

低解像度顔画像群を用いた集団の注目位置推定に関する予備的検討

A Preliminary Study on Gaze Target Localization from Low Resolution Faces of Group of People

児玉 祐樹¹ 川西 康友¹ 平山 高嗣¹ 出口 大輔¹ 井手 一郎¹
Yuki Kodama Yasutomo Kawanishi Takatsugu Hirayama Daisuke Deguchi Ichiro Ide
村瀬 洋¹ 永野 秀尚² 柏野 邦夫²
Hiroshi Murase Hidehisa Nagano Kunio Kashino

名古屋大学¹
Nagoya University

日本電信電話株式会社 NTT コミュニケーション科学基礎研究所²
NTT Communication Science Laboratories, NTT Corporation

1 はじめに

人々が見ている場所を知ることは、小売店における客の購買行動の分析や、デジタルサイネージの広告効果の測定など様々な用途で有用である。特に、各個人の傾向ではなく、全体としての傾向を分析したい場合には、複数の人物が同時に注目している対象を知ることが重要である。しかし、複数の人物を同時に観測するには広範囲の映像を撮影する必要があり、各人物画像の解像度は低くなる。そのため、各人物の高精度な視線情報の取得は困難となる。

この問題に対し、複数人が同一の対象を同時に注視する状況を想定し、集団としての大まかな視線情報の傾向を分析することによって、集団が注目している対象が存在する位置を高精度に推定する手法を提案する。本発表では、画像中の各人物の注目対象位置を推定した低精度な推定結果を統合することで集団の注目対象位置を精度良く推定する手法を提案する。

2 提案手法

まず、複数人が同一の注目対象を注視している画像から各人物の顔領域を検出し、各顔画像から注目対象位置を推定する。そして、得られた各推定結果を統合することで、集団の注目対象位置を推定する。本発表では、注目対象の位置推定を、予め定めた9つの視標表示位置(図1)のいずれかを認識する問題として定式化する。

顔領域の検出には HOG 特徴を利用した検出手法 [1] を用いる。得られた各顔画像を事前に学習した Convolutional Neural Network (CNN) に入力し、それぞれの顔画像に対する各注目対象位置の尤度を求める。最後に、各顔画像に対する注目対象位置の尤度を統合し、最終的な出力結果とする。具体的には、各顔画像に対する尤度の平均が最大となる注目対象位置を全体の推定結果とする。

3 実験

提案手法の有効性を確認するため、注目対象位置の推定精度に関する評価実験を行った。実験では、注目対象位置の尤度を統合する人数を増加させた際の認識率の変化を確認した。

まず、データセットを作成するため、12名の被験者による被験者実験を行った。事前に設定した6箇所の被験者位置から、前方に提示された視標を6名の被験者が同時に注視している様子を被験者の組み合わせを変えなが

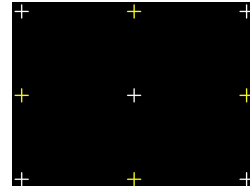


図1 視標表示位置



図2 画像例

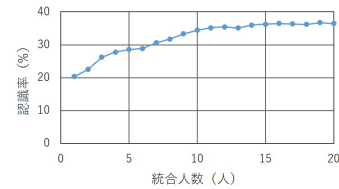


図3 実験結果

ら6,480枚撮影した。視標は、図1に示す9箇所にランダムに表示した。撮影した画像例を図2に示す。各被験者は全ての被験者位置で撮影した。検出された顔画像サイズは51×51～75×75画素であった。得られた被験者12名の顔画像を6名ずつCNNの学習データとテストデータに分割した。テストデータに使用した被験者6名は全ての被験者位置において顔が検出できた被験者であり、使用した顔画像は各被験者・被験者位置・視標表示位置の組み合わせごとに1枚ずつ選んだ324枚とした。統合人数が7人以上の場合では、仮想的に同じ被験者位置に複数の被験者が存在する状況を設定し、異なる位置に座った同一被験者を異なる被験者として扱った。

統合する人数の増加にともなう認識率の変化を図3に示す。この結果から、統合を行わない場合(統合人数1人)の認識率が約20%であるのに対し、統合人数の増加にともない認識率が向上し、統合人数15人以上では約1.8倍となる36%の認識率を達成できていることがわかる。

4 むすび

集団が注目している対象の位置を低解像度顔画像から高精度に推定する手法について検討した。今後は、推定結果の統合手法の改善や個人の注目対象位置の推定手法について検討する予定である。

謝辞 本研究の一部は科学研究費補助金による。

参考文献

- [1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", Proc. of CVPR 2005.