

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. 研究実施体制

(1) 「佐藤」グループ

- ① 研究代表者: 佐藤 洋一 (東京大学生産技術研究所、教授)
- ② 研究項目
 - カメラによる視線計測技術の開発
 - 注視推定と誘導のための視覚的顕著性モデルの開発

(2) 「杉本」グループ

- ① 主たる共同研究者: 杉本 晃宏 (国立情報学研究所コンテンツ科学研究系、教授)
- ② 研究項目
 - カメラによる視線計測技術の開発
 - 視覚的顕著性にもとづく注視誘導技術の開発

(3) 「久野」グループ

- ① 主たる共同研究者: 久野 義徳 (埼玉大学大学院理工学研究科、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 主たる共同研究者: 小池 英樹 (電気通信大学大学院情報システム学研究科、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

§2. 研究実施内容

(文中に番号がある場合は(3-1)に対応する)

注視推定技術の開発

カメラによる視線計測技術の開発

アピアランススペースの頭部姿勢推定において高精度な推定を実現するためには、推定器学習のための真値データセットを適切な条件下で獲得することが重要となる。従来研究で想定されていた手作業のデータセット構築に代わるアプローチとして、実際の設置環境で観測される歩行者の進行方向を顔向きのラベルとして用いる手法を提案した⁹⁾。提案手法により、別環境で撮影されたデータセットを用いるよりも高精度な学習が可能であることが示された。

また、カメラベースの視線推定は一般的に、目画像と注視点座標の関係性を学習することで行われるが、従来手法では学習のためのデータセットとして画面中の密な参照座標に対応する大量の目画像を必要としており、これが煩雑なキャリブレーション作業に繋がっていた。これに対し、今年度は L1 ノルム最適化の枠組みを用いて線形補間を行うためのデータを適応的に選択することで、少数の学習データから視線推定を行う手法を提案した⁸⁾。さらに同様の枠組みをテスト画像の位置合わせにも用いることで、簡便で高精度な視線推定を実現している。

一方、カメラに対する相対的な頭部姿勢によって目画像の見えは大きく変化するため、カメラベースの視線推定においては頭部姿勢変動の取り扱いも非常に重要な課題となる。この問題に対し今年度は、目画像と注視点座標のマッピング関数に頭部位置による注視点の変動と頭部回転による目画像の歪みの補償項を組み合わせ、頭部姿勢の変動に頑健な視線推定を実現する手法を提案した³⁾。補償項のための簡易なキャリブレーション動作を追加するだけで、従来手法以上の精度で視線推定が可能となる。

注視推定と誘導のための視覚的顕著性モデルの開発

中心視と周辺視での視野特性の違いは人間の注意のメカニズムにおいて重要な役割を果たしていると考えられる。しかしながら、既存の視覚的顕著性モデルの多くは、各画像特徴を対象画像全体で均一に扱うものがほとんどであり、このような特性が反映されていない。今年度は、視覚刺激を受容する網膜部位に応じて、その視覚特性が変化することを考慮した視覚的顕著性モデルの構築手法を提案した¹⁵⁾。従来モデルでは均一な重みが用いられていた特徴マップの統合段階において、各特徴マップを現在の注視位置からの距離・方向に応じた領域に分割し、それぞれの特徴・領域に対して異なる重みパラメータを割り当てることで統合を行う。それぞれの重みパラメータは、実際に計測した視線データに対する学習によって最適化することにより、視野特性に対応した顕著性モデルを実現する。評価実験により、既存モデルに比べてサッカード予測性能の高い顕

著性マップが計算できることを確認するとともに、獲得した重みパラメータの分析により、視野特性と視覚的顕著性モデルにおける最適な特徴・スケール選択の関わりを明らかにした。

注視誘導技術の開発

視覚的顕著性にもとづく注視誘導技術の開発

人の注視を遮ることなく、自然と視線を誘導する視線インタフェースの実現を目指して、視覚的顕著性に基づく画像加工を提案した⁵⁾。見る対象の様子によって注視のしやすさは変わるという発想に基づき、注視の引き込みやすさを示す指標として、視覚的顕著性マップの計算モデルを利用した。そして、画像中で指定した領域の視覚的顕著性が最も高くなるように輝度と色相を画素ごとに調整するという処理を繰り返し、入力画像を加工した。また、画像を用いた実験によって視覚的顕著性の変化を検証し、提案手法の有効性を確認した。

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、特定の人に必要なところを注視するように誘導する技術の開発を目指して研究を進めている。これまでに、注視誘導を行うためにはロボットから見れば、注視獲得、アイコンタクト、注視移動の3つの行動が必要なことを、それぞれの行動段階に対する実験で示した。まず今年度は、特定の条件において、ロボットが一連の動作で注視誘導ができることを、被験者を用いた実験で確認した¹⁾。どのような身体行動で注視誘導が可能かは、対象とする人間の注意レベル、その人間とロボットの位置関係、さらに周囲の状況などの条件により変わってくる。この条件のうち、今年度は人間とロボットの位置関係について検討した。位置関係は、ロボットが人間の視野のどこに捉えられるかで表すことにし、近周辺視、遠周辺視、視野外の条件で、注視誘導の実験を行った。その結果、近周辺視の場合は、ロボットが相手に顔を向ける程度で注視を獲得し誘導できるが、遠周辺視では、頭部を動かすなどのさらに動的な行動が必要なが分かった。視野外の場合は、当然ではあるが、視覚に訴える身体行動では注意を獲得できないので、音声などの他の行動が必要なが分かった。また、注視誘導の応用の例として、話に注意を払っていない相手を言語・非言語行動で引き付けることのできるガイドロボットの検討を行った^{12, 14)}。

注視反応モデルにもとづいた注視誘導技術の開発

ユーザの位置、姿勢、視線の向きを入力として、注視を誘導する対話型壁型ディスプレイシステムを開発している。今年度は、(1)ユーザの壁からの距離に応じて同時に複数の表示を実現する装置、(2)画像の解像度を動的に変化させることでユーザの視線を誘導する装置、の2つを主に開発した。前者はモザイク表示を応用したものである。1枚の大きな画像(遠距離画像)を複数枚の小さ

な画像(近距離画像)から構成すると、画面からの距離の大きなユーザには遠距離画像が認識され、距離の小さなユーザには近距離画像が認識される。本研究では、まず距離と認識できる画像解像度の関係を調べた。次に、近距離画像の色情報を微妙に変化させることでより自然な遠距離画像を実現する手法を試みた。後者は画像の解像度変化を利用するものである。画像の一部を高解像度とし他を低解像度にする、ユーザの視線が高解像度部分に誘導される、という仮定のもとに研究を行った。まず、一部高解像度部分のある画像を提示した場合のユーザの視線移動を調べた。結果は予想どおり高解像度部分に視線が誘導された。次に、低解像度部分の許容度を調べるため、異なる解像度の画像の同じ画像をユーザに提示し、その許容範囲、および認識の範囲を調べた。

§3. 成果発表等

(3-1) 原著論文発表

● 文詳細情報

1. M. M. Hoque, T. Onuki, Y. Kobayashi, Y. Kuno, Controlling Human Attention through Robot's Gaze Behaviors, 4th International Conference on Human System Interaction (HSI2011), pp.195-202, May 2011. (DOI: 10.1109/HSI.2011.5937366)
2. T. T. Le, Y. Kang, A. Sugimoto, S. T. Tran and T. D. Nguyen: Hierarchical Spatial Matching Kernel for Image Categorization, Proc. of Int. Conf. on Image Analysis and Recognition (ICIAR2011), LNCS 6753, June 2011. (DOI: 10.1007/978-3-642-21593-3_15)
3. F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "A Head Pose-free Approach for Appearance-based Gaze Estimation", in Proc. British Machine Vision Conference (BMVC2011), August 2011. (DOI: 10.5244/C.25.126)
4. 菅野裕介, 松下康之, 佐藤洋一, “視覚的顕著性を用いた自己較正による視線推定”, 電子情報通信学会論文誌 Vol.J94-D, No.8, August 2011. (DOI: 不明)
5. A. Hagiwara, A. Suimoto and K. Kawamoto: Saliency-Based Image Editing for Guiding Visual Attention, Proc. of 1st Int. Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI2011), September 2011. (DOI: 10.1145/2029956.2029968)
6. Y. Kang, H. Nagahashi and A. Sugimoto: Image Categorization Using Scene-Context Scale Based on Random Forests, IEICE Trans. Inf. and Syst., Vol. 94-D, No.9, September 2011. (DOI: 10.1587/transinf.E94.D.1809)
7. L. Su, S. Kumano, K. Otsuka, D. Mikami, J. Yamato, and Y. Sato, "Early Facial Expression Recognition with High-frame Rate 3D Sensing", in Proc. IEEE Int.

- Conf. Systems, Man, and Cybernetics (SMC2011), October 2011. (DOI: 10.1109/ICSMC.2011.6084179)
8. F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Inferring Human Gaze from Appearance via Adaptive Linear Regression", in Proc. IEEE International Conference on Computer Vision (ICCV2011), November 2011. (DOI: 10.1109/ICCV.2011.6126237)
 9. I. Chamveha, Y. Sugano, D. Sugimura, T. Siriteerakul, T. Okabe, Y. Sato, and A. Sugimoto, "Appearance-Based Head Pose Estimation with Scene-Specific Adaptation", in Proc. IEEE Workshop on Visual Surveillance (VS2011), November 2011. (DOI: 10.1109/ICCVW.2011.6130456)
 10. K. Yamada, Y. Sugano, T. Okabe, Y. Sato, A. Sugimoto, and K. Hiraki, "Attention Prediction in Egocentric Video using Motion and Visual Saliency", in Proc. Pacific-Rim Symposium on Image and Video Technology (PSIVT2011), November 2011. (DOI: 10.1007/978-3-642-25367-6_25)
 11. Y. Kang and A. Sugimoto: Scale-Optimized Textons for Image Categorization and Segmentation. Proc. of IEEE International Symposium on Multimedia (ISM2011), December 2011. (DOI: 10.1109/ISM.2011.48)
 12. M. A. Yousuf, Y. Kobayashi, A. Yamazaki, Y. Kuno, Implementation of F-Formation and "Pause and Restart" for a Mobile Museum Guide Robot, International Workshop on Multimodality in Multispace Interaction (MiMI2011), December 2011. (DOI: 不明)
 13. D. Thomas and A. Sugimoto: Illumination-Free Photometric Metric for Range Image Registration, Proc. of IEEE Workshop on Application of Computer Vision (WACV2012), January 2012. (DOI: 10.1109/WACV.2012.6163041)
 14. M. A. Yousuf, Y. Kobayashi, A. Yamazaki, K. Yamazaki, Y. Kuno, A Mobile Guide Robot Capable of Formulating Spatial Formations, Proc. 18th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV2012), February 2012. (DOI: 不明)
 15. H. Kubota, Y. Sugano, T. Okabe, Y. Sato, A. Sugimoto and K. Hiraki, "Incorporating Visual Field Characteristics into a Saliency Map", In Proc. Symposium on Eye Tracking Research and Applications (ETRA2012), March 2012. (DOI: 10.1145/2168556.2168629)