

平成28年度 卒業論文

顔とディスプレイの距離に応じた自動ズーム機能の提案

指導教員
舟橋 健司 准教授

名古屋工業大学 工学部 情報工学科
平成22年度入学 22115060 番

名前 佐竹 祐紀

目次

第1章	はじめに	1
第2章	顔とディスプレイの距離に応じたズーム	3
2.1	顔認識を利用した測定	4
2.2	目の認識を利用した測定	9
第3章	実験	12
3.1	実験システム	12
3.2	実験1：認識対象の変化による評価と考察	13
3.3	実験2：操作条件の変化による評価と考察	14
第4章	むすび	15
	謝辞	16
	参考文献	17
	発表論文リスト	18

第1章 はじめに

近年、入力インターフェースの種類は多様化している。キーボードやタッチパネルのように押すことで入力を行うインターフェース、マウスやゲームコントローラーに見られるような装置そのものを動かすことで入力を行うインターフェース、また音声入力インターフェースや視線入力インターフェース、ジェスチャーを直接入力するインターフェースなどがある。これらの入力インターフェースの研究はより扱いやすく正確に操作を行うという点に主眼をおいたものが多い。例えば視線入力インターフェースの簡略化 [1] や音声認識による候補数の選択 [2] に注目した研究がある。そういった点に主眼を置いた研究が多いのは、ユーザにとっては入力インターフェースは自身が意図したことを意図したとおりに行わせる事ができるということが重要だからである。

ところで人間は手に持った書類や本などの文字のサイズが小さくてそれを読むのが難しいと感じた場合、顔を書類や本に近づけるか手を動かすかしてそれらとの距離を縮めることで文字を見やすくして読もうとする。また、地図や新聞などの大きな紙面の全体像を見たいと思った場合、それらから離れそれらと顔との距離をとることで全体像を視界の中に収めることで全体を見る。一方でPCの画面やタブレットデバイスで同じように、文字が読みづらいので画面を拡大したい、全体像を見たいので画面の縮小を行いたいといった状況になると、人間はマウスを操作してズーム操作を行うかタッチによるピンチ操作をすることで拡大や縮小の動作を行う [3][4]。しかし、PCやタブレットで行うこれらの操作は日頃からこれらを扱い操作に慣れている人は気にしないが、そうでない人にとっては直感的な操作ができないので少し面倒でもある。

そこで本研究では、より直感的な操作を行うためにもPCとタブレットデバイスを使用している場合にも、本や書類を読むときに顔を近づけたり地図や新聞などから距離をとって全体像を見たりするのと同じように、それらの画面と顔との距離から画面の表示サイズを適切に拡大縮小する手法を提案する。顔とディスプレイの位置関係により表示を変更するという点で関連する研究に、その位置関係の変化から運動視差を考慮して立体視を実現しているものがある [5]。画面と顔の間の距離を利用して何らかの処理を行う手法を考えた場合、画面と顔の間の距離をそのまま操作のパラメータに設定し認識する度に操作を実行するようにすると、身じろぎのような姿勢の微細な変化やカメラの認識誤差に応じて常に変更が生じてしまい、期待通りの結果が得られない恐れがある。そこで姿勢の微細な変化やカメラの認識誤差による影響を避けるために、ある時間において平滑化を行い、しきい値以上の変化が生じた場合にのみ変更処理を実行することを提案する。本論文では、提案

する自動表示拡大縮小機能および、このしきい値や変更処理の度合いを決定するパラメータについて述べる。またこのアイデアの妥当性を確認するための実験について第3章で述べる。第4章では本研究のまとめ、今後の課題について述べる。

第2章 顔とディスプレイの距離に応じたズーム

顔と画面との間の距離変化を測定するためには距離センサを使用することも考えられるが、本研究では距離センサよりも簡単に準備ができるカメラを使用して顔と画面の距離の変化の測定を行うことを試みる。カメラはデスクトップパソコンの場合、モニターの上もしくは下部に設置することで使用者の顔を撮影する。ノートパソコンやタブレットの場合、画面側に内蔵された内臓カメラを使用することで使用者の顔を撮影する。これらのカメラを使って撮影した使用者の顔画像から使用者と画面の距離の変化を測定し、それを利用して画面の拡大縮小を行うか否かの判定を行う。ここでは顔全体を認識(図 2.1)してその大きさの変化から距離の変化を測定した上で拡大縮小処理を行う手法と、二つの目を認識(図 2.2)してそれらの間隔の変化から距離の変化を測定した上拡大縮小処理を行う手法の二種類の手法を提案する。またこの二つの手法を利用する際は画像からの認識をやりやすくするため、カメラに対して顔が正対しているものとする。

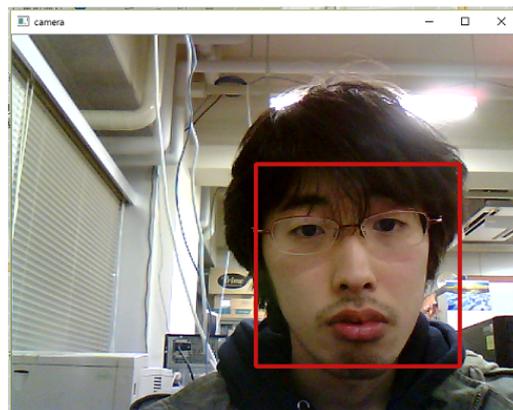


図 2.1: 顔認識

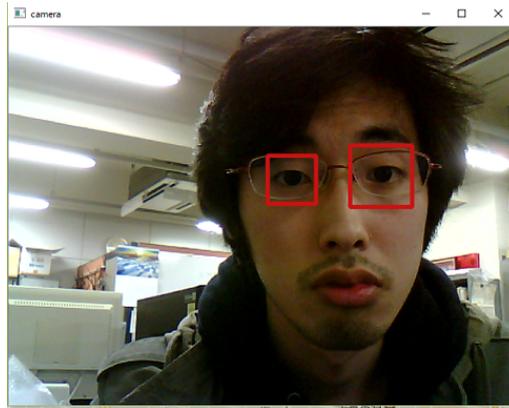


図 2.2: 両目の認識

2.1 顔認識を利用した測定

ある時刻 t においてキャプチャした画像に対して顔認識を行い, 認識した顔の高さを $h(t)$, 顔の幅を $w(t)$ とする (図 2.3). この二つ数値 $h(t)$ と $w(t)$ から, 体の微細な動きや認識誤差の影響を避けるために指定時間 D の間の高さと幅の平均値 $H(t)$ と $W(t)$ を求める. 次に平均値 $H(t)$ と $H(t-D)$, $W(t)$ と $W(t-D)$ の差を計算する (式 2.1). 時刻 t における高さ と幅の平均値と時刻 $t-D$ における高さ と幅の平均値を比較することで指定時間 D の間に画像内の顔の大きさがどれだけ変化したかを算出する. 当初は平均を用いずに時刻 t と時刻 $t-D$ における高さ と幅, すなわち $h(t)$ と $h(t-D)$, $w(t)$ と $w(t-D)$ を比較することも検討したがその場合には時刻 t に誤認識や顔を認識できなかった場合に, 拡大縮小操作を行わなかったり想定以上に変化したりと意図した操作が行えないといった問題が発生することが考えられ, また身じろぎなどの微細な変化や認識誤差による数値の影響もより大きく出てしまう. そのため指定時間 D における平均値を用いることとした.

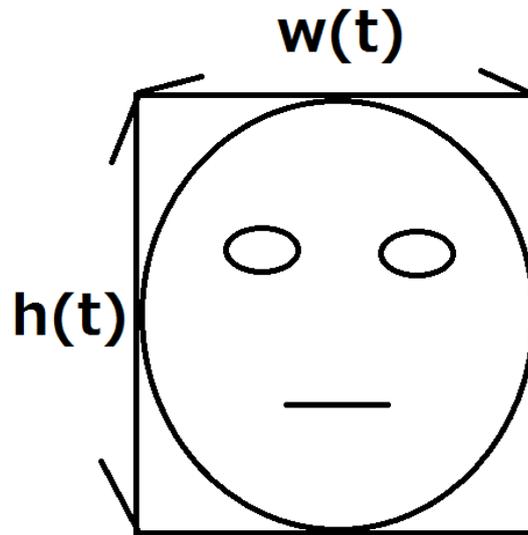


図 2.3: 顔の高さと幅

$$\begin{aligned} H'(t) &= H(t) - H(t - D) \\ W'(t) &= W(t) - W(t - D) \end{aligned} \quad (2.1)$$

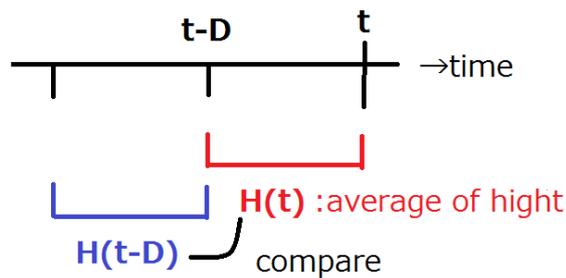


図 2.4: 平均と比較

この比較した数値 $H'(t)$ や $W'(t)$ も身じろぎなどの微細な変化や認識誤差による数値の変化を受けている。そのため、この影響による意図しない操作を減少させるために変化割合が一定値 c_1 ($\times 100[\%]$) 以上のときのみ拡大縮小処理を行うようにする。また姿勢を正すために座り直すような場合にも対応するために変化割合が一定値 c_2 以下のときのみ拡大縮小処理を行うようにする。次の式 (2.2) は画像の拡大操作の時の判定式である。指定時間 D 間での高さや幅の変化量の割合 $H'(t)/H(t - D)$ と $W'(t)/W(t - D)$ が共に c_1 以上 c_2 以下ならば、操作を行う意図があると言えるほど顔とカメラの距離が縮まったと判断し拡大操作を行う (図 2.5)。

$$c_1 < \frac{H'(t)}{H(t - D)} < c_2, \quad c_1 < \frac{W'(t)}{W(t - D)} < c_2 \quad (2.2)$$

縮小操作の時も同様に次の判定式 (2.3) を用いて操作の有無を判定する. 指定時間 D 間での高さと幅の変化量の割合 $-(H'(t)/H(t-D))$ と $-(W'(t)/W(t-D))$ が共に c_1 以上 c_2 以下ならば, 操作を行う意図があると言えるほど顔とカメラの距離が離れたと判断し縮小操作を行う (図 2.6).

$$c_1 < -\frac{H'(t)}{H(t-D)} < c_2, \quad c_1 < -\frac{W'(t)}{W(t-D)} < c_2 \quad (2.3)$$

これらの式で操作の有無を判定し操作を行う場合, その時の拡大縮小操作の倍率 s を次の式 2.4 を用いて設定する.

$$s = \frac{1}{2} \left(\frac{H(t)}{H(t-D)} + \frac{W(t)}{W(t-D)} \right) \quad (2.4)$$

この時, 時刻 t において即座に画像サイズを s 倍すると画像の変化が即座で滑らかで無いので時刻 t から時刻 $t+D$ にかけて段階的に拡大縮小を行うことでスムーズなアニメーションとして表示する. ここで $t = D/N$ (N は整数) とすると時刻 $t+i$ ($0 \leq i \leq N$) 時点における倍率を次の式から算出する.

$$s(t+i) = 1 + \frac{s-1}{D} \times i \quad (2.5)$$

時刻 $t+i$ における画像サイズが時刻 t における画像サイズの $s(t+i)$ 倍となるように画像を拡大縮小する. この操作を繰り返す事で時刻 $t+D$ 時点, すなわち $i=N$ の場合における画像サイズを時刻 t における画像サイズの s 倍とする.

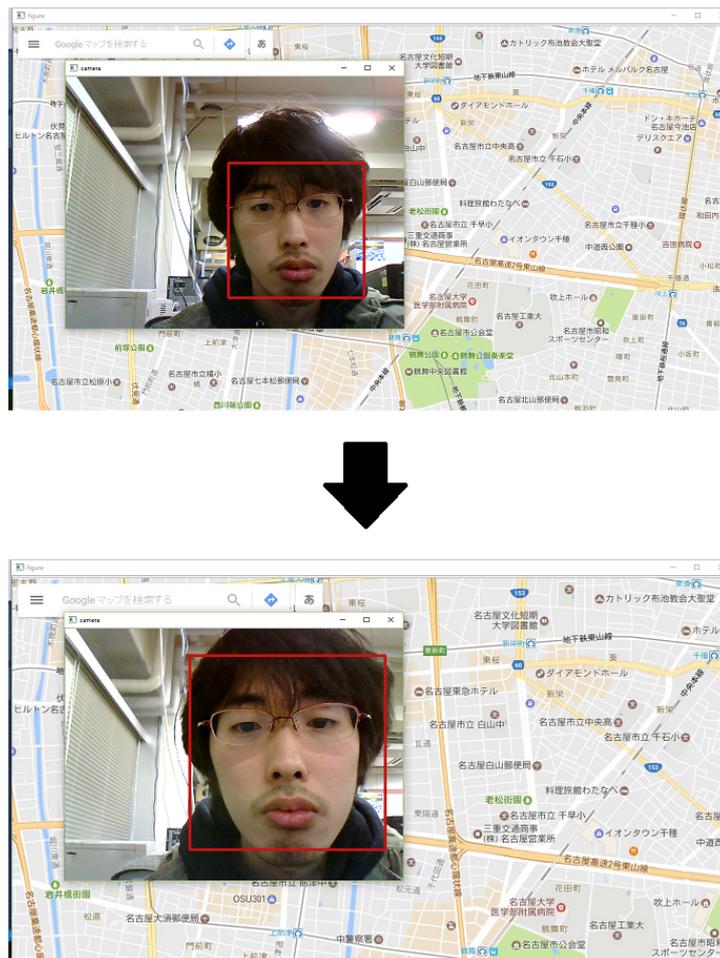


図 2.5: 拡大

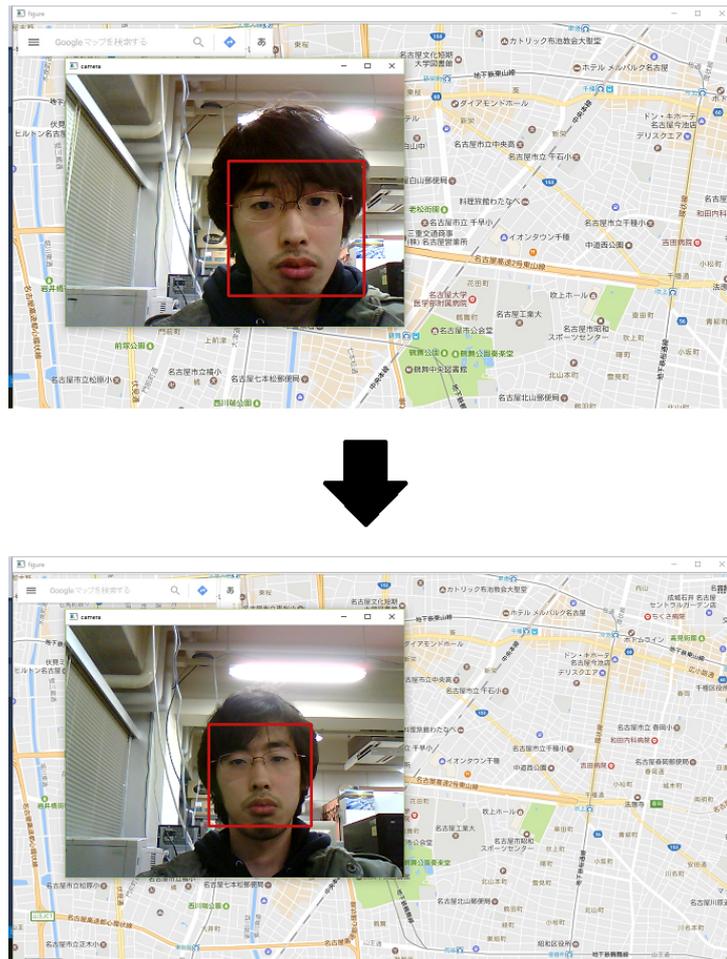


図 2.6: 縮小

2.2 目の認識を利用した測定

ある時刻 t においてキャプチャした画像より目の認識を行い、認識したそれぞれの目の中心座標を $\{x_1(t), y_1(t)\}$ と $\{x_2(t), y_2(t)\}$ とする (図 2.7). そして顔認識の場合と同様に身じろぎなどによる微細な変化や認識誤差の影響による意図しない操作を減少するために指定時間 D の間の中心座標の平均値 $\{X_1(t), Y_1(t)\}$ と $\{X_2(t), Y_2(t)\}$ を求める. この左右の目の中心点の平均値よりピタゴラスの定理を用いて次の式から指定時間 D 間における画像内の目の間隔の平均値 $E(t)$ を求める (式 2.6).

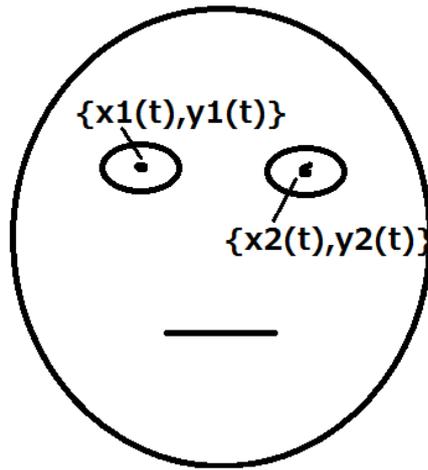


図 2.7: 目の中心点

$$E(t) = \sqrt{(X_2(t) - X_1(t))^2 + (Y_2(t) - Y_1(t))^2} \quad (2.6)$$

続いて平均値 $E(t)$ と $E(t - D)$ を比較することで画像内の目の間隔の変化を算出し、意図しない操作を減少させるために変化割合が一定値 c_1 ($\times 100$ [%]) 以上、一定値 c_2 以下のときのみ拡大縮小処理を行うようにする. そのための判定式は次の式 (拡大:式 2.7, 縮小:式 2.8) である.

$$c_1 < \frac{E(t) - E(t - D)}{E(t - D)} < c_2 \quad (2.7)$$

$$c_1 < -\frac{E(t) - E(t - D)}{E(t - D)} < c_2 \quad (2.8)$$

式中の $(E(t) - E(t - D))/E(t - D)$ が c_1 以上 c_2 以下ならば拡大操作を行う意図があるといえるほど顔とカメラの距離が縮まったと判断して拡大を行い (図 2.8), $-(E(t) - E(t - D))/E(t - D)$ が c_1 以上 c_2 以下ならば縮小操作を行う意図があるといえるほど顔とカメラの距離が離れたと判断して縮小を行う (図 2.9). またその時の倍率 s の値は次の式 (2.9)

で設定され, その段階的な変化は式 (2.5) により処理する.

$$s = \frac{E(t)}{E(t - D)} \quad (2.9)$$

顔認識の場合と同様に時刻 $t + i$ t における画像サイズが時刻 t における画像サイズの $s(t + i - t)$ 倍となるように画像を拡大縮小する. この操作を繰り返す事で時刻 $t + D$ 時点, すなわち $i = N$ の場合における画像サイズを時刻 t における画像サイズの s 倍とする.

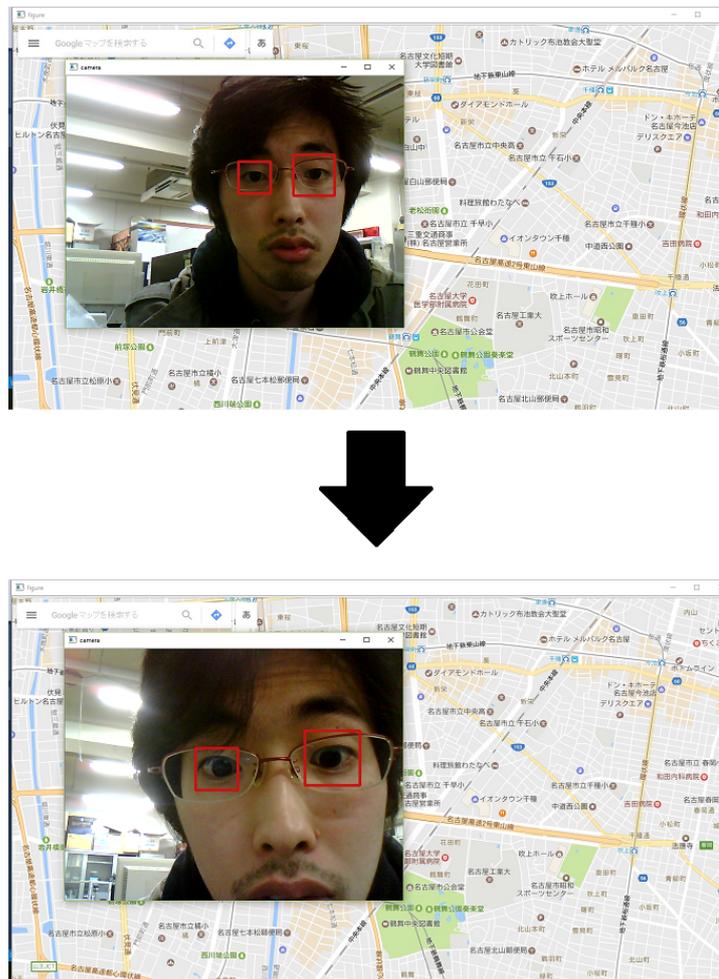


図 2.8: 拡大

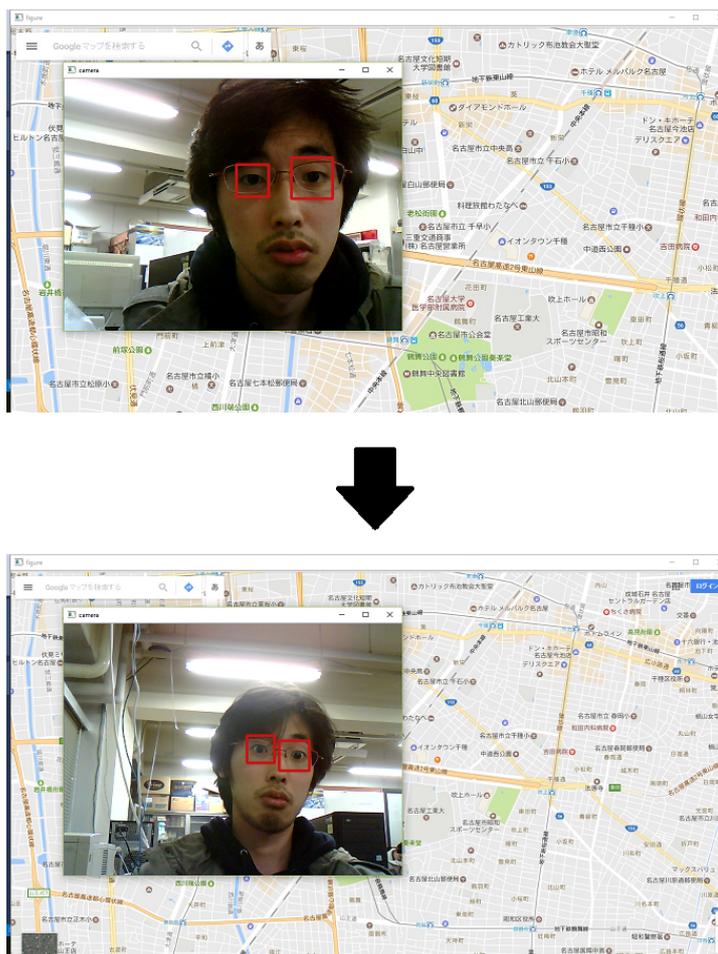


図 2.9: 縮小

第3章 実験

第2章で述べた提案手法に基づき、実験システムを構築し実験を行うことで本手法の妥当性の検証と各パラメータの考察をする。実験では20代の男女合計8名に、複数のパラメータを設定した実験システムに対してそれぞれのパラメータでの使用感想を聞くことで評価を行う。

3.1 実験システム

実験システムはCPU: Intel Atom CPU Z2760 1.80GHz, OS: Windows 10のタブレットPC上にC++を用いて実装した。またBUFFALO社製のUSBカメラBSW20K04を使用し画像サイズを 640×480 pixel, 最大30fpsに設定して撮影を行い認識をした(図3.2)。顔や目の認識をするためにOpenCVの標準ライブラリを使用し、提案手法中の指定時間 D を6フレームと設定した。

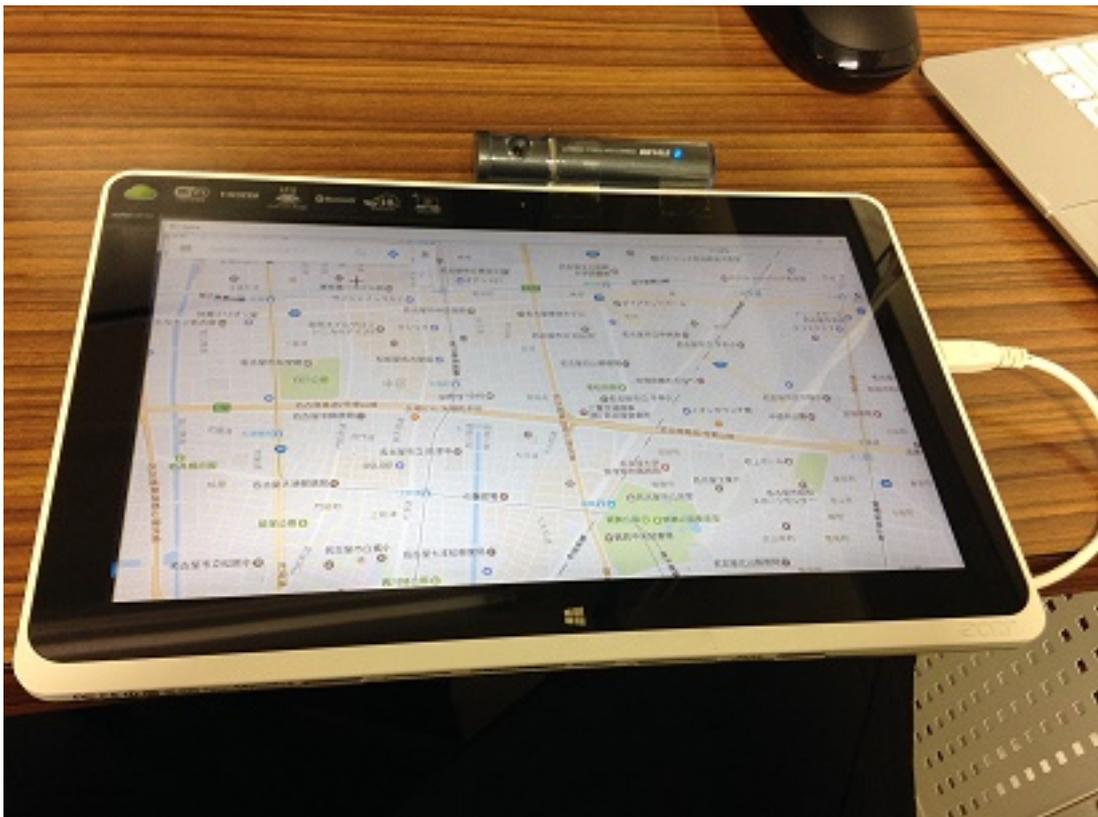


図 3.1: 使用したタブレット



図 3.2: 実験の様子

3.2 実験 1：認識対象の変化による評価と考察

顔を認識して作動する実験システムと目を認識して作動する実験システムのどちらの方が使用者の動作に追従できるか評価する。被験者にはそれぞれのシステムを利用してもらいどちらの方が操作しやすかったかを評価してもらった。本実験では操作条件を平等にするため提案手法中の操作条件 c_1 を予備実験よりどちらの実験システムに置いても 0.04, c_2 を予備実験よりどちらも 0.5 と設定した。表 3.2 はその結果で「こちらの方が扱いやすかった」と答えた人数を集計した。

顔認識	目の認識
0人	8人

表 3.1: 実験 1 結果

全員が顔認識で作動する実験システムよりも目の認識で作動する実験システムの方が扱いやすいと答えた。これは拡大する操作を行った際にタブレットと顔の距離が近くなり、顔全体をカメラの撮影範囲に収めることができず顔の認識を行うことができなくなり、操作に支障が出たためと思われる。目の認識の利点に目の部分がカメラの撮影範囲に入っていれば認識を行うことができ顔全体を撮影範囲に収める必要が無いという事があるが、こ

の実験結果はその点が大きく出た結果であるといえる。この結果からこの実験システムに置いては目の認識を用いた方が適切に使用者の動作に追従できていると言える。

3.3 実験2：操作条件の変化による評価と考察

提案手法中の操作条件 c_1 の値はそれが意図した操作であるか否かを識別するための値である。そのためこの値によってシステムの操作性は大きく変わることになる。そこで適切な操作条件 c_1 を探すために、複数のパターンの c_1 で拡大縮小操作を行ってもらい、どれが一番扱いやすかったかを集計した。なお c_2 の値は実験1と同じく予備実験から0.5と設定した。目を認識して作動する実験システムを使い c_1 の値を0.01, 0.04, 0.07の3パターン用意しこれらを実際に操作してもらい、扱いやすかった順を集計した。表3.3はその結果である。

C_1	0.01	0.04	0.07
A	2	1	3
B	2	1	3
C	1	2	3
D	3	1	2
E	1	2	3
F	2	1	3
G	2	1	3
H	3	1	2
一番の人数	2	6	0

表 3.2: 実験2 結果1

もっとも扱いやすかったと答えた人数が多かったのは c_1 の値が0.04の時であった。実験者の感想から0.01では反応が過敏すぎて静止しているつもりでも反応し、意図してないのに操作が行われ、0.07では反応が悪く操作する意図があつたにもかかわらず操作できなかったということがあったからと考えられる。0.01よりも0.07の方が扱いやすいという評価が多いのは反応しないよりも過敏でもいいから反応する方が操作することができるからであると思われる。この結果からこの実験システムに置いては c_1 の値は0.04が適切であるといえる。今後の課題として個人差も含め c_1 の値を0.025, 0.04, 0.055で比較する等、より詳細な調査をしていきたい。

第4章 むすび

本研究では、顔と画面の距離を適切に参照して操作を行う手法を提案し、その妥当性を検証するためのサンプルとして画像の拡大縮小を行うシステムを作成し実験を行った。作成したシステムでは指定時間の間にディスプレイに顔を指定範囲の速度で近づけると表示画像は拡大され、またディスプレイから指定範囲の速度で遠ざかると画像は縮小される。最初の実験では顔認識の場合と目の認識の場合のそれぞれを利用してもらい使用者の動作に追従できるかを評価した。その結果、タブレットのような手に持って扱うデバイスの場合カメラの撮影範囲に収まり切らなくても認識できる目の認識の場合の方が動作に追従することが検証できた。2つ目の実験では操作条件のパラメータを変更しながらシステムを使用してもらい使用者の動作に追従できるかどうかを評価をしてもらった。その結果、パラメータの違いによる操作性の違いを確認することができた。また反応しづらい数値より過敏に反応する数値の方が評価は高いことも確認できた。

今後の課題として、経験から設定した数値 (指定時間 D , 操作条件 c_1, c_2 , 倍率 s) を調整することにより操作性や追従性を向上させること、顔と画面間の距離を調整する動作から老眼であることを認識して老眼に対応したズームを行うこと、本研究の有効性や妥当性を更に確認、検証すると言ったことが挙げられる。またこの手法を拡大縮小操作以外の異なる操作へ応用させること、顔の動作とカメラによる認識の関係を使った車の電子ミラーへの応用といった他分野での研究に本研究が手助けになることを期待したい。

謝辞

本研究を進めるにあたって，日頃から多大な御尽力を頂き，御指導を賜りました名古屋工業大学，舟橋健司 准教授，伊藤宏隆 助教に心から感謝致します．最後に，本研究に多大な御協力頂きました舟橋研究室諸氏に深く感謝致します．

参考文献

- [1] 高 トククン, 中井 直輝, 板倉 直明, 水戸 和幸, 水野 統太, “Eye Gesture 入力インタフェースシステムの開発 (コミュニケーション支援及びヒューマンコミュニケーション一般)” 信学技報, HCS2012 - 6, pp.31-36, May 2012.
- [2] 趙 國, 山下 洋一, “N-best 音声認識における認識スコアを利用した候補提示数の決定” 信学技報, SP2003 - 182, pp.43-48, Jan 2004.
- [3] Pierre Wellner, “The DigitalDesk calculator: tangible manipulation on a desk top displa”, *Proceedings of the UIST '91*, pp.27-33, 1991
- [4] George W. Fitzmaurice, Hiroshi Ishii, William A. S. Buxton, “Bricks: laying the foundations for graspable user interfaces”, *Proceedings of the CHI '95*, pp.442-449, 1995
- [5] Jeremie Francone, Laurence Nigay, “Using the User’s Point of View for Interaction on Mobile Devices”, *Proceedings of the 23th ACM International Conference of the Association Francophone d’Interaction Homme-Machine (IHM2011)*, pp.25-31, 2011

発表論文リスト

国際会議

1. Yuki Satake, Kenji Funahashi, “Automatic Display Zoom Using Face Size of Camera Image”, *ICAT-EGVE 2016*, pp.1-2, LittleRock, USA, Decn. 2016