

# 画像を用いた人の選択行動の興味度合推定

Estimation of Human Interest Level in Choosing from Video Sequence

若井 祐介 †, 鷲見 和彦 †, 松山 隆司 †

Yusuke WAKAI†, Kazuhiko SUMI † and Takashi MATSUYAMA †

†:京都大学大学院情報学研究科知能情報学専攻

{wakai,sumi,tm}@vision.kuee.kyoto-u.ac.jp

2005 年 10 月 11 日

概要：人が対象を選ぶ状況で、個々の対象に対する興味の変化を定量的に画像計測する手法を提案する。ある対象に対する興味が高まると、近付いて、その対象を凝視するという性質を利用し、商品側に置かれたカメラで計測した人物像から体、顔、視線を抽出し、その変動を認識することで、興味の度合の変化が抽出できることを実例を用いて示す。

## 1 研究の背景と目的

言語、音声、画像を対象としたパターン情報処理とコンピュータビジョンの研究は、1970年代から着実な進歩を遂げてきた。従来のパターン認識技術は、主として二次元や三次元の物体を対象とし、カテゴリの識別・形状や属性の計測とモデル化を目的としたものが多い。人の内面的な状態を定量化しようとしたものとしては、表情認識において、怒喜哀楽など極端な表情を識別する試みは見られるが、日常のコミュニケーションで見られる微妙な感情の表現を認識するには至っていない。今後の計算機システムやロボットは、決められた動作を確実に行うためにセンサ情報を用いる知能ロボットを越えて、人とコミュニケーションをとりながら、人を助け、人の生活を豊かにする共生的パートナー [1] になることが期待されている。そのためには人の内面的な状態を推定しながら、それに応じて自律的に最適な行動をとるシステムを実現しなければならない。

では、コミュニケーションにおいて必要な人の内面状態の推定とは何であろうか。人対人のコミュニケーションにおいては、まず第一に相手に話を聞いてもらうことが必要である。話し手が話しかける際に、相手が聞いてくれるか拒否しそうかを判断し、好意的に聞いてくれるように工夫したり、話しかける

間合いを取ることがごく自然に行われている。また、話し手が何らかの意見や情報を相手に伝えたいとき、話題に対して、肯定的か否定的かを見極めて、相手が納得してくれるように話を誘導して行かねばならない。これらに共通な概念として、完全な受け入れから拒否まで段階的に変化するいくつかの感情の軸が存在する。それらは、たとえば次のようなものである。肯定的 — 否定的, 賛成 — 反対, 好感 — 反感, 興味がある — 興味がない。

これらの個別の感情の軸を互いに分離することは大変難しいが、例えば、「何かの行動を相手に説得してさせる」というような個々のシナリオの中では、肯定的か否定的かを識別することは我々人間にとっては容易である。

そこで、この研究ではコンピュータが観察によって聞き手の興味の度合いを数値化することを試みる。このような状況は日常生活において数多く存在するが、本研究では、その一つの事例として、多くの選択肢の中から欲しいものを選んでいいる人に対して、それらを比較しながら商品の説明を行っているというシナリオについて検討した。

従来、印象や好感度を定量化するためには、被験者に質問を行い、被験者の主観的感覚を何段階かで回答してもらう手法 [2] が広く用いられているが、現在進行中のコミュニケーションでは、質問を行うこと

なく、相手の様子や挙動からリアルタイムに、かつ、受動的に読み取ることが必要である。そこで、テレビカメラを複数用いて、被験者の映像を撮影しながら、映像から肯定的な要素と否定的な要素を抽出して定量化することを試みた。

以下、第2章では、人の内部状態や興味の度合いを推定するための関連研究について述べ、第3章では、本研究の提案である対象への近付き度合いと、視線の凝視度合いとを用いた、興味の定量化手法を説明する。第4章で実験について説明と考察を行い、第5章で結果と考察を示し、第6章で結論と今後の展望について述べる。

## 2 関連研究と本研究の主張

MIT Media Lab で Affective computing と名付けられた人の感情を定量化する研究において Mota らは椅子の座面と背面に取りつけた圧力センサーの圧力分布を入力とし、隠れマルコフモデル (HMM) を用いて、座学する被験者の興味の度合いを3つの状態に分類した [3]。この研究は人の興味の度合いを定量化したものとして価値が高いが、使用したセンサーは人と接触・装着するものであり、自由姿勢の人を非拘束・非接触で計測するには至っていない。

非接触で人の興味や意図を計ろうとする場合、まず第一に視線情報の利用が考えられる。大野は視線を使ったインタラクションの研究 [5] において、コンピュータが人の視線を精度良く計測し、それをインタフェースとして利用するアイデアを示した。しかし、視線だけでは、漫然と眺めていることと、食い入るように見ているという、熱心さの違いを定量化しにくいのが欠点である。山本らは注目した直後に瞬目の頻度が高まることに注目し、視聴映像に対する注目度の推定を行った [4]。しかし、人が注目する動作を終了した時に、初めて直前の状態がわかるため、完全なリアルタイム性が実現できていなかった。

一方、視線を用いない興味や集中の度合いの計測として、田中らは人の動作や姿勢を用いることを提案している [6]。人が完全にリラックスした状態を初期値として熱心に作業を行う場合には全身に身構える姿勢が現れることを自己負荷量という姿勢エネルギー消費により表現したが、作業対象は単一であり、複数の候補から興味の対象と度合いを識別・定量化するには至っていなかった。

そこで本研究では、非拘束・非接触なセンサ情報から人の興味の対象とその度合いとの両方を計測する手法を提案する。これを実現するために、複数のカメラを用いたシステムで人の上半身と顔を撮影し、映像から視線と姿勢との両方を計測して、それらの値を統合する興味の計測手法を考案した。人の視線が見ている対象が興味の候補であり、かつ、それを継続して凝視しているか、また、より良く見るためににじり寄って見ているかで、興味の程度を推定する。

## 3 興味の定量化

本研究において、人の興味は、顔の向きや視線という興味の選択を示すしぐさと、体の動きや姿勢という興味の強さを示すしぐさとの両方に現れると考えた。本章では、まず、視線と体の動きや姿勢との個別の計測とその解釈について述べ、ついで、それらの統合について説明する。

### 3.1 視線計測から得られる興味

人の視線は平均約 250ms の固視 (fixation) や追従 (tracking) と約 10ms の跳躍 (saccade) の繰り返しからなることがわかっている [7, 8]。ただし、これだけでは視線と興味の関係を定量化することは難しい。戸田らは、絵の情報を記憶するタスクにおける人の注視状態を観察し、視線の状態を注目対象模索状態と詳細情報獲得状態とに分類した [9]。詳細情報獲得状態とは、人がある対象を見つけ、その対象をより良く知ろうとしている状態であり、ほぼ一定の領域に視線が約 1.2 ~ 1.5s 以上滞留していることが特徴である。逆に、視線が一定の領域にとどまっても、その時間が 1.2s より短ければ、強い興味が発生しているかどうかを確認することはできない。

一方、滞留時間が長くなった場合、たとえば、全く刺激の入らない状態で何かに注目していれば、滞留時間と興味の程度とは比例するかも知れないが、コミュニケーションにおいては画像や音声など次々と新しい情報が提供されるので、興味の持続は新しい情報の提供によって更新されると考えた方がよい。

そこで、本研究では、視線の滞留時間と興味の強さとの間に図1のような関係があると仮定した。ここで、時刻  $t_1$  は、興味の発生を観測できる遅れ時間であり 1.2s とした。時刻  $t_2$  は、興味が飽和する(あ

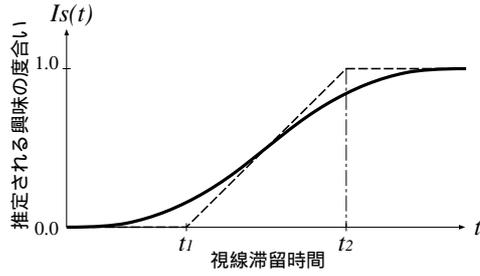


図 1: 視線の滞留時間と興味の強さとの関係

るいは情報の新提示によって興味が更新される) 時刻であり, 本研究では実験の設計によって決定される. これらの時刻をパラメータとして,  $t_1$  から  $t_2$  にかけてなめらかに興味が上昇するよう式 1 で視線から計測される興味の強さ  $I_s(t)$  の瞬時値を表した.

$$I_s(t) = \frac{1}{1 + \exp(-\alpha t + \beta)} \quad (1)$$

ただし,  $\alpha, \beta$  は, それぞれ時刻  $t_1, t_2$  から算出され,  $\alpha = 4/(t_2 - t_1)$ ,  $\beta = 2(t_2 + t_1)/(t_2 - t_1)$  である. 瞬時値を  $\Delta t$  の区間で平均し, 時刻ごとの平均興味  $\bar{I}_s(t)$  を算出する.

$$\bar{I}_s(t) = \int_{t-\Delta t}^t \frac{1}{1 + \exp(-\alpha t + \beta)} dt \quad (2)$$

### 3.2 動きと姿勢から得られる興味

一方, 興味の度合いが高まってくると, 対象をより良く見たい・知りたいという欲求によって, 対象へ体が引き付けられる現象が観測される. 観測される人の視点から見ると, 図 2 上段に示されるように, 人が近づくにつれ視野内での対象物体の占める面積  $S$  が増大して行き, それにつれて相対的な注意が上昇して行く. この, 状況を興味の対象である物体側から眺めると, 図 2 下段に示されるように, 人の顔や上半身の面積  $F$  が同じ比率で増大する. 人と対象との距離が  $d_n$  の時, 人の視野 (焦点距離  $f$ ) に入る対象の高さ, 幅, 対象の面積をそれぞれ  $h, w$  とし, カメラ (焦点距離  $f'$ ) が視野内の人の高さ, 幅, 面積をそれぞれ  $h', w'$  とすれば,

$$S \propto f^2 hw / d_n^2 \quad (3)$$

$$F \propto f'^2 h' w' / d_n^2 \quad (4)$$

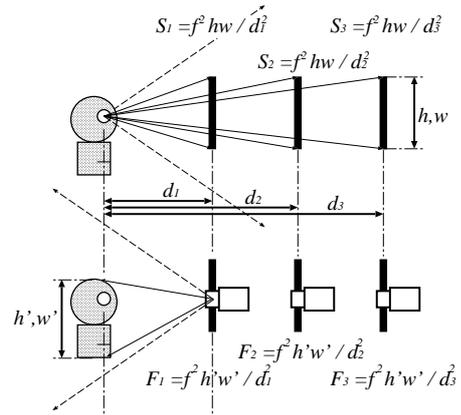


図 2: 人と対象物体の距離と人から見た対象の大きさの関係 (上段) および対象物体から見た人の大きさの関係 (下段).

であるから,  $S \propto F$  である. そこで本研究においては, 興味の対象候補となるすべての物体のそばにカメラを置き, カメラごとに時刻  $t$  における面積  $F(t)$  を計測して, 計測値と実験開始時の値  $F_0$  との比を人の姿勢から得られる興味の強さ  $I_p(t)$  とした. すなわち, 次式のように表せる.

$$I_p(t) = F(t)/F_0 \quad (5)$$

### 3.3 興味の度合いの統合

これまで述べた視線による興味  $\bar{I}_s(t)$  には, 滞留持続時間による興味の大きさという要因を持っているが, 山本 [4] らが指摘するように視線の滞留は意識の強さを示す必要条件でありながら十分条件ではない. 一方, 対象物体の距離や姿勢による興味  $I_p(t)$  には, となりの興味候補と見分けるだけの空間分解能がない. よって, これらの情報を相互に独立であると見なし, その積によって, 対象に対する興味  $I(t)$  を定義する.

$$I(t) = \bar{I}_s(t) I_p(t) \quad (6)$$

## 4 画像処理による実装方法と実験

### 4.1 視線計測アルゴリズムの実装

本章では, 画像から視線抽出と人物シルエット抽出を行い興味  $I_s(t)$ ,  $I_p(t)$  とを求める方法, および,

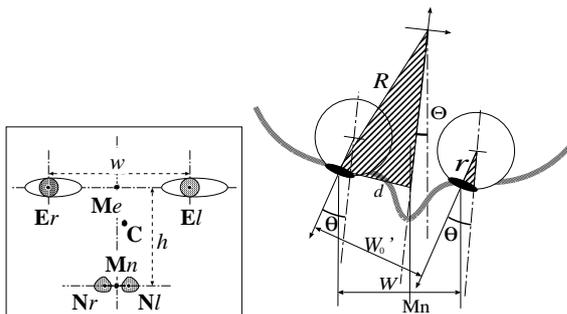


図 3: 顔座標の定義，正面図(左)と平面図(右)

それを用いた人の興味の計測実験とについて述べる。

#### 4.1.1 座標の定義

視線計測のために，まず，顔の位置を Leinhart の顔検出器 [10] で検出し，その位置を基準に林らの手法 [11] により瞳の位置と鼻の穴からなる 4 点の顔特徴点を検出した．これによって検出される瞳と鼻の穴の位置をそれぞれ  $E_r(x_{er}, y_{er})$ ,  $E_l(x_{el}, y_{el})$ ,  $N_r(x_{nr}, y_{nr})$ ,  $N_l(x_{nl}, y_{nl})$  とし，顔検出器が出力する顔の中心座標を  $C(c_x, c_y)$  とする．顔の視線方向を軸とする回転を考慮しない場合，両目の中点を  $M_e$ ，両鼻の中点を  $M_n$ ， $M_e$  から  $M_n$  までの高さを  $h$ ，両眼の幅を  $w$  とすれば，

$$w = x_{er} - x_{el} \quad (7)$$

$$h = \frac{1}{2}((y_{er} + y_{el}) - (y_{nr} + y_{nl})) \quad (8)$$

である． $w, h$  の正面時の値をそれぞれ  $w_0, h_0$  とする．

本来，顔の回転は，垂直軸廻りの横回転，水平軸廻りの縦回転，視線軸廻りの回転の 3 軸があるが，本研究では実験条件を工夫して被験者が縦回転と視線軸廻りの回転が現れにくいように工夫し，横回転のみを考慮した．

この実験では顔の縦回転を考慮しなくて良いので， $h$  の変化は距離の変化または顔の大きさの変化であると考え， $h$  で  $w$  を正規化し，最終的に顔の横回転角  $\Phi$  を求める．両目の視線方向成分の間の長さを  $w'_0$  とおいている．

$$\Phi = \cos^{-1}\left(\frac{w h_0}{h w'_0}\right) \quad (9)$$

最後に瞳の動きより視線方向を決定する．図 3 より，斜線部分の 2 つの三角形は相似の関係となり，そ

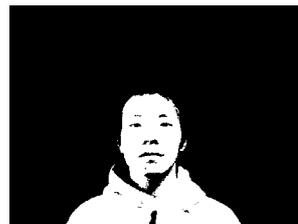


図 4: 背景差分を用いたシルエットの例

の比率関係より眼球の回転角  $\theta$  を求める．顔を円に近似した時の半径を  $R$ ，眼球を円に近似した時の半径を  $r$  とおいている．

$$\theta = \sin^{-1}\left(\frac{d}{R}\right) \quad (10)$$

$$d = x_{el} - \frac{x_{nl} + x_{nr}}{2} \quad (11)$$

それぞれのカメラは被験者に対して同距離にして 30 度の間隔をとる．対象からみて  $-15^\circ \sim +15^\circ$  に視線が検出された場合に「見ている」と判別する．

#### 4.2 シルエット抽出の実装

空間的占有度を求めるアルゴリズムとしては，実験開始直前の画像を背景とする背景差分を用いている．しきい値は固定であり図 4 のようなシルエットが検出される．実験開始時のシルエットの面積を  $F_0$  とし，時刻  $t$  における  $F(t)$  から式 5 により興味  $I_p(t)$  を推定する．

#### 4.3 実験の内容

以上述べたように実装した興味の計測手法が，人の変化する興味がどの程度正確に計測できるかを実証するため，次のような実験を行った．被験者は，図 5 に示されるように着席して，前面のパネルに張り出された 3 つの旅行広告を見ながら，3 分間で好きな対象を一つ選ぶことを指示される．左，正面，右の対象とカメラとを区別するために，それぞれ L, F, R と呼ぶことにした．3 分間の実験の間に，参考情報として 6 回の音声情報が流れる．音声情報は 20(s) 間隔で提示され，それぞれある一つの旅行に有利な情報または不利な情報が含まれている．与えた音声情報の例を図 6 に示す．実験中の被験者の顔と上半

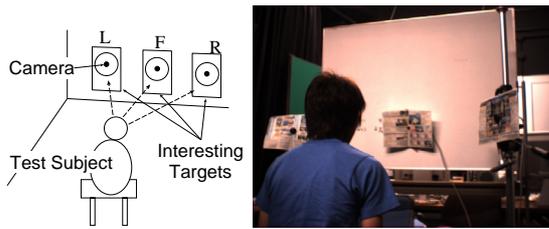


図 5: 被験者・対象・カメラの位置関係 (左) と実験風景 (右)

あなたはボーナスをもらいました。  
夏休み3つどののバック旅行をえらびますか？

- 沖縄
  - きれいな海でバカンス
  - ゴルフ、スキューバなど多彩なスポーツ
  - + 近くの好きな島で遊べる
  - スキューバはサメが出たため取りやめ中
- 北海道
  - 最高気温が30°Cに達しない涼しい気候
  - 函館や小樽、知床といった観光名所
  - + 札幌は計画都市なので広々としてきれいな街並み
  - 観光名所全てを回するには1週間近くの日数を要する
- 台湾
  - 古都の遺跡など観光スポットは多彩
  - 中華料理などの台湾グルメを堪能できる
  - + 足つぼマッサージで日頃の疲れを癒せる
  - 京都と変わらず夏は暑い

図 6: 与えられる情報の例

身の画像は、図中黒丸印に置かれたカメラで観察されている。実験終了後、被験者と実験主催者はビデオ映像を再生して、時々刻々の(どの対象にどの程度興味があったかという)被験者の心境を記録する。ただし、被験者が自己申告できる興味の変化の時間的解像度は粗いため、実験主催者が観察による主観を加えて時間解像度を高めた。5人の被験者に対して1人ずつ行った。

## 5 結果と考察

図7に興味の推定値  $\bar{I}_s(t)$  (Y軸) と  $I_p(t)$  (Z軸) の変化を示す。図において、5人のうち代表者1人分のFの結果であり、縦向き矢印は音声情報を与えた時刻を示している。音声情報が与えられると、 $\bar{I}_s$  と  $I_p$  が大きく変動しており、与えた情報によって被験者の挙動が影響されたことを示している。

次に、 $\bar{I}_s$  と  $I_p$  を統合して得られた興味の推定値  $I$  と、被験者の申告を元に観測者が時間的に補完した興味の度合いとの関係を調べた。図8は、横軸に5段階の申告に基づく興味の度合い、縦軸に本手法に

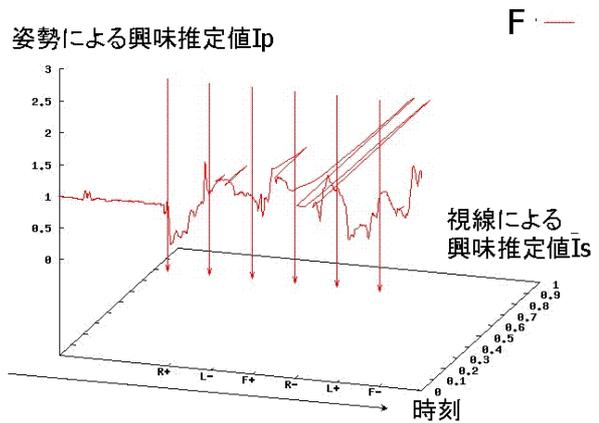


図 7: 実験によって得られた視線による興味  $\bar{I}_s(t)$  と、姿勢による興味  $I_p(t)$  の変化

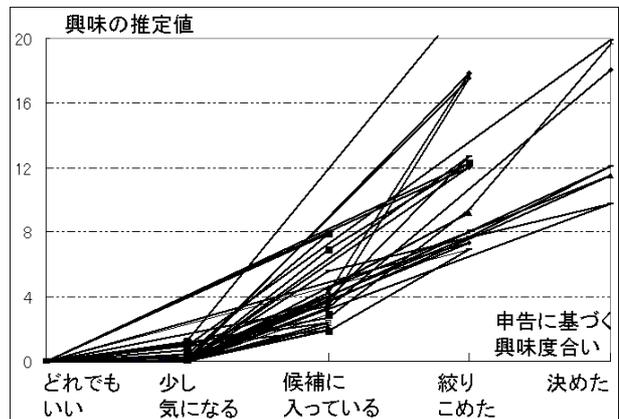


図 8: 視線と姿勢とを統合した興味推定値と、申告に基づく興味度合いとの関係

よる推定値  $I$  を、全被験者5人分プロットして隣接する時刻ごとに線分で結んだものである。

興味の表現の個人差があるため本手法で興味の度合いを絶対値として得ることは難しい。しかし、全データで、カメラ計測による興味の度合いと自己申告による興味の度合いとに強い相関が見られ、かつ、すべての線分が右上がりであるので、人が感じた興味の変化と、本手法によって検出された興味の変化とは、符号が常に一致することが確認できた。

一方、図8と視線だけから得られた興味の度合い  $\bar{I}_s$  をプロットした図9とを比較すると、両者は同じ傾向が見られるが、図9は線分の傾きのばらつきが大きく、一部に、水平/垂直の線分が観測される。これは、視線計測だけでは、必ずしも人の感覚とは一

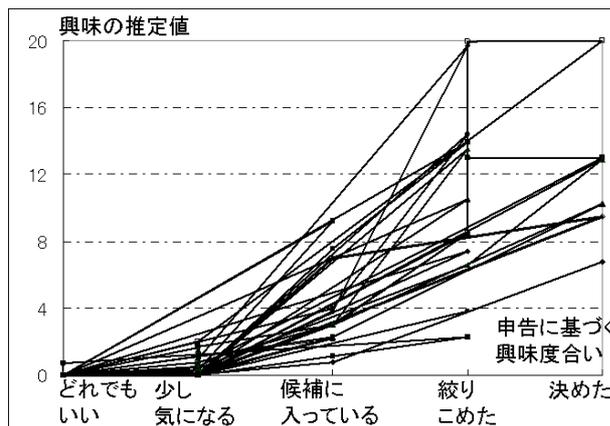


図 9: 視線のみから得られた興味推定値と、申告に基づく興味度合いとの関係

致しないことを示しており、本研究で視線と姿勢との観察結果を統合したことの有効性を示している。

## 6 結論

視線と姿勢とを観察することで、人が、どの対象にどの程度興味を持っているかを推定する手法を提案した。表現の個人差が大きいため興味の絶対値計測はできなかったが、音声情報を与えながら対象を選ぶ状態で、興味の変化方向、すなわち肯定的に変化したか否定的に変化したかを正しく検出できることが確認された。なお、本研究では顔検出と視線検出のアルゴリズムに入手しやすいアルゴリズムを利用したため、顔の回転軸に制約があったり、人が動き回れる空間が小さいなどの実装上の課題が多く残されている。また、姿勢検出についても、近寄りを面積で表現するなど簡略化してしまった部分も多い。今後、実装を改めて実用性を向上すると共に、視線と位置・姿勢だけでない細かな変化を読み取れる表現特徴を見つけ出すことが重要である。今回は繰り返し評価のために画像を一旦ファイルに記録してオフラインで処理したが、利用した顔検出・視線検出・背景差分はリアルタイム動作の実績があり、今後、実装を進めてリアルタイム処理を実現する予定である。

謝辞 本稿の一部は 21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」によるものである。また、本研究の遂行に当っては科学研究費補助金特定領域研究 13224051 の支援を受けた。

## 参考文献

- [1] 松山隆司, 杉本晃宏, 佐藤洋一, 川嶋宏彰: 人間と共生する情報システムの実現を目指して, 人工知能学会誌, Vol.19, No.2, pp.257-266, 2004.3
- [2] 松下豊彦著: SD 法によるイメージの測定, 川嶋書店, 1983
- [3] Selene Mota and Rosalind W. Picard: Automated Posture Analysis For Detecting Learner's Affective State Level, IEEE CVPR HCI 2003, Media Lab. Technical Report No.574, 2003
- [4] 山本哲也: 画像センシング情報を用いた視聴映像に対する注目度推定法, 情処研報 CVIM, No.144, pp.259-266, 2004
- [5] 大野健彦: 非言語情報に基づく Computer Mediated Communication: 情処研報 CVIM, No.150, pp.25-32, 2005
- [6] 鷲見 和彦, 田中 宏一, 松山 隆司: 三次元姿勢計測を用いた人の動作特徴の記述, 画像の認識理解シンポジウム MIRU2004, vol.1, pp.660-665, 2004.7
- [7] 池田光男: 目は何を見ているか-視覚系の情報処理, 平凡社, 1988
- [8] 乾敏郎: 視覚情報処理の基礎, サイエンス社, 1991
- [9] 戸田 真人, 鷲見 和彦, 松山 隆司: 装着型視覚センサを用いた注目対象映像の獲得と理解, 画像の認識理解シンポジウム MIRU2005, IS4-150, pp.1388-1395, 2005
- [10] Rainer Lienhart, Jochen Maydt: An Extended Set of Haar-like Features for Rapid Object Detection, IEEE ICIP 2002, Vol.1, pp.900-903, 2002
- [11] 林 健太郎, 橋本 学, 鷲見 和彦: 頑健性と精緻性を備えた顔特徴点追跡による顔方向推定, 電子情報通信学会論文誌 D-II Vol.J84-DII, No.8, pp.1762-1771, 2001.8

若井 祐介: '04 年北海道大学工学部電子工学科卒業, 京都大学大学院情報学研究科修士課程知能情報学専攻在籍, 卒業研究は遺伝的アルゴリズムの開発, 現在は人と共生するシステムの研究に従事。

鷲見 和彦: '84 同大学院修士課程了。'84 三菱電機株式会社を経て'03 より京大情報学研究科 COE 研究員・客員教授。専門は画像の認識理解の研究とロボット・ITS・セキュリティ・生体個人認証への応用研究。現在は観察に基づく人の状態の推定とヒューマンインタフェースへの応用に興味。'98 ロボット学会実用化技術賞, '03 画像センシングシンポジウム論文賞受賞。

松山 隆司: '76 年京大大学院修士課程了。京大助手, 東北大学助教授, 岡山大教授を経て'95 年より京大大学院電子通信工学専攻教授。現在同大学院情報学研究科知能情報学専攻教授。2002 年学術情報メディアセンター長, 評議員。工博。画像理解, 人工知能, 分散協調視覚の研究に従事。1980 年情報処理学会創立 20 周年記念論文賞, 1990 年人工知能学会論文賞, 1993 年情報処理学会論文賞, 1994 年電子情報通信学会論文賞, 1995 年第 5 回 ICCV Marr Prize, 1996 年 IAPR Fellow, 1999 年電子情報通信学会論文賞, 2000 年画像センシングシンポジウム優秀論文賞