研究テーマ　視線検出に関する研究

研究者　本郷仁志　財団法人ソフトピアジャパン　雇用研究員
加藤邦人　岐阜大学　共同研究員

フェーズI

1 研究の概要

フェーズIでは、知的室内環境パーセプトルームを構築するために、ジェスチャ認識技術とあわせ、視線を検出する技術の構築を行った。リビングのように部屋全体に配置された機器を対象とする場合、まず、操作対象の機器を選出する必要がある。この研究では、利用者がジェスチャで指示を出したときの注目機器を対象物として推定する方法として、室内の壁面に配置した複数のカメラから、手サインを提示したタイミングでの利用者の位置と顔方向から注目対象物を推察し、複数の候補がある場合は対象機器に備えたカメラで検出した瞳の形状から視線方向を判断して決定する方法を構築した。

2 研究の目標

ジェスチャ認識技術をインタフェースに応用し、機器の操作を容易にするなど、知的環境を構築する研究が盛んに行われている。知的空間の構築を目指すために、ジェスチャ認識の精度を向上させることも重要であるが、リビングのように室内の様々な場所に置かれた複数の機器を操作するには、まず操作する機器を選択することが重要となる。利用者が操作対象物に注目することは自然な動作であることから、視線方向を検出することで、注目対象物を推定できると考える。本研究では、人物が何に注目しているのか視線方向を検出し注目対象物を推定することで、その人物が提示するジェスチャに応じて対象機器を操作できる環境を構築することを目的としている。特に、人物の場所や方向に依存しない知的環境の構築を目指している。

注目対象物を推定するためには、まず視線方向を検出する必要がある。視線検出技術は古くから研究されており、すでに様々な視線インタフェースが開発されている。しかし、これらの装置は、視線検出精度を上げるために被験者の頭部の動きを制限しかつ、ユーザに大きな負担を強いるキャリブレーションが必要となる。ステレオカメラによりシステムとユーザとの距離を測ることで、被験者頭部の前後の動きに対して自由度を向上させた装置が開発されている。しかし、室内を歩き回る人を対象にすることは困難である。我々が必要とする視線検出は、画面上の正確な注視点を求めるよりもむしろ、会話している相手や注目対象物など視線を向けている先の対象物を検出できることである。さらに、室内の位置に依存せず、あらゆる方向の視線を検出する必要がある。既存システムを複数台同時に利用することで計測範囲を拡張しようとした場合、赤外光の干渉、眼部の高解像度画像の獲得などに問題が発生し、室内空間で視線方向を検出することは困難となる。

そこで、室内にある対象物に対して、何に注目しているかを判断する注目対象物推定方法を提案する。本手法は、壁に備えた複数のグローバルカメラと対象物に装備したターゲットカメラの協調により視線検出を行う。まず、グローバルカメラで人物位置と動作、および顔方向を推定し、対象物に備えたカメラを制御して人物の顔領域をズームし、その顔画像から検出した瞳の円形度を比較することで注目を推定する。つまり、円形度を比較することで視線方向を推定する方法となる。従来のワールド座標系に変換した視線方向を算出するのでなく、新しいパラダイムによる視線検出方法である。

3 実施内容

3.1 知的環境の構築

本研究では、知的環境としてリビングを想定し、室内にある家電製品をジェスチャで操作で

きる環境の構築を目指している。つまり、ルームが人物のジェスチャを察知し、その人の代わりに機器を操作することになる。このような環境下では、ユーザに負担をかけず、室内のどこからでも容易に機器を操作できることが課題となる。知的環境の想定シーンを図1に示す。利用者はソファに座り、ＴＶに向かって手サインを提示し、スイッチを入れたシーンである。

図1　知的空間内での想定シーン

このような場合、複数の機器が対象となるため、操作する機器を選択する必要がある。利用者が操作したい機器を注目することは自然な動作であることから、操作する機器に注目しながら手サインを提示することは、機器とコミュニケーションする上で自然な形と考えられる。よって、手サインを提示する際に利用者が注目している機器を判別することで、操作対象の機器を特定することが可能と考える。利用者にとっては機器の選択が意識せずに行えることになる。

本システムは、知的環境内の情報を集中管理するメインＰＣと、壁に備えた複数のグローバルカメラと、室内にある対象物に装備したターゲットカメラから構成される。各カメラはそれぞれＰＣに接続されている。

グローバルカメラは、室内空間の死角をできるだけ少なくなるように 16 台のカメラを図 2 に示すように配置した。主に、人物の位置および顔方向を検出するのに用いる。一方、ターゲットカメラは、パン・チルト・ズームの機能を備え、部屋周辺に配置された各対象物に配置する。グローバルカメラで得られた顔位置情報を元に、カメラを制御して顔領域を拡大した画像を獲得し、注目判定を行う。

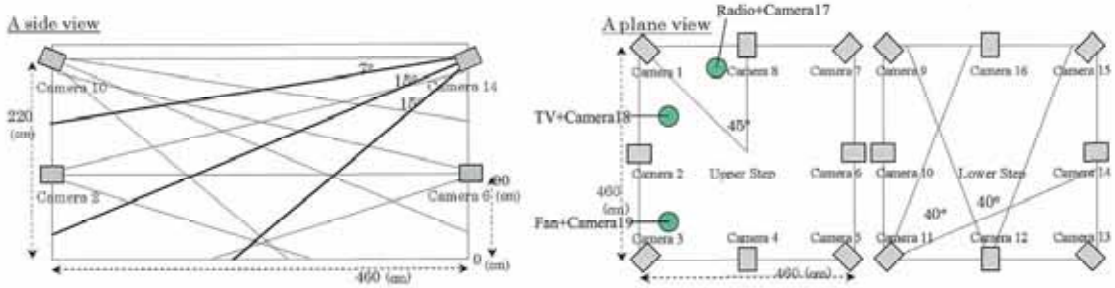


図2　カメラ配置

本システムでは、まずメインＰＣが、画面合成ユニットにより 16 台のグローバルカメラ映像を 1 つに統合した合成画像から、人物の位置および動作を識別する。画面合成ユニットにより、各カメラからの入力画像の解像度は、160×120 画素と低下するが、 16 台のカメラ映像を同時に処理することができ、検出情報の統合が容易となる。ここで得られる映像を図 3 に示す。図中の番号は図 2 に示したカメラ番号に相当する。

次に、頭部が撮像されていると予想されるすべてのグローバルカメラに、顔方向推定の命令を送信し、推定結果を収集し統合する。この時点で人物位置と顔方向から対象物を限定することができる。複数の候補がある場合、ターゲットカメラを制御して、拡大した顔画像を獲得し瞳検出を行う。検出した瞳の円形度を比較することで最終的にメインＰＣが注目対象物を推定する。検出情報は TCP/IP で伝送する。リビングの大きさは、4.6×4.6 m^2 である。

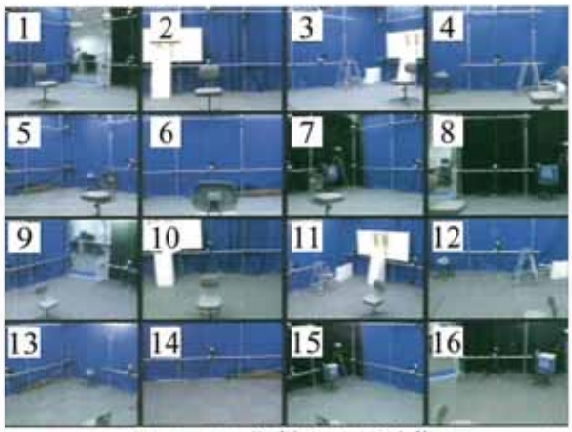


図 3　16 分割カメラ映像

3.2 注目対象物推定方法

本手法について、処理過程の順に説明する。

3.2.1 人物位置検出と手サイン提示検出

人物位置を求める方法として、ステレオマッチングや、視体積交差法、多視点カメラを用いた方法などがある。本システムでは、人物とカメラの位置関係によっては人物像がフレームからはみ出る場合があり、カメラ間で検出領域が同一人物なのかを判断する必要がある。そこで、人物の同定ならびに人物の位置、動作領域の抽出が容易に行える、背景差分とフレーム間差分との統合型視体積交差法を用いる。背景差分とフレーム間差分を個々に視体積交差法でボクセル空間に投票し、検出されたボクセルの位置関係を分析することで、人物領域のどの部位が動作しているか簡易に検出できる。

視体積交差法のボクセル投票により、図 3 のように人物ブロックを検出する。このときのカメラ 16 台の合成画面は図 4 であり、図 3 の人物ブロックから逆射影した、各カメラごとの人物領域を矩形で示している。検出された手領域を矩形で示している。

図 6 は、人物が手を上げた場合のフレーム間差分で得たボクセルであり、動領域ブロックが人物ブロックの上部に出現している。ここでは、動領域ブロックの重心が垂直方向にある閾値以上移動した場合、手を上下させたと判断する。手を上げたときの、人物ブロックの重心位置を人物位置として検出する。このときのカメラ 16 台の合成画面は図 7 であり、各カメラごに検出された手領域を矩形で示している。

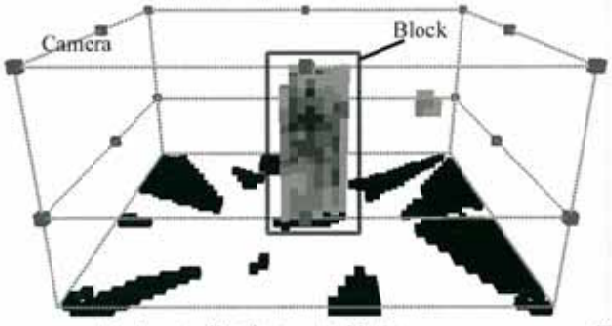


図 4　人物ブロック検出

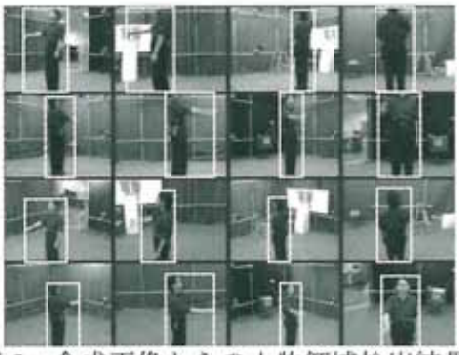


図 5　合成画像からの人物領域検出結果

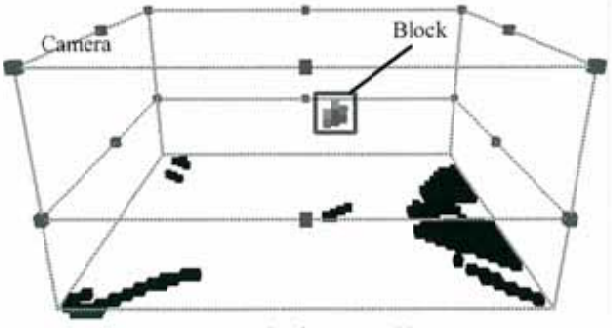


図 6　動領域ブロック検出

図 7　合成画像からの手領域検出結果

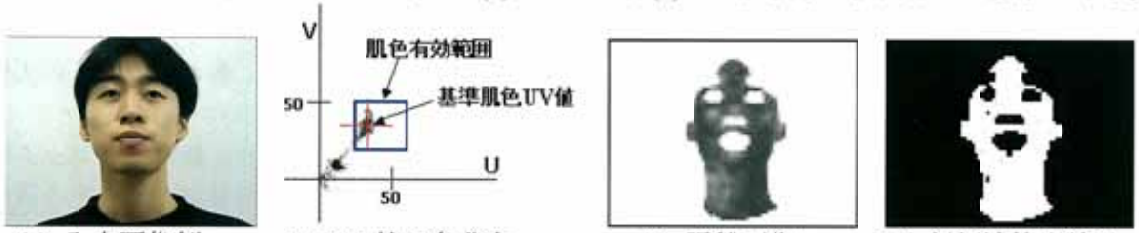
しかし、ボクセルだけでは人物がどの方向を向いているか分からない。そこで、人物ブロックの上部を頭部と仮定し、その領域を捉えるグローバルカメラに対して、その領域と顔方向推定の命令を送信する。

2.2.2 顔方向推定

メイン PC から命令を受けたグローバルカメラは、指定された領域に対して顔検出を行い、顔方向の推定結果を返す。室内全体を把握するメイン PC において、顔領域を特定し個々のグローバルカメラ画像上で対応する領域を指定することで、適切な領域に対して顔方向推定が行える。

グローバルカメラでは、顔方向を推定するために、まず肌色基準値手法を用いて顔領域を抽出し、その領域に対して 4 方向面特徴と線形判別分析を用いて、顔方向を推定する。

肌色基準値手法は、肌色の個人差やカメラの個体差による色の変化を肌色基準値が変動することで吸収する。なお、本実験ではルームの四方をブルーバックで囲み、顔領域が容易に抽出できるようにした。まず、入力画像の各画素の RGB 値を LUV 表色系の U, V 値へ変換し、2 次元の U, V 値ヒストグラムを作成する。入力画像図 8(a)の U, V 値の色分布を図 8(b)に示す。あらかじめ定めた肌色有効範囲内で画素数が最大の U, V 値を基準肌色 UV 値とする。なお、肌色有効範囲は日本人をサンプルとした予備実験により定めた。図 8(b)中の矩形は肌色有効範囲を示し、十字の中心は基準肌色 U, V 値を示す。次に、入力画像の各画素の U, V 値と基準肌色 U, V 値との距離を計算する。図 8(c)の画像は図 8(a)の各画素の U, V 値と基準肌色 U, V 値との距離を計算し、距離値を濃淡値に置き換えた画像である。色が黒いほど基準肌色 U, V 値に近いことを表している。そして、基準肌色 U, V 値からの距離の画像のヒストグラムを作成して判別分析法により 2 値化し、顔領域を抽出する。図 8(d)は入力画像(a)より顔領域が抽出された結果である。



(a) 入力画像例　(b) UV 値の色分布　(c) 距離画像　(d) 顔領域抽出結果

図 8 肌色抽出処理

顔方向推定には、顔領域から抽出した 4 方向面特徴を基に線形判別分析により生成した顔方向判別空間を用いる。特徴量は、顔領域から 4 方向面特徴を抽出した後、8×8 に低解像度化した 256 次元とする。

顔方向推定実験では、学習データとして 50 名から収集した顔画像を、水平方向 15 度間隔で ±45 度の範囲を 7 方向に分解したブロック単位で判別特徴空間を生成した。学習人物 50 人の未知データに対しては平均 86%、未知人物 50 人に対して平均 84%の識別率を得ている。

各グローバルカメラの位置関係が既知の場合は、推定結果を統合することで、推定精度を向上させることができる。また、手などで顔が遮蔽されている場合でも、顔全体を適切に捉えているカメラによる推定結果を統合することで、正しい顔方向を推定することができる。

3.2.3 瞳検出

顔領域から瞳を検出する方法として、分離度フィルタを用いる方法が提案されているが、候補カ所が多く検出されるため、室内に置かれた多様なカメラアングルからの顔画像に対応するのは困難と思われる。そのため、まず、目と口の領域を抽出する。目、口などの顔部品の抽出精度を向上させるために、顔方向に応じてテンプレートを切り替える色・エッジ統合型マルチテンプレートマッチング手法を用いる。

目・口領域抽出には、4 方向面特徴と色情報を用いたテンプレートを顔方向別に作成したマルチテンプレートにより検出する。テンプレートは図 9 に示すような右目、 左目、 口のそれぞれに 4 方向面特徴のテンプレート 4 枚と、基準肌色 UV 値からの U, V 値の差のテンプレート 2 枚とで合計 6 枚ずつ作成する。 4 方向面特徴とはエッジ成分を水平方向、右上がり方向、左

上がり方向、垂直方向の４方向面に分けてから、それぞれの面に対してガウシャンフィルタをかけ、ぼかし処理を施し低解像度化したものである。色情報としてU、V値をそのままテンプレートとして持つのではなく基準肌色 *UV* 値との差をとる。これは、テンプレート作成時とシステム稼動時の環境の違いを吸収するためである。

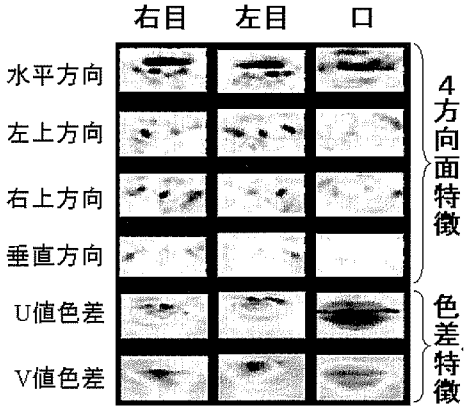


図９ 顔部品テンプレート

以上のようにして作成したテンプレートを用いてテンプレートマッチングを行う。先に求めた顔領域に対して、収縮、膨張処理を行い、ノイズを除去する。そして、ラベリング処理を行い面積が最大のものを抽出した後、膨張、収縮、穴埋め処理をする。以上の処理を施した後、その領域を対象に各顔部品の存在する可能性として評価値Sを計算する。評価値Sは４方向面特徴の相関値Eと色の情報の相関値Cの和で決定する。評価値S、相関値E、Cの計算式を(1)(2)(3)に示す。

$$E=\frac{\left\{\sum_{i=1}^{m}\sum_{j=1}^{n} I_{(x,y)}(i,j,v)\times T(i,j,v)\right\}^2}{\sum_{i=1}^{m}\sum_{j=1}^{n} I_{(x,y)}(i,j,v)^2\times\sum_{i=1}^{m}\sum_{j=1}^{n} T(i,j,v)^2} \tag{1}$$

$$C=1-\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(I_{u(x,y)}(i,j)-T_u(i,j)\right)^2+\sum_{i=1}^{m}\sum_{j=1}^{n}\left(I_{v(x,y)}(i,j)-T_v(i,j)\right)^2}{2m\cdot n\cdot U_{\max}\cdot V_{\max}} \tag{2}$$

$$S=E+C \tag{3}$$

式(1)の *I,T* はそれぞれ入力画像とテンプレートの各画素を表し、*v* はエッジの方向面（水平方向、垂直方向、右上がり方向、左上がり方向）を表す。式(2)の *u, v* はそれぞれ *U* 値 *V* 値を示し、*m ,n* はテンプレートのサイズを表す。式(3)で求められた評価値を、閾値で判別して各顔部品の候補領域を決定する。各顔部品の候補領域の全組合せに対して簡単なルールを適応し、存在し得ない組合せを排除する。例えば、右目は左目より右にある、口はどちらの目よりも下にあるといったルールを適用した。この様に制限された組合せの中から評価値の最も高い組合せを顔部品抽出結果とする。

実験では、3 方向のテンプレートを作成し、顔方向推定結果に応じてテンプレートを切り替えることで検出精度を向上させた。更に、顔画像のサイズによる解像度の正規化処理、顔輪郭部での背景部分のマスク化など改善を加えることで、目領域の検出精度を向上させた。

次に、検出された目領域から瞳を検出する。瞳の検出は視線検出への応用を考え、瞳の輪郭を円で近似し瞳の中心点を検出する。本研究では瞳の円を検出する手法としてHough変換を用いる。Hough変換の特徴として、画像情報を大局的にとらえ、パターンの一部が現れていなくても検出できることやノイズに強いことが挙げられる。瞳は個人差によらず円形で近似できること、また、瞳の一部が瞼に隠れている場合でも瞳の輪郭を円で復元し検出できると考えHough変換を用いた。しかし、Hough変換は特徴点１点につき１本のHough曲線を描くため特徴点の

数に比例して計算時間が多くなるといった欠点がある。本手法では瞳の位置を絞り込み、余分な特徴点を排除することで処理時間の短縮と精度の向上を図った。本手法による瞳の位置検出方法は以下の通りである。

抽出した目領域内で、肌の部分は彩度値が高いのに比べ、目の部分は彩度値が低いことに着目して、彩度値のヒストグラムを作成し判別分析法により２値化を行い、目の部分と肌の部分に分割する。図 10(a)は抽出した目領域であり、図 10(b)は目の部分と肌の部分に分割した結果である。次に、目の部分で白目の部分は輝度値が高いのに比べ黒目の部分は輝度値が低いことに着目し、目の部分の輝度値のヒストグラムを作成し、判別分析法により２値化を行い黒目の部分と白目の部分に分割する。図 10(c)は白目の部分と黒目の部分に分割した結果を示す。分割された黒目の形状は影の影響などにより、完全な瞳の形になりにくい。また、瞼のエッジが強く出るとき、瞳のエッジは抽出しにくくなる。そこで、本手法では分割した黒目の形状から瞳の領域を求め、入力画像の瞳領域を微分し、判別分析法により微分画像の２値化を行う。図 10(d)はエッジ抽出をした結果である。このエッジ点群を特徴点として Hough 変換により瞳の円検出を行う。

(a)抽出した目領域　(b)目と肌の分割　(c)白目、黒目の分割　(d)瞳のエッジ

図 10 瞳抽出処理

円の Hough 変換は、各特徴点に対してそれがある円の一部である可能性をパラメータ空間に投票する処理である。パラメータ空間は円を表現するのに最低限必要である円の中心座標 (a,b) と円の半径 r の３つのパラメータで構成される。パラメータ空間に投票される様子を図 5 に示す。画像上の x-y 座標に特徴点が一点ある状態を考える。その特徴点は円１、円２の一部である可能性が考えられる。このように、考えられる可能性のある円のパラメータ(a, b, r)を全てパラメータ空間に投票する。一つの特徴点につきパラメータ空間内では図 11 のようなコーン状の形で投票される。この投票処理をすべての特徴点に対して行った後、パラメータ空間内から投票数が最大のパラメータを選択する。このパラメータで表現される円が最も信頼度の高い円である。本手法ではパラメータ空間内から投票数の最大のパラメータで表される円を瞳の円として検出する。

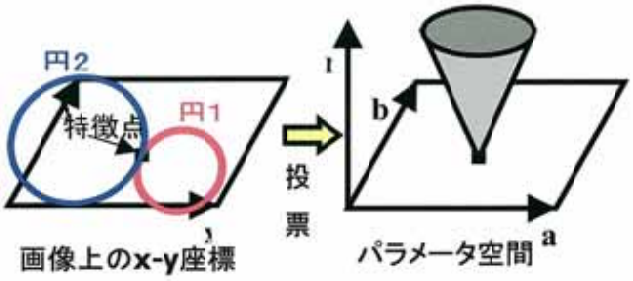


図 11　Hough 変換の処理

3.2.4 注目判定

注目判定手法は、操作対象物にカメラを備え、各対象物からの視点で得られた画像を元に最も注目している機器を選出する。これにより、注目対象物への方向を視線方向として推定する。

検出方法の概念図を図 12 に示す。瞳の形状を正円と仮定して、注目対象物から見た瞳の形状が最も円形であることから、円形度の高い結果を得たカメラに視線を向けていると判断する。つまり、人物の視線方向を直接測るのでなく、瞳の円形度を測ることで視線方向を推定することになる。

円形度は、式(4)に示すように、ハフ投票空間の投票値に基づき評価値（円形度）E を求める。

$$E = f(c_x, c_y) / \sigma_v \tag{4}$$

関数 f は円の中心点近傍の投票合計数である。σv は中心点を中心とした瞳中心近傍領域 ij の

投票値 v の分散値となる。つまり、瞳の形状が変形すると投票値がばらつき、評価値 E は低くなる。なお、本稿では中心点近傍として中心点とその４連結部分とし、中心点を 2、近傍を 1 の加重で合計した。瞳中心近傍領域 ij は 5×5 とした。

本手法により、従来、人物の視線方向をワールド座標系に変換するために必要であったカメラパラメータの校正が不要となる。また、本手法は基本的にキャリブレーションが不要であり、カメラ搭載機器を室内に増やすことで視線検出の精度向上が期待される。

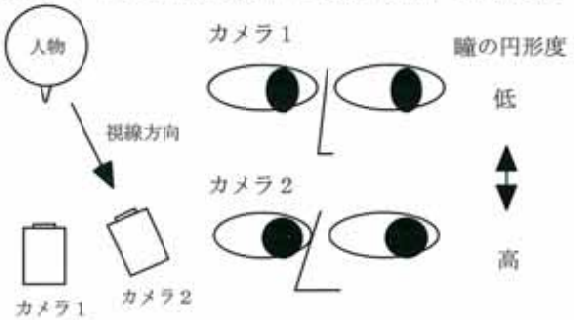


図 12　注目判定方法の概念図

4 結果

4.1 注目判定実験

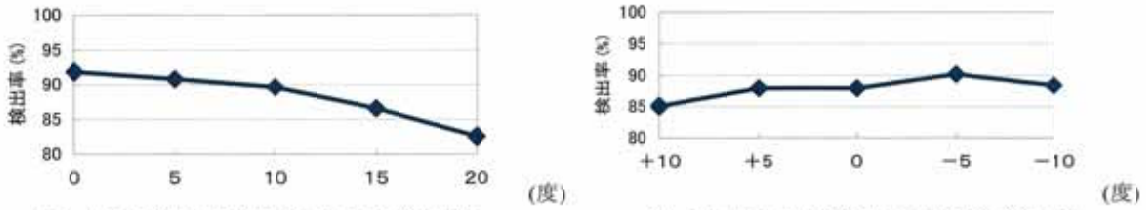
本手法による評価実験を行った。実験データは、図 13 に示す多方向顔画像収集システムを用いて収集した。中央には 15 台のカメラを 1 列 5 台ずつ 3 列に、縦横 5 度間隔で配置してあり、カメラから 3.7 m 離れた椅子に座った被験者の顔にカメラを向けて収集した。視票とカメラは全て 5 度間隔で置かれている。被験者は 13、 28、 48 の視標の方へ顔と鼻先を向ける。それぞれの方向を向いている時に、その周辺の縦横 5×5 の 25 個の視標の方を顔を動かさずに見ているときに、15 個のカメラで同時に撮影することで顔の向きが 45 方向、それぞれの方向を向いている時の視線方向は 25 方向と、合計 1125 種類の顔と視線の向きの画像を得る。

図 13 多方向顔画像収集システム

被験者は眼鏡を着用していない人 10 名、合計 11250 枚の画像を使用した。画像サイズは 640×480 である。なお、予備実験より円の Hough 変換のパラメータ空間の半径は 7～12（ピクセル）とした。

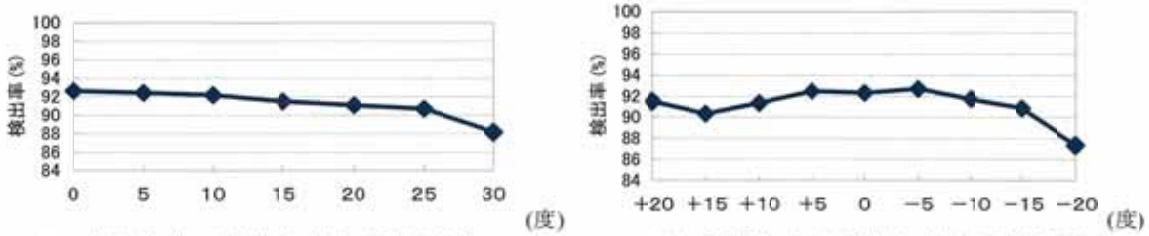
実験の結果、目領域の抽出率は 95.8%、瞳検出率は 91.7%であった。図 14(a)に顔の向きが正面から水平方向へ変化した時の瞳検出率を、図 14(b)は顔の向きが上から下へ変化した時の瞳検出率を示す。二つの結果より顔の向きが正面に近いほど検出率が良くなることがわかる。これは瞳の見え方に影響を受けるためと考えられる。

そこで、次に被験者の注視点の変化による瞳検出率を調べた。図 15(a) は注視点がカメラ側から水平方向に変化した時の瞳検出率を示す。図 15(b) は注視点が上下に変化した時の瞳検出率を示す。こちらの結果も注視点がカメラに近くなるほど良い結果となった。検出率が低下するのは図 15(a) の 30 度付近と図 15(b) の－20 度付近であった。また、図 15(b)の＋15 度付近も不安定な結果となった。



(a) 水平方向の変化に対する検出率　　(b) 垂直方向の変化に対する検出率

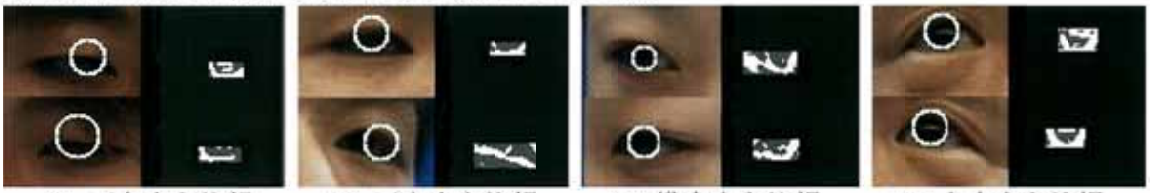
図 14　顔の向きの変化に対する瞳検出率



(a) 水平方向の変化に対する検出率　　(b) 垂直方向の変化に対する検出率

図 15　注視点の変化に対する瞳検出率

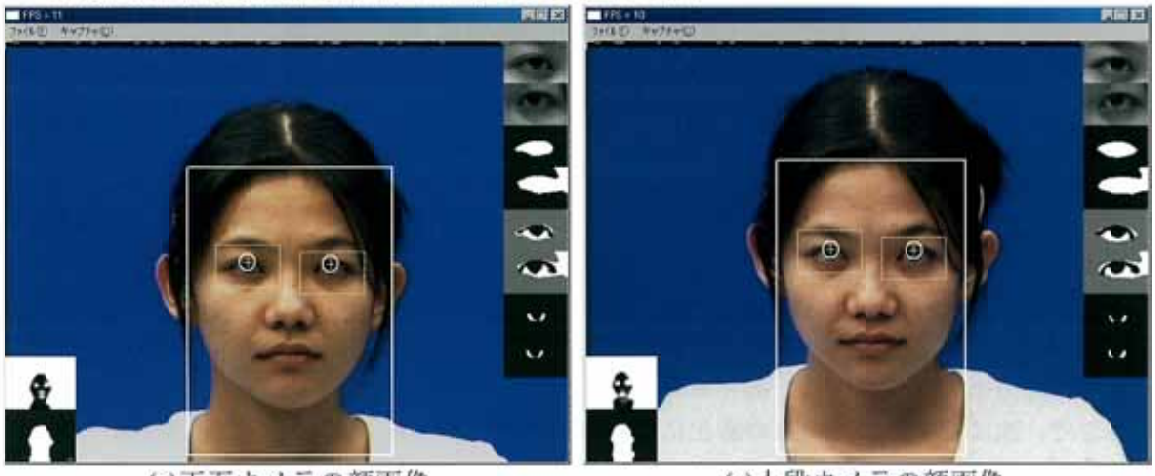
図 16(a)、 (b)は下の方を注視したときの結果例を示す。下の方を注視したとき、影の影響を受け瞳の輪郭の抽出が難しい。また、色を用いたセグメンテーションに失敗することもあった。図 16(b)は横の方向を注視した時の結果例を示す。横の方向を注視したとき、瞳の輪郭が手前側しか抽出できないことや目頭や目じりのエッジの影響を受けて検出が不安定であった。図 16(c)は上方を注視した時の結果例である。実環境ではこのように瞳に照明が写り込むことが考えられる。この場合、瞳の輪郭のエッジは抽出できるが、写り込んだ照明のエッジも抽出されてしまい、その影響で検出が不安定となった。しかしながら本手法ではカメラ側から上下方向に 15 度、左右に 25 度の範囲では安定した検出率であった。



(a) 下方向を注視　(b) 下方向を注視　(c) 横方向を注視　(d) 上方向を注視

図 16　瞳抽出実験結果例

図 17(a)は正面のカメラを注視しているとき、図 17(b)は 1 列目上段のカメラを注視しているときの、その注視したカメラより獲得した画像に対して瞳を検出した結果の一例である。検出した目領域、瞳、瞳エッジを入力画像にはめ込んでいる。瞼により瞳の一部しか見えないが、ハフ変換により瞳を適切に抽出していることを確認した。



(a)正面カメラの顔画像　　(a)上段カメラの顔画像

図 17　瞳抽出結果例

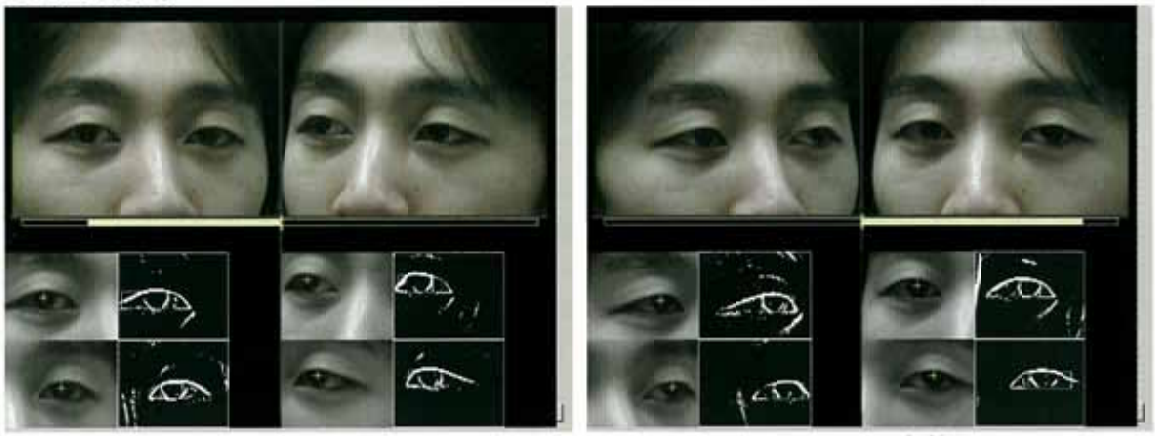
瞳検出が最も良かった被験者(99.8%)の注目度を判定した。注目の判定方法として、両目のうち円形度が高い方を注目判定に用いた。円形度があらかじめ設定した閾値以上の場合、注目し

ていると判定した。実験の結果、15 台のカメラに対して正しく注目されていると判定された正解率は 73%(11/15)、異なるカメラで注目していると誤って検出された誤検出率は約 19.6% (50/255)であった。瞳がある程度の大きさ以上で撮影されていなければ円形度に差が表れにくいが、この問題に関しては、ターゲットカメラを制御し拡大した画像を獲得することで対応する。

4.2 知的環境システムの評価

ターゲットカメラを制御して、獲得した顔画像から瞳を抽出し、円形度を評価する実験を行った。2 台のターゲットカメラ(SONY EVI-D30)の画角が約 30 度になるように設置した。距離は両方とも約2 m である。被験者には、瞳と顔方向を自由に動かせ、2 台のターゲットカメラを交互に注視させた。撮影時間は約 20 秒、約 2、3 秒間隔で交互にカメラを注視させた。但し、今回は眼部がカメラフレームからはみ出ないように顔を動かすこととした。また、同期の問題は、各カメラ画像を画面合成ユニットによる合成画像で処理することにより対処した。本来、ターゲットカメラは頭部の動きに対して拡大しながら追従する必要があるが、パンチルトカメラによる追跡は今後の課題としたい。

図 18 に抽出結果と円形度を比較した結果例を示す。上部左側がカメラ１の画像、右側がカメラ２の画像となり、下部はそれぞれの瞳抽出結果である。円形度が高い方にその比率をバーで表示した。瞳が注目しているときは、注目しているターゲットカメラの方が高い円形度を得ることができた。



(a)カメラ１を注目　　(b)カメラ２を注目

図 18　注目度比較実験

表１に瞳検出率と注目判定率を示す。収集した 92 フレーム中、5 フレームは、瞬きまたはどちらのカメラも見ていないと判断し、除外した。両方のカメラ共に４回ずつ注目されていた。しかし、カメラ１の瞳検出率がカメラ２に比べ悪かった。これは、カメラ１を見ていているある注視中に影の影響で連続して誤検出したのが原因である。注目判定率は、正しく瞳検出された結果に対して正しく注目判定が行われた場合を正解とした。本実験では、約 80%を得ることができた。

表１　注目度判定率

	フレーム数	注視回数	瞳検出率	注目度判定率
カメラ１	42	4	57.1 %	62.5 %
カメラ２	45	4	86.7 %	89.7 %
トータル	87/92	8	72.4 %	79.4 %

フェーズ II

1 研究の概要

フェーズ I では、知的室内環境パーセプトルームにおいて利用者の顔の向きと視線から操作対象を判断するシステムを構築した。しかし、顔の向きが合ったとしても利用者は必ずしもその機器を見ているとは限らないため、もう一歩踏み込んだ認識が求められる。そこで、フェーズ II では単眼カメラで撮影された利用者の顔全体画像から視線を求める研究を行った。単眼カ

メラを用いた場合、パン・チルト・ズームカメラを自動制御することで、部屋の中の利用者の顔全体を捉えた画像を撮影するシステムを構築している。ここでは、このシステムにより利用者の顔全体が撮影されていることを仮定し、目の特徴点の抽出による視線推定精度の検証を行う。

2 研究の目標

近年、人間とコンピュータとのインタフェースの重要性が高まる中、一般家庭において家電製品へのインタフェース入力面においては、家電製品上のボタン操作やリモコンでの操作に留まっている。これでは、利用者はボタンやリモコンの使用方法を理解することを余儀なくされる。また、家電製品が複数であればリモコンも複数になると考えられ、利用者は全てのリモコンの使用方法を理解し使い分ける必要がある。このように、インタフェース入力面においては、人間がコンピュータに合わせる形が一般的であり、これにより利用者に負担をかけるという問題が起こる。

このような問題に対して、人間からコンピュータへのインタフェース入力面においても利用者が直感的に使えるよう、コンピュータが人間の顔からその人の状態を把握し、操作を受け付けるという新しいインタフェースを実現しようとする研究として、パーセプトルームの構築を行っている。ここで、操作内容の認識の他に操作の対象機器の認識が重要となる。これについては、音声で操作対象を認識するシステムが提案されている。しかし、オーディオ等音を出力する機器の稼動時には人間の声と機器の音が混在し認識が難しくなる。また認識のための、集音機器の設置位置の決定が難しい。また、利用者の顔の向きから操作対象を判断するシステムが提案されている。しかし、顔の向きが合ったとしても利用者は必ずしもその機器を見ているとは限らない。そのため、もう一歩踏み込んだ認識が求められる。

そこで、ここでは単眼カメラで撮影された利用者の顔全体画像から視線を求めることを目標とする。単眼カメラを用いた場合、パン・チルト・ズーム等の制御が容易に実現できる。カメラを自動制御することで、部屋の中の利用者の顔全体を捉えた画像を撮影するシステムを構築している。このシステムにより利用者の顔全体が撮影されていることを仮定して研究を進める。この研究は、視線を求めるために必要となる目の特徴点の安定的な抽出手法の構築と、それを用いた視線検出の精度向上が目的である。

3 実施内容

3.1 目の特徴点抽出

虹彩の抽出には、フェーズIで述べた方法と同等である。

目尻、目頭の抽出には、目と肌とで大きく異なる彩度値の変化に着目した。目尻、目頭の領域付近では、彩度値の変化が顕著に現れると予想される。そこで、彩度値の変化している部分を抽出し、さらに変化の最も大きい点を目尻、目頭それぞれで求めることで、精密な特徴点抽出を行なう。

図19には目領域の画像を示す。図19(a)は目領域の入力画像であり、図19(b)は彩度値に変換した画像である。彩度値は低いほど黒く、高いほど白く表している。また、図20は目領域の彩度値の高低を等高線表示で示している。グラフのx軸、y軸は図19(b)の画像平面と同様のx-y座標を示し、z軸は彩度値Satを示す。図20で示されるように目と肌の境界では彩度値の高低差が著しく大きい。目尻と目頭の抽出では、目と肌の境界のなかでもx軸で最も左右の部分の彩度値の変化に着目する。この部分では、彩度値の高低差を描く曲線は3次曲線に近い形になる。目尻、目頭の抽出処理では、この高低差を描く曲線のなかから、特徴的である勾配の強い点を抽出する。以後、彩度値の高低差を描く曲線を彩度曲線と呼ぶ。

(a) 入力画像

(b) 彩度値画像

図 19　目領域の画像

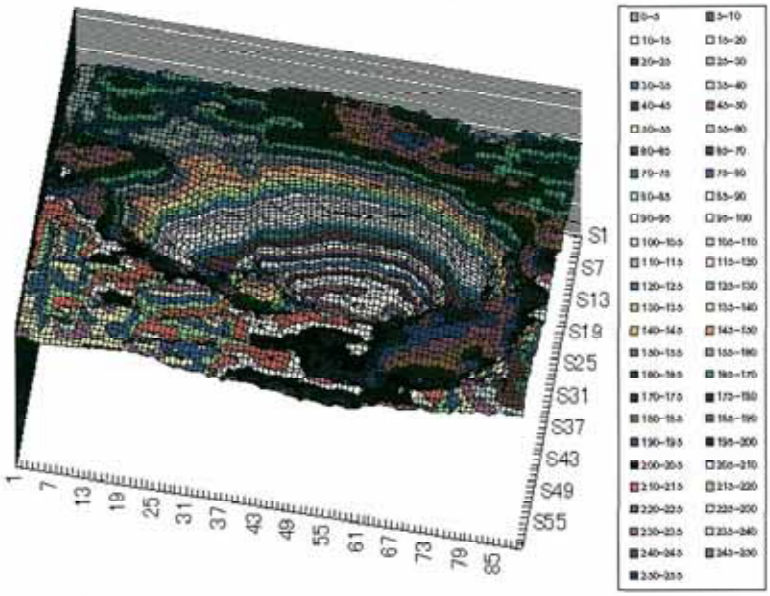


図 20　目領域の彩度値

勾配の強い点を抽出するために 2 つの処理を行なう。1 つ目は目尻、目頭付近の彩度曲線を抽出する。ここでは、図 20 に示すような x-y-Sat の 3 次元の彩度値の等高線から、勾配の強い部分の最も左右の彩度値の変化を抽出する。図 21 は、x-y-Sat の 3 次元の彩度値の等高線から勾配の強い部分の最も左右の彩度値の変化を抽出し、2 次元で表現した模式図である。中央の彩度値が低い部分は目であり、両側の彩度値の高い部分は肌である。次に彩度曲線の勾配の強い部分のみを抽出する。勾配が強い部分として、目と肌のどちらともとれる彩度値の範囲を決定し、彩度曲線からその範囲を抽出した。2 つ目は目尻、目頭付近の彩度曲線から勾配の最も強い点を決定する。本節では、これらの 2 つの方法を説明する。

最初に、x-y-Sat の 3 次元空間の情報から彩度曲線を抽出する。x-y-Sat の 3 次元空間の Sat の値を 1 つずつずらしながら x-y 平面にスライスし、現れた x-y 平面の境界線のうち、x 値で最も左右に位置する点を抽出する。これらの点は、いずれも目の両端点を示しており、目尻、目頭の候補と考えられる。彩度曲線はこれらの点から構成される。このようにして抽出された彩度曲線を図 22 に示す。なお、この彩度曲線は図 19 の目尻付近のものである。

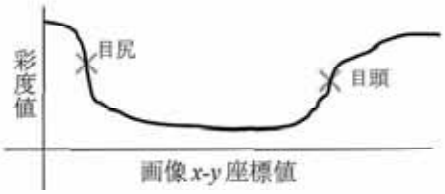


図 21　目領域彩度値の 1 次元画像

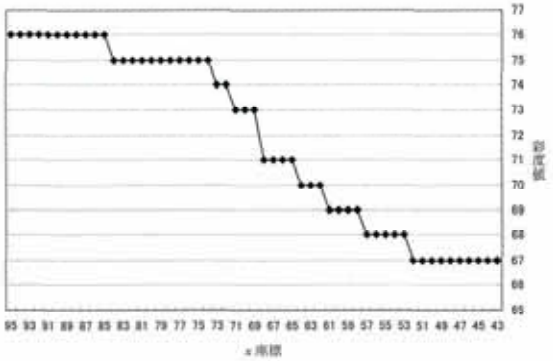


図 22　x-Sat 平面上の彩度曲線

ここで、どの範囲の彩度値でスライスするかという問題がある。スライスする範囲は彩度曲線の勾配の強い部分が、ある程度現れ、かつ最も勾配の強い点が含まれていることが必要条件となる。その彩度値の範囲は目と肌の境目の彩度値付近で、目と肌のどちらともとれる彩度値が存在する部分であると考えられる。そこで、目領域の彩度値の分布から大津の判別分析法により、統計的に目のクラスと肌のクラスに分け、両クラスを正規分布で表したときに共通する彩度値に着目する。図 23 には目領域の彩度値の正規化ヒストグラムを示し、これは彩度の確率分布とみなすことができる。図は大津の判別分析法により目のクラスと肌のクラスに分割し、各クラスに正規分布を当てはめたものである。目のクラスの正規分布と肌のクラスの正規分布の共通部分が、目と肌のどちらともとれる彩度値であると考えられる。しかし、2 つの正規分

布の重なる範囲は無限に存在するため、共通部分でも両者の境界として信頼性の高い範囲を決定する。本手法では信頼性の高い範囲 m から n を、共通部分全体に対する k 付近の面積の比により決定し、彩度曲線からその範囲を抽出する。ここで、肌である確率分布を Ps、目である確率分布を *Pe* とする。*Ps*、*Pe* は正規分布であり、式(5)、(6)で表される。

$$P_s(Sat) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left[-\frac{(Sat-\mu_s)^2}{2{\sigma_s}^2} \right] \tag{5}$$

ただし、σ_s、${\sigma_s}^2$、μ_s はそれぞれ肌のクラスの標準偏差、分散、平均を示す。

$$P_e(Sat) = \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left[-\frac{(Sat-\mu_e)^2}{2{\sigma_e}^2} \right] \tag{6}$$

ただし、σ_e、${\sigma_e}^2$、μ_e はそれぞれ目のクラスの標準偏差、分散、平均を示す。

今、スライスする範囲を m から n、共通部分で最大値となる彩度値を *k* とすると、共通部分の全面積 Sa とスライスする範囲の面積 *Sp* は式(7)、(8)で表される。

$$S_a = \sum_{t=-\infty}^{k} P_s(t) + \sum_{t=k}^{\infty} P_e(t) \tag{7}$$

$$S_p = \sum_{t=m}^{k} P_s(t) + \sum_{t=k}^{n} P_e(t) \tag{8}$$

このときスライスする範囲の信頼性 P は面積の比率で式(9)のように評価される。

$$P = S_p / S_a \tag{9}$$

P が大きいほど、その範囲に最も勾配の強い点が含まれる確率は高いが、余分なデータ点が多く含まれる。また *P* が小さいほど、余分なデータ点は少なくなるが、勾配の最も強い点が含まれる確率が低くなる。本研究では、*P* は実験的に 50%と定めた。

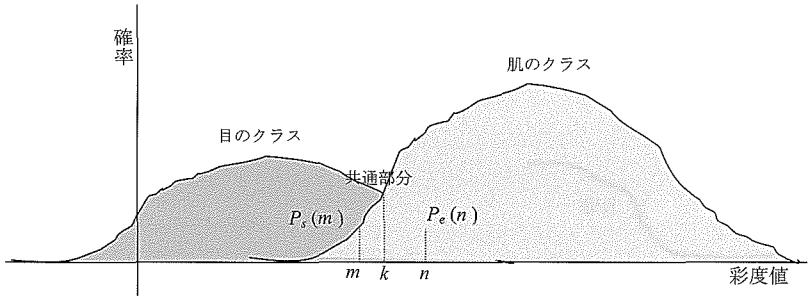
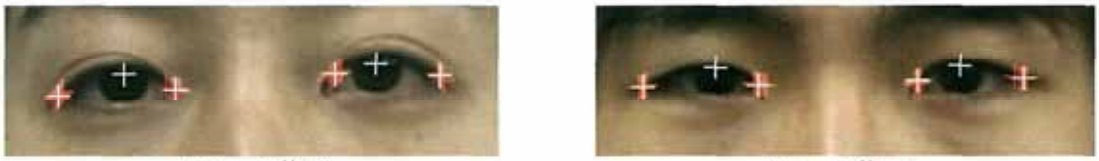


図 23　目領域の彩度値のヒストグラム

抽出された彩度曲線上の点群は、勾配の強い部分程密となり、その分布は勾配の強い点をピークとした正規分布に近い形になると考えられる。そこで、分布の中央値をとることで外れたデータ点の影響を抑えながら、勾配が最も強い点を抽出する。彩度曲線を構成する、目尻、目頭の彩度曲線上の点と、抽出された特徴点の結果を図 24 に示す。彩度曲線上の点は赤色の十字で示しており、多く現れるほど濃く表示している。白色の十字は抽出した各特徴点である。余分なデータ点が候補として抽出された場合にも、中央値をとることで特徴点を安定して抽出している。



(a)　人物A　　(b)　人物B

図 24　彩度曲線上の点と抽出点

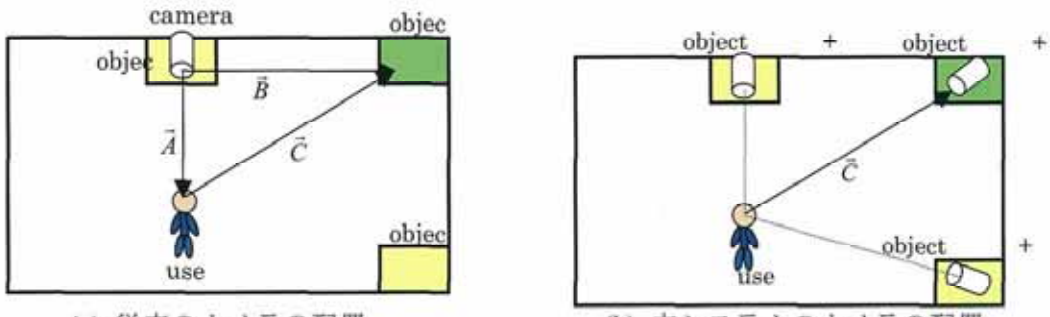
3.2　視線インタフェースシステム

現在までに視線方向を計測するシステムが研究されている．これらのシステムの多くは図25(a)に示すカメラ配置のものが多い．図 25(a)では、ユーザが注視するオブジェクトを決定するために□ユーザがどの位置から、□どの方向を注視しているか、□注視する方向にどのオブジェクトがあるか、という 3 つの情報が必要となる．すなわち、図 25 (a) の $\vec{A}$ (カメラからユーザへ向かうベクトル)、$\vec{B}$ (カメラからオブジェクトへ向かうベクトル)、$\vec{C}$ (ユーザからオブジェクトへ向かうベクトル)を求める問題に置き換えられる．そして 3 つのベクトルが式(10)を満たすときがオブジェクトを注視している状態となる．

$$\vec{A}+\vec{C}=\vec{B} \tag{10}$$

一方、本システムではカメラの配置に工夫をして、全く違った方法でユーザが注視するオブジェクトを推定する．本システムのカメラの配置を図 25 (b) に示す．図 1 (b) では、ユーザが注視する対象である各オブジェクトに 1 台ずつカメラが搭載されている．そして各々のカメラが、ユーザに注視されているかを判断することでユーザが注視するオブジェクトを決定する．図 25 (b) 中の $\vec{C}$ (ユーザからオブジェクトへ向かうベクトル) を求めることでユーザが注視するオブジェクトを決定できる．いずれかのカメラから検出される $\vec{C}$ が式 (11) を満たすとき、ユーザはそのオブジェクトを注視している状態となる．

$$\vec{C}=0 \tag{11}$$



(a) 従来のカメラの配置　　(b) 本システムのカメラの配置

図 25 空間内のカメラの配置

このようなカメラ配置にすることで、以下の 3 つのメリットが考えられる．

1 つめに、本システムでは $\vec{A}$ を精密に計測しなくてもよい点があげられる．カメラとユーザ間の距離が大きくなるほど、$|\vec{A}|$ を精密に計測することは難しくなる．そのため従来の研究では、ユーザの頭部を固定することで $|\vec{A}|$ の値を一定にしたり、ユーザが移動できる範囲は $|\vec{A}|$ の値を計測しやすい範囲に限定することがなされてきた．しかし、部屋の中での移動範囲を制限することはユーザにとって大きな負担となる．本システムでは、カメラとユーザ間の距離が大きい場合でも、人物追跡手法によりカメラのパン、チルト、ズーム制御を行ないユーザの顔画像を撮影する．また、$|\vec{A}|$ を精密に計測する必要がないため、カメラとユーザの距離計測の問題が解消される．

2 つめに、本システムでは $\vec{B}$ が不要になることがあげられる．従来システムでは、ユーザが注視する全てのオブジェクトの位置情報をあらかじめ調べて登録しておく必要がある．そのため、オブジェクトを追加、移動する度に位置情報を再登録しなくてはならない．本システムでは、$\vec{B}$ を検出しないためオブジェクトを自由に追加、移動させることが可能となる．

そして 3 つめに、本システムでは検出する $\vec{C}$ の範囲が小さいことがあげられる．従来のカメ

ラの配置では全てのオブジェクトの位置情報に対して条件式（10）を確認するため、$\bar{C}$は 360 度の範囲で検出する必要があった．そのため 360 度どこからでもユーザの顔を撮影できるようにカメラの配置をしなくてはならない．本システムでは、ユーザがオブジェクトの方を注視するときには必ずカメラの方を向いた顔が撮影されるため、カメラとオブジェクトの位置関係に制約はなく自由に配置することが可能である．

3.3 キャリブレーション

ユーザは視標を見ながらボタンを押して自分の顔画像を撮影するという作業を 2 通り行なうことでキャリブレーションを行なう．キャリブレーションを行なうことで個人差による目の大きさや形状などの違いを吸収する．キャリブレーションの方法を以下に述べる．

キャリブレーションを行なうときの環境を図 26 に示す．図中の円は眼球を表す．これはユーザの位置と眼球の状態を表す．ユーザから距離 D_1 離れた位置にはカメラが配置されている．カメラを視標 A とし、そこから D_2 離れたところに視標 B を配置する．ユーザはカメラの方に顔を向け、そのまま視標 A と B をそれぞれ注視する．そして、それぞれのタイミングで録画のボタンを押し、顔を撮影する．

こうして撮影されたそれぞれの顔画像から特徴点抽出処理により目尻、目頭、瞳の中心の 3 点の特徴点を抽出する．これらの特徴点の位置のパラメータと D_1、D_2 から式(12)により計算される□$_c$をキャリブレーションのパラメータとして視線検出に用いる．

$$\theta_c = \tan(D_1/D_2) \tag{12}$$

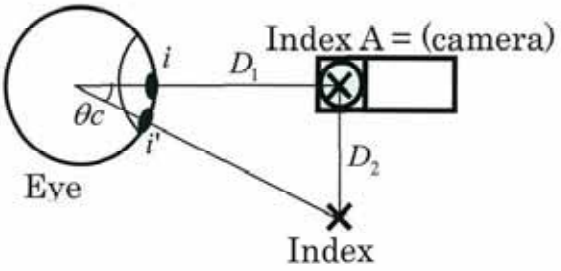


図 26 キャリブレーションの環境

3.3 視線検出処理

本システムでは、顔がカメラの方に向いたことを検出した後、視線検出を行なうので顔の向きは正面向きに限定される．そのため視線の計算方法が非常に簡単になる．本手法では、キャリブレーションにより得られた特徴点の位置と□$_c$のパラメータと入力画像より得られた特徴点の位置のパラメータを用いて視線を検出する．

ある時刻 t に撮影した入力画像の目の状態を図 27 (a) に示す．それぞれの×印は目尻、目頭、瞳の中心点の各特徴点の位置を表し、瞳の中心点を P_t とする．図 27 (b) にはキャリブレーションにより得られた各特徴点 C_l、C_r、I、I'の位置を示す．C_l、C_r は目尻、目頭の位置、I はカメラの方向を、I'は□$_c$度の視票を注視した状態の瞳の中心点を表す．これら(a)と(b)では、撮影された時のユーザとカメラの間の距離が異なることが予想される．そのため大きさと位置の正規化を行なう．

(a) 入力の状態　　(b) キャリブレーション

図 27 目の特徴点の状態

まず最初に大きさの正規化を行なう．(a)と(b)では、それぞれ目尻と目頭の点が検出されている．この目尻と目頭の 2 点間の距離を目の大きさとして、(a)と(b)の目の大きさの比率をもとめる．(b)に対する(a)の目の大きさを R とする．ここで、キャリブレーション時の特徴点 C_l、C_r、I の位置を図 28 に示す．図 28 は点 C_l を原点とした x-y 座標である．図 28 では瞳の中心点は C_i とする．C_l を始点として C_i と C_r へ向かうベクトルに対し、先にもとめた比率 R をかけることで C'_i、C'_r を算出する．こうして、(b)の状態を(a)の状態と同じ大きさに正規化した後の点 C_l、

C_i、C'_rを得る．

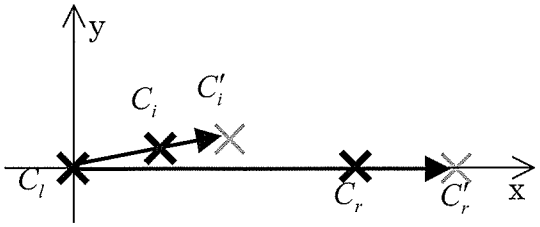


図 28 目尻の点を中心とした座標

次に位置の正規化を行なう．位置の正規化は(a)と同じ大きさに正規化した後の(b)と(a)のそれぞれの目尻と目頭の点を結ぶ線分の中点を合わせることで行なう．こうして(a)の大きさと位置に正規化された後のパラメータを用いて視線方向を検出する．

正規化された後の画像上での瞳の点と眼球の関係を、1 次元の平面上に記したものを図 29 に示す．図中の円は眼球を、下方にある直線は画像中の座標を表す．入力画像の瞳の位置を P_t、図 27(b)の I、I'に対して大きさと位置の正規化をした後の点を I_t、I'_tとする．これら 3 点は画像上で計測された位置を表し、そのときの眼球上の瞳の位置は p_t、i_t、i'_tで表される．また、$\square_c$はキャリブレーションの角度であり、□はもとめる角度である．ここで、画像上で計測できる値 M、L と$\square_c$は既知であるため、もとめる角度□は式(13)により計算される．

$$\theta = \sin^{-1}\left(\frac{M}{L}\sin\theta_c\right) \tag{13}$$

式(13)により縦方向成分の角度と横方向成分の角度をそれぞれ計算してから、オブジェクトを注視した状態との視線のずれ量を検出する．

なお、本研究では眼球は球であると仮定し、球の中心を回転中心とした．そのため、視線方向の垂直成分、水平成分は式(13)によりそれぞれもとめる．また、ユーザの視線ベクトルの大きさは考えないため、左右の目の視線方向を平均することでユーザの視線を検出する．

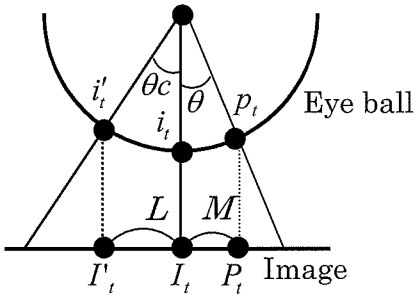


図 29 視線計測方法

4 結果

4.1 実験データ

様々な種類の視線のデータを撮影して視線検出の実験を行なった。被験者には正面に設置したカメラの方を向かせたままヘッドレストを用いて頭部が動かないように固定した。そして、その状態のままカメラの周辺に配置した視標をそれぞれ注視させた。視標の配置を図 30 に示す。視標は 49 個あり、中央の 25 番の視標はカメラのレンズの中心である。各視標は水平、垂直に 3 度間隔の格子状で、中央から水平、垂直方向±9 度の範囲に配置されている。被験者は 1 名、視線の方向は 49 種類、各 100 フレームで合計 4900 枚の画像を撮影した。入力画像の解像度は 640×480、瞳の半径は約 13pixel であった。

キャリブレーションは被験者と視標との距離を 60cm とし、正面と左 25 度の位置の 2 箇所に視標を配置し、それぞれを注視させた。各視標を注視した画像から任意に 10 フレームずつ選択し、それらの特徴点抽出結果の平均値から個人パラメータを獲得した。

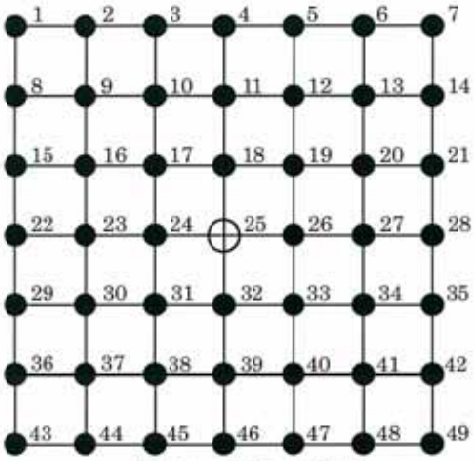


図 30　視標の配置

4.2　実験結果と考察

図 31 に各視線データの視線検出結果を示す。被験者からカメラへの方向を原点とし、x 軸は、正の値が被験者から見て右側を、負の値が左側をとしており、y 軸は、正の値が被験者から見て上方を、負の値が下方を示す。各視線データ 100 フレームの結果は平均されている。図 32 には特徴点抽出の結果を示す。白色の十字は各特徴点を抽出した結果を示し、円は虹彩を検出した結果である。また、緑色の十字は個人パラメータより求めた、正面を注視した場合の虹彩中心点を示す。もし、虹彩の点が緑色の十字より上にあれば、そのとき被験者は情報を注視していることを示す。表 2 には各視線データの水平成分、垂直成分別の理想値との誤差を示す。例えば水平+9 の値は視標番号 7、14、21、28、35、42、49 の視標を注視した視線データの視線検出結果の水平成分に着目しており、垂直+9 値は視標番号 1、2、3、4、5、6、7 の視標を注視した視線データの視線検出結果の垂直成分に着目している。

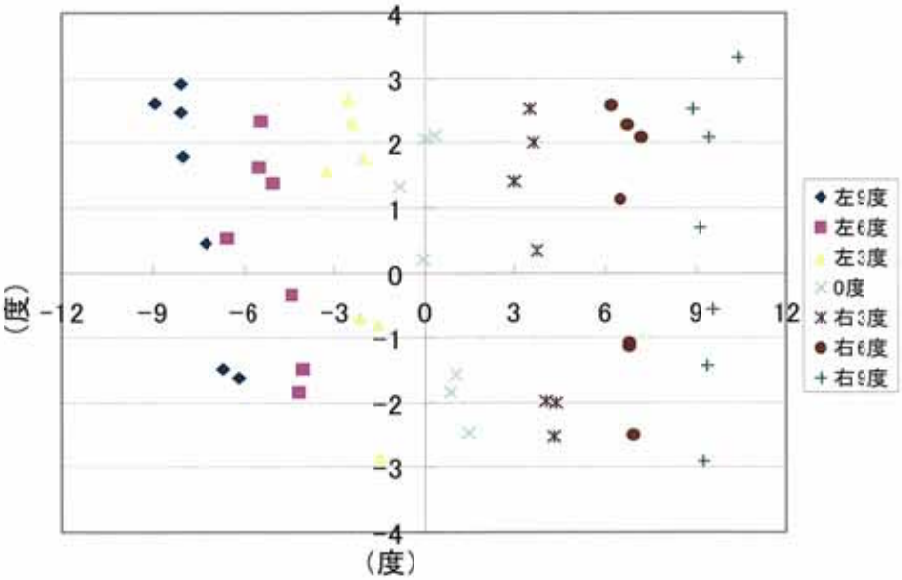


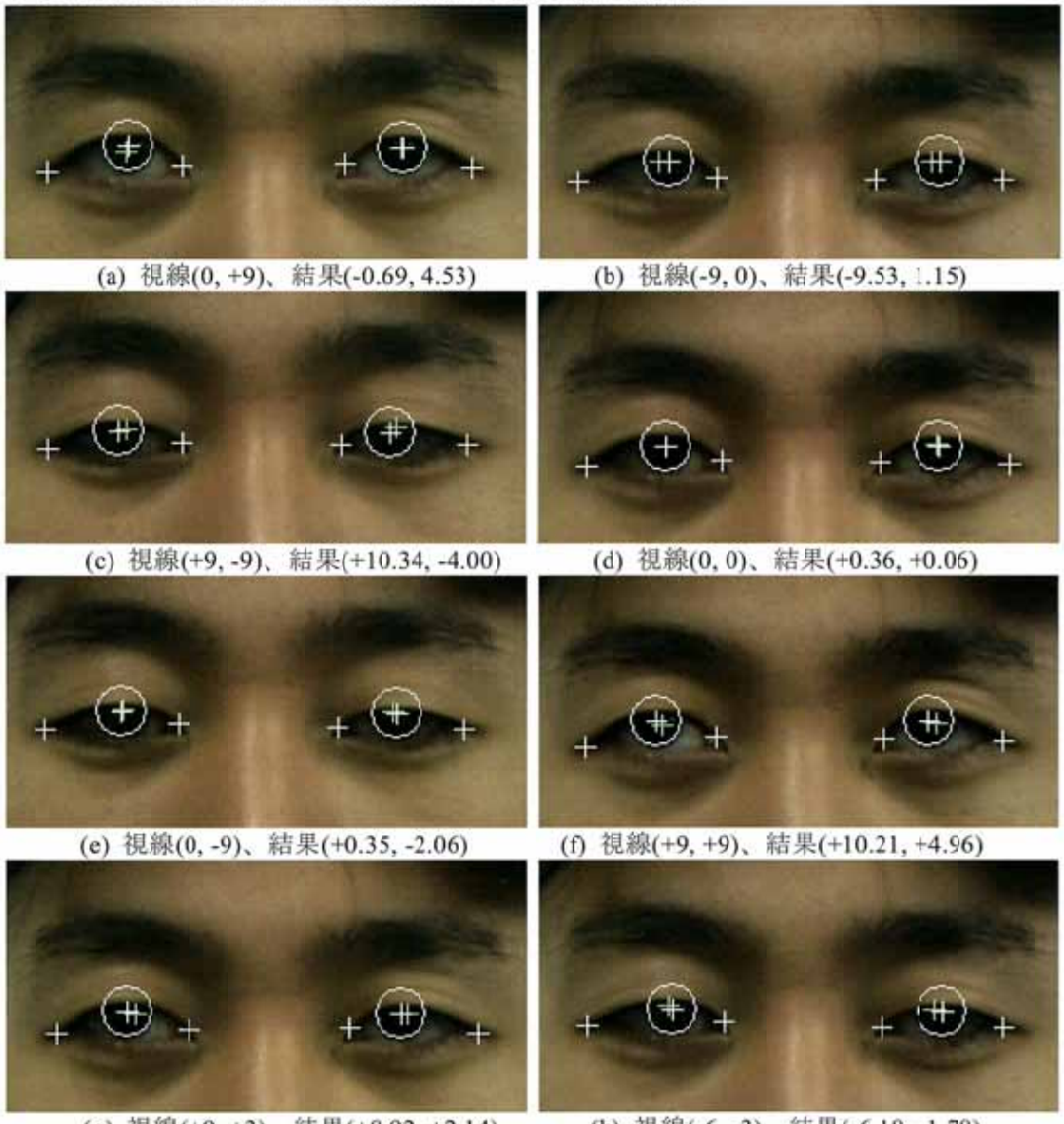
図 31　視線検出の結果

表 2　水平、垂直成分別の誤差（度）

	+9	+6	+3	0	-3	-6	-9	平均
水平	0.46	0.71	0.78	0.66	0.86	1.14	1.41	0.86
垂直	7.16	3.56	0.84	1.08	2.12	4.54	6.66	3.71

図 31 より、検出結果は理想値との誤差がまばらに現れていたが、3 度間隔と密な視線データでありながら各視線の違いを検出した。表 2 より、視線の水平成分の誤差は平均で 0.86 度と安

定した結果を得た。垂直成分の誤差は平均で 3.71 度であり、視線が上下になる程誤差が大きくなった。虹彩のエッジの上下部分が瞼に隠れたため、円の上下位置の検出精度が下がったと考えられる。虹彩のエッジ抽出方法の改善など、虹彩の円の上下位置の検出精度を向上させることは今後の課題になると考えられる。しかしながら、水平、垂直成分共に正面付近では誤差は小さかった。注視しているかどうかを推定するためには、正面付近の視線が重要となるため、本手法は注視判定をするには有効な手法であると考えられる。



(a) 視線(0, +9)、結果(-0.69, 4.53)　(b) 視線(-9, 0)、結果(-9.53, 1.15)

(c) 視線(+9, -9)、結果(+10.34, -4.00)　(d) 視線(0, 0)、結果(+0.36, +0.06)

(e) 視線(0, -9)、結果(+0.35, -2.06)　(f) 視線(+9, +9)、結果(+10.21, +4.96)

(g) 視線(+9, +3)、結果(+8.92, +2.14)　(h) 視線(-6, -3)、結果(-6.18, -1.78)

図 32　特徴点抽出結果

4.3　注視判定に向けた実験

利用者がカメラの方を注視しているかどうかを推定することは、利用者の視線が、ある角度以内であるかどうかを推定することに等しい。そこで、正面を注視した視線データとそれ以外の視線データで境界となる角度を定め、その角度以内の視線が検出されたデータ数を比較する実験を行なった。実験データとして前節の 49 種類の視線データを使用した。49 種類の視線データの視線方向を図 5.5 に示す。各点の番号は視線番号であり、それぞれ 3 度間隔である。例えば 25 番は正面を注視したデータであり、1 番は左へ 9 度、上へ 9 度を注視したデータである。各視線データは 100 フレームである。前章で示したように本システムでは 3～4 度の精度を目標としているため、本実験では図 33 の円で示されるような 3 度以内と検出されたデータを正面を注視したデータとしてカウントした。実際に 3 度以内を注視しているデータは視標番号 25 番を注視した 100 フレームのデータであり、これを Data1 と定義する。そして、3 度以上を注視し

たデータは視標番号 18、24、25、26、32 番以外を注視した 4400 フレームであり、これを Data2 と定義する。Data1 で視線検出結果が 3 度以内に検出された検出成功率と、Data2 で視線検出結果が 3 度以上に検出されたリジェクト成功率を表 3 に示す。

結果より、正面を注視したデータ Data1 の視線検出結果では、100 フレーム中 94 フレームが 3 度以内に検出された。また、3 度以上を注視したデータ Data2 の視線検出結果では、86%が 3 度以上と検出された。以上の結果より、本システムでは高い精度で注視判定が可能であることが示された。実際にシステムを稼動させるときには、数フレームの結果から最終的な判断を下すことでさらに信頼性が向上されると考えられる。

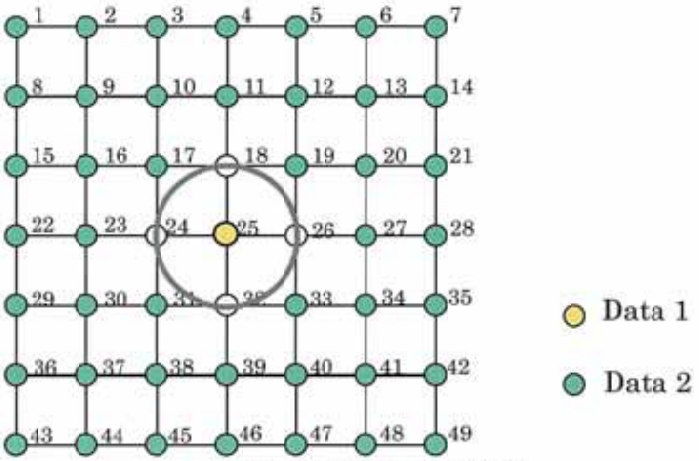


図 33　視線データの種類

表 3 検出成功率とリジェクト成功率

	成功率　(成功数)
Data 1	94% (94 / 100)
Data 2	86.0% (3782 / 4400)

フェーズ III

今後の取り組み

これまでの研究において、室内に固定したグローバルカメラと対象物に搭載したターゲットカメラを協調して、注目対象物を推定する知的環境を構築した。対象物に備えたカメラ画像から瞳の円形度を求め、注目を検出することで注目対象物を推定することができる。また、撮影された顔画像から目尻、目頭、瞳の中心点の 3 点の特徴点を抽出し、視線方向を検出することができる。

今後の課題としては、頭部の動きに対する目領域の拡大画像獲得のためのターゲットカメラの制御、瞳の検出精度およびロバスト性の向上が挙げられる。最終的には、実際のリビングへの応用に取り組む計画である。さらに、マーケティング調査のような応用展開においては、顧客の興味を計るための注目推定が重要となると考えられる。