

単眼距離カメラを用いた顔向き の 推定

内田奈穂子^{†1} 胡振程^{†1} 董延超^{†2}

交通事故の主要な原因はドライバによる漫然運転である。近年この問題を解決するために、連続画像を用いたドライバの顔向き の 推定に関する研究が行われている。本稿では、単眼距離カメラの赤外線エミッタを切り替えることにより同時に取得する赤外線画像と距離画像を用いた顔特徴点検出法および顔向き推定法について述べる。赤外線画像は顔特徴点検出と追跡に利用し、距離画像は顔領域検出および顔向き推定の際に用いる。ドライビングシミュレータ環境で提案手法の有効性を検証した。また、提案手法は、ドライバモニタリングシステムのみならず、セキュリティやアバターシステムなど他のインターフェースシステムに応用可能であると考えられる。

Face Orientation Detection with a Single Depth Camera

NAOKO UCHIDA^{†1} ZHENCHENG HU^{†1} YANCHAO DONG^{†2}

Careless driving is a major factor in most traffic accidents. In last decades, research on estimation of face orientation and tracking of facial features through a consequence images have shown very promising results of determining the level of driver's concentration. In this research, we propose a novel approach of facial features and face orientation detection by using a single uncalibrated depth camera by switching IR depth pattern emitter. With this simple setup, we are able to obtain both depth image and infrared image in a continuously alternating grab mode. Infrared images are employed for facial features detection and tracking while depth information is used for face region detection and face orientation estimation. This system is not utilized only for driver monitoring system, but also other human interface system such as security, avatar systems and so on.

1. はじめに

高速道路における交通事故の主な原因に、ドライバの漫然運転が挙げられる。高速道路安全局は、報告されている事故のうち約25%が、居眠り、よそ見などドライバの不注意が原因であると推定している。

近年、連続画像を用いた顔向き の 検出や顔特徴点の追跡に関する研究が盛んに行われている。特に、距離情報を用いた顔向き検出法が数多く提案されている[1][2]。従来、距離情報を取得するために、TOF(Time of Flight)カメラやステレオカメラがよく利用されている。TOFカメラは高精度な距離情報を取得することが可能であるが、高価であるため実応用には向かない。ステレオカメラは単眼カメラよりも設置にスペースを必要とし、高精度かつ密な奥行き情報を推定するための計算コストが高い。また、二次元画像から距離情報を推定する手法もあるが[3]、被写体の大きさと距離のCoupling問題で曖昧になる場合がある。

上記の問題を解決するために、本稿では単眼距離カメラの赤外線エミッタを切り替えることにより同時に取得する赤外線画像と距離画像を用いた顔特徴点検出法および顔向き推定法を提案する。赤外線画像は顔特徴点検出と追跡に利用され、距離画像は顔領域検出と顔向き検出に用いられる。

入力装置として、安価かつ高速な処理が可能である

Microsoft社のKinectを利用する[4]。可視光画像は照明変化に影響を受けやすいため、提案手法では利用せず、その代わりに赤外線画像を使う。赤外線画像は暗い車内でも利用することができる点で優れており、自動車メーカー等でも赤外線画像を利用した顔特徴点検出技術が開発されている[5][6]。また、赤外線画像と距離画像は単一のカメラで取得することができるため、カメラ間のキャリブレーションが不要であることも利点である。

2. 単眼距離カメラによる顔向き検出システム構成

本研究で使用されるKinectはMicrosoft社から販売されているRGB-Dセンサであり、距離データを得られる安価なセンサとして広く用いられている[7][8]。Kinectには可視光画像、距離画像、音声をそれぞれ取得する3種類のセンサが搭載されており、これらのデータを同時に取得することができる。図1にKinect本体を示す。図1のように、本体の中央部に赤外線光プロジェクタ、RGBカメラ、距離カメラが並んでいる。赤外線光プロジェクタから投影されたランダムドットパターンを距離カメラによって撮影し、ドットパターンの変化量を用いて距離画像が計算される。したがって、距離カメラは赤外線カメラと言い換えることができる。

提案手法では、Kinectの距離カメラを用いて、赤外線画像と距離画像という二種類の組を入力画像として取得する。提案手法で前提としている車内の環境光は、日中と夜間では大きく変化する。赤外線画像は、可視光画像と比較して環境光の変化に影響されにくいいため、提案手法での利用に適している。本研究では、一定の照明環境下における赤外

^{†1} 熊本大学大学院自然科学研究科
Kumamoto University, The Graduate School of Science and Technology
^{†2} 同済大学
Tongji University

線画像を取得するために、追加のリング型赤外照明装置を利用する。

図2に示すように、Kinectから取得される赤外線画像には、距離情報を計算するためのランダムドットパターンが写りこむ。このドットパターンは顔特徴点を検出する際のノイズとなるため、赤外線画像を撮影する場合は赤外線エミッタをオフにする。すなわち、距離画像を撮影するときはエミッタをオン、赤外線画像を撮影するときはエミッタをオフにする。一台のKinectで距離画像と赤外線画像を取得するためには、エミッタのオンとオフを毎フレームで切り替えながら撮影する必要がある。エミッタのオンとオフが確実に画像に反映されるのを待つために、エミッタの出力を切り替えるごとに一定の待機時間を設ける。本研究における実行環境は、Windows7 Intel® Core™ i5-3230M CPU2.60GHz メモリ4.00GBであり待機時間は80msとする。この場合のフレームレートは約2.5fpsである。

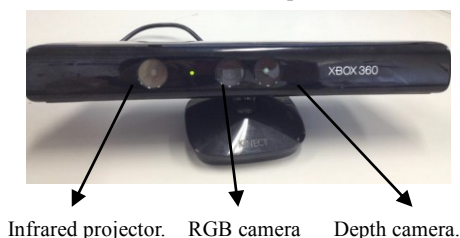


図1 Kinectセンサ



図2 赤外線エミッタのドットパターン

3. 単眼距離カメラによる顔向き推定手法

本稿では、単眼距離カメラの赤外線エミッタを切り替えることにより同時に取得する赤外線画像と距離画像を用いた顔特徴点検出法および顔向き推定法について述べる。赤外線画像は顔特徴点検出と追跡に利用し、距離画像は顔領域検出および顔向き推定の際に用いる。また、本稿では2次元画像の顔特徴点を追跡するため、顔の3次元運動より2次元特徴点の位置を予測し、従来では問題となる顔回転の際に特徴点の追跡ができなくなることを解決し、マッチング率を大幅に向上することができ、より正確な顔向きを推定できる。

3.1 提案手法の概要

図3に提案手法の概要を示す。2章の手法により取得する赤外線画像と距離画像の組を入力する。はじめに、入力した距離画像を用いて顔領域を検出する。顔領域を検出し

た後、顔特徴点の検出または追跡を行う。顔特徴点を検出する場合、まず目、口といった顔器官の領域を Haar-like 分類器で検出する[9][10]。顔特徴点は各領域内で検出され、提案手法では両目の内側および外側のコーナー点と口の左右のコーナー点を抽出する。各コーナー点の検出には、主にエッジ検出、二値化、コーナー検出、テンプレートマッチングを利用する。入力画像が2枚目以降の場合は、ピラミッドLK法により特徴点の追跡を行う。前後フレーム間で特徴点を追跡することで、毎フレームで特徴点を検出するよりも安定した特徴点を取得することができ、かつ処理時間を短縮することができる。特徴点検出または追跡の後、顔向きの推定を行う。顔向きは、各特徴点が同一平面上にあると仮定し、この平面の法線ベクトルと一致するものとして推定される。また、追跡失敗かつ連続追跡回数が一定以上の場合は、現在フレームを用いて特徴点を再検出する。

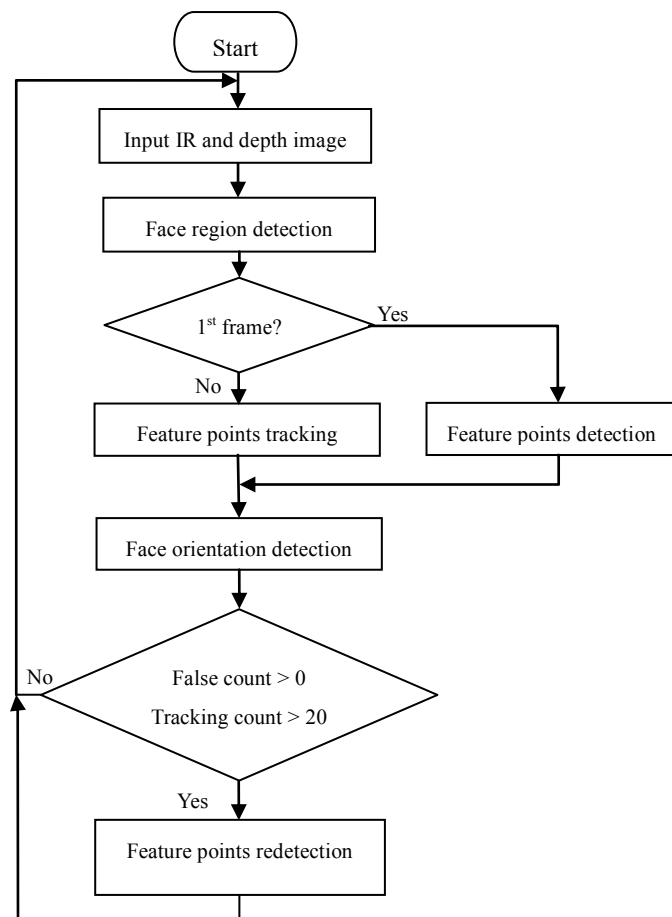


図3 システム概要図

3.2 顔領域検出

人物の顔領域は、距離情報のヒストグラムに基づいて検出される。Kinect センサは車内のメーターパネルの前に設置しており、図4に示すような画像が取得される。図4より、ドライバはセンサから最も近い位置にいると仮定できる。このことから、センサから1メートル以上距離が離れ

ているピクセルは背景画素として除去する。

背景を除去した後、垂直方向および水平方向の距離ヒストグラムを用いて顔領域を切り取る。画像に対して垂直方向に画素数を累積したヒストグラムを図5に示す。図5のように、人物の頭のある領域は、他の領域よりも高い値になる。このことを利用して、ヒストグラム中でしきい値以下の値を持つ領域の画素は除去し、顔領域の水平方向座標を得る。ただし、しきい値は繰り返し処理により画像ごとに決定する。

同様に、水平方向の距離ヒストグラムも作成し、顔領域の垂直方向座標を得る。ただし、垂直方向座標の決定には、ドライバの首や体の領域を除去する必要があるため、ヒストグラムに加えて領域の平均距離を利用する。ドライバが運転席に座っている状態を正面のカメラから撮影する場合、顔領域は首領域と比較して近くに位置するため、首領域距離情報は大きくなる。したがって、平均以上の距離情報を持つ画素を首の領域とみなし、除去する。次に、残った体の領域を除去する。本稿で提案する画像取得環境において、体領域は顔領域と比較して画面に写る範囲が狭い。したがって、画像上の最大連結成分を顔領域とみなし、それ以外の画素は体領域として除去する。

上述のプロセスで決定した顔領域には、髪の毛の影、鼻孔、眼鏡の反射などが原因で距離情報が欠損している部分がある。この欠損部分はバイリニア法により補間され、その後ガウシアンフィルタで平滑化を行う。図6に欠損補間と平滑化を行った顔領域を示す。

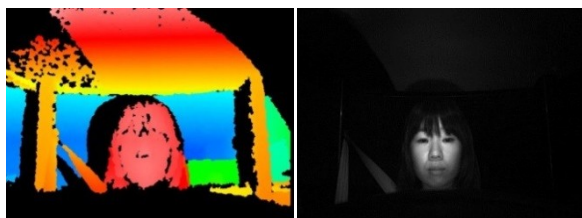


図4. (a) 赤外線画像, (b) 距離画像

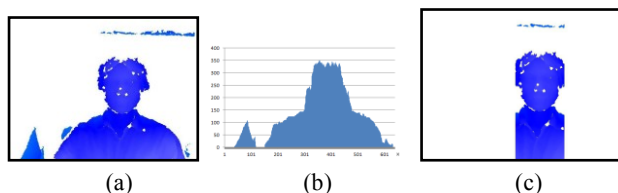


図5. (a) 背景除去後の距離画像 (b) (a)の垂直方向ヒストグラム. (c) x座標決定

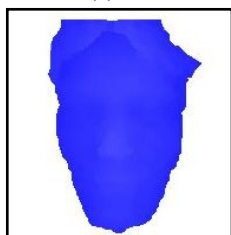


図6. 欠損補間, 平滑化後の顔領域

3.3 顔特徴点検出および追跡

検出した顔領域内で目領域および口領域を検出する。各領域の検出には Haar-like 特徴を用いた分類器を利用する。個々で用いる分類器はそれぞれ目領域用と口領域用にトレーニングされている。眉領域は、検出した目領域の上に位置する領域として定義する。顔特徴点は Y. Dong ら[11]によって提案された手法を利用する。Y. Dong らは入力として可視光画像を用いているが、本研究では赤外線画像を入力とする。

各顔器官の領域を検出した後に顔特徴点を検出する。顔特徴点として、両目の内側および外側のコーナー、眉頭のコーナー、および口の両口角を求める。以下にそれぞれの特徴点の検出について述べる。

3.3.1 両目コーナー点

両目の内側及び外側のコーナーは以下の手順で検出される。あらかじめ検出した目領域をガウシアンフィルタによって平滑化する。平滑化した目領域から、Prewitt フィルタによって上まぶたのエッジを抽出する。ここで Prewitt フィルタの方向は、上、右上、および左上方向である。その後、抽出されたエッジを二値化し、最も長いエッジの両端点を求める。その後、コーナー検出法[12]を用いて目のコーナーの候補を検出し、最終的にテンプレートマッチングによってコーナー点を決定する。図7はテンプレートマッチングに利用する両目のコーナーのテンプレートである。



(a) 右目 (b) 左目

図7.両目の内側および外側のテンプレート

3.3.2 口コーナー点

口のコーナー点として、左右の口角を検出する。口コーナー点は目のコーナー点とほぼ同様の手順で検出される。口領域を平滑化したあと、口のエッジを検出して二値化する。目のコーナー検出と異なる点は、口のコーナー検出は膨張収縮処理を行う点である。膨張収縮処理は口のエッジの欠損を補うために適用される。

3.3.3 特徴点追跡

3.3.1 および 3.3.2 で記述した両目および口のコーナー点検出は、処理の開始直後である1枚目のフレームで行われる。2枚目以降のフレームでは、追跡処理により特徴点位置を決定する。本研究ではピラミッドLK法を用い、ひとつ前のフレームと現在フレームの間で特徴点の追跡を行う[13]。LK法は、特徴点の輝度が移動前と移動後で等しいと仮定した追跡手法である。ピラミッドLK法は、LK法のノイズに対するロバスト性を向上させたアルゴリズムである。

提案手法において、顔特徴点は基本的にピラミッドLK法で追跡されるが、追跡結果にエラーが生じる場合がある。我々は、以下に挙げる3点に当てはまる場合は、信頼性が

低い特徴点とみなす。

- (1) 特徴点座標の距離情報が得られない場合。
- (2) 特徴点周辺の輝度値がしきい値以下の場合。
- (3) 追跡後の座標の標準偏差値が高い場合。

(1)について、3.2 節で述べた顔領域検出の際に背景を除去しているため、距離情報が得られない座標は背景である場合が多い。(2)については、特に特徴点が輪郭周辺にあるときは、毛髪による遮蔽や照明の光量不足などにより輝度値が低下し、追跡精度が低下する。また、顔の回転による特徴点の消失も考えられる。(3)については、提案法では運転中の頭の位置は極端に大きく変化しないものと仮定する。したがって前後フレーム間における特徴点位置も大きく変化しないため、特徴点が正しく追跡されたときは標準偏差値が低くなる。

追跡の結果、信頼性が低いとみなされた特徴点は、一定の条件を満たす場合にのみ再検出される。特徴点の再検出は3.3.1, 3.3.2 で記述した手法で行われる。以下に特徴点を再検出する際の条件を示す。下記(1)および(2)の条件をどちらも満たす場合、特徴点を再検出する。

- (1) 信頼性の低い特徴点が、現在フレームまでの累計で1つ以上ある場合。

- (2) 連続20フレーム以上で追跡している場合。

上記の(2)については、信頼性の低い特徴点が見つかるたびに特徴点の再検出をすると処理時間が長くなる。検証実験の結果、両目と口の計6点のコーナーのうち、最低3点あれば顔向きを推定できることがわかっているため、このような条件を定める。

3.4 顔向き推定

顔向きは顔特徴点の一つの平面上にあると仮定して推定される。本研究では両目の内側および外側のコーナー点、口の両コーナー点在同一平面上にあると仮定する。以下に顔向きの計算方法について述べる。

各特徴点座標とその距離情報から、3次元座標 (x, y, z) が得られる。この座標はカメラの位置を原点としたカメラ座標系である。カメラ座標は式(1)を用いて射影変換することで、2次元の平面座標に変換できる。

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \frac{x-c_x}{f_x} \\ \frac{y-c_y}{f_y} \\ 1 \end{pmatrix} z \quad \dots (1)$$

ただし (c_x, c_y) は中心座標、 (f_x, f_y) は焦点距離である。この平面の傾きは式(2)で求めることができる。

$$\begin{pmatrix} (X_1, Y_1, Z_1) \\ (X_2, Y_2, Z_2) \\ (X_3, Y_3, Z_3) \\ \vdots \\ (X_n, Y_n, Z_n) \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \dots (2)$$

式(2)において n は特徴点数である。式(2)の (a, b, c) が顔向きのパラメータとなる。

4. 検証実験

本章では提案手法の精度実験について述べる。顔向き推定の精度実験は、被験者が頭部に装着した加速度センサを用いて取得する真値との比較を行う。また、顔向きの推定精度には顔特徴点の検出精度が大きく影響するため、顔特徴点検出精度に関する実験についても述べる。

4.1 実験環境

本実験で用いる赤外線画像と距離画像はKinectを用いて撮影する。実験に用いるデータは、運転者がドライビングシミュレータの運転席に座っている場면을撮影したものである。被験者は男性9名、女性4名の計13名である。ドライビングシミュレータは室内に設置しており、窓をカーテンで覆っているため、太陽光による影響はないものとする。図8に示すように、Kinectをハンドルの後方に、円形の赤外照明装置をハンドルの中央とその垂直線上のフレームに設置している。赤外照明装置の波長840nmである。

4.2 使用するデータ

図9に被験者の顔向きを指示するマークを示す。マークはA, B, C, D, E, F, Gの7方向に設定する。被験者はドライビングシミュレータの運転席に座り、まず図9のAのマーク（正面方向）を見る。その状態からBのマークを見るように顔を動かし、その後再び正面方向を向く。この動作を2回繰り返す。その間、データを連続して撮影する。C, D, E, F, Gについても同じ動作をしている場면을撮影する。A方向については、正面方向のみを一定時間向いている場면을撮影する。A~Gの各マークは、サイドミラーやバックミラー、速度計など、運転者がよく見る箇所に設定されている。図10に取得した赤外線画像の例を示す。

顔向きの推定値を評価するために、本実験では3軸角速度センサを用いて頭部角度を取得し、真値とする。被験者は右耳の上方に加速度センサを装着し、Kinectで画像データを撮影すると同時に頭部運動の角速度を取得する。頭部角度は角速度の積分値として計算する。角速度センサのバイアス誤差を除去するために、静止時の平均角速度をバイアスとして減算する。また、角速度センサのドリフトの影響を抑えるため、データの連続取得時間は40秒以下とする。

予備実験として、上記の条件で角速度センサの精度実験を行う。センサをターンテーブル上に設置し、ターンテーブルを30度まで回転させたときの角度を求め、実測値と比較する。実験の結果、計測値は平均29.55度となり、誤差率は1.5%であった。



図 8. Kinect および赤外照明器



図 9. ドライビングシミュレータ上のマーク

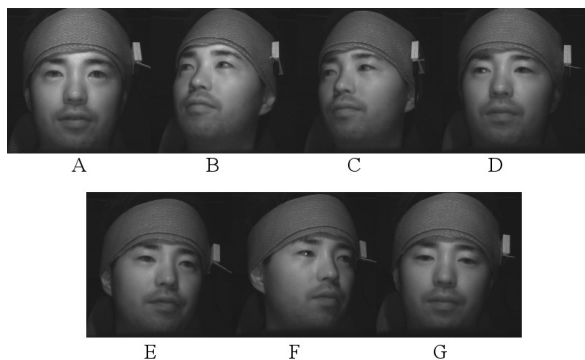


図 10. 各方向の赤外線画像

4.3 実験結果

4.3.1 顔特徴点検出実験

顔特徴点検出実験では、正面向きの顔と回転した顔における検出率を比較する。正面向きの顔として A 方向のデータを用い、回転した顔として C 方向のデータを利用する。表 1 に顔特徴点検出実験の結果を示す。表 1 において、顔領域および目、口領域は前ステップで正しく検出済みであるとする。検出した特徴点位置が真値から上下左右に各 3 ピクセル以上離れている場合、検出失敗とした。ただし真値は実験者が画像上で位置を確認し、定めたものとする。

表 1 より、正面向きである方向 A の場合は全ての特徴点を 95%以上の精度で検出することが可能である。一方、正面向きの顔と比較して、回転した顔である方向 C の場合は検出精度が低下している。特に、C 方向は左方向への回転であるため、回転によって消失する左目尻、左目頭の検出精度が低下している。

表 1 顔特徴点検出精度

	左目 尻	左目 頭	右目 頭	右目 尻	左口 角	右口 角
A[%]	99.18	99.12	97.13	96.51	97.74	98.77
C[%]	76.40	80.00	94.00	94.80	94.40	98.80

4.3.2 顔向き推定実験

次に顔向き推定の実験結果について述べる。真値との比較のため、真値および推定値の初期値を 0 度とする。また、4.3.1 の結果より、顔が回転している場合に特徴点検出精度が低下するため、本実験では特徴点の再検出を正面向き(ヨー方向、ピッチ方向回転がそれぞれ 7 度以下)の場合に限定する。図 11 に顔向き推定の出力結果の例を示す。図 11 の矩形が顔領域を示し、点が各特徴点、白線が顔向きのベクトルを示す。表 2 に A,B,C,D,E,F,G の各方向における真値と比較した平均誤差と平均最大誤差を示す。表 2 より、顔の回転角度が大きいほど、平均誤差、平均最大誤差ともに大きくなることが示される。平均最大誤差には特徴点の追跡失敗によって顔向き推定に失敗した結果が含まれる。また、図 12 に推定値と真値を比較したグラフを示す。図 12 (b)右図は 4 秒から 9 秒付近で特徴点の追跡に失敗したが、その後特徴点の再検出により再び顔向きの推定が可能になった例である。

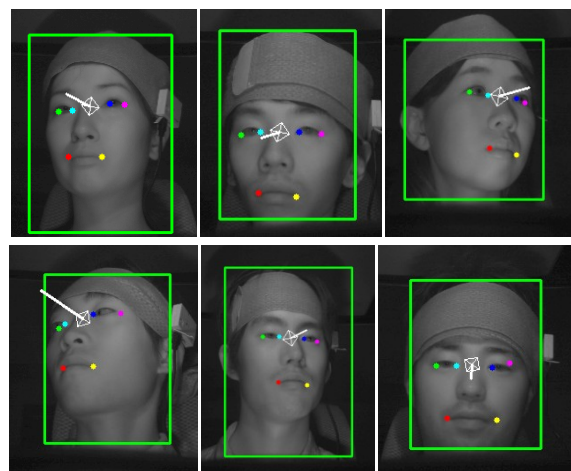


図 11 顔向き推定結果

表 2 推定した顔向きの平均誤差および平均最大誤差

	平均誤差[度]	最大誤差[度]
A	2.08	5.51
B	5.25	14.69
C	5.84	18.8
D	3.09	8.07
E	3.67	9.58
F	3.58	10.18
G	2.19	6.55
平均	3.67	10.48

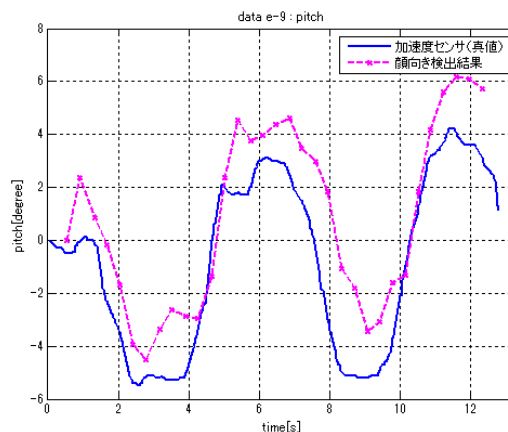
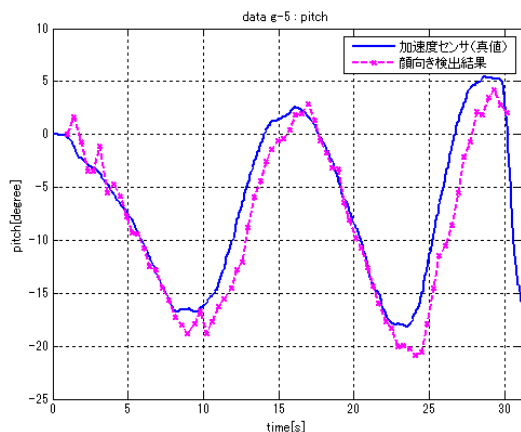


図 12(a) 推定角度と真値の比較 (ピッチ方向)

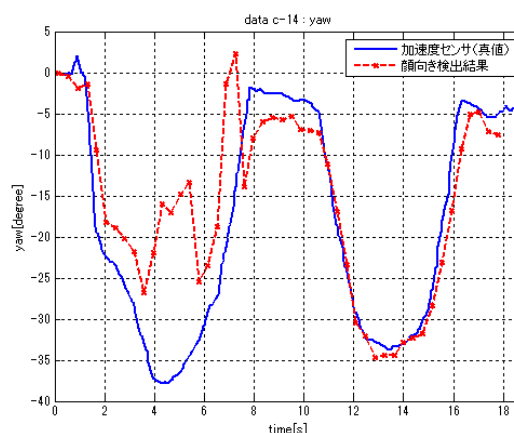
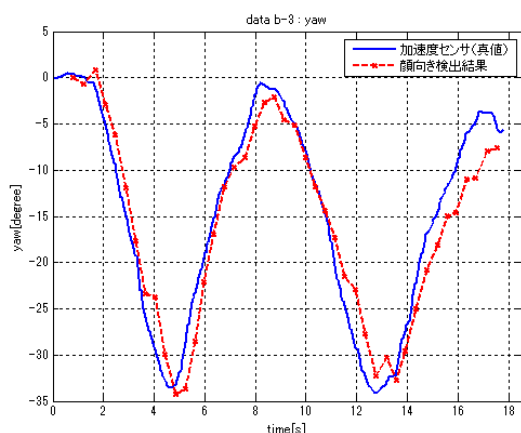


図 12(b) 推定角度と真値の比較 (ヨー方向)

5. 結論

本稿では Kinect センサを単眼距離カメラとして利用した顔向き検出法を提案した。入力画像である赤外線画像と距離画像は、Kinect の赤外線エミッタの ON/OFF を随時切り替えることによって取得した。実験の結果、平均誤差 3.67 度の精度で顔向きの推定が可能であった。今後の研究課題として、回転した顔における特徴点検出の精度向上が挙げられる。また、現段階では顔特徴点が全て平面状に存在すると仮定した簡易モデルであるため、何らかの 3 次元顔モデルを適用することで、特徴点の検出および追跡の精度向上が期待できる。

参考文献

- [1] Qin Cai, David Gallup, Cha Zhang, Zhengyou Zhang : 3D Deformable Face Tracking with a Commodity Depth Camera. European conference on computer vision conference, 229-242 (2010)
- [2] Gabriele Fanelli, Juergen Gall, Luc Van Gool : Real Time Head Pose Estimation with Random Regression Forests. CVPR, 617-624 (2011)
- [3] Dalong Jiang, Yuxiao Huc, Shuicheng Yand, Lei Zhang, Hongjiang Zhang, Wen Gao : Efficient 3D reconstruction for face recognition. Pattern Recognition 38 787 - 798 (2005)
- [4] Kinect -Xbox.com. <http://www.xbox.com/en-CA/kinect>
- [5] DENSO driver status monitor. http://www.denso.co.jp/ja/news/event/globalmotorshows/2013/files/its13_driver_status_monitor.pdf.
- [6] TOYOTA Pre-crash Safety System with Driver Monitor, http://www.toyota-global.com/innovation/safety_technology/safety_technology/technology_file/pre_crash/
- [7] Nikolai Smolyanskiy, Christian Huitema, Lin Liang, Sean Eron Anderson : Real-time 3D face tracking based on active appearance model constrained by depth data. Image and Vision Computing 32 860-869 (2014)
- [8] M. Pamplona Segundo, L. Silva, O.R.P. Bellon, S. Sarkar : Orthogonal projection images for 3D face detection. Pattern Recognition Letters 50 72-81 (2014).
- [9] P. Viola and M. Jones : Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE CVPR, (2001).
- [10] R. Lienhart and J. Maydt : An Extended Set of Haar-like Features for Rapid Object Detection. IEEE ICIP 2002, Vol. 1: 900-903, (2002).
- [11] Yanchao Dong : Study on 3D model based face pose and animation tracker for driver inattention monitoring system. Graduate School of Science and Technology, Kumamoto University, thesis for the degree of Doctor of philosophy (2011).
- [12] J. Shi and C. Tomasi : Good features to track. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR94), (1994)
- [13] J. Y. Bouguet : Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm. Technical report, OpenCV Document, Intel Corporation Microprocessor Research Labs (1999)