

映像監視システムにおける3次元姿勢推定に基づく 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 山崎賢人² 阿倍博信¹

概要: 近年、顔の3次元再構成への要求が高まっている。しかし、その品質は顔の向き、照明環境などによって左右される。そこで、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、最適な顔画像の抽出が可能となり、様々なシーンで顔の3Dモデルの活用が期待できる。このような背景のもと、筆者らは先行研究において正面画像との類似度に基づく顔画像のスコアリング方式を提案したが、着目する顔の特徴が不明瞭であった。そこで、本論文では、顔の姿勢がスコアリングに与える影響を評価するため、顔の向きと解像度に着目し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。

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

Face Image Scoring Method for Best Shot Extraction Based on 3D Pose Estimation in Video Surveillance Systems

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 in the video surveillance fields. In recent years, 3D face reconstruction has been introduced to several fields, so the demand for it is increasing. However, qualities of reconstructed faces are affected by face orientation, facial expression, or environments. Considering this issue, we have proposed the scoring method which enables the video surveillance systems extract a best face image for the 3D face reconstruction from video frames. In our previous work, we proposed a method focused on the face similarities with frontal face images. However, it was unclear which feature this method is focusing on. Against this background, we consider the new method which evaluates scores using facial orientations and resolutions. In this paper, we describe a scoring model using machine learning and evaluations to clarify the effect of facial orientations and resolutions on scoring.

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

1. はじめに

近年、映像監視システムの利用が広がりを見せており、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[1]。

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]。

一般的に、顔が正面を向き、解像度が高い画像から3次元顔再構成するほど品質が向上するが、監視カメラの映像では正面方向の解像度が高い顔の画像が存在するとは限らない。加えて、監視映像では顔の姿勢や照明条件、表情など、様々な要因が存在するため、3次元顔再構成に適した顔画像の定義が難しい。

このような背景のもと、本研究では監視映像中の顔画像に対して、3次元再構成への適性をスコアリングするため、機械学習によるスコアリングモデルを生成することにした。筆者らは先行研究において、正面画像との類似度に基づくスコアリング方式を提案した[3][4]。具体的には、図1に示すとおり、正面を向き、解像度の高い顔画像から3次元再構成した3Dモデルを正面に向けた画像（基準画像）と、任意の顔画像から3次元再構成した3Dモデルを正面に向けた画像（比較対象画像）との類似度を教師データにし、CNNを用いてスコアリングモデルを作成した。このスコアリングモデルでは、顔が正面を向き、解像度が高い顔画像であってもベストショットとして抽出されないケースがあった。前述したとおり、照明条件、表情など、様々な特徴が学習されていると想定できるが、ブラックボックスであ

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

ることから、想定外のケースに対して改善策を練ることが難しい。さらに、学習に使用した ChokePoint Dataset[5]は人物がカメラの奥から手前に歩く映像が多く、学習データに偏りがあったと予見される。

本論文では、複合的な特徴をひとつずつひも解くために、顔の向きと解像度により着目したスコアリングモデルを生成することにした。顔の向き、解像度によるスコアリングへの影響を評価するためには、様々な顔の向き、解像度を試行し、さらにスコアの偏りが少ないデータセットが必要となる。本論文では、データセットの生成方法から、そのデータセットから生成したスコアリングモデルを評価したことについて述べる。さらに、本研究では、スコアリングモデルを監視インタフェースに活用することを目的としている。このことから、監視インタフェースの活用例として、スコアリングした映像に対して最高スコアの顔画像（ベストショット顔画像）から3次元顔再構成した3Dモデルを、監視者に表示することを想定し、3Dモデルの視認性についても考察した。

2. 関連研究

前述のとおり、顔画像のスコアリングは、顔認証技術において品質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[6]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[7]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。さらに既存の顔認証システムでは、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[8]。

スコアの算出には本研究で対象とする3次元顔再構成の代表的な技術であるDengらのDeep3DFaceReconstruction[9]を使用した。Deep3DFaceReconstructionを用いた3次元顔再構成では、入力された1枚の顔画像から顔の特徴を検出し、顔の形状を予測する。さらに顔画像から肌の色を取得し、予測した顔の形状のメッシュデータとともに保存する。図2にDeep3DFaceReconstructionによる3次元顔再構成の例を、解像度が比較的大きい画像、小さい画像の2種の正面、斜め、横、下向きの顔画像を用いて示す。この例から3次元顔再構成の結果が顔の向きや解像度に依存していることがわかる。本研究では顔の向きと解像度の影響を計ることが目的であるため、顔の向きや解像度のデータが必要となる。

Albieroらのimg2pose[10]では、入力された画像から顔の6Dofパラメータを推定することできる。この6DofパラメータからPitch, Yaw, Rollと、顔領域のバウンディングボックスを数値データとして取得可能であるため、本研究ではimg2poseを使用して顔の3次元姿勢を推定した。ただし、img2poseではカメラに対してPitch, Yaw, Rollが正面である場合は0となる（図3）。本論文で示すPitch, Yaw, Rollの値はimg2poseである。また、顔領域が入力画像に占める割合を計算し、カメラからの距離に比例するようにScaleを定義した。Scaleの値は大きくなるほど、顔画像の大きさ（解像度）は小さくなる。

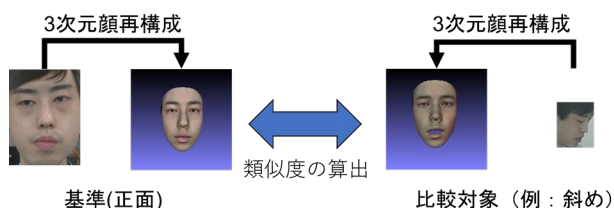


図1 評価基準の算出方式

Fig.1 Calculation method of the evaluation criteria

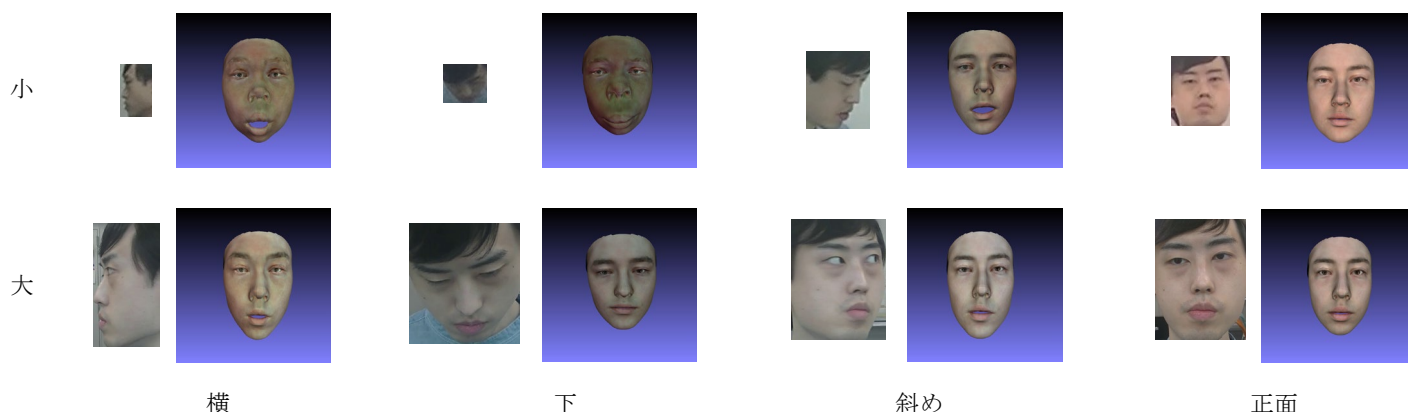


図2 Deep3DFaceReconstructionによる顔の向き、解像度ごとの3次元顔再構成の例

Fig.2 Example of 3D face reconstruction by Deep3DFaceReconstruction for each face orientation and resolution

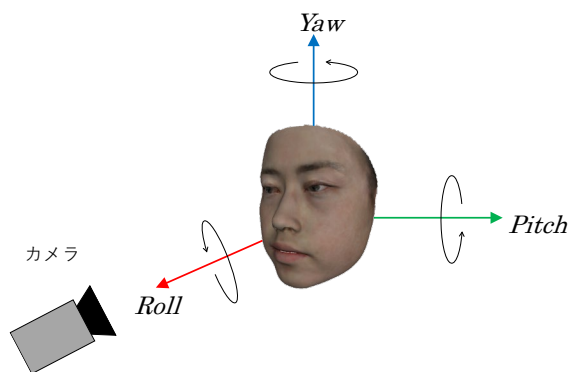


図 3 img2pose が表す Pitch, Yaw, Roll

Fig.3 Pitch, Yaw, and Roll ranges represented by img2pose

3. スコアリング方式の提案

3.1 基本方針

顔の特徴は複数存在するが、3次元顔再構成する場合、顔の向き、解像度の影響が大きいことは2章で述べた。本研究では、影響の大きな顔の向き、解像度によるベストショット顔画像のためのスコアリング方式を提案する。ただし、今後のことを考慮して、顔の向き、解像度以外の様々な指標が追加された場合でも、容易に対応可能なスコアリング方式を設計とする。

本論文で提案するスコアリング方式の方針は以下のとおりである。

- 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3Dモデルどうし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
- 顔の向き、解像度を定量的に表すため6Dofパラメータを基に抽出したPitch, Roll, Yaw, Scaleの4パラメータを使用（以後、4パラメータとは3次元姿勢推定値のPitch, Roll, Yaw, Scaleを指す）
- パラメータとスコアのデータセットから教師あり学習を用いてスコアリングモデルを生成

本研究の目的は監視カメラ映像での利用であるため、同一人物の様々な角度から撮影された顔画像が入手可能であることを前提に、基本方針に基づいてスコアリングモデルを生成する。生成したスコアリングモデルは映像中のフレームごとに、抽出した4パラメータを入力値にスコアを推定する。

3.2 データセットの作成

3.1節で述べたとおり、4パラメータとスコアのペアをデータセットとして生成するが、1章でも述べたとおり顔の向き、解像度およびスコアに偏りの少ないデータセットを作成することが望ましい。そこで、監視カメラの映像を意識して1人の人物が室内で移動する映像からデータセットを作成した。映像は室内を歩き回る様子や、椅子に着席する様子を上部方向、正面方向、下部方向から撮影したもの

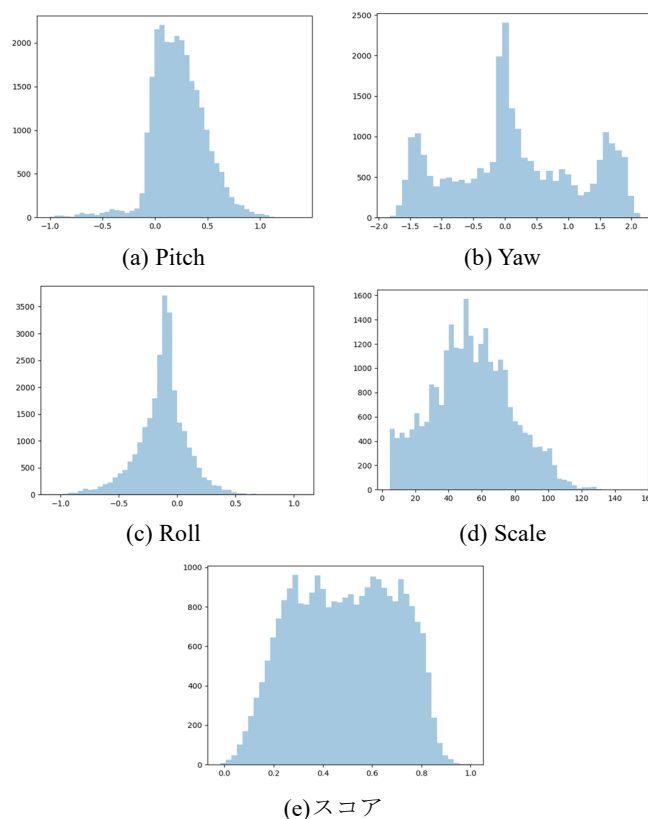


図 4 Pitch, Yaw, Roll, Scale, スコアのヒストグラム

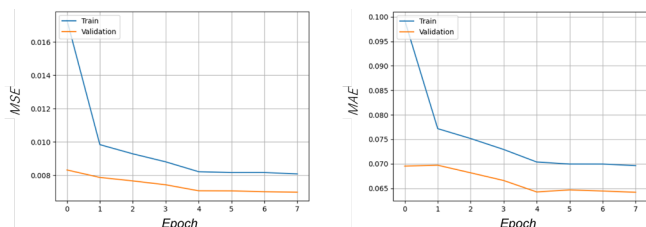
Fig.4 Histogram of Pitch, Yaw, Roll, Scale and Score

である。撮影した映像中の各フレームから推定した4パラメータにスコアを付加し、データセットを作成した。スコアは、文献[3][4]同様、正面を向いている画像であるPitch, Roll, Yaw値が最小の顔画像から生成した基準画像と、他の顔画像から生成した比較対象画像との類似度である（図1）。類似度の算出にはMXNet[11]ベースの顔分析ライブラリであるinsightface[12]を使用した。本論文におけるスコアリングモデルには、26,408組の4パラメータとスコアのセットを作成した。本来であれば、4パラメータと、それに対応するスコアの影響を見るためにスコアの分布が均一になるように、様々な4パラメータを試す必要がある。本論文では、試行の1例として図4に示すデータセット26,408組を用いる。図4に示すヒストグラムはデータセット内に含まれる顔画像の分布を示しており、縦軸は画像枚数、横軸はimg2poseの値である。

3.3 顔画像のスコアリングモデルの作成

3.1節で述べたとおり、4パラメータとスコアのデータセットからスコアリングモデルを作成する。3.2節で作成したデータセットを用いて、学習用データと検証用のテストデータを作成した。

本研究では、深層学習フレームワークとしてtensorflow[13], Keras[14]を選択し、言語はPythonを使用した。Pitch, Yaw, Roll, Scaleを説明変数、スコアを目的変数とする全結合層を3層、ドロップアウト層を3層のネッ



(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

トワークを構築し、モデルを作成した。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。また、活性化関数はReLUを選択し、学習の際は平均二乗誤差について Early Stopping を適用し、エポック数は7回であった。加えて、学習用データと検証用のテストデータに10分割のクロスバリデーションを適用した。

図5には、学習が完了した際の評価関数の値が最小であった学習用データ、テストデータの評価関数の値を示す。テストデータの最小の平均二乗誤差は約0.008であり、平均は約0.009であった。また、最小の平均絶対誤差は約0.070であり、平均は0.074であった。評価関数は共に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。

4. 評価

4.1 評価方針

評価1ではスコアリングモデルを評価し、評価2では、ベストショット顔画像から生成した3Dモデルの視認性を評価する。各評価の詳細は次のとおりである。

- 評価1：スコアモデルの評価

作成したスコアリングモデルの汎化性能を確認する。そのため撮影したテスト映像を用いて、スコアリングモデルのスコアの予測精度を評価する。さらに、3次元姿勢推定の結果がスコアに及ぼす影響を評価するため、作成したデータセットにおける4パラメータとスコアの関係や、テスト映像の各フレームにおけるスコアを比較する。

- 評価2：スコアと3Dモデルの視認性の評価

ベストショット顔画像のスコアと3Dモデルの視認性の関係性を評価する。まずデータセットのスコアと3Dモデルの関係を評価し、次に後述するテスト映像と3Dモデルの関係を評価する。

4.2 評価準備

評価用にテスト映像6本を撮影した。撮影した映像は1人の人物が室内を歩き回る様子を正面方向、上部方向から撮影した映像であり、これらの映像は3.2節で生成したデータセットとは異なる。テスト映像で撮影された人物は学

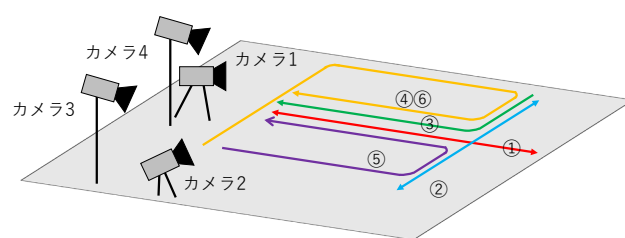


図6 カメラの設置位置と軌跡

Fig.6 Camera location and path

表1 撮影したテスト映像

Table 1 Kind of test video captured

映像	カメラ	詳細
①	1	カメラの手前と奥を往復
②	1	カメラを横切りながら往復
③	1	カメラの奥から手前方向に歩行
④	2	室内のテーブルの周りを周回
⑤	3 (俯瞰)	室内のテーブルを周回
⑥	4 (俯瞰)	室内のテーブルを周回

表2 テスト映像によるモデル評価

Table 2 Evaluation of Model from test video

映像	平均二乗誤差(MSE)	平均絶対誤差(MAE)
①	0.057	0.198
②	0.040	0.180
③	0.075	0.241
④	0.040	0.169
⑤	0.028	0.144
⑥	0.029	0.143
平均	0.036	0.156

習用データと同様の人物であり、監視カメラ映像を意識して、顔の横方向の傾きと解像度が変化するものを撮影した。映像1から4はスコアリングモデルの動作確認が目的であることから、人物の目線とほぼ同じ高さにカメラを設置して撮影した。本研究は監視カメラ映像におけるベストショット顔画像の抽出を目的としていることから、映像5,6では監視カメラの設置位置を想定した位置に設置し、俯瞰映像を撮影した(図6, 表1)。

4.3 評価1：スコアモデルの評価

スコアの予測精度の評価結果であるMSE, MAEの評価結果を表2に示す。すべての映像は様々な顔の向き、解像度が複合的に含んでいるにも関わらず、表2で示したとおりMSEとMAEがともに0に近い値のため、データセットから生成したスコアリングモデルは有用であると考えられる。

次に、3次元姿勢推定の結果がスコアに及ぼす影響を評価する。本研究における4パラメータは独立したパラメータではなく、相互に関連している。データセットにおける4パラメータと、各パラメータに対応するスコアとの関係をヒートマップで表した(図7)。ヒートマップの濃淡は各軸の値に対応したデータ数を表している。図7(a)では、Pitch

の正の値が増加するほどスコアが低下する相関関係が確認されたため、顔が下を向くほどスコアが低下するといえる。図 7(b)では、Yaw の正の値、負の値が増加するほどスコアが低下する強い曲線関係が確認されたため、顔の向きが左右に向くほどスコアが低下するといえる。図 7(c)では、Roll

の負の値が増加するほどスコアが低下する相関関係が確認されたため、顔が左に傾くほどスコアが低下する傾向があるといえる。図 7(d)では、Scale の値が増加するほどスコアが減少する弱い相関関係が確認されたため、顔領域の解像度が小さいほどスコアが低下する傾向があるといえる。

最後に、テスト映像における 4 パラメータとスコアの関係を示す。各テスト映像におけるベストショットのスコアは、映像 1 が 0.757、映像 2 が 0.630、映像 3 が 0.393、映像 4 が 0.843、映像 5 が 0.632、映像 6 が 0.691 であった。図 8 に示すのはテスト映像におけるスコアを箱ひげ図で表したものである。中でもカメラ位置が同一かつ軌跡の多くが類似している映像 1 と映像 3 を比較する。映像 1 と映像 3 は図 6 に示すとおり、カメラに対して奥から手前に近づいてくる軌跡は共通しているが、映像 3 は、近づく前に右から左への軌跡が含まれている。この右から左への軌跡は遠方かつ横向きであるため、スコアが高くないことは想定できるが、映像 1 のベストショットのスコアは、映像 3 のスコアと比較して約 2 倍である。この原因を探るべく、図 9 には映像 1 と映像 3 の各フレームにおける 4 パラメータを、図 10 には各フレームにおけるスコアを示す。

映像 1 は 200 フレーム付近まで人物がカメラに背を向けて遠ざかる。その後 200 フレーム付近で振り返って 400 フ

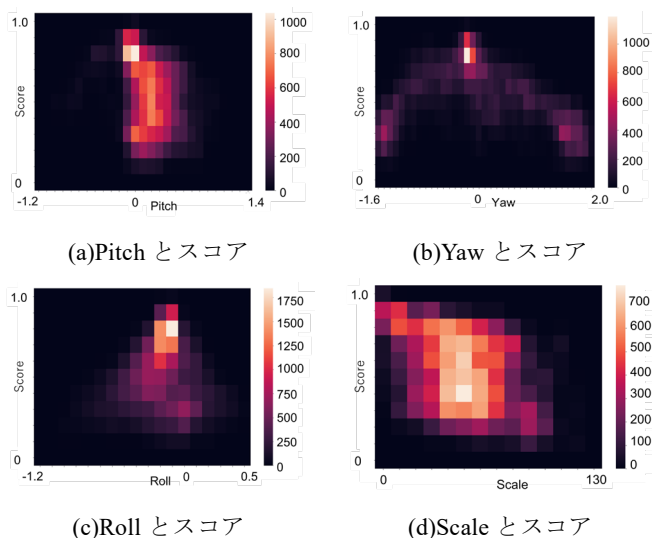


図 7 各パラメータとスコアのヒートマップ

Fig.7 Heatmap of each parameter and Score

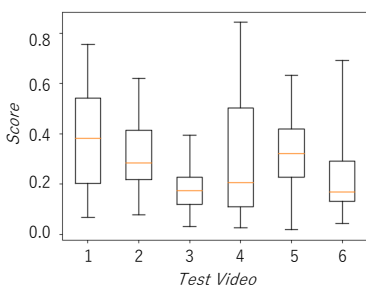
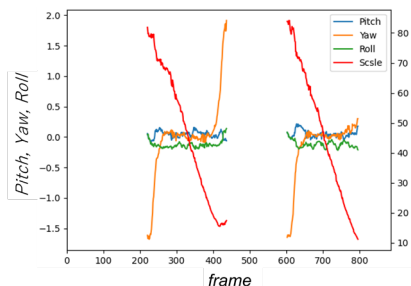
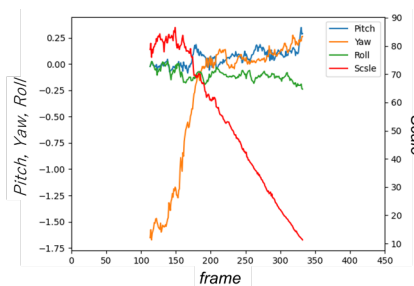


図 8 テスト映像のスコアの箱ひげ図

Fig.8 Box plot of score in test video



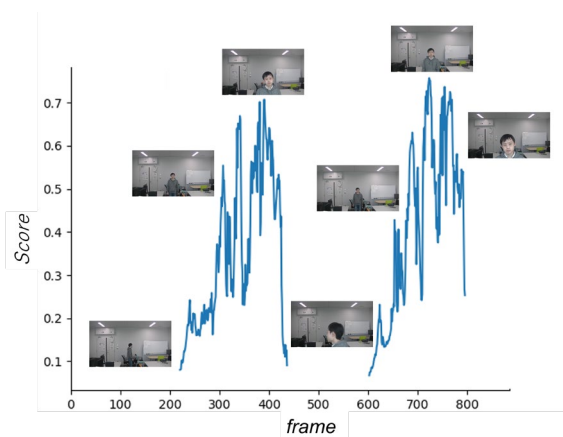
(a) 映像 1



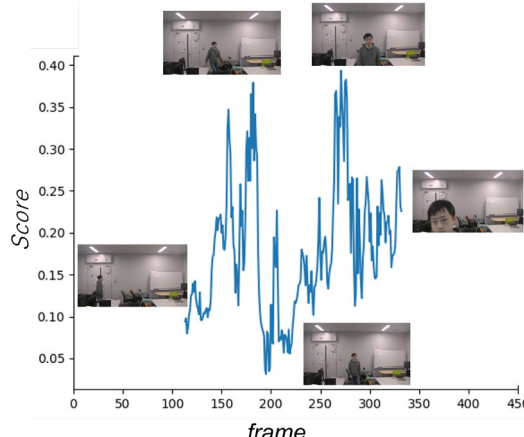
(b) 映像 3

図 9 3 次元姿勢推定値

Fig.9 3D pose estimation value



(a) 映像 1



(b) 映像 3

図 10 スコア

Fig.10 Score

フレーム付近までカメラに接近する．200 フレーム付近で振り返ることで顔領域が出現するとともに正面方向に変化するため Scale (距離) が大きく低下し，Yaw が負の方向から大きく上昇する．400 フレーム付近までは Scale の変化のみが大きい，Scale と Yaw が複合して大きく変化するフレームが存在したため，スコアの上昇幅も大きくなったと考えられる．また，Scale のみが大きく変化するフレームではスコアの振れ幅が大きい．600 フレーム以降も同様である．

映像 3 は人物がカメラを横切る方向に動作し，200 フレーム付近で右折する．その後，カメラに接近する．このときのスコアは映像 1 で同様の軌跡と相関がある．さらに，図 7(d)に示したヒートマップでは Scale とスコアに弱い相関関係が確認できた．これらのことから，Scale は他のパラメータ値の影響を受けやすいことがわかった．

さらに映像 3 の 200 フレーム付近に注目してみると，200 フレーム付近で右折する際，Scale の変化は少なく，Yaw のみが大きく上昇する．200 フレーム以降はカメラに接近するため，Scale (距離) のみが大きく低下する．図 7(c)に示したヒートマップでは Yaw とスコアに強い曲線関係が確認できており，Yaw の影響力が大きいことがわかった．ヒートマップから Pitch や Roll も同様と想定できる．これらのことから，Scale の重みが Pitch, Yaw, Roll より小さいことがわかった．

4.4 評価 2:スコアと 3D モデルの品質の評価

データセットのスコアと 3D モデルの関係を評価するため，データセットの 3D モデルを定性的に評価する．データセットにおけるスコアを 0.1 ごとに区切り，各区分からランダムに顔画像を 10 枚ずつ，計 100 枚を抽出した．抽出した顔画像から 3D モデルを生成し，目視で OK/NG を基準に 3D モデルの品質を評価した．この評価では，想定した 3D モデルの活用方法である監視者への表示に使用可能であるかを品質の定性評価の評価基準とした．監視者は監視カメラ映像で撮影された人物を特定する必要があるため，筆者の主観ではあるが，ベストショット顔画像から生成した 3D モデルが元の人物と特定可能であれば OK，不可能であれば NG と判断した．図 11 では，各区分の OK/NG の顔画像の枚数を示す．

次に，前述の判定の正当性を確認するため，テスト映像のベストショット顔画像から生成した 3D モデルを定性的に評価する．基準には同様に OK/NG を用いた．図 12 に示すのは，各映像におけるベストショット顔画像と生成した 3D モデル，ベストショットのスコア，OK/NG 判定結果である．映像 1, 3, 4, 5, 6 の 3D モデルは OK と評価し，映像 2 の 3D モデルは NG と評価した．

図 11 について，3D モデルの品質の傾向によりスコアは 3 つの範囲に分割できる．スコアが 0.3 未満の範囲ではほぼ全ての 3D モデルが NG と評価されたため，スコアがこの範囲の場合，視認性が低い可能性が高い．また，スコア

が 0.3 以上 0.7 未満の範囲では OK/NG の 3D モデルが両方存在したため，スコアがこの範囲の場合，視認性の可能なものと不可能なものが共存している．そして，スコアが 0.7 以上の範囲ではすべての 3D モデルが OK と評価されたため，スコアがこの範囲の場合，視認性が高い．これらのことから，スコアの範囲が 0.3 未満の場合を Low，0.3 以上 0.7 未満の場合を Middle，0.7 以上の場合を High と分類する．評価 2 における映像 1, 4 のベストショット顔画像は High であり，3D モデルも OK と評価されたため，スコアと視認性が一致したといえる．また，映像 2, 3, 5, 6 のベストショット顔画像は Middle に分類されるが，映像 2 が NG，映像 5, 6 は OK であった．Middle の範囲は 3 次元顔再構成に適した顔画像も適していない顔画像もベストショット顔

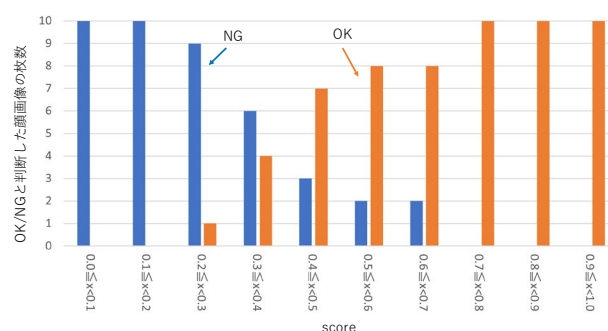


図 11 スコアと顔画像

Fig.11 Score and face image

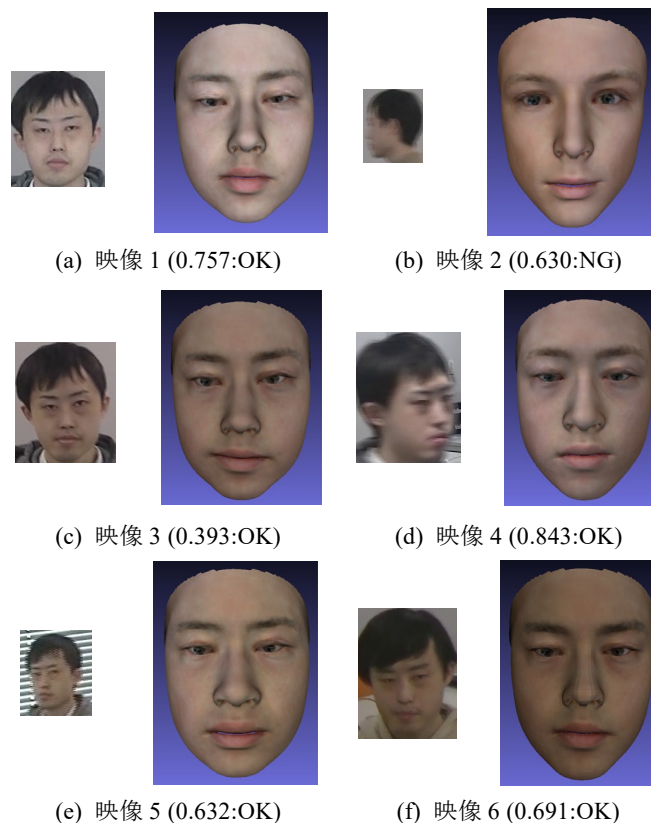


図 12 テスト映像のベストショット顔画像と 3D モデル

Fig.12 Best shot face image and 3D model of test video

画像として抽出される可能性がある。本定性評価は筆者の主観でOK/NGを分けているため、分類を高度にするには、サンプルを増やす必要がある。

ここからはテスト映像の中でも映像2の3DモデルのみがNGと評価された原因について考察する。図13には映像2のPitch, Yaw, Rollとフレームの関係、図14には映像2のスコアとフレームの関係を示す。映像2の予測されたスコアには、Rollが影響を与えたと考えられる。映像2は人物がカメラを横切りながら往復する映像であり、ほとんどが横顔のフレームであるが、300フレーム付近で振り返る。振り返るときに顔が傾き、Rollの変化量は少ないが負の方向に急速に変化している。図7(c)に示したヒートマップによるとRollが負の方向に変化するほどスコアが低下する相関関係が確認できるため、予測されたスコアはYawの影響により上昇したのち、300フレーム付近でRollの影響により下降していると考えられる。これらのことから振り返った瞬間のフレームがベストショット顔画像として抽出されると予想したが、図14に示すとおり、4パラメータの変化が少ないにもかかわらず、スコアの振れ幅が大きいことから、想定外のベストショット顔画像になった。顔の向きが急速な変化が存在したとき、変化量が小さい場合でもスコアに影響を与えることが分かったため、顔の向きが細かく変化する映像について、より詳細な評価が必要である。

5. おわりに

本論文では、顔の姿勢の変化による3次元顔再構成の品質を評価し、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、顔の向き、解像度に着目して設計し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、各パラメータの影響について評価した。その結果、Scaleの重みが小さいことがわかった。本来であれば、RollとPitchやRollとYawなどの4パラメータどうしの関係も評価する必要があるが、複雑に絡みあうパラメータどうしを評価することは難しい。さらに、画像類似度を教師データにしていることの妥当性も見直す必要がある。

今後は、4パラメータどうしの評価方法を検討するとともに、テスト映像における被写体の人数や、視認性を評価する人数を増やす。さらに、顔の姿勢だけではなく、照明条件や、髪の毛による遮蔽、目を閉じる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、スコアリングモデルの精度を向上させる予定である。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*

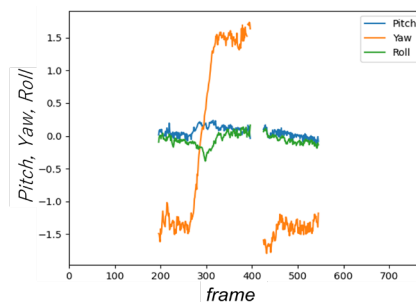


図13 映像2のPitch, Yaw, Rollとフレームの関係

Fig.13 Relation of frame and Pitch, Yaw and Roll of Video 2

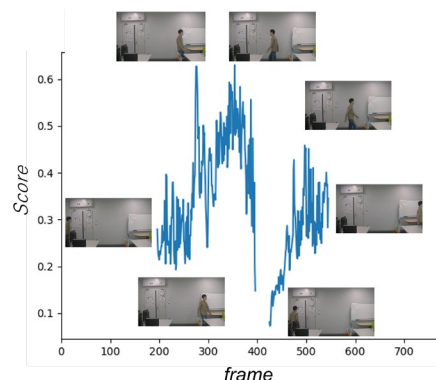


図14 映像2のスコアとフレームの関係

Fig.14 Relation of Score and frame of Video 2

Computer Vision, Vol. 64, No. 1, pp. 5-30 (2005).

- [3] 久米孝, 皆川純, 山崎賢人, 阿倍博信: 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, 第27回 情報処理学会 デジタルコンテンツクリエーション研究会(2021).
- [4] 久米孝, 皆川純, 山崎賢人, 阿倍博信: 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, 情報処理学会第83回全国大会 (2021).
- [5] 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).
- [6] 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).
- [7] 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).
- [8] 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).
- [9] 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).
- [10] Albiero, V., Chen, X., Yin, X., Pang, G., Hassner, T: img2pose: Face Alignment and Detection via 6DoF, Face Pose Estimation, *arXiv preprint arXiv:2012.07791* (2020).
- [11] 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).
- [12] 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).

- [13] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. et al.: Tensorflow: A system for large-scale machine learning, *12th {USENIX} symposium on operating systems design and implementation {OSDI} 16*, pp. 265-283 (2016).
- [14] Chollet, F. et al.: keras (2015).