

AggreGaze: 観客の注意の集合推定 公衆ディスプレイ上で

杉野祐介

ドイツ・マックス・プランク情報科学研究所 sugano@mpi-inf.mpg.de

張旭聡

ドイツ・マックス・プランク情報科学研究所 xczhang@mpi-inf.mpg.de

アンドレアス・ブルリング

ドイツ・マックス・プランク情報科学研究所 bulling@mpi-inf.mpg.de

要約

視線は観客の注意を監視・分析するための重要性から、公共ディスプレイ研究で頻りに検討される。しかし、現在の視線推定機能を備えた公衆ディスプレイのインターフェースは、特別な目的のアイ・トラッキング機器を必要とするか、個々のユーザーごとに明示的な個人キャリブレーションを必要とする。私たちは、公衆ディスプレイ上の時空間的な観客注意を推定する新しい手法AggreGazeを提案する。我々の手法は、ディスプレイに取り付けられた単一の市販カメラのみを必要とし、個人キャリブレーションを必要とせず、ディスプレイ全体にわたる視覚的注意推定を提供する。これは、1) 最先端の外観ベース視線推定法の誤差を現場の訓練データ収集によって補正し、2) 校正されていない、したがって不正確な複数ユーザーの視線推定を結合して共同注意推定に統合することで実現する。補正のためのさまざまな視覚刺激を提案する：標準的な9点キャリブレーション、動くターゲット、テキストおよびディスプレイ内容に埋め込まれた視覚刺激、さらには通常のビデオコンテンツ。公共空間での2週間の展開に基づき、注意マップを推定する際にground-truthの視線分布に極めて近い効果を示すことを実証する。

ACM分類キーワード

H.5.m. 情報インターフェースと提示 (例: HCI) : 雑多

著者キーワード

視覚的注意; 視線追跡; 公共ディスプレイ; 視線推定

はじめに

公衆ディスプレイの文脈における人間の視線には二重の目的がある。第一に、視線はマウスより指し示すのが速く、直感的に使えるため、インタラクションのモダリティとして魅力的である [15]。第二に、視線はユーザーが何に興味を持っているかを自然に示すため、観客の注意を監視するのに適している [2]。非インタラクティブ (パッシブ) ディスプレイの視覚的注意を測定することは特に relevantである。

この研究の部分または全体のデジタルまたは印刷物の個人または教室使用は、営利目的または商業的利益のために複製・配布されないこと、及びこの通知と第一ページの完全な引用を掲載することを条件に、料金なしで許可されます。他の著者が所有する本研究の構成要素の著作権を尊重してください。引用は許可されています。その他のコピーや改訂を行う場合は、事前の特別な許可および/または料金が必要です。Permissions@acm.orgへ許可を問い合わせてください。

UIST 2016、2016年10月16-19日、東京、日本
著作権は著作者が所有します。出版権はACMにライセンスされています。
ACM 978-1-4503-4189-9/16/10...\$15.00
DOI: <http://dx.doi.org/10.1145/2984511.2984536>

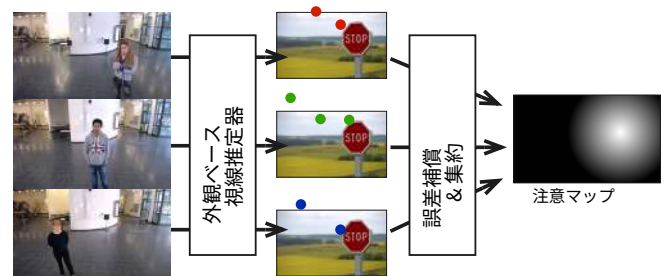


図1: AggreGazeは最先端の外観ベース視線推定法に基づいて構築され、現場の訓練データから得られたユーザー依存なしの誤差補正関数を適用します。補正された視線位置は2Dの観客注意マップを計算するために集約されます。

広告の効果分析 [35] やディスプレイブラインドネス効果 [11, 28] の研究など、さまざまな応用を可能にし、より一般的には広範囲にわたる注意性の高いユーザーインターフェースの開発の鍵となる。

インタラクティブなディスプレイは、通常、特別な用途の据え置き型または頭部装着型の視線追跡機器を各ユーザーごとに事前キャリブレーションすることを必要とします [24, 40, 44]。最近のインターフェースは、正確な視線点の推定や煩わしいキャリブレーションを必要としない、滑らかな追跡 [16, 46] や左右への短い凝視 [57, 59] などの操作技術に依存します。しかし、これらのインターフェースはすべて同時に1人のユーザーのみをサポートし、通常はユーザーがディスプレイの正面に固定された位置に留まることを要求します。

対照的に、公共ディスプレイ上で複数のユーザーの注意を受動的に監視することは、ユーザーが任意の距離や角度からディスプレイを見たり、移動中にも見ることができると、かなり困難です。最近の外観ベース手法の進展は、個人キャリブレーションなしで野外での視線推定を約束します [56, 50]が、ある設定で訓練された視線推定器を、ノートパソコンのような設定から公衆ディスプレイのような別の設定へ移行する方法は未解決のままです。

AggreGazeを提示します。公衆ディスプレイ上の観客の注意を推定する新しい方法です。方法はキャリブレーションフリーで、ディスプレイ全体の2D注意マップを提供し、ディスプレイに取り付けられた単一の市販のRGBカメラのみを必要とします。AggreGazeは公衆ディスプレイ用の最先端の外観ベース視線推定法 [50] の限られた推定精度を次の2つの方法で改善します：最初に視線推定器の上にマッピング関数を訓練して補正を行い、

訓練時と導入時のカメラ角度および照明条件の違いによって生じる誤差を補償します。この補償のためのデータ収集に向けて、異なる視覚刺激を検討します。たとえば、標準的な9点および移動目標の設計[19, 32, 48]から、通常の表示コンテンツに埋め込んだ文字および視覚刺激、さらに通常の動画コンテンツまでを扱います。加えて、本手法は、これらの推定が不正確であり、それゆえそれ自体では信頼できない場合でも、異なるユーザからの注視推定値を集約して全体の注意分布を算出します（図1参照）。このようにして、本手法は観客の注意の時空間ヒートマップを生成できます。これらのヒートマップはたとえば、コンテンツ提供者が、観客が意図された画面上の位置に注意を向けているかどうかを分析するために利用できる可能性があります。このような分析は、表示情報の見えやすさを向上させるための自動調整へとさらに拡張できます。

本研究の具体的な貢献は3点です。第一に、公的ディスプレイにおける集団的な注意推定のための新しい手法であるAggreGazeを提示します。本手法は1) 市販の既製カメラをディスプレイに1台取り付けるだけでよく、2) キャリブレーション不要であり、3) ディスプレイ全体にわたる注意推定を提供し、4) 複数ユーザをサポートします。第二に、カメラ角度と照明の違いに対処するための誤差補償の手法を導入し、この補償のための学習データを収集するために、表示コンテンツへ視覚刺激を埋め込むさまざまな方法を提案します。第三に、2週間にわたりシステムを公共空間に設置して、AggreGazeとさまざまな視覚刺激の実環境評価を示します。結果は、集約された注意マップが、人の注視のグラウンドトゥース分布に非常に近いことを示しています。

関連研究

本研究は、公的ディスプレイにおける注意の計測に関する先行手法、ならびに複数ユーザからの注視観測の分析および可視化に関する先行手法に基づいています。

公的ディスプレイにおける注意の計測

公的ディスプレイ上でユーザの注意を計測する方法については、多くの研究がなされています。頭部装着型のアイトラッカーは、注視インタラクションや制御された環境における注意の計測に成功裏に用いられてきました（例については[24, 39, 47]を参照）。しかし、ユーザに追加デバイスを装着させることは、制約のない状況で導入され、多数の未知のユーザに利用される公的ディスプレイには現実的ではありません。この問題に対処するため、別の研究では公的ディスプレイに搭載したアイトラッカーに依存しています[32, 46]。ただし、現在のアイトラッカーの追跡範囲が限られているため、ユーザにはディスプレイ正面の固定位置に立つことが求められます。単純な赤外線センサは、個人ごとのキャリブレーションなしで使用できます[36, 45]が、眼球接触（アイコンタクト）しか検出できません。

これに対して、市販のRGBカメラは容易にセットアップでき、はるかに広い視野をカバーできます。Zhangらは、公的ディスプレイのインタラクションの入力として、相対的な眼球運動を検出するシステムを提案しました[57, 58]。しかし、彼らの手法はディスプレイ中心から視線をそらした瞬間しか検出せず、画面全体の注意推定は提供しません。同時に、外観ベースの注視推定手法は、近年コンピュータビジョン分野において大きな進展が見られています。

現在のアプローチでは、大量の実データまたは合成データ[49]の学習データを活用することで、制御されていない「その場 (in-the-wild)」の状況[50, 56]でも、パーソナルなキャリブレーションなしでの注視推定[43]が可能になります。しかし、公的ディスプレイに必要なとされる、個人ごとのキャリブレーションなしでの遠距離注視推定は、最先端の手法であっても依然として難しい課題です。残された重要な課題は、たとえばラップトップ[56]のような別の環境で訓練された注視推定器を、公的ディスプレイへ移し替えることです。この課題に対処するために、最先端の外観ベースの注視推定手法に加えて、デバイス固有の追加的な誤差補償のための学習を提案します。

注視が取得できない状況では、頭部姿勢や身体の向きを代替として用いることができますが、提供できるのは大まかな注意推定のみです[3, 38, 52]。別の方法として、視覚的注意の計算モデルを用い、ボトムアップ方式で注意分布を予測することができます[6]。最近の研究では、これらのモデルの適用範囲をインタラクティブな状況[7, 51]へ拡張しています。しかし、これらのモデルは視覚情報のみを用い、ユーザのタスク・目標・意図といった注意へのトップダウンの影響を無視しているため、予測能力は依然として限定的です。

複数の注視観測の可視化

複数のユーザから記録した注視データを可視化することは、アイトラッキングおよび情報可視化の研究における中核的なテーマです。Blascheckら[5]は、さまざまな注視データの可視化手法を調査し、アイトラッキングデータの特性に基づいて9つのグループに分類しました。3D注視を可視化するためのいくつかの手法は提案されています[41, 31]が、大多数の研究は2D注視データの可視化に焦点を当てています[13, 22]。主要な手法の1つは、画像などの特定の刺激に対して記録された注視データを、注意ヒートマップとして要約することです。これらのヒートマップは通常、刺激のさまざまな部分における注視の頻度と持続時間をエンコードします。そのわかりやすい解釈可能性から、マーケティングやWebのユーザビリティ研究など、学術以外でも広く利用されています。

動画などの動的刺激に対する注意を可視化する方法は、公的ディスプレイの文脈でも関連しており、集中的に研究されてきました。たとえば、Duchowskiらは、ヒートマップを用いて複数の視聴者からの眼球運動を集約するリアルタイムの可視化手法を開発しました[12]。Kurzahlsらは、複数の視聴者のデータを用いて一般的な視聴行動の傾向を特定し、それを新しい時空間可視化[23]によって可視化しました。さらに後の研究では、複数の視聴者の注視をクラスタリングすることで時空間分析を可能にする別の手法を導入しました[21]。これらの研究はいずれも、個々の注視計測を集約して視聴者の注意の一般的な特性を表すものですが、このアプローチはこれまでのところ、注視の可視化の分野の外では用いられていません。

観客の注意の推定

公的ディスプレイにおける注意推定のための本手法は、多モーダル畳み込みニューラルネットワーク (CNN) に基づく最先端の外観ベース注視推定手法に依拠しています（図2参照）。外観ベースの手法は、例えば（以下のような）明示的な眼の特徴検出を行わずに、眼の外観から注視方向への対応関係を直接学習します。

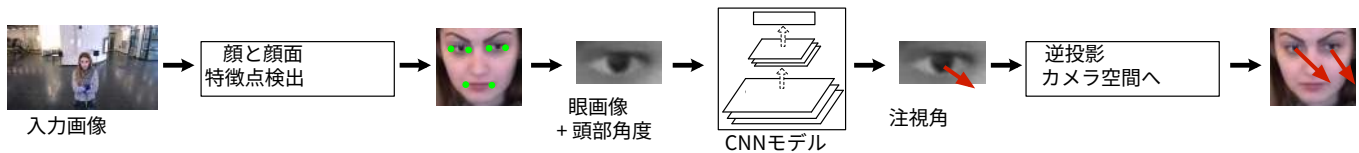


図2: AggreGazeで使用する外観ベースの注視推定パイプライン。まず、校正済みの単眼RGBカメラから取得した入力画像内の特徴点を特定するために、最先端の顔および顔面の特徴点検出手法を採用する。次に、汎用の3D顔モデルを当てはめて3D頭部姿勢を推定し、正規化技術を用いて頭部姿勢と眼画像を正規化空間へ切り出し・ワープする。最後に、最先端のマルチモーダル畳み込みニューラルネットワーク (CNN) を用いて、頭部姿勢 h および眼画像 e から注視方向 v への対応付けを学習する。

眼の外眼角または瞳孔中心の。である。私たちは、より深いネットワークアーキテクチャを用いたマルチモーダルCNNモデル[56]を、大規模な合成データセット[49]で学習する。これは特にパブリックディスプレイ環境を対象としている。しかし、パブリックディスプレイ環境では、カメラ角度や周囲照明など、学習と運用の間の相違に加え、異なるユーザー間の外観のばらつきにより、依然として大きな注視推定誤差が残る。この問題に対処するために、外観ベースの注視推定パイプライン (図3参照) の上に、誤差補償と集約の手順を提案する。

外観ベースの注視推定

図2は、外観ベースの注視推定パイプラインの概要を示す。パイプラインへの入力動画は、パブリックディスプレイに搭載された単眼RGBカメラから記録される。カメラの内在パラメータと、カメラ座標系におけるターゲットディスプレイの3D姿勢は、ミラーベースの手法[33]を用いて校正する。まず、入力画像内のすべての顔をHOG-SVMの顔検出器[18]で検出する。次に、顔の特徴点検出と追跡の手法を用いて、顔画像[4]における2Dの顔特徴点位置 (左右の眼角と口角) を出力する。検出された顔特徴点位置は、汎用の3D顔形状モデルにそれらを当てはめることで、ユーザーの3D頭部姿勢を推定するために用いられる。私たちは、[56]が提供する汎用の3D顔形状モデルを使用する。次に、[43]で説明されている画像正規化を適用し、眼画像 e を切り出し、頭部姿勢を頭部角度ベクトル h を用いた正規化空間へワープする。この正規化により、ロール角における頭部回転が除去され、画像が所定の距離にスケールされる。これにより、学習データ空間が注視角と頭部角の限られた範囲に実質的に制限される。

正規化処理は左右両方の眼に適用される。切り出した眼画像 e と頭部角度 h を、CNN注視推定モデルの入力として用い、このモデルが正規化空間における注視角 g を予測する。私たちは、元のモデル[56]のCNNネットワークを、より深いAlexNetアーキテクチャ[20]で置き換えた。ネットワークには5つの畳み込み層、2つの全結合層、および1つの出力層が含まれる。元のモデルでは入力画像サイズが 227×227 ピクセルであることを前提としているのに対し、私たちのパブリックディスプレイ環境では入力眼画像が低解像度になりがちなため、代わりに 60×36 ピクセルを用いる。入力サイズのこの変更を反映して、最初の畳み込み層のストライドパラメータも4から1へ変更する。[56]で提案されているように、最初の全結合層の出力の末尾に頭部角度ベクトル h を連結する。

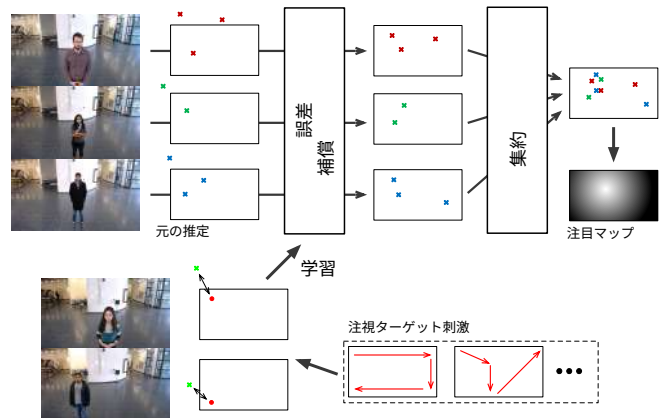


図3: 誤差補償および集約処理の概要。誤差補償関数は、注視ターゲット刺激を用いて収集した学習データで学習され、個々の推定結果に適用される。これらの補償済み注視位置は時間同期され、集約されて、最終的な注目マップ表現が作成される。

CNNモデルの学習では、LSVRC-2010 ImageNet学習セット[34]で事前学習されたモデルを用いて畳み込み層を初期化する。入力画像サイズが異なるため、2つの全結合層は最初から学習する。[50]の手法に従い、まず[49]の提供するツールで合成した100,000枚の画像でモデルを学習し、その後UTデータセット[43]で微調整する。ターゲットディスプレイ上部にカメラを取り付ける私たちの構成を対象とするため、カメラ座標系において垂直注視方向が0度未満となるように学習サンプルも制限する。このようにして、学習データにはカメラより下を見ている眼画像のみが含まれる。

最後の層の出力は、正規化空間における2次元の注視角ベクトル v である。注視角 v は、その後、眼の3D位置から伸びる注視レイとして、カメラ座標系における3Dの注視ベクトルへ逆投影される。注視レイがディスプレイ平面と交差することで、ディスプレイ上の最終的な注視点 g が得られる。

注視推定誤差の要因の取り扱い

実験セクションで述べるように、外観ベースの注視推定の性能は、より深いネットワークやターゲット化を行っても、大幅に低下する傾向がある。

学習データです。困難さは主に2つの要因に由来します：異なるユーザ間での外観のばらつきと、学習と運用の間に存在する相違です。最初の要因である外観のばらつきはユーザ固有であり、学習ベースの人非依存注視推定手法にとって基本的な課題を表しています[56]。個別のキャリブレーションなしでは、多くのユーザにわたって頑健で正確な推定を得ることは困難です。2つ目の要因も、現在の学習ベース手法の重要な制約です。訓練済み推定器を異なる環境で使用すると、例えばカメラを別の角度に配置した場合、別の画面上の注視範囲に対して使用した場合、または異なる照明条件下で使用した場合に、大きなグローバル誤差が観察されました。注視推定手法には、推定した3D注視レイをスクリーンと交差させる幾何学的計算が含まれるため、幾何学キャリブレーションの不正確さによって追加のグローバル誤差が生じる可能性があります。図3に示すように、環境誤差と個人的誤差にそれぞれ対処するため、2つのアプローチ、すなわち誤差補償と注視集約を提案します。

誤差補償

一般に、学習ベースの注視推定手法は学習データに含まれるケースのみを適切に扱うことができ、そのため推定結果は学習環境に偏る傾向があります。このような誤差要因は、本質的には、入力となる頭部位置の範囲および、照明条件や対象ディスプレイ平面の幾何形状といった環境特性に応じて、推定結果に対するバイアスとして作用します。

この問題に対処するため、まず外観ベースの注視推定モデルの上に、追加の環境誤差補償関数を適用します。訓練サンプルの集合 $\{(p, gl, gr), \boxtimes g\}$ があるとします。ここで (p, gl, gr) は外観ベースの注視推定パイプラインから得られる注視推定結果（それぞれユーザの3D頭部位置と、左右の眼から推定された注視位置）であり、 $\boxtimes g$ はグラウンドトゥルースの画面上の注視位置です。次に、誤差補償関数は、この学習データに対する回帰を用いて学習されます。

データ $\boxtimes g = f(p, gl, gr)$ 本研究では、サポートベクトルを用います。放射基底関数カーネルを用いた回帰（SVR）。

この補償関数は個々のユーザに依存せず、各運用（デプロイ）に対して1回だけ学習すると仮定します。最良の場合には、訓練データ収集のために参加者を募集することで実現できますが、期待されるグラウンドトゥルースの注視位置を持つ視覚刺激を挿入することで、実際の観衆から訓練データを収集できるかどうかを検討することは実務上重要です。以下のセクションでは、完全に制御された事前学習から、自然なコンテンツを用いた現地での訓練データ収集まで、訓練サンプルを得るための設計空間についても議論します。

注視集約

一方で、個人的誤差補償には最終的にユーザごとの訓練データが必要です。しかし、制約のない公共ディスプレイ環境では、そのような訓練データ収集が可能だと仮定することは現実的ではありません。公共ディスプレイ環境の重要な特性の1つは、同じユーザが数日、数週間、あるいは数か月の間に何度もそのディスプレイを訪れることに対応しなければならない点です。したがって、注意モニタリングシステムへの入力、推定された注視の集合となり、

多くの観衆から蓄積された位置です。この意味で、実務上最も重要な課題は、個々の注視推定誤差を減らすのではなく、これらの低精度な観測から注意の分布を推定することです。

したがって、個々の注視位置から注意マップを作成し、ディスプレイ上での観衆の注意分布を解析・可視化することを提案します。複数ユーザにわたる、時間同期された外観ベースの注視推定結果を集約して、時空間的な注意分布を復元します。個人的誤差は正規分布に従うと仮定でき、注視推定結果は真の注意分布から引かれたノイズを含むサンプルとみなせます（実験セクションでの詳細な分析を参照）。

公共ディスプレイに表示された映像刺激の各タイムフレーム t について、複数の観衆から（補償された）画面上の注視位置の集合 $\{\boxtimes g_t\}$ を蓄積します。次に、それらの平均と分散を計算することで、この観測集合を正規分布として近似します。これにより、観衆の注意の確率分布を近似したものが得られます。実験では、この集約された注意分布が、グラウンドトゥルースの人間の注視分布をよく表現できることを示します。

実装の詳細

外観ベースの注視推定パイプラインはC++で実装されています。顔面ランドマーク追跡のステップでは、dlibライブラリ[17]の顔検出モジュールを用いたCLM-framework1を使用します。3D顔形状モデルのフィッティングは、OpenCV[8]におけるPnPアルゴリズムの実装で行います。CNNベースの注視推定では、Caffeライブラリ[14]とともにAlexNetモデル2を学習し、使用します。記録された各ビデオ系列から、このパイプラインは検出されたすべての顔に対応する左右の眼の注視位置を出力します。

誤差補償および集約パイプラインはPythonで実装されています。誤差補償ステップにはscikit-learn[29]のSVR実装を使用し、ハイパーパラメータはランダム化サーチによって最適化します。学習データが与えられると、これによって外観ベースの注視推定パイプラインから得られる元の推定結果が、誤差補償された画面上の注視位置へと変換されます。これらの入力ビデオフレームと注視推定データは、公共ディスプレイに表示される対応する刺激の最も近いフレームと時間同期されます。続いて集約ステップでは、表示された刺激の各フレームについて、誤差補償されたすべての注視位置の平均と分散を計算し、保存します。

データ収集

公共ディスプレイのセットアップを用いて収集したin-the-wildデータを使って、私たちのAggreGazeアプローチを検討します。このデータ収集の目的は2つあります：1) グラウンドトゥルースの眼球トラッキングデータと比較して本手法を評価するためのテストデータを収集すること、そして2) 環境誤差補償のための学習データ収集に向けた、さまざまな視覚刺激デザインを調べることです。

1<https://github.com/TadasBaltrusaitis/CLM-framework>

2https://github.com/BVLC/caffe/tree/master/models/bvlc_alexnet

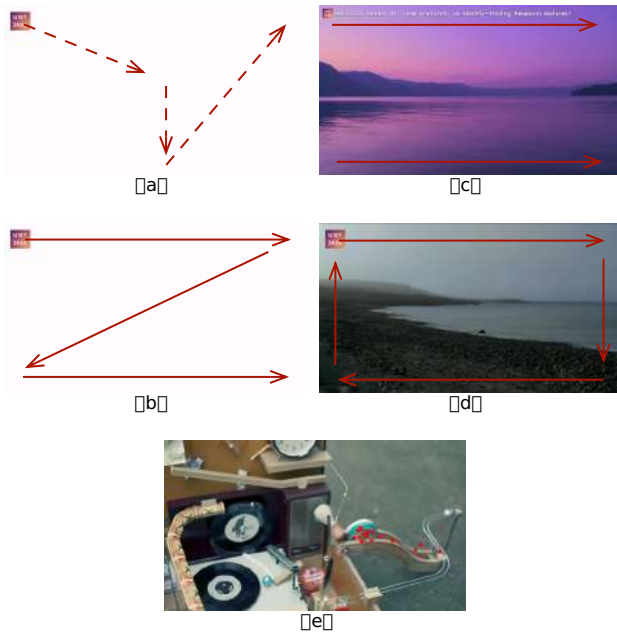


図4：本研究で誤差補償マッピングを訓練するために探索した異なる視覚刺激：離散的な9点 (a) と追跡 (b)、通常の表示内容に埋め込まれたテキスト (c) とロゴ (d)、および通常のビデオコンテンツ (e)。



図5：AggreGazeを公共空間で2週間展開しました。録画設定は、iiyama ProLite 46インチディスプレイ、Logitech C930eウェブカメラ、およびカスタム録画ソフトウェアを実行するホストコンピュータで構成されました。

自然なビデオ刺激

データ収集のため、2つの異なるビデオソースを使用しました。まず、UIST 2015で発表された論文の8本の30秒動画クリップを選びました。これらの動画は複雑な質感・視覚要素を含み、公共ディスプレイで表示される典型的なコンテンツを表しています。加えて、グラウンドトゥールの人間視線アノテーションを含む2つの既存のビデオデータセットを使用しました。最初のデータセットはCoutrotらによって作成され、4つの異なるカテゴリからなる72本の自然映像を含みます：動く物体1つ、動く物体が複数、風景、会話をする人々[10]。2つ目のデータセットはHollywood2データセットで、69の人気動画から抜粋された短いクリップで構成されています。

ハリウッド映画[26]。MatheらはHollywood2データセット[27]に視線追跡データを公開しました。これにより、商用視線追跡装置を用いて記録された15~20名の視聴者からのグラウンドトゥールの視線位置が、これらのデータセットの双方で利用可能になります。

誤差補償のための視覚刺激

前述のとおり、環境誤差補償のためのトレーニングデータを収集することが主要な課題です。本研究では、トレーニングデータ収集のためのさまざまな視線ターゲット設計を提案・評価します。ユーザーへの明示的な指示を伴う完全に制御されたデータ収集シナリオに加え、通常のディスプレイ内容に様々な程度で埋め込まれた動画刺激を用いた制約の少ないデータ収集シナリオも検討します。

静止ターゲット

トレーニングデータ収集の最も直接的な設計は、静止した視線ターゲットを表示することです。視線追跡システムのほとんどは、明確な注意の焦点を確保するためにアニメーションされたドット刺激を使用しますが、ここでは静止ターゲットを表示するだけとし、この刺激にすべての聴衆が焦点を合わせるには限らないと仮定します。図4 (a) に示すように、視線ターゲットマーカーは9つの事前定義された離散的なグリッド位置にランダムに現れ、各点で1.6秒間保持されます。赤いドットをターゲットマーカーとして使用する標準設計に加え、通常の高さの画像 (UIST 2016ロゴ) を使用する設計もテストします。

追跡ターゲット

視線追跡システムで使用される別の視覚刺激は、移動する視線ターゲット [19, 32, 48] です。静止設計とは異なり、表示領域全体を速く効率的にカバーできると期待されます。本研究では図4 (b) と同様のZ字型デザインを採用しました。ターゲットマーカーは左上の位置から開始し、各コーナーへ10 mm/秒の速度で移動します。コーナーに到達すると1秒間停留します。静止ターゲットの場合と同様に、単純なドットターゲットと小さな画像ターゲットの両方をテストします。

埋め込みターゲット

上記の2つの設計はトレーニングデータ収集のための専用動画を想定していますが、動画コンテンツに視覚刺激を埋め込むことは実務的にも重要です。埋め込み設計として、テキストティッカーとエッジロゴの2つを検討します。テキストティッカー設計ではACM technews ウェブサイトから短い見出しテキストを抽出し、図4 (c) と同様に動画の上部と下部の両方にランダムな見出しを表示しました。ロゴ画像は左端のコーナーに表示され、テキストは徐々に表示されます (1秒あたり25文字)。エッジロゴ設計では、同じUIST 2016のロゴ画像を使用し、図4 (d) と同様に画像の縁に沿って位置を動かしました (1秒あたり13 mm)。

動画

コンテンツ埋め込みトレーニングデータ収集のより高度なアプローチは、事前に記録された人間の視線パターン [1] またはボトムアップの顕著性モデル [42] からグラウンドトゥールの視線分布を推定できる動画を使用することです。これらの以前のアプローチは、テスト入力動画上の視線と顕著性パターンを用いた個人キャリブレーションに焦点を当てますが、主な焦点はこれらの点にあります

[3http://technews.acm.org/](http://technews.acm.org/)

この手法は人に依存しない誤差補償である。上記で述べたデータセットの動画から、図4(e)のとおり各フレームの注視位置を抽出する。マッピング学習には平均注視位置を用い、マッピング性能は他の動画で評価する。同様に、これらの動画に対して画像ベースの顕著性予測を適用し、顕著性予測結果の最大位置を注視ラベルとして用いる。顕著性予測モデルとして、ボトムアップ予測モデル（ブルマップ顕著性）[25]とトップダウンの顔検出[54]を組み合わせた。

データ収集プロトコル

データ収集のために、公共の場に iiyama ProLite 46インチのディスプレイを設置した。画角90度のフルHDウェブカメラ（Logitech c930e）をディスプレイ上部に取り付けた。映像刺激と撮影されたカメラ画像はいずれもホストコンピュータによってタイムスタンプされ、記録中に並行して保存された。この記録セットアップと上記の映像刺激を用いて、公共の場で2種類のデータ収集を行った。データ収集のために募集した年齢が20~29歳の19名の参加者（女性6名）に加えて、記録ディスプレイの前に現れたすべての顔を記録した。

制御条件

まず、12名の参加者による完全に制御されたデータ収集から開始した。参加者には、静止した注視目標を用いたトレーニングデータ収集を行うよう明確に依頼した。より具体的には、ディスプレイの前方に9つの位置を定義した。参加者には、それぞれの位置に順番に立って、ディスプレイ上で示された目標位置を見るよう求めた。静止設計と同様に、60（10×6グリッド）の位置に赤いドットを表示した。これにより、視線位置と立ち位置の両方について十分なばらつきを持つ、環境補償のための最も理想的な一回限りのトレーニングデータが得られる。

自然条件

続いて、データ収集用に設計したパターンを含む72本のビデオクリップを用いて、公共ディスプレイの視聴者の自然な反応を記録した。このシーケンスは以下のビデオクリップで構成される。

- ・明示的なキャリブレーションパターン 4個（2 静止、2 追従）
- ・埋め込み型キャリブレーションパターン 12個
- ・UIST ビデオ 8本
- ・Coutrot データセットの動画 24本（カテゴリごとに6本）
- ・Hollywood2 データセットの動画 24本

合計すると、全体のループにはおよそ25分かかった。

上記の12名の参加者も、この自然記録を同様に行った。これら72本のビデオを、位置に関する制限なしで視聴するよう大まかに指示した。別の7名の参加者は、この自然記録にのみ参加した。上記の12名の場合と異なり、我々は遠隔で、ディスプレイが設置されている場所に行き、一定時間ビデオを視聴するよう指示した。これらの記録の間、現地の調査担当者はいないため、制御された環境よりも自然に振る舞うことが期待される。

ビデオクリップのシーケンスをおよそ12時間×13日間再生した。これにより、ビデオあたり平均で25顔が検出され（SD=10.4）、したがって合計では $25 \times 72 = 1080$ 顔となった。図6は、我々の記録からのいくつかのサンプル画像を示している。

図から分かるように、記録中に注視位置は大きく変化した。通行人は、刺激を単独で見る場合もあれば、異なる人数規模のグループで見る場合もあり、また静止している場合も移動している場合もあった。図6では、外観ベースの注視推定手法への入力として使用したサンプルの眼領域画像もさらに示している。見てのとおり、眼領域画像は通常、低解像度で、ピントが外れ、ぼやけており、視聴者もメガネやメイクをしていた。これらの画像特性はいずれも、モデルベースの注視推定手法にとって非常に難しい状況を提示していると同時に、外観ベース手法を採用したことの利点を示している。

実験

Coutrot [10] および Hollywood2 [26, 27] のデータセットにある正解の注視分布と、推定した注意マップを比較することで AggreGaze を評価した。

ベースライン手法の性能分析

まず、外観ベースの注視推定手法のベースライン性能と、誤差補償の効果を評価した。図7は、制御された記録において、12名の参加者に対する公共ディスプレイ上での平均注視推定誤差を示している。各ドットは9つの立ち位置のいずれか（上から見たときで、下にカメラとディスプレイ位置がある）に対応し、その大きさや色は、正解の目標位置からのユークリッド距離として定義した注視推定誤差を表している。図7aは外観ベースの注視推定による元の推定結果に対応し、図7bは誤差補償後の結果に対応する。図7bでは、誤差補償関数を leave-one-person-out の方式で学習した。すなわち、他の参加者から得た学習データを用いた。

図7aに示すとおり、元の注視推定結果は誤差が大きく（ ≈ 30 cm）、立ち位置が近/中央の位置からさらに離れるほど誤差はますます大きくなる。我々の誤差補償アプローチにより、これらの推定誤差を大幅に低減できる（図7b）。

図8は、個人誤差の分布、すなわち補償後に残る誤差を示している。最初の5名の参加者から、誤差補償後の推定結果をランダムに100件選び、中央の赤いドットとして示された正解位置に対するそれらの位置をプロットした。先に述べたとおり、個人誤差は正解位置の周りで正規分布的に分布する傾向があり、明確な個人差（バイアス）は観察されなかった。

注意予測の性能

続いて、明示的な一回限りの学習を用いて、AggreGaze アプローチの性能を定量化した。具体的には、制御された記録セッションを学習データとして用い、データセットの動画に対する予測された注意分布を、正解の人間の注視（フィクサーション）と比較した。

本システムは個々の注視位置を集約し、注意分布を出力するため、視覚的顕著性マップを評価する際によく用いられる評価指標を採用した。受信者動作特性（receiver operating characteristics）の曲線下面積（AUC）測定は、最も一般的な評価基準の1つであり、



図6：2週間の展開中に記録されたサンプル画像。通行人は、さまざまな距離から、公的ディスプレイに表示された視覚刺激を、単独で、またサイズの異なるグループでも、さらに静止している場合と移動中の場合の両方で見ていました。赤色で示された検出された顔のバウンディングボックスも掲載しています（プライバシー上の理由から、ここでは顔のみピクセル化されています）。下段は、顔および顔のランドマーク検出を使用して抽出したサンプルの拡大眼領域画像を示します。非制約の設定では、低解像度で、焦点が外れ、かつぼやけた画像になります。追加の問題として、多くの視聴者が眼鏡やメイクを着用していることがあります。

サリエンシーマップ。シャッフルしたAUC [55] を用いました。ここでは、真陽性サンプルはグラウンドトゥルースの注視位置から取得し、偽陽性サンプルは他のすべてのフレームにおけるグローバルな注視分布に従って取得します。正規化スキャンパス・サリエンシー (NSS) は、もう一つの一般的な指標であり、グラウンドトゥルースの注視位置における正規化したゼロ平均注意マップの平均値として定義されます [30]。ベースラインとして、環境誤差補償なしでの元の外観ベースの視線推定結果、および純粋に画像ベースのサリエンシーモデルを用いた注意予測結果と比較しました。さらに、注意予測とボトムアップ・サリエンシー予測を共同で用いた場合も示します。この場合、我々の予測結果を視覚的サリエンシーの事前分布マップとして使用し、ボトムアップ・サリエンシーマップに乘算しました。

図9は、すべての手法のシャッフルAUCおよびNSSスコアを示します。法。Original は、計算された注意マップに対応し誤差補償なしの場合です。補償された位置 \bar{x}_g の代わりに、元の左目および右目出力 (g_l, g_r) の平均位置を取り、同じ手順でそれらを集約しました。Saliency は、画像ベースのサリエンシー予測結果に対応しており、ターゲット刺激の設計で議論したのと同じモデルを使用しました。Mean は、別のコンテンツベースのベースラインに対応し、注意マップはグラウンドトゥルースの視線位置の全体分布から作成しました。グラウンドトゥルースデータのすべてのフレーム/動画にわたって視線位置を統合し、平均注意マップを作成しました。Proposed は、我々の AggreGaze 手法に対応しています。提案手法はこれらすべてのベースライン手法を上回り、性能向上

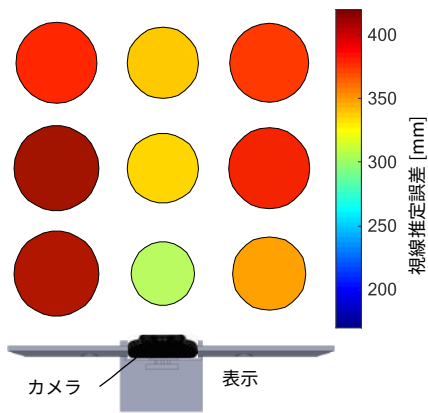
最良のベースライン (Mean) との差は、両方の指標で統計的に有意です (対応のあるt検定、 $p < 0.01$)。

図10では、我々の手法によって推定された注意マップの例を示します。各行は、それぞれサリエンシーマップ、平均注意マップ、集約された注意マップを示しています。重ねられた白い点は、グラウンドトゥルースの視線位置を表します。一般にサリエンシーマップは人の注視位置を表し得るものの、多くの偽陽性領域も含む傾向があります。平均マップは常に画像の中心に留まり、基本的に注意分布のダイナミクスを表現できません。対照的に、AggreGaze が予測した注意マップは、グラウンドトゥルースの視線位置とよく相関しています。

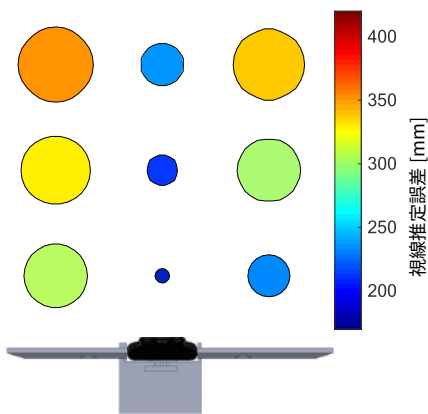
視線ターゲット設計

図11では、環境誤差補償に対する異なるターゲット設計を用いた注意予測性能について、さらに比較しました。自然な記録セッションのみを使用し、視線ターゲット刺激のそれぞれを用いて、異なる誤差補償関数を学習させました。上から下に向かって、各プロットはテンプレートのサリエンシー/視線、埋め込みテキスト・ティッカーおよびエッジロゴのパターン、追跡 (pursuit) と離散的な視線ターゲットに対応しています。最後のプロットは、上で議論した制御されたキャリブレーション結果に対応します。テンプレートのサリエンシー/視線を用いた結果では、マッピング関数を学習するために使用したビデオクリップをテストセットから除外しました。

ご覧のとおり、記録ループに埋め込まれた明示的な視線ターゲット (Stationary, Pursuit) は、完全に制御された記録と同等の質の訓練データを提供します。エッジロゴの設計 (Stationary) も同様に良好に機能しており、性能は



(a) 元の見積もり



(b) 誤差補償あり

図7：制御条件下での公衆ディスプレイ正面の異なる位置での注視推定誤差。バブルの大きさや色は右側に示されるミリメートル単位の注視推定誤差に対応します。(a) 元の外観ベースの注視推定法と (b) 私たちの誤差補償後の結果を示します。

テキストティッカー設計 (Text) により劣化。比較的明確な視線ターゲットを持つこれらのケースと比べて、テンプレート顕在性/視線パターンの性能は相対的に低いです。

ディスカッション

実験結果は、提案されたAggreGazeアプローチが、複数のユーザーが参加し、個人のキャリブレーションや特別な機器を必要としない、難易度の高い公衆ディスプレイ環境で、真の視線分布に非常に近い時空間的注意マップを推定できることを示しています。私たちのアプローチの核心は、現地のトレーニングデータから環境誤差を補償し、複数の観測を統合して注意分布を推定するというもので、制約の少ない自然環境で学習ベースの視線推定手法をデプロイし実用化する新しい方法を提供します。画像ベースの顕在性予測ベースラインは自然動画刺激に対してのみ良好な性能を示しますが、私たちの手法は予測可能です。

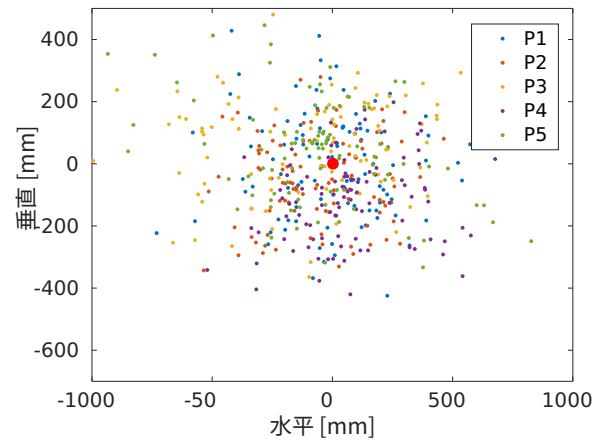


図8：制御条件下での最初の5人の参加者の個人注視推定誤差。誤差補正後に100の注視推定をランダムに選択し、真の注視点 (赤い点) に対する位置をプロットしました。

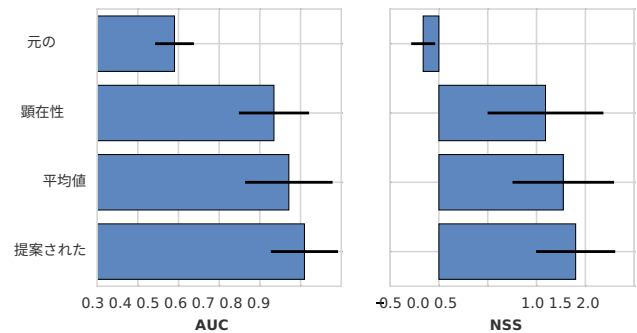


図9：AggreGazeと他のベースライン手法との性能比較。棒グラフは、誤差補償なしの元の推定結果、顕在性マップ、平均注視マップ、およびAggreGazeの集合注視マップのシャッフルされたAUCとNSSスコアを示します。

あらゆるディスプレイ内容の注視マップ。同様に、平均注視ベースラインは顕著性指標の点で驚くほど良好に機能しますが、そのような静的な注視予測の実用的な意味は本質的に制限されています。

さまざまな視線ターゲット設計を分析することにより、表示内容に視線ターゲットを埋め込むことで現地のトレーニングデータを使用して誤差補償機能を効率的に学習できることがわかりました。これはAggreGazeアプローチが現実世界の環境に直接展開可能である可能性を強調します。一方、テキストティッカーやテンプレート視線のようなより自然なパターンの性能はまだ制限されています。これは、人間の視線行動が明示的なターゲット位置よりも曖昧になる場合が多く、テンプレート視線パターンの強い中心バイアスが全体的な誤差補償に追加の課題をもたらすと考えられるからです。したがって、将来の研究の重要な方向性は、他の埋め込み手法を調査し、現地トレーニングデータ収集の手段を拡大することになるでしょう。

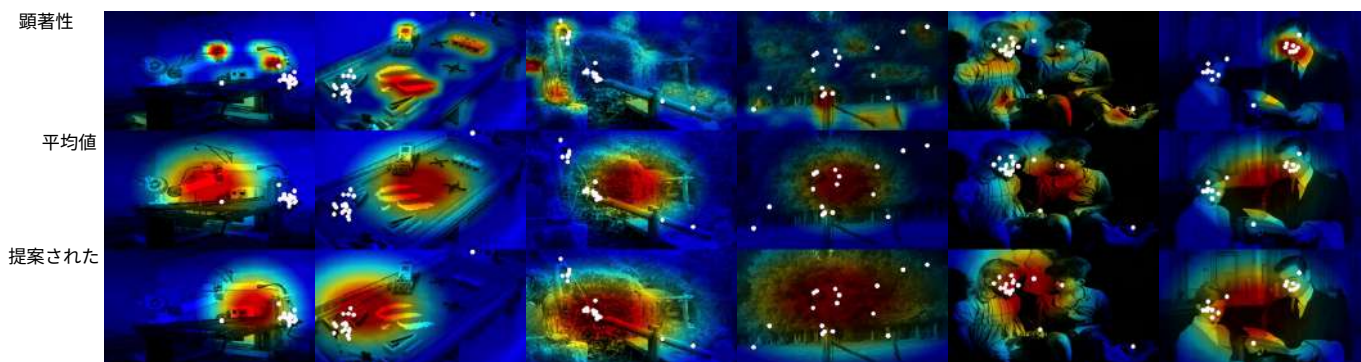


図10: 注意予測結果の例。各行は、顕著性マップ、平均注目マップ、および私たちの方法による集約注目マップを、それぞれ、地表真の注視位置を示す白い点が重ねられた状態で示します。画像はCoutrotデータセットから取得しています。

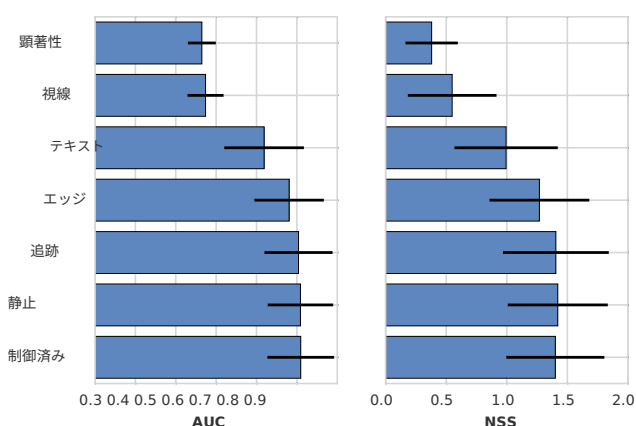


図11: 異なる訓練目標との性能比較。棒グラフは、動画の顕著性と視線、埋め込みテキストとエッジターゲット、追跡、および9点視線ターゲットと完全に制御された訓練データを用いて収集した訓練データを使用したAggreGazeのシャッフルAUCとNSSスコアを示します。

私たちの方法は、基礎となる視線推定手法の改善の恩恵をさらに受けます。視線推定の性能は、深層学習手法の活用により最近大幅に向上していますが、現在の手法は屋外の照明変化や低品質入力画像に依然として問題を抱えています。したがって、今後の研究では、実世界の条件下で頑健に機能する正確な視線推定手法を開発することが重要です。現在のシステムは単一の正規分布を仮定していますが、そのような推定改善は、マルチモーダル分布として観客の注意をより正確に近似することを可能にします。

結論

私たちはAggreGazeを提案しました。公衆ディスプレイ上の観客の注意を推定する新しい手法です。私たちの手法は、現場での訓練データ収集を通じて、最新の外観ベースの視線推定にデバイス特有の誤差補正を適用します。

およそ個々の観察をまとめて、結合注視マップを推定します。私たちの方法は、ディスプレイに取り付けられた市販のカメラ1台だけを必要とし、個別のキャリブレーションを要求せず、全体ディスプレイに対する視覚注目の推定を提供します。公共スペースでの2週間にわたる展開の結果、推定された注目マップは人間の注視の真実分布に密接に一致することが示されました。したがって、私たちの方法は、公衆ディスプレイ上での観客の注意を侵襲性を抑えつつ正確に監視する重要な一歩を表しており、広範な注意型ユーザーインターフェースの研究における新しい方向性を開くものです。

謝辞

この研究は、サール大学の多Modal計算と相互作用 (MMCI) 研究クラスター、アレクサンダー・フォン・フンボルト財団、およびJST CREST研究助成金の一部によって支援されました。

参考文献

1. Alnajar, F., Gevers, T., Valenti, R., および Ghebreab, S. 人間の視線パターンを用いた較正不要な視線推定. Proc. ICCV (2013), 137-144.
2. Alt, F., Bulling, A., Mecke, L., および Buschek, D. Attention, please! 広範なディスプレイに対する視聴者の注意を測る特徴の比較. Proc. DIS (2016).
3. Ba, S. O., および Odobez, J.-M. ヘッドポーズと会議文脈手掛かりからの複数人の視線焦点推定. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 1 (2011), 101-116.
4. Baltrušaitis, T., Robinson, P., および Morency, L.-P. 構造化帰帰のための連続条件付きニューラル場. Proc. ECCV. 2014, 593-608.
5. Blaschek, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., および Ertl, T. 視線追跡データの可視化の最先端. Proc. EuroVis, vol. 1. 2014 (2014).
6. Borji, A., および Itti, L. 視覚的注意のモデリングの最先端. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 1 (2013), 185-207.

7. ボルジー, A., シヒテ, D. N., and Itti, L. インタラクティブ環境における上位視覚注意の計算モデル化。In Proc. BMVC (2011), 1–12.
8. Bradski, G. The OpenCV ライブラリ。Dr. Dobb's Journal of Software Tools (2000).
9. Bulling, A. 広範囲な注意型ユーザーインターフェイス。IEEE Computer 49, 1 (2016), 94–98.
10. Coutrot, A., and Guyader, N. 銘結果、顔、音が動的な社会的場面での視線に与える影響は？ Journal of Vision 14, 8 (2014), 5–5.
11. Dalton, N. S., Collins, E., and Marshall, P. Display blindness?: 眼球追跡を用いて設置ディスプレイの可視性を再検討。In Proc. CHI (2015), 3889–3898.
12. Duchowski, A. T., Price, M. M., Meyer, M., and Orero, P. 実時間ヒートマップによる総視線可視化。In Proc. ETRA (2012), 13–20.
13. Dumais, S. T., Buscher, G., and Cutrell, E. ウェブ検索における視線パターンの個人差。In Proc. IliX (2010), 185–194.
14. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. Caffe: 畳み込みアーキテクチャによる高速特徴埋め込み。arXiv preprint arXiv:1408.5093 (2014).
15. Khamis, M., Bulling, A., and Alt, F. 視線を用いた対話型パブリックディスプレイの課題に取り組む。In Adj. Proc. UbiComp (2015), 763–766.
16. Khamis, M., Saltuk, O., Hang, A., Stolz, K., Bulling, A., and Alt, F. Textpursuits: パブリックディスプレイでの追跡型対話とキャリブレーションのためのテキスト活用。In Proc. UbiComp (2016).
17. King, D. E. Dlib-ml: 機械学習ツールキット。Journal of Machine Learning Research 10 (2009), 1755–1758.
18. King, D. E. Max-margin object detection. arXiv preprint arXiv:1502.00046 (2015).
19. Kondou, Y., and Ebisawa, Y. ヘッドフリー遠視線検出システムにおける動く視覚ターゲットを用いた簡易視線キャリブレーション。In Proc. VECIMS (2008), 145–150.
20. Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet分類深層畳み込みニューラルネットワーク。In Proc. NIPS, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. 2012, 1097–1105.
21. Kurzhals, K., Heimerl, F., and Weiskopf, D. Isecube: ビデオの視線データの視覚分析。In Proc. ETRA (2014), 43–50.
22. Kurzhals, K., Hlawatsch, M., Burch, M., and Weiskopf, D. 固視イメージチャート。In Proc. ETRA (2016), 11–18.
23. Kurzhals, K., and Weiskopf, D. ダイナミック刺激の視線データの時空間ビジュアル分析。IEEE Transactions on Visualization and Computer Graphics 19, 12 (2013), 2129–2138.
24. ランダー, C., ゲーリング, S., クルーガー, A., ボーリング, S., and ブリング, A. Gaze projector: 複数ディスプレイ間での正確な視線推定とシームレスな視線相互作用。In Proc. UIST (2015).
25. Li, J., and Zhang, Y. 高速かつ正確な物体検出のためのサーフカスケード学習。In Proc. CVPR (2013), 3468–3475.
26. Marszałek, M., Laptev, I., and Schmid, C. 文脈におけるアクション。In Proc. CVPR (2009), 2929–2936.
27. Mathe, S., and Sminchisescu, C. 目の中のアクション: 動的視線データセットと視覚認識のための学習銘頭モデル。IEEE Transactions on Pattern Analysis and Machine Intelligence 37, 7 (2015), 1408–1424.
28. Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., and Krüger, A. Display blindness: デジタルサイネージへの注意に及ぼす期待の影響。In Pervasive Computing. 2009, 1–8.
29. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Pythonの機械学習。Journal of Machine Learning Research 12 (2011), 2825–2830.
30. Peters, R. J., Iyer, A., Itti, L., and Koch, C. 自然画像における下向き視線配分の要素。Vision Research 45, 18 (2005), 2397–2416.
31. Pfeiffer, T. 空間内の注意を3Dアテンションボリュームで測定・可視化。In Proc. ETRA (2012), 29–36.
32. Pfeuffer, K., Vidal, M., Turner, J., Bulling, A., and Gellersen, H. Pursuit calibration: 視線キャリブレーションをより煩雑でなく、より柔軟に。In Proc. UIST (2013), 261–270.
33. Rodrigues, R., Barreto, J. P., and Nunes, U. 主平面鏡反射の画像を用いたカメラ姿勢推定。In Proc. ECCV. 2010, 382–395.
34. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., ほか. Imagenet 大規模視覚認識チャレンジ。International Journal of Computer Vision 115, 3 (2015), 211–252.
35. Schrammel, J., Mattheiss, E., Döhnt, S., Paletta, L., Almer, A., and Tscheligi, M. 移動中のペルベイシブ広告媒体への注意行動。In Pervasive Advertising. 2011, 287–307.
36. Shell, J. S., Vertegaal, R., and Skaburskis, A. W. Eyepliances: 視覚注意にตอบสนองする注目端末。In Ext. Abstr. CHI (2003), 770–771.
37. Sibert, L. E., and Jacob, R. J. 視線インタラクションの評価。In Proc. CHI (2000), 281–288.

38. Sippl, A., Holzmann, C., Zachhuber, D., and Ferscha, A. 公共ディスプレイのリアルタイム注視トラッキング。Ambient Intelligence にて。2010年、167-176頁。
39. Smith, J. D., Vertegaal, R., and Sohn, C. Viewpointer: 軽量なキャリブレーション不要の視線追跡による普遍的なハンズフリー指示。Proc. UIST (2005)、53-61。
40. Stellmach, S., and Dachsel, R. Look & touch: 視線対応のターゲット獲得。Proc. CHI (2012)、2981-2990。
41. Stellmach, S., Nacke, L., and Dachsel, R. 3次元仮想環境の高度な視線ビジュアル化。Proc. ETRA (2010)、109-112。
42. Sugano, Y., Matsushita, Y., and Sato, Y. 視覚的顕著性を用いた外観ベースの視線推定。IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 2 (2013), 329-341。
43. Sugano, Y., Matsushita, Y., and Sato, Y. 学習-合成による外観ベースの3D視線推定。Proc. CVPR (2014)、1821-1828。
44. Turner, J., Bulling, A., Alexander, J., and Gellersen, H. デバイス間の視線対応のポイント間コンテンツ転送。Proc. ETRA (2014)、19-26。
45. Vertegaal, R., Shell, J. S., Chen, D., and Mamuji, A. 拡張注意の設計: 集中型ユーザーインタフェースの枠組みへ。Computers in Human Behavior 22, 4 (2006), 771-789。
46. Vidal, M., Bulling, A., and Gellersen, H. Pursuits: 滑らかな追跡視線運動と移動ターゲットに基づく表示との自発的相互作用。Proc. UbiComp (2013)、439-448。
47. Walter, R., Bulling, A., Lindlbauer, D., Schuessler, M., and Müller, J. 公共ディスプレイとの全身インタラクション中の視覚注意を分析。Proc. UbiComp (2015)、1263-1267。
48. Williams, O., Blake, A., and Cipolla, R. s3gpを用いたスペースおよび半教師付き視覚マッピング。Proc. CVPR, vol. 1 (2006), 230-237。
49. Wood, E., Baltrušaitis, T., Morency, L.-P., Robinson, P., and Bulling, A. 100万枚の合成画像から外観ベースの視線推定器を学習する。Proc. ETRA (2016)、131-138。
50. Wood, E., Baltrušaitis, T., Zhang, X., Sugano, Y., Robinson, P., and Bulling, A. 目の形状登録と視線推定のための瞳のレンダリング。Proc. ICCV (2015)。
51. Xu, P., Sugano, Y., and Bulling, A. グラフィカルユーザーインタフェースにおける視覚注意の時空間モデル化と予測。Proc. CHI (2016)。
52. Yoo, B., Han, J.-J., Choi, C., Yi, K., Suh, S., Park, D., and Kim, C. 大規模ディスプレイのための視線と手のジェスチャを組み合わせた3Dユーザーインタフェース。Ext. Abstr. CHI (2010)、3709-3714。
53. Zhai, S., Morimoto, C., and Ihde, S. 手動と視線入力のカスケード(マジック)ポインティング。Proc. CHI (1999)、246-253。
54. Zhang, J., and Sclaroff, S. 顕著性検出: ブールマップ手法。Proc. ICCV (2013)、153-160。
55. Zhang, L., Tong, M. H., Marks, T. K., Shan, H., and Cottrell, G. W. Sun: 自然統計を用いたサリエンスのベイズフレームワーク。Journal of Vision 8, 7 (2008), 32-32。
56. Zhang, X., Sugano, Y., Fritz, M., and Bulling, A. 自然環境下での外観ベース視線推定。Proc. CVPR (2015)、4511-4520。
57. Zhang, Y., Bulling, A., and Gellersen, H. Sideways: 現場配置ディスプレイとの自発的相互作用の視線インタフェース。Proc. CHI (2013)、851-860。
58. Zhang, Y., Chong, M. K., Müller, J., Bulling, A., and Gellersen, H. 野外での公共ディスプレイ向け視線追跡。Personal and Ubiquitous Computing 19, 5 (2015), 967-981。
59. Zhang, Y., Müller, H. J., Chong, M. K., Bulling, A., and Gellersen, H. Gaze horizon: 視線で歩行者が公共ディスプレイと対話できるようにする。Proc. UbiComp (2014)、559-563。