

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4725884号
(P4725884)

(45) 発行日 平成23年7月13日(2011.7.13)

(24) 登録日 平成23年4月22日(2011.4.22)

(51) Int.Cl.

F I

G 0 6 T 1/00 (2006.01)

G 0 6 T 1/00 3 4 0 Z

請求項の数 9 外国語出願 (全 14 頁)

(21) 出願番号 特願2005-91690 (P2005-91690)
 (22) 出願日 平成17年3月28日(2005.3.28)
 (65) 公開番号 特開2006-277022 (P2006-277022A)
 (43) 公開日 平成18年10月12日(2006.10.12)
 審査請求日 平成20年2月28日(2008.2.28)

(出願人による申告)平成16年度独立行政法人情報通信研究機構、研究テーマ「人間情報コミュニケーションの研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願

(73) 特許権者 393031586
 株式会社国際電気通信基礎技術研究所
 京都府相楽郡精華町光台二丁目2番地2
 (74) 代理人 100099933
 弁理士 清水 敏
 (72) 発明者 サブリ・グルプズ
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内
 (72) 発明者 川戸 慎二郎
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内
 (72) 発明者 木下 敬介
 京都府相楽郡精華町光台二丁目2番地2
 株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54) 【発明の名称】 口画像に対する任意変形に関するパラメータを推定する方法

(57) 【特許請求の範囲】

【請求項1】

ユーザのビデオ画像に視覚的音声前処理を行なうシステムにおいて、ユーザの口画像に対する任意変形に関するパラメータを推定する方法であって、

前記ビデオ画像の各々に対し、

前記ユーザの前記画像中の目印となる位置2箇所を決定するステップと、

前記ビデオ画像中のユーザの口の輪郭を決定するステップと、

前記口の輪郭の位置を利用することにより、前記任意変形と逆の変形が前記ビデオ画像に対し適用可能となるように、前記ユーザの口画像に関する変形パラメータを推定するパラメータ推定ステップとを含み、

前記パラメータ推定ステップは、前記画像中の前記ユーザの前記口画像に対する主成分分析の結果を利用して、前記ユーザの前記画像の回転量を推定する第1の回転量推定ステップを含む、方法。

【請求項2】

前記推定するステップは、さらに、

前記ユーザの画像中の、前記目印を利用して、ユーザの画像の回転量を推定する第2の回転量推定ステップと、

前記第1の回転量推定ステップで推定された回転量と、前記第2のステップで推定された回転量とを組合せることにより、前記ユーザの画像の回転角を推定するステップとを含む、請求項1に記載の方法。

10

20

【請求項 3】

前記第 1 の回転量推定ステップが、

前記目印の位置を基準として前記ユーザの画像中の口画像の領域を推定する口領域推定ステップと、

推定された前記口画像中のエッジを検出することにより前記口唇の外側輪郭の座標位置を決定するステップと、

前記唇の外側輪郭のサンプル点の平均と共分散行列とを計算するステップと、

前記共分散行列の第一の固有ベクトル e_1 を求めるステップと、

角度 $\alpha = \arctan(e_y / e_x)$ を計算するステップとを含み、

ここで e_x および e_y は前記第一の固有ベクトルの x 軸成分および y 軸成分であり、前記 x 軸及び前記 y 軸はそれぞれ、前記画像中の標準座標系を規定する、請求項 1 または請求項 2 に記載の方法。

10

【請求項 4】

前記口領域推定ステップが、

前記画像中の前記目印間の距離を、前記画像中の前記目印の前記位置と、推定された変形とを利用して計算するステップと、

前記目印を基準として定められる領域であって、さらに計算で求められた前記目印間の距離と同じ幅となるような距離で選択された幅を有する、口領域を選択するステップとを含む、請求項 3 に記載の方法。

【請求項 5】

20

前記口領域推定ステップがさらに、所定の定数と、前記計算で求められた前記目印間の距離との比率によって、前記口画像のスケール係数を推定するスケール係数推定ステップとを含む、請求項 4 に記載の方法。

【請求項 6】

前記スケール係数推定ステップにおいて推定された前記口画像のスケール係数によって、前記選択された口領域をスケールするステップをさらに含む、請求項 5 に記載の方法。

【請求項 7】

目印となる位置 2 箇所を決定する前記ステップは、前記ユーザの前記画像中の瞳の位置を決定するステップを含む、請求項 1 ~ 請求項 6 のいずれかに記載の方法。

30

【請求項 8】

前記ユーザの画像中における、前記口領域に対する平行移動の影響を除去するために、前記口領域の画像の移動量を計算するステップをさらに含む、請求項 1 ~ 請求項 7 のいずれかに記載の方法。

【請求項 9】

視覚的前処理に用いるプログラム製品であって、

プロセッサを用いたシステムに、請求項 1 ~ 請求項 8 のいずれかに記載のステップのすべてを行なわせる命令を記憶したコンピュータ読取可能媒体を含む、プログラム製品。

【発明の詳細な説明】

【技術分野】

40

【0001】

この発明は、自然な発声条件下で一貫した視覚的な音声特徴量の抽出を行なうことに関する。より正確には、この発明は頭部の画像の平行移動、回転およびスケールの微妙な違い(similarity)をすぐさま修正することによって、話者が自身の頭を自由に動かしても差し支えないようにする、視覚的な音声特徴量の抽出に関する。

【背景技術】

【0002】

視覚的な音声処理は、マン - マシンインターフェイスの研究において重要な課題の 1 つである。例えば、聴覚的情報に視覚的情報を加えることは少なくとも 2 つの実用的な基準に合致する。すなわち、これは音声認識における人の視覚的な知覚を模倣し、さらに、これは

50

音響信号に常に存在するとは限らない情報を含み得るからである〔非特許文献 2、8、15 を参照〕。

【 0 0 0 3 】

この統合手法は主に、非特許文献 9 及び非特許文献 13 に示される初期段階での統合と最終段階での統合とに分類され、視覚的音声特徴量抽出アルゴリズムは主に、モデル（輪郭、スネーク等）を用いる方法と、外観（画素）を用いる方法とに分類される。詳細は非特許文献 11、12、及び 16 を参照されたい。

【 0 0 0 4 】

モデルを用いるアプローチでも外観を用いるアプローチでも、自然な発声条件下においては、一貫した視覚的音声特徴量を抽出することは、制御された条件下におけるよりも困難である。報告されたさまざまな視覚的特徴量抽出方法及び 2 つの統合手法では、実験作業は制御された環境で行なわれるものとされ（例えば非特許文献 9、12 及び 16 を参照）、話者には頭部を自然に動かしながら話すことが許されない。

【 0 0 0 5 】

これに対して、自然な条件下では、話者は自然に動く。従って、口の画像の微妙な違いを修正することは、視覚的音声処理のために一貫した視覚的特徴量を抽出するには決定的に重要な課題である。

【非特許文献 1】S. グルブズ、K. キノシタ及び S. ヤノ、「トレーニング可能な多変量ガウス分類器を用いたビデオシーケンスからの口トラッキング」、PRMU 2003、仙台、日本、2003 年 12 月 (S. Gurbuz, K. Kinoshita and S. Yano. Mouth Tracking from Video Sequences using Trainable Multivariate Gaussian Classifiers. PRMU 2003, Sendai, Japan, December 2003.)

【非特許文献 2】A. アジャウダーニ及び C. ベノイト、「hmm ベースの asr における聴覚及び視覚パラメータの統合について」、人と機械による音声読取：モデル、システム及び応用、NATO ASI シリーズ、シリーズ F、コンピュータ及びシステムサイエンス、150 号、461 - 471 ページ、1996 年 (A. Adjoudani and C. Benoit. On the integration of auditory and visual parameters in an hmm-based asr. In Speechreading by Humans and Machines: models, systems, and applications, NATO ASI Series. Series F, Computer and Systems Sciences, number 150, pages 461-471, 1996.)

【非特許文献 3】C. ブレグラー、H. ヒルド、S. マンケ、及び A. ワイベル、「唇の読取による、連続文字認識の改良」、音響音声及び信号処理国際会議予稿集、1993 年 (C. Bregler, H. Hild, S. Manke, and A. Waibel. Improving connected letter recognition by lipreading. In Proceedings of International Conference on Acoustics Speech and Signal Processing, 1993.)

【非特許文献 4】G. イェンガー、G. ポタミアノス、C. ネティ、T. ファリク、及び A. ヴァーマ、「自動音声読取のための頑健な視覚的 roi 検出」、IEEE マルチメディア信号処理第 4 回ワークショップ、79 - 84 ページ、2001 年 (G. Iyengar, G. Potamianos, C. Neti, T. Faruque, and A. Verma. Robust detection of visual roi for automatic speechreading. In IEEE Fourth Workshop on Multimedia Signal Processing, pages 79-84, 2001.)

【非特許文献 5】A. J. ジェイン、「デジタル画像処理の基礎」、プレントイスホール、1997 年 (A. K. Jain. Fundamentals of Digital Image Processing. Prentice-Hall, 1997.)

【非特許文献 6】M. カス、A. ウィトキン及び D. ターゾプーロス、「スネークス：活性輪郭モデル」、コンピュータビジョンに関する第 1 回国際会議予稿集、259 - 268 ページ、1987 年 (M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In Proceedings of 1st International Conference on Computer Vision, pages 259-268, 1987.)

【非特許文献 7】S. カワト、N. テツタニ、「SSR フィルタ及びサポートベクトルマ

10

20

30

40

50

シンを用いたリアルタイムのスケール適応顔検出及びトラッキング」、コンピュータビジョンに関するアジア会議、2004年(S. Kawato, N. Tetsutani. Scale adaptive face detection and tracking in real time with SSR filter and support vector machine. In Asian Conference on Computer Vision, 2004.)

【非特許文献8】H. マクガーク及びJ. マクドナルド、「唇を聴き声を見る」、ネイチャー、264:746-748、1976年(H. McGurk and J. MacDonald. Hearing lips and seeing voices. Nature, 264:746-748, 1976.)

【非特許文献9】S. ナカムラ、「聴覚的-視覚的音声処理のための統計的マルチモーダル統合」、IEEE 神経回路網トランザクション、13(4)、2002年(S. Nakamura. Statistical multimodal integration for audio-visual speech processing. IEEE Transactions on Neural Networks, 13(4), 2002.)

10

【非特許文献10】E. パターソン、S. グルブズ、Z. チュフェキ、及びJ. ガウディ、「移動する話し手、話者独立の特徴量研究とCuaveマルチモーダル音声コーパスを用いたベースライン結果」、EURASIP 応用信号処理ジャーナル、2002(11)、2002年(E. Patterson, S. Gurbuz, Z. Tufekci, and J. Gowdy. Moving-talker, speaker-independent feature study and baseline results using the cuave multimodal speech corpus. EURASIP Journal on Applied Signal Processing, 2002(11), 2002.)

【非特許文献11】E. パタジャン、B. ビショフ、D. ボドフ及びN. ブルック、「音声認識を向上するための改良された自動唇読取システム」、コンピュータシステムにおけるヒューマンファクタに関する会議、19-25ページ、1988年(E. Petajan, B. Bischoff, D. Bodofo, and N. Brooke. An improved automatic lipreading system to enhance speech recognition. In Conference on Human Factors in Computing Systems, pages 19-25, 1988.)

20

【非特許文献12】G. ポタミアノス、J. ルティン、及びC. ネティ、「聴覚的-視覚的lvcsrのための階層的判別特徴量」、音響音声及び信号処理国際会議予稿集、2001年。(G. Potamianos, J. Luettin, and C. Neti. Hierarchical discriminant features for audio-visual lvcsr. In Proceedings of International Conference on Acoustics Speech and Signal Processing, 2001.)

【非特許文献13】D. ロイ、「相互情報を用いた音声と視覚の統合」、音響音声及び信号処理国際会議予稿集、2000年(D. Roy. Integration of speech and vision using mutual information. In Proceedings of International Conference on Acoustics Speech and Signal Processing, 2000)

30

【非特許文献14】A. W. シニア、「顔認識システムのための顔と特徴量の検出」、聴覚及び視覚ベースのバイオメトリック人物認証(AVBPA)予稿集、154-159ページ、1999年(A. W. Senior. Face and feature finding for face recognition system. In Proceedings of Audio-and Video-Based Biometric Person Authentication (AVBPA), pages 154-159, 1999.)

【非特許文献15】E. バティキオティス-ベイツィン、K. G. ムンハル、M. ヒラヤマ、Y. V. リー、及びD. ターゾポラス、「音声におけるオーディオビジュアルな挙動の力学」、人と機械による音声読取：データ、モデル及びシステム、D. G. ストーク及びM. E. ヘンケ編、NATOSプリンガーフェアラーク、ニューヨーク、NY(1996)、150巻、1996年(E. Vatikiotis-Bateson, K. G. Munhall, M. Hirayama, Y. V. Lee, and D. Terzopoulos. The dynamics of audiovisual behavior in speech. In Speechreading by Man and Machine: Data, Models and Systems, D. G. Stork and M. E. Hennecke Eds. NATOSpringer-Verlag, New York, NY (1996), volume 150, 1996.)

40

【非特許文献16】X. ツァン、R. M. マーセロー、M. クレメンツ、及びC. C. ブラウン、「改良された音声認識のための視覚的音声特徴量の抽出」、音響音声及び信号処理国際会議、2002年(X. Zhang, R. M. Mersereau, M. Clements, and C.

50

C. Brown. Visual speech feature extraction for improved speech recognition. In International Conference on Acoustics Speech and Signal Processing, 2002.)

【発明の開示】

【発明が解決しようとする課題】

【0006】

発明者らが知る限りでは、視覚的音声処理のために口画像の微妙な違いを修正することに関する研究はこれまでなされていない。

【0007】

この発明は、一貫した視覚的な音声特徴量の抽出を容易にすることにより、聴覚と視覚とを統合して自然な音声のための音声認識装置の性能を改良するためになされた。この目的のためには、口画像への平行移動の影響、スケーリングの影響及び回転の影響をなくして、視覚的音声処理用に一貫した視覚的特徴を抽出可能な技術が必要である。

【0008】

従って、この発明の目的の1つは、自然な条件下で話している話者の口画像に対する微妙な違いの影響を、すぐさまなくすためのコンピュータプログラムを提供することである。

【0009】

この発明の別の目的は、口の境界を検出し、口画像領域の歪のパラメータを推定して逆の歪を適用することにより、自然な条件下で話している話者の口画像に対する微妙な違いの影響を、すぐさまなくすためのコンピュータプログラムを提供することである。

【課題を解決するための手段】

【0010】

この発明の第一の局面は、ユーザのビデオ画像に視覚的音声前処理を行なうシステムにおいて、ユーザの口画像に対する任意変形に関するパラメータを推定する方法に関する。この方法は、ビデオ画像の各々に対し、ユーザの画像中の目印となる位置2箇所を決定するステップと、ビデオ画像中のユーザの口の輪郭を決定するステップと、目印の位置と口の輪郭の位置とをあわせて利用することにより、任意変形と逆の変形がビデオ画像に対し適用可能となるように、ユーザの口画像に関する変形パラメータを推定するステップとを含む。

【0011】

好ましくは、推定するステップは、画像中の目印の位置を利用して、ユーザの画像の回転量を推定するステップを含む。

【0012】

より好ましくは、推定するステップは、ユーザの口画像の、主成分分析に基づく回転量の推定を利用して、ユーザの画像の回転量を推定するステップと、目印の位置を利用して推定した画像の回転量と、主成分分析を利用して推定された画像の回転量とを組み合わせることにより、ユーザの画像の回転角を推定するステップとを含む。

【0013】

さらに好ましくは、推定するステップが、ユーザの口画像の、主成分分析に基づく回転量の推定を利用して、ユーザの画像の回転量を推定するステップを含む。

【0014】

主成分分析に基づく回転量の推定を利用してユーザ画像の回転量を推定するステップは、画像中の口唇の外側輪郭を決定するステップと、唇の外側輪郭のサンプル点の平均と共分散行列とを計算するステップと、共分散行列の第一の固有ベクトル e_1 を求めるステップと、角度 $\alpha = \arctan(e_y / e_x)$ を計算するステップとを含んでもよく、ここで e_x および e_y は第一の固有ベクトルの x 軸成分および y 軸成分である。

【0015】

好ましくは、推定するステップが、画像中の目印間の距離を、画像中の目印の位置と推定された変形とを利用して計算するステップと、口区域を含みさらに目印間の計算された

10

20

30

40

50

距離と同じ幅となるような距離で選択された幅を有する、口領域を選択するステップと、所定の定数の、計算された目印間の距離に対する比率によって、口画像のスケーリング係数を推定するステップと、をさらに含む。

【0016】

より好ましくは、方法は、口画像のスケーリング係数によって、選択された口領域をスケーリングするステップをさらに含む。

【0017】

さらに好ましくは、目印となる位置2箇所を決定するステップは、ユーザの画像中の瞳の位置を決定するステップを含む。

【0018】

この方法はさらに、口領域を所定の場所にシフトすることにより、口に対する平行移動の影響を除去するステップを含んでもよい。

【0019】

この発明の第2の局面に従えば、視覚的前処理に用いるプログラム製品は、プロセッサを用いたシステムに、上述の方法のいずれかに記載のステップのすべてを行なわせる命令を記憶したコンピュータ読取可能媒体を含む。

【発明を実施するための最良の形態】

【0020】

< 第一の実施例 >

視覚的前処理を行なう動機は、視覚的音声処理のための一貫した視覚的特徴量の抽出を容易にすることである。すなわち、この前処理によって、話者の頭の回転、平行移動及びカメラスケールのパラメータを口画像から除去する。図1はこの発明の視覚的前処理アルゴリズムの概要を示すブロック図である。

【0021】

[一般的構造]

図1を参照して、この実施例の口画像修正装置30は、画像フレーム32の各々から口画像を抽出し、話者の頭部の回転、平行移動、及びカメラスケールパラメータによって引起こされる歪のパラメータを推定し、微妙な違いを修正した口画像34を出力する。

【0022】

口画像修正装置30は、画像フレーム32中の口画像をトラッキングする口トラッキングモジュール40と、画像フレーム32の各々において、話者の口エッジを検出する口輪郭線検出モジュール42と、検出された口画像へのPCA (Principal Component Analysis: 主成分分析) に基づき、口画像の回転量を推定するPCAによる回転量推定モジュール44と、各画像フレーム32において目の画像をトラッキングし、各画像フレーム32について話者の瞳画像の座標を出力する両眼トラッキングモジュール46と、両眼トラッキングモジュール46から与えられる両眼の位置に基づいて、顔の回転量を推定する両眼位置による回転量推定モジュール47と、モジュール44及び47で推定された回転量を組み合わせるための回転量推定組合せモジュール48とを含む。

【0023】

口画像修正装置30はさらに、瞳間の距離を修正して回転による歪を除去するための両眼距離修正モジュール49と、回転量推定組合せモジュール48によって推定された回転量に基づき、話者の頭部の回転によって引起される口画像の歪を修正するための、口領域回転量修正モジュール50と、口画像を含む顔画像の予め定められた領域を選択するための口ROI (Region of Interest: 関心領域) 選択モジュール52と、口画像を予め定められた基準サイズ (N×M画素) に正規化するためのスケール修正モジュール54と、正規化された口画像のサイズ (N及びM) を記憶するためのメモリ56と、ROI中の口画像の平行移動の影響をなくし、微妙な違いを修正した口画像34を出力するための平行移動修正モジュール58とを含む。

【0024】

10

20

30

40

50

[A . 修正ステップ]

一般に、観察された口画像 x と基準となる口画像 x° との関係は、以下のように表せる。

【 0 0 2 5 】

$$x = s R(\alpha) x^{\circ} + t \quad (1)$$

ここで、 s はスケーリングパラメータであり、 $R(\alpha)$ は 2 D 回転行列 (α は回転角) であり、 t は平行移動ベクトルである。画像フレームは全て、任意の s 、 $R(\alpha)$ 及び t の影響を除去するために以下のように前処理される。

【 0 0 2 6 】

1 . モジュール 4 4 及び 4 7 で推定された PCA の結果と標準座標系に対する目の角度とを組み合わせることにより、モジュール 4 8 で $R(\alpha)$ の推定と修正とを行なう。

10

【 0 0 2 7 】

2 . 両眼の間の距離と上唇の輪郭とを基準として用いて、モジュール 5 0 及び 5 2 で口境界を選択する。

【 0 0 2 8 】

3 . 両眼の間の距離に基づいて、スケール修正モジュール 5 4 でスケール s を修正する。

【 0 0 2 9 】

4 . 固定された座標位置へ口の中心を移動することにより、平行移動修正モジュール 5 8 において平行移動の影響を除去する。

20

【 0 0 3 0 】

[B . 口境界の選択]

口境界は、口トラッキングアルゴリズムと目のトラッキングアルゴリズムとを合わせて利用することによって選択される。この実施例で用いるアルゴリズムは、本件発明者らによって開発されたものであるが、ここでは入手可能などのようなアルゴリズムを用いてもよい。

【 0 0 3 1 】

口境界の選択は、特にスケール修正のため、この微妙な違いを修正する処理において重要な手順である。発声の間に口の形状はしばしば変形するので、セグメント化した唇の外輪郭又は口角にもとづいた口境界の選択は不正確なスケール修正につながる。このため、顔上で (少なくとも 2 点の) 変形しない目印を測定基準として選択し、セグメント化された口領域が存在する口境界を選択する必要がある。

30

【 0 0 3 2 】

発明者らは、両眼が好ましい目印となることを見出し、モジュール 4 7 及び 4 9 でそれぞれ推定し修正した瞳間の距離を、口境界選択時の基準距離として利用した。すなわち、口境界の幅は、話者の瞳間の距離と距離方向に同じ幅となるように選択され、セグメント化された口がその中の水平方向中心に位置する。同様に、高さも同じ距離で設定されるが、口の上方境界は上唇の外輪郭の先端で始まる。なぜなら、上顎は目の位置に対し移動しないからである。こうして、 $L_t \times L_t$ の口画像が得られる。ここで、 L_t は時間 t における、時間依存の瞳間の画素距離である。

40

【 0 0 3 3 】

この実施例のモジュール 4 0 で用いられる口トラッキングアルゴリズムはベイズのフレームワークに基づくものである (非特許文献 1 を参照)。ここでは予めラベリングする必要のないカラー画像を利用する。これは現在のフレーム中の口領域をセグメント化し、対象となる領域を選択することによって、次のフレームでの検索空間を限定する。両眼トラッキングモジュール 4 6 で用いられる目トラッキングアルゴリズムは両眼の中点を検出することに基づく。ここで用いられるアルゴリズムの詳細は非特許文献 7 に見られる。これらのアルゴリズムはともに、通常のパブリックドメインのオペレーティングシステムで 1 . 9 GHz で動作する最新のコンピュータ上でリアルタイムに動作する。

【 0 0 3 4 】

50

〔 C . 微妙な違いのパラメータ除去 〕

一方で、口の外觀の歪は一般に任意の 3 D 回転によって引起される。他方で、人とコンピュータのインタフェイスの応用を考えれば、話者にはカメラに向き合うことが要求される。話者はカメラに向き合うが、頭部を自由に動かすことはできる。

【 0 0 3 5 】

カメラの光軸に平行な話者の頭部の動きは、画像のスケーリングを引起す。カメラの光軸に垂直な話者の頭部の動きは、平行移動を引起す（この場合投影的な歪は避ける）。光軸に対して回るような頭部の動きは、口画像の 2 D 回転を引起す。一般に、頭部の動きは上述の動きのいずれか、またはその組合せである。もし話者がカメラの光軸から目をそらせば、それはその時点で話者がコンピュータと意思の疎通をはかる意図がない、という意味である。

10

【 0 0 3 6 】

自然な発声の場合、このような歪の大きさはカメラの光軸に対する話者の頭部の角度に依存する。このような動きは図 2 に見られるような遠近感（奥行き方向）の歪を生じさせる（3 D の頭部回転による）。この実施例で用いるアルゴリズムは、顔がこのような部分的にしか見えない場合、話者がコンピュータと意思の疎通をはかる意図がないものと考え、このような遠近感（奥行き方向）の歪を除くことはしない。

【 0 0 3 7 】

人とコンピュータのインタフェイスの応用設計では、セグメント化された口画像は全てのフレームについて、視覚的特徴量抽出処理に先立って、特に外觀に基づく視覚的特徴量抽出処理に先立っては、前処理を行なって正準変形を施しておく必要がある。ユーザに求められるのは、発声の間、カメラの焦点にほぼ顔を向けていることなので、剪断変形の効果はなしと見ることができ、x 方向及び y 方向のスケーリングを同一とみなすことができる。こうして、セグメント化された口画像データに生じうる微妙な違いは、平行移動、スケーリングまたは回転のみか、これらを組合わせたものとなる。

20

【 0 0 3 8 】

上述の仮定に基づき、観察された口画像データ x と基準口画像データ x° との関係は次のように表される。

【 0 0 3 9 】

$$x = s R(\alpha) x^\circ + t \quad (2)$$

30

ここで s はスケーリングパラメータであり、 $R(\alpha)$ は α の 2 D 回転行列であり、 α は 2 D 回転角であり、 t は平行移動ベクトルである。

【 0 0 4 0 】

固定された座標位置に口画像の中心を移動することにより、セグメント化された口画像から平行移動ベクトル t を容易になくすることができる。以下のサブセクションでは、回転角 α とスケーリングパラメータ s とをそれぞれどのように推定し除去するかを説明する。

【 0 0 4 1 】

〔 D . 唇の回転の問題 〕

口角または鼻や目等の他の顔の特徴のみを利用する場合、自然な発声の間にそのような顔の目印の位置を定めることに画素上の正確さが必要なため、回転を正確に修正するには問題がある〔非特許文献 4、7 及び 10 を参照〕。目印を使用する口回転修正では、トラッキングした瞳を用いることが合理的であって、PCA ベースの回転量推定及び修正方法と組合わせて上述の問題を克服できることがわかっている。

40

【 0 0 4 2 】

D . 1 . PCA を用いた回転量修正

PCA は、多変量データを分析して主成分として知られる新たな直交軸の組を特定する方法である〔画像処理における PCA の詳細については非特許文献 5 を参照〕。これは PCA による回転量推定モジュール 44 で行なわれる。第一の主成分は、データの最も主要な分散を示す軸であり、第二の主成分はデータの 2 番目に主要な分散を示す直交軸であり、以下同様である。

50

【 0 0 4 3 】

図3に示すように、有向エッジ検出フィルタを適用して、唇の外輪郭を抽出した。輪郭70上のi番目の点の座標を x_i とし、 μ と C とをそれぞれN個のサンプル点に対する平均と共分散とする。

【 0 0 4 4 】

【数1】

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (3)$$

$$C = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T \quad (4)$$

10

ここで、 T は転置を表す。ここでの課題は、新たな直交軸(u -及び v -軸)の組を見出し、標準座標系(x -及び y -軸)に対する回転角 α を推定し、口画素座標データの回転をなくすことである。

【 0 0 4 5 】

図4(A)を参照して、 x -軸80と u -軸82との間の回転角 α 84を推定するために、唇の外輪郭端の共分散行列 C の固有値 $\{\lambda_1, \lambda_2\}$ を求め、最大固有値に対応する固有ベクトル e_1 を見出す。この処理は以下のとおりである。

【 0 0 4 6 】

$$\det(C - \lambda_1) = 0 \quad (5)$$

20

次に、固有ベクトル(特性ベクトルともいう。)を次の式に従って求める。

【 0 0 4 7 】

$$C e_i = \lambda_i e_i, \quad i = 1, 2 \quad (6)$$

回転角 α は以下のように計算される。

【 0 0 4 8 】

$$\alpha = \arctan(e_y / e_x) \quad (7)$$

ここで e_x 及び e_y は e_1 の2成分である。

【 0 0 4 9 】

次に、口領域回転量修正モジュール50において、口画像データ x に $R^{-1} = R(-\alpha)$ を乗ずることにより、回転修正された口画像が得られる。図4(B)は回転修正後の口画像を示す。

30

【 0 0 5 0 】

D.2. スケーリングの問題

スケーリングの問題は、話者のカメラからの距離、カメラのズーム係数、及び話者の実際の口の寸法から生じる。この場合、唇画像の周波数内容を利用するDCT(Discrete Cosine Transformation: 離散的コサイン変換)またはウェーブレット変換法等の、画素に基づく視覚的特徴量抽出法は、上述の要因の1つまたは2つ以上によって引起されるスケールパラメータを除去することなく口画像を処理した場合、一貫性を欠く観察ベクトルを生じさせることになる。

40

【 0 0 5 1 】

図5に示すような回転した画像でスケーリングファクタを除去するためには、まず、セクションCで計算した R^{-1} を用いてモジュール50で回転を修正する。次に、両眼距離 L_t (100)を測定し、図5(B)に示すようにモジュール49で修正する。この画像を M/L_t でスケーリングすることにより図5(C)に示すように瞳間の距離110は M となる。ここで M は予め規定した正規化のための画素長である。 $N \times M$ サイズの正規化された口画像112がこのスケールの正規化された画像から切取られる。ここで、上方向の境界は上唇の輪郭(図5(C))である。これを、微妙な違いを修正した口画像34と呼ぶ。

【 0 0 5 2 】

50

〔コンピュータによる実現〕

この実施例に従う口画像修正装置 30 は、コンピュータハードウェアとその上で実行されるコンピュータプログラムで実現され得る。例えば、口トラッキングモジュール 40 は発明者らが開発したプログラムで実現できる。しかしながら、これは公に入手可能な目トラッキングプログラムでも実現可能である。同様に、モジュール 42 から 58 もコンピュータプログラムで実現可能である。

【0053】

図 6 は上述の適当なコンピュータプログラムとともに、この実施例の口画像修正装置 30 を実現するコンピュータシステム 330 の斜視図である。図 6 を参照して、このコンピュータシステム 330 は、CD-ROM (コンパクトディスク読出専用メモリ) ドライブ 350 及び FD (フレキシブルディスク) ドライブ 352 を有するコンピュータ 340 と、いずれもコンピュータ 340 に接続されたキーボード 346 と、マウス 348 と、モニター 342 とを含む。システム 330 はさらに、コンピュータ 340 に接続されて話者の頭部画像のフレームをキャプチャするビデオカメラ 374 を含む。システム 330 はさらに、図示しないマイクロフォンを含む。

10

【0054】

図 7 はコンピュータ 340 のハードウェアブロック図である。図 7 を参照して、CD-ROM ドライブ 350 及び FD ドライブ 352 に加えて、コンピュータ 340 は CPU (中央処理装置) 356 と、CPU 356 に接続されたバス 366 と、バス 366 に接続された読出専用メモリ (ROM) 358 と、バス 366 に接続されたランダムアクセスメモリ (RAM) 360 と、バス 366 に接続されたハードディスクドライブ 354 と、バス 366 に接続され、ビデオカメラ 374 の出力の各フレームをキャプチャするためのビデオキャプチャボード 368 とを含む。

20

【0055】

CD-ROM ドライブ 350 は挿入される CD-ROM 362 からデータを読み出し、FD ドライブ 352 は FD 364 からデータを読み出し、ここにデータを書込む。

【0056】

上述の口画像修正装置 30 を実現するプログラムは CD-ROM 362 または FD 364 に記憶され、さらにハードディスク 354 に転送される。プログラムが CPU 356 によって実行される時に、プログラムはハードディスクドライブ 354 から読出されて RAM 360 にロードされる。CPU は、RAM 360 のプログラムカウンタ (図示せず) という名前のレジスタによって指定されるアドレスから命令をフェッチし、命令中のオペランドによって指定される RAM 360 またはハードディスクドライブ 354 等の記憶装置からデータをフェッチし、結果を命令オペランドによって指定されるアドレスに書込む。

30

【0057】

コンピュータ 340 の動作は周知であるので、ここでは詳細は説明しない。

【0058】

〔動作〕

口画像修正装置 30 は以下のように動作する。カラーの画像フレーム 32 の各々について、口トラッキングモジュール 40 が口領域をセグメント化し、ROI を選択して次のフレームの探索領域を制限する。口輪郭線検出モジュール 42 は図 3 に示すような口の輪郭 70 を検出する。PCA による回転量推定モジュール 44 は口輪郭線検出モジュール 42 から出力される輪郭 70 で、PCA を利用した画像で予め規定された x - 及び y - 軸を基準に、回転角を推定する。推定された角度が回転量推定組合せモジュール 48 に与えられる。

40

【0059】

両眼トラッキングモジュール 46 は、各画像フレーム 32 の両眼の画像をトラッキングして、x - 座標及び y - 座標を基準とする話者の瞳画像の座標を、両眼位置による回転量推定モジュール 47 と両眼距離修正モジュール 49 とに出力する。

【0060】

50

回転量推定組合せモジュール 48 は PCA による回転量推定モジュール 44 からの角度と両眼位置による回転量推定モジュール 47 からの角度とを組合わせ、両眼距離修正モジュール 49 と口領域回転量修正モジュール 50 とに推定された角度 α を出力する。

【0061】

両眼距離修正モジュール 49 は話者の瞳間の距離 L_t を修正し、この距離 L_t を ROI 選択モジュール 52 とスケール修正モジュール 54 とに与える。

【0062】

口領域回転量修正モジュール 50 は $R^{-1}(\alpha)$ を利用し、回転量推定組合せモジュール 48 からの α を利用して、口領域周囲の話者の顔画像を修正する。口領域回転量修正モジュール 50 は修正された口領域の画像を ROI 選択モジュール 52 に出力する。

10

【0063】

ROI 選択モジュール 52 は話者の瞳間の距離と同じ幅の距離となるように選択され、セグメント化された口が水平方向の中央に位置するような ROI を選択する。同様に、ROI の高さも同じ距離に設定される。ROI の上方境界は上唇の外輪郭の先端で始まる。従って、 $L_t \times L_t$ の口画像が得られる。ここで、 L_t は時間 t における、時間に依存する瞳間の画素距離である。

【0064】

スケール修正モジュール 54 は ROI の画像を M/L_t でスケーリングし、この結果、瞳間の距離は図 5 (C) に示すように $M/110$ となる。ここで M は予め規定した正規化のための画素長である。 $N \times M$ サイズの正規化された口画像がこのスケールの正規化された画像からスケール修正モジュール 54 によって切取られる。

20

【0065】

最後に、画像の口の中心を画像の幅方向中央の固定された座標点に移動することにより、平行移動の歪を除去する。

【0066】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。

30

【図面の簡単な説明】

【0067】

【図 1】この発明の一実施例に従った口画像修正装置 30 を示すブロック図である。

【図 2】話者の頭部の回転により引起される口画像の奥行き方向の歪を示す図である。

【図 3】口の輪郭 70 を示す図である。

【図 4】口画像の回転角 α の PCA による推定を示す図である。

【図 5】この発明の一実施例による口画像のスケール修正を示す図である。

【図 6】この発明の第一の実施例を実現するコンピュータシステム 330 の斜視図である。

【図 7】図 6 に示すコンピュータ 340 のブロック図である。

40

【符号の説明】

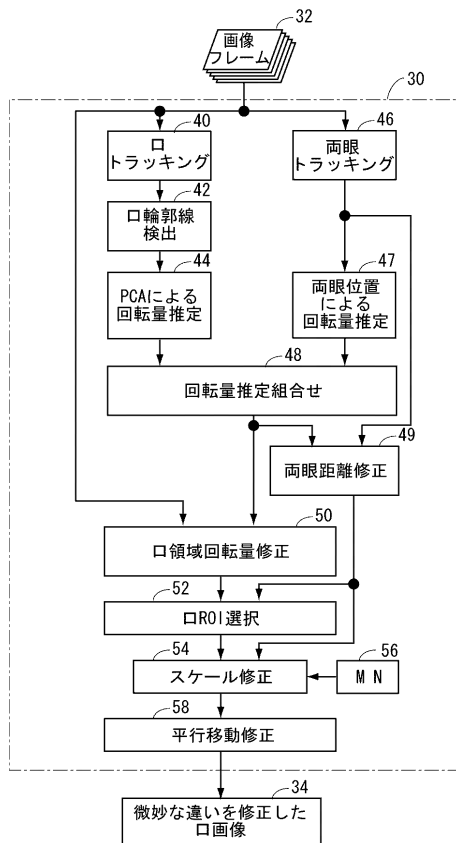
【0068】

- 30 口画像修正装置
- 32 画像フレーム
- 34 微妙な違いを修正した口画像
- 40 ロトラッキングモジュール
- 42 口輪郭線検出モジュール
- 44 PCA による回転量推定モジュール
- 46 両眼トラッキングモジュール
- 48 回転量推定組合せモジュール

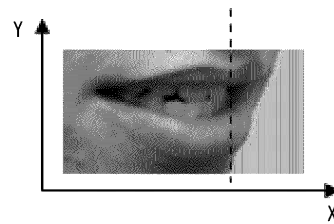
50

- 5 0 口領域回転量修正モジュール
- 5 2 口ROI選択モジュール
- 5 4 スケール修正モジュール
- 5 6 メモリ
- 5 8 平行移動修正モジュール

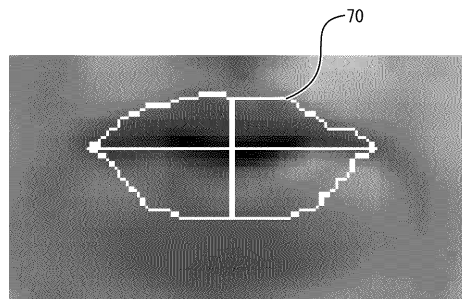
【図 1】



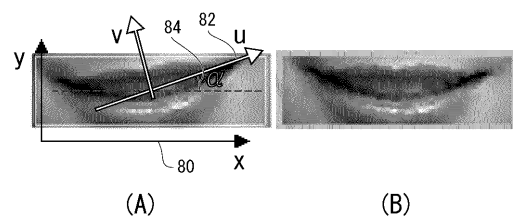
【図 2】



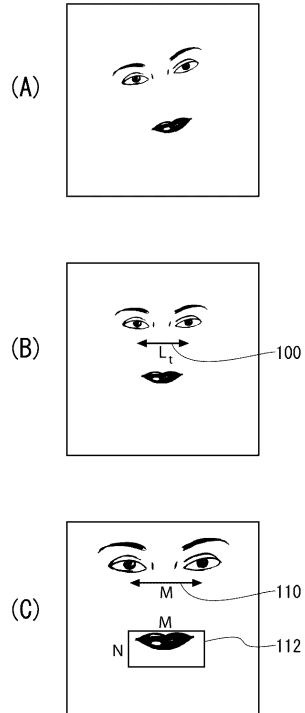
【図 3】



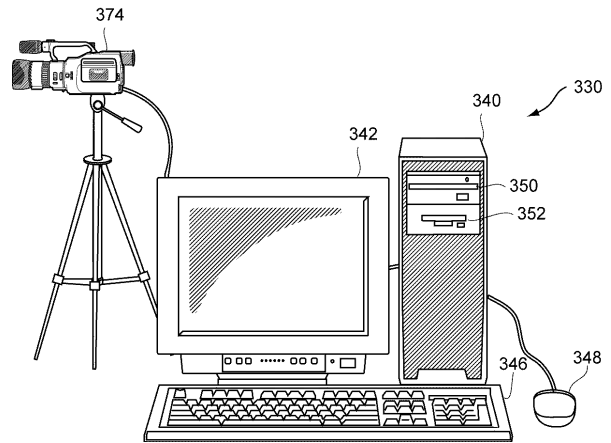
【図 4】



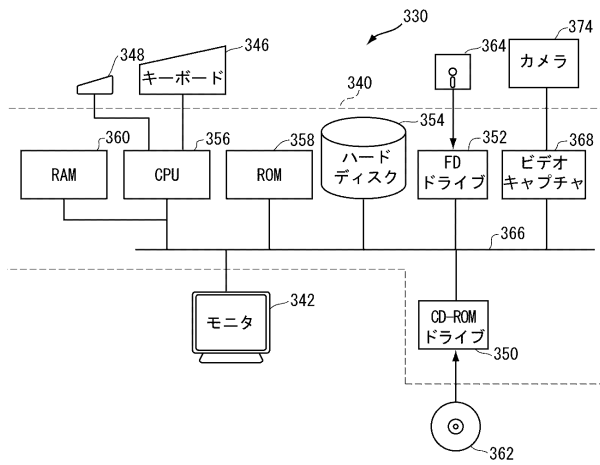
【図 5】



【図 6】



【図 7】



フロントページの続き

(72)発明者 矢野 澄男

京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

審査官 秦野 孝一郎

(56)参考文献 特開2003-281539(JP, A)

特開平09-171560(JP, A)

(58)調査した分野(Int.Cl., DB名)

G06T 1/00