

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

- ・環境埋め込みカメラによる視線計測技術の開発
- ・不完全な視線情報からの注視推定技術の開発
- ・視覚刺激に対する注視反応モデルの構築
- ・視覚刺激による注視誘導技術の開発

(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

- ・環境埋め込みカメラによる視線計測技術の開発
- ・不完全な視線情報からの注視推定技術の開発
- ・視覚刺激に対する注視反応モデルの構築
- ・視覚刺激による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

- ・ロボット動作に対する注視反応モデルに基づいた注視誘導技術の開発

(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

- ・視覚刺激による注視誘導技術の開発

- ・視覚適応型インタフェースの開発

§2. 研究実施の概要

● 環境埋め込みカメラによる視線計測技術の開発

本年度は、不特定人物を対象とした低解像度目画像からの視線推定技術の開発に取り組んだ。人物、および視線方向の大規模なサンプルを含む多視点画像データセットを新たに構築し、復元した 3 次元形状を元に大量の学習目画像を生成する。これを利用して、人物やカメラ姿勢に依存しない目画像と視線方向の関係を学習することで、個人毎のキャリブレーションを必要としない、純粋に画像ベースの視線推定を実現している。

● 不完全な視線情報からの注視推定技術の開発

これまで低解像度画像からの顔向き推定技術の開発に取り組んできたが、空間内での歩行者の顔向き／視線を空間的な注意マップに結び付けて理解することも重要な課題と言える。本年度はこうした取り組みの一環として、視線履歴の情報をを用いたグループ検出手法を開発した。二人の歩行者が同じグループに属するかどうかを識別する問題設定で、歩行者間の共同注視や相互注視の頻度を特徴として用いることで、従来手法よりも高い精度で空間内のグループを検出することができる。

● 視覚刺激に対する注視反応モデルの構築

これまでに開発した、人間の視野特性を考慮した視覚的顕著性モデルを元に、学習データセットの属性と学習によって得られるモデルとの関係性を定量的に検証した。学習用の画像の意味レベルや、視線計測時のタスク設定によって、学習結果のモデルには統計的に有意な差が生じることが確認でき、将来のモデルの高精度化や医療診断への応用の可能性を確認できた。

また、従来の視覚的顕著性マップは画像や映像から得られる画像特徴のみに基づいて計算されているが、注意の引かれやすさは視覚情報だけでなく聴覚情報などにも依存している。そこで、音特徴から顕著性を検出し、画像特徴と統合した視覚的顕著性マップを構築する手法を検討した。音と映像中の動きの相関によって音源の位置を推定し、音のサウンドスペクトルグラムから得られる顕著性によって、画像特徴による顕著性の高さや音源位置による顕著性の高さを統合した顕著性マップを考案し、その有効性を確認した。

● ロボット動作に対する注視反応モデルに基づいた注視誘導技術の開発

相手の行動および注意レベルを画像から認識し、適切な時点をとらえて注意を獲得し注視誘導を行うことのできるロボットを開発した。例えば、相手が本を読んでいるなら、ページをめくるときなどをとらえて視線を向けることで注意を獲得する。美術館等では特定の展示品を見ていて興味がありそうならガイドを申し出る。また、このような注視誘導を有効に行えるように、ロボットの目とその取りつけられている頭部形状との関係を調べた。その結果、ロボットの視線をとらえやすくするには、鼻のようなものがある方がよいことが分かった。そこで、そのような形状を持ち、レーザプロジェクトで目のCG像を投影することにより視線を高速に動かすことのできるロボット頭部を開発した。

● 視覚刺激による注視誘導技術の開発

これまでに開発した、静止画像加工による注視誘導技術を動画像に適用できるように拡張した。人間の知覚により近いとされている HSI 空間で加工を行うことで加工による違和感の軽減を図るとともに、動画像中の各フレームにおいて指定領域の顕著性が常に最大になるように動的に視覚的顕著性を制御し、その後フレーム間の制御量の変動を平滑化するというアプローチを採用した。また、人物追跡技術を取り入れることで、動画像中の初期フレームにおいて指定した領域を自動的に追跡できるようにした。

また、解像度制御に基づく視線誘導技術の提案と開発を行った。本手法は視線を誘導したい部分の解像度は高解像度のまま残し、他の部分の解像度を徐々に下げる。ただし、この解像度制御は視聴者の気づかない範囲で行われる。実験の結果、視聴者の視線は高解像度領域に誘導されることが確認された。

● 視覚適応型インタフェースの開発

公共空間における大型のデジタルサイネージはその表示領域やコンテンツのサイズにより最適な視聴距離がある。本研究ではモザイク画像の特徴を応用することにより、遠近2つの距離にいる視聴者に対してそれぞれ異なった情報を同時に提案する画像の生成手法を開発した。本年度は、本手法をさらに改善し、遠近各画像の視認性向上を行った。

§3. 成果発表等

(3-1) 原著論文発表

論文詳細情報(国内)

1. 畑元, 小池英樹, 佐藤洋一, 解像度制御を用いた視線誘導, 情報処理学会インタラクション 2014 予稿集, 2014 年 2 月

論文詳細情報(国際)

2. L. Su and Y. Sato, "Early Facial Expression Recognition with Early RankBoost", In Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition (FG 2013), pp. 1-7, April 2013. (DOI: 10.1109/FG.2013.6553740)
3. Y. Sugano, Y. Matsushita and Y. Sato, "Graph-based joint clustering of fixations and visual entities", *ACM Trans. on Applied Perception*, Vol. 10, No. 2, Article 10, May 2013. (DOI: 10.1145/2465780.2465784)

4. M. M. Hoque, T. Onuki, Y. Kobayashi, and Y. Kuno, "Effect of Robot's Gaze Behaviors for Attracting and Controlling Human Attention," *Advanced Robotics*, vol.27, no.11, pp.813-829, May 2013. (DOI:10.1080/01691864.2013.791654)
5. M. M. Hoque, K. Deb, D. Das, Y. Kobayashi, and Y. Kuno, "An Intelligent Human-Robot Interaction Framework to Control the Human Attention," in *Proc. International Conference on Informatics, Electronics & Vision (ICIEV)*, May 2013. (DOI:10.1109/ICIEV.2013.6572539)
6. D. Das, Y. Kobayashi, and Y. Kuno, "Observing Human's Face for Robot's Controlling His/Her Attention," in *Proc. International Conference on Quality Control by Artificial Vision (QCAV2013)*, pp.138-143, May 2013.
7. F. Lu, Y. Matsushita, I. Sato, T. Okabe and Y. Sato, "Uncalibrated photometric stereo for unknown isotropic reflectances", In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR2013)*, pp. 1490-1497, July 2013. (DOI: 10.1109/CVPR.2013.196)
8. Y. Sugano, H. Kasai, K. Ogaki, and Y. Sato, "Image Preference Estimation from Eye Movements with a Data-driven Approach", in *Proc. Int. Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI2013)*, August 2013.
9. M. A. Yousuf, Y. Kobayashi, Y. Kuno, A. Yamazaki, and K. Yamazaki, "How to Move towards Visitors: A Model for Museum Guide Robots to Initiate Conversation," in *Proc. International Symposium on Robot and Human Interactive Communication (Ro-Man2013)*, pp.587-592, August 2013. (DOI:10.1109/ROMAN.2013.6628543)
10. D. Thomas and A. Sugimoto, "Range Image Registration Using a Photometric Metric under Unknown Lighting," *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol.35, No. 9, pp. 2252-2269, September 2013. (DOI:10.1109/TPAMI.2013.21)
11. I. Chamveha, Y. Sugano, Y. Sato and A. Sugimoto, "Social Group Discovery from Surveillance Videos: A Data-Driven Approach with Attention-Based Cues", in *Proc. British Machine Vision Conference (BMVC 2013)*, pp. 1-12, September 2013. (DOI: 10.5244/C.27.121)
12. I. Chamveha, Y. Sugano, D. Sugimura, T. Siriteerakul, T. Okabe, Y. Sato, and A. Sugimoto, "Head Direction Estimation from Low Resolution Images with Scene

- Adaptation", *Computer Vision and Image Understanding*, vol. 117, No. 10, pp. 1502-1511, October 2013. (DOI:10.1016/j.cviu.2013.06.005)
13. J. Nakajima, A. Sugimoto and K. Kawamoto, "Incorporating Audio Signals into Constructing a Visual Saliency Map," In *Proc. Pacific-Rim Symposium on Image and Video Technology (PSIVT2013)*, pp.468-480 (LNCS 8333, Springer), October 2013.
 14. T. Shi and A. Sugimoto, "Video Saliency Modulation in the HSI Color Space for Drawing Gaze," In *Proc. Pacific-Rim Symposium on Image and Video Technology (PSIVT2013)*, pp.206-219 (LNCS 8333, Springer), October 2013.
 15. D. Das, Y. Kobayashi, and Y. Kuno, "Attracting Attention and Establishing a Communication Channel Based on the Level of Visual Focus of Attention," in *Proc. International Conference on Intelligent Robots and Systems (IROS2013)*, pp.2194-2201, November 2013. (DOI:10.1109/IROS.2013.6696663)
 16. Y. Suzuki and H. Koike, "MultiMosaic: simultaneous visualization of multiple images for near and far viewers using image mosaic and saliency map", In *SIGGRAPH Asia 2013 Technical Briefs (SA '13)*, Article 19, 4 pages, November 2013. (DOI: 10.1145/2542355.2542379)
 17. D. Thomas and A. Sugimoto, "A Flexible Scene Representation for 3D Reconstruction Using an RGB-D Camera," In *Proc. IEEE International Conference on Computer Vision (ICCV2013)*, pp.2800-2807, December 2013. (DOI: 10.1109/ICCV.2013.348)
 18. P. K. Anasosalu, D. Thomas and A. Sugimoto, "Compact and Accurate 3-D Face Modeling Using an RGB-D Camera: Let's Open the Door to 3-D Video Conference," In *Proc. IEEE Workshop on Consumer Depth Cameras for Computer Vision (CDC4CV2013)*, pp.67-74, December 2013. (DOI: 10.1109/ICCVW.2013.16)
 19. T. Onuki, T. Ezure, T. Ishinoda, Y. Kobayashi, and Y. Kuno, "Static and Dynamic Robot Gaze Expressions to Communicating with Humans," in *Proc. Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV2014)*, February 2014.
 20. B. Ye, Y. Sugano, and Y. Sato, "Influence of stimulus and viewing task types on a learning-based visual saliency model", in *Proc. Symposium on Eye Tracking Research and Applications (ETRA2014)*, March 2014.

21. F. Lu, T. Okabe, Y. Sugano, and Y. Sato, "Learning gaze biases with head motion for head pose-free gaze estimation", *Image and Vision Computing*, vol 32, issue 3, pp. 169-179, March 2014. (DOI: 10.1016/j.imavis.2014.01.005)
22. Y. Sugano, Y. Ozaki, H. Kasai, K. Ogaki, and Y. Sato, "Image preference estimation with a data-driven approach: a comprehensive study between gaze and image features", *Journal of Eye Movement Research*, Vol. 7, Issue 3, No. 5, pp. 1-9, March 2014.
23. F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Adaptive Linear Regression for Appearance-Based Gaze Estimation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Online Preprint, March 2014. (DOI: 10.1109/TPAMI.2014.2313123)
24. T. Onuki, T. Ishinoda, E. Tsuburaya, Y. Miyata, Y. Kobayashi, and Y. Kuno, "Designing Robot Eyes for Communicating Gaze," *Interaction Studies* (accepted)
25. Y. Sugano, Y. Matsushita, and Y. Sato, "Learning-by-Synthesis for Appearance-based 3D Gaze Estimation", in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR2014)*. (accepted)