

# Gaussian Processes for Regression と AAM のパラメータによる視線方向 認識

高谷 学<sup>†</sup> 滝口 哲也<sup>††</sup> 有木 康雄<sup>††</sup>

<sup>†</sup> 神戸大学大学院工学研究科情報知能学専攻 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

<sup>††</sup> 神戸大学自然科学系先端融合研究環 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: <sup>†</sup>takatani\_\_m@me.cs.scitec.kobe-u.ac.jp, <sup>††</sup>{takigu,ariki}@kobe-u.ac.jp

あらまし 視線推定は、コンピュータビジョンの分野において重要な技術であり、インターフェースやロボット、人間の監視、支援など、幅広い応用が期待されている。近年では、比較的デバイスのコストも低く、被験者への負担の小さい単眼カメラによる視線推定手法が提案されている。本研究では、Active Appearance Models のパラメータを特徴量として抽出し、それらの特徴量と顔、視線方向の幾何的な関係をモデル化し、Gaussian Processes for Regression を用いて顔、視線方向の同時推定手法を提案する。実験の結果、提案手法は従来の線形回帰分析を用いた場合に比べ、悪条件下で撮影した被験者の視線方向精度が向上した。

キーワード 視線推定, Active Appearance Models, Gaussian Processes for Regression

## Simultaneous Estimation of Face and Gaze Orientation by Regression Analysis With Information Criterion

Manabu TAKATANI<sup>†</sup>, Tetsuya TAKIGUCHI<sup>††</sup>, and Yasuo ARIKI<sup>††</sup>

<sup>†</sup> Graduate School of Engineering, Kobe University Rokkodai 1-1, Nada-ku, Hyogo, 657-8501 Japan

<sup>††</sup> Organizatoin of Advanced Science and Technology, Kobe University Rokkodai 1-1, Nada-ku, Hyogo, 657-8501 Japan

E-mail: <sup>†</sup>takatani\_\_m@me.cs.scitec.kobe-u.ac.jp, <sup>††</sup>{takigu,ariki}@kobe-u.ac.jp

**Abstract** It is an important technique to estimate gaze orientation automatically in Computer Vision. The application using gaze estimation for interface, robot, and human interaction is expected. In recent years, many approaches for gaze estimation using only a monocular camera with low cost were proposed. In these approaches, the problem was the gaze orientation estimation error mainly caused by the face orientation estimation. To solve this problem, we propose simultaneous estimation of face and gaze orientation by GPR(Gaussian Processes for Regression) using AAM parameters. In the experimental result, the average estimation error of subjects in not good condition was mainly improved.

**Key words** Gaze estimation, Active Appearance Models, Regression analysis, Gaussian Processes for Regression

### 1. はじめに

人間は主に視覚から周囲の環境を知覚、認識する [1]。人間の視線を推定する技術はロボットビジョン、エージェント開発などの研究において非常に重要であり、ノンバーバルコミュニケーションを行うロボット [2] や、視線情報に基づく動画像解析 [3]、ドライバの眠気感知 [4] など、幅広いアプリケーションへの応用が期待される技術である。

視線推定の研究は近年盛んに行われており、これらの研究では高精度に視線を推定するために様々なアプローチが提案されてきた。これらのアプローチは大きく 2 種

類に分類できる。一つはカメラ以外に、赤外線カメラやヘッドマウントカメラ等の特殊なデバイスを用いて、視線を高精度に推定する手法である [6] [7]。さらに、もう一つのアプローチとして、単眼カメラ画像からの視線推定手法が挙げられる。前者のアプローチに比べ、後者は推定制度が低い、デバイスのコスト、被験者への負担という点においては優れており、近年様々な手法が提案されている [8] [9] [16]。本研究では、デバイスのコストや、人間への身体的負担の少ない単眼カメラによる視線の高精度な推定を目的とする。

従来の単眼カメラによる視線推定の手法として、黒眼移動量から視線を推定する手法 [5]、3 次元眼球モデルに

よる視線推定 [10] [11] などが挙げられる。これらの手法では、顔方向推定の後に視線方向推定を行うため、顔方向の推定誤差が視線方向の推定誤差に影響し、視線推定誤差が大きくなり、高精度な視線推定が行えないという問題がある。

そこで我々は、その問題を解決するため、Cootes らの提案する Active Appearance Models (以下 AAM と略す) [12] の combined パラメータ  $c$  を特徴量として抽出し、その特徴量と顔、視線方向の関係を幾何的な位置関係を基にモデル化する。次に、このモデルを基に回帰分析を行うことで、顔、視線方向を同時推定し、より高精度に視線推定を行う手法を提案した [24]。

AAM は顔、視線方向の学習画像列を与えることで、比較的次元の低いパラメータで顔、視線の方向変化に関する顔画像の形状とテクスチャの変化を制御するモデルである。そのため、顔の方向が変化しても特徴量を比較的頑健に抽出可能であるため、本研究では AAM を特徴量抽出の手法として採用した。

しかし、我々が以前提案した手法では線形回帰分析が用いられており、学習画像によって生成された AAM のパラメータと視線の関係が複雑で線形モデルで近似しにくいデータの場合、精度が低下するという問題があった。本研究では、その問題を解決するために線形な回帰分析ではなく、Gaussian Process Regression を用いて視線方向認識を行った。実験の結果、線形回帰分析を用いた手法に比べ、良好でない照明条件下で実験が行われた被験者について視線方向認識の精度が向上した。

本稿の構成は次の通りである。2章で提案手法の流れについて述べ、3章で Viola らの提案する顔領域探索手法と顔、視線方向同時推定のための特徴量を抽出する手法である AAM について述べる。4章では、得られた特徴量と、顔、視線の関係をモデル化する方法を提案する。さらに5章では以前我々が提案した手法と比較を行うための評価実験について述べる。

## 2. 提案手法の流れ

図1はこれまでに述べた提案手法における処理の流れの概略図である。提案手法では入力画像が与えられると、Viola の提案する手法 [14] により顔領域探索を行う。次に、構築しておいた AAM を用いて、入力画像と最も類似する画像を生成する AAM のパラメータを最急降下法により決定し、このパラメータを特徴量として抽出する。そして AIC によって選択された回帰式を用いて回帰分析を行い、図2のように視線と顔方向を同時に推定する。図2において緑色の矢印は視線方向、赤色の矢印は顔方向、矢印の大きさは角度の大きさを表している。

## 3. 特徴量抽出

入力画像から AAM のパラメータを決定する時、パラ

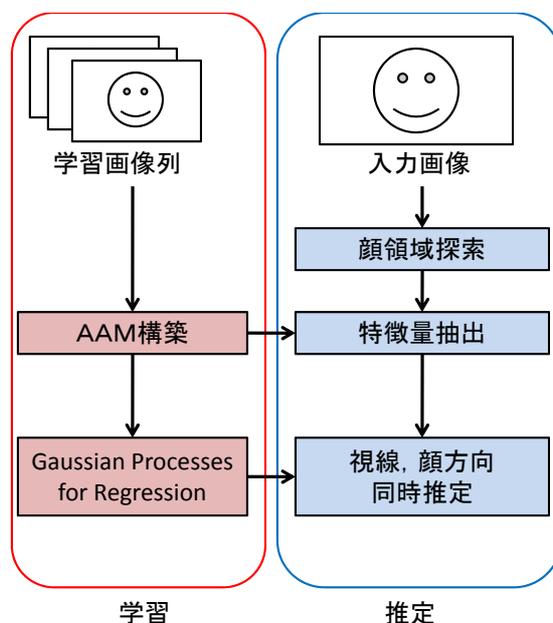


図1 提案手法における処理の流れ



図2 顔、視線方向同時推定の推定結果

メータの値を最急降下法によって変化させながら、入力画像とモデルのマッチングを繰り返し行って探索を行っている。従って、入力画像中における初期探索位置が、顔領域から大きく離れていると安定して AAM のパラメータを決定することができない。

本研究では安定して特徴量を抽出するために、Viola らの提案する顔領域探索手法 [14] を用いて顔領域を探索し、得られた顔領域を初期探索点として AAM のパラメータ探索を行っている。

本章では、Viola らの顔領域探索手法と、AAM によるモデル構築、及びパラメータの決定方法について述べる。

### 3.1 顔領域探索

AAM を用いて安定に特徴量を抽出するために、Viola らの提案する手法 [14] を用いて顔領域を探索する。この手法は顔領域探索において一般的に用いられている手法であり、安定して比較的高速に顔領域の探索が可能である [15]。

この手法では、図3に示す矩形領域を、図4のように当てはめ、白と黒の各矩形領域間の平均輝度の差を haar-like 特徴量として抽出し、顔を判別するのに有効な矩形の位置、種類、アスペクト比、スケールを弱識別機



図 3 Haar-like 特徴量抽出に用いられる矩形

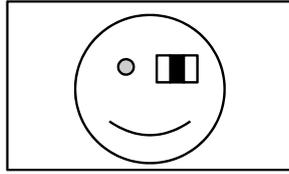


図 4 Haar-like 特徴量抽出のための矩形の当てはめ

として AdaBoost によって学習し、それらを線形結合して強識別機を構成し、顔領域を検出する。

### 3.2 Active Appearance Models

顔画像から特徴量を抽出する際、被験者が必ずしも正面を向いているとは限らない。Cootes らの提案する Active Appearance Models(AAM) [12] は、視線、顔方向の変化のある画像からモデルを構築するため、被験者の顔方向の変化に対して比較的精度良く特徴量を抽出できる。本研究では、AAM を用いて顔画像から特徴量の抽出を行う。以下では具体的に AAM について述べる。

AAM では、モデルの形状 (Shape) $s$  とテクスチャ (Texture) $g$  の変化を低次元のパラメータ  $c$  で表現する。 $s, g$  は式 (1), (2) で表現されるベクトルである。

$$s = (x_1, y_1, \dots, x_n, y_n)^T \quad (1)$$

$$g = (g_1, \dots, g_m)^T \quad (2)$$

ここで、 $x_i, y_i$  ( $i \leq n$ ) は各特徴点における座標を表している。学習画像に与えられたベクトル  $s$  を、回転、拡大縮小、位置の正規化を行い、学習画像列から平均形状  $\bar{s}$  を求める。 $g_j$  ( $j \leq m$ ) は、平均形状  $\bar{s}$  に画像を正規化したときの  $\bar{s}$  内部での各画素の輝度値であり、学習画像列から平均テクスチャ  $\bar{g}$  を求める

平均形状  $\bar{s}$ 、平均テクスチャ  $\bar{g}$  を用いて学習画像に対して主成分分析を行い、正規直交行列  $P_s, P_g$  を求め、得られた行列  $P_s, P_g$  を用いて、式 (3)(4) のように  $s, g$  をモデル化する。

$$s = \bar{s} + P_s b_s \quad (3)$$

$$g = \bar{g} + P_g b_g \quad (4)$$

ここで、 $b_s, b_g$  はそれぞれ主成分の係数であり、部分空間内で平均からの変化を表すパラメータである。さらに、 $b_s, b_g$  を式 (5) のように連結し  $b$  とする。

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} \quad (5)$$

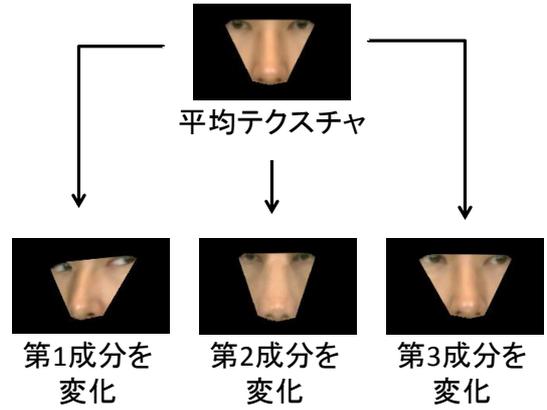


図 5 combined パラメータの各成分を変化させた時のモデル画像の変化 (第 1 成分は顔方向と視線方向の変化、第 2 成分は顔の縦の形状の変化、第 3 成分は視線方向の変化がそれぞれ表出している)

$W_s$  は形状とテクスチャの単位を正規化するための行列である。さらに、このベクトル  $b$  に対しても主成分分析を行い正規直交行列  $Q$  を求める。得られた行列  $Q$  を用いて  $b$  を式 (6) のように表現する。

$$b = Qc = \begin{pmatrix} Q_s \\ Q_g \end{pmatrix} c \quad (6)$$

$c$  は主成分の係数であり、combined パラメータベクトルと呼ばれる。式 (5) と式 (6) から式 (7) と式 (8) が得られる

$$W_s b_s = Q_s c \quad (7)$$

$$b_g = Q_g c \quad (8)$$

ここで、式 (7) と式 (8) より、形状  $s$  とテクスチャ  $g$  を  $c$  で表現すると式 (9)(10) のようになる。

$$s(c) = \bar{s} + P_s W_s^{-1} Q_s c \quad (9)$$

$$g(c) = \bar{g} + P_g Q_g c \quad (10)$$

このようにして、低次元のパラメータベクトル  $c$  によって、顔や視線の方向を変化させた顔画像の形状  $s$  とテクスチャ  $g$  をモデル化できる。

よって、 $c$  には顔、視線方向に関する情報が含まれていると考えられる。図 5 はパラメータベクトル  $c$  の各成分を変化させた場合に、顔画像がどのように変化するか、その様子を表したものである。本研究では、入力画像からパラメータベクトル  $c$  を抽出し、顔、視線方向の同時推定を行う。以下では、入力画像から特徴量を抽出する方法について述べる。

入力画像  $I$  に対して、モデル画像  $g(c)$  との誤差  $e$  を式 (11) のように求め、誤差を最小にするパラメータベクトル  $c$  及び、 $p$  を最急降下法によって求める。ここで、 $W$

は入力画像に対してアフィン変換を行う関数である．

$$e(\mathbf{c}, \mathbf{p}) = \|\mathbf{g}(\mathbf{c}) - I(\mathbf{W}(\mathbf{p}))\| \quad (11)$$

ただし， $\mathbf{p}$  は，入力画像をアフィン変換するための拡大，回転，平行移動のパラメータベクトルであり，poseパラメータベクトルと呼ばれる． $I(\mathbf{W}(\mathbf{u}; \mathbf{p}))$  は入力画像をパラメータ  $\mathbf{p}$  によってアフィン変換したときの画像である．

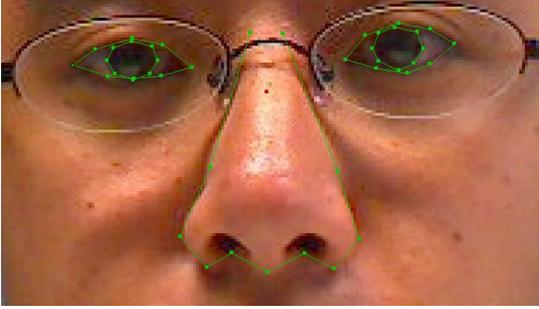


図 6 AAM 構築に用いた 43 点の特徴点 (緑色の点が特徴点)

このようにして，入力画像からパラメータベクトル  $\mathbf{c}$  を特徴量として抽出できる．本研究では，図 6 のように眼と鼻を中心とした 43 点の特徴点を用いて AAM を構築した．

#### 4. Gaussian Processes for Regression

Gaussian Processes for Regression(以下 GPR と略す) [25] は，説明変数ベクトル空間上でのデータ間の距離を利用して確率的に目的変数の値を推定する非線形な手法として近年注目を浴びている．線形でないデータに対して，有効であることからこの GPR を本研究では採用した．以下にその詳細を述べる．

GPR において，データ  $\mathbf{x}$  とデータ  $\bar{\mathbf{x}}$  間の距離は Covariance Function として以下のように表現される．

$$k(\mathbf{x}, \bar{\mathbf{x}}) = \sigma_f^2 \exp \frac{-(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T}{2l^2} + \sigma_n^2 \delta(\mathbf{x}, \bar{\mathbf{x}}) \quad (12)$$

ここで， $\delta(\cdot)$  はクロネッカーのデルタ関数である．また  $\sigma_f^2$  はデータ空間上での分散， $\sigma_n^2$  はデータの観測誤差を表現しており，これらは学習データを用いて値が決まる．

いま，学習画像から得られた  $N$  個の AAM の combined parameter である  $\mathbf{c}_1, \dots, \mathbf{c}_N$  と正解の視線方向のデータ  $\mathbf{y} = y_1, \dots, y_N$ 、そしてテストデータの combined parameter  $\mathbf{c}_*$  を観測したとき，テストデータの視線方向  $y_*$  を推定する場合，GPR では， $[\mathbf{y}, y_*]^T$  のデータが以下のような正規分布に従うと仮定する．

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathbf{N}(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}) \quad (13)$$

ただし，

$$K = \begin{bmatrix} k(\mathbf{c}_1, \mathbf{c}_1) & \dots & k(\mathbf{c}_1, \mathbf{c}_N) \\ \vdots & & \vdots \\ k(\mathbf{c}_N, \mathbf{c}_1) & \dots & k(\mathbf{c}_N, \mathbf{c}_N) \end{bmatrix} \quad (14)$$

$$K_* = \begin{bmatrix} k(\mathbf{c}_*, \mathbf{c}_1) & \dots & k(\mathbf{c}_*, \mathbf{c}_N) \end{bmatrix} \quad (15)$$

$$K_{**} = k(\mathbf{c}_*, \mathbf{c}_N) \quad (16)$$

である．このとき， $p(y_* | \mathbf{y})$  は以下のような正規分布で表される．

$$y_* | \mathbf{y} \sim \mathbf{N}(K_* K^{-1} \mathbf{y}, K_{**} - K_* K^{-1} K_*^T) \quad (17)$$

よって，尤も良いとされる  $y_*$  の推定値は以下になる．

$$\mathbf{y}_{\text{est}} = K_* K^{-1} \mathbf{y} \quad (18)$$

本研究では，このようにして得られた  $\mathbf{y}$  を推定値として出力する．

#### 5. 評価実験

線形回帰分析を用いた場合より，本研究の提案手法が AAM を用いた視線方向認識において有効であることを示すために評価実験を行った．実験に用いたデータベースの概要を表 1 に示す．AAM と回帰分析の学習には 63 枚の学習画像が用いられる．被験者数は 4 名だが，2 名は眼鏡を着用しており，照明条件が良好でない環境下で撮影を行っている．

表 1 実験に用いたデータベースの概要

被験者数	4 名
学習画像枚数	63 枚 / 人
テスト画像枚数	252 枚 / 人
画像の解像度	640 × 480
被験者の顔角度	-20, -15, ..., 15, 20 (deg)
被験者の視線角度	-20, -15, ..., 15, 20 (deg)

AAM の構築に一人の人間の画像を使用し，学習人物と同一人物の画像を入力画像として与えなければ，AAM による特徴量抽出の精度が低下する [22]．よって実験では各個人のデータに対してそれぞれ AAM を構築し，学習人物と同一の人物の画像を入力画像として与え，実験を行った．

比較する従来手法として，Ishikawa らによって提案されている手法 [16] と我々が以前提案した線形回帰分析による手法を採用し，これらの手法と比較を行った．この手法では，従来の単眼カメラにおける視線推定において，本実験と同一条件で実験を行った場合，最も精度が高いと考えられる．

実験の結果を図 7 に示す。図は手法による平均誤差の違いである。図中の AIC, MDL, BIC など赤色で示される部分の結果は我々が以前提案した手法の結果である。そして GPR と書かれてある青色の部分今回提案した手法の結果である。この結果から全体として平均誤差が 0.2 度ほど改善されている。

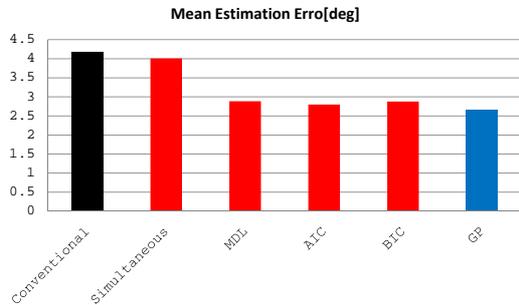


図 7 手法による平均誤差の結果の違い

さらに図 8 は図 7 の実験をさらに被験者ごとに結果の違いを分析したものである。この図中の被験者 m02 と m04 は眼鏡を着用し、さらに目領域の判別が困難な照明状況下で撮影された被験者である。この結果を見てわかるように、非線形な回帰分析を用いることで複雑な構造のデータにも対応でき、照明条件の悪い被験者について精度が約 1 度と格段に向上している。

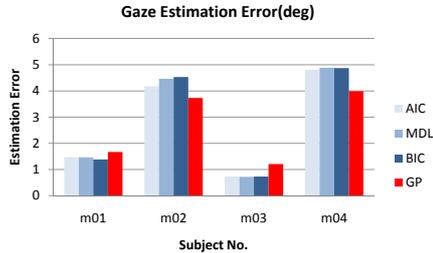


図 8 被験者の違いによる各手法の結果の違い

## 6. 結 論

本研究では、Active Appearance Models によって、特徴量を抽出し、抽出した特徴量を Gaussian Processes for Regression を用いて視線方向の認識を行った。線形回帰分析を用いる場合に比べ、良好でない環境下で撮影した被験者について認識精度が向上した。しかし、Active Appearance Models が学習に用いてない人物をテストデータとして与えた場合、特徴点を高精度に追跡できない問題があるため、今後はその問題を解決する必要がある。

- [1] Mehrabian Albert, "Silent Message: Implicit Communication of Emotions and Attitudes," Wadsworth Pub Co, 1980
- [2] Tomoko Yonezawa, Hirotake Yamazoe, "Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze tracking," Proceedings of the 9th international conference on Multimodal interfaces, pp.140-145, 2007
- [3] Derec Pang, Akisato Kimura, Tatsuto Takeuchi, Junji Yamamoto, Kunio Kashino, "A STOCHASTIC MODEL OF SELECTIVE VISUAL ATTENTION WITH A DYNAMIC BAYESIAN NETWORK," Proc, International Conference on Multimedia and Expo, pp.1073-1076, June, 2008
- [4] 白井了, 井東通昌, 乗松有, 足立和正, 中野倫明, 山本新, "ドライバの運転状態検知のための視線方向の検出", 電子情報通信学会技術研究報告 HIP ヒューマン情報処理, Vol.103, No.455, pp.67-72
- [5] 北川洋介, 加藤丈和, 呉海元, 和田俊和, "アイモデルを用いた視線推定のための黒目追跡," MIRU, 2005
- [6] Takehiko Ohno, Naoki Mukawa, Atsushi Yoshikawa, "FreeGaze: a gaze tracking system for everyday gaze interaction", Proceedings of the 2002 symposium on Eye tracking research & applications, pp.125-132, 2002
- [7] C.H.Morimoto, "Eye gaze tracking techniques for interactive applications," Computer Vision and Image Understanding, Vol.98, Number.1, pp.4-24, 2005
- [8] 三宅哲夫, 今村孝, 章忠, "単眼カメラ画像を用いた視線計測", MSRC, 2007
- [9] 小野泰弘, 岡部孝弘, 佐藤洋一, "低解像度画像からの視線方向推定", MIRU 2006
- [10] Hirotake Yamazoe, Akira Utsumi, Tomoko Yonezawa, Shinji Abe, "Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions," Proceedings of the 2008 symposium on Eye tracking research and applications, pp.245-250, 2008
- [11] Takahiro Ishikawa, Simon Baker, Iain Matthews, and Takeo Kanade, "Passive Driver Gaze Tracking with Active Appearance Models," CMU, 2004
- [12] T.F.Cootes, G.J.Edwards, C.J.Taylor, "Active Appearance Models", European Conference on Computer Vision 1998
- [13] Akaike H, "A new look at the statistical model identification" Automatic Control, IEEE Transactions Vol.19 p.716-p.723, 1974
- [14] Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Feature", Fourth IEEE International Conference pp.40-45 (2000)
- [15] 林伸治, 長谷川修, "低解像度画像からの顔検出", 画像電子学会誌, 第 34 巻, 2005
- [16] Takahiro Ishikawa, Simon Baker, Iain Matthews, Takeo Kanade, "Passive Driver Gaze Tracking with Active Appearance Models", CMU, 2004
- [17] T.F.Cootes, K Walker, C.J.Taylor, "view-based Active Appearance Models", Automatic Face and Gesture Recognition, 2000
- [18] Peter Huisman, Ruud van Munster, Stephanie Moro-Ellenberger, Raymond Veldhuis, Asker Bazen, "Making 2D face recognition more robust using AAMs for pose compensation", Automatic Face and Gesture Recognition, Vol.10-12, April, 2006
- [19] Hiroshi YAMAGISHI, Kenji AMAYA, "EFFICIENT BOUNDARY ELEMENT METHOD BY MULTI-POLE EXPANSION AND GENERALIZED IN-

VERSE MATRIX FOR ANALYSING TARGET REGION ”, 計算数理工学論文集, Vol8, No.03-081128

- [20] J. Rissanen, “ An Introduction to the MDL Principle ”, Available online at [www.mdl-research.org](http://www.mdl-research.org)
- [21] 下平英寿, 伊藤秀一, 久保川達也, 竹内啓, “ 統計科学のフロンティア3 モデル選択 予測・検定・推定の交差点 ”, 岩波書店, pp.3-26, 2005
- [22] 小林亮博, 佐竹純次, 平山高嗣, 川島宏彰, 松山隆司 “ AAM の動的選択に基づく不特定人物の顔追跡 ”, 情報処理学会研究報告 (CVIM-161), pp.35-40, 2008
- [23] Yusuke SAKASHITA, Hironobu FUJIYOSHI, Yutaka HIRATA “ Measurement of 3D Eye Movements Using Image Processing ”, 実験力学, Vol.6, No.3, September, 2006
- [24] 高谷学, 滝口哲也, 有木康雄, “過学習を考慮した AAM パラメータの選択と回帰分析による顔・視線方向同時推定”, 画像の理解・認識シンポジウム (2009), IS1-60
- [25] M.Ebden “Gaussian Processes for Regression: A Quick Introduction”, Available on Internet “<http://www.robots.ox.ac.uk/~mebden/reports/GPtutorial.pdf>”