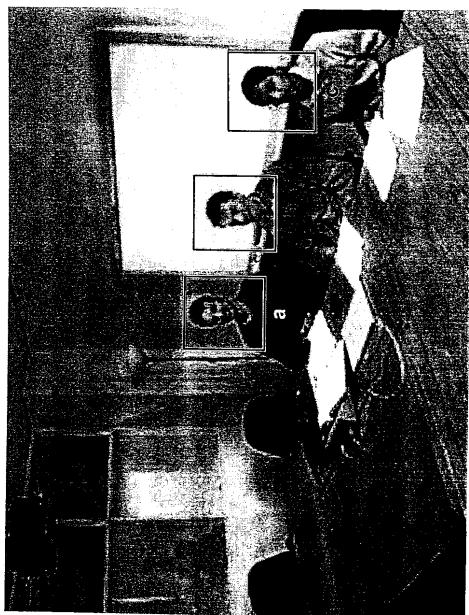
【図27】



【図28】



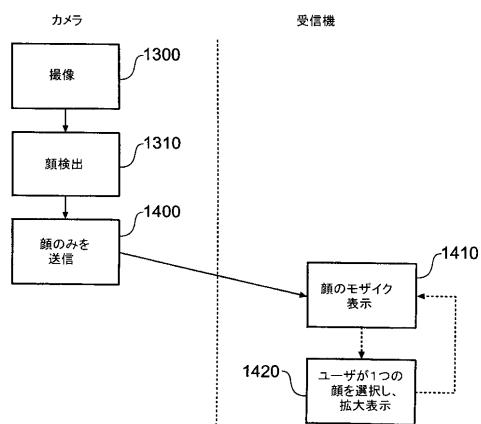
【図29】



【図30】



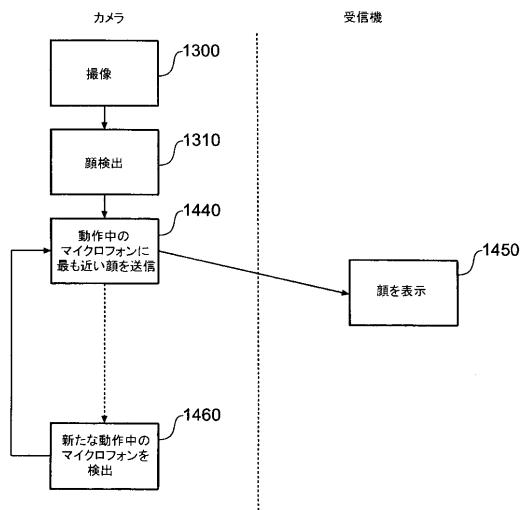
【図31】



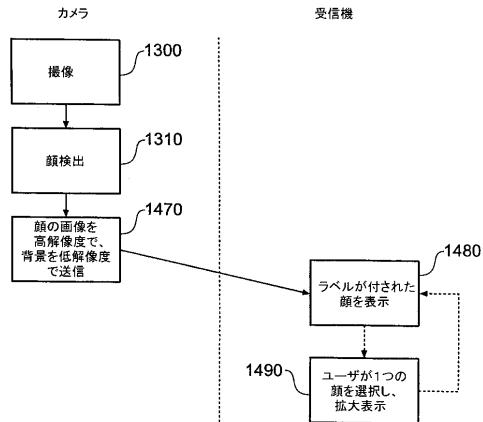
【図32】



【図3-3】



【図3-4】



【国際調査報告】

INTERNATIONAL SEARCH REPORT		International Application No PCT/GB 03/05224
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 H04N7/14 H04N5/232		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 H04N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 408 301 B1 (ASHE PHILIP R ET AL) 18 June 2002 (2002-06-18)	1,2,6,9, 10,24-28
Y	column 3, line 43 - column 5, line 8 column 6, line 42 - column 7, line 3	19-22 3-5
A	GB 2 361 097 A (SONY UNITED KINGDOM LTD) 10 October 2001 (2001-10-10) page 13, line 23 - page 23, line 24	3-5
Y	EP 0 330 455 A (TOKYO SHIBAURA ELECTRIC CO) 30 August 1989 (1989-08-30) column 1, line 3 - line 5 column 1, line 60 - column 2, line 6 column 8, line 9 - column 9, line 16	19,20,22
		-/-
<input checked="" type="checkbox"/>	Further documents are listed in the continuation of box C.	<input checked="" type="checkbox"/> Patent family members are listed in annex.
* Special categories of cited documents:		
A document defining the general state of the art which is not considered to be of particular relevance		
E earlier document but published on or after the international filing date		
L document which may throw doubt on priority, claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)		
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		
T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention		
W document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone		
V document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art		
& document member of the same patent family		
Date of the actual completion of the international search	Date of mailing of the international search report	
20 August 2004	04.10.2004	
Name and mailing address of the ISA European Patent Office, P.O. 5618 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Wentzel, J	

INTERNATIONAL SEARCH REPORT		International Application No PCT/GB 03/05224
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99/60788 A (PICTURETEL CORP) 25 November 1999 (1999-11-25) page 1, line 17 - page 2, line 15 page 3, line 13 - page 4, line 14 page 7, line 1 - line 35 page 11, line 22 - page 13, line 28 page 14, line 3 - page 16, line 32 page 36, line 32 - page 45, line 22 figure 5	1,9,10, 21,22 11-15
X	DE 196 00 958 A (FRAUNHOFER GES FORSCHUNG) 23 January 1997 (1997-01-23) column 1, line 24 - column 2, line 38 claims 6,14,15	1,9,10, 23
A	US 5 812 193 A (TOMITAKA TADAFUSA ET AL) 22 September 1998 (1998-09-22)	1,7
Y	column 17, line 40 - column 19, line 62	21
X	US 2001/053292 A1 (NAKAMURA KYOKO) 20 December 2001 (2001-12-20)	1
A	paragraph '0038! - paragraph '0044! paragraph '0052! paragraph '0064! - paragraph '0078!	8
X	US 2002/031262 A1 (YOSHIMURA TETSUYA ET AL) 14 March 2002 (2002-03-14)	1-4,9, 10,13, 14,18
A	paragraph '0052! - paragraph '0127!	7,11
X	EP 1 158 786 A (SONY CORP) 28 November 2001 (2001-11-28)	1,9,10, 13,14, 18,22
	paragraph '0062! - paragraph '0129! paragraph '0132! - paragraph '0172!	
A	US 2002/113862 A1 (CENTER JULIAN L ET AL) 22 August 2002 (2002-08-22) paragraph '0018! - paragraph '0025! paragraph '0048! - paragraph '0051!	1,8,15, 21,22

INTERNATIONAL SEARCH REPORT

International application No. PCT/GB 03/05224

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this International application, as follows:

see additional sheet

As a result of the prior review under R. 40.2(e) PCT,
no additional fees are to be refunded.

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/ GB 03/05224

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-6,9,10,24-28

Video camera arrangement recording metadata provided by face detector efficiently and related method

2. claims: 7,8,11,12

Video camera arrangement having improved face detection performance

3. claim: 13 to 21

Video camera arrangement comprising logic for further operation of said video camera arrangement in response/dependance of face detection

4. claims: 22,23

Applications of video camera arrangement

INTERNATIONAL SEARCH REPORT

Information on patent family members				International Application No PCT/GB 03/05224	
Patent document cited in search report	Publication date	Patent family member(s)		Publication date	
US 6408301	B1	18-06-2002		NONE	
GB 2361097	A	10-10-2001	AU EP WO JP US	4264701 A 1300019 A2 0176252 A2 2003530034 T 2002136531 A1	15-10-2001 09-04-2003 11-10-2001 07-10-2003 26-09-2002
EP 0330455	A	30-08-1989	JP JP JP JP EP US	1213086 A 1198884 A 2044881 A 2081590 A 0330455 A2 4951140 A	25-08-1989 10-08-1989 14-02-1990 22-03-1990 30-08-1989 21-08-1990
WO 9960788	A	25-11-1999	US AU EP JP WO	6593956 B1 6308799 A 1004204 A1 2002516535 T 9960788 A1	15-07-2003 05-03-2001 31-05-2000 04-06-2002 25-11-1999
DE 19600958	A	23-01-1997	DE WO DE	19600958 A1 9704428 A1 19701186 A1	23-01-1997 06-02-1997 16-07-1998
US 5812193	A	22-09-1998	JP JP JP JP US	3487436 B2 6165016 A 3399995 B2 6153047 A 5355163 A	19-01-2004 10-06-1994 28-04-2003 31-05-1994 11-10-1994
US 2001053292	A1	20-12-2001	JP	2001357404 A	26-12-2001
US 2002031262	A1	14-03-2002	CN JP	1344084 A 2002176619 A	10-04-2002 21-06-2002
EP 1158786	A	28-11-2001	EP JP US	1158786 A2 2002051315 A 2002054032 A1	28-11-2001 15-02-2002 09-05-2002
US 2002113862	A1	22-08-2002	NONE		

フロントページの続き

(51)Int.Cl. F I テーマコード(参考)
H 0 4 N 9/04 B

(72)発明者 ポーター、ロバート、マーク、ステファン
イギリス国、S O 2 2 5 A J ハンプシャー ウィンチエスター ウエスタン ロード 2 8
(72)発明者 ランバルス、ラテュナ
イギリス国、G U 2 8 D N サリー ギルフォード サウスウェイ 2 1 6
(72)発明者 ヘインズ、サイモン
イギリス国、R G 2 1 3 J X ハンプシャー ベージングストーク チャンドラー ロード 1
2
(72)発明者 ジラード、クライブ
イギリス国、G U 3 4 5 P F ハンプシャー アルトン メッドステッド ハッセル レーン
クリープランド
F ターム(参考) 5B057 CA01 CA08 CA12 CA17 CB01 CB08 CB12 CB17 CD05 DA08
DA16 DC25
5C053 GB06 GB21 HA29 JA03 JA21 LA06 LA14
5C065 AA02 BB16 BB48
5C122 DA08 DA11 EA42 EA68 FH10 FH14 FH16 FH23 FH24 FK34
FK39 FK42 GA17 GA23 GA34 HA87 HB01 HB05 HB09 HB10