

ステレオ時系列画像から頭の姿勢の検出 Detecting Head Pose from Stereo Image Sequence

○徐 明, 赤塚孝雄
Ming XU, T. AKATSUKA

山形大学工学部 電子情報工学科
Electronic and Information Dept., Engineering Faculty, Yamagata University

キーワード：姿勢検出 (Pose Detection), 立体視 (Stereo Vision), 顔認識 (Face Recognition)

連絡先：〒992米沢市城南4-3-16山形大学工学部 電子情報工学科 赤塚研究室 徐明
Tel: 0238-26-3323, Email: xum@eat72.yz.yamagata-u.ac.jp

1. はじめに

人間の頭の姿勢情報はコンピュータヒューマンインタアクション, 映像通信, 顔認識など様々な分野で使われている。頭の画像からその姿勢を求めるには, 主に頭の幾何特性を利用する方法^{1, 2, 3)}とモデルに基づく方法^{4, 5)}がある。他に, 入力顔画像と頭の三次元モデルから生成した顔画像との相関による方法⁶⁾もある。幾何特性による方法は, 顔要素の長さなどの測定は既知とされる。モデルによる方法は, 頭の画像と頭のモデルの間に, 少なくとも三つの対応点が要求される。これらの方法は一枚の二次元画像をしか利用しないので, 三次元の姿勢を求めるには, 足りない情報を既知知識として与えなければならない。しかし, 顔のパラメータや画像とモデルの間の対応点などの既知情報が得られない場合がよくあるので, これらの方法の応用は限られている。相関法は, 既知知識を必要としないが, 精度は十分ではない。

そこで, 我々はステレオ画像から頭の姿勢を高速かつ精度良く検出する方法を提案する。頭に関する三次元情報を直接抽出されるので, 既知知識がなくて済むという利点がある。本報告では, 立体視による頭の姿勢の検出法を説明し, その有効性を解析及び実験によって示す。最後に, この方法が顔認識への応用について検討する。

2. 立体視による頭の姿勢の検出

2.1 姿勢検出の原理

平面の傾きの方向を, その面に位置する非共線の三点の空間座標を分かれば決める。人間の頭は剛体と考えられ, その姿勢は顔の正面の傾きから唯一に決められる。人間の顔の一般的な特徴から見ると, 目の瞳孔, 口端は同じ平面に位置し, その平面は顔の傾きを最もよく反映している(口の極端の変動する場合を除く)。従って, 瞳孔と口端の中の三点の空間座標を分かれば, その三点が成す平面の法線を求めて, 顔の傾き, そして頭の姿勢を検出することが出来る。

瞳孔と口端を選ぶ理由は二つがある。まず, 瞳孔と口端は鼻の穴, 下唇の陰, 鼻の脇など良く使われる顔特徴より, 頭の姿勢や照明の変動に影響されにくく, 安定的な抽出が出来る。また, 特徴三角形の頂点は顔の中軸から離れるので, 正面姿勢の測定精度が向上した(第三節)。

2.2 顔特徴の抽出

顔特徴の抽出は姿勢検出の性能に大きく関わっているため, 位置精度が良く, 処理時間が短い抽出方法が望まれている。ここで, 我々は形状抽出とパターンマッチングを組み合わせた方法^{7, 8)}を利用して, 瞳孔と口端の抽出を行う。まず, 顔画像に対して, 二つの同心円からなるテンプレートをかけて, 各点の輝度分離度を計算し, 出力が最大の点を特徴の候補点として選ばれる。テンプレートの半径は特徴点の半径とする。そして, 各候補点は瞳孔と口端のパターンと照合し, 類似度が

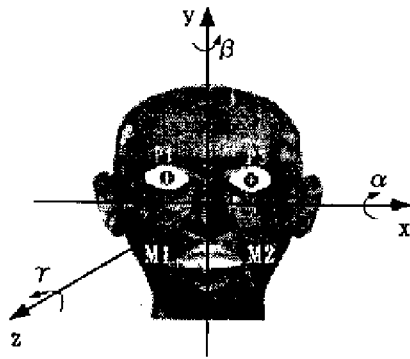


fig.1 Feature Points of Face

一番高い候補は特徴点とする。

2.3 三次元位置の計算

ステレオカメラを平行に設置され、顔の特徴点とその左右画像面にある投影点の関係はfig. 2に示される。ワールド座標系は0-X-Y-Zであり、左側のカメラの座標系と一致させる(ただ、Z方向で焦点距離を離れた)。両カメラの光軸間の距離はLである。

左右両画像から抽出された特徴点の対応点は同じ水平の走査線に位置し、視差は対応点のX座標の差である。

$$D_p = x_l - x_r \quad (1)$$

三角測量の原理とカメラのパラメータから、特徴点が画像面までの距離Zが求められ、

$$Z = \frac{Lf\mu}{D_p} \quad (2)$$

f はカメラの焦点距離、 μ はカメラのCCDセンサの空間分解率。

そして、特徴点のXとY座標も次のように求められる、

$$X = \frac{Z+f}{f} X_l, Y = \frac{Z+f}{f} Y_l \quad (3)$$

2.4 顔の傾きの計算

顔の傾きは瞳孔と口端からなる三角形の法線方向によって表わされる。顔は正面から大きく回転する場合を除けば、四つの三角形P1-P2-M1, P1-P2-M2, P1-M2-M1とP2-M2-M1の中の少なくとも一つを検出できる。特徴点A, B, Cのそれぞれの三次元座標を (a_1, b_1, c_1) , (a_2, b_2, c_2) と (a_3, b_3, c_3) とすれば、三角形の法線ベクトル

$$N = a\hat{x} + b\hat{y} + c\hat{z} \quad (4)$$

その中、

$$\begin{aligned} a &= a_1c_2 + b_2c_3 + b_3c_1 - b_1c_3 - b_2c_1 - b_3c_2 \\ b &= a_3c_2 + a_1c_3 + a_2c_1 - a_3c_1 - a_1c_2 - a_2c_3 \\ c &= a_1b_3 + b_2a_3 + a_2b_1 - a_1b_2 - a_2b_3 - a_3b_1 \end{aligned} \quad (5)$$

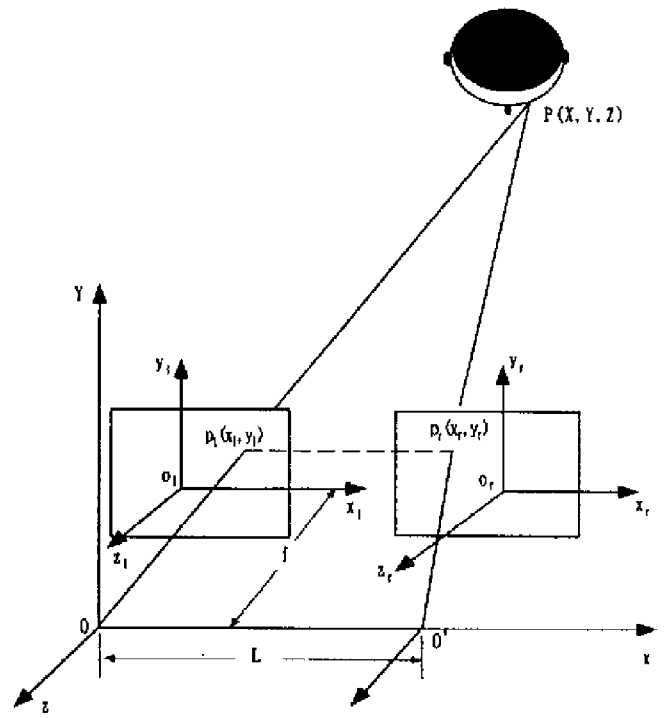


fig.2 Configuration of Stereo Camera

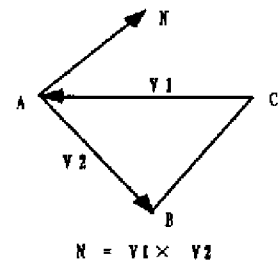


fig.3 Normal Direction of Triangle

さらに、顔の傾きは正面位置からX, YとZ軸に回って回転した角度 α , β , γ で表示する。 α と β は三角形の法線ベクトルと面 X-0-Z及び面 Y-0-Z成す角度に等しいので、

$$\alpha = \arcsin \frac{b}{\sqrt{b^2 + c^2}} \quad (6)$$

$$\beta = \arcsin \frac{a}{\sqrt{a^2 + c^2}} \quad (7)$$

角度 γ は画像面と平行する回転を表わすので、瞳孔の画像座標から直接求められる。

$$\gamma = \arctan \frac{Y_{p2} - Y_{p1}}{X_{p2} - X_{p1}} = \frac{y_{lp2} - y_{lp1}}{x_{lp2} - x_{lp1}} \quad (8)$$

x_{lp1} , y_{lp1} と x_{lp2} , y_{lp2} は特徴点P1とP2の左画像面での座標。

3. 精度の分析

特徴点の三次元座標は顔の傾きの検出精度を直接に影響している。その中、最も大きな役割を果

しているZ座標の精度に関係する要因は三つがある：特徴点を抽出する精度，撮像ステレオカメラのパラメータ及び顔とカメラの相対位置。

3.1 距離分解度と角度分解度

ステレオカメラの距離分解率 ΔZ はカメラのパラメータによって決める：

$$\Delta Z = \frac{C}{d} - \frac{C}{d+1} = \frac{C}{d(d+1)} \quad (9)$$

その中，

$$C = Lf\mu \quad (10)$$

d は左右画像での視差である。

$d \gg 1$ のため，

$$\Delta Z \approx \frac{C}{d^2} \quad (11)$$

(2) を (11) に代入して，

$$\Delta Z \approx \frac{C}{d^2} = \frac{Z^2}{C} = \frac{Z^2}{Lf\mu} \quad (12)$$

頭が回転する時，特徴点を球面に半径 r の円周に回転すると近似でき，球面に位置する特徴点は Z 軸 (Y 軸の時も) と成す角度 θ とする。特徴点は角度 $\Delta\theta$ の回転をする時の奥行きの変動は Δd とすれば，

$$\Delta d = r(\cos\theta - \cos(\theta + \Delta\theta)) \quad (13)$$

円周の半径は特徴点の位置によって変わる。顔が正面向きの場合，円周の半径 r を 65mm，Z を 500mm に仮定すれば， θ は $-85^\circ \sim 85^\circ$ の範囲に， $\Delta\theta$ は 1 度，3 度，5 度の場合 Δd の数値を fig. 4 に示し，さらに，(12) から計算した $\Delta\theta$ の角度分解率を持つために必要な位置検出精度を fig. 5 に示されている。

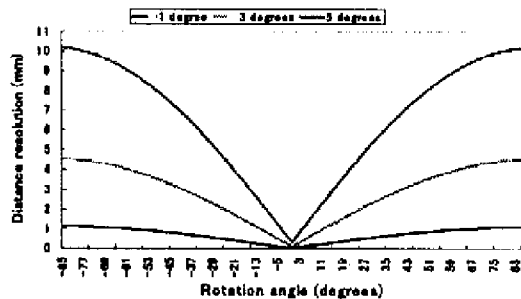


fig.4 Depth resolution

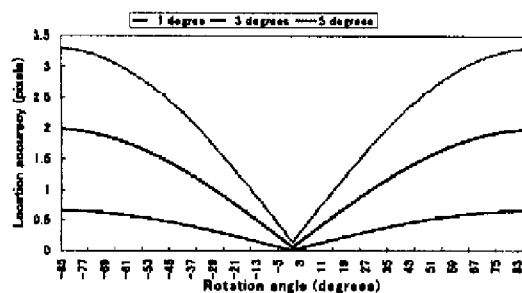


fig.5 Location accuracy

3.2 位置による誤差

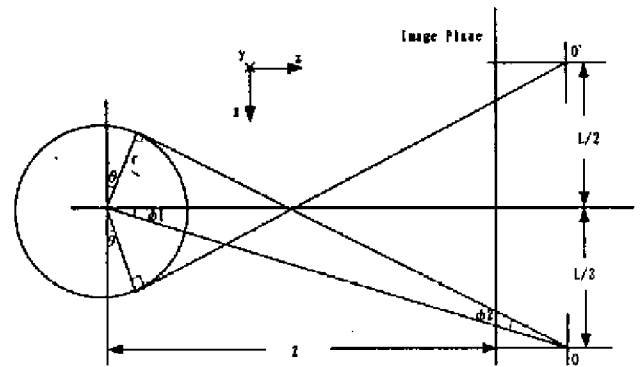


fig.6 Position-related error

特徴点は Y 軸を回って角度 θ に至る時，画像面にその投影点の X 座標は極値になり，この角度の周りに X 座標の変化は最も少なくなり，求めた視差の誤差が大きくなり，奥行き値，そして，検出した傾き角度の誤差も大きくなる。 θ の値は球の半径，カメラの画像面までの距離，カメラの光軸間の距離に影響される。Fig. 6 から，

$$\theta = \varphi_1 + \varphi_2 \quad (14)$$

その中，

$$\varphi_1 = \arctan \frac{L/2}{Z+f} \quad (15)$$

$$\varphi_2 = \arcsin \frac{r}{\sqrt{(Z+f)^2 + (L/2)^2}} \quad (16)$$

$L=200\text{mm}$ ， $r=65\text{mm}$ ， $f=6\text{mm}$ の時， θ は 18.5° である。

同様に，特徴点は X 軸を回って θ に至る時，投影点の Y 座標の変化は最も少なくなり，角度検出の誤差が大きくなる。

以上の分析によると，特徴点は Z 軸と成す角度はゼロまたは θ の時，角度検出誤差が最も大きいことが分かった。顔が Z 方向に向う時，三角形の頂点である特徴点は誤差の大きい領域外にあるので，傾きの角度を精度良く検出できる。Fig. 5 から，瞳孔や口端の検出誤差は 2 pixel の時でも，姿勢の角度の誤差は 5 度以内に留まることが分かった。頭の回転と共に，特徴点は誤差の大きい領域に入り，姿勢の検出誤差が大きくなることが予想できる。

4. 実験

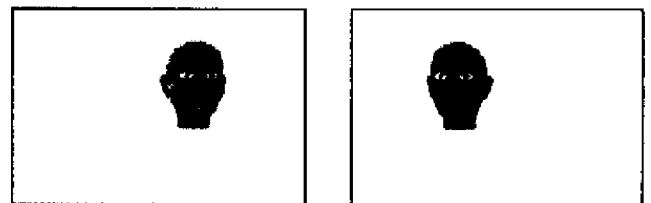


fig.7 CG Stereo image

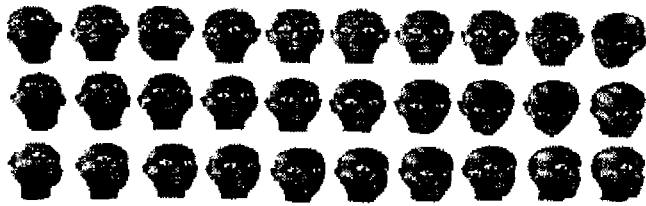


fig.8 CG head images with different poses

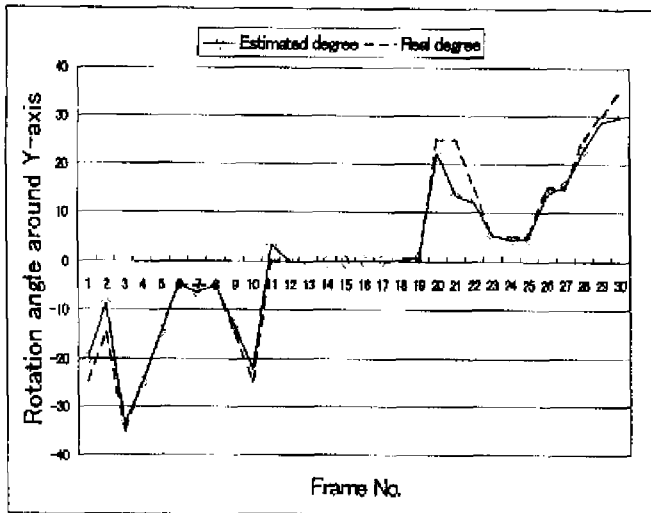


fig.9 Rotation around Y-axis

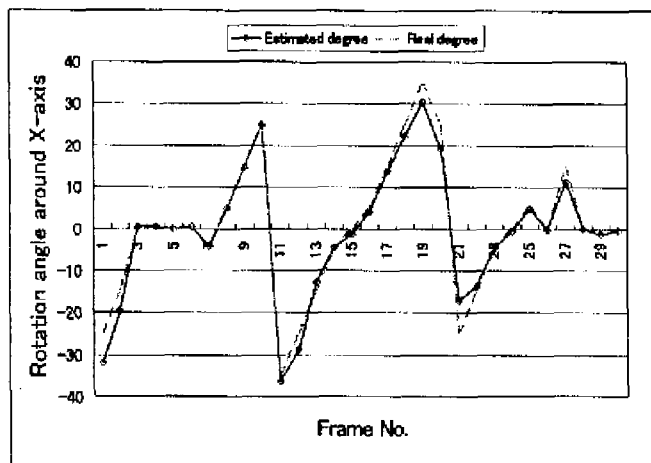


fig.10 Rotation around X-axis

提案した姿勢検出法の精度を検証するために、CG画像と実画像を利用して、実験を行った。まず、CGツールによって制作された頭のモデルを用いて、顔のテクスチャーをモデルに張り付ける。頭のモデルを回転させ、実のステレオカメラと同じパラメータを持つ仮想ステレオカメラによってモデルの各方向への投影像を撮る。Fig. 7は α 、 β はゼロである時のステレオ画像。Fig. 8は様々な姿勢で生成したCGの顔画像。瞳孔と口端の位置を抽出し、第二節で述べた方法を利用し頭の姿勢を検出する。モデルはX軸とY軸を回って -35 度から 35 度まで、 5 度の歩合で回転させ、 49 組のステレオ画像を撮った。これらの画像は顔が正面から上下左右への回転に対応している。 β と α の検出結果と実際

の値の比較をfig. 9とfig. 10に示し、誤差の統計結果をTable. 1に示す。

Table.1 Error data of CG image

	Min.	Max.	Average
α	0	11.4	1.82
β	0.1	7.79	1.79

さらに、頭のCGモデルから作られた頭の紙積層モデルを利用し、実のステレオカメラで実験を行った。モデルは回転ステージに置かれ、ステージの回転角度をPCで制御する。モデルをY軸に回って 5 度ずつ回転させる。Fig. 11には、 0 度、 10 度、 20 度、 30 度と 40 度の位置での左画面の像を表示している。



fig. 11 Example of head images (left camera)

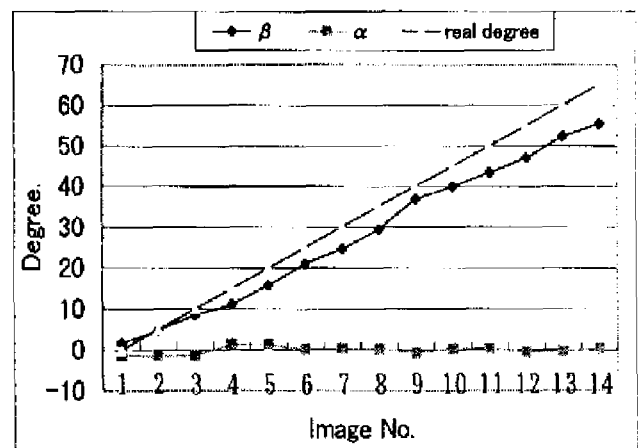


fig.12 Estimated rotation angle

姿勢検出の結果はfig. 12 に示し、誤差の統計値はTable. 2に表示される。

Table.2 Error data of model image

	Min.	Max.	Average
α	0.046	9.5	4.77
β	0.258	1.395	0.724

モデル画像の場合、誤差の平均値はCG画像より大きい、これはレンズの誤差によるものと考えられる。誤差の分布はCG画像と同じ結果がある。即ち、回転角度が小さい時、誤差が小さい。

実験に使われるステレオカメラは光軸間の距離は 200 mm、モデルまでの距離は大体 500 mm、レンズ

の焦点距離は6mm, CCDの空間分解度は9.7 μ m/pixel (Y方向), 8.3 μ m/pixel (X方向), そしてステレオカメラの距離分解率は1.9mm/pixel (Z方向)である。

精度を上げるには, 特徴点の抽出誤差を減らすか, (9)によってステレオカメラの距離分解率を上げるか二つのルートを考えられる。

5. 応用

実人物と写真の照合に基づく顔認識システムには, 頭姿勢の変動によって起こった顔画像の変化認識システムの性能を大きく影響している⁹⁾。それを対応するため, 人に決まった姿勢を取らせて画像を撮るのは可能であるが, インタフェースの柔軟性がなくなる。また, 頭が様々な姿勢での顔画像の実写を撮るか, 少枚数の顔画像から新しい視点での仮想像を生成するかによって姿勢の変動をカバーする学習空間を作るアプローチもある。このアプローチの問題点は様々な姿勢での実写を撮るのは現実で困難であり, 仮想画像の生成は複雑な計算が要る。

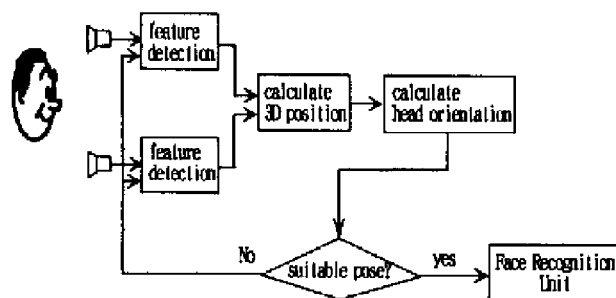


fig.13 Scheme of active face recognition

ここで, 我々は顔の時系列ステレオ画像を認識の入力として使うというアクティブ認識方法を提案する。Fig.13はこのアプローチの構成を示す。この方法は膨大な学習空間を作る必要を無くし, インタフェースの柔軟性を保つという利点がある。

提案した姿勢検出法は精度が良い, 特に正面像に近い姿勢の検出精度が高い, しかも処理速度が速いので, アクティブな顔認識に適用であると考えられる。

6. おわりに

本文はステレオ画像から頭の姿勢を検出する方法を提案した。検出の精度を分析及び実験で検討し, 高い精度を得られることが分かった。時系列画像を利用し, アクティブ的な顔認識アプローチへの適用を議論した。

今回の実験では, スナップショット写真だけを使った。ビデオレートの時系列画像に対する実験

は今後の課題として進めていくつもりである。

参考文献

- 1) 大村 和典, 伴野 明, 小林 幸雄: “単眼視画像による顔の向き検出法の指示入力への応用”, 信学論 Vol. J72-D-II, pp.1441-1447 1989年9月
- 2) 間瀬 健二, 渡辺 保日児, 末永 康仁: “ヘッドリーダ: 画像による頭部動作の実時間検出”, 信学論 Vol. J74-D-II pp.398-406, 1991年3月
- 3) Thanarat Horprasert, Yaser Yacoob and Larry S. Davis, “Computing 3-D Head Orientation from a Monocular Image Sequence”, Proceeding SPIE Vol. 2962, p. 244-252, 25th AIPR Workshop: Emerging Applications of Computer Vision
- 4) Thomas Maurer, “Estimation of Face Position and Pose with Labeled Graphs”, Proceeding of the BMVC 1996
- 5) Andrew Gee and Robert Cipolla: “Fast Visual Tracking By Temporal Consensus”, Image and Vision Computing, 14(2), 105-114, 1996
- 6) 塚本 明利, 李 七雨, 辻 三郎: “複数のモデル画像による顔の動き推定”, 信学論, Vol. J77-D-II pp.1582-1590 1994年8月
- 7) 福井 和宏, 山口 修: “形状抽出とパターン照合の組み合わせによる顔特徴点抽出”, 信学論 Vol.80-D-II pp.2170-2177 1997年8月
- 8) Alan Yuille and Peter Hallinan: “Deformable Templates”, pp.21-38, Active Vision, The MIT Press 1992
- 9) 赤松 茂: “コンピュータによる顔の認識 ——サーベイ——”, 信学論 Vol. J80-A, pp.1215-1230 1997年8月