H-003

マスク着用に対応した番組出演者の一覧提示システム

A list presentation system of masked performers on TV

河合 吉彦 †　　望月 貴裕 †
Yoshihiko Kawai†　Takahiro Mochizuki†

1　はじめに

テレビ放送局では，よりよい番組作りのために，番組出演者の年代や性別，出演時間，出演時期などを解析したいという要望がある．しかしながら，これらの情報を人手で解析するためには，多くの作業時間が必要になるという課題がある．そこで本稿では，顔検出および顔認識技術を活用した番組出演者の一覧提示システムを提案する．なお近年は，番組出演者がマスクを着用しているケースが多いため，顔検出，顔認識にはマスク着用に対応した手法を利用する．実験では，試作したシステムを実際のテレビ番組に適用し，有効性を検証する．

2　番組出演者の一覧提示システム

提案システムにおける処理の流れを図 1 に示す．始めに，番組映像から一定の時間間隔でフレーム画像を抽出し，各画像から顔が映る位置を検出する．本稿の実験では抽出間隔を 1 秒に設定した．次に，検出した顔画像から顔を認識するための特徴量を算出する．最後に，算出した特徴量をクラスタリングし，各クラスタの代表画像をユーザーに一覧提示する．

2.1　マスク着用への対応

マスク着用に対応した顔検出・顔認識用の深層ニューラルネットワーク（DNN）を学習するためには，マスク着用者の画像が含まれた大規模な学習データが必要となるが，数十万から数百万という規模の実画像を収集することは容易ではない．そこで本手法では，マスク未着用者の顔画像にマスクを合成する手法 [1] を利用する．

図 2 にマスクの合成手順を示す．合成には，マスク着用者が映ったテンプレート画像を利用する．まず手順 1 では，テンプレート画像と入力画像から 23 点の顔特徴点を検出する [2]．手順 2 では，顔特徴点を頂点としてテンプレート画像をドロネー三角形に分割する．手順 3 では，テンプレート画像における各々の三角形領域を，入力画像の対応する三角形領域の形状にあわせてアフィン変換していく．手順 4 では，マスクの彩度と明度を重畳先の入力画像にあわせて調整し，合成時の違和感を軽減する．また，必要に応じて色相をランダム変換することで，学習データのバリエーションを増やす．最後に，入力画像に重畳して出力結果とする．本手法では，鼻や輪郭の形状を考慮してマスクを細かく変形することで，実画像におけるマスクの変形を模擬している．

この合成画像を，学習データに一定の割合だけ混入することで，マスク着用に対応した DNN を実現する．

図 1　提案システムにおける処理の流れ

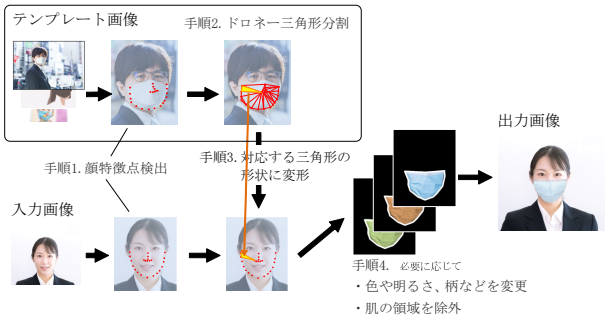


図 2　マスクの合成手順

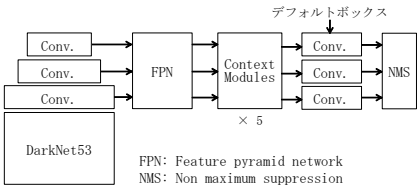


図 3　顔検出用 DNN の構造

2.2　顔検出処理

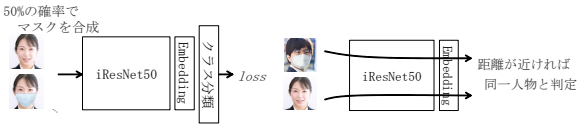
顔検出は，フレーム画像から顔が映る位置（バウンディングボックス，bbox と記載）を特定する処理である．提案システムでは，RetinaFace[3] をベースとした DNN 構造を採用する．図 3 に，顔検出用 DNN の構造を示す．バックボーンには，ResNet の代わりに DarkNet53[4] を利用する．バックボーンからの出力を，FPN に入力し，解像度の異なる特徴マップの情報を統合する．次に，context module[3] と呼ばれる構造の畳み込み層に入力する．最後に，デフォルトボックスからの移動量を畳み込み層で回帰することで顔の bbox を推定する．

学習データは，NHK で放送された約 2 週間分の番組映像から作成した．具体的には，約 16 万枚のフレーム画像に対して，約 65 万の bbox 情報を付与した．学習の際は，各画像の bbox に対して，30% の確率でマスクを合成した．表 1 に，マスク実画像データセット [5] に対する評価結果を示す．従来技術に比べ，マスク着用者に対する顔検出の精度が向上していることが確認できた．

† NHK 放送技術研究所

表 1　顔検出技術の評価結果

手法	再現率	適合率	F 値
Viola-Jones	14.73	66.96	24.15
RetinaFace[3]	81.36	37.25	51.10
提案手法	80.60	84.20	82.36



(a) 学習時　　(b) 推論時

図 4　顔認識用 DNN の構造

表 2　顔認識技術の評価結果

手法	TPR@FAR=0.2%
Anwar[7]	82.78
ElasticFace-Aug[8]	92.45
提案手法	96.38

2.3　顔認識処理

顔認識は，検出された顔が誰であるかを特定する処理である．図 4 に，顔認識用 DNN の構造を示す．学習時は，特徴抽出を担うバックボーンの後段に，クラス分類用の全結合層を連結する．学習データは，人物 ID が付与された顔画像であり，クラス分類層から正しい人物 ID が出力されるように学習する．推論時は，バックボーンの最終出力を，人物の特徴表現が埋め込まれた embedding として利用する．具体的には，算出された embedding の距離が近ければ同一人物，遠ければ別人と判定する．提案手法では，バックボーンに iResNet[6] を利用する．iResNet は非線形化や正規化のタイミングや頻度が調整されており，安定した学習が期待できる．

学習データは，NHK で放送された約 10 か月分の番組映像から作成した．作成した学習データは約 660 万枚，約 11 万人分となった．学習の際は，各画像に 50% の確率でマスクを合成した．

表 2 にマスク実画像データセット [7] に対する実験結果を示す．評価指標は，他人を同一人物と誤認識する割合（FAR）を 0.2% とした際に，同一人物を正しく認識できる割合（TPR）とした．実験の結果，既存手法に比べて認識精度が向上することが確認できた．

2.4　クラスタリング処理

顔認識用 DNN から出力された embedding をクラスタリングすることで，番組出演者の顔画像をクラスタに分割する．出演者数は未知であるという想定のため，提案手法ではクラスタ数をあらかじめ設定する必要がない DBSCAN アルゴリズムを採用する．

2.5　一覧提示システムの画面

図 5 に提案システムの画面を示す．画面（a）がメイン画面であり，人物（クラスタ）の代表画像が，番組内での登場回数，平均サイズの情報とともに一覧表示される．画面（a）の顔画像をクリックすると画面（b）に遷移し，その人物の顔画像が登場時刻順に表示される．画面（b）で顔画像をクリックすると画面（c）に遷移し，その顔が映っているフレーム画像が表示される．本システムにより，番組に誰が出演していたのか，どの程度の頻度で登場していたのか，どの時刻に出演していたのか，などが簡単に把握できるようになる．

(a)　　(b)　　(c)

図 5　提案システムの画面

3　評価実験

実際の放送番組を用いて，提案手法の有効性を検証した．実験には NHK で放送された 20 分のドキュメンタリー番組を利用した．

実験の結果，番組に登場した 35 人のうち 33 人を正しく検出することができた．残りの 2 人については，別人のクラスタに包含されていた．また，同一人物が複数のクラスタに分割されるケースもあり，提案システムによるクラスタ総数は 57 となった．多い場合は，1 人が 7 つのクラスタに分割されていた．背景にたまたま小さく映りこんでいた場合や，マスクを着用している人物が真横を向いている場合などにおいて，embedding 間の距離が大きくなりクラスタが分割されるケースがあった．すべてを人手で解析することに比べると，本システムによる作業コストの軽減は有効といえるが，利用目的にあわせた改善が必要と考える．今後，DBSCAN のパラメータ調整や顔認識精度の改善などを検討したい．

4　おわりに

本稿では，顔検出・顔認識技術を活用した番組出演者の一覧提示システムを提案した．実験では，本システムを実際の放送番組に適用し，一定の有効性を確認した．

参考文献

[1] 河合, 望月: “コロナ時代の顔画像解析に向けたマスク着用学習データの合成手法”, 映メ冬大, 21A-7 (2021)
[2] A. Bulat, and G. Tzimiropoulos: “How far are we from solving the 2D & 3D Face Alignment problem?”, arXiv: 1703.07332 (2017)
[3] J. Deng, *et. al.*: “RetinaFace: Single-shot multi-level face localisation in the wild”, in Proc. CVPR (2020)
[4] J. Redmon, and A. Farhadi: “YOLOv3: An incremental improvement”, arXiv: 1804.02767 (2018)
[5] Make ML: “Mask dataset”, https://www.kaggle.com/datasets/andrewmvd/face-mask-detection
[6] I. C. Duta, *et. al.*: “Improved residual networks for image and video recognition”, arXiv:2004.04989 (2020)
[7] A. Anwar, and A. Raychowdhury: “Masked face recognition for secure authentication”, arXiv:2008.11104 (2020)
[8] M. Huber, *et. al.*: “Mask-invariant face recognition through template-level knowledge distillation”, in Proc. IEEE ICAFGR (2021)