

視線推定の頭部自由化のための 距離画像を利用した頭部姿勢推定

高田 憲一¹ 北須賀 輝明¹ 有次 正義¹

概要：本稿では頭部の姿勢を固定しない視線推定のための距離画像を用いた頭部姿勢の推定方法を提案する．提案手法はハードウェアとして Kinect for Windows を使用し，距離画像の中から鼻の位置を追跡しながら，その鼻の位置を基準として頭部姿勢を推定する．鼻の位置を追跡するための特徴は，簡単な処理により現れ，顔の向きが変わっても現れる．距離動画像を使用するため，カラー画像ではわからないカメラからの距離がピクセルごとにわかり，また照明条件による色の変化がないといったメリットがある．また，提案手法による頭部姿勢の推定では機械学習を用いないので学習データを必要としない．推定した結果として，頭部の垂直方向での回転角度は平均で約 3.5 度，水平方向の回転角度は平均で約 1.65 度の誤差となった．また，頭部を固定した状態での視線推定では誤差は約 2 度となった．

Head Posture Estimate Using Depth Image for Head Liberalization of the Eyes Estimation

KENICHI TAKADA¹ TERUAKI KITASUKA¹ MASAYOSHI ARITSUGI¹

1. はじめに

人の目は，他の動物の目と異なり，白目がはっきりと見える．白目があることで，視線がどこを向いているかがわかる．これは，動物にとって敵に自分の視線が感知されてしまうため弱点になるが，人はこれをコミュニケーションに使っており，重要な情報を持っている．そのため，視線の情報を利用しようとする動きがある．たとえば，視線からどこを見ているかが分かると，何に興味があるかを推測することができるため，その情報を元に効果的な広告を提供しようとするものがある [1]．また，視線を使用すると，簡単に 3 次元上の位置を指定できるため，ロボット操作の入力装置としても使用が期待されている．一般ユーザにとっても手を使わずにポインティングデバイスのような入力デバイスとして使用できるため，手がふさがっている場合や手がぬれている場合などにもコンピュータへと入力ができる．そのため，視線を推定する研究は盛んに行われている．また，実際に販売しているものもあり，筋萎縮性

側索硬化症のために体があまり動かない人などが使用している．しかし，視線を推定する研究では高価で特殊なハードウェアを必要とするものや，頭を固定するように要求したり，頭の動きや姿勢を制限するなどのユーザへの負担により，一般家庭に視線推定は普及していない．そのため，ハードウェアの価格や入手のしやすさ，頭の姿勢においてユーザの負担が小さい視線を推定するシステムが望ましい．

本研究では，ハードウェアとして距離画像を撮影できる Microsoft の Kinect for Windows を使用することで，これらの問題に対処する．カラー画像では照明条件の変化による色の変化や 顔の凹凸のような奥行きが分からないといった課題がある．本研究では，ピクセルごとにカメラからの距離が分かる距離画像を使用して頭部姿勢を推定する．カラー画像は頭部姿勢推定後の視線推定のみに用いる．

Kinect から取得した距離画像からシンプルな処理により抽出した特徴から鼻の位置を追跡していき，その鼻の位置を基準として頭部姿勢を推定する．そして，頭部姿勢を正面向きに変換する．頭部姿勢を推定する方法として，カラー画像と距離画像を組み合わせモデルを使うもの [2] があるが，精度の高いモデルを作成するためには膨大な数

¹ 熊本大学大学院 自然科学研究科
Graduate School of Science and Technology, Kumamoto University

の顔画像が必要になるため、本研究では使用しない。

また、視線推定では頭部が正面向きのカラー画像を使用し、画面のどこを見ているかを推定する。本稿では頭部姿勢を推定し、頭を正面向きに見えるように変換する方法と、頭が正面向きでの視線推定方法について説明するが、それらを組み合わせた頭部姿勢の自由な視線推定については扱わない。

以降、2章では、視線推定方法や頭部姿勢推定についての関連研究を紹介する。3章では、提案手法である距離画像を利用した頭部姿勢推定の手順について述べる。4章では提案手法を使用した実験とその結果について示す。5章では本稿のまとめと今後の課題を述べる。

2. 関連研究

これまでに視線推定のための多くの方法が提案されている。商品として販売されているものもあり、筋萎縮性側索硬化症のために体があまり動かすことができない人が使用している。しかし、高価で一般人にとっては入手するのは難しい。また赤外線を瞳に当てその反射を見ることによって視線を推定するもの[3]があるが、特殊なハードウェアを必要とするため、これもまた現時点では一般人にとって入手することが難しい。そこで、入手しやすいハードウェアとして Web カメラを使用して視線を推定しようとする研究[4]がある。それらの研究の課題に頭部姿勢の変化による目の外観の多様化がある。つまり、頭部姿勢が変わるごとに目の画像の見え方が変わることによって、視線推定が難しくなる。その変化に対応するために姿勢が変わるごとに学習データを収集し、大規模な学習データを必要とするものがある。しかし、大規模な学習データを必要とすると、学習に必要な時間が長くなり、利用者にとって負担になる。また、目の見え方が変わらないように頭部を固定するものや、頭部姿勢を厳しく制限するものもある。しかし、これも一般のユーザにとって負担が強い。

目の画像の見え方は頭部姿勢に依存するため、視線推定において頭部姿勢を推定することが重要になる。頭部姿勢推定のために AAM(Active Appearance Model)[5] や 3DMM(3D Morphable Model)[2] などのモデルを使うものがある。これらのモデルを使用するには学習データが大量に必要となる。

3. 提案手法

本研究では、以下の手法により頭部姿勢の自由な視線推定方法を想定している。はじめに、取得した距離画像から頭部姿勢を推定し、頭を正面向きにした画像へと変換する。そして、その正面向きにしたカラー画像の目の部分から、正面向きの顔に対しての視線方向を推定する。その視線方向のデータを元の頭部姿勢の方向へと戻し、画面のどこを見ているのかを推定する。

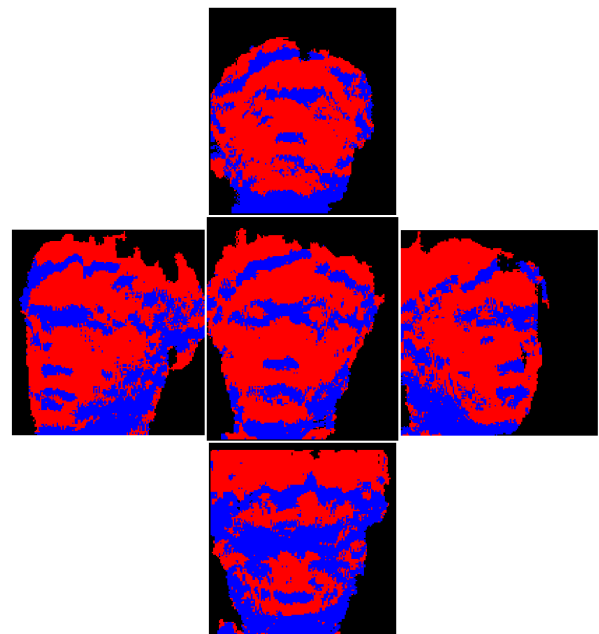


図 1 さまざまな向きでの特徴画像
1 段目中央：上向き、2 段目左：右向き、中央：正面、右：左向き、3 段目中央：下向き

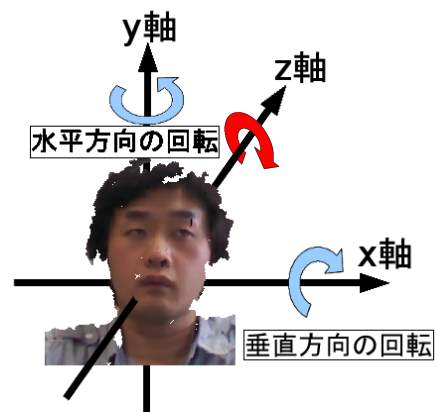


図 2 xyz 軸

このうち、本研究では距離画像を利用して頭部姿勢を推定し頭部姿勢が正面向きになるように変換する部分と、頭部が正面向きのカラー画像を用いて視線を推定する部分についてのみ扱う。元の頭部姿勢に戻し、画面のどこを見ているかを推定する方法については扱わない。

提案手法では、距離画像から鼻の特徴を見つけ、鼻の位置を追跡しその位置情報を元に頭部姿勢を推定する。ハードウェアには距離画像とカラー画像を取得できる Kinect を使用する。距離画像からピクセルごとに傾きを求め、その傾きを元に図 1 のように赤と青の領域の 2 つに分割した。そして、図 1 のように顔の向きを変えても鼻の領域に青の部分が現れるため、その部分を追跡することで鼻の位置を見つける。見つけた鼻の位置を基準に周りの深度を見ることで頭部姿勢を推定し、正面向きに見えるように画像を変換する。ただし、図 2 に示す 3 つの回転のうちで水色の矢

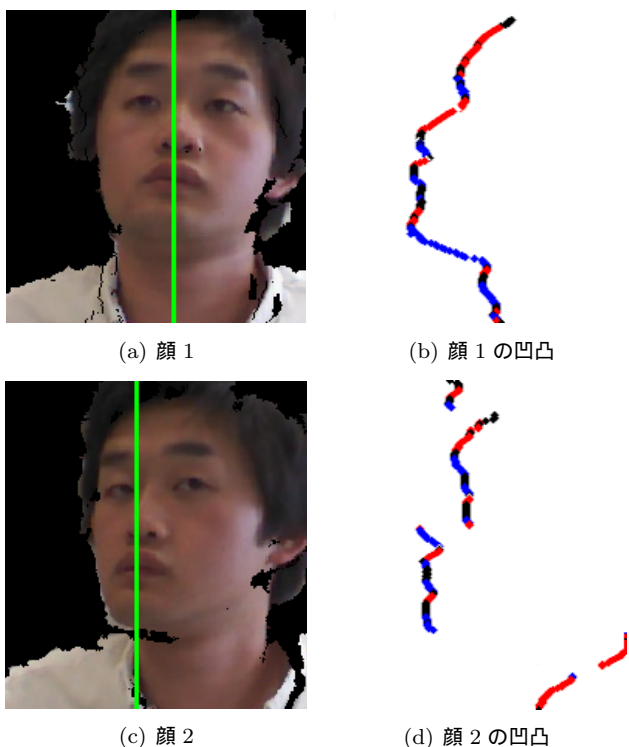


図 3 顔画像とその凹凸

印がある x 軸と y 軸周りの回転角度のみ推定し、赤色の矢印がある z 軸周りの回転角度については取り扱わない。 z 軸周りの回転は、首をかしげる動作に対応する回転である。

また、頭部を固定した状態での視線を推定する。canny エッジフィルタを使い、目の端を見つけ目の画像を切り出す。そして、目の画像を横に 3 分割、縦に 5 分割し、各ブロックごとに輝度値の合計を求め、全ブロックの合計が 1 になるように調整し、これを特徴とする。この特徴を使いサポートベクター回帰により視線を推定する。

3.1 顔の向きの推定および回転

本研究では、頭部の 3 次元のモデルを使い頭部姿勢を推定するのではなく、顔の中の特徴的な場所を見つけ、そこから頭部姿勢を推定する。顔の向きが変化しても距離画像に現れ続ける特徴として、鼻の周りに着目する。そして、顔の水平方向と垂直方向の傾きを求める。

はじめに、顔の距離画像から頭部の水平方向のおおよその傾きを求めるために、顔の凹凸に着目する。図 3(a) と図 3(c) のカラー画像の緑の線があるラインの凹凸を、図 3(b) と図 3(d) に示す。図 3(b) を見ると鼻の部分以外にも唇の部分も前に出ていることがわかる。また、図 3(d) は少し顔を右に向けたとき凹凸だが、図 3(b) とは違い、図 3(d) では垂直方向にみただけでは鼻のほとんどの部分が表示されていない。そのため、縦に凹凸を見るだけでなく、わかりやすい特徴が必要になると考えた。

そこで、ピクセルごとの傾きに注目した。各ピクセルごとに垂直方向の距離を注目しているピクセルとその上下 2

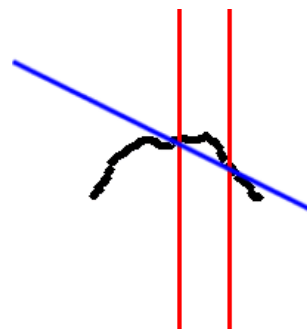


図 5 上から見た図

ピクセルずつ、計 5 ピクセル分ずつ取り出し最小二乗法により傾きを求める。そして、顔全体の傾きを平均し、平均より奥向きに傾きの大きい部分を (青) とし、手前向きに傾きの大きい部分を (赤) に分ける。このように分けると図 1 のようになる。図 1 で、鼻の一部分が青色になっていることがわかる。この特徴は図 1 で示すように、正面だけでなく上下左右のどの向きでも、青くなっており特徴が現れていることがわかる。これは、図 3(b) で示すとおり鼻の一部である。このように、鼻が丸みを帯びており徐々に傾きが大きくなっているため、このように特徴として現れる。これは図 1 のように頭部の姿勢を変えても現れる。Kinect からのデータは毎秒約 30 フレームなので、フレーム間での頭の大きな動きはないと仮定し、鼻の下部分を追いかけることで鼻の位置を追跡する。

次に、この鼻の下範囲の左端と右端から頭部の水平方向の傾きを求める。青の範囲の左右の端から外側に 5 ピクセル分の奥行きデータを取得し、最小二乗法により傾きを求める。青の範囲の一番下の横のラインの凹凸を示したものを図 5 に示す。図 5 は、黒が顔の凹凸で、赤線が青の範囲の両端を示し、青線が傾きを示している。そして、その傾きに従って、水平方向で回転させてカメラと平行にする。回転させた画像を図 4(c) に示す。回転させる際に、 xy 方向はピクセル単位だが、 z 方向は mm 単位になっているため、単位を合わせる。式 1 によって変換させる。

$$\begin{cases} x' = (x - cx) * (0.001472 * z + 0.1606) \\ y' = (y - cy) * (0.001662 * z + 0.0255) \\ z' = z \end{cases} \quad (1)$$

最後に、次に頭部の垂直方向の傾きを求める。これには、鼻の頭から上下に一定の距離の間の傾きを求めることにした。実験では、その距離を上下に 50 ピクセルずつとした。そうして、求めた傾きから垂直方向で回転させる。回転させた画像は図 4(d) となる。

本研究では、 z 軸周りの回転は扱わない。

3.2 視線推定

3.1 節の手順で頭部姿勢を推定したあと、頭部姿勢が正面向きになるようにカラー画像を回転させる。回転した力

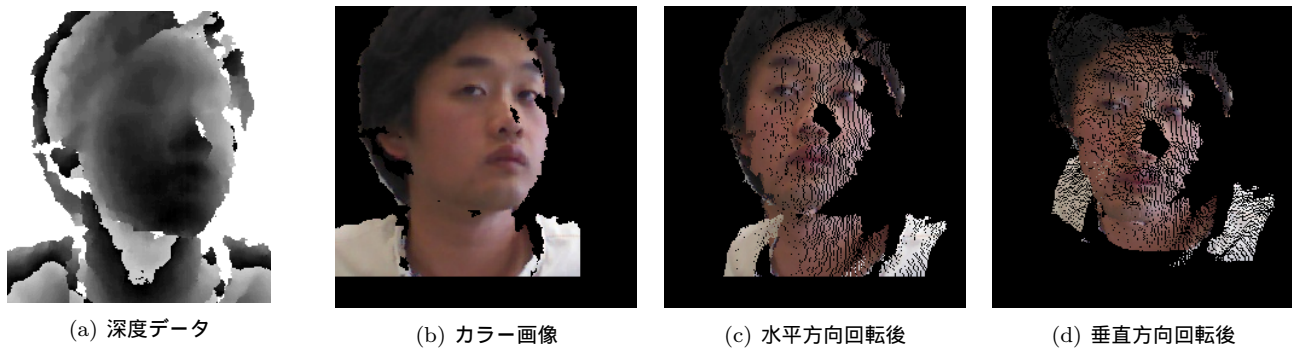


図 4 頭部回転



図 6 特徴抽出

ラー画像に対して，Feng ら [6] の方法を参考に特徴を抽出する．

まず，目のあたりの画像を切り取る．次に，目の位置を推定するために図 6(b) のように canny エッジフィルタを使い，目の端を見つける．そして，図 6(c) のように，画像の横方向は目の両端で切り取り，画像の縦方向では目頭が 3 分の 2 になる位置になるようにし，横幅の 5 分の 3 となるようにする．そのあと，図 6(d) のように目の画像を 3×5 の 15 ブロックに分割し，各ブロックごとに輝度の合計を求める．そして，全ブロックの合計が 1 になるように調整し，これを特徴とする．頭を動かすと，カメラと頭の距離や角度が変わり，目の部分の画質が変わるが，この特徴ではブロックごとに平均を取るため，その影響を受けにくい．

学習にはサポートベクター回帰を使用する．学習データは次のように取得する．図 7 の赤の点のような点がいくつかあり，それぞれの点を見たときの目の画像をクリックするごとに取得し特徴を抽出する．この点のそれぞれを見ているときの特徴を学習データとする．そして，サポートベクター回帰のパラメータの調整のために，図 7 の青の点のように学習に使用した点と異なる位置を見たときの目の画像から特徴を取得する．そして，パラメータ C と γ の 2 つを調整する．グリッド探索により，推定した視線の位置と実際に見ていた視線の位置の誤差が小さくなるように値を定める．はじめは粗く探索し，徐々に探索範囲を狭めながら探索する．

4. 実験

頭部姿勢推定と頭部を固定した視線推定の 2 つの実験をした．データの収集には Kinect を使用した．カメラのフ

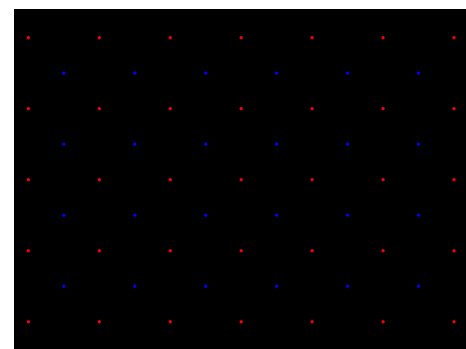


図 7 学習用 35 点，調整用 24 点

表 1 水平方向の回転

回転方向	回転角度 [度]	推定角度 1 [度]	推定角度 2 [度]	平均誤差 [度]
左回転	30	24.15	25.16	5.35
右回転	-30	-28.17	-28.51	1.66
平均				3.51

レームレートは約 30fps，カラー画像と深度データは 640×480 ピクセルで撮影した．

4.1 頭部姿勢推定

カメラから 600mm ～ 900mm 離れた部分で頭部を撮影した．カメラと頭部のなす角度を変えて撮影するために，頭部を固定した状態で，カメラを回転させ撮影した．水平方向では左右に 30 度回転させ撮影した．垂直方向では上下に 15 度回転させ撮影した．撮影は 2 回に分けた．その結果を表 1，表 2 に示す．

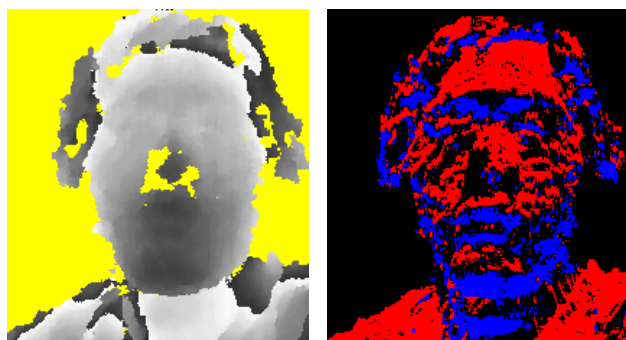
水平方向では誤差は平均で 3.51 度になった．そして，垂直方向では誤差が平均で 1.65 度になった．モデルを使用

表 2 垂直方向の回転

回転方向	回転角度 [度]	推定角度 1 [度]	推定角度 2 [度]	平均誤差 [度]
上回転	5	6.40	0.68	2.9
	10	9.06	9.60	0.67
	15	10.92	13.22	2.93
下回転	-5	-7.07	-3.03	2.01
	-10	-9.93	-9.46	0.31
	-15	-13.56	-14.20	1.12
平均				1.65



図 8 鼻の範囲推定のばらつき



(a) 距離画像

(b) 特徴抽出

図 9 追跡失敗画像

した場合 [2], [7] の頭部姿勢の誤差が 1 度以下に抑えられているため、それに比べると大きいものの、演算がシンプルであることと学習のためのデータを必要としないというメリットがある。図 8 は、黒の線は顔の水平方向における凹凸を 2 種類示しており、赤の線はそれぞれの鼻の端と推定された位置を示している。水平方向で誤差が大きくなった原因としては、図 8 に示す例のように、同じような顔の向きでも鼻の端と推定された位置が異なり、それによって傾きを求める場所にばらつきが生まれてしまうことが考えられる。

カメラを動かしたときの顔の距離画像からの鼻の部分の追跡は多くの場合で成功した。追跡に失敗した場合の距離画像を図 9(a) に、図 9(a) から赤と青に分けた画像を図 9(b) に示す。図 9(a) ではデータが取得できなかった場所を黄色で示している。この図 9 の場合に失敗した原因は、鼻の周りで距離データを取得できなかったことで鼻の下

表 3 視線推定結果

	横方向誤差 [pixel]	縦方向誤差 [pixel]	ディスプレイと 顔の距離 [mm]	角度差 [度]
実験 1	64.42	39.93	600	1.93
実験 2	51.50	71.62	600	2.25
実験 3	49.50	70.30	700	1.88
平均				2.02

特徴をうまく取れなくなったことだと考えられる。

4.2 頭部を固定した視線推定

学習用に使用した点は、図 7 に示す赤の点で、200 ピクセル間隔で 35 個配置した。パラメータ調整用に使用した点は、図 7 に示す青の点で、学習用に使用した点から xy 方向でそれぞれ 100 ピクセルずつずらした 24 個配置した。そして、パラメータを調整後、ランダムに配置した 15 個の点を見たときの視線を推定した。その推定した視線位置と指定した点の位置の距離の平均を求める。データは 3 度に分けて収集した。結果を表 3 に示す。推定した視線と実際の視線の角度差は平均で 2.02 度となった。

5. おわりに

本研究では、距離画像による頭部姿勢の推定方法を提案した。提案手法は、モデルを使うものではなく、顔の距離画像の中から特徴となる点から鼻の位置を追跡していき、その位置を基準として、頭部姿勢を推定した。そして、その頭部姿勢を元に顔の画像を正面向きになるように回転させた。

提案手法による実験では頭部姿勢の誤差は垂直方向で平均約 4 度、水平方向で約 1.7 度となった。この結果はモデルを使用したものに比べると精度は悪いが、シンプルな演算のみで、また学習用にデータを必要としないというメリットがある。

また、頭部を固定した視線推定では、目の画像を分割し、そのブロックごとの輝度値を特徴とし、サポートベクター回帰により学習した。その結果として、推定した視線の誤差は 2.02 度となった。

今後の課題

今後の課題として、まず頭部姿勢推定の精度向上のために、傾きを求めるために使用する点に、ばらつきが生まれないように定めるための方法を、考える必要がある。また、提案手法に含まれていない、z 軸周りの頭部姿勢推定の方法を検討する必要がある。前のフレームの情報がない場合でも、鼻の位置を見つけることができるような特徴により、追跡していなくても鼻の位置を見つけれられるようにする。そして、視線推定と組み合わせて精度を調べる必要がある。

参考文献

- [1] 杵渕哲也, 新井啓之, 他. 画像処理による広告効果測定技術. NTT 技術ジャーナル, pp.16-19, 2009.
- [2] Kenneth Alberto Funes Mora, Jean-Marc Odobez. Gaze estimation from multimodal Kinect data. CVPR Workshops 2012: pp.25-30
- [3] 大野権彦, 武川直樹, 吉川厚. 眼球形状モデルに基づく視線測定システム: 視線入力デバイスの実現に向けて. 情報処理学会研究報告. HI, ヒューマンインタフェース研究会報告 2001(38), pp.47-54, 2001.
- [4] Ba Linh Nguyen, Youssef Chahir, Michèle Molina, Charles Tijus, François Jouen. Eye gaze tracking with free head movements using a single camera. SoICT'10 , pp.108-113, 2010.
- [5] 中松由香梨, 滝口哲也, 有木康雄. 3次元 ActiveAppearanceModel を利用した視線方向推定. 画像の認識・理解シンポジウム, MIRU2011, IS2-11, pp.572-579, 2011-07.
- [6] L. Feng, Y. Sugano, O. Takahiro, and Y. Sato. Inferring Human Gaze from Appearance via Adaptive Linear Regression. In *ICCV: International Conference on Computer Vision*, Barcelona, Spain, 2011.
- [7] G. Fanelli, T. Weise, J. Gall, and L. V. Gool. Real Time Head Pose Estimation from Consumer Depth Cameras. In *33rd Symposium of the German Association for Pattern Recognition (DAGM'11)*, Sept. 2011.