アイコンタクト解析のための CNN を用いた目検出

小宮 凛子　Warapon Chinsatit　齊藤 剛史
(九州工業大学)

1　はじめに

英語スピーチやプレゼンテーションなどのパブリックスピーキングの指導に関して，現状は教材作成者の主観に基づく記述が多く，音声や映像データなどを用いた定量的分析には多大な労力が必要とされてきた．本研究ではスピーチシーンに対して，映像情報から話者の動作を自動的に解析し，スピーチ指導に役立つ指標を抽出することを目的としている．

これまで Active appearance model（AAM） により検出された顔特徴点を用いて頭部姿勢を推定することで，頭部姿勢が新たな評価指標として利用可能か検討した [1]．一方，スピーチシーンの従来の評価項目にはアイコンタクトがある．頭部姿勢を既存の評価指標と比較するためには，視線を推定するために黒目を検出して目の動きを解析できることが望ましい．そこで本稿では黒目位置の検出に取り組む．画像中から黒目や瞳孔を検出する研究は古くから取り組まれているが，本研究で解析対象とするシーンは目検出の研究対象とされている一般的な画像に比べて，画像フレームにおける目のサイズが小さい．そのため，目が鮮明ではない，目が細いなどの問題がある．本稿では近年注目されている Convolutional Neural Network（CNN）を用いた手法を検討する．

2　CNN を用いた目検出手法

本稿では，著者らは瞳孔検出を目的として提案した，CNN を用いた目検出法 [2] を適用する．CNN のモデルは，入力層 1 層，畳み込み層 5 層，プーリング層 2 層，出力層 1 層の全 9 階層で構成されている．プーリング層には正規化処理を適用し，活性化関数は Rectified Linear Unit を用いる．入力画像として目画像を与えることで，CNN により検出された黒目座標 P_{CNN} を出力する．

3　評価実験

CNN の学習には，先行研究 [2] と同様に公開データセットである SynthesEyes を用いる．SynthesEyes は 10 名の被験者の 3D 頭部スキャンモデルから生成された画像サイズ 120×80 画素，11,382 枚の目画像から構成されている．SynthesEyes には瞳孔輪郭の特徴点が用意されており，その重心座標を正解位置として与えて学習した．

本実験では，[1] で用いた福岡県内の高校で開催された英語弁論・暗唱大会における 9 名のスピーチシーンを評価実験に利用した．画像サイズは 854×480 画素であり，全フレーム画像数は 64,926 枚である．[1] では AAM により目周囲の 4 特徴点を検出している．これらを用いて目の重心点 P_G を求め，P_G を中心として 24×16 画素のサイズで切り取り，5 倍に拡大して 120×80 画素の目画像を生成した．これを CNN に入力し，P_{CNN} を求めた．

黒目検出結果を定量的に評価するために，スピーチシーンの全フレーム画像から 1 秒間隔で 1,559 枚の目画像を選出した．選出した目画像には目を閉じている状態があった．閉じている目は評価が難しいため，これらを除いた 1,336 枚の目画像を選び，正解黒目座標 P_{GT} を与えた．ここで P_{GT} は 4 人が目視で正解黒目座標を与え，その平均とした．評価基準として，P_{CNN} と P_{GT} 間のユークリッド距離 $error = |P_{CNN} - P_{GT}|$ を求めた．その結果を図 1 に示す．図中，横軸は誤差，縦軸はその誤差以下の画像枚数の比率を示す．誤差 5 画素以下を検出成功とすると成功率は 90%であった．1,336 枚の平均誤差は 3.13 画素であった．

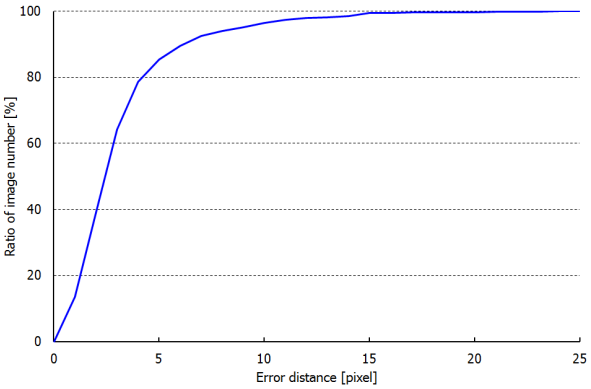

図 1: 検出結果


図 2: P_{CNN} と P_{GT}

黒目検出結果例を図 2 に示す．図中の赤点は P_{GT}，緑点は P_{CNN} である．各画像の $error$ は左から 2.3 画素，1.9 画素，5.3 画素であった．図 2 左は黒目が少し左側に位置しており，図 2 中は右上に黒目が位置している．これらの結果より，黒目と白目が区別できる場合は黒目の位置にかかわらず正しく検出できている．一方，図 2 右は目の中央に黒目が位置しているが，目が細く黒目が不鮮明であるため誤差が大きくなったと考えられる．

4　まとめ

本稿ではスピーチシーンに対して話者のアイコンタクトを解析するために，CNN を用いた黒目検出に取り組んだ．その結果，高い検出精度が得られ，有効性を確認した．今後は検出された黒目位置を用いた視線，アイコンタクトの解析に取り組む．

謝辞

本研究の一部は，JSPS 科研費 15K12416 の助成によるものである．

参考文献

[1] Rinko Komiya, Takeshi Saitoh, Miharu Fuyuno, Yuko Yamashita, and Yoshitaka Nakajima, “Head Pose Estimation and Movement Analysis for Speech Scene”, Proc. of 15th IEEE/ACIS International Conference on Computer and Information Science (ICIS2016), pp.409–413, 2016.6.

[2] Warapon Chinsatit, Takeshi Saitoh, “Pupil Center Estimation by using Deep Convolutional Neural Network”, MIRU2016, 2016.8.