

戦略的創造研究推進事業 CREST

研究領域

「共生社会に向けた人間調和型情報技術の構築」

研究課題

「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」

研究終了報告書

研究期間 平成21年10月～平成27年3月

研究代表者:佐藤 洋一
(東京大学 生産技術研究所、教授)

§ 1 研究実施の概要

(1) 実施概要

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に取り組んだ。

前者の注視推定について、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」(佐藤グループ、杉本グループ)では、特殊な照明などの装置を用いることなく、視線を高精度かつ簡略に取得可能とする技術の開発に取り組み、重要な成果を得た。まず、遠隔から撮影された低解像度映像のみしか得られない場合であっても、撮影環境の違いに影響を受けることなく人物の顔向きから注視方向を得ることに成功している。また、目画像のみを入力として用いるアピアランススペースの注視推定において、少数の学習用目画像を適応的に選択し視線推定に用いるアルゴリズムを新たに開発し、高精度な視線推定には多数の学習データが必要になるという主要な課題を解決した。さらに、目画像のアピアランスが頭部姿勢や個人に依存するために、不特定人物・任意頭部姿勢の条件下ではアピアランススペースの視線推定が上手く働かないという課題に対して、さまざまな条件下での目画像を生成し視線推定に利用するという独自のアプローチにもとづく一連の手法を開発し、その効果を実験により検証した。研究項目「不完全な視線情報からの注視推定技術の開発」(佐藤グループ)では、人物に事前キャリブレーションを強いることなく、より自然な条件下での注視推定を実現するために、被注視対象の視覚的顕著性の分布を手掛かりとした視線推定技術を他に先駆けて開発した。研究項目「視覚刺激に対する注視反応モデルの構築」(佐藤グループ、杉本グループ)では、視覚刺激による注視引付けのメカニズムについて検討し、これまでにない新たなモデルを提案することに成功した。観察者の視点から撮影された一人称視点映像を用いた注視予測は、ディスプレイに提示された映像を受動的に観察する場合の注視予測とは本質的に異なることに着目し、視覚的顕著性と観察者の自己運動の両方を考慮した注視予測モデルを提案し、注視予測精度の向上を確認した。また、人の視野特性は中心窩と周辺視で大きく異なることを踏まえ、人の注視を実測した大規模データベースを用いた学習アプローチにより、不均一な視野特性を適切に表現可能な視覚的顕著性モデルを新たに構築した。さらに、このモデルを用いて属性の異なる画像や人物間で視覚的顕著性モデルにどのような違いがあるのかを検証している。さらに、視覚と聴覚の異なるモダリティ間の同期が視覚的注意に強く影響するという心理物理学的知見を踏まえ、視覚刺激と聴覚刺激を統合した視覚的顕著性モデルを新たに提案し、聴覚刺激の考慮により視覚的注意予測の精度が向上することを確認した。

後者の注視誘導について、視覚刺激による注視誘導とロボットの身体動作による注視誘導の2つのアプローチについて研究を進めた。研究項目「視覚刺激による注視誘導技術の開発」(佐藤グループ、杉本グループ、小池グループ)では、先の研究項目で成果を得た視覚的顕著性モデルを活用し、視覚的顕著性を手掛かりにして対象領域の色や明るさを適切に変化させることにより、さ

りげなくそこに注視を誘導する技術を開発し、静止画と動画の両方についてその効果を実証した。また、被写体深度の変化が人の注視に影響するという知覚心理学分野の知見を踏まえ、画像解像度の制御による注視誘導というアプローチを新たに提案し、その効果を実証した。また、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」(久野グループ)では、人の行動分析に基づき、ロボット動作に対する注視反応モデルに関する仮説を立て、ロボットを用いた実験によりその効果を検証した。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」(久野グループ)では、先の研究項目で構築したモデルに基づいて動作するロボットを開発し、ロボットの身体動作による自然な注視誘導への有効性を実証した。その際、ユーザ実験を通じて、ユーザにとっての親しみやすさと視線の読み取りやすさを兼ね備え、注視誘導に適したロボットの目及び頭部のデザインを明らかにした上でロボットに実装し実験を行っている。最後に、研究項目「視線適応型インタフェースの開発」(小池グループ)では、公共空間に配置されるデジタルサイネージを想定し、ユーザの位置や視線に適応した対話型ディスプレイシステムを構築し、基礎的な実験を通じその有効性と課題を検証した。

(2) 顕著な成果

<優れた基礎研究としての成果>

1. Appearance-based Gaze Estimation using Visual Saliency

人間の視覚特性から注視分布を予測する視覚的顕著性の計算モデルを手掛かりにして注視推定を行う手法を開発した。映像とそれを見ている人物の目画像の組のみを入力として、顕著性から予測した注視点の存在分布を元に目画像と注視点座標の関係性を学習する。さらに推定された注視点座標を視覚的顕著性モデルの最適化に用いる枠組みも提案し、明示的なキャリブレーションを一切要することなく平均約 3.5 度の誤差で注視点を推定できることを示した。情報分野において最も高いインパクトファクタを持つ論文誌の一つである IEEE Trans. PAMI に採択された。また、観察者の自己運動、不均一な視野特性、音響情報とのマルチモーダル統合を考慮することで視覚的顕著性にもとづく注視予測自体の精度向上も実現した。

2. Adaptive Appearance-Based Gaze Estimation

目画像のみを入力として用いるアピアランスベースの注視推定において、少数の学習データから高精度な出力を行うためのアルゴリズムを提案した。目画像と注視点座標の組として与えられる学習データの中から、入力目画像を補間できる最小の組み合わせを適応的に選択し、その補間重みを注視点座標の補間に用いる。目画像の位置合わせや東部姿勢変動への対応を行うための枠組みも組み合わせることで、従来手法と比べてごく少数の学習データから、高精度な推定を行うことができる。上記成果 1 と同様に、代表論文が IEEE Trans. PAMI に採択された。さらに、アピアランスベースの注視推定における主要課題である頭部姿勢の変動への対応についても解決し、その成果が主要国際論文誌と国際会議で発表された。

3. Learning-by-Synthesis for Appearance-based 3D Gaze Estimation

対象人物に特化した学習データの存在を仮定できない、不特定多数の人物を対象としたアプリケーションスペース注視推定において、多数の人物の目画像を含む大規模なデータベースを用いて推定器を学習する手法を提案した。学習時点で未知の人物と任意の頭部姿勢方向に対応するために、多数のカメラ姿勢から生成した大量の目画像を学習に用いるアプローチを取り、既存手法を上回る性能を実現している。コンピュータビジョン分野において最もランクの高い国際学会であるCVPR に採択された。

< 科学技術イノベーションに大きく寄与する成果 >

1. 遠隔カメラによる人物顔向き推定

遠隔カメラにより得られる低解像度の画像から人物・歩行者の顔向きを推定し、行動を解析する技術は、情報支援以外にも防犯カメラ映像解析など様々な応用の可能性を持つ重要な基礎技術であると言える。本研究課題では、より簡便な顔向き推定のための、人物の進行方向を利用した適応的学習の枠組みや、注視行動履歴を特徴として用いる人物グループ検出手法の開発を行い、国際学会・論文誌において複数の成果発表を行った。

2. 視覚的顕著性に基づく注視誘導

人間の視覚特性を考慮してさげない注視誘導を行うための、視覚的顕著性にもとづく誘導技術は、自然な情報提示を行う上で必要不可欠の技術であり、情報環境における支援の枠組みに大きな変化をもたらし得る。本研究課題で検討した注視誘導モデルは、デジタルサイネージにおける情報提示や安全運転支援のための注意喚起など、実環境での応用を想定した技術開発・実地検証に取り組んだ。

3. ロボット身体動作にもとづく注視誘導

ロボットによる情報支援は、今後の新たな産業基盤としても重要な領域であり、人物の注意を考慮したインタラクション設計は自然な支援を行う上で必要な技術となる。本研究課題ではロボットによる注意誘導のためのモデル構築、および効果的な注意誘導を行うためのロボットデザインの検討に取り組み、如何にロボットが人間の注意を踏まえたインタラクションを行うかについて包括的な議論・研究を行った。

§ 2 研究実施体制

(1) 研究チームの体制について

① 佐藤グループ

研究参加者

氏名	所属	役職	参加時期
佐藤 洋一	東京大学生産技術研究所	教授	H21.10～
開 一夫	東京大学総合文化研究科	教授	H22.4～
岡部 孝弘	東京大学生産技術研究所	特任准教授	H21.10～
菅野 裕介	東京大学生産技術研究所	特任助教	H21.10～
Feng Lu	東京大学生産技術研究所	特任研究員	H23.4～
島野 美保子	東京大学生産技術研究所	特任研究員	H26.4～
米谷 竜	東京大学生産技術研究所	博士研究員	H26.4～
杉村 大輔	東京大学生産技術研究所	特任研究員	H21.10～H23.3
Teera Siriteerakul	東京大学大学院情報理工学 系研究科電子情報学専攻	D1～2	H22.4～H23.7
Shuai Han	東京大学大学院情報理工学 系研究科電子情報学専攻	D1～3	H21.10～H24.3
Lumei Su	東京大学大学院情報学環	D1～3	H21.10～H25.3
Isarun Chamveha	東京大学大学院情報理工学 系研究科電子情報学専攻	M1～D3	H22.4～
Wiennat Mongkulmann	東京大学大学院情報理工学 系研究科電子情報学専攻	D2～	H25.4～
山田 健太郎	東京大学大学院情報理工学 系研究科電子情報学専攻	M2	H22.4～H23.3
窪田 秀行	東京大学大学院情報学環	M1～2	H22.4～H24.3
大垣 慶介	東京大学大学院情報理工学 系研究科電子情報学専攻	M1～2	H23.4～H25.3
尾崎 安範	東京大学大学院情報理工学 系研究科電子情報学専攻	M2	H25.4～H26.3
Binbin Ye	東京大学大学院情報理工学 系研究科電子情報学専攻	M2	H25.4～H26.3
岡 晃平	東京大学大学院情報理工学 系研究科電子情報学専攻	M2	H26.4～
Peng Yu	東京大学大学院情報理工学 系研究科電子情報学専攻	M2	H26.4～

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

研究参加者

氏名	所属	役職	参加時期
杉本 晃宏	国立情報学研究所	教授	H21.10～
佐藤 いまり	国立情報学研究所	准教授	H21.10～
Thomas Diego	国立情報学研究所	D2～3, 特任研究員	H22.4～
姜 有宣	国立情報学研究所	特任研究員	H22.4～H23.3
望月 義彦	千葉大学融合科学研究科情報科学専攻	D3	H22.1～H23.3
萩原 愛子	千葉大学融合科学研究科情報科学専攻	M1～2	H23.4～H25.3
中島 次郎	千葉大学融合科学研究科情報科学専攻	M1～2	H25.4～

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
- ・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

研究参加者

氏名	所属	役職	参加時期
久野 義徳	埼玉大学大学院理工学研究科	教授	H21.10～
小林 貴訓	埼玉大学大学院理工学研究科	助教	H21.10～
Dipankar Das	埼玉大学大学院理工学研究科	産学官連携研究員	H23.10～H26.3

Mohammed Moshiul Hoque	埼玉大学大学院理工学研究科		H22.4～H24.9
Mohammad Abu Yousuf	埼玉大学大学院理工学研究科		H23.4～H25.9
福田 悠人	埼玉大学大学院理工学研究科	D3	H24.4～
Anjuman Ara	埼玉大学大学院理工学研究科		H25.10～H26.3
小貫 朋実	埼玉大学大学院理工学研究科	D2	H23.4～
Md. Golam Rashed	埼玉大学大学院理工学研究科	D1	H26.4～
佐野 要	埼玉大学大学院理工学研究科	M2	H26.4～

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
- ・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

研究参加者

氏名	所属	役職	参加時期
小池 英樹	東京工業大学 大学院情報理工学研究科	教授	H21.10～
佐藤 俊樹	電気通信大学大学院情報システム学研究科	助教	H24.4～
畑 元	電気通信大学大学院情報システム学研究科	M2	H24.4～
福本 将悟	電気通信大学大学院情報システム学研究科	M2	H23.4～H24.3
鈴木 裕太	電気通信大学大学院情報システム学研究科	M2	H24.4～H25.3
高橋 陽一	電気通信大学大学院情報システム学研究科	M2	H24.4～H25.3

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) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

人間のセンシングに基づく高度な情報環境の実現という背景・目的を共有する同 CREST 研究領域の武田チーム、河原チーム、黄瀬チーム、八木チーム、神田チームとの積極的な情報交換やワークショップの共同開催を通して、国内ネットワーク形成および研究の促進を図った。特に、注視推定技術や視覚的注意モデルへの関心を共有する武田チームとは密接な連携を行い、合同研究会を通して頻繁な意見交換を行った。(国際ワークショップ、合同研究会等の詳細については本報告書 § 5.1を参照)

§ 3 研究実施内容及び成果

3.1 環境埋め込みカメラによる視線計測技術の開発(佐藤グループ、杉本グループ)

既存の視線計測手法はアクティブな照明を伴う特殊な撮影装置が必要であるなど、適用範囲が限定されていた。本研究では、視線計測・注視方向推定の適用範囲を大幅に拡大することにより、日常生活空間内における人の注視をセンシング可能とすることを目指す。具体的には、防犯カメラ映像などで撮影された低解像度画像から人の注視を推定する技術、ならびにアピランススペースの視線推定手法の頑健化・高精度化に取り組んだ。

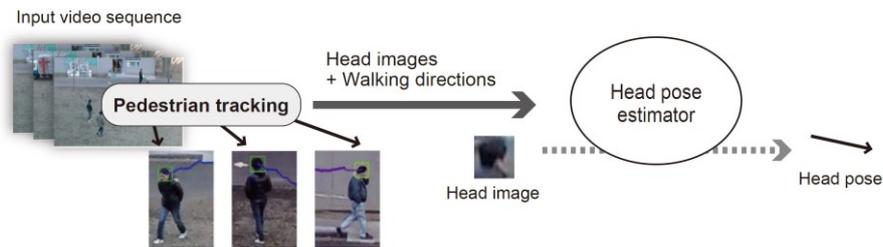


図 1 進行方向を利用した適応学習による顔向き推定

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ランススペースの手法がいくつか提案されてきた。アピランススペースの推定手法では、顔向きの正解ラベルが付与された大量の顔画像を学習データとして用いることで、画像の見え方と顔向きの関係性を学習するアプローチが取られる。特定のモデルを前提としないため低解像度画像に対しても比較的頑健に動作することが知られているが、一方で学習用の正解画像を多数用意しなければ十分な精度が得られ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。処理の概要を図 1 に示す。歩行者追跡手法を入力となる映像に適用することで獲得した歩行軌跡に対して直線あてはめ処理を行い、信頼できない追跡結果を除去した上で抽出した代表顔画像群に進行方向のラベルを与える。このようにして得られる学習データを用いることで、設置環境に適応した顔向き推定器を構築することが可能となる。顔画像を直接入力特徴とした SVM (Support Vector Machine) による顔向き識別をベースライン手法とした実証実験では、異なる撮影環境で撮影された学習データを学習に用いた場合に比べて、顔向き推定精度が向上することが確認できた。

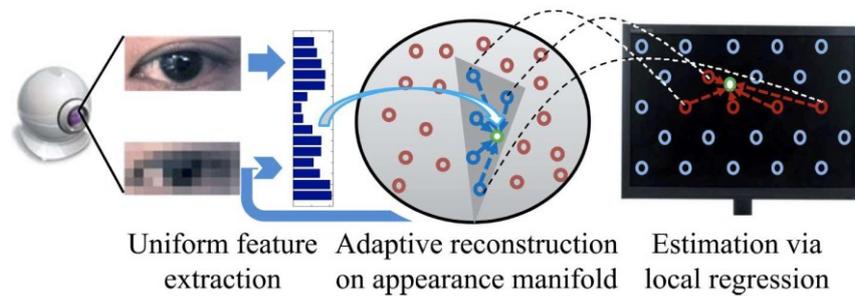


図 2 Adaptive Linear Regression による注視推定

また、アピアランススペースの視線推定も顔向き推定の場合と同様に、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。これに対し本研究では、L1 ノルム最小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。すなわち、学習データに含まれる目画像の線形結合として入力目画像を再構成する際に、「できる限り少数の学習サンプルを用いて再構成誤差を一定以下に抑える」問題としてこれを定式化することで、入力目画像を近似するのに適したできる限り少数の学習サンプルの組を適応的に発見する。このようにして得られた結合係数と同一の結合係数を用いて学習データの注視点座標を線形結合することにより、入力目画像に対応する未知の注視点座標を高い精度で出力することができる。さらに同様の枠組みを入力目画像切り出し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。これにより画面上9点という限られた学習サンプル数でも平均推定誤差1度程度の推定精度を実現でき、キャリブレーション作業の簡易化に貢献することができる。

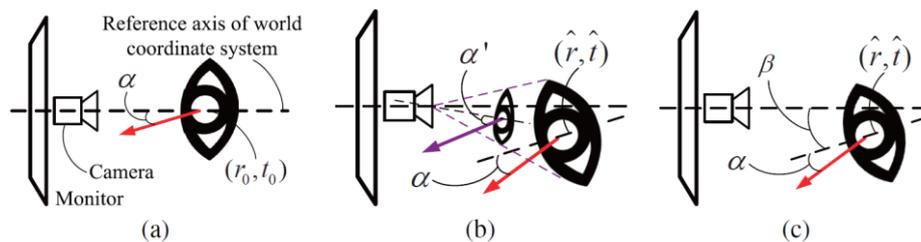


図 3 注視推定における頭部姿勢変動の補償

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、アピアランススペ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。従来手法は頭部位置の固定を前提としたものがほとんどであり、実環境に適用する上ではこれが大きな技術的ハードルになっている。この問題に対し本研究では、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。従来手法で考えられていた学習時頭部姿勢での視線推定結果(図 3(a))に加え、視線を固定して顔向き方向のみを変化させた新たな学習

データから獲得した画像歪み補償項(図 3(b))および頭部位置変化に対する幾何的な補償項(図 3(c))を組み合わせることで、自由な頭部位置変動を許容するアピランススペース視線推定が可能となる。評価実験では、頭部位置が変動するテストデータに対して従来手法を適用した場合に発生する 10 度程度の推定誤差を、提案手法により 2 度程度に抑えることができることを確認した。

さらに、上記の手法で画像歪み補償項の学習のために必要となっていた学習データをさらに削減するための新たなアプローチとして、任意の頭部姿勢に対応する目画像を、少数の頭部姿勢に対応する目画像から疑似的に合成する手法を提案した。すなわち、基準となる頭部姿勢で獲得した学習目画像群の見え方が入力頭部姿勢ではどのように変化するかを推定し、これを元に実際には撮影されていない、入力頭部姿勢での学習データを生成する。提案手法では、このような画像変形を、画素ごとに独立したフローとして定義し、未知の頭部姿勢に対するフローを、基準となる頭部姿勢以外で撮影された 4 枚の追加画像のみを用いて推定する。図 4 に示すように、提案手法によって精度よく画像変形を推定することが可能であり、少数の追加データのみを用いて前述の手法と同等かそれ以上の推定精度を示すことが実験を通して確認できた。



図 4 注視推定のための目画像生成



図 5 生成型学習のための 3 次元視線データセット

本研究項目では、これら特定人物目画像からの視線方向推定に加え、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も取り組んだ。前述のように、低解像度目画像を入力とした視線推定は、学習データに基づく回帰推定の問題として捉えることができる。しかし、これまでの研究では個人及びシーンに特化した学習データを前提としていたため、学習環境に一致した人物・カメラ環境では高精度な推定が実現できる一方、システム利用の都度、人物ごとの学習データが必要となる欠点があった。

本研究では、多人数・多視点の目画像を含む大規模な学習データセットを構築し、人物に寄らない推定器を学習するアプローチを取ることでこの問題に対する解決を図った。構築したデータセットは合計 50 人物を対象に 8 台のカメラで撮影した 160 視線方向の画像を含んでおり、従来研究で公開されてきたデータセットと比較しても最大の規模を誇る。本研究では更にこれら多視点画像から目領域の 3 次元形状を復元し、仮想的なカメラ姿勢から見た目画像を大量に生成する(図 5)。

これにより、人物、カメラ姿勢、視線方向すべてに関して密な学習データを用いて視線推定器を学習することが可能となる。この学習データセットを用いて Random Forest アルゴリズムにより学習した推定器は、個人毎のキャリブレーションを一切必要とせず、任意の姿勢から撮影した 15×9 画素の目画像のみを入力として平均誤差 6.5 度の視線推定性能を実現している。

3.2 不完全な視線情報からの注視推定技術の開発(佐藤グループ)

日常生活空間内に設置されたカメラによる視線計測では、記述の通り視線計測装置を利用した場合のように高い精度で視線情報を得ることができないことが問題となる。本研究項目では、視線の時間的な停留に加えて、視線方向にある環境の属性として視覚的顕著性やグループの注視方向の相関などの情報を用いることでこの問題に対する解決をはかる。人間側を観測するだけではなく、環境中の各部分に視線がどの程度引き寄せられやすいかを定量的に予測することで、人間と環境の双方向から注視方向を推定することを試みる。

人物にキャリブレーション動作を強いることなく自然な動作の中での注視推定を実現するために、明示的なキャリブレーション情報を必要としない注視推定技術を提案した。鍵となるアイデアは、映像から抽出した視覚的顕著性マップを注視点座標の存在確率分布として捉えることにある。視覚的顕著性マップとは、動画像中のどの領域が人間の目を惹きやすいかを表現した分布であり、人間の側を観測する一般の注視推定とは逆に、注視される側の動画像を解析することにより人間の注視行動を予測する。人物の実際の注視点を顕著性のみで完全に決定することは本質的に困難であるものの、これまで提案された数々の計算モデルにより、ある程度実際の注視点に一致した分布が得られることが実験的に証明されている。注視点推定と視覚的顕著性は密接な関わりを持つが、二つを直接結びつける研究例は決して多くなく、顕著性マップを注視点推定器の学習に利用するのは、本研究により初めて試みられるアプローチである。

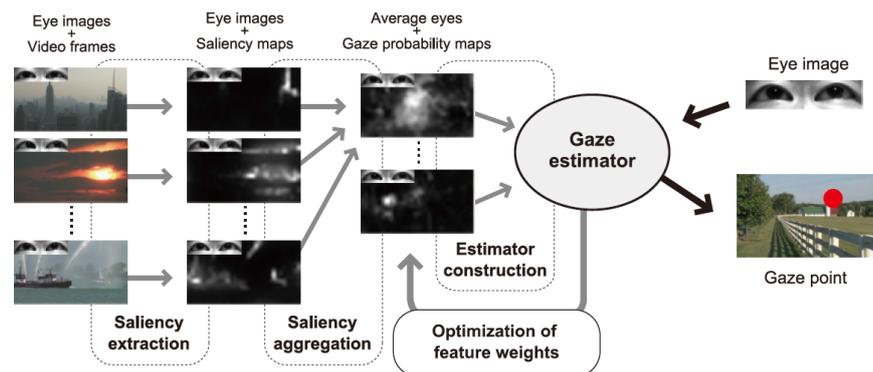


図 6 視覚的顕著性を利用した視線推定

手法の概要を図 6 に示す。提案手法では、ある映像と、頭部姿勢固定の条件下でそれを鑑賞している人物の目画像のみを入力とし、顕著性マップと目画像が同期したデータセットを入力とする。前述のとおり、ここで顕著性マップを注視点の存在確率と捉える。個々の顕著性マップそれ自体の注視点予測性能は決して高くないが、多数の目画像と顕著性マップが与えられた際、平均的

には両者間に高い相関がある。提案手法では、目画像の類似度を元に顕著性マップを統合することで少数の代表目画像とその注視点存在確率分布を生成し、最終的に目画像と注視点座標の関係性を学習する。これにより、推定対象となる入力データ自体が持つ情報を元に注視点推定器の学習を行い、注視点座標を推定することが可能となる。

さらに提案手法では、ここで一度推定された注視点座標を元に顕著性マップモデルのパラメータ最適化を行う。図7に示すように、本手法で用いる顕著性マップは色や輝度など様々な基礎特徴から計算されたマップを足し合わせることで算出されている。したがって、leave-one-out方式で推定された注視点座標に最も一致するようにこの統合重みを最適化することで、顕著性マップ自体の注視点予測性能が向上し、さらにはこのフィードバックループによって最終的な注視点推定性能も向上することが期待できる。既存の映像を入力とした評価実験では、提案手法により平均誤差3.5度での視線推定を実現している。

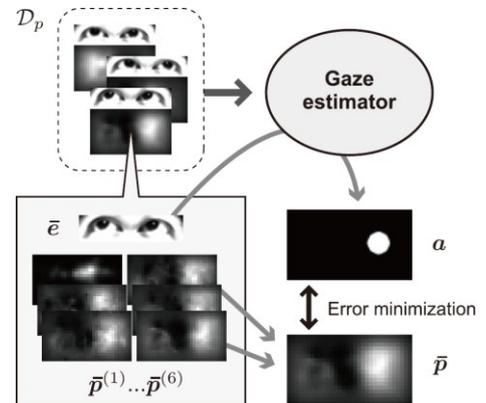


図7 注視点推定結果の顕著性モデルへのフィードバック

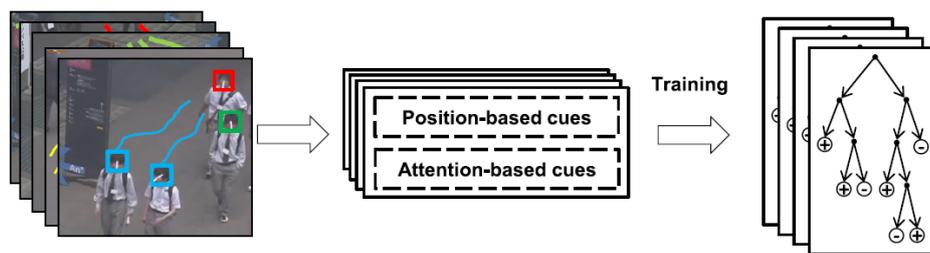


図8 注視点行動履歴に基づくグループ検出

また、上記研究項目「環境埋め込み型カメラによる視線計測技術の開発」では、低解像度画像からの顔向き推定技術に取り組んだが、これまでに得られた顔向き推定技術を元に、歩行者の共同・相互注視を手掛かりとしたグループ検出手法の開発を行った。歩行者間の注視点行動や環境自身の誘目性を手掛かりに空間全体の注視・注意マップ及び歩行者行動理解を行うことは、空間理解の観点から重要なタスクとなる。本研究では、ある二人の歩行者が同一のグループに属しているかを識別するタスクに対して、図8のように二つの特徴を用いる。一つは従来研究でも用いられている、歩行軌跡に関する特徴であり、もう一つが我々の提案する、注視点履歴に関する特徴である。具体的には、過去の顔向き推定履歴を元に、二者の注視点方向の差(=共同注視)および互いの位置関係と注視点方向のずれ(=相互注視)のヒストグラムを算出する。進行速度や進行方向と顔向きとのずれなどと共にこれらの特徴を連結し、Random Forest アルゴリズムによるグループ識別機の学習を行った。複数のデータセットを用いた実験を通して、歩行軌跡のみを特徴として用いる手法よりも優位な性能が得られることが確認できた。

3.3 視覚刺激に対する注視反応モデルの構築(佐藤グループ、杉本グループ)

視覚刺激に対する注視引付けのメカニズムを検討し、新たな計算モデルを提案する。視覚的顕著性マップの高精度な計算モデルを構築することで他研究項目における注視推定の性能向上を目指すと同時に、認知科学や心理学の専門家の協力を仰ぎながら、人間の視覚的注意機構に対する理解を深めるための基礎的貢献を目指す。当初の研究計画から研究ターゲットを更新した本研究項目であるが、従来研究では扱われていなかった新たなモデルを提案することに成功した。

人間の注視行動を推定することの重要性は、近年様々な応用が期待される一人称視点映像においても同様に大きい。しかし、視覚的顕著性モデルに関する従来研究では、ほとんどの場合人間が受動的にディスプレイ

上に提示された動画像を見るような環境下で注視予測性能の評価を行っていた。撮影者の頭部に固定されたカメラによって得られる一人称視点映像において、カメラの動きは撮影者本人の頭部移動によって生じるため、映像の動的変化と撮影者本人の注視との関係は上記のような受動的な条件とは大きく異なる。

本研究では、一人称視点映像のための新たな注視予測モデルとして、撮影者の自己運動を用いる手法を提案した。上述の通り、一人称視点映像におけるカメラの動きは撮影者の顔向きの変化に対応し、これ自体が注視行動に深く関連している。図 9 に示すように、提案手法では自己運動を元に注視の空間的なバイアスを予測し、これを視覚的顕著性マップに足し合わせることでより一人称視点映像に適した視覚的注視推定を実現する。これらの効果は従来研究で実験的には示唆されていたが、本研究では一人称映像のみを用いて予測した自己運動により注視マップを計算し、特に回転マップを用いた場合に実際の注視点との一致精度が大きく向上することを確認した。

また、人物の注視停留点の分布は視覚的顕著性マップの計算モデルによってある程度精度良く予測できることが報告されてきたが、視線移動を予測することは困難であることが指摘されている。既存のモデルは、そのほとんどが視覚刺激を空間的に一様なものとして扱っていたが、人間の視細胞は網膜全体に不均一に分布しており、網膜部位の中心窩からの距離に応じて異なる特性を持つことが知られている。こうした特性の違いが視覚的注意のメカニズムに深く関連することは従来研究でも実験的に確認されており、視野特性を考慮した視覚的顕著性モデルを構築することで視線移動の予測精度が向上する可能性がある。本研究では、実際に計測した人間の視線移動の大規模なデータベースを用いた学習アプローチによりこれを実証し、視野特性と視覚的顕著性との

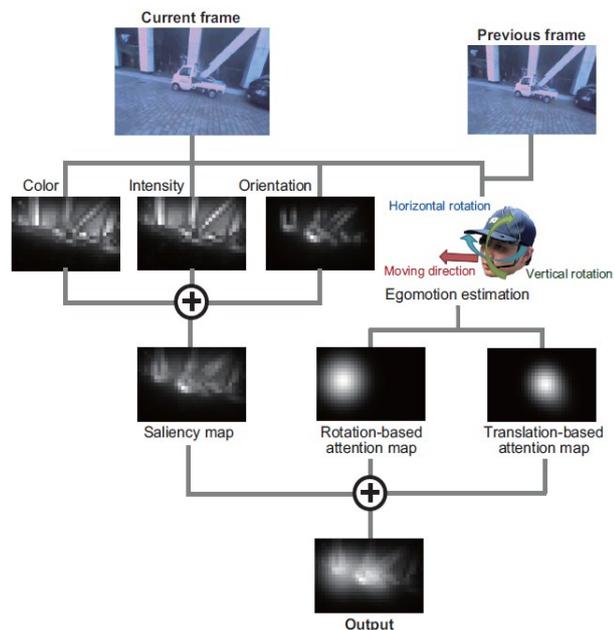


図 9 視覚的顕著性と自己運動による注意推定

関わりを明らかにした。

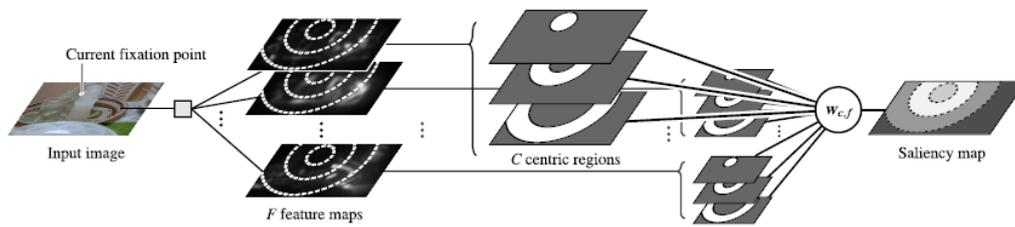


図 10 視野特性を考慮した視覚的顕著性モデル

図 10 に示すとおり、提案モデルでは、中心窩からの距離と方向に応じた特性の違いを捉えるために、各特徴マップを現在の注視点を中心とした円環領域に分割する。最終的な顕著性マップはこれらの分割マップを線形統合することによって得られるが、各領域に対する重みはそれぞれ異なるものとする。この統合重みを変化させることによって、人間の視野特性を顕著性マップに組み込むことが可能となる。しかし、特徴毎に異なる視野特性と視覚的顕著性の関係を全て明示するのは困難であるため、本研究では画像データセットに対する実際の人間の視線移動分布を元に統合重みを最適化する。すなわち、多数の自然画像と各画像に対する実際の視線移動計測データが与えられたとき、最終的な顕著性マップが視線移動の終点にピークを持つようなマップに最も近づく統合重みを算出することで視野特性をモデル化している。

このようなモデルを用いることで、視線移動予測性能が向上することが実験的にも確認できた。さらに、最適化された統合重みは各特徴が距離・方向に応じて人間の注視行動に与える影響の強さに対応するため、従来のアプローチでは解明が困難な視野特性と視覚的顕著性の関係を定量的に示すことができた。

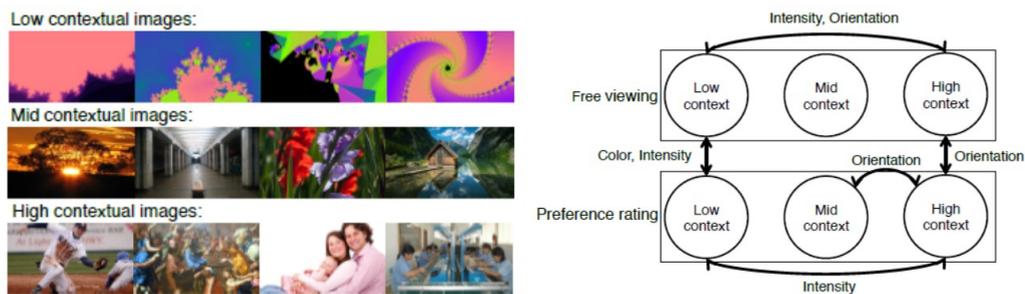


図 11 画像および閲覧タスクと視覚的顕著性モデルの関係性

さらに、医療診断応用等への可能性やモデルの高性能化の可能性を踏まえ、属性の異なる画像・人物間で視覚的顕著性モデルにどのような違いがあるかを検証した。具体的には、人工的に生成した意味のない画像から人物のインタラクションを含む高次の意味を持つ画像まで、コンテキストの強さが異なる三段階の画像データセットを準備し、さらに 3 セットそれぞれを純粋な自由閲覧タスク下と各画像の好み度合を一枚ごとに評価する選好評価タスク下の二つの条件下に分け、12

人の被験者に提示しながら視線計測を行った。これら 3 コンテキストレベル×2 タスクそれぞれのデータセットからそれぞれ視覚的顕著性モデルを学習した時、仮にこれらの要因が顕著性の傾向に影響を与えるならば、学習されたモデル自体にも統計的に有意な差が生じることになる。図 11 の矢印はモデル間の統計的検定を行った際に有意差が生じた対を表現しており、タスクと画像コンテキスト固有の差がモデルに生じていることがわかる。これらの研究課題に用いたデータセットは、従来研究で公開されてきたデータセットと比べても独自のものであり、これらを一般公開することで更なる調査・研究を促進することが期待できる。

一方、映像には通常、音響情報も付与されているが、従来の視覚的顕著性マップの計算では、音響情報は破棄され、画像情報のみが用いられてきた。しかし、人がどこに視覚的注意を向けているかは、視覚刺激のみではなく、聴覚刺激にも依存している。そこで、映像から得られる音響信号を考慮した顕著性マップの構築を提案した。提案手法では、時系列画像のフレームに対して、どこからどれだけ顕著な音響が出ているのかを表す音源特徴を定義し、これを新たに視覚的顕著性マップに組み込む。音源特徴は、空間的な音源位置の情報を時

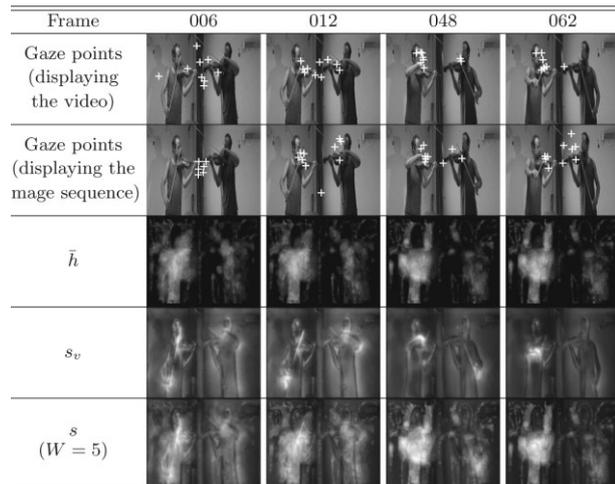


図 12 従来の視覚的顕著性マップと音響信号を考慮した視聴覚顕著性マップの比較

間的に変化する聴覚的顕著性によって重み付けたものである。提案手法では、輝度、色彩、方向、点滅、運動の画像情報から得られる画像・時系列画像特徴に、音源特徴を加えた 6 つの特徴から、顕著性マップを作成する。音源特徴を定義し組み込むことにより、既存の視覚的顕著性マップでは破棄されていた音情報を顕著性マップに反映させることができる。被験者による実際の視線データに基づく比較実験を通して、聴覚刺激の影響を音源特徴という形で取り入れることにより、顕著性マップ上に反映させることができ、その性能を向上させることができることを確認した(図 12)。

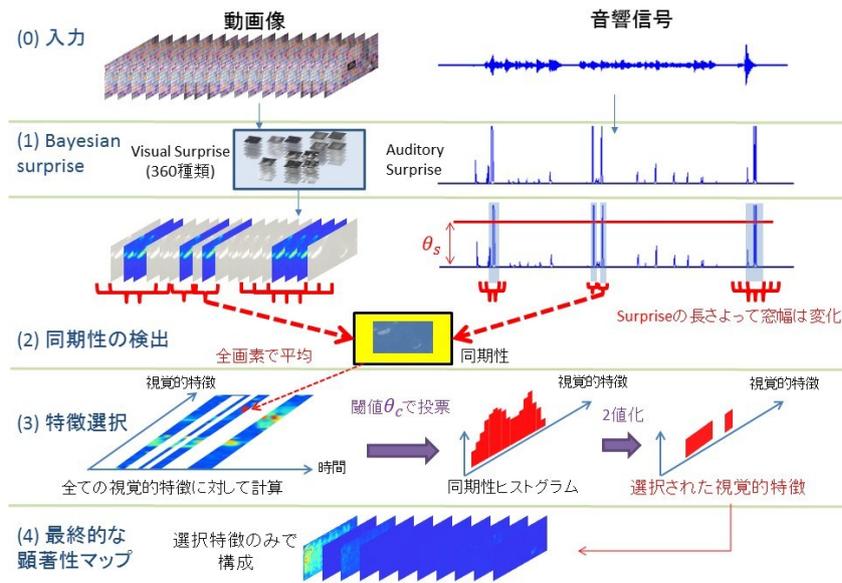


図 13 視覚情報と聴覚情報の同期に基づく顕著性マップモデル

一方、視覚情報と聴覚情報の時間的な同期が視覚的注意に強い影響を与えるという心理物理学的知見がある。そこで、画像特徴と音響特徴を統合するのではなく、知覚的に顕著な聴覚的特徴と同期性の高い画像特徴を選択し、選択した画像特徴を強調し、そのみを用いた顕著性マップを構築する手法を考案した。ここでは、画像特徴の時間的変化をベイズ理論の枠組みで捉え顕著性を計算する Bayesian surprise を採用し、図 13 に

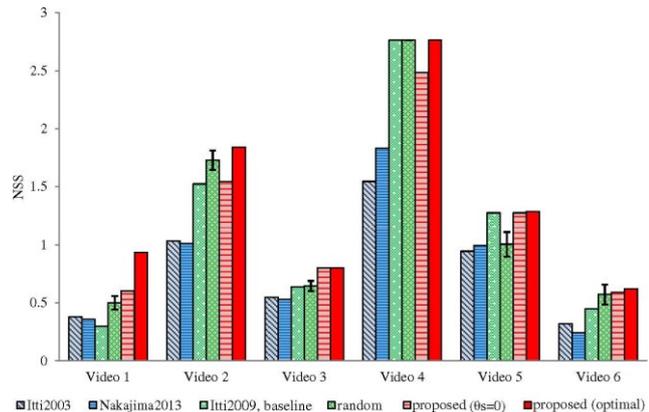


図 14 既存手法との性能比較

示すように4段階で構成されるモデルを構築した。提案手法における4ステップは、(1) Bayesian surprise モデルによる visual surprise マップの抽出、及び Bayesian surprise を音響信号に応用した auditory surprise signal の抽出、(2)聴覚的に顕著なイベントの区間に対し、visual surprise と auditory surprise の正規化相関計算に基づく同期性の検出、(3)聴覚イベントと同期性の高い視覚的特徴の選択、(4)選択された特徴のみを用いた顕著性マップの作成、である。動画データセットを用いた比較実験により、提案手法は顕著性マップを計算するために用いる画像特徴の数が少ないにもかかわらず、従来手法より高精度に注視推定を実現していることが確認できた(図 14)。

3.4 ロボットの身体動作に対する注視反応モデルの構築(久野グループ)

本研究項目では、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行った。この研究項目では、どのような状況のときにどのようなロボットの身体行動が人間の注視反応をもたらすかのモデルを検討した。

特定の人に必要なところを注視するように誘導するためには、一般には、まず、その人の注意をロボットの方に向けさせ(注視をロボットに向けさせ)、アイコンタクトにより互いの間のコミュニケーションチャンネルを確立し、そして共同注視を実現する(ロボットが相手の注視を向けさせたいものを見ると相手もそれを見る)の3段階が必要である。ロボットがどのような身体動作でこれを行うことができるかは、いろいろな要因で変わってくる。ロボット側で決まってくるものとしては、この注視誘導の緊急度がある。すぐに相手の注視を向けさせなければならないような状況、例えば、相手に危険が迫っているような状況では、この研究ではできるだけ特定の相手だけの注視を誘導し、他の人には影響を与えないことを考えているが、そのようなことは言っていない。声や音を出して、

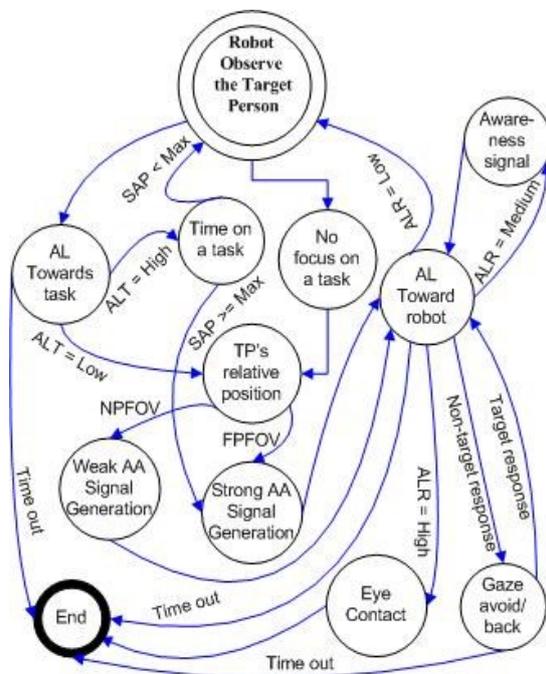


図 15 注視獲得のモデル

相手の注意を引く必要がある。これは当然のことなので、今回の研究ではこのような状況は考えない。ロボットにはある程度の時間の余裕があって、適切なタイミングで、できるだけ関係ない人には影響を与えない行動で、注視を誘導することを考える。発話も身体行動と言えれば身体行動であるが、ここでは音を出さない身体の動き、特に頭部と視線の動きを中心に検討した。

時間に余裕がある状況も2つに分けられる。時間に非常に余裕がある場合は、ロボットは積極的な行動をせずに相手がロボットの方に向かって視線を向けたときを捉えて、アイコンタクトから共同注視へと向かえばよい。ここではこれを reactive な方法と呼ぶ。時間に余裕があるといっても、ある程度の時間以内に注視を誘導したい場合は、ロボットの方から注意をロボットに向けさせるための行動を積極的に行わなければならない。ここではこれを proactive な方法と呼ぶ。reactive な方法は先に述べた3段階の最初の段階を成り行きにまかせるということなので、モデルの検討では proactive な方法を中心に考える。ただし、reactive な方法の実現には、技術的には相手が注意をロボットに向けたときを逃さずに捉えることを検討する必要がある。

検討したモデルを図 15 に示す。これはロボットへの注視の獲得を中心に示したものである。図にはアイコンタクトの段階が一つのノードとして示されているが、このアイコンタクトが成立した場合、その後、ロボットが視線を対象物に向ければ、人間がそちらを向くことが期待でき、注視誘導が完了する。

人間の注意の状態はさまざまな要因が関連する。それらを網羅的に検討することはできないの

で、このモデルではロボットから観察・推察可能で、ロボットが注視を誘導するのに大きく関わると思われるものを中心に考えた。具体的には、人間がどのような作業中で、その作業にどれくらい注意を向けているかと人間とロボットの位置関係を中心に考えた。後者は、ロボットが人間の視野のどこに捉えられているかを表している。図中の NPFOV はロボットが近周辺視野、FPFOV は遠周辺視野にいる場合である。ロボットが中心視野に入っている場合は、先に述べた reactive な方法の場合と同様の状況で、アイコンタクトの成立を図るところから考えればよいので、この図には示していない。ロボットが行う身体動作は、関係のない人の注意を引くのを避けるために、可能な限り弱い動作(weak AA(attention attraction) signal)を用いる。ここでは対象の人に顔を向ける動作を用いた。それでは不十分な状況ではより強い動作を用いる。一般にはある程度継続した動きが注意を引くことが分かっている。実験では頭部だけのロボットを用いたので、モデルの検証の実験では、ロボットが頭を振る動作を用いた。

ロボットは対象となる人間 TP を観察し、何をしているか認識する。特に特定の仕事をしていなければ、位置関係を調べ、NPFOV なら弱い動作、FPFOV なら強い動作で人間の注意の獲得を図る。ロボットはそれにより人間のロボットへの注意レベル ALR がどうなったかを調べる。人間がロボットの方を向いてきていれば ALR は中程度と考え、顔を上げて相手と正対するようにする。これで相手が目をそらさなければ ALR は高いと判断し、アイコンタクトを行う。なお、実験では、アイコンタクトは目が合ったならば、瞬きを少しして相手の視線に気づいたことを示すことにより完成するようにしている。心理学では、アイコンタクトには gaze crossing と gaze awareness が必要であるとされているので、瞬きにより gaze awareness を明示的に示すようにした。なお、対象となる人間以外の人々がロボットを見てきたときは、視線をそらせて、その人とのコミュニケーションの意思がないことを示すようにしている。

人間が読書など何らかの作業をしている場合は、ロボットはできるだけ、その人の作業を邪魔しないで注意を獲得することを目指す。人間がいろいろな作業をしているところをビデオ撮影した結果、人間は一つの作業にずっと集中していられるわけではなく、ある程度時間がたつと、視線を作業対象からはずして注意レベルが下がることが分かった。注意が集中していられる最長の時間は人や状況により異なるが、平均的な時間を観察実験から定めた(図 15 中の Max)。その時間までは、人間の作業への注意レベルを認識し、そのレベルが低くなったときに、位置関係に応じて注意獲得の動作を行う。最長時間を過ぎてもそのようなときがなかった場合は、強い動作を用いて注意獲得に向かう。

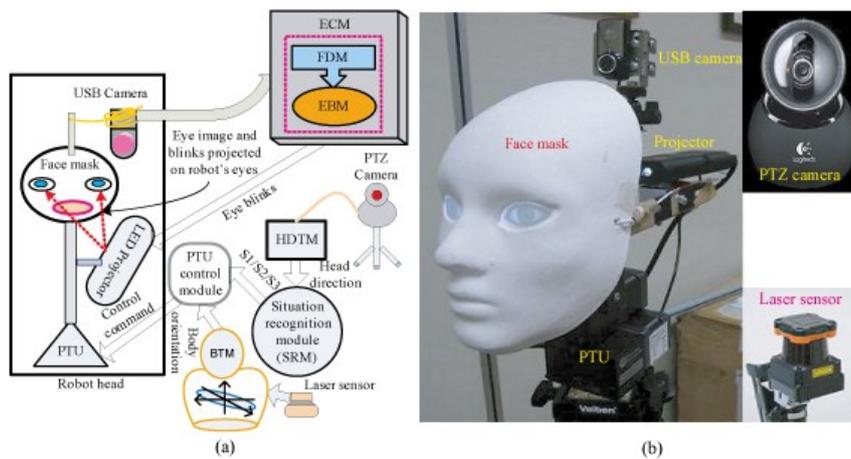


図 16 実験用ロボット頭部

さらに、図 16 に示すようなロボット頭部を開発し、被験者を用いた実験でモデルの有効性を検証した。ロボット頭部は小型プロジェクタで目の CG 像を投影することにより図 17 のように視線を動かしたり瞬いたりすることができるほか、頭部全体はパンチルトユニットにより動かすことができる。レーザーレンジファインダで人の位置を求め(図 17 左上)、そちらに PTZ(pan-tilt-zoom)カメラを向け、その画像から顔の向きなどを求める。また、人間がしている作業を認識することができる。

このロボットを用いて、モデルの各部分ごと、また全体について被験者を用いた実験を行った。図 18 は全体についての実験結果の一部である。実験の詳細は割愛するが、3 つのロボットが 20 人の被験者に対して 3 回ずつ視線誘導を試みた結果である。図の縦軸は 3 回の試行のうち視線誘導に成功した平均回数を示している。AC-robot が提案のモデルに基づくロボットである。HT-robot は弱い行動しか取らないロボット、HS-robot は提案のロボットとほぼ同じだが、アイコンタクトのときに瞬きで gaze awareness を明示的に示さないロボットである。実験結果は提案のモデルに基づくロボットが他より有意に好成績で、80%以上の場合、視線誘導が可能であったことを示しており、提案のモデルの有効性が確認できた。

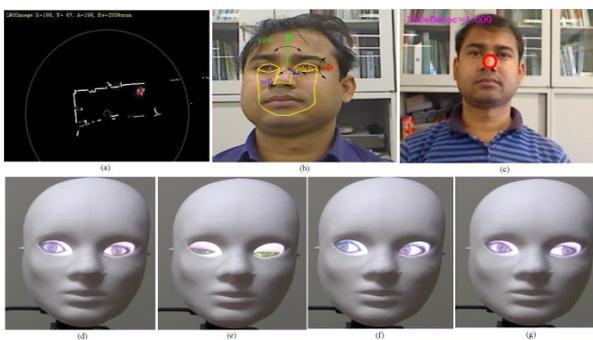


図 17 顔検出と目の表示

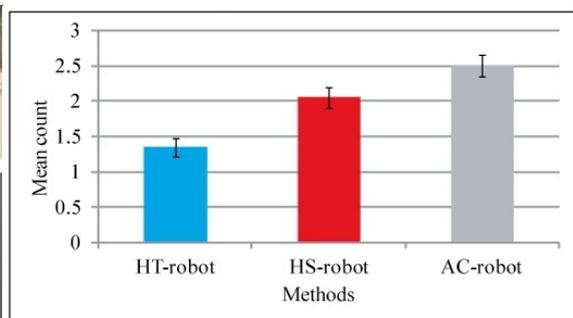


図 18 視線誘導実験結果

3.5 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(久野グループ)

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」で構築したモデルに基づいて動作するロボットを開発し、有効性を実証した。はじめに注視誘導に適したロボットのみおよび頭部について検討し、その結果に基づいてロボット頭部を開発した。そして、それを用いて注視誘導を行うロボットを検討した。

前の研究項目で述べたように、ロボットが特定の人に必要などころを注視するように誘導するためには、アイコンタクトや共同注視を人とロボットの間で行えることが必要である。アイコンタクトや共同注視を実現するためには、ロボットが何を見ているかが分かるように、ロボットの視線の読み取りやすさが重要である。さらに、日常生活で人と共存するロボットでは、外見的な親しみやすさも重要である。特に目の外観は、ロボットの印象を大きく左右するため、視線の読み取りやすさと親しみやすさを兼ね備えたロボットの目のデザインについて検討した。

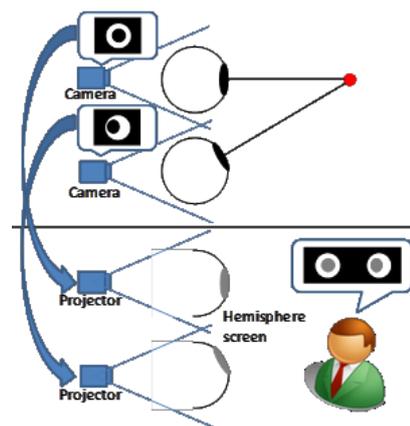


図 19 目の表示方法

その検討のために、プロジェクタで目の CG 像を投影することにより、様々な目を表現できる方法を開発した。近年、プロジェクタで顔を投影するタイプのロボットがいくつか提案されているが、それらは、正確な視線を表現するような配慮がなされたものではなかった。視線方向は、一般に黒目の中心の法線方向であるが、顔や目を非平面スクリーンに投影する場合には、投影面の形状を幾何学的に考慮して像を作成しないと、黒目の形状が歪むなど、視線方向を正確に表現できない。そこで、スクリーンに半球を用い、そこに CG で作成した黒目を投影することで、眼球を表現することとした。こうすることで、CG の世界で、ある一点を注視した眼球を、そのままの形状、見た目で実世界に出現させることができる。さらに、CG の世界と実世界の座標系を対応づけておけば、実世界の注目点を正確に向いた眼球を描画し、スクリーンに投影することができる(図 19)。実際に、予備実験では、非球面スクリーンに投影するよりも球面スクリーンに投影した方が、視線の読み取り誤差が少なかった。

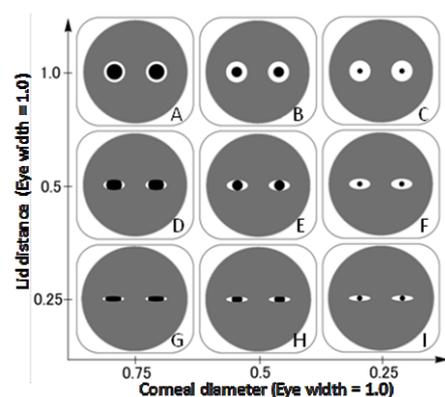


図 20 目のデザインの候補

次に、実際に視線の読み取りやすさと親しみやすさを兼ね備えたロボットの目のデザインについて検討を行った。視線の読み取りやすさには、目の外部形状と黒目の位置の関係が大きく影響することが、認知科学の分野において明らかになっている。そこで、目の横幅を基準に縦幅を 1、0.5、0.25 倍に変化させた外部形状と、目の横幅を基準に直径を 0.75、0.5、0.25 倍に変化させた黒目を作成し、図 20 のようにロボットの目のデザイン案を 9 種類作成した。これを用いて、視線の読み取りやすさと親しみやすさの調査を行った。

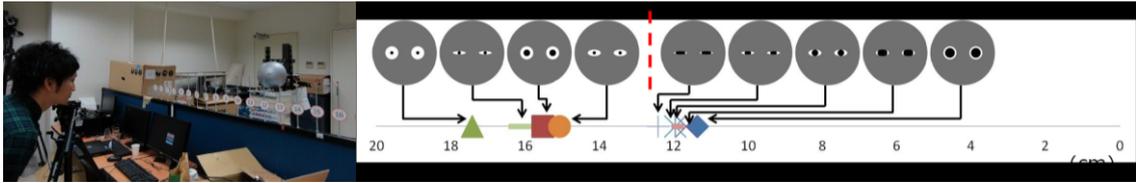


図 21 視線の読み取り実験の様子とその結果



図 22 目の親しみやすさの一对比較法による評価

視線の読み取りやすさの調査は、図 21(左)に示すようにロボットと被験者の間にマーカーを設置し、ロボットの目がどのマーカーを見ているかを被験者に答えてもらう形式で行った。図 20 に示した 9 種の目ごとに正解値からの誤差をとり、平均値をとったところ、図 21 (右)のような結果となった。図 21 (右)を見ると、破線の部分で平均誤差の値に隔たりがあり、目のデザインを平均誤差の大きなグループと小さなグループに分けることができる。そこで、黒目の面積と白目の面積を調査したところ、平均誤差の小さなグループは黒目の面積と白目の面積が同じ、または黒目の面積の方が白目の面積より大きいものだった。平均誤差の大きなグループはその反対に、黒目の面積が白目の面積より小さくなっていった。このことから、黒目の面積と白目の面積が同じ、または黒目の面積の方が白目の面積より大きい目が視線方向を読み取りやすいことが分かった。さらに、正面からマーカーまで、目を動かした場合の実験を行ったところ、目を動かした方が視線読み取り誤差が小さくなることが分かった。

さらに、9 種の目のデザインの親しみやすさを調べるため、サーストンの一对比較法による評価実験を行った。実験には、ランダムに提示される一对のロボットの目の画像のどちらかを選択できるインタフェースを備えた Web System を作成して用いた。被験者には、図 22 (左)のような入力端末を操作することで回答してもらった。「どちらのロボットの顔が親しみやすいですか?」という問いに対して、20 歳前後の男女 105 人の日本人被験者に回答してもらった。Case III の構成尺度によりグラフ化した結果が図 22 (右)である。図 22 (右)を見ると、上位 2 つは目の外部形状が丸のものであり、その次には目の外部形状が中くらいのもので 2 つ続いている。この結果を踏まえ、上位 4 つの白目の面積と黒目の面積を比較すると、図 20 の B を除き白目より黒目の方が面積が大きいという結果となった。この結果より親しみやすさを感じる要因として、1 つ目の要素として目の縦幅があり、2 つ目の要素として白目に対する黒目の面積比が考えられる。今回の評価結果からは、目の縦幅が大きく、黒目の面積と白目の面積が同じ、または黒目の面積が白目の面積より大きい目が親しみやすいと考えられる。

以上のように目についての知見は得られたが、人間は概略の視線として頭部の向きの情報を利用しているので、頭部形状についても検討を行った。図 23 のように(a)平面、(b)半球面、(c)半球面に鼻のような突起があるという 3 種類の頭部を用いて視線読み取りの実験を行った。目の形状の場合と同様に、ロボットの前方にマーカーを並べ、被験者にロボットがどのマーカーの方を向いて

いるかを答えてもらった。16人の被験者による実験の結果、(c)、(a)、(b)の順に視線方向が読み取り易いという統計的に有意な結果が得られた。



図 23 視線読み取り実験に用いた 3 種類の顔

さらに、頭部や目の動かし方についても検討を行った。頭部について

は、人間が振り向くときの動作を調べた結果、等速ではなく、遅い・速い・遅いというふうに加速度運動をしていることが分かった。ロボットの頭部もこのように動かした方が、等速で動かすよりも自然な感じを与えることが分かった。また、ある方向へ視線を向けるのに、頭部と目の両者を動かすことが、頭部だけを動かすより自然に感じられることが分かった。以上の結果をもとに、図 24 のように、鼻のような突起があり、目形状も実験により良いとされたものをもつロボット頭部を製作した。

さらに、このようなロボットの実際の応用の例としてミュージアムのガイドロボットを想定し、効果的に人間にとっても感じの良い注意獲得、コミュニケーションの開始法を検討した。人間の視線のパターンには、spontaneous looking、scene-relevant looking、orientation of thought looking の 3 種類がある。これらは、特に目的があるのではなくシーンを見ている、何か特定の意図があって見ている、何か考えていて特にものは見ていないという3つの状況に対応する。今回は最後のパターンは除外して、初めの2つのパターンだけを考えることにした。図 25(a)はディスプレイを見ている scene-relevant looking のときの視線の動きであり、(b)は特に何かを見ている場合ではない



図 24 開発したロボット頭部

spontaneous looking のときの視線の動きである。ミュージアムでは、訪問者が特定の展示物に興味をもって見ているときには scene-relevant looking が、そういう展示物が特にまだないときには spontaneous looking が示されると考えられる。ガイドロボットとしては、scene-relevant looking を示す訪問者を見つけたら、その展示物のガイドをすることを申し出ることが考えられる。そこで、画像から視線を求め、そのパターンを認識する方法を開発した。そして、それをを用い、ある展示に興味をもつような人を見つけると、そこに行き、ガイドを申し出ることのできるロボットを実現した。実際の美術館の一室を模した環境で実験を行い、絵の説明をして欲しそうな人を見つけ、ロボットが説明することができることを確認した。



図 25 視線のパターン (a) Scene-relevant looking, (b) Spontaneous looking

3.6 視覚刺激による注視誘導技術の開発(佐藤グループ、杉本グループ、小池グループ)

人が注視する対象を選択する主な要因は、ユーザの興味や意図であると考えられるが、必ずしもそれのみではない。見る対象の様子によっても注視のしやすさは変わってくる。そこで、視覚的顕著性を踏まえて、人に注視させたい対象の色や明るさを適切に変化させることにより、さりげなく人の視線をひきつける技術の構築を目指した。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入力画像を加工することによって、指定領域に注視を誘導する手法を開発した。

提案手法では入力画像に対して、視覚的顕著性が指定領域内で高くなり、領域外で低くなるように明るさと色を調整することを繰り返す。顕著性の値を制御するために、各ピクセルに対して、そのピクセルの顕著性の値に貢献したスケールを適応的に選択し、さらに顕著性の値への各特徴量の貢献の割合を計算し、それに応じて各ピクセルの特徴量の強度を変更した。また、特徴量の強度変化をそのピクセルの顕著性の値に応じて重みづけした。これによって、顕著性の値を効率的に更新することができる。しかし、入力画像の各ピクセルの色特徴を独立に変更する操作によって加工画像と入力画像との間に違和感が生じるため、入力画像と加工画像との間で色ヒストグラムになるべく差が出ないように輝度値の補正を画像全体で一括して行う機能を組み込んだ。これらの操作を、指定領域の顕著性の値が最も大きくなるまで繰り返すことによって、指定領域に注視を自然に誘導することができる。



図 26 交通事故多発地点の画像

提案手法による画像の加工例を図 26 に示す。それぞれ、入力画像、指定領域の顕著性を高めるため加工を施した画像、加工前後の画像に対する顕著性マップを示してある。また、輝度値の補正の効果を見るために、補正なしの加工画像とそれに対応する顕著性マップも示してある。これらは広島県警察¹による事故多発地点での画像であり、入力画像の指定領域を運転者が注視していないことが原因で事故が発生したとされている。いずれの画像でも指定領域の顕著性の値が最も高くなる画像を生成できていることが見て取れる。これらの画像を用いて被験者による注視実験を行った結果を図 27 に示す。指定領域における注視発生が加工前後で優位な差となっていることがわかる。図 28 は、加工前後での輝度差を示すが、輝度補正によって入力画像と加工画像との輝度差が大きく軽減できていることが読み取れる。これは、誘導率はほとんど同じであるにもかかわらず、輝度値の補正によって加工による違和感が軽減されていること、すなわち、加工がより目立たなくなっていることを意味する。これらの結果から、視覚的顕著性を用いた画像加工によって注視を指定領域に誘導可能であることが確認された。

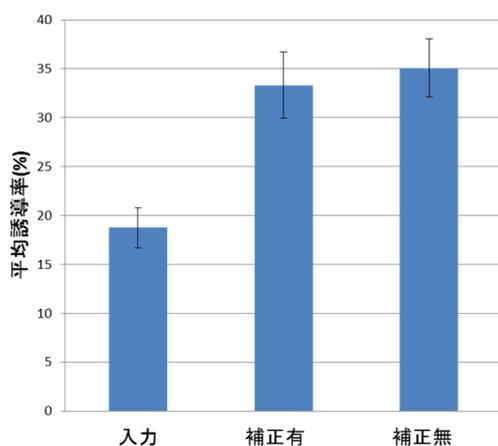


図 27 被験者を用いた注視実験結果

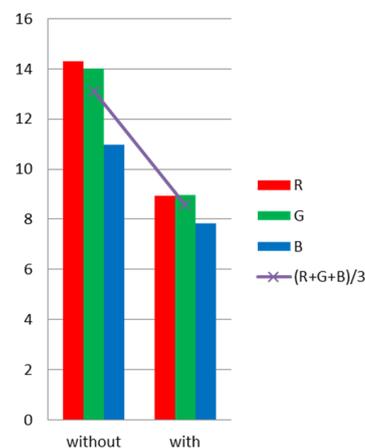


図 28 加工前後の輝度差

さらに、この手法を動画にも適用できるように拡張した。動画を対象とした場合は、指定した対象領域(物体)が、時間とともに画像中の位置が移動するという問題が発生する。これに対して、有効な物体追跡技術として知られている Tracking-Learning-Detection (TLD)を採用し、最初のフレームで領域を指定すれば、以後のフレームでも安定にその領域を自動的に検出することができるようにした。フレーム単位で独立に画像加工を行うと、隣接フレームにおける加工の度合いの違いから、加工後の動画にはちらつきが出る。これを抑えるため、各フレームにおける加工を平滑化する処理を導入した。具体的には、各フレームにおける指定領域の境界をガウスクーネルによって目立たなくする処理、領域内の顕著性を高める重みと領域外の顕著性を下げる重みを隣接フレームにおいて滑らかに変化させる処理、を導入した。さらに、加工によって生じる動画像の違和感を軽減するために、人の知覚により近いとされている HIS 空間による色表現を採用し、加工はこの空間内で行った。

¹ <http://www.pref.hiroshima.la.jp/site/police>

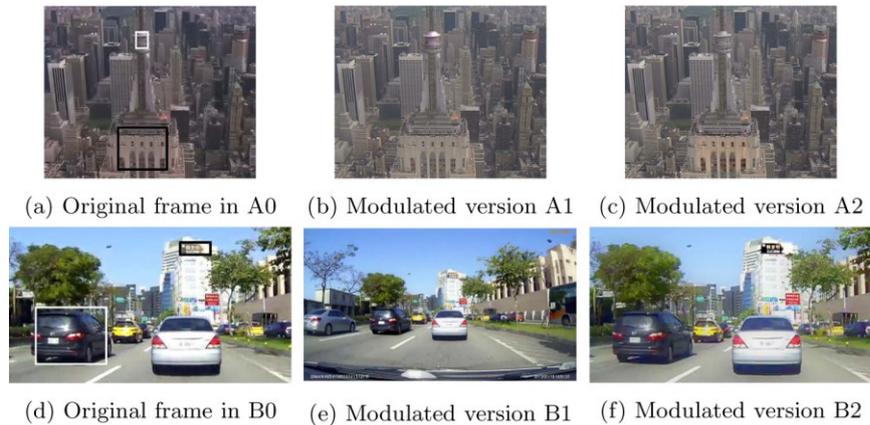


図 29 実験に用いた動画のフレーム例

構築した手法の有効性を確認するために被験者 15 名による実験を行った。実験では、2 種類の動画を用い、それぞれに対して 2 箇所の異なる領域を指定して加工した動画を作成し、オリジナルを含め、計 6 種類の動画を被験者に提示した(図 29)。また、被験者による注視点と指定領域との距離のヒートマップの例を図 30 に示す。図 30 はオリジナル動画Bと黒いバンの顕著性が上がるように加工した動画B1のヒートマップの結果である。この例の場合、オリジナルの動画で黒いバンがもともと注視をひきつけやすかったことが読み取れるが、加工によって、さらに注視がひきつけられていることがわかる。動画提示中に発生した全注視に対して、指定領域内に発生した注視の割合を求めたものが表 1 である。いずれの加工の場合でも、オリジナルの動画に比べ、加工動画の方が指定領域内で発生した注視の割合が格段に上昇していることがわかる。以上のように、提案手法によって動画に対しても、初期フレームにおいて領域を指定さえすれば、その領域を自動で追尾し、その領域の顕著性が高まるように動画を加工することで視線を誘導することが可能であることが確認された。

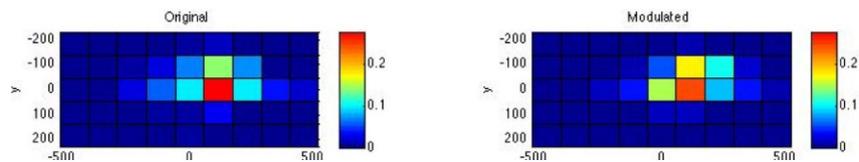


図 30 15名の被験者による注視点と指定領域との距離のヒートマップ

表 1 全注視に対して注視が指定した領域で発生した割合

Video	A1	A2	B1	B2
Original	0.0188	0.0350	0.0989	0.0076
Modulated	0.0813	0.1020	0.1497	0.0594

さらに、本項目では画像解像度に着目した注視誘導技術の開発も行った。知覚心理学の分野における知見の1つとして被写界深度の変化による人の注視点の変化が指摘されている。そこで本研究では、画像内に高解像度部分と低解像度部分を動的に生成し、被験者の注視点が高解像度部分に誘導されるかどうかの実験を行った。

図 31 に実験環境を示す。被験者はあご台にあごを固定し、約 60cm 離れた位置にある 23 インチディスプレイを見る。ディスプレイ下部には視線測定器(Tobii TX300)が設置され、被験者の視線を 300Hz で計測する。この状態で、被験者に 100 枚の画像を連続して提示した。まず、被験者の視線を画像中心に置くために、黒背景の中心に白い十字カーソルの画像を2秒間提示した。次に画像を提示するが画像1枚の提示時間は 10 秒である。1枚の提示が終了すると被験者は画像の印象を(良い、普通、悪い)で評価する。これは被験者の眠気を防止するために行われた。その後、最初の十字カーソル画像を提示して被験者の視線を再び中心に戻す。これを繰り返し行い、100 枚の画像を提示した。

画像セットとしては2セット準備した。1つ目の画像セットは高解像度部分を含まない画像である。これは原画像に対して標準偏差 $\sigma=\{7.0, 9.0\}$ 、カーネルサイズ 13x13 でガウシアンフィルタを適用することで得られたものである。もう1つの画像セットは、高解像度部分を含む画像を 100 枚中に 40 枚含むものである。残り 60 枚は最初の画像セット同様ガウシアンフィルタを適用しただけの画像で、これは実験の意図を知られないためのダミー画像である。

実験結果の一部を図 32 に示す。左図は高解像度部分のない画像、右図は一部高解像度部分(図中左のヒートマップが赤い部分)のある画像である。図には被験者の視線の滞留度を示すヒートマップがオーバーレイされている。これを見ると、高解像度のある部分には被験者の視線が滞留していることがわかる。他の図でも同様の結果が得られた。この実験から、画像中に解像度の高い部分と低い部分を作ると、高解像度の部分に視線が誘導されることがわかった。

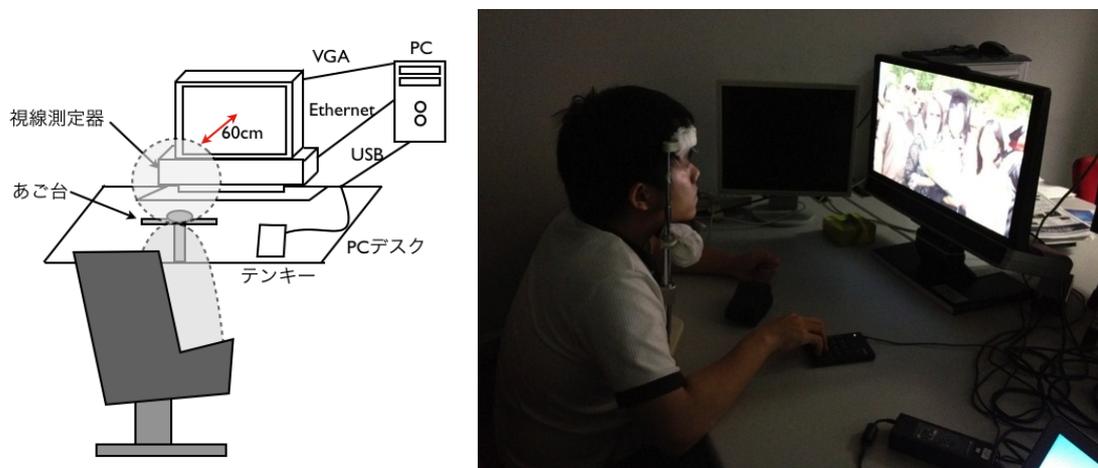


図 31 実験環境



図 32 高解像度部分のない画像とある画像における視線滞留分布の比較



図 33 輝度変調による注視誘導の例（左・変調前、右・変調後）

また、本研究項目で検討したような微かな視覚刺激による視線誘導に関しては、その「さりげなさ」をどのように評価するかが重要な課題となるが、視覚的顕著性による視覚刺激の強度評価は、誘導のさりげなさを予測する上でも重要な尺度となることが予想される。そこで、車載映像を用いた輝度変調による注視誘導の実験(図 33)において、実際に目的の箇所に注視が誘導されたかどうかだけではなく、被験者が人工的な視覚刺激提示に気付いたかどうかを実験後のアンケートを通して確認し、視覚的顕著度との比較を行った。刺激の顕著度が高ければ高いほど被験者は刺激に気づき、注視しやすくなる傾向が確認され、顕著度を尺度として刺激の強度を最適化することの有効性が確認できた。一方、実験結果ではユーザに気づかれずにさりげなく注視を誘導するためには現在の注視位置からの距離など他の要因を考慮した適応的な制御を行う必要があることも示唆され、より柔軟な情報提示技術を実現するためには今後更に注視誘導モデルを高度化する必要があると言える。

3.7 視線適応型インタフェースの開発(小池グループ)

本研究では、公共空間に配置される大型のデジタルサイネージを対象とし、ユーザの位置や視線に適応した対話型ディスプレイシステムに関する研究を行った。

図 35 は開発した視線適応型壁ディスプレイシステムである。スクリーンには単焦点プロジェクタを用いて大画面画像(80 インチ)を提示する。スクリーン上部には USB カメラが設置され、スクリーン前方を撮影する。スクリーン前方にユーザが現れるとユーザの顔を認識し、3 次元頭部姿勢を推定・追跡する(図 36)。得られた顔の3次元の位置・姿勢情報からユーザのスクリーンからの距離と顔の向きを推定する。

スクリーンとユーザの距離はそのインタラクションの形態に大きな影響を及ぼす。つまり、スクリーンからある程度離れているユーザは、スクリーン上に提示される情報を単純に見るという比較的受動的なインタラクションしか行わない。これに対し、スクリーンに近いユーザは、提示される情報のより詳細な情報を得ようとして、能動的なインタラクションを行う。本システムではユーザのスクリーンからの距離に応じて、3つのインタラクションレベルを設定した。第1は距離が遠い場合で、この時、画面には3x3のマトリクス状に本システムが持つ情報コンテンツの概観を提示する。第2は距離が中間的な場合で、システムはユーザの顔の向きを追跡し、上記 3x3 のマトリクスに提示されたコンテンツのうち、どれか1つへの滞留時間がある一定の閾値以上になると、そのコンテンツを拡大表示する。第3は距離が近い場合で、この時システムはユーザがさらなる詳細な情報を得たいと判断し、

当該コンテンツに関するより詳細な情報を提示する。図 34 にこれら3つのモードの遷移を示す。

顔の法線ベクトルの向きを利用したポインティングは必ずしも視線の方向と正確に一致しないが、視線計測装置を装着しなくても良いこと、顔の位置にある程度の自由度があることから、公共空間での使用を考慮した場合より実用的である。また今回は1ユーザのみへの対応であったが、多人数への拡張も可能である。

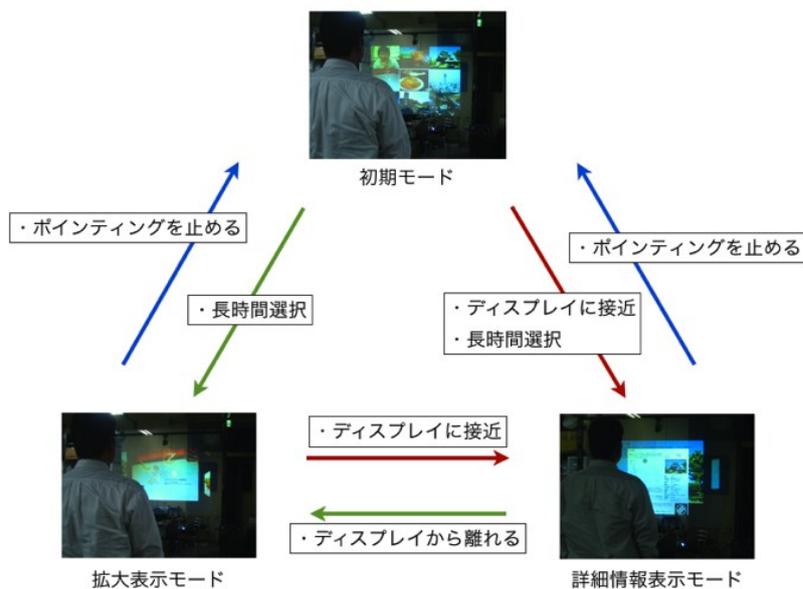


図 34 距離に基づく適応型情報表示

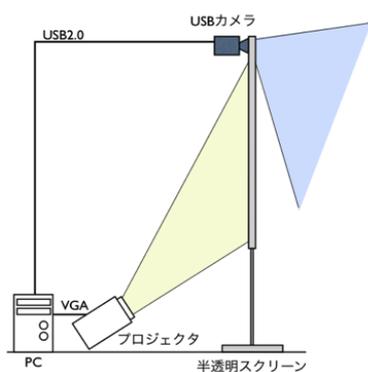


図 35 壁型ディスプレイのシステム構成



図 36 頭部姿勢の推定

一方、ディスプレイに表示される映像やスクリーンに投影される映像にはそのサイズに応じた最適な視聴距離がある。例えば、公共スペースに普及する大型のデジタルサイネージやポスターは比較的遠方にいる人々に情報を提供する一方、近距離にいる人々にとっては映像全体を把握することが難しい。一方、サイネージに表示された小さな文字や絵は、近距離の人々には認識できるが、遠方の人には認識が難しい。これに対する従来研究としては、Photomosaic のように大量の

画像で1枚の大きな画像を構成する手法や、Hybrid Images のように周波数の異なる2つの映像を重畳表示することで、遠距離と近距離の両方にそれぞれ異なる画像を提示するものがある。しかし、Photomosaic では近距離用画像が非常に小さく、画像のごく近くに寄らないとこれらを認識することができない。Hybrid Images では遠近両画像の形状が似ていることが必要で、かつカラー画像には対応できない。

以上の問題を解決するために、本研究ではモザイク画像の特徴を利用した遠近両画像の重畳手法を実現した(図 37)。基本的なアイデアは、まずモザイク化した遠画像の上に透過度を高めた複数の近画像を重畳表示する。この結果、遠近両画像の同時提示が可能となるが、この単純な手法の問題点は、近画像の透過度を高くすると、遠画像が見やすくなる一方、近画像が見づらくなり、逆に透過度を低くすると、近画像が見やすくなり遠画像が見づらくなる。この重畳画像の視認性を高めるために、我々は以下の視認性改善処理を行った。

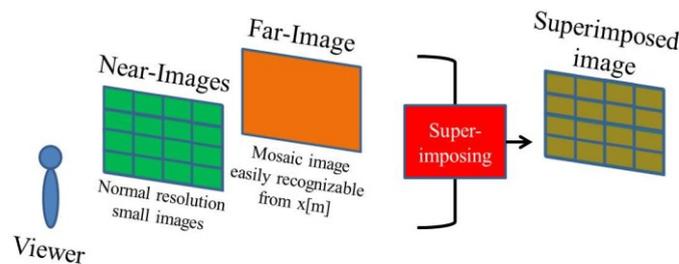


図 37 モザイクを利用した遠近画像の同時提示

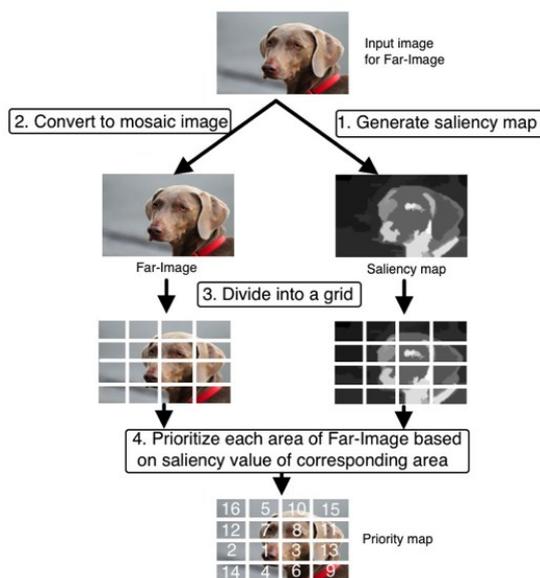


図 38 視覚的顕著性に基づく遠画像のモザイク各領域の優先度決定

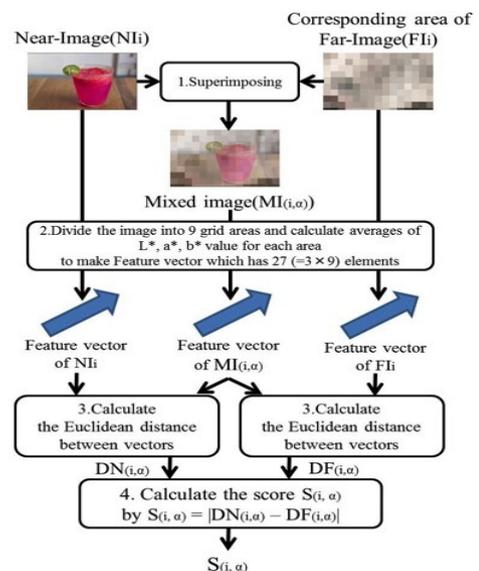


図 39 近画像の透過度の設定方法

第1に、遠画像から視覚的顕著性マップを生成し、モザイク化された各領域の優先度を顕著性に基づき計算した。そして、優先度の高い領域から順にその色特徴に近い近画像を配置していく。第2に、遠画像と近画像の両者ができるだけ高い視認性を得るために、近画像の透過度を自動計算する処理を行った。第3に、個々の近画像の顕著性マップを計算し、顕著性の高い部分の透過度を上げ、低い部分の透過度を下げた。第4に、近画像の高顕著性領域に対応する遠画像のモザイクタイルを平滑化し、色勾配を滑らかにする処理を行った。以上の視認性改善処理を行った結果、単純に透過度を変化させて重畳させた場合に比べ、遠近両画像の視認性が高まることが示された(図 40)。



図 40 遠近画像の重畳表示 (左：視認性改善処理前、右：視認性改善処理後)

§ 4 成果発表など

(1) 原著論文発表 (国内(和文)誌 6件、国際(欧文)誌 59件)

(国内)

- [1] 菅野裕介, 松下康之, 佐藤洋一, “視覚的顕著性を用いた自己校正による視線推定”, 電子情報通信学会論文誌 Vol.J94-D, No.8, August 2011.
- [2] 原田一馬, 菅野裕介, 佐藤洋一, “運動視差を用いた直感的なマルチタッチインタラクション”, 情報処理学会シンポジウム インタラクション 2012, March 2012.
- [3] 窪田秀行, 菅野裕介, 岡部孝弘, 佐藤洋一, 杉本晃宏, 開一夫, “人間の視野特性を考慮した学習に基づく視覚的顕著性モデル”, 画像の認識・理解シンポジウム (MIRU 2012), August 2012.
- [4] 鈴木裕太, 小池英樹, モザイク画像を利用した遠近画像の同時提示手法, 日本ソフトウェア科学会インタラクティブシステムとソフトウェア(WISS2012), December 2012
- [5] 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, pp. 57-64, 2014年2月
- [6] 畑元, 小池英樹, 佐藤洋一, “解像度制御を用いた視線誘導”, 情報処理学会論文誌, Vol.56, No.4, pp.1-10, April, 2015.

(国際)

- [1] Y. Kobayashi and Y. Kuno, “People tracking using integrated sensors for human robot interaction”, In Proc. IEEE ICIT 2010, pp.1597-1602, March 2010.
(DOI: <http://dx.doi.org/10.1109/ICIT.2010.5472444>)
- [2] Y. Sugano, Y. Matsushita and Y. Sato, “Calibration-free gaze sensing using saliency maps”, Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR2010), pp. 2667-2674, June 2010.
(DOI: <http://dx.doi.org/10.1109/CVPR.2010.5539984>)
- [3] Y. Kobayashi, T. Shibata, Y. Hoshi, Y. Kuno, M. Okada, and K. Yamazaki, “I Will Ask You: Choosing Answerers by Observing Gaze Responses Using Integrated Sensors for Museum Guide Robots”, 19th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN2010), 2010.
(DOI: <http://dx.doi.org/10.1109/ROMAN.2010.5598721>)
- [4] T. Siriteerakul, D. Sugimura, Y. Sato, "Head Pose Classification from Low Resolution Images using Pairwise Non-Local Intensity and Color Differences", Proc. Pacific-Rim Symposium on Image and Video Technology (PSIVT2010), November 2010.
(DOI: <http://dx.doi.org/10.1109/PSIVT.2010.67>)
- [5] T. Siriteerakul and Y. Sato, "Investigating image metric features for classifying head pose orientation", International Technical Conference on Circuits/Systems,

- Computer and Communications (ITC-CSCC 2010), July 2010.
- [6] K. Yamada, Y. Sugano, T. Okabe, Y. Sato, A. Sugimoto and K. Hiraki, "Can Saliency Map Models Predict Human Egocentric Visual Attention?", International Workshop on Gaze Sensing and Interactions, November 2010.
(DOI: http://dx.doi.org/10.1007/978-3-642-22822-3_42)
- [7] M. M. Hoque, T. Onuki, E. Tsuburaya, Y. Kobayashi, Y. Kuno, T. Sato, and S. Kodama, "An Empirical Framework to Control Human Attention by Robot," International Workshop on Gaze Sensing and Interactions (ACCV2010 Workshop), 2010.
(DOI: http://dx.doi.org/10.1007/978-3-642-22822-3_43)
- [8] 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: <http://dx.doi.org/10.1109/HSI.2011.5937366>)
- [9] 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: <http://dx.doi.org/10.5244/C.25.126>)
- [10] 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: <http://dx.doi.org/10.1145/2029956.2029968>)
- [11] 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: <http://dx.doi.org/10.1109/ICSMC.2011.6084179>)
- [12] 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: <http://dx.doi.org/10.1109/ICCV.2011.6126237>)
- [13] 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: <http://dx.doi.org/10.1109/ICCVW.2011.6130456>)
- [14] 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), pp. 277-288,

November 2011.

(DOI: http://dx.doi.org/10.1007/978-3-642-25367-6_25)

- [15] 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.
- [16] 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: <http://dx.doi.org/10.1541/ieejeiss.133.28>)
- [17] 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), pp. 333-336, March 2012.
(DOI: <http://dx.doi.org/10.1145/2168556.2168629>)
- [18] A. Yamazaki, K. Yamazaki, T. Ohyama, Y. Kobayashi and Y. Kuno, "A Techno-Sociological Solution for Designing a Museum Guide Robot: Regarding Choosing an Appropriate Visitor," International Conference on Human-Robot Interaction (HRI2012), pp.309-316, March 2012. (DOI: <http://dx.doi.org/10.1145/2157689.2157800>)
- [19] K. Ogaki, K. Kitani, Y. Sugano, and Y. Sato, "Coupling Eye-Motion and Ego-Motion Features for First-Person Activity Recognition", Proc. IEEE Workshop on Egocentric Vision (in conjunction with CVPR2012), pp. 1-7, June 2012.
(DOI: <http://dx.doi.org/10.1109/CVPRW.2012.6239188>)
- [20] M. Hoque, D. Das, T. Onuki, Y. Kobayashi and Y. Kuno, "Robotic System Controlling Target Human's Attention," International Conference on Intelligent Computing (ICIC2012), Lecture Notes in Computer Science, vol.7390, pp.534-544, July 2012.
(DOI: http://dx.doi.org/10.1007/978-3-642-31576-3_68)
- [21] M. Yousuf, Y. Kobayashi, Y. Kuno, A. Yamazaki and K. Yamazaki, "Development of a Mobile Museum Guide Robot That Can Configure Spatial Formation with Visitors," Intelligent Computing Technology, International Conference on Intelligent Computing (ICIC2012), Lecture Notes in Computer Science, vol.7389, pp.423-432, July 2012.
(DOI: http://dx.doi.org/10.1007/978-3-642-31588-6_55)
- [22] M. Hoque, D. Das, T. Onuki, Y. Kobayashi and Y. Kuno, "Model for Controlling a

- Target Human's Attention in Multi-Party Settings," Proc. Ro-Man2012, September 2012.
(DOI: <http://dx.doi.org/10.1109/ROMAN.2012.6343797>)
- [23] D. Das, M. Hoque, T. Onuki, Y. Kobayashi and Y. Kuno, "Vision-Based Attention Control System for Socially Interactive Robots," Proc. Ro-Man2012, September 2012.
(DOI: <http://dx.doi.org/10.1109/ROMAN.2012.6343800>)
- [24] M. Hoque, D. Das, T. Onuki, Y. Kobayashi and Y. Kuno, "An Integrated Approach of Attention Control of Target Human by Nonverbal Behaviors of Robots in Different Viewing Situations," Proc. IROS2012, pp.1399-1406, October 2012.
(DOI: <http://dx.doi.org/10.1109/IROS.2012.6385480>)
- [25] Y. Sugano, K. Harada and Y. Sato, "Touch-consistent perspective for direct interaction under motion parallax", in ACM ITS2012 (Interactive Tabletops and Surfaces Conference), November 2012.
(DOI: <http://dx.doi.org/10.1145/2396636.2396695>)
- [26] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Head Pose-Free Appearance-Based Gaze Sensing Via Eye Image Synthesis", in 21st International Conference on Pattern Recognition (ICPR2012), pp. 1008-1011, November 2012.
- [27] T. Onuki, T. Ishinoda, Y. Kobayashi and Y. Kuno, "Designing Robot Eyes for Gaze Communication", Frontiers of Computer Vision (FCV2013) Workshop, January 2013.
(DOI: <http://dx.doi.org/10.1109/FCV.2013.6485468>)
- [28] M. Yousuf, Y. Kobayashi, Y. Kuno, K. Yamazaki and A. Yamazaki, "A Mobile Guide Robot Capable of Establishing Appropriate Spatial Formations," IEEJ Transactions on Electronics, Information and Systems, vol.133-C, no.1, January 2013.
(DOI: <http://dx.doi.org/10.1541/ieejieiss.133.28>)
- [29] Y. Sugano, Y. Matsushita, and Y. Sato, "Appearance-based Gaze Estimation using Visual Saliency," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 35, no. 2, pp. 329-341, February 2013.
(DOI: <http://dx.doi.org/10.1109/TPAMI.2012.101>)
- [30] 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: <http://dx.doi.org/10.1109/FG.2013.6553740>)
- [31] 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: <http://dx.doi.org/10.1145/2465780.2465784>)

- [32] 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: <http://dx.doi.org/10.1080/01691864.2013.791654>)
- [33] 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: <http://dx.doi.org/10.1109/ICIEV.2013.6572539>)
- [34] 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.
- [35] 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.
- [36] 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: <http://dx.doi.org/10.1109/ROMAN.2013.6628543>)
- [37] 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: <http://dx.doi.org/10.5244/C.27.121>)
- [38] 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: <http://dx.doi.org/10.1016/j.cviu.2013.06.005>)
- [39] 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.
(DOI: http://dx.doi.org/10.1007/978-3-642-53842-1_40)
- [40] 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.
(DOI: http://dx.doi.org/10.1007/978-3-642-53842-1_18)
- [41] 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: <http://dx.doi.org/10.1109/IROS.2013.6696663>)
- [42] 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: <http://dx.doi.org/10.1145/2542355.2542379>)
- [43] 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: <http://dx.doi.org/10.1109/ICCVW.2013.16>)
- [44] T. Onuki, T. Ishinoda, E. Tsuburaya, Y. Miyata, Y. Kobayashi, Y. Kuno, “Designing robot eyes for communicating gaze” Interaction Studies 14:3 (2013), pp.451-479.
(DOI: <http://dx.doi.org/10.1075/is.14.3.07onu>)
- [45] A. Yamazaki, K. Yamazaki, K. Ikeda, M. Burdelski, M. Fukushima, T. Suzuki, M. Kurihara, Y. Kuno, Y. Kobayashi, “Interaction between a quiz robot and multiple participants” Interaction Studies 14:3(2013), pp.366-389.
(DOI: <http://dx.doi.org/10.1075/is.14.3.04yam>)
- [46] 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.
- [47] 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.
(DOI: <http://dx.doi.org/10.1145/2578153.2578199>)
- [48] 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: <http://dx.doi.org/10.1016/j.imavis.2014.01.005>)
- [49] 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.

- [50] 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), June 2014.
(DOI: 10.1109/CVPR.2014.235)
- [51] A. Kanda, M. Arai, R. Suzuki, Y. Kobayashi, Y. Kuno, "Recognizing Groups of Visitors for a Robot Museum Guide tour" 7th international Conference on Human System Interaction (HSI2014), pp. 123 - 128, June 2014.
(DOI: <http://dx.doi.org/10.1109/HSI.2014.6860460>)
- [52] T. Onuki, K. ida, T. Ezure, T. Ishinoda, K. Sano, Y. kobayashi, Y. kuno, "Designing Robot Eyes and Head and Their Motions for Gaze Communication" International Conference on Intelligent Computing (ICIC2014), pp. 607-618, August 2014.
(DOI: http://dx.doi.org/10.1007/978-3-319-09333-8_66)
- [53] M. Arai, Y. Sato, R. Suzuki, Y. Kobayashi, Y. Kuno, S. Miyazawa, M. Fukushima, K. Yamazaki, A. Yamazaki, "Robotic Wheelchair Moving with Multiple Companions" International Symposium on Robot and Human interactive Communication (Ro-Man2014), August 2014.
- [54] M. M. Hoque, Y. Kobayashi, Y. Kuno, "A Proactive Approach of Robotic Framework for Making Eye Contact with Humans" Hindawi Publishing Corporation Advances in Human-Computer Interaction, Volume 2014, Article ID 694046, 19 pages.
(DOI: <http://dx.doi.org/10.1155/2014/694046>)
- [55] 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. Vol. 36, Issue 10, pp. 2033-2046, October 2014.
(DOI: <http://dx.doi.org/10.1109/TPAMI.2014.2313123>)
- [56] J. Nakajima, A. Kimura, A. Sugimoto and K. Kashino, "Visual attention driven by auditory cues: Selecting visual features in synchronization with attracting auditory events," Proc. of the 21st International Conference on MultiMedia Modeling, pp. 74-86, January 2015.
(DOI: http://dx.doi.org/10.1007/978-3-319-14442-9_7)
- [57] M.G. Rashed, R. Suzuki, A. Lam, Y. Kobayashi, Y. Kuno, "Toward Museum Guide Robots Proactively Initiating Interaction with Humans," International Conference on Human-Robot Interaction (HRI2015), March 2015. (DOI: 10.1145/2701973.2701974)

- [58] K. Sano, K. Murata, R. Suzuki, Y. Kuno, D. Itagaki, Y. Kobayashi, "Museum Guide Robot by Considering Static and Dynamic Gaze Expressions to Communicate with Visitors," International Conference on Human-Robot Interaction (HRI2015), March 2015. (DOI: 10.1145/2701973.2702011)
- [59] Yusuke Sugano, Yasuyuki Matsushita, Yoichi Sato, and Hideki Koike, "Appearance-based Gaze Estimation with Online Calibration from Mouse Operations", IEEE Transactions on Human-Machine Systems, 2015. (accepted)

(2) その他の著作物(総説、書籍など)

- [1] 久野義徳、“ヒューマンインタフェースとロボット:ヒューマンロボットインタラクションって何? ”、ヒューマンインタフェース学会誌、Vol.14, No. 4, pp.259-262, 2012.
- [2] 姜, 長橋, 杉本: 画像分類のためのランダムフォレストを用いたシーンコンテキストスケールの開発、画像ラボ, Vol. 23, No. 8, pp. 5-11, 2012.

(3) 国際学会発表及び主要な国内学会発表

① 招待講演 (国内会議 12 件、国際会議 10 件)

(国内)

- [1] 久野義徳, “相互行為分析に基づくロボットの開発,” HCG シンポジウム, 2009. (12 月 11 日, 札幌)
- [2] 佐藤洋一、ユビキタスな注視センシングに向けて、精密工学会 動画像処理実用化ワークショップ DIA2011、徳島、2011 年 3 月 4 日
- [3] 佐藤洋一、“人の視覚的注意をさぐる: 人の行動と視覚的顕著性に基づくアプローチ”、計測自動制御学会北陸支部学術講演会、富山、2012 年 10 月 27 日
- [4] 佐藤洋一、“人の視覚的注意をさぐる: 人の注意をひきつけるとは? ”、2nd Microsoft Kinect for Windows Workshop (情報処理学会シンポジウム インタラクション 2013 併催)、東京、2013 年 2 月 28 日
- [5] 久野義徳、“人間行動の社会的分析に基づくサービスロボットの開発”、第 8 回感性ロボティクスワークショップ、日本感性工学会、東京、2013 年 3 月 18 日
- [6] 佐藤洋一、物体の見えのモデリングから人の視覚へ、第 12 回情報科学技術フォーラム 受賞者フォーラム:メディア認識・理解の最前線と若手へのメッセージ、2013 年 9 月
- [7] 佐藤洋一、視覚的顕著性の計算モデルの高度化と視線推定への応用、生理研視覚研究会「視知覚の現象・機能・メカニズム - 生理学的、心理物理学的、計算論的アプローチ」、2013 年 9 月
- [8] 菅野裕介、データ駆動型アプローチによる視線・視覚的注意の推定、情報処理学会コンピュータビジョンとイメージメディア研究会、2013 年 9 月
- [9] 佐藤洋一、Sensing, predicting, and utilizing visual attention, 応用脳科学コンソーシ

アム ニューロアテンション研究会, 2013 年 12 月

- [10] 杉本晃宏: 視覚的顕著性に基づく注視の検出とその応用、九州大学大学院システム情報科学研究所講演会、2014 年 6 月
- [11] 久野義徳, “人の気持ちの分かる生活支援ロボット”, 日本総研 地域におけるヘルスケア産業創出 さいたま地域, 2015 年 2 月
- [12] 佐藤洋一, “Sensing and learning from human visual attention”, 大川賞記念コンピュータビジョンシンポジウム, 2015 年 3 月

(国際)

- [1] Y. Sato, Toward Calibration Free Gaze Estimation, 2011 China-Japan Symposium on Visual Computing, September 17, 2011.
- [2] Y. Sato, Toward Calibration Free Gaze Estimation, Taiwan Symposium on Computer Graphics, July 13, 2011.
- [3] Y. Kuno, Robotic Wheelchair and Care Robot Based on Sociological Interaction Analysis, Workshop on Social Robots for Assisted Living – Challenges and Responsibilities, Aalborg (Denmark), November 28, 2011.
- [4] Y. Kuno, Computer Vision for Human-Robot Interaction, and Vice-Versa, 18th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV2012), Electric Power Historical Museum (Kawasaki), February 4, 2012.
- [5] Y. Sato, “Sensing, predicting, and utilizing visual attention”, International Symposium on Computational Behavioral Science, September 2013.
- [6] Y. Sato, “Sensing, Predicting, and Utilizing Human Visual Attention”, the 17th SANKEN International Symposium, January 2014.
- [7] Yoichi Sato, “Sensing, Predicting, and Utilizing Human Visual Attention”, 4th International Conference on Image Processing Theory, Tools and Applications, October 2014.
- [8] Yoichi Sato, “Sensing and Utilizing Human Visual Attention”, JSPS UK/Japan Symposium on Computer Graphics and Virtual Reality, November 2014.
- [9] Yoshinori Kuno, “Assistive Robots Based on Sociological Interaction Analysis,” The 9th International Forum on Strategic Technology (IFOST 2014), October 2014.
- [10] Yoshinori Kuno, “Assistive Robots Based on Observation of Humans,” The 21st Korea-Japan Workshop on Frontiers of Computer Vision (FCV2015), January 2015.

② 口頭発表 (国内会議 19 件、国際会議 26 件)

(国内)

- [1] 柴田高志, 鴫田憲, 星洋輔, 小林貴訓, 久野義徳, “複数鑑賞者に適応的な身体的行動を用いて解説をするミュージアムガイドロボット,” 計測自動制御学会システムインテグレーション部門講演会論文集(SI2009), pp1252-1254, 2009. (12月25日, 東京)
- [2] 柴田高志, 星洋輔, 鴫田憲, 小林貴訓, 久野義徳, "頭部動作の計測に基づき質問相手を選択するガイドロボット," 情報処理学会全国大会, pp.5.21-22, 2010. (3月1日, 東京)
- [3] 新川祥史, 杉村大輔, 佐藤洋一, 小池英樹, ”人の視覚特性に基づく大型ディスプレイによる複数人物への選択的情報提示”, 情報処理学会インタラクション 2010、 pp. 29-32, 2010. (3月1日, 東京)
- [4] 小貫朋実, 圓谷恵美, 小林貴訓, 久野義徳, 自然な動作で人間の注視を獲得するロボット 頭部動作の検討, 2011年電子情報通信学会総合大会, 東京, 2011年3月15日(開催中止)
- [5] 姜有宣, 杉本晃宏: シーンコンテキストスケールを用いた画像分類, 情報処理学会コンピュータビジョンとイメージメディア研究会, SIG-CVIM-175, 2011.
- [6] 小貫朋実, 宮田雄規, 小林貴訓, 久野義徳, 親しみやすさと視線コミュニケーション機能を考慮したロボットの目のデザイン, HAI シンポジウム 2011, 京都工芸繊維大学, 2011年12月3日.
- [7] 窪田秀行, 菅野裕介, 岡部孝弘, 佐藤洋一, 杉本晃宏, 開一夫, ”視野特性を考慮した視覚的顕著性モデルの構築”, 日本視覚学会 2012年冬季大会, 2012年1月.
- [8] 小貫 朋実, 宮田 雄規, 小林 貴訓, 久野 義徳, "視線コミュニケーションを考慮したロボット頭部の開発," 画像センシングシンポジウム(SSII2012), 2012.
- [9] M. Yousuf, Y. Kobayashi, Y. Kuno, K. Yamazaki and A. Yamazaki, "Model of Guide Robot Behavior to Explain Multiple Exhibits to Multiple Visitors," Proc. International Session of 30th Annual Conference of the Robotics Society of Japan (RSJ2012), pp.83-86, 2012.
- [10] 窪田秀行, 菅野裕介, 岡部孝弘, 佐藤洋一, 杉本晃宏, 開一夫, ”人間の視野特性を考慮した学習に基づく視覚的顕著性モデル”, 画像の認識・理解シンポジウム(MIRU2012), 福岡, August 2012.
- [11] 鈴木裕太, 小池英樹, モザイク画像を利用した遠近画像の同時提示手法, 日本ソフトウェア科学会インタラクティブシステムとソフトウェア(WISS2012), 青森, December 2012
- [12] 萩原, 杉本, 川本: 誘目性を活用した画像加工による注視誘導, 電子情報通信学会パターン認識・メディア理解研究会, 大阪, February 2013.
- [13] 笠井啓史, 大垣慶介, 菅野裕介, 佐藤洋一, ”視線情報を用いた画像選好の識別”, 映像情報メディア学会 HI 研究会, 東京, March 2013.
- [14] 中島, 杉本, 川本, 移動音源下における視覚的顕著性モデルの構築, 情報処理学会コンピュータビジョンとイメージメディア研究会, 東京, 2013年5月
- [15] Ye, 菅野, 佐藤, "Investigating individual differences in learning-based visual saliency models", 情報処理学会コンピュータビジョンとイメージメディア研究会, 鳥取,

2013年9月

- [16] 畑, 小池, 佐藤, 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オン 2014, 東京, 2014年2月
- [17] 尾崎, 菅野, 佐藤, 視線情報と画像特徴に基づく画像の選好推定, 電子情報通信学会パターン認識・メディア理解研究会, 東京, 2014年3月
- [18] 佐野, 小貫, 井田, 小林, 久野, 人の注視を誘導する目と連動したロボットの振り向き動作の検討, 2014年電子情報通信学会総合大会, 新潟, 2014年3月
- [19] 中島, 木村, 杉本, 柏野, ”音響信号との時空間同期を利用した視覚的顕著性の計算モデル”, 情報処理学会研究会報告 SIG-CVIM, 2015年1月.

(国際)

- [1] T. Siriteerakul, D. Sugimura, Y. Sato, "Head Pose Classification from Low Resolution Images using Pairwise Non-Local Intensity and Color Differences", Proc. Pacific-Rim Symposium on Image and Video Technology (PSIVT2010), November 2010.
- [2] M. M. Hoque, T. Onuki, E. Tsuburaya, Y. Kobayashi, Y. Kuno, T. Sato, and S. Kodama, "An Empirical Framework to Control Human Attention by Robot," International Workshop on Gaze Sensing and Interactions (ACCV2010 Workshop), Queenstown, New Zealand, November 2010.
- [3] 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, Yokohama, Japan, May 2011.
- [4] A. Hagiwara, A. Sugimoto and K. Kawamoto, "Saliency-Based Image Editing for Guiding Visual Attention", Int. Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI2011), Beijing, September 2011.
- [5] 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), Kagawa, Japan, December 2011.
- [6] 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.
- [7] 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.
- [8] 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), Kawasaki, Japan, February 2012.
- [9] K. Ogaki, K. Kitani, Y. Sugano, and Y. Sato, "Coupling Eye-Motion and Ego-Motion Features for First-Person Activity Recognition", Proc. IEEE Workshop on Egocentric Vision (in conjunction with CVPR2012), Rhode Island, USA, June 2012.
- [10] M. Hoque, D. Das, T. Onuki, Y. Kobayashi and Y. Kuno, "Robotic System Controlling Target Human's Attention," International Conference on Intelligent Computing (ICIC2012), Huangshan, China, July 2012.
- [11] M. Yousuf, Y. Kobayashi, Y. Kuno, A. Yamazaki and K. Yamazaki, "Development of a Mobile Museum Guide Robot That Can Configure Spatial Formation with Visitors," Intelligent Computing Technology, International Conference on Intelligent Computing (ICIC2012), Huangshan, China, July 2012.
- [12] D. Das, M. Hoque, T. Onuki, Y. Kobayashi and Y. Kuno, "Vision-Based Attention Control System for Socially Interactive Robots," Proc. Ro-Man2012, Paris, France, September 2012.
- [13] M. Hoque, D. Das, T. Onuki, Y. Kobayashi and Y. Kuno, "Model for Controlling a Target Human's Attention in Multi-Party Settings," Ro-Man2012, Paris, France, September 2012.
- [14] M. Hoque, D. Das, T. Onuki, Y. Kobayashi and Y. Kuno, "An Integrated Approach of Attention Control of Target Human by Nonverbal Behaviors of Robots in Different Viewing Situations," IROS2012, Vila Moura, Algarve, Portugal, October 2012.
- [15] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Head Pose-Free Appearance-Based Gaze Sensing Via Eye Image Synthesis", in 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, November 2012.
- [16] T. Onuki, T. Ishinoda, Y. Kobayashi and Y. Kuno, "Designing Robot Eyes for Gaze Communication", Frontiers of Computer Vision (FCV2013) Workshop, Incheon, Korea, January 2013.
- [17] M. M. Hoque, K. Deb, D. Das, Y. Kobayashi, and Y. Kuno, "An Intelligent Human-Robot Interaction Framework to Control the Human Attention," International Conference on Informatics, Electronics & Vision (ICIEV), Bangladesh, May 2013.
- [18] D. Das, Y. Kobayashi, and Y. Kuno, "Observing Human's Face for Robot's Controlling His/Her Attention," International Conference on Quality Control by Artificial Vision (QCAV2013), Fukuoka, May 2013.

- [19] 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,” International Symposium on Robot and Human Interactive Communication (Ro-Man2013), Korea, August 2013.
- [20] H. Koike, “Vision-based Human Computer Interaction”, NTU-UEC Tokyo Joint Workshop, Taiwan, September 2013.
- [21] J. Nakajima, A. Sugimoto and K. Kawamoto, “Incorporating Audio Signals into Constructing a Visual Saliency Map,” 6th Pacific-Rim Symposium on Image and Video Technology (PSIVT2013), Guanajuato, October 2013.
- [22] Y. Sato, “Modeling, sensing and controlling visual attention”, International Joint Workshop on Advanced Sensing/Visual Attention and Interaction (ASVAI2013), Okinawa, November 2013.
- [23] D. Das, Y. Kobayashi, and Y. Kuno, “Attracting Attention and Establishing a Communication Channel Based on the Level of Visual Focus of Attention,” International Conference on Intelligent Robots and Systems (IROS2013), Tokyo, November 2013.
- [24] 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,” 3rd IEEE Workshop on Consumer Depth Cameras for Computer Vision (CDC4CV2013), Sydney, December 2013.
- [25] T. Onuki, K. Ida, T. Ezure, T. Ishinoda, K. Sano, Y. Kobayashi, and Y. Kuno, “Designing Robot Eyes and Head and Their Motions for Gaze Communication” International Conference on Intelligent Computing (ICIC2014), Taiyuan, China, August 2014.
- [26] J. Nakajima, A. Kimura, A. Sugimoto and K. Kashino, “Visual attention driven by auditory cues: Selecting visual features in synchronization with attracting auditory events,” The 21st International Conference on MultiMedia Modeling, January, 2015.

③ ポスター発表（国内会議 8 件、国際会議 30 件）

（国内）

- [1] 圓谷、小貫、小林、久野、親しみやすさと視線の読みとりやすさを兼ね備えたロボットの目のデザイン、2011 年電子情報通信学会総合大会学生ポスターセッション、東京、2011 年 3 月 16 日（開催中止）
- [2] 萩原、杉本、川本：注視誘導のための視覚的顕著性に基づく画像加工，情報処理学会 コンピュータビジョンとイメージメディア研究会，SIG-CVIM-177，2011.
- [3] 山田、菅野、岡部、佐藤、杉本、開，“顕著性と自己運動に基づく一人称視点における視覚

的注意推定”, 画像の認識・理解シンポジウム (MIRU 2011), July 2011.

- [4] 萩原, 杉本, 川本: 視覚的顕著性を利用した画像加工による注視誘導, 2012 電子情報通信学会総合大会 ISS 特別企画学生ポスターセッション, 2012.
- [5] 大垣, 木谷, 菅野, 佐藤, “自己運動と視線運動を用いた一人称視点映像からの自己動作認識”, 画像の認識・理解シンポジウム (MIRU 2012), August 2012.
- [6] 小貫, 江連, 石野田, 小林, 久野, 注視を演出するコミュニケーションロボット, 画像センシングシンポジウム (SSII2013), 横浜, 2013 年 6 月
- [7] 中島, 木村, 杉本, 柏野, “音響信号との同期を手がかりとする視覚的顕著性の計算モデル”, 画像の認識・理解シンポジウム (MIRU2014), 2014 年 7 月.
- [8] 中島, 木村, 杉本, 柏野, “音響信号との時空間同期を利用した視覚的顕著性の計算モデル”, 情報処理学会コンピュータビジョンとイメージメディア研究会, 2015 年 1 月.

(国際)

- [1] Y. Kobayashi, T. Shibata, Y. Hoshi, Y. Kuno, M. Okada and K. Yamazaki, "Choosing Answerers by Observing Gaze Responses for Museum Guide Robots," HRI2010 LBR, March 2010.
- [2] Y. Kobayashi, T. Shibata, Y. Hoshi, Y. Kuno, "Selective function of speaker gaze before and during questions - Towards developing museum guide robots," CHI2010 Extended Abstracts, April 2010.
- [3] Y. Sugano, Y. Matsushita and Y. Sato, "Calibration-free gaze sensing using saliency maps", Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR2010), June 2010.
- [4] T. Siriteerakul and Y. Sato, "Investigating image metric features for classifying head pose orientation", International Technical Conference on Circuits/Systems, Computer and Communications (ITC-CSCC 2010), July 2010.
- [5] K. Yamada, Y. Sugano, T. Okabe, Y. Sato, A. Sugimoto and K. Hiraki, "Can Saliency Map Models Predict Human Egocentric Visual Attention?", International Workshop on Gaze Sensing and Interactions, November 2010.
- [6] 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.
- [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.
- [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.
- [9] M. M. Hoque, T. Onuki, E. Tsuburaya, Y. Kobayashi, and Y. Kuno, Controlling

- Human Attention by Robot's Behavior Depending on his/her Viewing Situations, International Conference on Social Robotics (ICSR2011), Amsterdam, November 24, 2011.
- [10] 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.
- [11] M. M. Hoque, T. Onuki, Y. Kobayashi, Y. Kuno, Attracting and Controlling Human Attention through Robot's Behaviors Suited to the Situation, 7th ACM/IEEE International Conference on Human-Robot Interaction, Boston, March 6, 2012.
- [12] M. A. Yousuf, Y. Kobayashi, Y. Kuno, A. Yamazaki, and K. Yamazaki, Establishment of Spatial Formation by Mobile Guide Robot, 7th ACM/IEEE International Conference on Human-Robot Interaction, Boston, March 6, 2012.
- [13] Y. Sugano, K. Harada and Y. Sato, "Touch-consistent perspective for direct interaction under motion parallax", in ACM ITS2012 (Interactive Tabletops and Surfaces Conference), Boston, USA, November 2012.
- [14] T. Onuki, T. Ishinoda, Y. Kobayashi and Y. Kuno, "Design of Robot Eyes Suitable for Gaze Communication," HRI2013, Tokyo, Japan, March 4, 2013.
- [15] D. Das, M. Hoque, T. Onuki, Y. Kobayashi and Y. Kuno, "Attention Control System with Suitable Cue and Time Depending on the Level of Attention," HRI2013, Tokyo, Japan, March 4, 2013.
- [16] L. Su and Y. Sato, "Early Facial Expression Recognition with Early RankBoost", IEEE Int. Conf. Automatic Face and Gesture Recognition (FG 2013), Shanghai, April 2013.
- [17] L. Feng, Y. Matsushita, I. Sato, T. Okabe and Y. Sato, "Uncalibrated photometric stereo for unknown isotropic reflectances", IEEE Conference on Computer Vision and Pattern Recognition (CVPR2013), Portland, July 2013.
- [18] Y. Sugano, H. Kasai, K. Ogaki, and Y. Sato, "Image Preference Estimation from Eye Movements with a Data-driven Approach", Int. Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI2013), Lund, August 2013.
- [19] I. Chamveha, Y. Sugano, Y. Sato and A. Sugimoto, "Social Group Discovery from Surveillance Videos: A Data-Driven Approach with Attention-Based Cues", British Machine Vision Conference (BMVC 2013), Bristol, September 2013.
- [20] H. Hata and H. Koike, "Visual Attention Guidance Using Image Resolution Control", NTU-UEC Tokyo Joint Workshop on Human-Computer Interaction, Taiwan, September 2013.
- [21] T. Shi and A. Sugimoto, "Video Saliency Modulation in the HSI Color Space for

Drawing Gaze,” Pacific-Rim Symposium on Image and Video Technology (PSIVT2013), Guanajuato, October 2013.

- [22] J. Nakajima, A. Sugimoto and K. Kawamoto, “Constructing a Visual Saliency Map Combined with Audio Signals,” International Joint Workshop on Advanced Sensing/Visual Attention and Interaction (ASVAI2013), Okinawa, November 2013.
- [23] Y. Sugano, Y. Matsushita, and Y. Sato, “Automatic detection of regions of interest via graph-based joint clustering”, International Joint Workshop on Advanced Sensing/Visual Attention and Interaction (ASVAI2013), Okinawa, November 2013.
- [24] I. Chamveha, Y. Sugano, Y. Sato, and A. Sugimoto, “Data-driven detection of social groups via attention-based cues”, International Joint Workshop on Advanced Sensing/Visual Attention and Interaction (ASVAI2013), Okinawa, November 2013.
- [25] Y. Kobayashi, T. Onuki, D. Das, Y. Kuno, “Designing robot's behavior and eyes to effectively attract and shift user's attention”, International Joint Workshop on Advanced Sensing/Visual Attention and Interaction (ASVAI2013), Okinawa, November 2013.
- [26] T. Onuki, T. Ezure, T. Ishinoda, Y. Kobayashi, and Y. Kuno, “Static and Dynamic Robot Gaze Expressions to Communicating with Humans,” Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV2014), Okinawa, February 2014.
- [27] B. Ye, Y. Sugano, and Y. Sato, "Influence of stimulus and viewing task types on a learning-based visual saliency model", Symposium on Eye Tracking Research and Applications (ETRA2014), Safety Harbor, March 2014.
- [28] D. Das, M. G. Rashed, Y. Kobayashi, and Y. Kuno, “Recognizing Gaze Pattern for Human Robot Interaction,” International Conference on Human-Robot Interaction (HRI2014), Bielefeld, March 2014.
- [29] Y. Sugano, Y. Matsushita, Y. Sato, “Learning-by-Synthesis for Appearance-based 3D Gaze Estimation”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR2014), Columbus, June 2014
- [30] K. Oka, Y. Sugano, Y. Sato, A. Sugimoto, T. Kondo, K. Hachisuka, K. Abe, E. Okuno, “Studying effects of shifting gaze with subtle visual stimuli during driving,” The 10th Asia-Pacific Conference on Vision, Kagawa, July 2014

(4) 知財出願

① 国内出願(0件)

なし

② 海外出願(0件)

なし

③ その他の知的財産権

なし

(5) 受賞・報道等

① 受賞

- [1] 萩原愛子: 情報処理学会 CVIM 研究会卒論セッション優秀賞 (2011.05)
- [2] 萩原愛子: 2012年電子情報通信学会 総合大会 ISS 特別企画「学生ポスターセッション」優秀ポスター賞(2012.03)
- [3] M.M. Hoque, T. Onuki, Y. Kobayashi, and Y. Kuno: Best Presentation Award, IEEE HSI2011 (4th International Conference on Human System Interaction) (2011.05)
- [4] * M.A. Yousuf, Y. Kobayashi, Y. Kuno, A. Yamazaki, and K. Yamazaki: Best Paper Award, The 2012 Eighth International Conference on Intelligent Computing (2012.07)
- [5] 原田一馬, 菅野裕介, 佐藤洋一, 情報処理学会シンポジウム インタラクション 2012 インタラクティブ論文賞 (2012.03)
- [6] * T. Onuki, K. Ida, T. Ezure, T. Ishinoda, K. Sano, Y. Kobayashi, and Y. Kuno, Best Paper Award, International Conference on Intelligent Computing (ICIC2014) (2014.08)

② マスコミ(新聞・TV等)報道

- [1] TBS テレビ「夢の扉～My Goal～」(平成23年3月20日 18:30-19:00)にて電気通信大学小池研究室の特集番組が放映され、本プロジェクトの成果システムが紹介された。
- [2] テレビ放送: テレビ埼玉 埼玉ビジネスウォッチ 2013年7月14日 コミュニケーションロボット

③ その他

なし

(6) 成果展開事例

① 実用化に向けての展開

- ・ 視覚刺激による注視誘導技術に関して、株式会社デンソーとの共同研究を実施

- ・ 視聴覚刺激による注視反応モデルの構築とその応用について、NTT との共同研究を実施
- ② 社会還元的な展開活動
- ・ 研究成果の一部として得られたデータセットは、研究目的の利用に限り一般公開している。
(<http://www.hci.iis.u-tokyo.ac.jp/datasets/>)

§ 5 研究期間中の活動

5.1 主なワークショップ、シンポジウム、アウトリーチ等の活動

年月日	名称	場所	参加人数	概要
22年11月8日	International Workshop on Gaze Sensing and Interactions	Queenstown, New Zealand	38人	視線の検出とインタラクションに関する国際ワークショップ
25年3月8日	合同シンポジウム	東北大学	30人	第75回情報処理学会全国大会企画セッション (CREST 八木チーム、武田チーム、黄瀬チームとの合同開催)
25年11月5日	International Joint Workshop on Advanced Sensing / Visual Attention and Interaction - Toward Creation of Human-Harmonized Information Technology-	LOISIR HOTEL NAHA, Okinawa, Japan	50人	ACPR2013 併設国際ワークショップ(神田チーム、八木チーム、河原チーム、武田チーム、黄瀬チーム合同開催)

§ 6 最後に

本研究では、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に取り組んだ。コンピュータビジョン、ロボティクス、ヒューマンインタラクションの各分野を専門とする研究メンバーが集い、各グループが自律的に研究を推し進めつつ、グループ間で有機的に連携することで、全体として研究目標をほぼ達成し、多くの重要な成果を挙げることが出来た。例えば、人の視覚特性の一つである視覚的顕著性を手掛かりにして事前キャリブレーションを必要としない視線推定を行うという全く新しい発想に基づく手法を考案し、実際に高い精度での視線推定が可能であることを示した。また、適応的線形回帰にもとづく視線推定アルゴリズムを新たに考案し、視線推定における事前準備の大幅な簡略化と高精度化を同時に実現している。これらの研究成果は情報分野において最も高いインパクトファクタを持つ論文誌の一つである **IEEE Trans. PAMI** に採択されており、国内外で高く評価されている。また、不特定人物を対象とし、事前準備不要な視線推定という困難な課題を達成するために、多人数・多視線方向における大規模目画像データベースを他に先駆けて構築し、大量データからの統計的学習に基づく視線推定という、従来とは全く異なる発想にもとづく手法の有効性を実証した。また、視覚刺激に対する注視反応のモデル化に関しても、従来の視覚的顕著性モデルでは人の視野特性と自己運動への考慮が欠如しているという点に着目し、既存モデルよりも高い予測精度を持つ視覚的顕著性モデルの構築に成功している。また、人の行動を阻害することなく、さりげなく人の注視を誘導するために、視覚的刺激による自然な注視誘導とロボットの身体動作による注視誘導という2つのアプローチを提唱し、それらが人の注視誘導に有効であることを検証した。この視覚刺激による注視誘導はデジタルサインエージや運転者への注意喚起による事故防止などの応用への展開が期待される。また、ロボットの目と頭部動作による注視の引付と誘導に関する技術はヒューマンロボットインタラクションの基盤をなす技術として重要な意味を持つと考える。

一方、ロボットの身体動作による注視誘導に関して、当初想定したよりも実験・評価のためのロボットの開発に時間を要したため、具体的な応用シナリオを想定したプロトタイプシステムによる効果検証が本報告書作成の時点では未完了となっている。これについては本研究が終了する27年度末までには終えることを予定している。また、視覚刺激による注視誘導を用いた視線適応型インタフェースの開発に関しても、プロトタイプシステムによる予備的な実験による性能評価を終えたところであり、今後より規模の大きな実験を通じたシステムの洗練化が求められる。

また、本研究を進める中で、本 CREST 領域内の多くのチームとワークショップや合宿形式の研究会の開催、実験データやプログラムの共有など、様々な形での連携を進められたことは当初想定していなかった予想外の収穫であり、本研究を推進するにあたり大きな支えとなった。定期的で開催されている領域会議や領域シンポジウムはこのようなチーム間の連携を促す上で、非常に役立っており、CREST の枠組みでの研究推進の重要な特長の一つと考える。



合宿形式の合同研究会（平成 26 年 9 月 25 日～26 日、国際高等セミナーハウス、軽井沢市）の様子