Vol.2021-GN-112 No.17
Vol.2021-CDS-30 No.17
Vol.2021-DCC-27 No.17
2021/1/25

# 監視カメラ映像における 3 次元顔再構成のための顔画像スコアリング方式の提案

久米孝 [1]　皆川純 [2]　山﨑賢人 [2]　阿倍博信 [1]

**概要**：顔認証システムなど，様々なところで顔の 3 次元再構成への要求は高まっている．しかし，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．そこで，映像中の画像に対して顔の向きなどの様々な観点から 3 次元顔再構成のためにスコアリングすることで，最適な画像の抽出が可能となり，様々なシーンで，顔の 3 次元モデルを活用できると考えた．本論文では，第 1 弾として，顔の向きに注目し，データセットを用いた機械学習によるスコアリングモデルを生成する方式について述べる．また，生成したスコアリングモデルを評価することで，最も 3 次元顔再構成に適した顔画像の抽出方法に対して考察した．

**キーワード**：映像監視システム，顔認証，3 次元顔再構成，スコアリング

# A Face Image Scoring Method for 3D Face Reconstruction of Surveillance Videos

TAKASHI KUME[†1]　JUN MINAGAWA[†2]
KENTO YAMAZAKI[†2]　HIRONOBU ABE[†1]

***Abstract***: This paper describes a face image scoring method for 3D face reconstruction. The needs for 3D face reconstruction is increasing in various fields, such as face recognition systems. However, the accuracy of extracting 3D face information is greatly affected by the orientation of the face, the lighting environment, facial expressions, and so on. Therefore, by scoring the images in the video for 3D face reconstruction from various viewpoints such as the orientation of the face, we thought that we could extract the best images and utilize the 3D model of the face in various scenes. In this study, as a first step, we designed a method to generate a scoring model by machine learning using a dataset, focusing on the orientation of the face. By evaluating the generated scoring model, we discuss how to extract the most suitable face image for 3D face reconstruction.

***Keywords***: Video surveillance system, Face authentication, 3D face reconstruction, Scoring

## 1. はじめに

近年，安全・安心に対する意識が高まる社会環境の中，映像監視システムの利用が広がりを見せている[1]．特に，ディープラーニングを用いた画像解析機能は，性能が飛躍的に向上したことから，用途も侵入検知や動線解析などの防犯面だけでなく，人数カウントや滞留・混雑検知，人の属性抽出などのマーケティングや業務改善など，様々な場面で利用されつつある．

同様に，顔認証技術の性能も飛躍的に向上し，商業施設での利用や入国審査，犯罪者の割り出しへの活用のほか，行動パターンや振る舞いに基づく犯罪の防止への検討も急速に進んでいる．

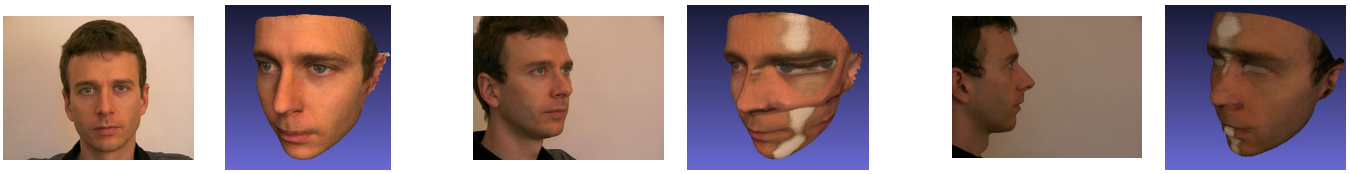
顔認証技術は，画像内の顔を検出する顔検出技術と，検出した顔が誰なのかを特定する顔照合技術の 2 つの技術から構成される．どちらも顔の特徴を抽出するが，2000 年ごろからは，顔の特徴を 3 次元情報として抽出する研究が盛んである[2]．顔の 3 次元情報の抽出精度は顔の向き，照明環境，表情などによって，大きく左右される．そのため，顔認証システムでは，顔照合の観点から顔の向き，照明環境，表情などの顔の特徴を用いて 3 次元顔再構成のためにスコアリングを行い，スコアが最大のものをベストショットとし，ベストショットに対して顔照合を行うことにより顔認証精度の向上を図っている[3]．また，顔の特徴を用い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．

このような背景のもと，本研究では，監視カメラ映像から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，この方式を用いることで，最大スコアの顔画像（ベストショット顔画像）の抽出を目指している．

本論文では，監視カメラ映像から顔を検出した結果に対してベストショット顔画像を抽出するために，顔画像のスコアリング方式について提案するが，スコアリング方式は様々な指標を基に設計する必要がある．そこで，本論文ではプロトタイプとして，指標を顔の向きにのみ注目して設計したことについて報告する．以下，2 章にて関連研究，3 章にてスコアリング方式の提案，4 章にて評価し，評価に対して考察する．そして，最後にまとめを行う．

1 東京電機大学
Tokyo Denki University
2 三菱電機株式会社
Mitsubishi Electric Corporation

Vol.2021-GN-112 No.17
Vol.2021-CDS-30 No.17
Vol.2021-DCC-27 No.17
2021/1/25



正面　　斜め　　横

図 1 PRNet による顔の向きごとの 3 次元顔再構成の例

Fig.1 Example of 3D reconstruction for each face orientation from PRNet

## 2. 関連研究

前述の通り，顔画像のスコアリングは，顔認証技術において精度向上の鍵となる技術のひとつである．今岡らの提案した顔認証技術[4]において，精度低下の要因となる照合画像と登録画像の顔の向き，照明環境の変化に対して，摂動空間法[5]を用いて 1 枚の画像から様々な顔の向き・照明条件の異なる画像を生成することで対応している．摂動空間法において，顔の向きの変動については，顔の 3 次元形状を推定し，様々な向きの顔をレンダリングすることにより画像を生成することで対応する．また，照明の変動については，拡散反射モデルをベースに構築した顔の照明基底モデルを用いて，照明条件の異なる画像を生成することで対応する．顔認証技術において，摂動空間法の適用により，顔の向き，照明環境の変動に対するロバスト性の向上は期待できるが，さらに先行研究[3]では，監視カメラで顔検出し，検出結果の顔画像に対して顔照合の観点から顔の向き，照明環境，表情などをスコアリングし，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．

次に，本論文で対象とする顔画像から再構成する 3 次元顔再構成技術は，代表的なものとして，Feng らの開発した PRNet[6]や Deng らの開発した Deep3DFaceReconstruction[7]があげられる．

図 1 に The MIT-CBCL face recognition database [8]の正面，斜め，横（右）向きの顔画像を用いて PRNet による 3 次元顔再構成の例について示す．PRNet における 3 次元顔再構成では，1 枚の画像から 3 次元頂点に対応する色を取得し，結果をメッシュデータとして保存する．図 1 を見ても分かるとおり，元の顔画像の顔の向きが正面から外れてしまうほど，再構成後の 3 次元モデルは目に見えない領域のテクスチャが自己閉塞のための歪みが大きくなってしまっていることが分かる．一方，Deep3DFaceReconstruction は顔の向きが正面から外れても，PRNet よりは視認可能な 3 次元顔再構成が可能であるため，本研究では，Deep3DFaceReconstruction の使用を試みる．

## 3. スコアリング方式の提案

### 3.1 基本方針

2 章の結果を踏まえ，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下記のとおり整理する．

- 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
- 顔の特徴による指標は，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機械学習による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比較的崩れていない，または推測が少ないため，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からのスコアリングに使用する
- スコアリング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により算出した類似度をスコアとして使用する
- 今回は顔の向きに注目したが，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

前節で設定した基本方針に基づき，本研究で提案する監視カメラ映像に対するベストショット抽出のためのスコアリング方式を設計した．

まず，スコアリング方式の提案にあたり，対象とする監視カメラ映像の前提条件として以下の条件を設定した．

- 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用いることができる映像とする

以下に，その提案内容について説明する．

**(1) 前処理：**

最初に，監視カメラ映像を入力し，フレームを抽出する．次に，抽出したフレームから顔検出処理により顔領域を切り出す．切り出した顔領域に対して，リサイズ処理により画像サイズを一定に変換し，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作成：**

基本方針に基づき，前処理後の顔領域画像に対して，基

Vol.2021-GN-112 No.17
Vol.2021-CDS-30 No.17
Vol.2021-DCC-27 No.17
2021/1/25

準画像との類似度の比較結果をスコアとして算出し，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作成する．

本研究のスコアリング方式に用いる顔画像の評価基準を算出する．図 2 には，The MIT-CBCL face recognition database を用いた例を示す．前述のとおり，スコアリングの際には，顔の向き，照明環境，表情など様々な指標が考えられるが，今回は，3 次元顔再構成により生成された 3 次元モデルの品質が顔の向きに連動する点に着目し，顔の向きの観点から学習用データを準備する．前処理後の基準画像と学習用データの顔領域画像に対してそれぞれ 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．この顔領域画像とスコアのペアを CNN により学習することでスコアリングモデルを作成する．

**(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)で作成したスコアリングモデルに入力することで，連続的にスコアリング処理を行う．この中で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．

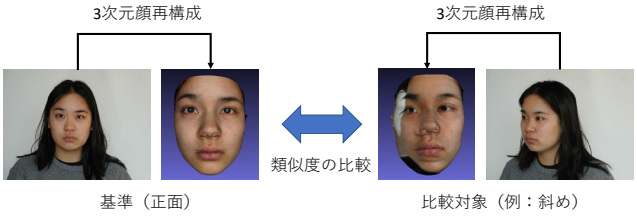



図 2 評価基準の算出方式

Fig.2 Calculation method of the evaluation criteria

### 3.3 3 次元モデルの評価

本研究では， Deep3DFaceReconstruction により顔の 3 次元モデルを生成し，類似度の計算には MXNet[9]をベースとする顔分析ライブラリである insightface[10]を用いて試作するため，Deep3DFaceReconstruction の適正を確認する．

まず，Deep3DFaceReconstruction を用いて，正面を向いた顔画像 1 枚と，画像に映る顔が小さいものや，正面を向いていないものなど，3 次元モデル作成には適さないと考えられる画像 2 枚の合計 3 枚を 1 セットとして，3 次元顔再構成をした結果を目視で確認した（図 3）．生成された 3 次元モデルは顔領域が小さくとも鮮明な 3 次元モデルが生成される．また，写っていない部分が存在する場合でも，同様に鮮明な 3 次元モデルが生成される．

次に，定量的に評価するため，監視映像から抽出した連続画像を 3 次元顔再構成し，スコアを算出した．本評価で使用した画像は，監視カメラ映像からクロップされた画像のデータセットである ChokePoint Dataset[11]である．このデータセットから単独の人物が室内でカメラの奥から手前方向に向かって歩いて来る様子が撮影されている連続画像 94 枚を用いた．この 94 枚中 79 枚目が最も正面を向いた顔画像であったため，この画像を基準として類似度を計算した．

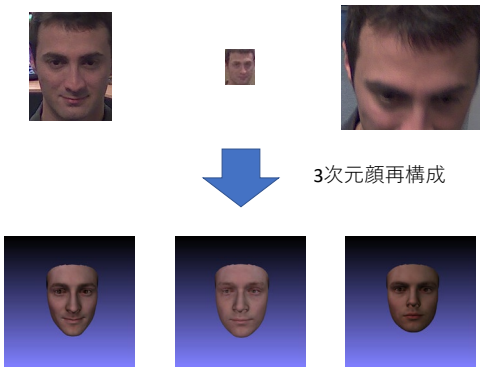



図 3 3 次元モデルの比較

Fig.3 Comparison of 3D models

3.2 節で述べたとおり，各画像から顔の領域を切り出し，切り出した顔領域の画像から 3 次元顔再構成を行った．この再構成された 3 次元モデルを用いて類似度をスコアリングした．その結果は図 4 に示すとおり，正面を向いた顔画像である 79 枚目に向かってスコアが 1.0 に近づいていき，その後，下降していることがわかる．したがって，想定したスコアが算出されたため，Deep3DFaceReconstruction は本提案方式において有効な 3 次元顔再構成手法であると判断した．

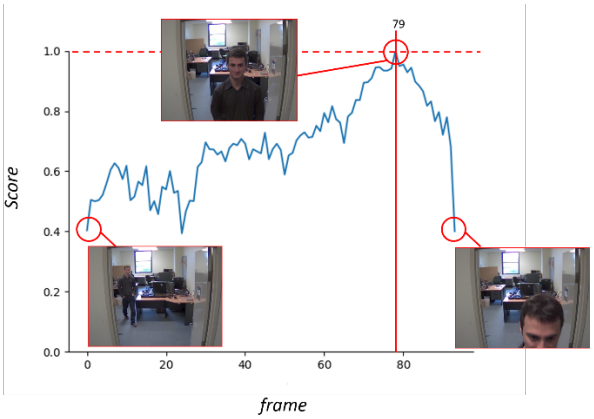



図 4 3 次元モデルのスコアの値

Fig.4 Values of 3D model scores

### 3.4 顔画像のスコアリングモデルの作成

提案方式に基づく 3 次元顔再構成結果による顔照合の類似度を機械学習にかけることで，顔画像をスコアリングするモデルの作成を行う．Deep3DFaceReconstruction により生成した 3 次元モデルを用いて，学習データと検証用のテストデータを作成した．このうち，学習データから機械学習

Vol.2021-GN-112 No.17
Vol.2021-CDS-30 No.17
Vol.2021-DCC-27 No.17
2021/1/25

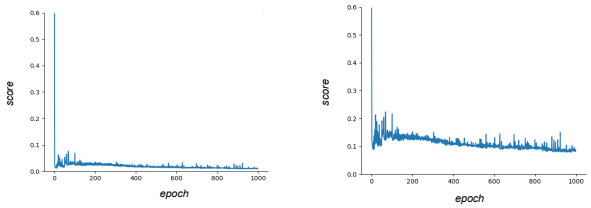
によりスコアリングモデルを作成した．

学習データの作成に使用した画像は 3.3 節同様，ChokePoint Dataset である．このデータセットより，単独の人物がカメラの奥から手前方向に歩いて来る様子が撮影されている連続画像を 11 人分，計 495 枚を抽出し，学習データとして使用した．また，テストデータには学習データとは別の人物の同様の画像を 1 人分，106 枚使用した．

学習データとテストデータの作成にあたり，3.2 節と同様の方法を用いて画像から顔の 3 次元モデルを生成した．次に，生成した 3 次元モデルを正面方向に補正し，2 次元画像に投影した．データセットから同一人物の正面を向いた画像を 1 枚目視で選択し，この画像との比較によりスコアを算出した．

作成した学習データを用いて学習するが，本研究では，深層学習フレームワークとして tensorflow[12]，Keras[13]を選択し，言語は Python を使用した．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，モデルを作成した．また，活性化関数は ReLU を選択し，学習回数は 1000 回とした．バッチサイズは 165 である．評価関数は平均二乗誤差(Mean Squared Error: MSE)と平均絶対誤差(Mean Absolute Error: MAE)を用いた．

図 5 に示すのは，テストデータの評価関数の値である．平均二乗誤差は約 0.018, 平均絶対誤差は約 0.111 であった．平均二乗誤差，平均絶対誤差は共に 0 に収束していることから，作成したスコアリングモデルは有用であると判断した．




(a) MSE　　(b) MAE

図 5 スコアリングモデルの評価関数の値

Fig.5 Value in evaluation functions of scoring model

## 4. 評価と考察

### 4.1 モデル評価

提案したスコアリング方式の有用性を確認するため，まずモデルを評価した．3.4 節では検証用に 1 人分のテストデータを用いたが，本評価では 10 人分のデータを用いて評価する．本評価で用いたデータは ChokePoint Dataset であるが，3.4 節で使用した学習データと検証用テストデータとは別のデータで試行している．

評価結果は表 1 に示すとおりである．全体の人物の MSE は約 0.042 であり，MAE は約 0.159 であった．また，人物ごとの値においても，MSE と MAE は共にばらつきが少なく，値も 0 に近い．したがって監視カメラで撮影された画像において，提案したスコアリング方式は有用であると考えられる．

使用したデータセットである ChokePoint Dataset にはカメラの奥から手前に歩く様子が撮影された画像が多く含まれている．これは監視カメラ映像特有である撮影された画像中の顔の領域サイズが安定しないことに対してロバストであるとがわかった．

表 1 監視カメラ画像によるモデル評価

Table. 1 Evaluation of Model from surveillance camera

| 人物 | 平均二乗誤差 | 平均絶対誤差 |
|---|---|---|
| 1 | 0.028 | 0.139 |
| 2 | 0.035 | 0.145 |
| 3 | 0.067 | 0.194 |
| 4 | 0.045 | 0.175 |
| 5 | 0.049 | 0.172 |
| 6 | 0.047 | 0.157 |
| 7 | 0.037 | 0.147 |
| 8 | 0.036 | 0.155 |
| 9 | 0.045 | 0.165 |
| 10 | 0.033 | 0.143 |
| 全体 | 0.042 | 0.159 |

### 4.2 ベストショット抽出評価

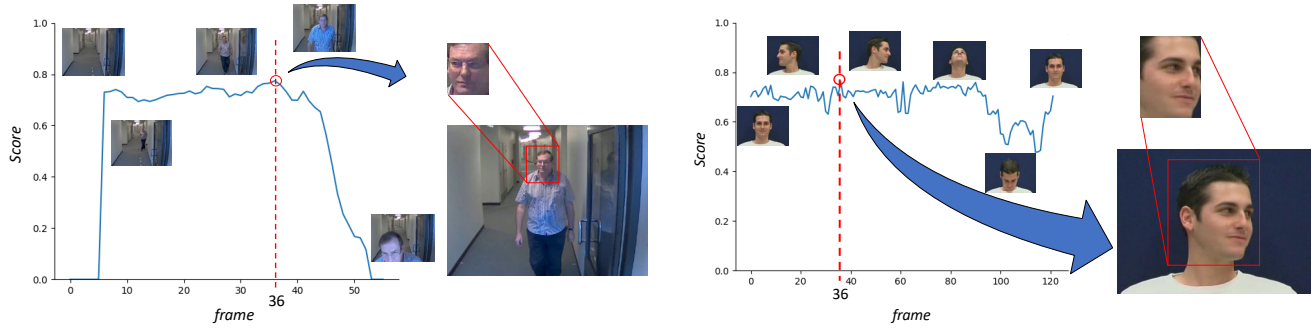
生成したスコアリングモデルを用いて，ベストショットを抽出し，評価した．使用した映像は ChokePoint Dataset から 56 枚 1 セット，人物の行動を撮影したデータセット VidTIMIT Audio-Video Dataset[14]から 122 枚 1 セットの合計 2 セットである．ChokePoint Dataset を用いたセット 1 は室内をカメラの手前方向に歩いて来る様子が撮影されている映像であり，VidTIMIT Audio-Video Dataset を用いたセット 2 は人物のバストアップにおいて正面方向から上下左右に顔の向きを変えている様子が撮影されている映像である．両セットともに画像中の人物は 1 名のものを使用した．評価方法は抽出した画像が正面を向いている顔画像かを目視で評価する．

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た結果とスコアを図 6 に示す．セット 1 のベストショットは 36 枚目であり，セット 2 のベストショットも 36 枚目であった．

セット 1 においては，正面に近い顔画像が抽出できた．しかし，セット 1 には，もっと解像度の高い正面を向いた顔が映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 枚目を境に下がっていったと考えられる．

次にセット 2 について考察する．セット 2 中に，明らか

Vol.2021-GN-112 No.17
Vol.2021-CDS-30 No.17
Vol.2021-DCC-27 No.17
2021/1/25




(a) セット 1 (ChokePoint Dataset)　　(b) セット 2 (VidTIMIT Audio-Video Dataset)

図 6 ベストショット抽出の評価結果

Fig, 6 Evaluation results of bestshot detection

に正面を向いた顔画像が含まれていたが，抽出された画像は正面より右向きの顔画像であった．これは，学習データとして用いた ChokePoint Dataset は画像内に大きく顔の向きを変化させた顔画像が少なかったことが原因であると考えられる．

ChokePoint Dataset に含まれる画像は人物がカメラの手前方向に歩いてくる画像が多いため，顔の向きの変化はあるものの，セット 2 のように，バストアップで撮影された顔の向きが大幅に変化するものに対しての学習が不十分であったと考えられる．ただし，監視カメラにバストアップが映る可能性は低い．これらのことから，今後は監視カメラ映像の特徴を考慮した，学習方法の変更も必要であると考えられる．

## 5. おわりに

本論文では，監視カメラ映像から検出した顔画像に対して，ベストショット顔画像を抽出するため，顔画像のスコアリング方式を提案し，指標を顔の向きに注目して設計を行ったことについて報告した．また，提案方式によるスコアリングモデルを生成し，精度評価とベストショット抽出評価を行い，評価に対して考察した．その結果，監視カメラ映像の特徴を整理する必要があることがわかった．今後は，顔の向きだけではなく，照明条件や，髪の毛による遮蔽，瞬きなどの監視カメラ映像を意識した指標による学習データを作成し，より高精度なスコアリングモデルを生成する予定である．

同時に，提案したスコアリングモデルを応用した監視カメラシステムにおけるアプリケーション例も検討・提案する．

## 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] 今岡仁，早坂昭裕，森下雄介，佐藤敦，広明敏彦： 顔認証技術とその応用，NEC 技報， Vol. 63, No. 3, pp. 26-30 (2010).
[5] 今岡仁，佐藤敦：判別分析と摂動空間法を用いた顔照合アルゴリズム，情報科学技術フォーラム一般講演論文集，Vol. 4, No. 3, pp. 31-32 (2005).
[6] Feng, Y., Wu, F., Shao, X., Wang, Y. and Zhou, X.: Joint 3d face reconstruction and dense alignment with position map regression network, *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 534-551 (2018).
[7] 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).
[8] Weyrauch, B., Heisele, B., Huang, J. and Blanz, V.: Component-based face recognition with 3D morphable models, *2004 Conference on Computer Vision and Pattern Recognition Workshop*, IEEE, p.85 (2004).
[9] Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C. and Zhang, Z.: Mxnet: A exible and efficient machine learning library for heterogeneous distributed systems, *arXiv preprint arXiv:1512.01274* (2015).
[10] 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).
[11] 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).
[12] 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).
[13] Chollet, F. et al.: keras (2015).
[14] 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).