

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6549734号
(P6549734)

(45) 発行日 令和1年7月24日 (2019.7.24)

(24) 登録日 令和1年7月5日 (2019.7.5)

(51) Int. Cl.

F I

G 0 6 T 7/00 (2017.01)
A 6 1 B 5/00 (2006.01)G 0 6 T 7/00 6 6 0 Z
A 6 1 B 5/00 C

請求項の数 18 (全 27 頁)

(21) 出願番号	特願2017-559799 (P2017-559799)	(73) 特許権者	590000248
(86) (22) 出願日	平成28年5月20日 (2016.5.20)		コーニンクレッカ フィリップス エヌ ヴェ
(65) 公表番号	特表2018-515854 (P2018-515854A)		KONINKLIJKE PHILIPS N. V.
(43) 公表日	平成30年6月14日 (2018.6.14)		オランダ国 5656 アーエー アイン ドーフエン ハイテック キャンパス 5
(86) 国際出願番号	PCT/EP2016/061376		High Tech Campus 5, NL-5656 AE Eindhoven
(87) 国際公開番号	W02016/185004		
(87) 国際公開日	平成28年11月24日 (2016.11.24)	(74) 代理人	100107766
審査請求日	平成30年12月12日 (2018.12.12)		弁理士 伊東 忠重
(31) 優先権主張番号	15168566.6	(74) 代理人	100070150
(32) 優先日	平成27年5月21日 (2015.5.21)		弁理士 伊東 忠彦
(33) 優先権主張国	欧州特許庁 (EP)		
早期審査対象出願			

最終頁に続く

(54) 【発明の名称】 ビデオ・シーケンスにおける生体皮膚組織の識別

(57) 【特許請求の範囲】

【請求項 1】

ビデオ・シーケンスにおいて生体皮膚組織を識別するための装置であって、

当該装置はプロセッサおよびコンピュータ可読コードを有するコンピュータ可読媒体を有し、前記コンピュータ可読コードは、前記プロセッサによって実行されると；

ビデオ・シーケンスを受け取る段階であって、前記ビデオ・シーケンスは複数の画像フレームを含む、段階と；

前記画像フレームのそれぞれを複数のフレーム・セグメントに分割することを、ピクセルを、該ピクセルの色および空間的類似性に基づいてフレーム・セグメントにグループ分けすることによって実行する段階であって、各フレーム・セグメントは前記画像フレームにおける近隣のピクセルの群である、段階と；

複数のビデオ・サブシーケンスを形成することを、各ビデオ・サブシーケンスについて、現在の画像フレームに対する次の画像フレームにおける、現在の画像フレームにおける対応するフレーム・セグメントに対するクロミナンスの最小の差をもつフレーム・セグメントを識別することによって、そのビデオ・サブシーケンスにおいて、フレーム・セグメントどうしのクロミナンスの間の差が最小化されるように、前記複数の画像フレームのそれぞれからフレーム・セグメントを選択することによって実行する段階であって、各ビデオ・サブシーケンスは前記複数の画像フレームの二つ以上からのフレーム・セグメントを含む、段階と；

前記複数のビデオ・サブシーケンスを解析して、各ビデオ・サブシーケンスについて

10

20

脈信号を決定する段階と；

それぞれの決定された脈信号について、他の決定された脈信号のそれぞれとのペアごとの類似性に基づく類似性行列を決定する段階と；

前記類似性行列から前記ビデオ・シーケンスにおける生体皮膚組織の領域を識別する段階とを実行するよう機能する、装置。

【請求項 2】

脈信号と他の脈信号の一つとについてのペアごとの類似性は：

それら二つの脈信号の周波数スペクトルの少なくとも一部の間の相関；

それら二つの脈信号の周波数スペクトルの少なくとも一部の間の規格化された相互相関

10

；

それら二つの脈信号の周波数スペクトルの少なくとも一部の間の相関のエントロピーの指標；および

それら二つの脈信号の内積の結果

のうちの一つまたは複数を含む、請求項 1 記載の装置。

【請求項 3】

ペアごとの類似性は周波数ベースのペアごとの類似性を含む、請求項 1 記載の装置。

【請求項 4】

前記プロセッサは、前記類似性行列から前記ビデオ・シーケンスにおける生体皮膚組織の領域を識別することを、前記類似性行列に対して行列分解を実行することによって実行するよう構成されている、請求項 1 記載の装置。

20

【請求項 5】

前記プロセッサは、複数のビデオ・サブシーケンスを形成することを、各ビデオ・サブシーケンスについて、そのビデオ・サブシーケンスにおいて、フレーム・セグメントどうしの位置の間の空間距離の間の差が最小化されるように、前記複数の画像フレームのそれぞれからフレーム・セグメントをさらに選択することによって実行するよう構成されている、請求項 1 記載の装置。

【請求項 6】

前記プロセッサはさらに、前記ビデオ・シーケンスにおける生体皮膚組織の識別された領域に関連付けられた一つまたは複数の脈信号から、一つまたは複数の生理的特性を決定するよう構成されている、請求項 1 記載の装置。

30

【請求項 7】

前記ビデオ・シーケンスを捕捉するためのカメラをさらに有する、

請求項 1 記載の装置。

【請求項 8】

ビデオ・シーケンスにおいて生体皮膚組織を識別するための方法であって：

ビデオ・シーケンスを取得する段階であって、前記ビデオ・シーケンスは複数の画像フレームを含む、段階と；

前記画像フレームのそれぞれを複数のフレーム・セグメントに分割することを、ピクセルを、該ピクセルの色および空間的類似性に基づいてフレーム・セグメントにグループ分けすることによって実行する段階であって、各フレーム・セグメントは前記画像フレームにおける近隣のピクセルの群である、段階と；

40

複数のビデオ・サブシーケンスを形成することを、各ビデオ・サブシーケンスについて、そのビデオ・サブシーケンスにおいて、フレーム・セグメントどうしのクロミナンスの間の差が最小化されるように、前記複数の画像フレームのそれぞれからフレーム・セグメントを選択することによって実行する段階であって、各ビデオ・サブシーケンスは前記複数の画像フレームの二つ以上からのフレーム・セグメントを含む、段階と；

前記複数のビデオ・サブシーケンスを解析して、各ビデオ・サブシーケンスについて脈信号を決定する段階と；

それぞれの決定された脈信号について、他の決定された脈信号のそれぞれとのペアごと

50

の類似性に基づく類似性行列を決定する段階と；

前記類似性行列から前記ビデオ・シーケンスにおける生体皮膚組織の領域を識別する段階とを含む、
方法。

【請求項 9】

ペアごとの類似性は周波数ベースのペアごとの類似性を含む、請求項 8 記載の方法。

【請求項 10】

前記類似性行列から前記ビデオ・シーケンスにおける生体皮膚組織の領域を識別する前記段階は、前記類似性行列に対して行列分解を実行することを含む、請求項 8 記載の方法。

10

【請求項 11】

ビデオ・サブシーケンスを解析して脈信号を決定することは：

フレーム・セグメント内の各ピクセルについての諸ピクセル値を平均し；

ビデオ・サブシーケンスについての脈信号を、そのビデオ・サブシーケンス内の各フレーム・セグメントについての諸平均ピクセル値から形成することを含む、
請求項 8 記載の方法。

【請求項 12】

ピクセル値を平均することは：

フレーム・セグメント内のピクセルのピクセル値に重み付けする段階であって、ピクセル値はそのフレーム・セグメント内でのそのピクセルの空間位置および／またはそのフレーム・セグメントの中心または中心近くのピクセルまたはピクセル群との色の差に基づいて重み付けされる、段階と；

20

フレーム・セグメント内のピクセルの重み付けされたピクセル値を平均する段階とを含む、

請求項 11 記載の方法。

【請求項 13】

コンピュータ可読コードが具現されている非一時的なコンピュータ可読媒体であって、前記コンピュータ可読コードは、コンピュータまたはプロセッサによって実行されると、該コンピュータまたはプロセッサが；

ビデオ・シーケンスを取得する段階であって、前記ビデオ・シーケンスは複数の画像フレームを含む、段階と；

30

前記画像フレームのそれぞれを複数のフレーム・セグメントに分割することを、ピクセルを、該ピクセルの色および空間的類似性に基づいてフレーム・セグメントにグループ分けすることによって実行する段階であって、各フレーム・セグメントは前記画像フレームにおける近隣のピクセルの群である、段階と；

複数のビデオ・サブシーケンスを形成することを、各ビデオ・サブシーケンスについて、そのビデオ・サブシーケンスにおいて、フレーム・セグメントどうしのクロマティック平面におけるクロミナンスの間の差が最小化されるように、前記複数の画像フレームのそれぞれからフレーム・セグメントを選択することによって実行する段階であって、各ビデオ・サブシーケンスは前記複数の画像フレームの二つ以上からのフレーム・セグメントを含む、段階と；

40

前記複数のビデオ・サブシーケンスを解析して、各ビデオ・サブシーケンスについて脈信号を決定する段階と；

それぞれの決定された脈信号について、他の決定された脈信号のそれぞれとのペアごとの類似性に基づく類似性行列を決定する段階と；

前記類似性行列から前記ビデオ・シーケンスにおける生体皮膚組織の領域を識別する段階とを含む、

含む方法を実行させられるよう構成されている、非一時的なコンピュータ可読媒体。

【請求項 14】

ペアごとの類似性は周波数ベースのペアごとの類似性を含む、請求項 13 記載の非一時

50

的なコンピュータ可読媒体。

【請求項 15】

前記類似性行列から前記ビデオ・シーケンスにおける生体皮膚組織の領域を識別することは、前記類似性行列に対して行列分解を実行することを含む、請求項 13 記載の非一時的なコンピュータ可読媒体。

【請求項 16】

ビデオ・サブシーケンスを解析して脈信号を決定することは：

フレーム・セグメント内の各ピクセルについての諸ピクセル値を平均し；

ビデオ・サブシーケンスについての脈信号を、そのビデオ・サブシーケンス内の各フレーム・セグメントについての諸平均ピクセル値から形成することを含む、
請求項 13 記載の非一時的なコンピュータ可読媒体。

10

【請求項 17】

ピクセル値を平均することは：

フレーム・セグメント内のピクセルのピクセル値に重み付けする段階であって、ピクセル値はそのフレーム・セグメント内でのそのピクセルの空間位置および／またはそのフレーム・セグメントの中心または中心近くのピクセルまたはピクセル群との色の差に基づいて重み付けされる、段階と；

フレーム・セグメント内のピクセルの重み付けされたピクセル値を平均する段階とを含む、

請求項 16 記載の非一時的なコンピュータ可読媒体。

20

【請求項 18】

前記方法が：

前記ビデオ・シーケンスにおける生体皮膚組織の識別された領域に関連付けられた一つまたは複数の脈信号から、一つまたは複数の生理的特性を決定する段階をさらに含む、
請求項 13 記載の非一時的なコンピュータ可読媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ビデオ・シーケンスにおける生体皮膚組織を識別するための装置および方法に関する。

30

【背景技術】

【0002】

近年、リモート光体積変動記録法（photoplethysmography）（リモートPPGまたはrPPG）を実行する技法が開発されている。これらの技法は、撮像ユニット（たとえばカメラ）を使って捕捉された画像フレームのビデオ・シーケンスからPPG信号が得られることを可能にする。ビデオ・シーケンスが処理されrPPG信号が抽出されるのが自動的であることが望ましく、それにより被写体が自動的にモニタリングできる。しかしながら、これはビデオ・シーケンスにおいて生体の皮膚組織の領域が自動的に識別されることを要求する。

【0003】

ビデオにおいて被写体を検出するタスクは、コンピュータビジョンにおける根源的なトピックの一つとして、過去数十年において幅広く研究されている。被写体を含んでいるビデオ・シーケンスが与えられたとき、目標は、被写体の身体部分に対応する領域を位置特定することである。たいていの既存の研究は、教師付きトレーニング機構において、被写体と背景との間で区別するために人間の外観の特徴を利用する。しかしながら、これらの方法での共通の問題は、トレーニングされた特徴が人間に特有ではなく、人間の外観に似た任意の特徴が誤分類されうるということである。さらに、教師付き方法は通例、事前にわかっているサンプルに制約され、予測不可能なサンプルが生じるときには失敗する傾向がある。たとえば、正面の顔でトレーニングされた顔検出器は横から見た顔を位置特定できず、一方、明るい肌でトレーニングされた肌分類器は暗い肌の被写体では失敗する。

40

【0004】

50

rPPG技法の開発に基づき、物理的な外観特徴に比べて、目に見えない生理的な特徴（たとえば脈）が、ビデオ・シーケンスにおいて人間を人間以外からよりよく区別できることが観察されている。自然環境では、生きている被写体の皮膚組織のみが脈動性を示すので、脈信号を示さないいかなるオブジェクトも、安全に人間以外のカテゴリーに分類できる。これは、たとえば図1に示されるように、人間に似た外観をもつオブジェクトの誤検出を防止できる。

【0005】

図1は、生体組織検出技法がいかに成功裏に動作すべきかの二つの例を与えている。左側の画像では、人間の顔および人工的な顔がカメラに正面を呈しており、人間の顔のみが識別されるべきである（人工的な顔が人間の顔と似た物理的な外観特徴をもつにもかかわらず）。このことは、生体皮膚組織に対応する領域の破線の四角および輪郭線によって示されている。右側の画像では、人間の顔と人工的な顔がカメラに側面を向けていて、人間の顔だけが識別されるべきである。

【0006】

非特許文献1の論文では、より高い周波数スペクトル・エネルギーをもつセグメント分割された局所領域（たとえばグリッド、三角形またはボクセル）を皮膚領域として選択するよう、硬閾値が設定される。非特許文献2の論文では、事前定義されたクラスタリング・パラメータが、類似性を共有する領域を皮膚領域としてクラスタリングするために使われる。

【0007】

しかしながら、これらの方法はパラメータなしではなく、実地的な使用事例からの現実的な困難に対して堅牢ではない。たとえば、成人で調整された硬閾値は新生児には適用可能でなく、事前定義されたクラスター数または密度は複数被写体検出には柔軟ではない。

【先行技術文献】

【非特許文献】

【0008】

【非特許文献1】G.Gibert and D.D'Alessandro, and F. Lance, "Face detection method based on photoplethysmography", 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp.449-453, (2013)

【非特許文献2】R. van Luijckelaar, W.Wang, S.Stuijk, and G. de Haan, "Automatic ROI detection for rPPG measurements", ACCV 2014, Singapore

【非特許文献3】Achanta et al., "SLIC Superpixels Compared to State-of-the-art Superpixel Methods", IEEE Transactions on Pattern Analysis & Machine Intelligence 2012 vol.34 Issue No.11, Nov. 2012, pp.2274-2282

【非特許文献4】G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG", TBME, 60(1):2878-2886, 2013

【非特許文献5】G. de Haan and A. van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature", Physiol. Meas. 35 1913, 2014

【非特許文献6】Haar, in P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", CVPR, 2001, 1

【非特許文献7】HOG, in N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", CVPR, 2005, 1

【非特許文献8】Y. Zhang, A. d'Aspremont and L. Ghaoui, "Sparse PCA: Convex relaxations, algorithms and applications", International Series in Operations Research & Management Science Volume 166:915-940, 2012

【非特許文献9】A. Argyriou, M. Signoretto and J. Suykens, "Hybrid conditional gradient-smoothing algorithms with applications to sparse and low rank regularization", arXivプレプリント: 1404.3591, 2014

【発明の概要】

【発明が解決しようとする課題】

【0009】

したがって、ビデオ・シーケンスにおいて生体皮膚組織を識別するための改善された方法および装置を提供することが目的である。

【課題を解決するための手段】

【0010】

第一の側面によれば、ビデオ・シーケンスにおいて生体皮膚組織を識別するための方法が提供される。本方法は、ビデオ・シーケンスを取得する段階であって、前記ビデオ・シーケンスは複数の画像フレームを含む、段階と；前記画像フレームのそれぞれを複数のフレーム・セグメントに分割する段階であって、各フレーム・セグメントは前記画像フレームにおける近隣のピクセルの群である、段階と；複数のビデオ・サブシーケンスを形成する段階であって、各ビデオ・サブシーケンスは前記複数の画像フレームの二つ以上からのフレーム・セグメントを含む、段階と；前記複数のビデオ・サブシーケンスを解析して、各ビデオ・サブシーケンスについて脈信号を決定する段階と；それぞれの決定された脈信号について、他の決定された脈信号のそれぞれとのペアごとの類似性に基づく類似性行列を決定する段階と；前記類似性行列から前記ビデオ・シーケンスにおける生体皮膚組織の領域を識別する段階とを含む。

10

【0011】

いくつかの実施形態では、類似性行列を決定する段階は、前記ペアごとの類似性を決定し；決定されたペアごとの類似性を組み合わせることによって前記類似性行列を決定することを含むことができる。

20

【0012】

いくつかの実施形態では、前記ペアごとの類似性を決定することは、各脈信号について、その脈信号と他の脈信号のそれぞれとの間のペアごとの類似性の指標を決定することを含むことができる。

【0013】

代替的な実施形態では、前記ペアごとの類似性を決定することは、各脈信号について、他の脈信号のそれぞれとのペアごとの類似性の二つ以上の指標を決定し；各脈信号についてのペアごとの類似性の前記二つ以上の指標を組み合わせ、各脈信号について、他の脈信号のそれぞれとの前記ペアごとの類似性を表わす距離メトリックを形成することを含むことができる。ここで、前記類似性行列は前記距離メトリックに基づいて決定される。

30

【0014】

いくつかの実施形態では、脈信号と他の脈信号の一つとについてのペアごとの類似性は、それら二つの脈信号の周波数スペクトルの間の相関の指標を含むことができる。代替的なまたはさらなる実施形態では、脈信号と他の脈信号の一つとについてのペアごとの類似性は、それら二つの脈信号の周波数スペクトルの間の規格化された相互相関の指標を含むことができる。代替的なまたはさらなる実施形態では、脈信号と他の脈信号の一つとについてのペアごとの類似性は、それら二つの脈信号の周波数スペクトルの間の相関の規則性の指標を含むことができる。代替的なまたはさらなる実施形態では、脈信号と他の脈信号の一つとについてのペアごとの類似性は、それら二つの脈信号の内積の結果を含むことができる。

40

【0015】

いくつかの実施形態では、ペアごとの類似性は周波数ベースのペアごとの類似性を含むことができる。

【0016】

いくつかの実施形態では、前記類似性行列から画像の前記シーケンスにおける生体皮膚組織の領域を識別する前記段階は、前記類似性行列に対して行列分解を実行することを含むことができる。

【0017】

50

いくつかの実施形態では、行列分解を実行することは、特異値分解SVD、QR分解、疎SVD、インクリメンタルSVD、主成分分析PCAおよび独立成分分析ICAを含むことができる。

【0018】

各フレーム・セグメントは、典型的には、画像フレームにおける近隣のピクセルの群である。いくつかの実施形態では、各フレーム・セグメントは同じ形をもつ。代替的な実施形態では、前記画像フレームのそれぞれを複数のフレーム・セグメントに分割する前記段階は、ピクセルを、該ピクセルの色および空間的類似性に基づいてフレーム・セグメントにグループ分けすることを含む。

【0019】

いくつかの実施形態では、複数のビデオ・サブシーケンスを形成する前記段階は、前記複数の画像フレームにおいて対応する空間位置にあるフレーム・セグメントから各ビデオ・サブシーケンスを形成することを含むことができる。代替的な実施形態では、複数のビデオ・サブシーケンスを形成する前記段階は、各ビデオ・サブシーケンスについて、そのビデオ・サブシーケンス内のフレーム・セグメントどうしの間のクロマティック・エネルギーおよび/または空間距離エネルギーが最小化されるように、前記複数の画像フレームのそれぞれからフレーム・セグメントを選択することを含むことができる。

【0020】

いくつかの実施形態では、ビデオ・サブシーケンスを解析して脈信号を決定する前記段階は、フレーム・セグメント内の各ピクセルについての諸ピクセル値を平均し；ビデオ・サブシーケンスについての脈信号を、そのビデオ・サブシーケンス内の各フレーム・セグメントについての平均された諸ピクセル値から形成することを含むことができる。

【0021】

いくつかの実施形態では、ピクセル値を平均する前記段階は、フレーム・セグメント内のピクセルのピクセル値に重み付けする段階であって、ピクセル値はそのフレーム・セグメント内でのそのピクセルの空間位置および/またはそのフレーム・セグメントの中心または中心近くのピクセルまたはピクセル群との色の差に基づいて重み付けされる、段階と；フレーム・セグメント内のピクセルの重み付けされたピクセル値を平均する段階とを含む。

【0022】

いくつかの実施形態では、本方法はさらに、ビデオ・シーケンスにおける生体皮膚組織の識別された領域に関連付けられた一つまたは複数の脈信号から、一つまたは複数の生理的特性を決定する段階をさらに含む。

【0023】

第二の側面によれば、コンピュータ可読コードが具現されているコンピュータ可読媒体を有するコンピュータ・プログラム・プロダクトが提供される。コンピュータ可読コードは、好適なコンピュータまたはプロセッサによって実行されると、該コンピュータまたはプロセッサに上記で記載された方法のいずれかを実行させるよう構成されている。

【0024】

第三の側面によれば、ビデオ・シーケンスにおいて生体皮膚組織を識別するための装置が提供される。本装置は処理ユニットを有し、該処理ユニットは、ビデオ・シーケンスを受け取る段階であって、前記ビデオ・シーケンスは複数の画像フレームを含む、段階と；前記画像フレームのそれぞれを複数のフレーム・セグメントに分割する段階であって、各フレーム・セグメントは前記画像フレームにおける近隣のピクセルの群である、段階と；複数のビデオ・サブシーケンスを形成する段階であって、各ビデオ・サブシーケンスは前記複数の画像フレームの二つ以上からのフレーム・セグメントを含む、段階と；前記複数のビデオ・サブシーケンスを解析して、各ビデオ・サブシーケンスについて脈信号を決定する段階と；それぞれの決定された脈信号について、他の決定された脈信号のそれぞれとのペアごとの類似性に基づく類似性行列を決定する段階と；前記類似性行列から前記ビデオ・シーケンスにおける生体皮膚組織の領域を識別する段階とを実行するよう構成されている。

10

20

30

40

50

【 0 0 2 5 】

いくつかの実施形態では、前記処理ユニットは、類似性行列を決定する段階を、前記ペアごとの類似性を決定し；決定されたペアごとの類似性を組み合わせることによって前記類似性行列を決定することによって実行するよう構成される。

【 0 0 2 6 】

いくつかの実施形態では、前記処理ユニットは、前記ペアごとの類似性を決定することを、各脈信号について、その脈信号と他の脈信号のそれぞれとの間のペアごとの類似性の指標を決定することによって実行するよう構成される。

【 0 0 2 7 】

代替的な実施形態では、前記処理ユニットは、前記ペアごとの類似性を決定することを、各脈信号について、他の脈信号のそれぞれとのペアごとの類似性の二つ以上の指標を決定し；各脈信号についてのペアごとの類似性の前記二つ以上の指標を組み合わせ、各脈信号について、他の脈信号のそれぞれとの前記ペアごとの類似性を表わす距離メトリックを形成することによって実行するよう構成される。ここで、前記類似性行列は前記距離メトリックに基づいて決定される。

10

【 0 0 2 8 】

いくつかの実施形態では、脈信号と他の脈信号の一つとについてのペアごとの類似性は、それら二つの脈信号の周波数スペクトルの間の相関の指標を含むことができる。代替的なまたはさらなる実施形態では、脈信号と他の脈信号の一つとについてのペアごとの類似性は、それら二つの脈信号の周波数スペクトルの間の規格化された相互相関の指標を含むことができる。代替的なまたはさらなる実施形態では、脈信号と他の脈信号の一つとについてのペアごとの類似性は、それら二つの脈信号の周波数スペクトルの間の相関の規則性の指標を含むことができる。代替的なまたはさらなる実施形態では、脈信号と他の脈信号の一つとについてのペアごとの類似性は、それら二つの脈信号の内積の結果を含むことができる。

20

【 0 0 2 9 】

いくつかの実施形態では、ペアごとの類似性は周波数ベースのペアごとの類似性を含むことができる。

【 0 0 3 0 】

いくつかの実施形態では、前記処理ユニットは、前記類似性行列から画像の前記シーケンスにおける生体皮膚組織の領域を識別することを、前記類似性行列に対して行列分解を実行することによって実行するよう構成される。

30

【 0 0 3 1 】

いくつかの実施形態では、前記処理ユニットは、行列分解を、特異値分解SVD、QR分解、疎SVD、インクリメンタルSVD、主成分分析PCAおよび独立成分分析ICAのうちの一つまたは複数をを使って実行するよう構成される。

【 0 0 3 2 】

各フレーム・セグメントは、典型的には、画像フレームにおける近隣のピクセルの群である。いくつかの実施形態では、各フレーム・セグメントは同じ形をもつ。代替的な実施形態では、前記処理ユニットは、前記画像フレームのそれぞれを複数のフレーム・セグメントに分割することを、ピクセルを、該ピクセルの色および空間的類似性に基づいてフレーム・セグメントにグループ分けすることによって実行するよう構成される。

40

【 0 0 3 3 】

いくつかの実施形態では、前記処理ユニットは、複数のビデオ・サブシーケンスを形成することを、前記複数の画像フレームにおいて対応する空間位置にあるフレーム・セグメントから各ビデオ・サブシーケンスを形成することによって実行するよう構成される。代替的な実施形態では、前記処理ユニットは、複数のビデオ・サブシーケンスを形成することを、各ビデオ・サブシーケンスについて、そのビデオ・サブシーケンス内のフレーム・セグメントどうしの間のクロマティック・エネルギーおよび/または空間距離エネルギーが最小化されるように、前記複数の画像フレームのそれぞれからフレーム・セグメントを

50

選択することを含むことによって実行するよう構成される。

【0034】

いくつかの実施形態では、前記処理ユニットは、ビデオ・サブシーケンスを解析して脈信号を決定することを、フレーム・セグメント内の各ピクセルについての諸ピクセル値を平均し；ビデオ・サブシーケンスについての脈信号を、そのビデオ・サブシーケンス内の各フレーム・セグメントについての平均された諸ピクセル値から形成することによって実行するよう構成される。

【0035】

いくつかの実施形態では、前記処理ユニットは、ピクセル値を平均することを、フレーム・セグメント内のピクセルのピクセル値に重み付けする段階であって、ピクセル値はそのフレーム・セグメント内でのそのピクセルの空間位置および／またはそのフレーム・セグメントの中心または中心近くのピクセルまたはピクセル群との色の差に基づいて重み付けされる、段階と；フレーム・セグメント内のピクセルの重み付けされたピクセル値を平均する段階とを実行することによって実行するよう構成される。

10

【0036】

いくつかの実施形態では、前記処理ユニットはさらに、ビデオ・シーケンスにおける生体皮膚組織の識別された領域に関連付けられた一つまたは複数の脈信号から、一つまたは複数の生理的特性を決定するよう構成される。

【0037】

いくつかの実施形態では、本装置はさらに、ビデオ・シーケンスを捕捉するための撮像ユニットを有する。

20

【図面の簡単な説明】

【0038】

本発明のよりよい理解のために、また本発明がどのように実施されうるかをより明瞭に示すために、ここで単に例として付属の図面を参照する。

【図1】生体皮膚組織検出技法の所望される動作を示す図である。

【図2】本発明の実施形態に基づく装置のブロック図である。

【図3】本発明の実施形態に基づく方法を示すフローチャートである。

【図4a】図4(a)～(d)はビデオ・シーケンスから、複数のビデオ・サブシーケンスについての脈信号がどのように得られるかを示す図であり、図4(a)は、ビデオ・シーケンスが一連の画像フレームから構成される様子を示す図である。

30

【図4b】各画像フレームが複数のフレーム・セグメントに分割される様子を示す図である。

【図4c】画像フレーム内の同じ空間位置にあるフレーム・セグメントを使って二つのビデオ・サブシーケンスが形成される様子を示す図である。

【図4d】そのように形成された二つのビデオ・サブシーケンスについて例示的な脈信号を示す図である。

【図5】例示的なボクセル脈スペクトル法における処理段階を示す図である。

【図6】画像フレームの三つの異なるスケールでのセグメント分割を示す図である。

【図7】ペアごとの類似性の四つの異なる指標および結果として得られる類似性行列を示す図である。

40

【図8】インクリメンタル疎PCAを使つての類似性行列分解の例を示す図である。

【図9】固有ベクトルの階層的ボクセルへの投影と、ビデオ・シーケンスのどの部分が生体皮膚組織に対応するかを示す融合マップとを示す図である。

【発明を実施するための形態】

【0039】

本発明の実施形態に従って生体皮膚組織を識別するために使用できる装置2が図2に示されている。装置2は、ある時間期間にわたってビデオ・シーケンスを捕捉する撮像ユニット4を有する。撮像ユニット4は、rPPG測定のために使用できるカメラ、たとえばRGBカメラであるまたはかかるカメラを含むことができる。撮像ユニット4は、複数の画像フ

50

レームを含むビデオ・シーケンスを処理ユニット 6 に提供する。

【 0 0 4 0 】

処理ユニット 6 は、装置 2 の動作を制御し、本稿に記載される生体皮膚組織識別技法を実装するための一つまたは複数のプロセッサ、マルチコア・プロセッサまたは処理モジュールを有することができる。いくつかの実施形態では、処理ユニット 6 は、複数の処理モジュールとして実装されることができる。各モジュールが、本稿に記載される生体皮膚組織識別技法の特定の部分または段階を実行するよう構成される。

【 0 0 4 1 】

装置 2 はさらに、本発明に基づく方法を実行するために処理ユニット 6 によって実行されることのできるコンピュータ可読プログラム・コードを記憶するためのメモリ・ユニット 8 を有する。メモリ・ユニット 8 は、処理ユニット 6 による処理の前、間および後の撮像ユニット 4 からのビデオ・シーケンスならびに処理の何らかの中間産物を記憶またはバッファリングするために使うこともできる。

【 0 0 4 2 】

いくつかの実施形態では、装置 2 は、統合されたまたは別個の撮像ユニット 4 をもつ汎用コンピュータ（たとえばデスクトップ PC）または統合されたまたは別個の撮像ユニット 4 を有するポータブル・コンピューティング装置（たとえばラップトップ、タブレットまたはスマートフォン）を含むことができることは理解されるであろう。いくつかの実施形態では、装置 2 はビデオ・シーケンスにおいて生体皮膚組織を識別する目的および / または生体皮膚組織に対応すると識別されたビデオ・シーケンスの領域から抽出された rPPG 信号から被写体の生理的特性を測定するための専用にされることができる。

【 0 0 4 3 】

実際的な実装では、装置 2 は、図 2 に示され上記したもの以外の他のまたはさらなるコンポーネント、たとえば主体が装置 2 を作動させるおよび / または操作することを許容するユーザー・インターフェースおよび装置 2 に電力を与えるための電源、たとえばバッテリーまたは幹線電源への接続を有していてもよい。ユーザー・インターフェースは、主体が装置 2 と対話し、装置 2 を制御することを許容する一つまたは複数のコンポーネントを有していてもよい。例として、前記一つまたは複数のユーザー・インターフェース・コンポーネントは、装置 2 を作動させ、作動停止させるためおよび / または生体皮膚組織識別プロセスのための、スイッチ、ボタンまたは他の制御手段を有することができる。ユーザー・インターフェース・コンポーネントは、さらに、または代わりに、装置 2 の動作について主体に情報を提供するために、ディスプレイまたは他の視覚的インジケータ（たとえばライト）を有することができる。同様に、ユーザー・インターフェース・コンポーネントは、装置 2 の動作について主体に可聴フィードバックを提供するためにオーディオ源を有することができる。

【 0 0 4 4 】

図 3 のフローチャートは、ある実施形態に基づくビデオ・シーケンスにおける生体皮膚組織を識別する方法を示している。

【 0 0 4 5 】

段階 1 0 1 では、撮像ユニット 4 がビデオ・シーケンスを取得する。ビデオ・シーケンスは、一連の画像フレームから構成される。一連の画像フレーム 2 0 は図 4 (a) に示される。

【 0 0 4 6 】

次に、画像フレーム 2 0 のそれぞれが複数のフレーム・セグメント 2 2 に分割される（段階 1 0 3）。各フレーム・セグメント 2 2 は、画像フレーム 2 0 における近隣のピクセルの群である。例示的なセグメント分割が図 4 (b) に示されている。この図では、各フレーム 2 0 は等しいサイズの正方形または長方形に分割されている。代替的な実施形態では、セグメントは異なる形、たとえば三角であってもよい。他の（好ましい）代替的な実施形態では、セグメント 2 2 の形状は、画像フレーム 2 0 内の画像によって決定されることができる（たとえば、前記形状の境界は、画像フレーム内の異なる色の間の境界をなぞ

10

20

30

40

50

ることができる)。しかしながら、各実施形態において、各フレーム・セグメント 22 が、各画像フレーム 20 における空間的に関係した(すなわち近隣の)ピクセルの群を含むことは理解されるであろう。前記好ましい実施形態では、フレーム・セグメント 22 は当技術分野において「スーパーピクセル」としても知られる。これはたとえば非特許文献 3 に記載される。

【0047】

ピクセルのグループ分け/セグメント 22 の形状が画像フレームの内容によって決定される、上記の好ましい「スーパーピクセル」実施形態では、フレーム・セグメント 22 は画像フレーム 20 内のピクセルをピクセルの色および空間的類似性に基づいてグループ分けすることによって決定できる。このようにして近隣のまたは隣接したピクセルであって同様のまたは一貫した色をもつものが単一のフレーム・セグメント 22 にグループ化される。

10

【0048】

上記の諸実施形態において、画像フレーム 20 は、単にその画像フレーム 20 の解析のみに基づいてフレーム・セグメント 22 に分割される。しかしながら、特定の画像フレーム 20 をその画像フレーム 20 および一つまたは複数のその後の画像フレーム 20 の解析に基づいてフレーム・セグメント 22 に分割することが可能である。換言すれば、上記した画像フレーム 20 の空間的および/または色ベースのセグメント分割は、時間領域において外観(たとえば色)および空間的類似性を共有するピクセルと一緒にグループ化されるよう、時間領域に拡張される。

20

【0049】

画像フレーム 20 をセグメント分割した後、フレーム・セグメント 22 から複数のビデオ・サブシーケンスが形成される(段階 105)。いくつかの場合には、各ビデオ・シーケンスは画像フレーム 20 のそれぞれからのフレーム・セグメント 22 を含むことができる。他の場合には、ビデオ・サブシーケンスは、ビデオ・シーケンスにおける画像フレーム 20 の、二つ以上の(好ましくは連続する)画像フレーム 20 を含む部分集合内のそれぞれの画像フレーム 20 からのフレーム・セグメント 22 を含むことができる。さらなる場合には、ビデオ・サブシーケンスは、一緒に連結された画像フレーム 20 の複数の部分集合からのフレーム・セグメント 22 を含むことができる(各部分集合は二つ以上の画像フレーム 20 を含む)。ビデオ・サブシーケンスがビデオ・シーケンスにおける画像フレーム 20 の部分集合のそれぞれからのフレーム・セグメント 22 (たとえば2~9個の画像フレーム 20 からのフレーム・セグメント 22)を含む実施形態では、ビデオ・サブシーケンスは本稿ではボクセルとも称される。

30

【0050】

いくつかの実施形態では、ビデオ・サブシーケンスは、適宜ビデオ・シーケンスにおけるそれぞれの画像フレーム 20 内または画像フレーム 20 の部分集合におけるそれぞれの画像フレーム 20 内で、同じ空間位置にあるフレーム・セグメント 22 を使って形成される。たとえば、諸画像フレーム 20 の左上隅にあるフレーム・セグメント 22 から一つのビデオ・サブシーケンスが形成でき、諸画像フレーム 20 の左下隅にあるフレーム・セグメント 22 からもう一つのビデオ・サブシーケンスが形成でき、などとなる。このことは、四つの画像フレーム 20 のそれぞれにおける二つの特定のフレーム・セグメント 22 について、図 4(c)に示されている。このように、第一のビデオ・サブシーケンス 24 は、諸画像フレーム 20 のそれぞれにおいて同じ空間位置にある第一のフレーム・セグメント 26 から形成され、第二のビデオ・サブシーケンス 28 は、諸画像フレーム 20 のそれぞれにおいて別の空間位置にある第二のフレーム・セグメント 30 から形成される。

40

【0051】

しかしながら、ある好ましい実施形態(フレーム・セグメント 22 がピクセルを色および空間的類似性に従ってグループ分けすることによって形成されるときに特に好ましい実施形態)では、ビデオ・シーケンスの内容における変化(たとえば被写体の動きに起因する変化)に対する本方法の堅牢さを改善するために、ビデオ・サブシーケンスは、互いに

50

整合する（たとえば、画像フレーム内の空間位置において概して整合するおよび色において概して整合する）、諸画像フレーム20からの諸フレーム・セグメント22を選択することによって、形成されることができる。これは、諸画像フレーム20を通じて「曲がりながら進む」ビデオ・サブシーケンス、よって、被写体の特定の部分についてのフレーム・セグメントを含むビデオ・サブシーケンスを生じることができる（たとえば、ビデオ・シーケンスにおいて被写体が左から右に動く際、ある特定のビデオ・サブシーケンスは、空間的および色の類似性のために被写体の類に対応する各画像フレーム中のフレーム・セグメント22によって形成されることができる）。ビデオ・サブシーケンスを形成する一つの好ましい方法は、画像200における特定のフレーム・セグメント22について、次の画像フレーム20においてその特定のフレーム・セグメント22から最小のクロマティック・エネルギー（すなわち、クロミナンスにおける最小の差）および空間距離エネルギー（すなわち、最小の空間距離）をもつフレーム・セグメント22を同定することである。クロマティック・エネルギーが、当該フレーム・セグメント22と次の画像フレーム20におけるあるフレーム・セグメント22におけるピクセルのクロミナンス値に基づくエネルギー関数を指すこと、よって、ビデオ・サブシーケンスを形成するためにクロマティック・エネルギーを最小化することは、特定のフレーム・セグメント22について、次の画像フレーム20において、考えているフレーム・セグメント22に対する最小のクロマティック・エネルギーをもつフレーム・セグメント22を同定することを含むことができることは理解されるであろう。考えているフレーム・セグメント22におけるピクセルについてのクロミナンスに比べてあるフレーム・セグメント22におけるピクセルについてのクロミナンスが異なるほど、クロマティック・エネルギーは高く、よってそのフレーム・セグメント22がそのビデオ・サブシーケンスのために選択される可能性は低くなることは理解されるであろう。空間距離エネルギーが、当該画像フレーム20における当該フレーム・セグメント22の空間位置と次の画像フレーム20におけるフレーム・セグメント22の空間位置に基づくエネルギー関数を指すこと、よって、ビデオ・サブシーケンスを形成するために空間距離エネルギーを最小化することは、特定のフレーム・セグメント22について、次の画像フレーム20において、考えているフレーム・セグメント22における最小の空間距離を与えるフレーム・セグメント22を同定することを含むことができることは理解されるであろう。次の画像フレーム20におけるあるフレーム・セグメント22の位置から考えているフレーム・セグメント22までの距離が大きいほど、空間距離エネルギーは高く、そのフレーム・セグメント22がそのビデオ・サブシーケンスのために選択される可能性は低くなることは理解されるであろう。代替的な手法では、時刻 t における新たな諸ボクセルを、時刻 $t-1$ における諸ボクセルの中心を使って初期化することも可能である。諸ボクセルが時間的に重なり合う場合には、ボクセルの最後のフレーム・セグメント22における中心が、次のボクセルの最初のフレーム・セグメント22における中心を決定する。

【0052】

次いで、各ビデオ・サブシーケンスは、各ビデオ・サブシーケンスについての脈信号を決定するために解析される（段階107）。脈信号は、ビデオ・サブシーケンスにおける諸フレーム・セグメント22の色または色の変化を表わす。ビデオ・シーケンスから（よってビデオ・サブシーケンスから）脈信号を決定するためのさまざまな技法が当技術分野において知られており、ここで詳細に述べることはしない。しかしながら、いくつかの例示的な技法が下記に呈示されるボクセル脈スペクトル（VPS）法の記述において言及される。

【0053】

いくつかの実施形態では、各フレーム・セグメント22のピクセル値（たとえばRGB値）が平均され、サブシーケンス内の各フレーム・セグメント22についての平均値の時系列から、脈信号が形成される。いくつかの実施形態では、ピクセル値は、フレーム・セグメント22におけるピクセルの空間位置および/またはフレーム・セグメント22の中心または中心近くのピクセルまたはピクセル群との色の差に基づいて重み付けされ、重み付

10

20

30

40

50

けされた値の平均が決定される。たとえば、ピクセル値は、そのピクセルからフレーム・セグメント 2 2 の空間的境界までの距離および / またはそのピクセルからそのフレーム・セグメント 2 2 の中心までの距離に基づいて重み付けされることができる。好ましくは、重み付けは、セグメント 2 2 境界に近いピクセルがより小さな重みをかけられることにつながる。境界に近いピクセルは、隣り合うセグメント 2 2 の間のジッタ・アーチファクトのためセグメント 2 2 の中央に近いピクセルほど信頼できないからである。

【 0 0 5 4 】

例示的な脈信号が、二つのビデオ・サブシーケンス 2 4、2 8 について図 4 (d) に示されている。この例では、ビデオ・サブシーケンス 2 4 は生体皮膚組織の領域を含み、よってこのビデオ・サブシーケンスから決定された脈信号は PPG 信号に典型的な特性（すなわち、心臓の拍動のため被写体の皮膚における血液灌流の変化と整合した振幅変動）を示す。ビデオ・サブシーケンス 2 8 は、生体皮膚組織の領域を含まず、よって、このビデオ・サブシーケンスから決定された脈信号は PPG 信号に典型的な特性を示さない（そして、ビデオ・シーケンスにおける周辺照明の変化がなければ、サブシーケンス 2 8 についての脈信号は概してノイズ信号に対応しうる）。

10

【 0 0 5 5 】

ひとたび各ビデオ・サブシーケンスについて脈信号が決定されたら、ビデオ・シーケンスについて類似性行列が決定される（段階 1 0 9）。類似性行列は、同様の脈信号が互いに相関付けられている行列である。類似性行列の使用は、何らかのパラメータ（たとえば皮膚色調またはクラスタリングに基づくパラメータ）があらかじめ定義されることを要求しないので、有利である。

20

【 0 0 5 6 】

類似性行列は好ましくは、各脈信号と他の脈信号とについてのペアごとの類似性を決定することによって形成される。すなわち、各脈信号について、他の脈信号のそれぞれとの類似性の指標が決定される。これらのペアごとの類似性指標が次いで組み合わされて、類似性行列を形成する。

【 0 0 5 7 】

ペアごとの類似性指標は好ましくは、周波数ベースのペアごとの類似性指標を含む、または、周波数ベースのペアごとの類似性指標である。同じ特定の被写体の生体皮膚組織の領域に対応する異なるフレーム・セグメント 2 2 は、相関において、類似の（または同じ）周波数ピーク・インデックス、位相または低いエントロピーをもつ脈信号を示すはずなので、これは有利である。

30

【 0 0 5 8 】

脈信号と他の脈信号のうちのひとつとについてのペアごとの類似性の指標は、その脈信号と他の脈信号のうちの前記一つとの周波数スペクトルの少なくとも一部の間の相関の指標（これは本稿ではスペクトル・ピークと称する）、その脈信号と他の脈信号のうちの前記一つとの周波数スペクトルの少なくとも一部の間の規格化された相互相関の指標（これは本稿ではスペクトル位相と称される）、それら二つの脈信号の周波数スペクトルの少なくとも一部の間の相関の規則性の指標（これは本稿ではスペクトル・エントロピーと称される）および / またはそれら二つの脈信号の内積の結果（脈信号は任意的には内積が計算される前にフィルタリングされることができる）であることができる。これらのペアごとの類似性指標のさらなる詳細は、下記の VPS 法の記述に見出される。

40

【 0 0 5 9 】

当業者は、ペアごとの類似性について上記のものに対する代替的なまたはさらなる指標であって、決定され類似性行列形成に使用できるものを認識するであろう。

【 0 0 6 0 】

いくつかの実施形態では、ペアごとの類似性の複数の指標（たとえばスペクトル・ピークおよびスペクトル位相）が、各脈信号と他の脈信号のそれぞれとについて決定でき、それらの指標が組み合わされて、各脈信号と他の脈信号のそれぞれとについてのペアごとの類似性を表わす距離メトリックを形成することができる。次いで、それらの距離メトリッ

50

クに基づいて類似性行列が決定されることができる。

【 0 0 6 1 】

ひとたび類似性行列が決定されたら、類似性行列から、ビデオ・シーケンスにおいて、生体皮膚組織の領域が識別される（段階 1 1 1）。生体皮膚組織の領域は、類似性行列の行列分解を実行することによって同定される。いくつかの実施形態では、行列分解は、特異値分解（singular value decomposition）（SVD）、QR分解、疎SVD、インクリメンタルSVD、主成分分析（principal component analysis）PCAおよび独立成分分析（independent component analysis）ICAを含むことができる。これらの技法は当技術分野において一般に知られており、本稿では詳細に述べることはしない。下記でより詳細に述べるある好ましい実施形態では、類似性行列はインクリメンタル疎PCAを使って分解される。

10

【 0 0 6 2 】

段階 1 1 1 は、同じ被写体に属するビデオ・シーケンスの諸部分を見出すために、類似性行列を直交基底に因子分解（分解）することを含むことができる。この因子分解の結果、異なる種類の類似性が独立な方向に分離される。その結果、同じ被写体に属するフレーム・セグメント 2 2 は同じ方向にクラスタリングされる。

【 0 0 6 3 】

図 3 には示していないが、ひとたびビデオ・シーケンスにおいて生体皮膚組織の領域が識別されたら、被写体（識別された生体皮膚組織をもつ被写体）の一つまたは複数の生理的特性がビデオ・シーケンスから決定できる。

【 0 0 6 4 】

20

いくつかの実施形態では、生理的特性（単数または複数）は、生体皮膚組織の識別された領域に関連付けられた前記一つまたは複数の脈信号から決定できる。この場合、前記一つまたは複数の脈信号は個々に解析されて生理的特性を決定し、複数の生理的特性が組み合わせられて（たとえば平均されて）被写体についての生理的特性の全体的指標を与えることができる。あるいはまた、前記一つまたは複数の脈信号が組み合わせられて（たとえば平均されて）単一の脈信号を与え、その脈信号が解析されて生理的特性を決定することができる。

【 0 0 6 5 】

他の実施形態では、ビデオ・シーケンスは再処理されて生体皮膚組織として識別された領域から脈信号（単数または複数）を抽出し、その脈信号が処理されて生理的特性を決定

30

【 0 0 6 6 】

ビデオ・シーケンスから決定される脈信号は、PPGセンサーを使って得られる信号と同様であり、よって被写体の前記一つまたは複数の生理的特性は、被写体のPPG信号または血液灌流（または血液灌流の変化）の他の指標から導出できる任意の特性、たとえば心拍数、心拍数変動率、拍間間隔、呼吸数、呼吸信号、SpO2値（すなわち、血液の動脈酸素飽和レベル）などを含むことができる。このように、PPG信号からそのような生理的特性を導出するために当技術分野で知られている諸技法（たとえば、心拍数を決定するための周波数領域におけるピーク検出）が、生体皮膚組織であると識別された領域から得られる脈信号から生理的特性についての値を導出するために使用できる。

40

【 0 0 6 7 】

本稿に記載される技法のある好ましい実施形態を下記に述べる。これは本稿では、ボクセル脈スペクトル（VPS）法と称される。

【 0 0 6 8 】

ボクセル脈スペクトル（VPS: Voxel-Pulse-Spectral）法

カメラに基づく脈抽出

人間の心臓血管系では、身体を通じて伝搬する血液脈は皮膚組織における血液体積を変える。血液中のヘモグロビンの光吸収は光スペクトルを通じて変化するので、皮膚反射の色変動を検出することで脈拍数を明らかにできる。近年のリモート光体積変動記録法（rPPG）技法は、通常のRGBカメラを使って人間の皮膚での脈に誘起された色変動を検出する

50

ことを可能にする。脈を見出すために時間的なRGB信号を因子分解するためにブラインド信号源分離方法（たとえばPCAに基づくもの、ICAに基づくもの）が提案されている。標準化された皮膚色調の想定のもとでRGBチャネルの線形結合として脈を定義するためのクロミナンスに基づくrPPG法も提案されており、これは現実的な困難（たとえば異なる被写体皮膚色）に対処することにおいて最も正確なrPPG法の一つである。

【0069】

脈に基づく関心領域の検出

人間の脈がビデオ・シーケンスにおいてrPPGによって測定できるという事実を与えられ、脈信号は被写体検出を支援するために使用できる。すなわち、生きている被写体を、その生体皮膚組織を位置特定することによって検出するのである。既存の技法は、局所的な脈抽出のために、ビデオを、固定した剛体的なグリッドにスライスする、脈信号に基づく顔検出方法を提案する。同方法は、硬閾値を設定して高いスペクトル・エネルギーをもつグリッドを見出し、それらのグリッドに顔領域とラベル付けする。同方法は、静止した顔がカメラからあらかじめ定義された距離のところに置かれている必要があるビデオに制限される。本稿に記載されるVPS法は、こうした制限を受けない。もう一つの既存の技法では、rPPGモニタリングを向上させるための顔に対する最適関心領域（RoI: Region of Interest）選択法が提案された。しかしながら、RoIはあらかじめ定義された顔ランドマークに制約され、これは被写体検出のための一般的解決策ではない。すなわち、ビデオにおいて見える可能性のある他の身体部分（たとえば手）を検出できない。対照的に、本稿に記載されるVPS法は、そのような想定をせず、脈動する血液体積をもつすべての身体部分を検出できる。

【0070】

VPS法の概観が図5に示されている。これは、画像フレーム20を含む入力ビデオ・シーケンスを受けて、被写体RoI 32（すなわち、生体皮膚組織に対応するビデオ・シーケンスの領域）を出力する。図3に示した方法と整合して、ビデオ・シーケンスは複数のサブシーケンスにセグメント分割され（階層的ボクセル段34 段階103および105）、各サブシーケンスについて脈信号が決定され（脈抽出段36 段階107）、RoIを同定するために類似性行列が決定されて解析される（スペクトル解析段38 段階109および111）。各段は下記で詳細に述べる。

【0071】

階層的ボクセル

被写体についての（あるいは内容一般についての）何らの事前情報もなしにビデオ・シーケンスを与えられて、ビデオ・シーケンスはまず、独立して脈が測定できる密な局所的領域にセグメント分割される（これは図5では階層的ボクセル段34である）。ビデオ・シーケンスは固定した剛体的グリッドにスライスされることができ、これは被写体サイズがグリッド幾何によって量子化されることを意味し、そのことは被写体が小さいときまたは身体動きがあるときは難儀するまたは失敗する。したがって、VPS法においては、「階層的ボクセル」と呼ばれる、脈抽出のための、より優れたビデオ・セグメント分割方法を使うことが好ましい。階層的ボクセルは、好ましくは複数のスケールでの（スケールは各画像フレーム20が分割されるクラスター/セグメントの数を決定する）、空間時間的にコヒーレントなクラスター（フレーム・セグメント）からなる。ここで、時間領域における空間的類似性および外観を共有する、諸画像フレーム20内の諸ピクセルが一緒にグループ化される。複数のスケールが好ましく、以下におけるVPS法の記述でも使われるが、VPS法において単一のスケールを使うことも可能である。

【0072】

一つのスケールから始めて、ボクセル（連続する画像フレームの部分集合（たとえば2, 3, 4など）からのフレーム・セグメントを含むビデオ・サブシーケンス）の構築は、短い期間 $T = \{2n+1, n \in \mathbb{N}^+\}$ 内の時間的に隣り合うスーパーピクセル/フレーム・セグメントの間のクロマティック・エネルギー E_c （すなわち、諸フレーム・セグメントについてのクロミナンスの差を最小化）および空間距離エネルギー E_s （すなわち、フレーム・セグメン

ト間の空間距離の差を最小化)を

【0073】

【数1】

$$\arg \min \left(\sum_{t=\frac{T-1}{2}}^{t+\frac{T-1}{2}} \sum_{p \in P(t)} (1-\lambda) E_c^t(p, k) + \lambda E_s^t(p, k) \right) \quad (1)$$

として最小化する手順として定義される。ここで、 $p \in P(t)$ は t 番目のフレームにおけるピクセルの集合である。 p の表現は4次元特徴ベクトル (x, y, u, v) であり、 (x, y) および (u, v) はそれぞれ画像平面およびクロマティック平面(たとえばYUV空間のUV平面、皮膚セグメンテーションのための経験的空間)における座標である。 T の間の全エネルギーを最小にするようピクセルを k 個のクラスターに割り当てるために、 K 平均クラスタリングが実行される。 λ は二つのエネルギー項の間のバランスを制御するパラメータである。

【0074】

さらに、単一スケールのボクセルは、式(1)における異なる k を同時に初期化することによって、複数スケールに拡張される。ここで、各スケールは独立なクラスタリングである。別個のスケールでのボクセルは異なる分解能およびエネルギー変動をもつことを考えると、 i 番目のスケールにおける λ_i は、 t でのそれ自身のエネルギー変動に基づいて、

【0075】

【数2】

$$\lambda_i^t = \log(k) \sqrt{\frac{\sigma(\phi(u_i^t)) \cdot \sigma(\phi(v_i^t))}{\sigma(\phi(x_i^t)) \cdot \sigma(\phi(y_i^t))}} \quad (2)$$

として適応的に自己調節される。ここで、 (\cdot) は標準偏差演算子を表わし； (\cdot) はクラスター平均の集合を表わし； $\log(k)$ はボクセルのコンパクトさを制御する、すなわちより高い分解能(より大きな k)をもつボクセルほどコンパクトであるべきである。異なるスケールでのリアルタイム調節は、安定しない、うつろいやすいクラスタリングを回避し、これは細かい粒度のセグメント分割を保存する。

【0076】

上記のような階層的ボクセルを使うことには、四つの恩恵がある。そのことを例解する図6は、生体被写体および人工的な顔を含む二つの異なる画像について三つの異なる分解能/スケール($k=16$ 、 $k=36$ 、 $k=64$)を示している。図6において、第一の画像40は図1の左側の画像に対応し、いずれもカメラを向いている人間の顔および人工的な顔を示している。第二の画像42は図1の右側の画像に対応し、カメラに側面を向けている人間の顔および人工的な顔を示す。第一に、階層的ボクセルを使うことは、脈測定のための、空間時間的にコヒーレントな「チューブ」(ビデオ・サブシーケンス)を確立する。第二に、ビデオにおけるスケール不変な被写体検出を可能にする。第三に、被写体形状の高い境界再現(recall)を維持する(フレーム・セグメントの境界が画像中の形をなぞっていることによって示される)。第四に、異なる分解能をもつボクセルから測定された脈は異なる量子化された品質をもつので、皮膚領域の統計的観察が生成される。

【0077】

脈抽出

本節は図5の脈抽出段36を記述する。階層中の各ボクセル(すなわち、ビデオ・サブシーケンス)は、並列的な脈抽出における独立な脈センサーであると想定される。ある好ましい実施形態では、脈測定のために、非特許文献4に記載される、クロミナンスに基づく方法(CHROM)が使われる。しかしながら、当業者は、ボクセルについての脈信号を決定するために使用できる他の技法を認識するであろう。たとえば、上記のように、ボクセルから脈信号を抽出するために、PCAまたはICAに基づく技法が使用でき、あるいは非特許

10

20

30

40

50

文献5に記載される「PBV」法が、RGBトレースから脈を抽出するために使用できる。局所的な領域／クラスターにおける脈信号を導出するためにすべてのピクセルの空間的に平均されたRGBを使うCHROMとは異なり、ボクセル中のピクセルRGB値がボクセル境界への距離に基づいて重み付けすることによって組み合わされる。すなわち、ボクセル境界に近いピクセルは、隣り合うボクセルの間の時折のジッタ・アーチファクトのため、それほど信頼できず、よってより小さな重みかけられるべきである。ピクセルkからボクセル境界までの最も近いユークリッド距離が d_k であるとする、tにおけるi番目のスケールでのj番目のボクセルの平均RGBは

【0078】

【数3】

$$(\bar{R}_{ij}^t, \bar{G}_{ij}^t, \bar{B}_{ij}^t) = \frac{\sum_{k=0}^N (d_k \cdot (R_{ijk}^t, G_{ijk}^t, B_{ijk}^t))}{\sum_{k=0}^N d_k} \quad (3)$$

10

として組み合わされる。ここで、Nはj番目のボクセルにおけるピクセル数を表わす。一定した照明の環境では、人間の皮膚組織は同じ相対PPG振幅を示すが、諸ボクセルにおけるクロマティックな差は脈振幅における変動を生じる。よって、CHROMとは異なり、ボクセルにおける平均RGBの時間微分が、そのクロミナンス信号を導出するために使われる。すなわち、

【0079】

【数4】

$$dC_{ij}^t = C_{ij}^t - C_{ij}^{t-1}, C \in \{\bar{R}, \bar{G}, \bar{B}\}$$

20

期間T(式1に関して定義した)において、規格化されたクロミナンス微分は

【0080】

【数5】

$$\begin{cases} \vec{dX}_{ij}^T = 3 \frac{\vec{dR}_{ij}^T}{\sum_{t=0}^T dR_{ij}^t} - 2 \frac{\vec{dG}_{ij}^T}{\sum_{t=0}^T dG_{ij}^t} \\ \vec{dY}_{ij}^T = 1.5 \frac{\vec{dR}_{ij}^T}{\sum_{t=0}^T dR_{ij}^t} + \frac{\vec{dG}_{ij}^T}{\sum_{t=0}^T dG_{ij}^t} - 1.5 \frac{\vec{dB}_{ij}^T}{\sum_{t=0}^T dB_{ij}^t} \end{cases} \quad (4)$$

30

として計算される。ここで、 $(dR_{ij}^t, dG_{ij}^t, dB_{ij}^t)$ は二つの画像フレームの間でのボクセル中のRGBの時間微分を表わす。各区間において推定されたクロミナンスの微分は線形結合されて脈微分とされ、さらに積分される。その後、異なる脈期間が重複加算されて、長さLをもつ完全な脈信号

【0081】

【数6】

$$\vec{S}_{ij}^L$$

40

にされる。この手順は、次のように解釈できる。

【0082】

【数7】

$$\vec{S}_{ij}^L = \sum_{t=0}^{L-T+1} \vec{S}_{ij}^{t+T} + w \cdot \text{csum}(\vec{dX}_{ij}^T - \frac{\sigma(\vec{dX}_{ij}^T)}{\sigma(\vec{dY}_{ij}^T)} \vec{dY}_{ij}^T) \quad (5)$$

ここで、 $\text{csum}(\cdot)$ は時間微分信号の累積和(cumulative sum)を表し；wは重複加算を平滑化するためのハニング窓である。結果として、(i番目のスケールでのj番目のボクセルからの)並列抽出された脈信号

50

【 0 0 8 3 】

【 数 8 】

$$\vec{S}_{ij}^L$$

は次のように中心化および規格化される。

【 0 0 8 4 】

【 数 9 】

$$\vec{S}_{ij}^L = \frac{\vec{S}_{ij}^L - \mu(\vec{S}_{ij}^L)}{\sigma(\vec{S}_{ij}^L)} \quad (6)$$

10

ここで、 $\mu(\cdot)$ は平均演算を表わす。この方法では使用された唯一の特徴は脈信号であることを注意しておく。色またはテクスチャーのような他の外観特徴は使われていない。

【 0 0 8 5 】

スペクトル解析

本節は、スペクトル解析段 3 8 を記述する。これは、類似性行列を形成し、類似性行列に対してインクリメンタル疎PCAを実行し、階層的融合を使ってビデオ・シーケンスにおける生体皮膚組織の領域を識別するという三つのサブ段を含む。

【 0 0 8 6 】

同じ被写体に属する皮膚領域から抽出された脈信号が位相および周波数のような多くの側面において類似性を共有するが着目されている。一方、皮膚以外の領域（たとえば背景）から抽出された脈信号は相関のないランダム・ノイズである。よって、階層的ボクセルから脈信号を抽出した後、脈信号のペアごとの類似性が、生きている被写体を見出すために使われる。これは、ビデオ・シーケンスに複数の被写体がいる場合にも適用可能である。異なる被写体から測定された脈も位相および周波数で区別できる。

20

【 0 0 8 7 】

類似性行列 この段階では、測定された脈に基づいて階層的ボクセルを相互に関連させる類似性行列 $= (D, C)$ が生成される。において、対角線トレースにおける諸要素Dは種々のスケールにおけるすべてのボクセルを含み；残りの諸要素Cは任意の対のボクセルの間のペアごとの連関を表わす。そのような類似性行列を構築するために、脈類似性を測るための距離メトリックが定義される必要がある。最も一般的に使われる距離メトリック、すなわちL1およびL2距離は、脈特徴には適用可能でない。しかしながら、他の外観特徴（たとえば非特許文献 6、7）に比べて、脈特徴における本質的で独特な性質は、周期性を含むということである。同じ被写体からの脈信号は以下の関係を示すことが着目されている：（1）類似した周波数をもち、よってその相互相関は有意なスペクトル・ピークを呈する；（2）有意な位相シフトをもたない；（3）周波数相関が規則的であり、乱れがより少ない；（4）脈信号を多次元ベクトルとして考える場合、二つの類似したベクトルには含まれる角度は小さい。したがって、脈信号についての類似性行列を構築するために使われる好ましい距離メトリックは、上記の連関（connections）を強調し、四つの異なる測定値から構成される。

30

40

【 0 0 8 8 】

・ スペクトル・ピーク 周波数領域において、諸ボクセルが連通する（communicate）ための脈拍数帯域 f [40,240]BPM（Beats Per Minute [拍毎分]）が定義される。これは、新生児およびスポーツをしている被写体を含む健康な被写体についての広い範囲である。二つの相互相関を取った脈信号のスペクトル・ピークは次のように定義される。

【 0 0 8 9 】

【 数 1 0 】

$$F = \arg \max_{f \in [40, 240]} (\mathcal{F}(\vec{S}_{ij}^L) \circ \mathcal{F}(\vec{S}_{i'j'}^L)^*) \quad (7)$$

50

ここで、白丸は要素ごとの積を表わし、*は共役を表わし、 $F(\cdot)$ は高速フーリエ変換（FFT）を表わす。

【 0 0 9 0 】

・ スペクトル位相 二つの類似した脈信号は同じ位相でもあるので、その規格化された相互相関（normalized cross-correlation）は

【 0 0 9 1 】

【 数 1 1 】

$$P = \max(\mathcal{F}^{-1}(NCC)) \quad (8)$$

ここで

$$NCC = \frac{\mathcal{F}(\vec{S}_{ij}^L) \circ \mathcal{F}(\vec{S}_{i'j'}^L)^*}{\|\mathcal{F}(\vec{S}_{ij}^L) \circ \mathcal{F}(\vec{S}_{i'j'}^L)^*\|_2} \quad (9)$$

10

としての時間領域における強い応答を示すはずである。ここで、 $\|\cdot\|_2$ はL2ノルムであり、 $\mathcal{F}^{-1}(\cdot)$ は逆FFTを表わす。

【 0 0 9 2 】

・ スペクトル・エントロピー 用語「エントロピー」は、二つの脈信号の間の相関の規則性を

【 0 0 9 3 】

【 数 1 2 】

$$E = \frac{\sum_{f=40}^{240} NCC(f) \log(NCC(f))}{\log(240 - 40)} \quad (10)$$

20

として測るために使われている。ここで、Eの解釈は、他の測定と整合する。すなわち、大きなEほどよりよい相関を表わす。

【 0 0 9 4 】

・ 内積 時間領域において、二つの脈信号の間の余弦角を測るために内積を

【 0 0 9 5 】

【 数 1 3 】

$$I = \left\langle \frac{\vec{S}_{ij}^L}{\|\vec{S}_{ij}^L\|_2}, \frac{\vec{S}_{i'j'}^L}{\|\vec{S}_{i'j'}^L\|_2} \right\rangle \quad (11)$$

30

として使う。ここで、 \langle, \rangle は内積演算を表わす。

【 0 0 9 6 】

最後に、これら四つの測定値が範囲[0,1]に規格化されて、ガウシアン・カーネルを用いて

【 0 0 9 7 】

【 数 1 4 】

$$\Sigma = 1 - \exp\left(-\frac{(F \circ P \circ E \circ I)^2}{2\sigma_{I,F,P,E}^2}\right) \quad (12)$$

40

として一緒に融合される。ここで、 $\sigma_{I,F,P,E}$ は四つの行列の間の、要素ごとの標準偏差を表わす。四つの測定値は互いに完全に独立ではないことを注意しておくべきである。測定値の間の冗長性は、類似性推定における不確定性を減らすために有益である。図7は、二つのビデオ・シーケンスについての、四つの測定値およびその融合された類似性行列

44、46の例を示している。ビデオ・シーケンスの一方は単一の生きている被写体を含み、他方は二人の生きている被写体を含む。より高いエネルギーをもつ要素が、その階

50

層における類似しているボクセルのインデックスを表わす。

【 0 0 9 8 】

本稿で使われる距離メトリックでは、二つのよく揃った脈信号は、相互相関の際に増強された周波数エネルギーを示し、これは事実上、ノイズ要素（たとえば、脈のないボクセル）を抑制することができる。対照的に、以前の距離メトリックはみな、比較における類似した要素の間の連関を高めることができない客観的な測定値である。結局、階層におけるすべてのボクセルは、類似性行列において相互に関連している。諸ボクセルにおいて生きている被写体を検出するタスクは、同じ部分空間内の要素が同一の類似性方向をもつような類似性行列の部分空間分割を見出すこととして、再定式化できる。

【 0 0 9 9 】

インクリメンタル疎行列分解 (incremental sparse matrix decomposition) 類似性行列 4×4 、 4×6 は、 $x_1 x_1^T + x_2 x_2^T + \dots + x_n x_n^T$ の線形結合として解釈できる。ここで、 x_i は多次元空間における直交ベクトルの集合である。同じ被写体に属するボクセルを見出すために、 X を X に因子分解するために行列分解技法が使われる。ここで、異なる被写体が異なる固有ベクトルに分離される。 X は多くの 0 の要素をもつ疎な行列である（たとえば、背景を指すボクセルは類似性を共有しない）から、表現力とデータ解釈可能性との間のトレードオフを追求することによって X に分解するために疎 PCA が適用される。

【 0 1 0 0 】

疎 PCA は非特許文献 8 に記載されている。疎 PCA は、次の凸でない目的関数を最適化することによって、 X における最大分散をもつ第一の疎な固有ベクトルを見出す：

【 0 1 0 1 】

【数 1 5】

$$\arg \max_X (X^T \Sigma X) \text{ subj.to } \|X\|_2 = 1, \|X\|_1 \leq n \quad (13)$$

ここで、 $\|\cdot\|_2$ は L1 ノルムであり、 $n > 0$ は X の濃度 (cardinality) を制御する。しかしながら、最大分散をもつ疎な固有ベクトルを計算することは組み合わせ問題であり、数値的に解くのが困難である。よって、

【 0 1 0 2 】

【数 1 6】

$$\arg \max_{\hat{\Sigma}} \text{Tr}(\Sigma \hat{\Sigma}) - \rho \|\hat{\Sigma}\|_1 \text{ subj.to } \text{Tr}(\hat{\Sigma}) = 1, \hat{\Sigma} \succeq 0 \quad (14)$$

のように、 l_1 ペナルティ付与を用いた半定符号の (semidefinite) 緩和のためのリフティング (lifting) 手順に続いて、式 (13) における凸でない階数制約条件が落とされる。ここで、 $\text{Tr}(\cdot)$ は行列トレース演算を表わし、 $\rho > 0$ は疎さを制御し、

【 0 1 0 3 】

【数 1 7】

$$\hat{\Sigma} = X X^T$$

は第一の先頭の固有ベクトルによって近似される対称行列である。この時点で、式 (14) を解くために、ハイブリッド条件付き勾配平滑化 (HCGS: Hybrid Conditional Gradient Smoothing) というアルゴリズムを使うことができる。HCGS は非特許文献 9 に記載されている。HCGS の利点は、条件付き勾配手法を使った凸緩和 (convex relaxation) における速い収束である。

【 0 1 0 4 】

しかしながら、実際上は、 X は、複数の被写体の場合、多重疎固有基底 (multiple sparse eigenbasis) からなることがある。一方、式 (14) は第一の先頭の固有ベクトルにおける疎さを促進するだけである。この問題に対処するため、その後の疎な固有ベクトル x_i は、以前の疎な固有ベクトル x_1, x_2, \dots, x_{i-1} を使い、ホテリングの収縮 (Hotelling's

10

20

30

40

50

deflation) を使って

【 0 1 0 5 】

【 数 1 8 】

$$\Sigma_i = \Sigma_{i-1} - (x_i^T \Sigma_{i-1} x_i) x_i x_i^T, i \in [1, m] \quad (15)$$

ここで

$$m = \arg \max_i \left(\frac{x_{i-1}^T \Sigma_{i-1} x_{i-1}}{1 + x_i^T \Sigma_i x_i} \right) \quad (16)$$

として、逐次的に を収縮させることによって推定される。ここで、

【 0 1 0 6 】

【 数 1 9 】

$$x_i \in X$$

はHGCSにおいて冪逐次反復 (power iteration) によって導出できる。mは、最も表現力のある (expressive) 固有ベクトルの自動的に見出された数であり、これはビデオ・シーケンスにおける被写体の数をも含意する。すなわち、mは通例、最大の固有値ギャップのところに見出される。

【 0 1 0 7 】

図 8 は、図 7 に示される二つの類似性行列 4 4、4 6 について、インクリメンタル疎PC Aを使っての類似性行列分解の例を示している。ここで、類似するボクセルは、選択された固有ベクトルにおける同じ方向に因子分解される。このように、図 8 は、類似性行列からの、因子分解され選択された固有基底を示している。もとの 4 4、4 6 におけるノイズのある要素は ^ 付きの 4 8、5 0 では消去されている。固有値 (それぞれグラフ 5 2 および 5 4 に示される) は、(それぞれグラフ 5 6 および 5 8 に示される) ^ 付きのにおける最も表現力のある固有ベクトルの数を明確に示している。

【 0 1 0 8 】

実のところ、いくつかの内在的 (たとえば脈拍数変動) および外来的 (たとえばルミナンス変化) 要因が、その後のフレームにおいて時折類似性行列を変化させることがある。このことは、それぞれの単一フレームから推定される疎な固有ベクトルの不安定性につながる。この問題を解決するために、 x_i Xを時間領域におけるリアルタイム変化になめらかに適合させるために、インクリメンタルな部分空間更新が用いられる。基本的には、これは時間変化する類似性行列

【 0 1 0 9 】

【 数 2 0 】

$$\hat{\Sigma}_{new}$$

を新たな観察として考慮し、異なるフレームからの複数の観察

【 0 1 1 0 】

【 数 2 1 】

$$[\hat{\Sigma}_{old}, \hat{\Sigma}_{new}]$$

を、部分空間モデルを

【 0 1 1 1 】

【 数 2 2 】

$$[U, D, V] = \text{SVD}([\hat{\Sigma}_{old}, \hat{\Sigma}_{new}]) \quad (17)$$

のように豊かにするために使う。ここで、SVD(・)は、特異値分解を表わし; UおよびDはインクリメンタルに更新された固有ベクトルおよび固有値である。時間変化する類似性行列から複数の疎な固有ベクトルをインクリメンタルに推定するための例示的なアルゴリズム

10

20

30

40

50

ムが下記のアルゴリズム 1 に示されている。

【 0 1 1 2 】

【 表 1 】

入力: 類似性行列 $\Sigma \in \mathbb{R}^{n \times n}$, 固有ベクトル U ,
固有値 D

1: $\rho = 0.25$ (疎さ), $N = 100$ (反復回数)

2: **for** $k = 1, 2, \dots, N$ **do**

3: $\beta_k = \frac{1}{n} \text{Tr}(\Sigma \Sigma_k) + \frac{\rho}{n} \|\Sigma_k\|_1$

4: $Z_k = \Sigma - \frac{\sqrt{k}\rho}{2\sqrt{2}} \Sigma_k + \frac{\sqrt{k}\rho}{2\sqrt{2}} \text{sign}(\Sigma_k) \odot (|\Sigma_k| - \frac{2\sqrt{2}}{n\sqrt{k}})$

5: $X = \{x_1, x_2, \dots, x_m\} \leftarrow Z_k$ の複数の固有ベクトル。

m は式(15)および式(16)によって決定される

6: $\hat{\Sigma} = X X^T$

7: $\Sigma_{k+1} = (1 - \frac{2}{k+1}) \Sigma_k + \frac{2}{k+1} \hat{\Sigma}$

8: **if** $\frac{|\beta_k - \beta_{k-1}|}{\beta_k} < 10^{-3}$ **then**

9: **break**

10: **end if**

11: **end for**

12: **if** $U, D == 0$ **then**

13: $[U, D, V] = \text{SVD}(\hat{\Sigma})$

14: **else**

15: $[U, \hat{\Sigma}'] R = \text{QR}([UD, \hat{\Sigma}]) \leftarrow \text{QR 分解によって解かれる}$

16: $[U', D', V'] = \text{SVD}(R)$

17: $U'_m \in U', D'_m \in D' \leftarrow$ 上位 m 個の固有ベクトルおよび
固有値を選択。ここで、 m は式(16)によって決定される

18: $U = \text{sign}(U) \odot |[U', \Sigma'_{new}] U'_m|, D = D'_m \leftarrow$ 部分空間
モデルを更新

19: **end if**

出力: U および D を更新

アルゴリズム 1

【 0 1 1 3 】

階層的融合 推定された疎な固有ベクトル 5 6、5 8 を階層的ボクセルに投影することによって、複数のスケールにおけるボクセルに基づく人間のオブジェクトネス・マップが得られる。ここで、各スケールは被写体に対する異なる量子化された記述をもつ。図 7 および図 8 からの二人の被写体を含むビデオ・シーケンスについてのこの投影が図 9 に示されている。固有ベクトル 5 8 は諸部分空間における被写体方向（符号）を決定するのみならず、対応する皮膚領域の脈動性（振幅）をも決定する。すなわち、額および頬は、投影において相対的に高い脈動性を示す。最終段階は、複数のオブジェクトネス・マップを単一の出力に融合するものである。階層的測定が皮膚領域についての統計的な観察を生成するという事実のため、基本的な発想は、この冗長性を活用して、すべてのスケールが最高の一致をもつ単一の出力を導出するということである。この意味で、階層的な融合は、マルチスケール・オブジェクトネス・マップの間のエネルギー最小化として、次のように書ける。

【 0 1 1 4 】

10

20

30

40

【数 2 3】

$$\arg \min_{\hat{o}} (\gamma E_1 + (1 - \gamma) E_2) \quad (18)$$

ここで

$$\begin{cases} E_1 = \sum_i \sum_j \|o_{ij}, \hat{o}\|_2 \\ E_2 = \sum_i (\sum_{\substack{j \\ o_{ij} \subseteq o_{i-1,j}}} \|o_{ij}, o_{i-1,j}\|_2 + \sum_{\substack{j \\ o_{i+1,j} \subseteq o_{ij}}} \|o_{ij}, o_{i+1,j}\|_2) \end{cases} \quad (19)$$

ここで、 o_{ij} は、固有ベクトル要素によって決定される*i*番目のスケールにおける*j*番目のボクセルのオブジェクトネス値に対応し； \wedge 付きの0は融合されたオブジェクトネス・マップを表わし； γ は二つのエネルギー項の間のバランスを制御する。式(19)において、 E_1 は入力と出力の間のエネルギーを最小化し、一方、 E_2 は異なるスケールにおける空間的に重なったボクセルどうしの間のエネルギーを最小化する。すなわち、暗黙的な木構造である。図9は、二人の生きた被写体がいるビデオ・シーケンスにおける融合された結果の例を示している。ここで、識別された被写体のそれぞれについて、別個の出力60、62が提供される。

10

【0115】

上記の方法およびVPS法の好ましい実施形態は、ビデオ・シーケンスにおいて生体皮膚組織を識別するための改善された方法および装置を提供する。特に、本方法は、通常の技法に比べて、改善された生体皮膚組織検出率を提供し、検出は生体組織を検出するために脈信号を使うことのみに基づく。これらの改善は、ビデオ・シーケンスにおける被写体または被写体の背景のスケール（撮像ユニット4からの距離）、姿勢、位置、皮膚色調、可視の身体部分または動き、被写体が撮像ユニット4から部分的に隠蔽されているかどうか、ビデオ・シーケンスに人工的な顔または身体部分が存在するかどうかまたはビデオ・シーケンスに複数の生きた被写体があるかどうかに関わりなく得られる。

20

【0116】

図面、本開示および付属の請求項を吟味することから、開示される実施形態に対する変形が、特許請求される発明を実施する当業者によって理解され、実施されることができる。請求項において「有する／含む」という語は他の要素や段階を排除するものではなく、単数形の表現は複数を排除するものではない。単一のプロセッサまたは他のユニットが請求項において記載されているいくつかの項目の機能を充足してもよい。ある種の施策が互いに異なる従属請求項に記載されているというだけの事実がこれらの施策の組み合わせが有利に使用できないことを示すものではない。コンピュータ・プログラムは、光記憶媒体または半導体媒体のような、他のハードウェアと一緒にまたは他のハードウェアの一部として供給される好適な媒体上で記憶／頒布されてもよいが、インターネットまたは他の有線もしくは無線の遠隔通信システムを通じてなど、他の形で配送されてもよい。請求項に参照符号があったとしても、範囲を限定するものと解釈されるべきではない。

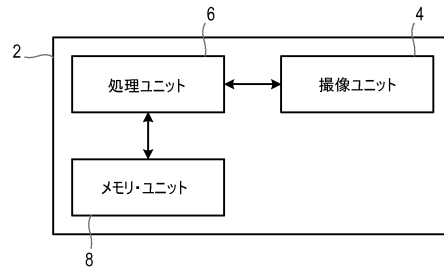
30

【図 1】

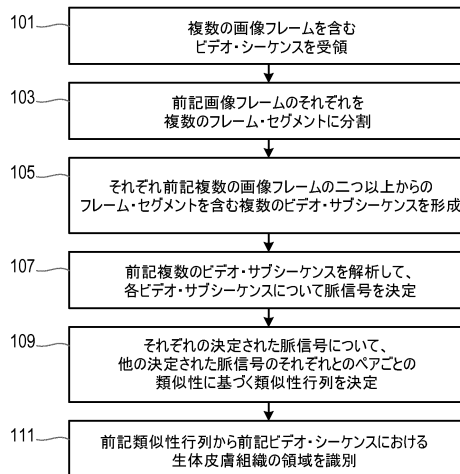


Figure 1

【図 2】



【図 3】



【図 4 a】

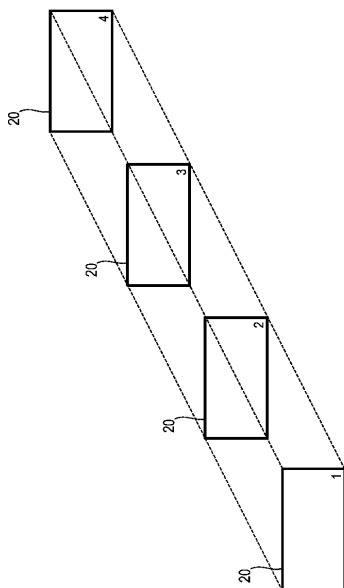


Figure 4(a)

【図 4 b】

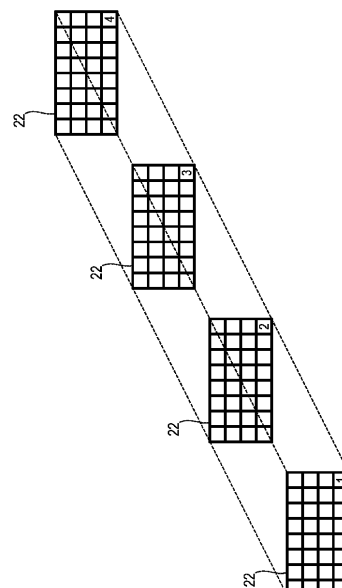


Figure 4(b)

【図 4 c】

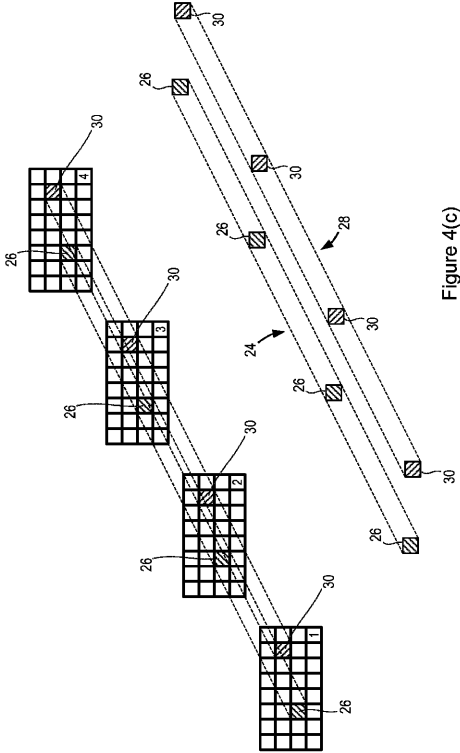
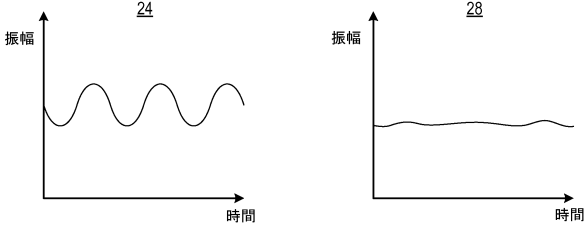
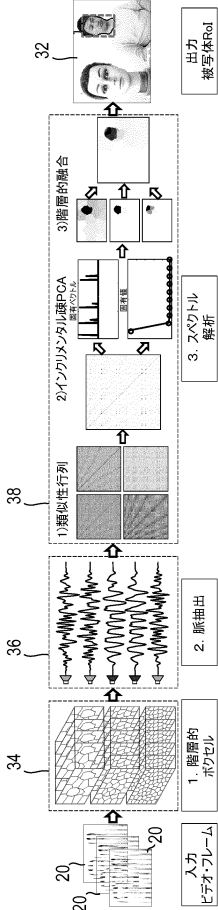


Figure 4(c)

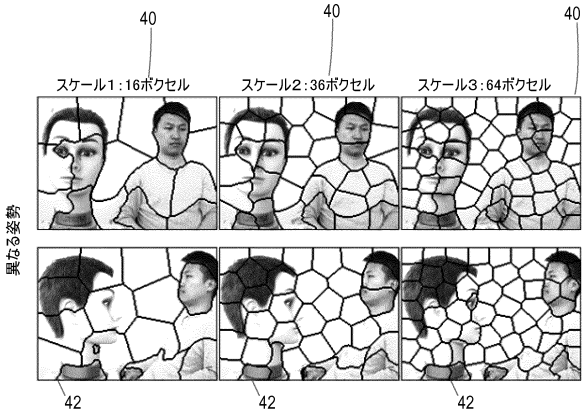
【図 4 d】



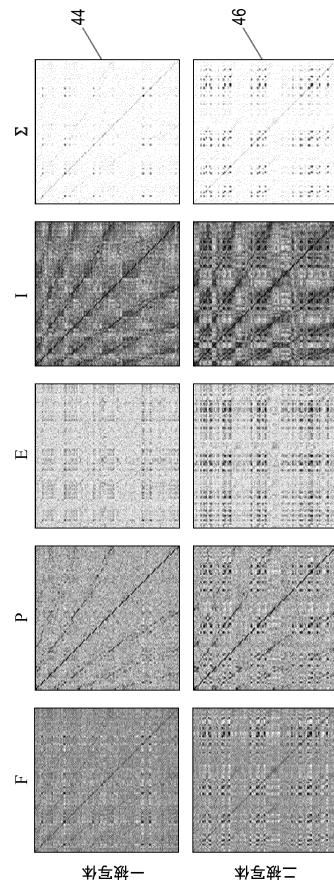
【図 5】



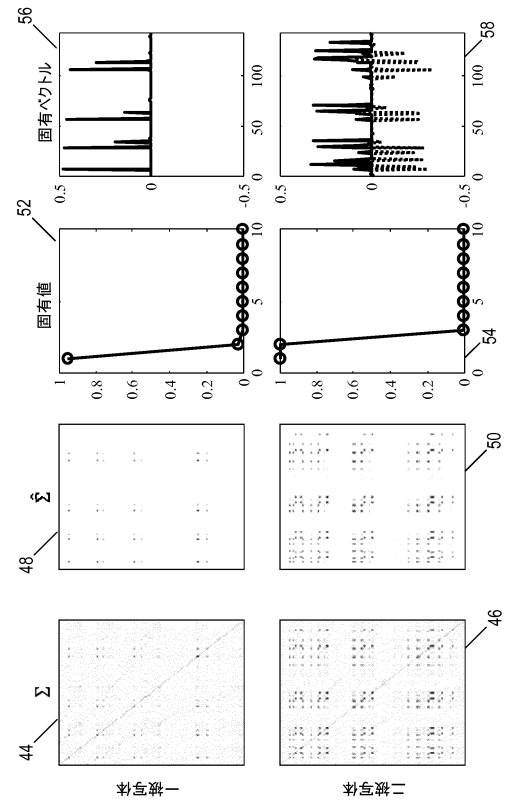
【図 6】



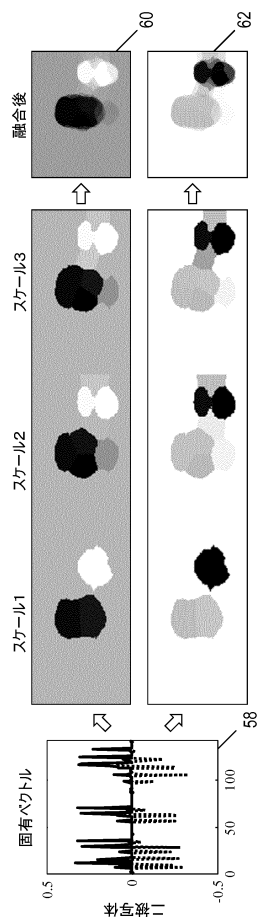
【図 7】



【図 8】



【図 9】



フロントページの続き

(74)代理人 100091214

弁理士 大貫 進介

(72)発明者 ワン, ウェンジン

オランダ国 5 6 5 6 アーエー アインドーフエン ハイテック キャンパス 5

(72)発明者 デ ハーン, ヘラルト

オランダ国 5 6 5 6 アーエー アインドーフエン ハイテック キャンパス 5

審査官 笠田 和宏

(56)参考文献 国際公開第 2 0 0 7 / 0 4 3 3 2 8 (W O , A 1)

国際公開第 2 0 1 1 / 0 4 2 8 3 9 (W O , A 1)

特開 2 0 0 7 - 2 9 5 9 4 6 (J P , A)

国際公開第 2 0 1 4 / 0 4 5 2 1 7 (W O , A 1)

(58)調査した分野(Int.Cl., D B 名)

I P C A 6 1 B 5 / 0 0

G 0 6 T 7 / 0 0 - 7 / 9 0