

単眼カメラを用いたお辞儀の正確さ評価システム

中島 清貴¹ 長谷川 達人² 木村 春彦¹

概要: お辞儀は日本人の挨拶文化の象徴であり、円滑なコミュニケーションを行う上で重要な動作である。現在、お辞儀の習得方法として、独学で本やインターネットの知識をもとに練習することや、ビジネスマナー講座を受講することなどがある。しかし、前者は指導者がいない状況で自分のお辞儀が正しいかどうか判断することができない、後者は費用や時間がかかり手軽ではないという課題がある。本研究は、これらの課題を解決するために、単眼カメラで撮影したユーザのお辞儀動画からお辞儀の正確さを評価するシステムの開発を行い、お辞儀を独習できる環境を構築することを目的としている。その実現に向け、本稿ではお辞儀を評価するためのパラメータである、「角度」、「時間」、「姿勢」を撮影された動画像から取得し、正確さを評価するシステムの開発を行った。また、専門家によるお辞儀評価と比較することでシステムの評価を行った。

キーワード: お辞儀学習支援システム, 画像処理, HOG 特徴量, オプティカルフロー

Evaluation System of the Exactness Bow using a Monocular Camera

Abstract: The bow is a symbol of greeting culture in Japan, and it is an important action for smooth communication. People can learn a correct bow by some methods, such as to use textbooks or internet by oneself, and to attend manner lectures. However, the former method has a problem that the learner cannot evaluate the correctness of own bow without an instructor. The latter method has a problem that the manner lectures is not easy to attend. In addition, it takes cost and time. In this paper, we develop a system which evaluates the user's bow from videos taken by a monocular camera, realize an environment that people can learn correct bow by oneself easily. Towards implementation of this system, we defined "posture", "angle", and "time" as parameters for evaluation of the user's bow. We developed the system which obtains above-mentioned parameters from videos by a monocular camera. Moreover, we evaluated the accuracy of our proposed system by comparing with the expert's bow evaluation.

Keywords: bow learning system, Image processing, Histogram of Oriented Gradients, Optical flow

1. はじめに

お辞儀は伝統的な挨拶、お礼、謝罪の行為である。日本人の挨拶文化は、まずお辞儀から始まると言われ、正しいお辞儀を習得することはビジネスマナーに必要不可欠である。お辞儀には会釈、敬礼、最敬礼という3種類があり、場所、時、相手との関係性などにより使い分けなければならない。正しいお辞儀ができる人物は礼儀正しいという印象を与え、相手に好感を持たれやすい。それに対して杜撰なお辞儀しかできない人物は相手に悪い印象を与える可能性

がある。そのため正しいお辞儀をすることは円滑なコミュニケーションを取るために重要である。

お辞儀の基本動作は腰から上体を曲げるという一見単純なものだが、角度、時間、姿勢など気をつけなければならないポイントが多い。首だけを動かす、猫背になっている、適切な角度や早さで行われていないなど、自分では気づかない間違ったお辞儀をしている状況が存在する。

正しいお辞儀の習得には、独学で本やインターネットの知識を基に練習するという手段があるが、指導者がいない状況では自分のお辞儀が正確であるかどうかかわからないという課題がある。また、ビジネスマナー講座に参加するという手段もあるが、費用や時間がかかり手軽とは言えない。

現在、お辞儀の既存研究として、お辞儀の種類を識別するもの [1]、マーカーレスでお辞儀の角度を推定するもの [2]

¹ 金沢大学大学院自然科学研究科
Graduate School of Natural Science & Technology,
Kanazawa University

² 東京医療保健大学
Tokyo Healthcare University

があるが、お辞儀学習を支援するシステムの開発を行っているものはない。そのため、我々はお辞儀学習支援システムの開発を目標として研究を行っている。これは特別なセンサなどを使わずに単眼カメラだけでお辞儀学習ができるように支援するシステムである。お辞儀学習支援システムを開発することによって、個人が自宅で容易にマナーの習得ができる環境の構築が可能になる。それにより、全国の社会人や就職活動生が自主学習の範囲でマナー力を向上させることができるなど、ビジネスマナー習得の方法を多様化させることに繋がる。

お辞儀学習支援システムを構築するためには次の機能が必要と考える。

- (1) 正しいお辞儀を教示する機能。
 - (2) 学習者のお辞儀を客観的に評価し、改善点を指摘する機能。
- (1) に対しては本や Web テキストの整備で実現できるが、(2) は現状、専門家が判断せざるを得ない。そこで (2) の前段階として、お辞儀動画を自動で点数化する機能を実現する。それによって、学習者は自身のお辞儀を客観的に評価できるようになり、悪い点を改善し易くなる。

本研究ではお辞儀学習支援システムの実現に向けて、お辞儀の正確さ評価システムを開発しその評価を行う。本システムは単眼カメラで撮影された動画からお辞儀の「角度」、「時間」、「姿勢」パラメータを取得し、各パラメータを点数化することで正確さの評価を行う。

なお、本研究はお辞儀の専門家の協力の下進めている。専門家には、お辞儀についての情報提供、正しいお辞儀動画の撮影、お辞儀動画の客観的評価をしてもらった。

2. お辞儀

2.1 お辞儀の手順と種類

正しいお辞儀の手順 [3] を図 1 と以下に示す。

- (1) 気を付けをし相手の顔を見る。
- (2) 首や背筋をまっすぐに伸ばし上体を傾ける。
- (3) 上体を傾けたまま一呼吸置く。
- (4) ゆっくりと上体を起こす。

お辞儀には会釈、敬礼、最敬礼の 3 種類があり、状況に応じて使い分けなければならない [3]。3 種類のお辞儀を図 2(a)~(c) に示す。

(a) 会釈

会釈は挨拶を表現するために用いられるお辞儀で、目線を落とし頭を軽く下げ、一瞬静止した後に上体を上げることが基本動作である。上体は約 15 度で屈折する。

(b) 敬礼 (普通礼)

敬礼は取引先への訪問時や客を出迎える時に用いるお辞儀で、最も頻繁に使われる。上体の屈折角度は約 30 度である。

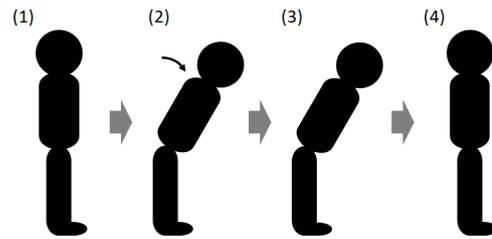


図 1 お辞儀の手順

Fig. 1 Process of the bow.

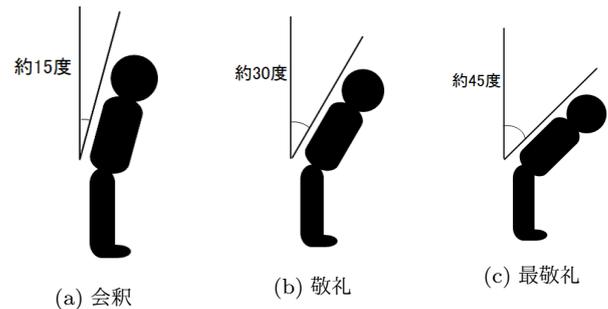


図 2 お辞儀の種類

Fig. 2 Types of the bow.

表 1 お辞儀の時間

Table 1 Times of the bow.

お辞儀の種類	屈体時間	静止時間	伸展時間
会釈	1 秒	ほんの一瞬	1 秒
敬礼	1 秒	1 秒	2 秒
最敬礼	1 秒	3 秒	4 秒

(c) 最敬礼

最敬礼は感謝を表現する時、謝罪をする時に用いるお辞儀である。上体を曲げた後にゆっくり戻ることによって丁寧さが表現できる。これを残心という。最敬礼は約 45 度で上体を屈折することで行う。

2.2 お辞儀の評価指標

お辞儀の正確さを評価するための指標として、「角度」、「時間」、「姿勢」という 3 種類のパラメータを定義した。既存研究の調査を行った結果、お辞儀を評価するシステムは現在存在していなかった。そのため、以下の 3 つのパラメータは専門家の意見を参考に我々が独自に定義したものである。それぞれのパラメータについての詳細を以下に示す。

(1) 「角度」パラメータ

気を付けをし背筋を伸ばした状態を直立状態、頭から腰までの上体を最も傾けた状態を屈体状態とする。お辞儀をした時の直立状態と屈体状態の角度差を「角度」パラメータと定義する。単位は度とする。

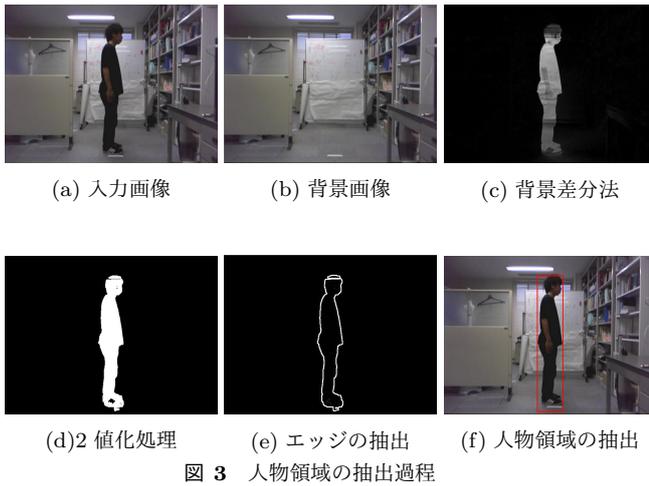


Fig. 3 Extraction process of human region.

(2) 「時間」パラメータ

直立状態から屈体状態までの時間を「屈体時間」、屈体状態でしばらく静止する時間を「静止時間」、屈体状態から直立状態までに戻る時間を「伸展時間」とし、お辞儀の「時間」パラメータと定義する。単位は秒とする。屈体時間、静止時間、伸展時間はお辞儀の種類ごとに長さが定義されている。表 1 にお辞儀の時間をまとめた。

(3) 「姿勢」パラメータ

首の曲がりや背筋の伸びなどを考慮できるように人物の上半身のシルエットを姿勢とした。お辞儀を直立時間(段階 1)、静止時間(段階 2)、直立時間(段階 3)に分け、各段階で姿勢を取得し、それを 1 つにまとめたものを「姿勢」パラメータと定義する。

その他、礼儀正しいお辞儀の要素として、目線の位置や笑顔、先に挨拶の言葉を言うことが挙げられるが、相手に良い印象を与えるための追加要素であり、正確さ評価における優先度は高くない。本研究ではお辞儀の基礎的な技術の評価することを重視するために、これらについては考慮していない。また、お辞儀は行う人の立場によって違いがある。例えば、性別によって違い、男性でも女性でも基本的に手の位置は腰の横であるが、女性は腰の前に手を重ねることが礼儀正しいとされる場合もある。職種による違いもあり、旅館の女将やキャビンアテンダントは通常より深いお辞儀をする。以上のようにお辞儀の種類は多種多様であるが、今回は一般的なユーザを想定し、2.1 節で示した最も基本的なお辞儀を対象とする。

3. 提案手法

3.1 概要

お辞儀は相手の正面に立って行うため、カメラの設置場所も被験者の正面にするべきである。しかし、お辞儀の角度、時間、姿勢を取得するためには、被験者の側面から動画を撮影する必要がある。今回は実環境におけるお辞儀評価

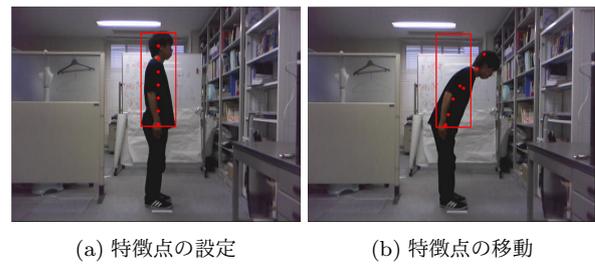


図 4 特徴点の設定と移動

Fig. 4 Setting and moving of feature points.



図 5 「角度」パラメータの取得

Fig. 5 Obtaining "angle" parameter.

を目的としておらず、個人においてお辞儀を習得する際の正確さ評価システムを構築することが目的である。そのためカメラは被験者の側面に設置し、お辞儀の各種パラメータ取得と正確さ評価を行う。

3.2 お辞儀のパラメータ取得

3.2.1 人物領域の抽出

人物領域の抽出過程を図 3(a)~(f) に示す。図 3(a)はお辞儀動画の 1 フレーム目の画像であり、これを入力画像とする。また、背景画像として図 3(b) を事前に撮影する。入力画像と背景画像から背景差分法によって図 3(c) の差分画像を取得する。その後、図 3(d) の 2 値化処理を行い、得られた画像から図 3(e) のエッジ抽出をする。画像のエッジを矩形で囲み、図 3(f) の人物領域抽出を行う。

3.2.2 「角度」パラメータの取得

まず、お辞儀の動きを動画の各フレームにおけるオプティカルフロー (3.3 節) で表現する。図 3(f) の人物領域の縦幅を半分にし、人物上半身領域を抽出する。その後、図 4(a) のように人物上半身領域内に特徴点 7 点を均一に設定し、特徴点ごとにオプティカルフローを算出する。オプティカルフローの算出範囲を画像全体から限られた特徴点に限定することで、高速化とノイズ除去を実現できる。図 4(b) はお辞儀動作中の特徴点の動きを示したものである。オプティカルフローを用いたトレース処理により、特徴点の動きを求める。次に上半身の特徴点 7 点の直線近似を最小二乗法によって行い、各フレームにおける人物の上半身の傾きと直立状態の角度差を求める。角度差が最も大きくなったフレームを屈体状態と決定し、その時の角度差を「角度」パラメータとして取得する。図 5 は屈体状態における上半身の傾きを直線で可視化したものである。



(a) 入力画像 (b) HOG 特徴量可視化画像

図 6 HOG 特徴量の抽出

Fig. 6 HOG extraction.

3.2.3 「時間」パラメータの取得

お辞儀が始まってから終わるまでの時間を屈体時間、静止時間、伸展時間に分け、「時間」パラメータとして取得する。それぞれの時間の区別にはお辞儀動画のフレーム間の角度変化を用いる。フレームごとに現在のフレームと1つ前のフレームの角度差を求め、3フレーム分の角度差を合計し角度変化値とする。一定の閾値を決め、角度変化値が閾値を上回った時、お辞儀の始まりと決定する。そのフレームから角度変化値が閾値を下回るまでを屈体時間、その後にもう一度閾値を上回るまでを静止時間、上回ってから角度の変化がなくなるまでを伸展時間とする。

3.2.4 「姿勢」パラメータの取得

人物の姿勢を推定する手法には、複数のカメラを用いて身体の形状を推定するもの [4]、TOF カメラから得られる距離情報を用いて回帰分析により姿勢推定を行っているもの [5] がある。しかし、これらの手法では動画撮影に複数のカメラや特殊なカメラが必要となる。本研究の目的は個人が自宅でお辞儀習得できる環境を構築することであり、一般的な家庭にある単眼カメラのみを用いるべきである。そこで本研究では、大西らの研究 [6] を参考に、物体の形状を表すことが可能な HOG 特徴量 (3.4 節) を用いて、人物姿勢推定を行った。なお、大西らの研究では 3 次元姿勢の推定をしているが、お辞儀の正確さを評価するためには人物の背の伸び具合などを考慮する必要があるため、本研究では HOG 特徴量の抽出までのみを行った。

図 6(a) は入力画像、図 6(b) は入力画像から抽出した HOG 特徴量を可視化した画像である。3.2.1 節で示したエッジ抽出までの処理過程を同様に行い、入力画像からエッジ画像を作成し、HOG 特徴量の抽出をする。さらに入力画像の人物上半身領域を抽出し、その領域のみの HOG 特徴量を使用することにする。体型差や身長差を考慮するために、人物上半身領域のサイズは一定の大きさに正規化する。正規化には全お辞儀動画の人物上半身領域の平均サイズを用いる。以上の HOG 特徴量抽出をお辞儀の段階 1～段階 3 において行い、得られた値を 1 つのデータとしてまとめ、「姿勢」パラメータとする。

3.3 オプティカルフロー

お辞儀の動きを抽出するために、オプティカルフロー [7] を求める。オプティカルフローは動物体解析法の 1 つであり、画像中の輝度情報から動きを分析し、動きベクトルにより物体の運動を表す。本研究ではオプティカルフローを求める手法としてブロックマッチング法 [8] を用いる。

ブロックマッチング法では画像を一定の大きさの小領域 (ブロック) に分割し、それぞれのブロックの動きベクトルを求める。これは各ブロックが前のフレームのどこに対応するかを探し、対応するブロックの位置の差を動きベクトルとするものである。 t フレーム目の $N \times N$ の大きさのブロックの動きベクトルを求めるものとする。ブロックの左上端の画素の位置を (i_0, j_0) として、このブロックの画像と直前のフレームの対応する位置から (p, q) だけずらした位置のブロックの画像との差を比較する。本研究ではマッチングの評価に画素値の差の絶対値の和 SAD (Sum of Absolute Difference) を用いた。以下に SAD の式を示す。

$$SAD(p, q) = \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} |f_t(i_0 + i, j_0 + j) - f_{t-1}(i_0 + i + p, j_0 + j + q)| \quad (1)$$

ブロックの位置 (p, q) を変えて $SAD(p, q)$ を計算し、それが最小となる (p_m, q_m) を動き検出をしたい物体の画像が移動したブロックの位置とする。求められた (p_m, q_m) との差を求めることで動きベクトルを導く。

3.4 HOG 特徴量

お辞儀の姿勢を取得するために、HOG 特徴量を求める。HOG (Histogram of Oriented Gradients) [9] は画像の局所領域から輝度勾配と勾配強度を取り出す特徴量であり、物体の大まかな形状を表現することができる。以下に HOG 特徴量の算出手順について述べる。

まず、画像の位置 (x, y) における輝度値 $I(x, y)$ として輝度勾配を次式で求める。

$$\begin{cases} f_x(x, y) = I(x + 1, y) - I(x - 1, y) \\ f_y(x, y) = I(x, y + 1) - I(x, y - 1) \end{cases} \quad (2)$$

ここで f_x, f_y はそれぞれ画像の横方向、縦方向の輝度勾配を表す。得られた輝度勾配から、勾配強度 $m(x, y)$ とその勾配方向を次式によって算出する。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (3)$$

$$\theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (4)$$

次に算出された輝度勾配画像を、 $N_c \times N_c$ 画素からなるセルという小領域に分割する。それぞれのセル領域において輝度勾配方向の重み付けを行うことにより輝度の勾配方向ヒストグラムを作成する。さらに各セルで作成した輝度

勾配ヒストグラムを、セルより大きな領域 $N_b \times N_b$ セルを 1 ブロックとして正規化を行う。正規化前の特徴量を f とし、あるブロックの特徴ベクトルを \mathbf{v} としたとき、次式により正規化を行う。

$$v = \frac{f}{\sqrt{\|\mathbf{v}\|_2^2 + \epsilon^2}} \quad (\epsilon = 1) \quad (5)$$

v は正規化後の特徴量である。このように得られた特徴ベクトルは、照明や影の影響を受けにくく、局所的な幾何学変化に頑健となる。本研究では最後に、正規化された HOG 特徴量のブロックごとの勾配強度最大値を 1 列のベクトルにまとめ、特徴量としている。

3.5 お辞儀の正確さ評価

「角度」、「時間」、「姿勢」パラメータを取得し、0~9 点の 10 段階で点数化することで正確さ評価を行う。点数は正しいお辞儀であるほど高くなる。各パラメータを点数化する手法を以下で説明する。

3.5.1 「角度」、「時間」パラメータの点数化

2.1 節と表 1 で示したように角度と時間はお辞儀の種類ごとに定義されている。そのため、「角度」、「時間」パラメータの点数化はその定義と比較することで行う。例外として、会釈の角度だけは専門家のお辞儀と比較する。会釈の角度の定義は 15 度であるが、撮影した専門家のお辞儀動画では平均角度が 24.5 度となり、大きな違いが生じた。このことから、実際の会釈は定義値よりも角度が大きくなる傾向があることがわかった。そのため、会釈のみは専門家のお辞儀の平均を正解データとし点数化を行う。

式 (6) によって点数を算出する。Error は定義値との誤差、Max は最大誤差を意味する。「角度」パラメータでは Max を 15 度、「時間」パラメータでは 1 秒とし、Error が Max を越えた時は点数を 0 点とした。また、「時間」パラメータは屈体時間、静止時間、伸展時間に分かれているため、それぞれで点数化を行い、平均点を最終的な点数とした。

$$Score = \begin{cases} 0 & (Error > Max) \\ (1 - \frac{Error}{Max}) \times 9 & (Error \leq Max) \end{cases} \quad (6)$$

3.5.2 「姿勢」パラメータの点数化

「姿勢」パラメータは専門家のお辞儀と比較することで点数化を行う。専門家のお辞儀の「姿勢」パラメータの平均を算出し正解データとする。点数化したいお辞儀の「姿勢」パラメータをテストデータとして、正解データとのユークリッド距離を求める。複数のテストデータでユークリッド距離を求め、その中で距離が最も大きいものが 0 点、小さいものが 9 点となるように正規化し点数化を行う。以上の処理をお辞儀の段階 1~段階 3 で行い、それぞれの点数の平均を「姿勢」パラメータの点数とした。

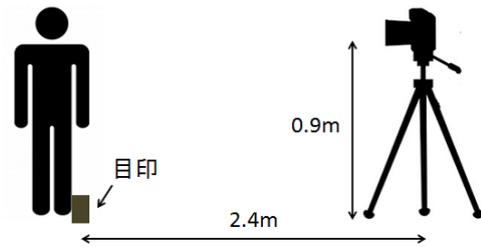


図 7 実験環境

Fig. 7 Experiment environment.

表 2 お辞儀の種類とシチュエーション

Table 2 Types of the bow and situation.

お辞儀の種類	角度	シチュエーション
会釈	15 度	知人とすれ違う時
敬礼	30 度	来客時
最敬礼	45 度	謝罪時

4. 評価実験と考察

4.1 実験概要

お辞儀動画から「角度」、「時間」、「姿勢」パラメータを取得し、点数化を行った。また、専門家に被験者のお辞儀動画を見せ、3 つのパラメータについてシステムと同様に 10 段階で点数をつけてもらった。この客観的評価の結果と本システムによる点数化の結果を比較することでシステムの評価を行った。

実験環境を図 7 に示す。被験者の側面に Web カメラを設置し、お辞儀動作を撮影する。被験者と Web カメラの距離は 2.4m、Web カメラの高さは 0.9m とした。Web カメラの解像度は 640 × 480 であり、フレームレートを 30fps に設定した。

4.2 データセット

被験者はお辞儀についての指導を受けた経験がない 20 代男性 10 名で、4.1 節の実験環境でお辞儀動画の撮影を行った。データセット中のお辞儀動画に点数の偏りがないようにするため以下の 3 段階で撮影した。

- (1) 3 種類のお辞儀の角度と使用するシチュエーションのみを伝え、動画を撮影 (3~6 点想定)。
- (2) お辞儀の指導を行った後に動画を撮影 (7~9 点想定)。
- (3) 被験者が意図的に行った悪いお辞儀を撮影 (0~2 点想定)。

(1) では表 2 の内容のみを被験者に伝えた。(2) では専門家の正しいお辞儀動画を見せた後、3 種類のお辞儀における時間を説明し、被験者が納得するまで練習をしてもらった。練習の際、自己の角度がわかるように、壁に 15 度、30 度、45 度を直線で描いたシートを設置した。姿勢については撮影中に適宜指導を行った。(3) では (2) で指導したお辞儀

表 3 各パラメータの平均絶対誤差

Table 3 Mean absolute error of each parameter.

	角度	時間	姿勢
平均絶対誤差	2.23 点	2.21 点	2.24 点

表 4 一定範囲内の点数化精度

Table 4 Accuracy of scoring within a ceration range.

	角度	時間	姿勢
± 0 点以内	15.6 %	14.8 %	10.3 %
± 1 点以内	44.8 %	41.5 %	30.4 %
± 2 点以内	65.9 %	63.7 %	60.0 %
± 3 点以内	77.4 %	78.1 %	83.0 %
± 4 点以内	85.9 %	88.5 %	93.7 %

に反する悪いお辞儀を被験者の任意で行ってもらった。以上の3段階で3種類のお辞儀を3回ずつ撮影し、合計270動画を取得した。

4.3 実験結果と考察

本システムによる点数化結果と専門家による客観的評価結果の誤差の絶対値をお辞儀動画ごとに求め、その平均を平均絶対誤差とし、各パラメータで算出した。その結果を表3に示す。さらに、平均絶対誤差が±0~4点以内で評価されている動画を正しく評価できているとして、推定精度を算出した。その結果を表4に示す。

表3より、どのパラメータにおいても平均絶対誤差が約2点であることがわかる。点数化は10段階で行っているため、±2点の誤差は比較的大きい値だと言える。また、表4を見ても、±0点以内で点数化できているものは約10%、±2点以内でも約60%程度となっており、精度として十分ではない。各パラメータで誤差が大きくなったお辞儀動画を調べ、原因の考察を行った。その結果を以下にまとめる。

4.3.1 「角度」パラメータの考察

被験者ごとに「角度」パラメータの平均絶対誤差を算出し、値を比較した。それを図8のように客観的評価を行った順番に並べた結果、序盤に行った2人の誤差が大きくなっていることがわかった。専門家に客観的評価の難易度を伺ったところ、10段階の評価は難しく、その時々によって細かい評価が変わるかもしれないと回答していた。特にお辞儀の角度は目視で評価する場合、細かい角度まで評価することは難しい。これらのことから、専門家がこの評価方法に慣れるまでに時間がかかり、序盤に評価を行った2人の誤差が大きくなってしまったのではないかと考えられる。今回は1人の専門家が客観的評価を行ったが、今後は複数の専門家が任意の被験者の順番で評価を行うことで改善が見込める。

4.3.2 「時間」パラメータの考察

お辞儀の種類ごとに「時間」パラメータの平均絶対誤差

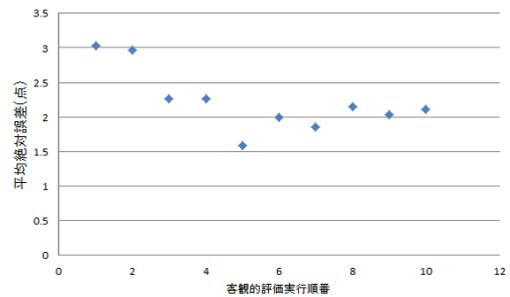


図 8 角度の平均絶対誤差

Fig. 8 Mean absolute error of angle.

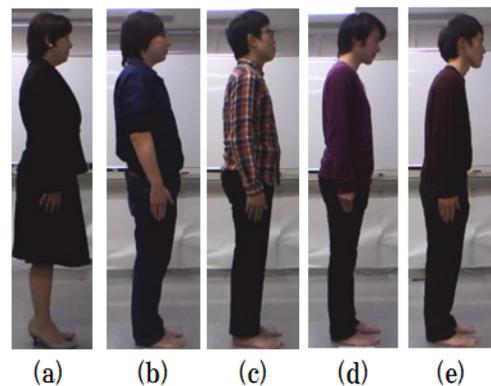


図 9 姿勢の比較

Fig. 9 Comparison of posture

を算出し、値を比較した。その結果、敬礼の誤差が1.69点と小さく、会釈、最敬礼が2.20点、2.74点と大きくなっていることがわかった。「時間」パラメータの点数化は式(6)で行っており、どのお辞儀の種類でもMaxを1秒としている。お辞儀の時間は表1で示したように、お辞儀の種類によって長さが違う。会釈の静止時間が最も短く、最敬礼の伸展時間が最も長い。そのため、1秒というMaxの値は会釈の静止時間に対して大きく、最敬礼の伸展時間に対しては小さい値となってしまう、適切な点数化をできていない可能性がある。以上のことから、会釈と最敬礼の平均絶対誤差が大きくなったと考えられる。お辞儀の種類ごとにMaxを変化させることで適切な点数化が可能になり、誤差を小さくできるのではないかと考える。

4.3.3 「姿勢」パラメータの考察

被験者ごとに「姿勢」パラメータの平均絶対誤差を算出し、値を比較したところ、被験者の体型によって誤差に違いが生じることがわかった。図9(a)は専門家、(b)~(e)は被験者の画像である。(b),(c)は(a)と比較的に体型や被服の形が似ており、平均絶対誤差は小さかった。それに対し(d),(e)は(a)と体型差や被服の形に違いがあり、平均絶対誤差は大きかった。このことから人物の体型や被服によって点数化の精度が変化するということがわかる。本システムは人物の体型差を考慮するために人物上半身領域の抽出時に領域のサイズの正規化を行っているが、それではまだ

表 5 客観的評価における各パラメータの相関

Table 5 Correlation of parameters in objective assessment.

	角度	時間	姿勢
角度	1.00		
時間	0.81	1.00	
姿勢	0.79	0.81	1.00

表 6 本システムにおける各パラメータの相関

Table 6 Correlation of parameters in this system.

	角度	時間	姿勢
角度	1.00		
時間	0.25	1.00	
姿勢	0.32	0.22	1.00

不十分で、今後はさらに体型差や被服の違いを考慮できる手法が必要だと考えられる。

4.3.4 各パラメータの相関

客観的評価と本システムによって得られた各パラメータ間の相関を調べた。その結果を表 5 と表 6 に示す。表 5 より客観的評価で得られたパラメータ間にはそれぞれ強い相関があることがわかる。相関が強いということは、他のパラメータの影響が大きいということである。例えば、本来「角度」パラメータの点数が高いときでも、「姿勢」パラメータが低ければ、「角度」パラメータも影響を受けて低くなってしまい、正しく点数化ができない可能性がある。これに対して本システムでは、表 6 より、得られたパラメータ間の相関は弱いことがわかる。つまり各パラメータは独立して点数化されているということである。以上のように、専門家の客観的評価と本システムでは、得られた各パラメータ間の相関が大きく異なり、点数化の手法に違いが生じている。これが平均絶対誤差が大きくなった最も大きな原因であると考えられる。今回の客観的評価の手法では、各パラメータを独立して点数化をすることは難しい。そのため、今後は本システムに適した評価方法を考えることが必要である。

5. まとめ

本研究では、単眼カメラのみを用いてお辞儀の正確さを評価するシステムの開発を行った。オプティカルフローを用いることで「角度」、「時間」パラメータを、HOG 特徴量を用いることで「姿勢」パラメータを取得し、点数化を行う手法を提案した。また、専門家によるお辞儀の客観的評価とシステムによる点数化の結果を比較することで本システムの評価も行った。その結果、それぞれのパラメータで平均絶対誤差が約 2 点、± 2 点以内で点数化できているお辞儀動画は約 60 % 程度であり、システムの精度としては不十分だった。結果の調査を行ったところ、専門家の客観的評価の点数化結果と本システムの点数化結果で各パラメータ間の相関に大きな差があった。これがシステムの精

度を下げた最も大きな原因ではないかと考えられる。今後はパラメータ取得と点数化の手法の改善、システムの評価方法の見直しを行い、より精度の高いお辞儀の正確さ評価システムの開発を目指す。

参考文献

- [1] 新田雅道, 梶原祐輔, 木村春彦: 機械学習を用いた映像内に現れるお辞儀動作の認識システム, 日本設備管理学会誌, 27(3), pp.109-115(2015)
- [2] 梶原祐輔, 木村春彦: 人の輪郭と局所特徴量を用いた立礼教育システムにおける屈体角度の推定, 電気学会論文誌 C, 135(12), pp.1555-1564(2015)
- [3] ANA ラーニング: [図解] これで仕事がうまくいく! ビジネスマナーの基本ルール, 成美堂出版 (2011)
- [4] 中澤篤志, 平亮介, 竹村治雄: Boosting を用いたボリュウムデータからの人体姿勢推定, 情報処理学会研究報告, CVIM, [コンピュータビジョンとイメージメディア研究会研究報告], Vol. 164, pp.143-148(2008)
- [5] 縣禎輝, 藤吉弘巨: 距離情報を用いた回帰分析による人物姿勢推定, 情報処理学会研究報告, CVIM, [コンピュータビジョンとイメージメディア研究会研究報告], Vol. 177, No16, pp.1-8(2011)
- [6] 大西克則, 滝口哲也, 有木康雄: HOG 特徴に基づく単眼画像からの人体 3 次元姿勢推定, MIRU, 画像の認識・理解シンポジウム (2008)
- [7] J.L.Barron, D.J.Fleet, S.S.Beauchemin: Performance of Optical Flow Techniques, International Journal of Computer Vision, 12-1, pp.43-77(1994)
- [8] D.I.Barnea, H.F.Silverman: "A Class of Algorithm for Fast Digital Image Registration", IEEE Trans.on Computers, 21,pp.179-186(1972)
- [9] N.Dalal and B.Triggs: Histograms of Oriented Gradients for Human Detection, IEEE Computer Vision and Pattern Recognition, pp.886-893(2005)