

顔部品の自動検出結果による擬人化呈示に関する研究

山下裕史^{1,a)} 鹿嶋雅之^{1,b)} 福元伸也^{1,c)} 佐藤公則^{1,d)} 渡邊睦^{1,e)}

概要: 現在、ゲームや医療などで3次元アバターを用いた対話型ゲームや医療支援システムなどが使われている。しかし、アバターの多くは人間であり、用途に合わせた多様なキャラクターが必要となる。従来の研究では、動物の顔を動かす研究であった。本研究では、Kinect v2で取得した距離データとカラー画像から顔部品を自動検出し、顔部品を非生物の物に呈示するシステムを提案する。検出した顔部品の座標を目視で確認した座標と比較し、誤差10ピクセル以内になった。また、西瓜や林檎、大根などの様々な物に対して様々なユーザーの顔部品呈示が成功した。これらの実験により、擬人化呈示の基本的な有用性を確認することができた。

キーワード: 画像認識・理解, 3次元形状, アニメーション

Research on Personification Presentation by Automatic Facial Parts Detection

HIROFUMI YAMASHITA^{1,a)} MASAYUKI KASHIMA^{1,b)}
SHINYA FUKUMOTO^{1,c)} KIMINORI SATO^{1,d)} MUTSUMI WATANABE^{1,e)}

Abstract: Currently, three-dimensional avatar is used in games and medical application. However, most of avatars are in limited human, they are required various avatars. Previous research was proposed the method which move the face of animal. This research proposes the new method which face parts is detected face parts automatically from distance and color, and synthesize them to abiotic. Accuracy was within 10 pixels by comparing position of the detected parts and the confirmed parts visually. Successfully, various objects such as watermelon, apples, radish was presented parts of various users. Those experimental results have shown the basic effectiveness of the propose method.

Keywords: Image Recognition, Three-Dimensional Shape, Animation

1. はじめに

現在、ゲームや映画、医療など様々な場面で3次元のキャラクターやアバターが使用されている。そして、コンピュータを介したユーザーとアバターとのコミュニケーションを利用した研究やシステムが開発されている。具体的には、アバターを介した銀行窓口システム^[1]やアバターとの対話型ゲーム、医療支援システムが挙げられる。また、自分の顔をデフォルメして画面上に表示するようなアプリケーションも存在する。しかし、これらの研究やシステムでは、人間の顔のアバターであることが多く、このアバターがサービスの案内や提供をしている。アバターには様々な用途が求められており、用途に合わせて、3次元アバターの多彩な動きが必要となる。また、人を楽しませるようなゲームやアプリケーションでは、人間だけでなく、動物や物など多様なアバターを使うことで、よりエンターテインメント性のあるゲームやアプリケーションを作ることができる。

そこで、本研究では非生物を擬人化させたアバターを様々なゲームやアプリケーション、システムに応用することを目的とする。そして、Kinect v2(Kinect for Windows v2)で撮影した人の顔部品を自動検出し、顔部品を非生物に呈示するシステムを提案する。

2. 関連研究

2.1 顔部品検出

顔部品検出は大きく3種類あり、距離データのみを利用した検出、カラー画像のみを利用した検出、距離データとカラー画像の両方を利用した検出に大別される。距離データのみを利用した顔部品検出では、顔部品の3次元の特徴を利用して検出している^[2]。しかし、顔の向きに対応できていない。カラー画像のみを使用した検出では、顔部品のYIQ表色系の特徴を利用して顔部品を検出している手法がある^[3]。しかし、耳や頬など表色系の特徴が似ている箇所を誤検出してしまふ。カラー画像と距離データの両方を用いた研究では、顔領域から顔の特徴量を抽出し、その特徴量を機械学習させることで、個人の判定を行い、セキュリティに応用する研究がある^[4]。この研究では、顔領域の画像からWebAPIにより顔部品の座標を最大50点抽出し、その50点の座標から個人の特徴量を求める。この特徴量を機械学習させ、個人の判定を行う。しかし、顔の位置のズレ

1 鹿児島大学大学院理工学研究科
Graduate School of Science and Engineering, Kagoshima University
a) k8650193@kadai.jp
b) kashima@ibe.kagoshima-u.ac.jp
c) fukumoto@ibe.kagoshima-u.ac.jp
d) kimi@ibe.kagoshima-u.ac.jp
e) mutty@ibe.kagoshima-u.ac.jp

やカメラとの距離がずれると特徴量の差異が大きくなってしまふ。これらの顔部品検出においてカラー画像のみの検出には主に RGB カメラが使われており、距離データを利用する検出には Kinect が使用されている。どちらのカメラも市場で容易に入手可能である。

本研究では、カラー画像と距離データの両方が取得できる Kinect を用いて、色情報と距離データのそれぞれのメリットを生かし、それぞれの情報から分かる顔部品の特徴を利用した顔部品検出を行うことで、従来研究に対する精度向上を図る。なお、検出した鼻を利用して顔の向きを推定する。

2.2 顔部品呈示

アバターの作成に関する研究では、顔のモデリングを行い、アバターに呈示する研究や顔部品を与えてその顔部品を元にアバターの顔を動かす研究などがある。顔のモデリングを行ってアバターに呈示する手法では、人の顔を撮影し、その顔のモデリングを行い、デフォルメしたアバターを作成する研究がある^[5]。この研究では、まず初めに顔特徴点抽出を行い、特徴量を算出する。その特徴量から顔の 3 次元モデリングを行う。モデリングされた顔に対して、ユーザーは顔部品をデフォルメできる。

顔部品を与えてそれを元にアバターの顔を動かす研究では、動物やモンスターに人の顔部品を合成させ、人の顔のように動物やモンスターの顔を動かす研究がある^[6]。この研究では、人の顔の表情を動物やモンスターの顔に表情筋などを考慮して合成する。しかし、これらの手法では動物や人を対象に顔部品を呈示しているが、非生物には呈示していない。また、この研究では、呈示対象である動物やモンスターに顔部品の座標が与えられている。

本研究では、検出した顔部品を非生物である野菜や果物などに呈示する。非生物の物には目や鼻、口といった顔部品が存在しないため、呈示対象となる非生物の物に最適な顔部品の位置を推定し、顔部品を呈示する。

3. 提案手法

3.1 研究の概要

本研究では、人の顔部品の 3 次元的特徴から顔部品を検出し、非生物を対象に顔部品を呈示するシステムを提案する。図 1 にシステム全体の流れを示す。

まず初めに、人の顔を Kinect v2 で撮影し、取得した距離データとカラー画像（図 2）から顔部品を検出する。顔部品検出は、初めに距離データを用いて鼻検出を行う。目検出と口検出において、検出した鼻の位置による探索制限をかけ、目はカラー画像のみを用い、口はカラー画像と距離データを用いて検出する。

次に顔部品呈示を行う。顔部品呈示は、非生物のデータを取得してサイズを解析し、解析結果を元に顔部品を呈示する。なお、顔部品検出と顔部品呈示は実時間で行う。図

3 は、取得した距離値を 256 で割り、余りを数値化することで距離値の大きさを輝度値 0 から 255 に制限し、画像化した距離画像である。検出した顔部品を非生物に合成させ、呈示する。

なお、本研究の初期条件として図 3 の紫色の枠のように顔の領域を制限し、その領域に本システムのユーザーの顔が入るようにして撮影する。また、人物だけを切り出すために、距離値に閾値を設け、閾値以上の距離値となった場合、背景とみなし、そのピクセルの輝度値を 0 にする。図 3 のように背景を黒にしている。

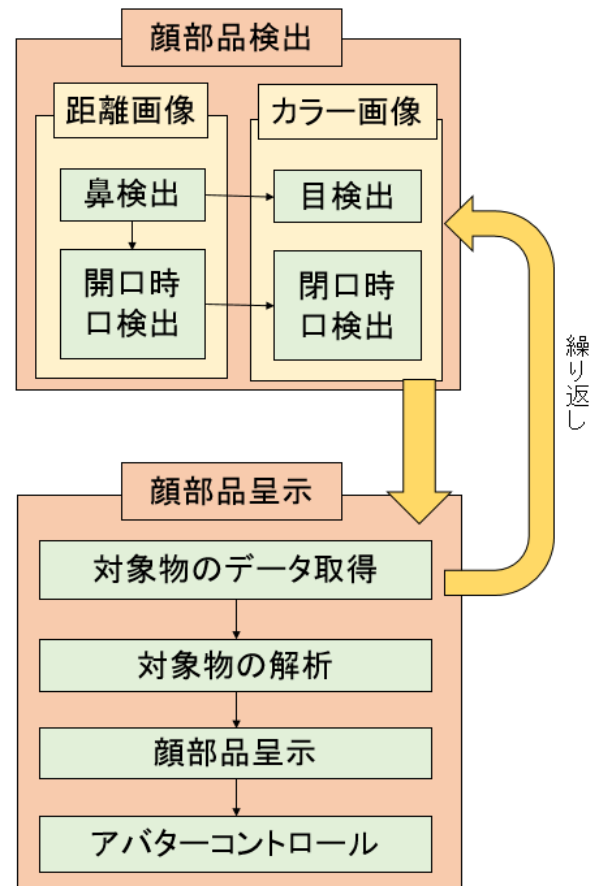


図 1 全体の流れ



図 2 カラー画像

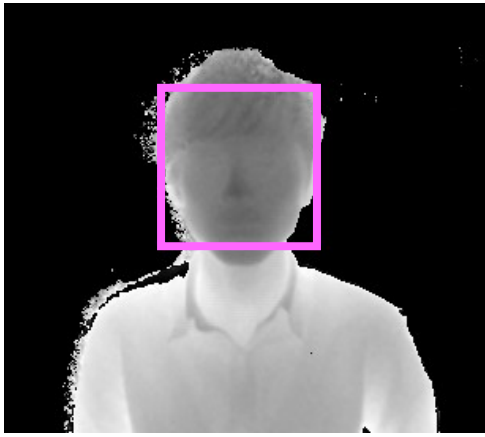


図3 距離画像

3.1.1 Kinect for Windows v2

Kinect v2はRGBカメラと赤外線カメラと赤外線を投光するプロジェクタを搭載している。Kinect v2は解像度1920×1080のカラー画像と解像度640×480の距離画像を取得できる。Kinect v2はTOF (Time of Flight)方式を採用している。この方式は、プロジェクタから投光した赤外線が反射して戻ってくる時間を計算し、距離情報を取得する。Kinect v1に比べて、RGBカメラの解像度が640×480から1980×1080に上がり、人物の取得、距離情報の取得できる範囲も拡大した。また、その他の性能も向上している。

3.2 顔部品検出

3.2.1 鼻検出

初めに、鼻を検出する。鼻頭は、顔の領域においてKinect v2からの距離が最短距離となる。距離画像では輝度値が小さいほどKinect v2までの距離が近くなるため、図4のAの鼻の領域は距離が最も近いことが分かる。この特徴を利用して、距離画像の顔領域から最短距離となる箇所を探索し、鼻を検出する。その後、鼻を探索する領域を検出した鼻の位置の周囲10ピクセルに制限する。鼻の位置は検出時に毎回更新する。

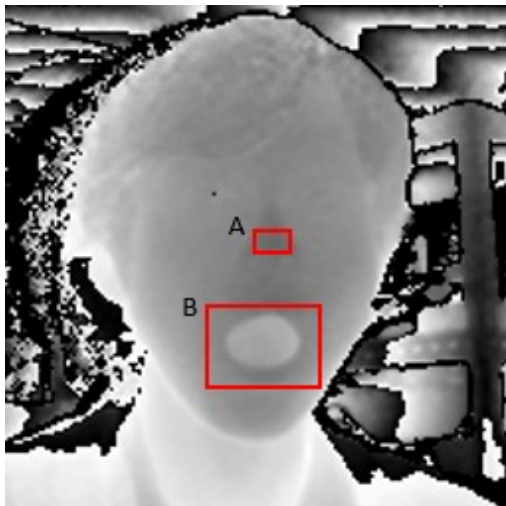


図4 顔の距離画像

3.2.2 テンプレートマッチング

3.2.3節、3.2.4節での目検出と口検出では、テンプレ

トマッチングを使用する。今回用いた比較方法は、NCC (Normalized Cross-Correlation) である。テンプレートマッチングは、テンプレート画像と画像データの類似度を計算し、類似度が高い場合、マッチングする手法である。(1)の式を用いて、類似度を計算している。Rが類似度、Tがテンプレート画像の画素値、Iが画像データの画素値、x,yは画像のピクセルの位置を表している。

$$R(x, y) = \frac{\sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2}} \quad (1)$$

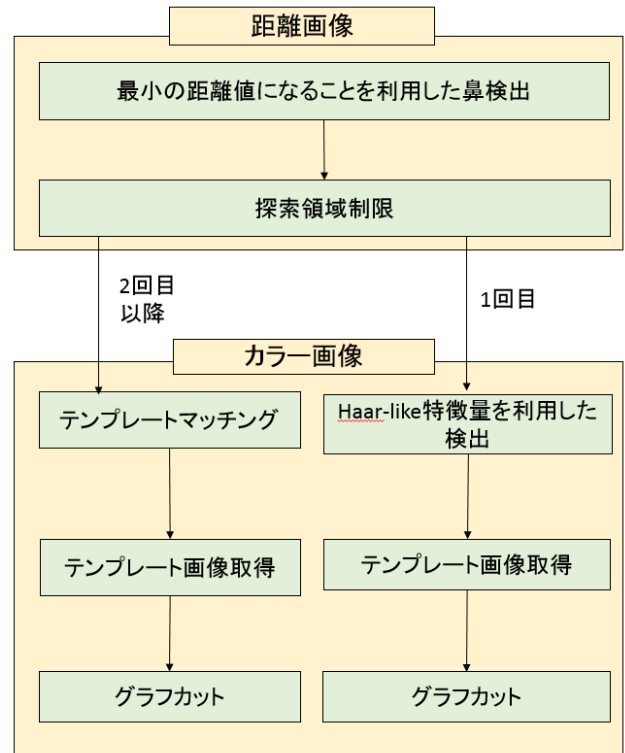


図5 目検出の流れ

3.2.3 目検出

次は目を検出する。図5に目検出の全体の流れを示す。本研究では、顔部品検出と顔部品提示を実時間で繰り返すため、目検出も繰り返すこととなる。その際、目検出は1回目と2回目以降の場合で手法が異なる。

1回目はHaar-like特徴を利用して検出する。まず、目が鼻より上であることを考慮し、目の探索領域を、検出した鼻の領域より上の領域に制限する。図6のAの矩形内が目を探索する領域を示している。制限した領域からHaar-like特徴を利用して、目を検出する。なお、Haar-like特徴とは、画像から局所的に一部分を切り出し、その画像から取得できる局所的な明暗差のことである。検出時、検出した目の領域の画像をテンプレート画像として切り出して、取得する。しかし、ここで検出した目の領域内には、目以外の余分な部分(眉毛や肌など)が含まれている。そこで、グラフカットによって削除して、目のみを検出する。図7ではグラフカットによる目のみの検出を示している。

2回目以降の検出では、取得したテンプレート画像とのテンプレートマッチングを行って目を検出する。まず、探索領域は1回目の検出と同様で、図6のAの矩形内に制限して目の探索を行う。なお、目は右目と左目あるので、テンプレートマッチングは2回行う。その際、2回とも片目だけマッチングしてしまう可能性がある。それを防ぐため、検出した鼻の位置を境にして、左目と右目の探索領域を制限して目の検出を行う。図8では、鼻を境とした左目と右目の探索領域を示している。検出後、1回目と同様の理由からグラフカットにより目のみを図7のように検出する。

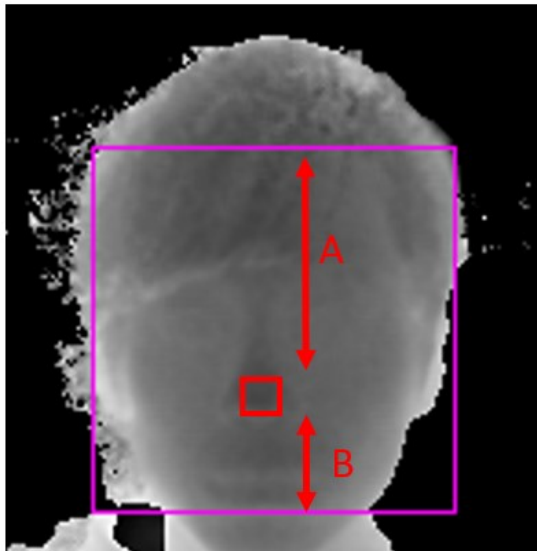


図6 探索領域

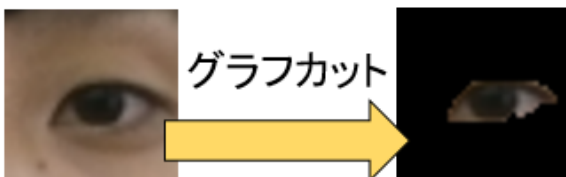


図7 グラフカット

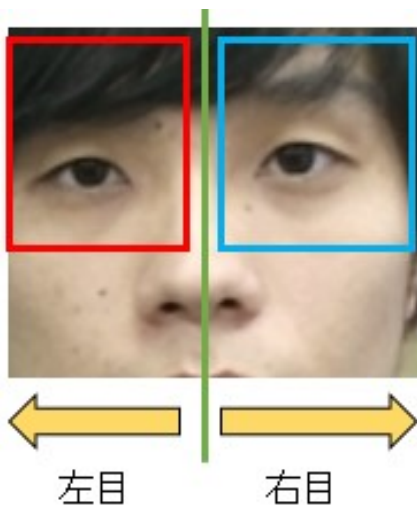


図8 鼻による探索領域制限

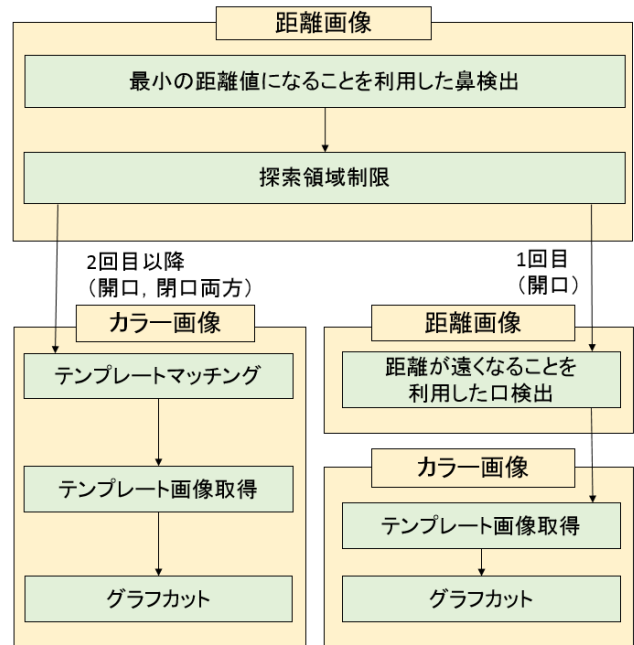


図9 口検出の流れ

3.2.4 口検出

最後に、口を検出する。図9に口検出の全体の流れを示す。本研究では、顔部品検出と顔部品表示を実時間で繰り返すため、口検出も繰り返すこととなる。その際、口検出は、1回目と2回目以降の場合で手法が異なる。

1回目は、ユーザーに開口してもらい、開口時の口を検出する。まず、開口時の口が鼻より下であることを考慮し、口の探索領域を、検出した鼻の領域より下の領域に制限する。図6のBの矩形の領域は探索領域を示している。次に、本システムのユーザーには開口してもらい、開口時の口を検出する。開口時の口内の領域は、唇や肌などの周囲の距離よりも Kinect v2 からの距離が遠くなる。図4のBの開口している領域は周囲と比べて輝度値が高いため、距離が遠いことが分かる。この特徴を利用して口を検出する。検出した鼻の位置から下に探索していき、急激に距離が遠くなる位置、すなわち上唇の位置を検出する。検出後、さらに下へ探索していき、急激に距離が近くなる位置、すなわち下唇の位置を検出する。これらの位置を囲んだ領域を口の領域として検出する。このとき、検出した口の領域の画像をテンプレート画像として切り出し、取得する。しかし、目検出のときと同様に、検出した口の領域には口以外の余分な部分（肌など）が含まれている。よって、グラフカットにより余分な部分を削除して、口のみを検出する。

検出2回目以降は取得したテンプレート画像とのテンプレートマッチングにより口を検出する。まず、探索領域を1回目と同様に図6のBの矩形内の領域に制限する。その後、テンプレートマッチングを行う。1回目は開口時の口の検出だけ行ったが、2回目以降は開口時の口と閉口時の

口の両方をテンプレートマッチングにより検出する。検出後、1回目と同様の理由から図7のようにグラフカットにより口のみを検出する。

3.3 顔部品呈示

3.3.1 初期条件

顔部品呈示の全体の流れを図10に示す。顔部品を呈示する対象物をKinect v2で撮影し、距離データとカラー画像を取得し、3次元データを取得する。なお、背景を削除するために対象物以外の点群は削除している。

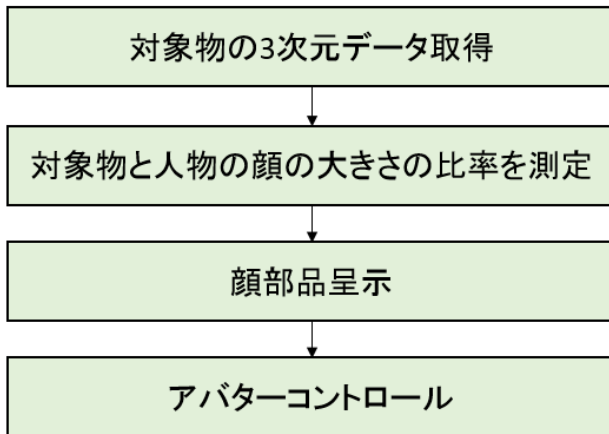


図10 顔部品呈示の流れ

3.3.2 比率計算

削除した際に設定した閾値の座標 x, y, z から対象物の位置、幅、高さを調べる。これを用いて対象物の幅・高さ Kinect v2 で撮影された人の顔領域の幅と高さから比率を計算し、それぞれの顔部品の x 座標と y 座標に比率を掛けることで、対象物に最適な顔部品の位置を計算する。同様にして、検出した顔部品の幅、高さに比率を掛けることで対象物に最適な幅・高さを計算する。計算式は式(2)(3)(4)(5)であり、 Ox, Oy, Ow, Oh は対象物の領域の左上の座標と幅と高さ、 Fx, Fy, Fw, Fh は Kinect v2 で撮影された人の顔部品の位置と幅と高さ、 Px, Py, Pw, Ph は顔領域の位置と幅と高さ、 x, y, w, h は対象物の顔部品を呈示する位置と幅と高さを示している。図11では上記に示した変数を図式化している。

$$x = Ox + (Fx - Ox) \cdot \frac{Ow}{Pw} \quad (2)$$

$$y = Oy + (Fy - Oy) \cdot \frac{Oh}{Ph} \quad (3)$$

$$w = Fw \cdot \frac{Ow}{Pw} \quad (4)$$

$$h = Fh \cdot \frac{Oh}{Ph} \quad (5)$$

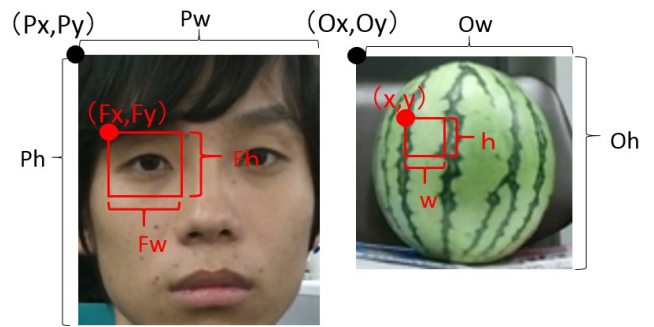


図11 比率計算

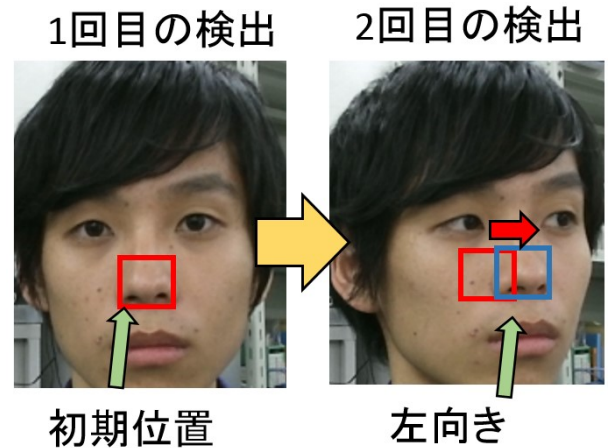


図12 向きの検出

3.3.3 顔部品呈示

3.3.2で計算した位置、幅、高さを元に物に顔部品を呈示する。実時間で顔部品検出・呈示を繰り返しているため、ユーザーの口の開閉、瞬きなどの顔部品の動きに合わせて呈示対象物の顔部品も連動できるようにしている。

また、本システムのユーザーの顔の向きも同時に推定している。まず、最初に検出した鼻の位置を初期値として取得し、ユーザーの顔の正面とする。その後、ユーザーの顔の向きが変化すると鼻の位置も変化するため、その位置を初期値と比較し、顔の向きを推定している。図12ではこれによりユーザーの顔の向きに合わせて、アバターも上下左右に向きを変える。

4. 実験及び考察

4.1 実験環境

本実験では、Microsoft製のKinect v2を用いる。被験者にはKinect v2から約50[cm]離れた椅子に座ってもらい、Kinect v2で撮影する。顔部品呈示の対象物のデータは、Kinect v2から約50[cm]離して、撮影し取得する。距離データとカラー画像を取得し、640×480の距離画像と1920×1080のカラー画像を表示できる。実験で用いる顔部品を呈示する対象物は西瓜と林檎と大根とする。本実験では、顔部品検出の精度の調査と複数の被験者における顔部品呈示、複数種類の対象物への顔部品呈示の3パターンの実験を行

う。

4.2 実験 1

4.2.1 実験内容

実験では、開口時の場合と閉口時の場合で顔部品検出を行った。検出の精度の評価として、目視で確認した顔部品の位置と検出結果との比較を行った。図 13 は開口時と閉口時の目視で確認した顔部品の位置を赤色の点でマークしている。そして、検出結果として顔部品それぞれの矩形が描画されるが、その中心座標を取得した。また、目視で確認した顔部品の位置を選択してその座標を取得した。取得した 2 つの座標を比較した。開口時と閉口時の場合で行った。

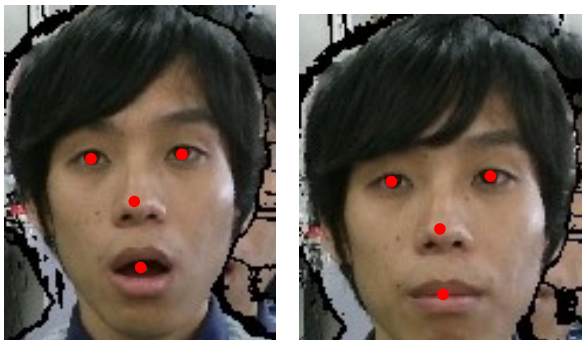


図 13 開口時と閉口時の比較画像

4.2.2 結果

顔部品検出結果を図 14 に示す。また、目視で確認した顔部品の位置 (図 13 の赤色の点) と検出した顔部品の位置を比較した結果を表 1, 表 2 に示す。

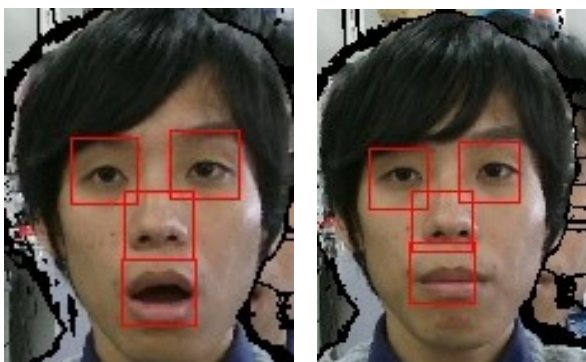


図 14 開口時と閉口時の顔部品検出

表 1 開口時の顔部品検出の誤差[pixel]

	検出結果	比較用画像	誤差
鼻	263,188	264,191	-1,-3
口	262,218	269,221	-7,-3
左目	285,168	284,171	+1,-3
右目	241,170	246,173	-5,-3

表 2 閉口時の顔部品検出の誤差[pixel]

	検出結果	比較用画像	誤差
鼻	263,196	264,195	-1,+1
口	263,219	267,222	-4,+3
左目	284,174	286,172	-2,+2
右目	243,177	244,174	-1,+3

4.2.3 考察

表 1 と表 2 を見ると、誤差が数ピクセルとなっており、良好な検出結果となっていると考えられる。しかし、閉口時に検出した口の領域外に口がはみ出している。閉口時の口は開口時の口の横幅に比べてサイズが大きくなっている。したがって、閉口時の口を検出した場合、検出結果の領域を大きめにする必要がある。

4.3 実験 2

4.3.1 実験内容

実験 2 では、様々な被験者で検出した顔部品を物に呈示する。本実験では、呈示する対象物は西瓜とする。被験者同士の顔部品が呈示された西瓜の比較を行い、呈示された顔部品に被験者の特徴 (目が大きいなど) が反映されているかを確認する。

4.3.2 結果

顔部品呈示結果を図 15~17 に示す。左の図が被験者で、右の図が顔部品呈示結果である。なお、処理時間は、撮影ボタンを押した瞬間、実時間で検出を行う。

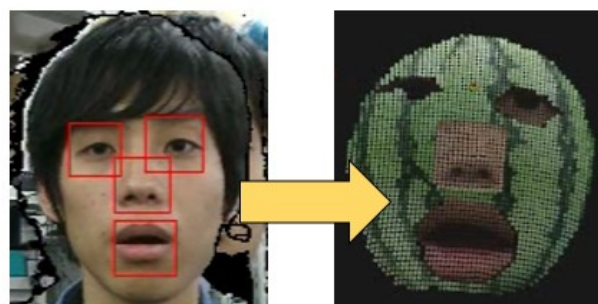


図 15 1 人目の顔部品呈示の結果

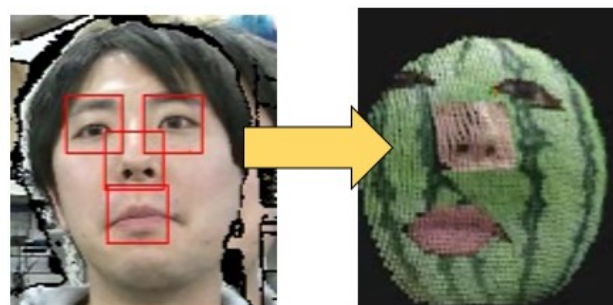


図 16 2 人目の顔部品呈示の結果

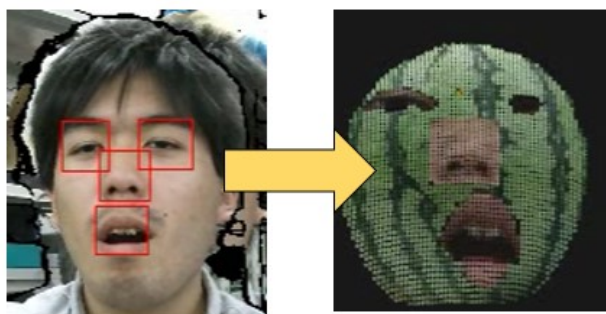


図 17 3人目の顔部品呈示の結果

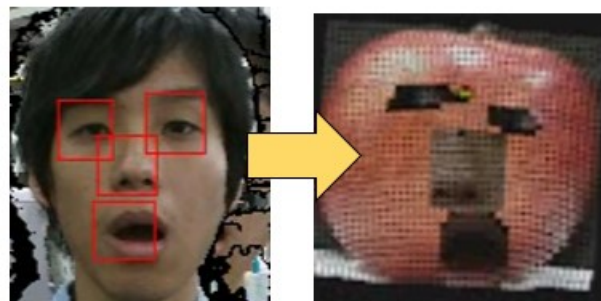


図 18 林檎への呈示結果

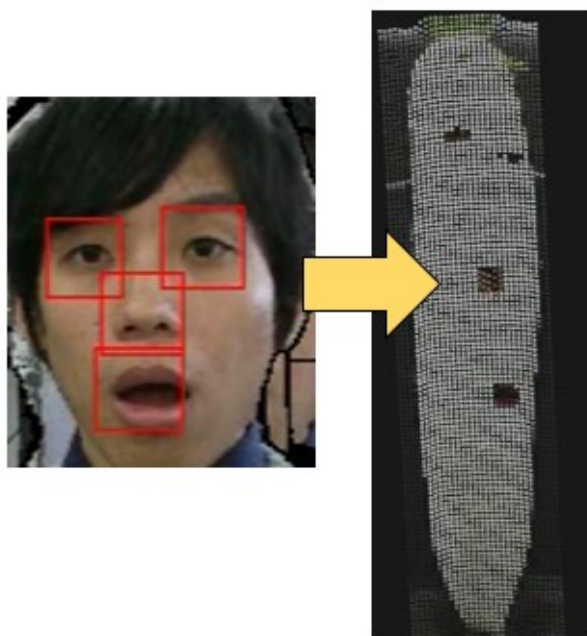


図 19 大根への呈示結果

4.3.3 考察

それぞれのユーザーで試した結果、図 15~17 から検出した顔部品が反映されたことが分かる。目や口が大きい人は呈示後、目や口が大きくなり、それぞれのユーザーの顔部品の特徴が表れていることが分かる。

以後、使用してもらったユーザーにアンケートをとることでシステムの実用性を調べる。主な項目として、「個人性が反映されているか」「システムの使用感の快適度」などを5段階評価にして調べる予定である。

4.4 実験 3

4.4.1 実験内容

実験 2 では、検出された顔部品を複数種類の物に呈示する。呈示する物は形や大きさがそれぞれ異なる物を使用するため、本研究で使用する物は林檎、大根とする。

4.4.2 結果

顔部品呈示結果を図 18、図 19 に示す。左の図が被験者で、右の図が顔部品呈示結果である。

4.4.3 考察

図 18, 19 から分かるように林檎や人参などの野菜や果物に呈示することができた。林檎は西瓜に比べてサイズが小さく、大根は西瓜に比べて細長い形状をしている。しかし、図 18, 19 を見るとそれぞれの物のサイズに合わせて顔部品の大きさや位置が呈示されていることが分かる。

5. まとめ

本研究では、Kinect v2 で撮影した人の顔部品を自動検出し、顔部品を非生物に呈示するシステムを提案した。本システムのユーザーの顔部品の動きに基づいて、呈示対象物の顔部品も実時間で動かすことができる。また、ユーザーの顔の向きの変化にも実時間で対応して、呈示対象物の向きも変化させることができる。

しかし、目検出のグラフカットを行う際に、安定して目のみを切り出すことができない。これは、眉毛等を含んで検出してしまふ際に起きる。この対策として、目を検出した際に眉毛を含んでいる場合、検出矩形の上部の画素が黒になるため、その黒を除く処理を行う。その処理により、目のみを綺麗に除くことができると考えられる。

また、呈示した対象物の鼻に関して、鼻以外の肌も呈示しているという問題が挙げられる。この対策として、勾配が急になっている部分を鼻とし、勾配が平坦になっている部分を肌とすることで鼻のみを呈示できると考えられる。

参考文献

- [1] 東大貴, 赤津裕子, 小松原明哲. "操作支援アバターを用いたインターフェイスの設計法の提案." 人間工学(2015): S248-S249.
- [2] Sanghun. Nam and Youngho. Chai. "Facial component detection using a Kinect for an Interactive Media Art." (2014).
- [3] 石田貴正, 平野晃昭, 中村納. "表情認識のための顔部品の特定とその抽出." 工学院大学研究報告(2012): 105-112.
- [4] 齊藤文哉, 馮軒昂, 原亜珠紗, 宮地美希, 北栄輔. "Kinect を用いた顔認識について." 設計工学・システム部門講演会講演論文集 2013.23 (2013).
- [5] 森大樹, 本田新九郎, 安野貴之, 石橋聡. "人物顔形状を反映した 3D デフォルメキャラクタ生成システムの構築." 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解(2001): 19-26.
- [6] Feng. Xu, Jinxiang. Chai, Yilong. Liu and Xin. Tong. "Controllable high-fidelity facial performance transfer." ACM Transactions on Graphics (TOG)(2014): 42.
- [7] Opencv.jp 物体検出
http://opencv.jp/opencv-2svn/cpp/object_detection.html