

佐藤 洋一

東京大学生産技術研究所・教授

日常生活空間における人の注視の推定と誘導による情報支援基盤の構築

## §1. 研究実施の概要

人と調和する情報環境を実現するためには、情報環境側が人の注意が何に向けられているのかを理解したうえで、適切なタイミングで適切な支援を提供できることが必要となる。本研究では、人の注意と密接に関係する注視に着目し、人と調和する情報環境実現のための基盤技術として、日常生活空間内における人の注視を推定する技術、および情報環境からの適切な働きかけにより人の注視をさりげなく誘導する技術の開発を目指している。今年度は、注視推定技術に関しては、低解像度画像からの顔向き推定に関して問題となる学習画像の獲得の困難さに対して、歩行者の進行方向を顔向きのラベルとして利用することにより自動的に学習を行う手法を提案した。さらに、視覚的顕著性を手掛かりとした視線推定について手法の改良による精度向上を図ったほか、一人称視点映像における視覚的顕著性マップモデルの働きの検証と一人称視点に適した顕著性マップモデルの提案を行った。また、注視誘導技術の開発に関しては、ロボットの身体動作による注意誘導のためのロボット頭部デザインの検討・開発を行い、大規模な被験者実験を行った。ここで得られた知見をもとに、注視獲得および状況認識手法の開発を継続していく。さらに、注視誘導のための情報提示システムを実現するために、注視方向により情報操作を行うディスプレイシステムを開発した。より自然な情報提示を実現するために、スクリーン上の注視誘導技術の開発を継続していく予定である。

## §2. 研究実施体制

### (1)「佐藤」グループ

- ① 研究分担グループ長:佐藤 洋一 (東京大学生産技術研究所、教授)
- ② 研究項目
  - 環境埋め込みカメラによる視線計測技術の開発
  - 不正確で不安定な視線情報からの注視推定技術の開発

### (2)「杉本」グループ

- ① 研究分担グループ長:杉本 晃宏 (国立情報学研究、教授)
- ② 研究項目
  - 環境埋め込みカメラによる視線計測技術の開発
  - 不正確で不安定な視線情報からの注視推定技術の開発

### (3)「久野」グループ

- ① 研究分担グループ長:久野 義徳 (埼玉大学、教授)
- ② 研究項目
  - ロボットの身体動作に対する注視反応モデルの構築

### (4)「小池」グループ

- ① 研究分担グループ長:小池 英樹 (電気通信大学、教授)
- ② 研究項目
  - 注視反応モデルにもとづいた注視誘導技術の開発

### §3. 研究実施内容

(文中に番号がある場合は(4-1)に対応する)

#### 注視推定技術の開発

##### 環境埋め込みカメラによる視線計測技術の開発

日常生活空間において人物の視線を推定する際、必ずしも顔領域の高解像度画像が撮影できるとは限らないため、状況設定に応じて適切な入力情報を使った推定手法を選択することが重要となる。特に監視映像などの低解像度画像を入力とした視線推定においては、目領域の情報を得ることはほぼ不可能になるため、視線の近似として顔向きを推定することが重要になる。過去の研究では、低い解像度の顔画像から顔向きを推定するアピアランスベースの手法がいくつか提案されているが、学習用の正解画像を多数用意しなければならないという問題があった。人間の顔の見え方はシーンやカメラに依存して大きく変化するため、推定精度を高めるためには実際に撮影する状況下で正解画像を獲得する必要があるが、現実的には不可能な場合も多い。

この問題を解決するために、設置された環境下で自動的に正解画像を獲得し、手作業のラベル付け無しで学習を行う顔向き推定技術の開発に取り組んだ。基本的なアイデアは、歩行中の人物はほとんどの場合進行方向を向いているという仮定のもと、人物の進行方向を正解顔向き方向として使うことにある。しかしこの仮定には、ほとんどのシーンにおいて人物の進行方向には偏りが見られ、全ての顔向きに対して均等な学習サンプルが得られないという問題がある。そのため、本研究では進行方向をラベルとして得られたシーン特有の学習データと異なる環境下で撮影された一般学習データを転移学習の枠組みで統合することを提案している。これにより、一般学習データを直接使った場合に比べて顔向き推定精度が向上することが確認できた。

##### 不正確で不安定な視線情報からの注視推定技術の開発

人間の注意について知るためには、人間側から得られる視線情報だけではなく、シーンの側から得られる視覚的顕著性などの情報も重要な役割を果たすと考えられる。本研究ではこれまで、視覚的顕著性のメカニズムに注目しながら、それを応用した視線推定手法に関して一定の成果を収めてきた。既に発表を行った、映像を見ている人物の目画像と映像コンテンツの視覚的顕著性の対応から、アピアランスベースのアプローチにより視線推定器を学習するという枠組みに関しては、その後推定精度の向上を図り、さらに視覚的顕著性の計算モデル学習も組み込んだシステムとして改良した。

また、視覚的顕著性の計算モデルは、現在でも未解決の問題が多数残されており、活発な研究がなされる分野となっている。本研究では、今年度、特に一人称視点での視覚的顕著性メカニ

ズムに注目して既存モデルの検証、およびより高精度な注視予測を行うことができるモデルの提案を行ってきた。視野カメラによる一人称視点映像中における視線・注意推定は、ライフログや **Augmented Reality** 等の観点からも重要な技術となるが、顕著性マップモデルに関する研究は主にディスプレイ上の映像提示を元に行われてきた面があり、実際の人間の視野に近い、一人称映像における視覚的注意のメカニズムには直接当てはまらない部分も多い。その一つの例としては、人間は自己運動により生じる視覚の変化を相殺した上で視野内の運動物体を知覚していると言われているが、既存の顕著性マップモデルにはこの働きが含まれていない点が挙げられる。このため、一人称映像においては運動特徴に基づく顕著性マップが視線と一致しないことが予想され、これは実際に我々の実験によって確認された。さらに本研究では、一人称視点映像における新しい視覚的顕著性のモデルとして、自己並進運動の消失点を使ったマップ算出の提案・評価を行った。自己運動の消失点に注意が向きやすいという仮定のもと、提案モデルでは一人称視点映像を元にカメラの自己運動を推定し、並進運動の消失点に顕著性を発生させる。このモデルに注視予測精度を向上させる効果があることが、実験によって確認できた。今後はより実用に即した高精度な顕著性マップモデルに関する研究を引き続き継続していく予定である。

## 注視誘導技術の開発

### ロボットの身体動作に対する注視反応モデルの構築

ロボットの身体動作により特定の人々の注視を誘導できるようにすることを目標に研究を進めている。今年度は、身体動作のうち頭部と視線の動きを中心に検討した。そのために、図1に示すようなロボットの頭部を開発した<sup>2)</sup>。これは人間の顔状のマスクの裏側からプロジェクタでCGで生成した目を投影するものである。全体をパンチルト運動機構に載せて頭部全体を動かすことができる。目の表示はCGなので、図1のように瞬きをしたり視線を動かすことなどがスムーズにできる。まず、54人の被験者を用いた実験により、このロボットが頭部や視線を動かすことで、ロボットに見られたと感じることを確認した<sup>2)</sup>。次に、同じロボット頭部を2台用意し、ロボットの頭部運動が人間の注視を獲得できるかを12人の被験者を用いた実験により調べた。ロボットが置かれていると、何もしなくても注目をかなり集めると思われる。そこで、同じロボットを2台製作し、それらを図2のように被験者の右前方と左前方に配置し、その内の1台はときどき頭部を動かすが、もう1台は静止したままとした。ロボットの背後の壁には絵画を水平に5点並べ、被験者にはそれらを見回すように指示した。そして、ロボットが頭部を動かしたときに注意を集めるかを調べた。実験の結果、頭部を動かすと静止している場合に比べ、有意に人間の注意を獲得できることが分かった<sup>2)</sup>。ただし、頭部運動で注意を集めても、そのままではまた注意が他に向かってしまう。そこで、人間がロボットを振り向いてきたときに合わせてロボットも人間の方を向くようにした。人間の頭部の動きはロボットとは別に設置したカメラの画像から認識するようにした。そして、ロボットと人間が相互に正面を向いた後にロボットが目を瞬きするようにした。これにより、人間同士が相互を意識してアイコンタクトを

したと感じるのと同様の効果が得られるかを調べた。比較のために、そのような目を合わせる動作をしないロボットも用意して、12人の被験者を用いて実験を行った。その結果、このような動作により、アイコンタクトをしたと人間に感じさせられることが分かった 2)。さらに、アイコンタクトの後、ロボットが他の方向の物体を見ると、人間もその物体を見るという共同注視が確立して注視誘導ができるかどうかを調べた 3)。その結果、一度ではうまくいかない場合もあるが、再度相手を見て繰り返せば、注視誘導が可能であることが分かった。

以上のように、注視誘導を行うためには、ロボットから見れば、①注視獲得、②アイコンタクト(注視保持)、③注視移動という3段階の行動が必要ながわかった。今年度は、人間の視野内にロボットがいて、人間も特に特定のものに注意を集中していないという状況を仮定して検討を行った。実際には、このような状況以外にも多くの場合がある。その際に主要な問題になるのは注視獲得の部分である。注視が獲得できれば、以後はかなり共通だと考えられる。そこで、今後は状況に応じた注視獲得の方法と、状況の認識法を中心に研究を進めていく。

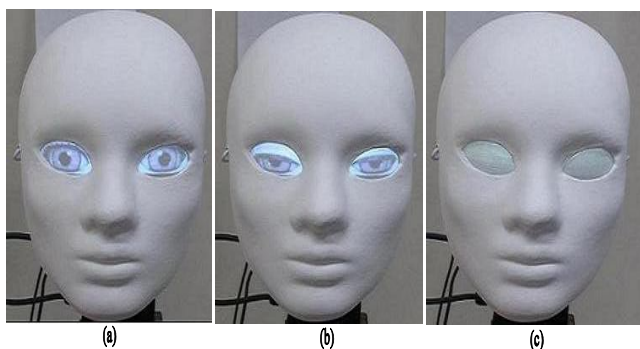


図1 実験用ロボット頭部



図2 注視獲得の実験

### 投影映像に対する注視反応モデルの構築

ユーザの位置、姿勢、視線の向きを入力として、注視を誘導するために適当な位置に情報を提示するシステムを開発している。今年度は、ユーザが提示された情報を顔の向きだけで操作することのできる図1に示すシステムを開発した。

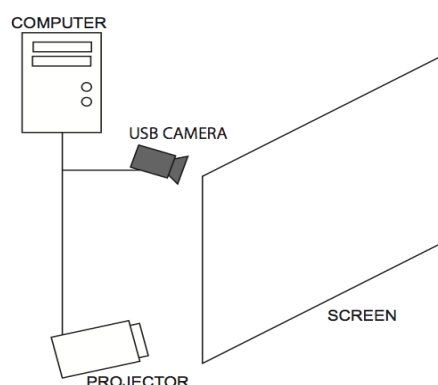
システムは半透明スクリーン(Denso Hologram Screen)と映像投影用プロジェクタ、ユーザの顔撮影用カメラ(Point Gray DragonFly)、画像処理用PCからなる。スクリーン手前のユーザの顔画像をカメラが撮影し、これを Seeing Machines 社製 faceAPI によって実時間特徴点追跡を行う。得られた特徴点から計算される顔の3次元姿勢から顔の正面方向を視線の向きと近似して使用している。実際には顔の向きと視線の向きには若干の差が生じるが、本研究が対象とする壁型ディスプレイにおいては、この近似は我々の以前の研究により妥当であることが示されている。ただし実時間で得られる顔の向きをそのままスクリーン上の注視点とすると、顔の揺れや計測ノイズ等の関係で注視点が不安定となる。この解決策として20枚程度のフレームから得られる顔の向きを平均化することで安定化に成功した。スクリーン上にはマトリクス状に写真が表示されており

(図2), ユーザが特定の写真をある一定時間注視すると, システムはユーザがその写真に興味があると判定し, その写真を拡大表示する. また顔の向きを左右に移動させる自然な動作で, 前後の写真にスクロールすることもできる. さらに特定の写真を拡大した状態でユーザがスクリーンに近づくと, システムはユーザがその写真にさらなる興味があると判定し, その写真に関するより詳細な情報, 特に文字による情報を表示する. つまり, ユーザのスクリーンからの距離に応じて提示情報の粒度を変化させている.

本システムの定量的評価はまだ行っていないが, 研究室でのインフォーマルな実験では顔による画像選択は安定しており, また前後画像への変更, 詳細な情報の取得などが, 自然で直感的な操作でできることがわかっている. 今後は被験者実験を含めた定量的評価を行っていく予定である.

なお本システムに関する学会発表はまだであるが, 平成23年3月20日放送のTBSテレビ「夢の扉」にて本システムは詳しく紹介された.

今後はシステムのマルチユーザ対応を進めるとともに, スクリーン上の特定の位置にユーザの視線を誘導する技術を開発していく予定である.



## §4. 成果発表等

### (4-1) 原著論文発表

#### ● 論文詳細情報

1. Diego Thomas and Akihiro Sugimoto: Robustly Registering Range Images Using Local Distribution of Albedo, Computer Vision and Image Understanding, Vol. 115, No. 5, 649-667, 2011. (DOI: 10.1016/J.CVIU.2010.11.016)

### (4-2) 知財出願

- ① 平成22年度特許出願件数(国内 0件)
- ② CREST 研究期間累積件数(国内 0件)