7T-03

映像監視システムにおけるベストショット抽出のための顔画像のスコアリング方式

久米孝[†1]　皆川純[†2]　山﨑賢人[†2]　阿倍博信[†1]
東京電機大学[†1]　三菱電機株式会社[†2]

1. はじめに

近年，映像監視システムの利用が広がっている[1]．特にディープラーニングを使用した画像解析機能や，顔認証技術はその性能が向上しており，防犯やマーケティングへの活用が進んでいる．顔認証技術は顔の特徴を抽出して使用しているが，2000 年頃からは顔の特徴を 3 次元情報として抽出する研究が盛んである[2]．顔の 3 次元情報の抽出精度は顔の向き，照明環境，表情などによって大きく左右される．そのため顔認証システムでは，監視カメラの顔の検出結果である顔画像を顔の向き，照明環境，表情などの顔の特徴を用いてスコアリングを行い，スコアが最大のベストショットに対して顔照合を行うことで顔認証精度の向上を図っている[3]．また，顔の特徴を用いて再構成された顔の 3 次元モデルは顔の向きを自由に変更できるため，監視インタフェースにも利用可能である．

本研究では 3 次元顔再構成による顔の 3 次元モデルを活用した映像監視システムの構築を目的として，監視カメラ映像から顔検出した顔画像に対して 3 次元顔再構成に適したスコアリング方式を設計し，本方式を用いた最大スコアの顔画像（ベストショット顔画像）の抽出を対象とする．抽出したベストショット顔画像に対して 3 次元顔再構成処理を行い顔の 3 次元モデルを構築することにより，監視インタフェースや顔認証システムでの活用を図る．本論文ではベストショット顔画像を抽出するためのスコアリング方式について提案する．

2. 関連研究

本論文で対象とする顔画像に対する 3 次元顔再構成技術は Deng らの開発した Deep3DFaceReconstruction[4]が代表的なものとしてあげられる．Deep3DFaceRecognition における 3 次元顔再構成では，入力された 1 枚の顔画像と顔の特徴点から顔の形状を予測する．さらに顔画像から肌の色を取得し，結果をメッシュデータとして保存する．そのため，顔の向きが正面から外れ，写っていない顔の部分が大きいほど人物の元の顔と異なる 3 次元モデルが生成される．生成される 3 次元モデルの差異をスコアリングに使用可能であると判断し，本研究では Deep3DFaceReconstruction の使用を試みる．

A Facial Image Scoring Method for Best Shot Extraction in Video Surveillance System
†1 TAKASHI KUME, HIRONOBU ABE, Tokyo Denki University
†2 JUN MINAGAWA, KENTO YAMAZAKI, Mitsubishi Electric Corporation

3. スコアリング方式の提案

3.1 基本方針

2 章を踏まえ，本論文にて提案するスコアリング方式の基本方針について以下の通り整理する．

- 顔照合の観点から顔の向き，照明環境，表情などの顔の特徴を指標にスコアリング方式を設計する
- 顔の特徴による影響は，独立しておらず，互いに影響を及ぼしているが，今回は基本設計のため，ひとつの指標のみでスコアリングモデルを生成する
- 3 次元顔再構成において，元の顔画像の顔の向きが正面を向いているほど元の顔に近い 3 次元モデルが生成されている点に着目し，3 次元顔再構成により生成した 3 次元モデルを顔の向きの観点のスコアリングに使用する
- スコアリングには対象画像と基準画像に対する 3 次元再構成結果の 3 次元モデル同士を顔照合することにより算出した類似度をスコアとして使用する
- 顔の向き以外の様々な指標が追加された場合でもスコアリングできるような方式を設計する

3.2 スコアリング方式の設計

基本方針に基づき，提案方式について方式設計を実施した．まず，スコアリング方式の提案にあたり，本研究で対象とする監視カメラ映像の前提条件として，以下を設定した．

- スコアリング対象の監視カメラ映像は，1 人の人物を抽出可能な映像とする
- スコアリングモデルのためには，同一人物の様々な角度から撮影された顔画像を使用可能な映像とする

以下に，その提案内容について説明する．

（1）前処理：

入力された監視カメラ映像からフレームを抽出する．抽出されたフレームから顔検出処理により顔領域を切り出し，リサイズ処理により画像サイズが一定である，顔領域に余白を付与した画像（顔領域画像）を作成する．

（2）スコアリングモデルの生成：

顔領域画像とスコアのペアを畳み込みニューラルネットワーク（CNN）を用いてモデル化することによりスコアリングモデルを生成する．図 1 に例として VidTIMIT Audio-Video Dataset[5]を用いた本研究の提案方式における顔画像の評価基準の算出方式について示す．基準画像と比較対象に対してそれぞれ 3 次元顔再構成を行い，生成された 3 次元モデ

ルに対して顔照合を行うことにより算出された類似度をスコアとして使用する．

（3）スコアリング処理，ベストショット抽出：

入力された監視カメラ映像を前処理にかけ，切り出した全ての顔領域画像を（2）で生成したスコアリングモデルに入力し，連続的にスコアリング処理を行う．スコアが最大となった顔領域画像をベストショットとして抽出する．

本研究では，3 次元顔再構成に Deep3DFaceReconstruction，類似度の計算には顔分析ライブラリである insightface[6]を用いて試作した．

図 1 評価基準の算出方式

3.3 顔画像のスコアリングモデルの生成

スコアリングモデル生成のため用いた画像として，今回は監視カメラ映像からクロップされたデータセットである ChokePoint Dataset[7]から学習データには 11 人分，495 枚，テストデータには学習データとは別の人物の画像を 1 人分，106 枚抽出した．また，本研究では深層学習フレームワークとして tensorflow[8]，Keras[9]を選択した．CNN で畳み込み層 3 層，プーリング層 3 層のネットワークを構築し，モデルを生成した．学習回数 1000 回の時点においてテストデータの平均二乗誤差は約 0.018，平均絶対誤差は約 0.111 であった．値は共に 0 に近いため，生成したスコアリングモデルは有用であると判断した．

4. 評価と考察

提案したスコアリング方式の有用性を確認するため，生成したモデルと，映像から抽出されるベストショットについて評価した．

生成したモデルの評価として，ChokePoint Dataset から 3.3 節で用いた学習用データ，テストデータとは別のテストデータを 10 人分作成し，評価した．全体の平均二乗誤差は約 0.042 であり，平均絶対誤差は約 0.159 であった．共に 0 に近いため，監視カメラで撮影された画像において，提案方式は有用であると考えられる．

提案方式により映像から抽出されるベストショットの評価として，生成したスコアリングモデルを用いて，ベストショットを抽出し，評価した．使用した映像は ChokePoint Dataset から 56 枚 1 セットであり，人物が室内をカメラの手前方向に歩いてくる様子が撮影されている．また，画像中の人物は 1 名である．評価方法は目視において，抽出した画像が正面を向いている顔画像か評価した．図 2 にベストショット抽出の評価の結果を示す．抽出したベストショットは 36 枚目であり，正面に近い顔画像であった．しかし，映像中にはより解像度の高い正面を向いた顔が写った画像を含んでいるが，その画像はベストショットとして抽出されていない．これは，照明の影響により，近づいてきた顔が青く光ったため，スコアが 36 枚目を境に下降したと思われる．

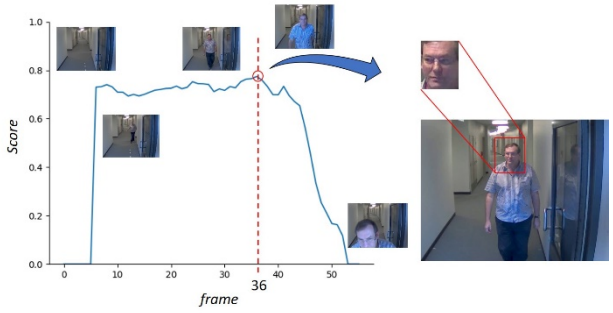

図 2 ベストショット抽出の評価結果

5. おわりに

監視カメラ映像から顔検出された顔画像のベストショットを抽出するため，顔画像をスコアリングする方式を提案した．加えて，提案方式によるスコアリングモデルを生成し，評価した．その結果，スコアリングモデルのスコア予測精度，ベストショット抽出の評価について，監視カメラ映像に対する有用性を確認した．本論文ではプロトタイプとして設計したため，今後は映像監視システムを構築するためシステム構成を検討し，システムに適したスコアリング方式を設計する．また，顔向き以外の指標を追加する．

参考文献

[1]. 山中秀昭： CCTV 監視システム技術の変遷と今後の展望，三菱電機技報，Vol. 88, No. 9, pp. 572-575 (2014).
[2]. Bronstein, A. M., Bronstein, M. M. and Kimmel, R.: Three-dimensional face recognition, *International Journal of Computer Vision*, Vol. 64, No. 1, pp. 5{30 (2005).
[3]. Xiong, L., Karlekar, J., Zhao, J., Cheng, Y., Xu, Y., Feng,J., Pranata, S. and Shen, S.: A good practice towards top performance of face recognition: Transferred deep featurefusion, *arXiv preprint arXiv:1704*.00438 (2017).
[4]. Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y. and Tong, X.: Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2019).
[5]. Sanderson, C. and Lovell, B. C.: Multi-region probabilistic histograms for robust and scalable identity inference, *International conference on biometrics*, Springer, pp. 199-208 (2009).
[6]. Deng, J., Guo, J., Yuxiang, Z., Jinke, Y., Irene, K., and Zafeiriou, S.: RetinaFace: Single-stage Dense Face Localisation in the Wild, *arXiv preprint arXiv:1905.00641* (2019).
[7]. Wong, Y., Chen, S., Mau, S., Sanderson, C. and Lovell, B. C.: Patch-based Probabilistic Image Quality Assessment for Face Selection and Improved Video-based Face Recognition, *IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops*, IEEE, pp. 81-88 (2011).
[8]. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. et al.: Tensorow: A system for large-scale machine learning, *12th {USENIX} symposium on operating systems design and implementation {fOSDI} 16)*, pp. 265-283 (2016).
[9]. Chollet, F. et al.: keras (2015).