

機械学習による視線推定と その実世界応用

菅野裕介

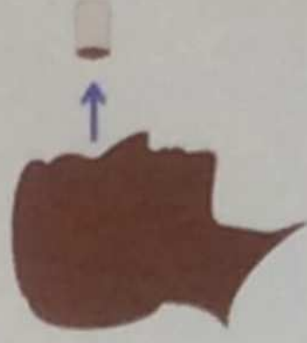
大阪大学大学院 情報科学研究科

視線推定・アイトラッキング

- 人物がどこを見ているかを計測・推定する
 - ▶ Alternative User Interfaces
 - ▶ 広告・マーケティング調査
 - ▶ 人間の内部状態を推定するための特長として
- 従来の手法は専用のハードウェアを必要とする



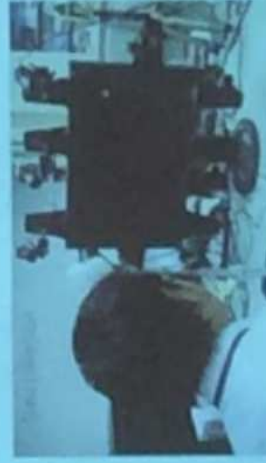
カメラベースの視線推定



- 通常のカメラから、画像中に写っている人物の (3次元) 視線方向を推定する
 - ウェブカメラ、ウェアラブルカメラ
- 従来手法では実現できないアプリケーションが多数
 - HRI, 一人称視点映像解析、公共空間での注視推定

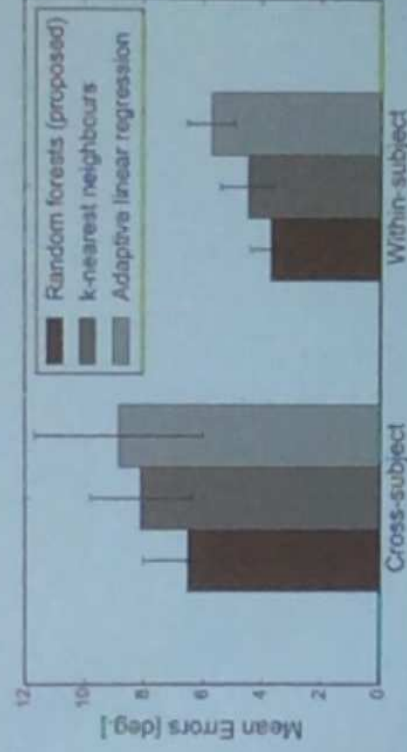
Multi-view gaze dataset [Sugano et al., CVPR14]

- 8台のカメラを同期して撮影した視線真値付きの目画像データセット
 - $50人 \times 160 (16 \times 10)$ 視線方向
- 3D reconstruction can be used to synthesize training data
 - $50 \times 160 \times 144$ 姿勢 = 1,152,000枚の目画像
- 人物非依存の視線推定関数を学習することが可能

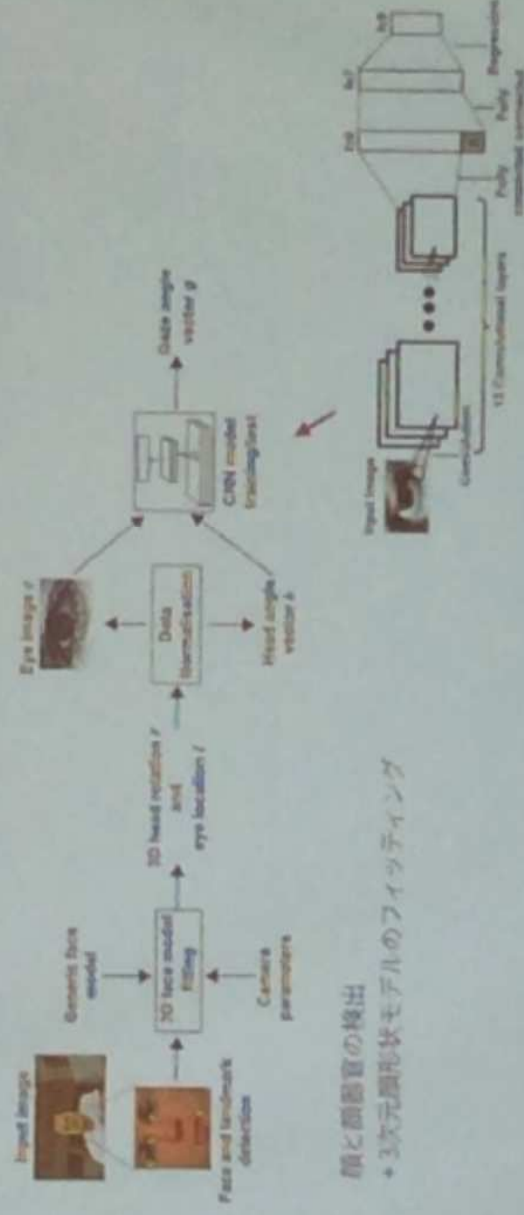


データセット上での推定性能

- 平均誤差: 6.5 度
 - 3-fold cross-person validation: 生成データで学習、実データで推定
 - 個人ごとのキャリブレーションを行っていないことを考えると十分な性能?



畳み込みニューラルネットワークによる視線推定



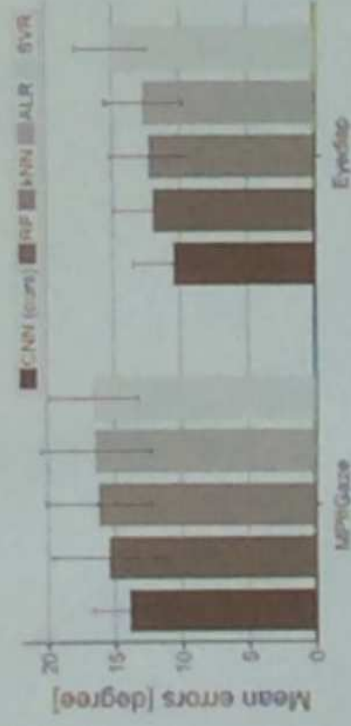
顔と顔形状モデルのフィッティング
+ 3次元顔形状モデルのフィッティング

CNNアーキテクチャ

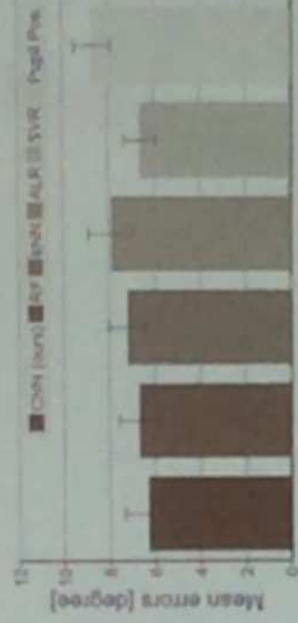
(目画像入力、全結合層で顔部姿勢ベクトルを結合)

性能の比較

Cross-domain evaluation (Training on UT Multiview)



Within-domain evaluation (on MPIiGaze)



• 学習データセットの違いによる性能の違いが顕著

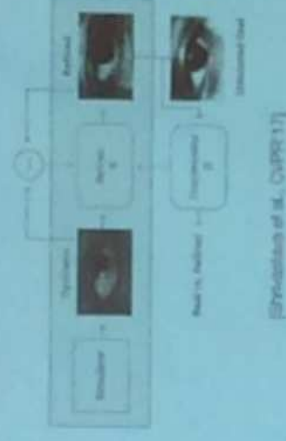
- 学習データが対象ドメインから得られている場合、同等の性能が得られる
- 異なるデータセットでの学習条件の場合にCNNベースのアプローチが有利

まとめ

- 機械学習アプローチによる視線推定の試み
 - ▶ 大規模なデータセットから、人物非依存の視線推定器を学習する
 - ▶ 学習・評価データセット構築が重要な研究タスクの一つ
 - ▶ 3DCGによる生成画像の利用
 - ▶ 顔画像全体を入力特徴とした視線推定のためのCNNアーキテクチャ
- 実環境応用に向けたシステム設計
 - ▶ HCIの観点から学習環境とテスト環境の違いに対応する
 - ▶ 対象ユーザ・設置環境から自動的に学習データを獲得するための枠組み

今後の展望

- シミュレーション・生成データによる学習
 - ▶ 学習データ不足。ドメイン適応の困難さはCV全体でも重要な課題の一つ
 - ▶ GANによる画像変換を行う例
- 顔画像からの視線推定
 - ▶ 顔全体をCG生成して学習することは可能?
 - ▶ 実画像データセットも不足している
- 学習データ獲得のためのシステムデザイン
 - ▶ ユーザ・環境の多様性にHUI的な観点から取り組むことは他の認識タスクでも重要
 - ▶ 学習アルゴリズムからユーザインタフェースまで一貫した視点で取り組むことで新たな地平が開けるのでは?



[Presentation of ML, CVPR'17]