

令和 6 年 5 月 31 日現在

機関番号：12601
研究種目：基盤研究(C)（一般）
研究期間：2021～2023
課題番号：21K11932
研究課題名（和文）顔形状復元によるデータ生成と自己教師型補助タスクに基づく視線推定器のドメイン適応
研究課題名（英文）Gaze Estimator Domain Adaptation Based on Data Generation Through Face Shape Reconstruction and Self-Supervised Auxiliary Tasks
研究代表者
菅野 裕介（Sugano, Yusuke）
東京大学・生産技術研究所・准教授
研究者番号：10593585
交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究では、顔形状の3次元復元に基づく学習データ生成手法と特徴分離を用いたドメイン適応手法を組み合わせることで、未知の環境でも頑健に動作する視線推定モデルの構築に取り組んだ。単眼画像から顔形状を復元し様々な顔向きでレンダリングすることで学習データの多様性を拡張し、教師なしドメイン適応により生成データと実データの差を吸収した。さらに、複眼カメラ入力を用いたアピアランスベース視線推定モデルを開発し、カメラ間の相対姿勢を手がかりとした特徴変換・融合により高い汎化性能を実現した。

研究成果の学術的意義や社会的意義

本研究の学術的意義は、顔形状の3次元復元とドメイン適応を組み合わせた新しい視線推定手法を提案し、未知の環境でも高い精度を実現したことにある。また、任意の複数カメラを用いて視線推定を行うことのできる手法には前例がなく、カメラの位置関係を拘束条件として用いる特徴融合は他の課題にも応用できる可能性がある。提案手法により、多様な姿勢や環境での視線推定が可能となり、自然なインタラクションを必要とする様々なアプリケーションに活用できる。例えば、対話システムやデジタルサイネージ、自動車の運転支援など、ユーザの視線情報を用いることで、よりシームレスで直感的なインターフェースの実現が期待できる。

研究成果の概要（英文）：In this study, we combined a data generation method based on 3D face shape reconstruction with a domain adaptation technique using feature separation to develop a robust gaze estimation model that operates effectively in unknown environments. By reconstructing face shapes from monocular images and rendering them in various orientations, we enhanced the diversity of the training data. Unsupervised domain adaptation was employed to bridge the gap between generated data and real data. Additionally, we developed an appearance-based gaze estimation model using multi-camera input, achieving high generalization performance through feature transformation and fusion based on the relative orientation between cameras.

研究分野：コンピュータビジョン

キーワード：視線推定 機械学習 ドメイン適応

1. 研究開始当初の背景

機械学習ベースの視線推定手法は、低画質・低解像度な顔画像のみで頑健な視線推定を実現できる可能性を秘めている。しかし他の多くの画像認識手法と同様に、学習時と推論時の画像撮影環境が異なると推定性能が大きく低下するという課題を抱えていた。視線推定の性能に影響を与える条件は複数存在するが、中でもカメラから見た相対的な頭部姿勢および視線方向の分布の違いが最も本質的な性能低下要因となる。学習時のデータに含まれていない視線方向が推論時に現れた場合、外挿補間が必要になるため推定精度が大きく低下する傾向にある。カメラ座標系における計測対象人物の頭部姿勢や視線方向の分布は応用シナリオや設置環境に依存して大きく異なるため、これは深刻な精度低下要因になり得る。しかし未知の設置環境に適応的に対応しながら高精度な視線推定を行うための手法は未だ確立されていなかった。

2. 研究の目的

本研究の目的は、頭部姿勢や視線方向の分布が異なる対象ドメインに対して人物非依存の視線推定モデルを適応するための枠組みを実現することである。具体的には、3次元視線方向のラベルが付与された2次元画像から復元・レンダリングした画像を視線推定モデルの学習に使用し、視線ラベルのない通常の顔画像を手がかりに特徴学習やドメイン適応を行う。これら二つの要素を統合することで、教師情報のない対象ドメインへの視線推定モデルの適応を目指す。

3. 研究の方法

本研究では以下の2つのアプローチを主軸に進める。第一に、3次元視線方向のラベルが付与された既存の2次元顔画像データセットに3次元顔形状復元手法を適用し、対象ドメインの頭部姿勢分布に一致した学習データを生成する。第二に、生成画像と実画像の見え方の違いを吸収するために、自己教師型の擬似タスクを用いて特徴抽出器をドメイン適応する手法を開発する。さらに、対象ドメインや対象タスクの教師ラベルを用いることなく、視線推定モデルの適応を行う手法について広く検討する。

4. 研究成果

生成データによる訓練とドメイン適応

本研究では単眼3D顔再構成に基づく、学習による合成を用いた新しい外観ベースの視線推定手法を提案した。従来の視線推定用データセットは、頭部姿勢や視線分布の多様性が不足していることが課題であった。我々は既存のデータセットから3D顔形状を再構成し、新しい頭部姿勢での顔画像を合成することでこの課題に取り組んだ。本手法の新規性は、限られたデータセットを拡張し、実環境に近い多様な頭部姿勢と視線分布を持つ大規模な訓練データを生成できる点にある。

提案手法の主要な特長として以下の3点が挙げられる。第一に、投影的整合と呼ぶ手法を導入し、3D顔再構成結果を入力画像のカメラ座標系に正確に位置合わせできるようにした。新しい視点から顔画像を合成する際にも正確な視線ラベルを維持できる。第二に、合成データ生成時に得られる顔領域マスクを教師信号として用いることで、合成訓練データの特性を十分に活用しモデルの推定精度を高める手法を提案した。第三に、画像のレンダリング時に照明や背景をランダムに変化させる手法を導入し、合成データの画像的な多様性を高めた。

提案手法の有効性を検証するため、複数の公開データセットを用いた実験を行った。その結果、提案手法により元のデータセットの視線範囲を大きく拡張できることを確認した。また実画像のみを用いた場合や既存手法による合成データを用いた場合と比較して、提案手法により高い視線推定精度が得られることを示した。特に頭部姿勢と視線分布が大きく異なるデータセット間のクロス評価において、提案手法による精度向上が顕著であった。本手法は視線推定モデルの訓練データを拡張し、実環境での頑健性を高めるための有力なアプローチになり得ると考えら

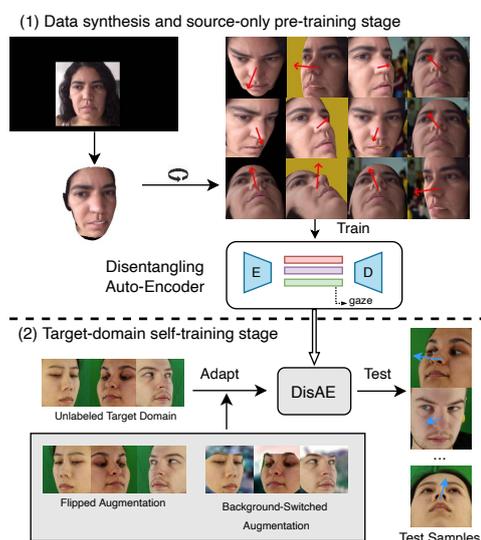


図1 生成データによる訓練

れる。

さらにこの研究を元に、合成データと実データの差を埋めるための教師なしドメイン適応手法を新たに提案した(図1)。本研究では合成データによるソースドメインから視線関連の表現を学習するための、新しい特徴分離オートエンコーダ(DisAE)を提案した。さらに、DisAEと自己学習を組み合わせた教師なしドメイン適応のための新しいアプローチを提案し、実験によりその有効性を示した。DisAEでは、まず合成ソースデータを用いて外観、頭部姿勢、視線の特徴を分離して学習する。次に、ラベルなしのターゲットデータに対して、背景切り替え整合性損失を用いた自己学習によりDisAEをさらに適応させる。これは背景をランダムに切り替えた画像でも視線推定の結果は一致するという拘束条件をモデルに課すことを意味し、結果として視線推定モデルが顔領域に注目することが期待される。さらに実験では多視点顔再構成との詳細な比較を行い、単眼再構成アプローチの性能を分析した。多視点顔再構成には高品質な形状再現という利点がある一方、同期された多視点が必要といった実用上の制約もある。単眼再構成は簡便でありながら多視点に近い性能が得られる可能性を秘めている。

アイコンタクト検出モデルの教師なし学習

本研究では、実世界の動画におけるビデオ非依存のアイコンタクトセグメンテーションという新しい課題に取り組んだ(図2)。任意の動画に適用でき、事前に注視対象を知らず、対象人物のアイコンタクトの瞬間を得ることができる統一モデルの学習を目指した。

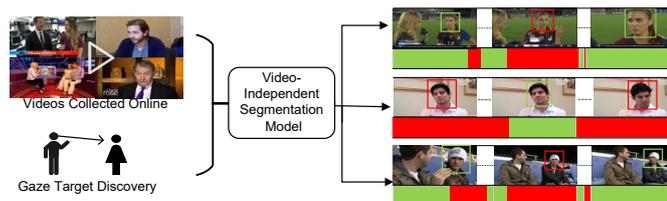


図2 アイコンタクト検出

提案手法の主な特長は以下の3点である。

第一に、制約のない動画でフレーム単位のアイコンタクトラベルを取得するための新しい注視対象発見手法を提案した。これによりセグメンテーションモデルを教師なしで学習できる。第二に、手動でアノテーションを行った52本の動画からなる評価用データセットを公開した。第三に、提案手法により従来のビデオ依存の訓練手法を上回り、特に非正面の顔のトラックレットにロバストなビデオ非依存のアイコンタクト検出器を実現できることを示した。

具体的には、まずアピアランススペースの視線推定結果からノイズの多い疑似ラベルを生成するための、教師なしの注視対象発見手法を提案した。オンライン動画にはカメラの動きや人工的な編集が含まれることが多いため、カメラを基準に注視対象を特定するのは簡単ではない。そこで、注視対象と人物の相対位置が固定されていると仮定し、体の座標系で人間の視線を分析し、高密度の注視点領域を陽性サンプルとして扱う手法を提案した。次に、VoxCeleb2データセットに注視対象発見手法を適用し、大規模な疑似ラベル付き学習データセットを取得した。初期の疑似ラベルに基づき、オリジナルの顔特徴を入力として、セグメンテーションモデルを反復的に学習した。

実験では、提案手法が評価データセットで71.88%の検出精度を達成し、従来のビデオ依存のアイコンタクト検出器を上回ることを示した。特に非正面の顔を含むデータにおいて良好な性能を示すことが確認できた。一方で、移動する注視対象や人物への対応、極端な頭部姿勢での性能など、いくつかの課題も明らかになった。以上のように、本研究ではビデオ非依存のアイコンタクトセグメンテーションという新しい課題を提案し、教師なしの注視対象発見とセグメンテーションモデルの反復学習による解決手法を示した。提案手法は多様な環境下での人間の社会的行動理解に役立つと期待される。

未知の環境に適応可能なマルチビュー視線推定

本研究では、マルチビューのアピアランススペース視線推定のための新しい手法を提案した(図3)。入力画像間の相対的な回転行列を制約条件として用いて、ある画像の特徴を他の画像へ転送する際の特徴融合アプローチを提案した。提案手法は未知のカメラ設置環境や頭部姿勢に対する汎化性能を向上させる利点がある。

提案手法の主要な特長は以下の3点である。第一に、相対的な回転行列を追加の入力情報として利用するカメラ非依存のマルチビュー視線推定という新しい課題に取り組んだ。第二に、マルチビュー視線推定において、相対的な回転行列を特徴間の転送時の制約として組み込む新しい特徴融合アプローチを提案した。提案手法は回

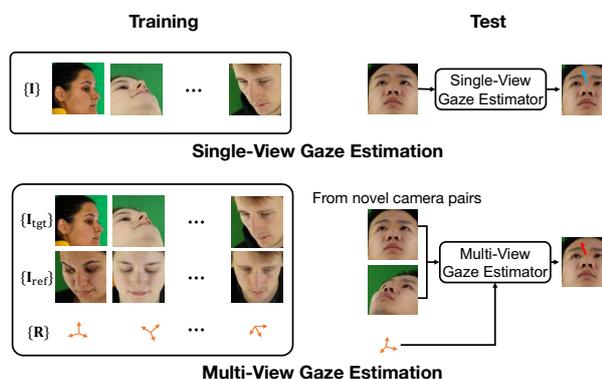


図3 マルチビュー視線推定

転行列を用いて特徴を画像間で転送する際の制約とし、内部の特徴表現について詳細な分析を行った。第三に、マルチビュー視線推定により未知の頭部姿勢に対する汎化性能が向上することを示した。

具体的には、入力画像の相対的な回転行列を追加の入力として、複数の回転制約付き特徴融合ブロックを積み重ねたネットワークを提案した。各ブロックでは、一方の特徴に回転行列を乗算することで、他方の画像の特徴空間に転送する。本来物理的な回転を特徴空間に直接適用することはできないが、回転操作を明示的に組み込んだ学習プロセスを通じて、回転可能な特徴表現を学習できると期待される。

提案手法の有効性は複数のデータセットを用いた実験により検証した。最先端のドメイン汎化手法を含むベースラインよりも優れた汎化性能が得られることを確認した。また回転制約を通じて得られた内部特徴表現について詳細な分析と可視化を行った。未知のカメラ位置に対する汎化性能は実用上重要な指標であり、提案手法はこの点でベースラインを上回る性能を示した。以上のように、本研究では相対的な回転行列を制約として用いる新しいマルチビュー視線推定手法を提案し、その有効性を実証した。提案手法は多様な環境下での頑健な視線推定の実現に貢献すると期待される。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 Qin Jiawei, Shimoyama Takuru, Sugano Yusuke	4. 巻 -
2. 論文標題 Learning-by-Novel-View-Synthesis for Full-Face Appearance-Based 3D Gaze Estimation	5. 発行年 2022年
3. 雑誌名 Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)	6. 最初と最後の頁 4977-4987
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/CVPRW56347.2022.00546	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Wu Tianyi, Sugano Yusuke	4. 巻 13844
2. 論文標題 Learning Video-Independent Eye Contact Segmentation from In-the-Wild Videos	5. 発行年 2023年
3. 雑誌名 Lecture Notes in Computer Science (ACCV2022)	6. 最初と最後の頁 52-70
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-031-26316-3_4	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Hisadome Yoichiro, Wu Tianyi, Qin Jiawei, Sugano Yusuke	4. 巻 1
2. 論文標題 Rotation-Constrained Cross-View Feature Fusion for Multi-View Appearance-based Gaze Estimation	5. 発行年 2024年
3. 雑誌名 Proc. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)	6. 最初と最後の頁 5973-5982
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/WACV57701.2024.00588	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計7件（うち招待講演 2件/うち国際学会 3件）

1. 発表者名 菅野裕介	
2. 発表標題 人にひらかれたメディア理解に向けて 人を理解する、人と理解する	
3. 学会等名 電子情報通信学会 パターン認識・メディア理解研究会（PRMU）（招待講演）	
4. 発表年 2022年	

1. 発表者名 Yoichiro Hisadome、Yusuke Sugano
2. 発表標題 View-consistent Feature Alignment for Multi-view Appearance-based Gaze Estimation
3. 学会等名 第25回 画像の認識・理解シンポジウム
4. 発表年 2022年

1. 発表者名 Qin Jiawei、Shimoyama Takuru、Sugano Yusuke
2. 発表標題 Learning-by-Novel-View-Synthesis for Full-Face Appearance-Based 3D Gaze Estimation
3. 学会等名 4th International Workshop on Gaze Estimation and Prediction in the Wild (国際学会)
4. 発表年 2022年

1. 発表者名 Wu Tianyi、Sugano Yusuke
2. 発表標題 Learning Video-Independent Eye Contact Segmentation from In-the-Wild Videos
3. 学会等名 16th Asian Conference on Computer Vision (ACCV2022) (国際学会)
4. 発表年 2022年

1. 発表者名 下山拓流、菅野裕介
2. 発表標題 ラベル分布の異なるドメインに対するアピアランスベース視線推定モデルの教師無し適応
3. 学会等名 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU)
4. 発表年 2021年

1. 発表者名 Hisadome Yoichiro, Wu Tianyi, Qin Jiawei, Sugano Yusuke
2. 発表標題 Rotation-Constrained Cross-View Feature Fusion for Multi-View Appearance-based Gaze Estimation
3. 学会等名 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (国際学会)
4. 発表年 2024年

1. 発表者名 Yusuke Sugano
2. 発表標題 Toward Appearance-Based Gaze Estimation Open To Diverse People And Environments
3. 学会等名 2023 ACM Symposium of Eye Tracking Research & Applications (ETRA) (招待講演)
4. 発表年 2023年

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 学習モデル生成プログラム、情報処理装置及び学習 モデル生成方法	発明者 菅野 裕介、久留 陽 一郎、呉 天一、秦 嘉偉	権利者 同左
産業財産権の種類、番号 特許、2023-077924	出願年 2023年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	秦 嘉偉 (Qin Jiawei)		
研究協力者	呉 天一 (Wu Tianyi)		

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	久留 陽一郎 (Hisadome Yoichiro)		

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関