

受動型ステレオビジョンを用いた指差し認識に関する検討 A Study on Pointing Gesture Recognition Using Passive Stereo Vision

栗山 貴好[†] 宮澤 一之[†] 青木 孝文[†]
Takayoshi Kuriyama Kazuyuki Miyazawa Takafumi Aoki

1. まえがき

ジェスチャは、文字や音声などと同じく人間同士の有効なコミュニケーション手段の一つである。人間とコンピュータとのコミュニケーションにおいても、近年ジェスチャの有効性が指摘されている。その中でも指差しジェスチャは最もよく使われるジェスチャの一つであり、コンピュータに指差しを理解させることで人間とコンピュータとの新しいインターフェースを開発する試みが数多くなされている [1]-[5]。

指差しジェスチャが多用される状況の例として、ディスプレイを用いた会議が挙げられる。近年、会議では大型のディスプレイなどに資料を映し、これを見ながら議論を行うという形式が一般的になっている。このような状況においては、レーザポインタなどを用いてディスプレイ上の位置を指示することが頻繁に行われる。参加者の指差しジェスチャを認識し、指示位置をディスプレイ上に提示するようなシステムが構築できれば、レーザポインタを用いなくともより簡便にディスプレイ上の位置を指示することができるようになる。

現在までに様々な指差し認識アルゴリズムが提案されており、その一つとしてデータグローブなどの接触型センサを利用する手法がある [6]。この手法では、データグローブを装着することにより、手の3次元位置に加えて指の曲がり具合などの情報が得られ、指差しに限らず多くのハンドジェスチャの認識が可能になる。しかしながら、この手法ではデータグローブを装着する煩わしさや、利用者に不自由を強いてしまうという欠点がある。

これに対して、赤外線センサなどの非接触型センサ [7]-[9] を用いる方法も提案されている。これらのセンサから得られる3次元情報を利用して高精度に指差し認識を行うことができるが、上記のグローブと同様に高価なデバイスを用意する必要がある。

これらのように特殊なデバイスを用いる手法に対して、カメラなどの比較的簡便なデバイスを用いて指差し認識を行う手法も数多く提案されている。一つのカメラのみを利用する手法としては、利用者に指サックを着用してもらい、その色情報から指先を検出する手法や、肌色情報から指先を検出する手法がある [1], [5]。しかし、一般的な室内において色情報のみで指先を高精度に検出することは難しく、安定性を向上させるために多くの複雑な処理が必要となる。

また、複数のカメラを用いることで対象空間の3次元情報を取得し、これを利用して指差し認識を行う手法も提案されている [2], [4]。これらの手法は、利用者が立っていることを想定した一般的な室内での指差し

による室内空間制御を目的としており、部屋全体をカバーするためにきわめて多くのカメラを必要とする。また、利用者の位置を特定するために背景差分などを用いており、事前に背景を取得しておく必要がある。

これらの手法に対し、本論文では、会議などディスプレイへの指差しが行われる状況を想定し、簡便な一台のステレオカメラのみで利用者がディスプレイのどこを指しているのかを認識するための手法を提案する。利用者がディスプレイを指す場合、その利用者は必ずディスプレイの方向を向いている。これを利用し、提案手法ではディスプレイの上部にステレオカメラを取り付け、得られた画像に対して顔検出を適用することで利用者の位置を特定する。そして、ステレオ画像の対応付けによって対象空間の3次元情報を取得し、利用者の指先を検出することでディスプレイのどこが指されているのかを高精度に推定する。本論文で利用するステレオカメラは基線長を短くしているため非常にコンパクトであり、ディスプレイへの取り付けも容易である。基線長の短いステレオカメラを用いる場合は対応付け誤差が3次元計測精度に大きな影響を及ぼすことが問題となるが、提案手法では位相限定相関法 (Phase-Only Correlation: POC) と呼ばれる画像マッチング手法を利用することでサブピクセル精度でのステレオ対応付けを行い、高精度な3次元情報を取得している。POCに基づくステレオ対応付けは、フーリエ変換を多用するためその計算時間が問題となるが、提案手法ではディスプレイと利用者の位置関係および色情報を利用して画像中で対応点探索を行う領域を削減し、処理時間の大幅な短縮を達成している。本論文では、実際にディスプレイとステレオカメラを組み合わせた指差し認識システムを構築し、実験を通して利用者が複数の場合でも提案手法によって高精度かつリアルタイムに指差しを認識できることを示す。

2. 位相限定相関法に基づくステレオ画像の高精度対応付け手法

筆者らは、位相限定相関法と呼ばれる高精度画像マッチング手法を用いた受動型ステレオビジョンシステムを構築してきた [10]。平行化されたステレオ画像ペアのように、画像ブロック間の移動ベクトルが1次元方向に制限されている場合は、1次元のPOC関数を利用することが可能となり、大幅に計算量を削減することができる [11]。本章では、1次元POCに基づくステレオ画像の対応付け技術について簡潔に述べる。詳細は [11] を参照されたい。

[†]東北大学 大学院情報科学研究科, GSIS, Tohoku University

2.1 位相限定相関関数

N 点の二つの 1 次元画像信号 $f(n)$ 及び $g(n)$ が与えられたとする．ここで，1 次元画像信号の離散時間インデックス (整数) を，便宜上， $n = -M, \dots, M$ とする．ただし， M は正の整数であり， $N = 2M + 1$ である．

1 次元画像信号 $f(n)$ 及び $g(n)$ の 1 次元離散フーリエ変換 (Discrete Fourier Transform: DFT) を次式で定義する．

$$\begin{aligned} F(k) &= \sum_n f(n) W_N^{kn} \\ &= A_F(k) e^{j\theta_F(k)} \end{aligned} \quad (1)$$

$$\begin{aligned} G(k) &= \sum_n g(n) W_N^{kn} \\ &= A_G(k) e^{j\theta_G(k)} \end{aligned} \quad (2)$$

ここで， $k = -M, \dots, M$ は離散周波数インデックス (整数) であり，回転因子を $W_N = e^{-j\frac{2\pi}{N}}$ と定義する． $A_F(k)$ 及び $A_G(k)$ は振幅スペクトルであり， $\theta_F(k)$ 及び $\theta_G(k)$ は位相スペクトルを表す．また， \sum_n は，インデックス全域にわたる加算 \sum_n^M を表す．このとき， $F(k)$ と $G(k)$ の正規化相互パワースペクトルを次式で与える．

$$R(k) = \frac{F(k)\overline{G(k)}}{|F(k)\overline{G(k)}|} \quad (3)$$

ここで， $\overline{G(k)}$ は $G(k)$ の複素共役を表す．また， $\theta_F(k) - \theta_G(k)$ は，二つの 1 次元画像信号の位相差スペクトルであることに注意されたい．画像マッチングにおいて，この位相差スペクトルは重要な性質を有しているが，これを直接的に利用するよりは，次のような相関関数を定義する方が便利である．

1 次元 POC 関数 $r(n)$ を，正規化相互パワースペクトルの 1 次元逆離散フーリエ変換 (Inverse Discrete Fourier Transform: IDFT) として定義する．

$$r(n) = \frac{1}{N} \sum_k R(k) W_N^{-kn} \quad (4)$$

二つの信号が類似している場合，POC 関数は，デルタ関数に近い極めて鋭いピークを有する．この相関ピークの高さは二つの信号の位相差スペクトルが周波数に対して線形であれば，相関ピークの高さは 1 となる．この相関ピークの高さは信号の類似度の尺度として有用であり，「画像照合」の用途で用いられる．一方，相関ピークの座標は二つの信号の相対的な位置ずれに対応し，主として「画像レジストレーション」の応用で重要になる．

以下では，1 次元画像信号 $g(n)$ が 1 次元画像信号 $f(n)$ を δ だけ微小に平行移動させた 1 次元画像信号である場合を考える．ここで， δ は n 方向のサブピクセルレベルの移動量を表している．このとき， $f(n)$ と $g(n)$ の POC 関数は次式で与えられる．

$$r(n) \approx \frac{\alpha \sin(\pi(n + \delta))}{N \sin(\frac{\pi}{N}(n + \delta))} \quad (5)$$

ここで， $\alpha = 1$ である．上式は，1 次元画像信号が δ だけ微小に平行移動した場合の POC 関数の一般形を表している． α は，相関ピークの高さを表現するために導入されたパラメータである．画像に対して，無相関なノイズが加わると α の値が減少するため，実際には $\alpha \leq 1$ となる．この相関ピークのモデルに基づく関数フィッティングにより，パラメータ α, δ を推定することで，信号の類似度 (位相差スペクトルの線形性) とサブピクセル精度の移動量を求めることができる．

2.2 ステレオ画像マッチングへの応用

以下では，POC 関数を用いて，平行移動量の推定を行う際に重要になる各種の高精度化手法について述べる．

2.2.1 窓関数の適用

1 次元 DFT では，信号が周期的に循環することを仮定するため，端点での不連続性が問題となる．この不連続性の影響を軽減するため，1 次元画像信号に対して窓関数を適用することが重要である．本論文ではハニング窓を適用する．

2.2.2 スペクトルの重み付け

一般に，自然画像のエネルギーは低周波領域に集中し，高周波成分のエネルギーは相対的に小さいことが知られている．このため，エイリアシング，ぼけ，雑音，ひずみなどの外乱が加わると，高周波成分の SN 比が大幅に劣化する．そこで，信頼性の低い高周波成分の影響を抑制するために，正規化相互パワースペクトルの計算の際に，低域型のスペクトル重み付け関数を適用することにより，大幅な精度向上が可能である．3 次元計測などの応用では，ガウス関数などがよく用いられる．この場合，相関ピークモデルもガウス形になる．

2.2.3 相関ピークモデルのフィッティング

一般に，移動量 (δ) は実数値であり，POC 関数のピーク座標がサンプリング格子点の間に存在するため，そのままでは正確に移動量を推定することが困難である．そこで，相関ピークのモデルが式 (5) で与えられることを考慮し，実際に計算された POC 関数の数値データに対して本モデルをフィッティングすることで，画像のピクセル間に存在するピークの位置を推定する．上記 2.2.2 のスペクトル重み付け関数を適用した場合，その重み付け関数に応じた相関ピークモデルをフィッティングする必要がある，このとき α, δ がフィッティングパラメータとなる．今回は，フィッティング手法として Levenberg-Marquardt 法を用いた．

このように，POC 関数の相関ピークモデルは，入力画像によらないため，関数フィッティングによって高精度な移動量推定が可能である．

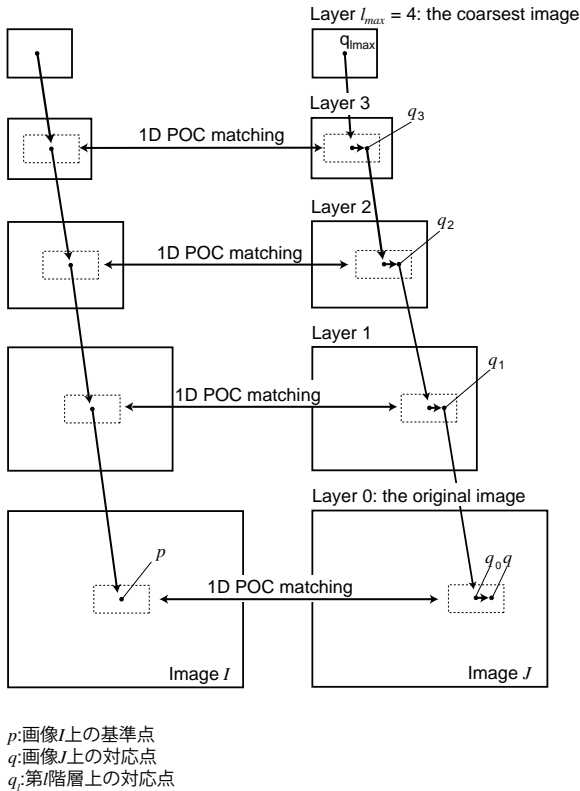


図 1: 画像ピラミッドを用いた粗密探索

2.2.4 サブピクセル対応点探索

上記 2.2.1 ~ 2.2.3 の平行移動量推定手法を局所的な画像ブロックマッチングに適用し、画像ピラミッドによる粗密探索と組み合わせることにより、ステレオ画像の対応点をサブピクセル精度で検出できる。この手法は、画質や輝度値の変化に対してロバストであり、ステレオ画像のサブピクセル対応付けに有効である。画像ピラミッドを用いた粗密探索の様子を図 1 に示す。

3. 指差し認識アルゴリズム

本章では、指差し認識アルゴリズムについて述べる。提案手法では図 2 で示されるように、ディスプレイ上部にステレオカメラを取り付け、3次元情報を利用して利用者の位置と指先の位置を求める。また、提案手法では事前にディスプレイ四隅の3次元座標を求めておく必要がある。図 3 のようにディスプレイの各コーナーからマーカーをつけたロープを二回以上引き、マーカーの3次元座標から直線の方程式をそれぞれ求め、各直線の交点としてディスプレイのコーナーの座標を求める。提案手法は、視点の検出、3次元復元、手の抽出、外れ点除去、指差し位置推定の5つの処理から構成されている。以下に各処理の詳細について述べる。

3.1 視点の検出

本論文では、ディスプレイ上の指し位置を利用者の視点から指先へ向かう直線 (図 4) と、ディスプレイ平面との交点と定義した。そのため、利用者の視点の3次元座標が必要となる。提案手法では視点の座標を得るために顔検出を利用する。

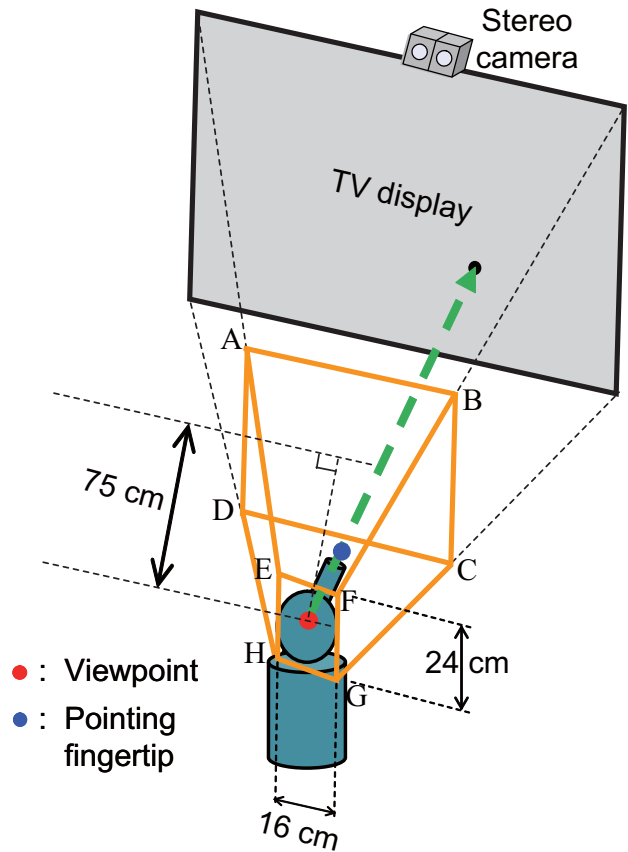


図 2: 復元領域を限定するための六面体

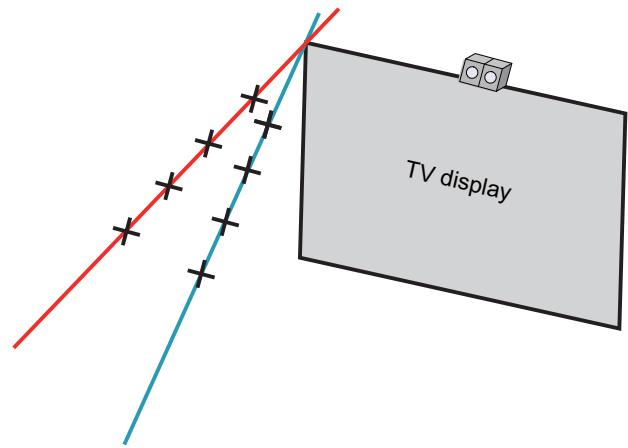


図 3: ディスプレイ四隅の座標の推定方法

3.1.1 顔検出

まず、カメラから得られた画像より、OpenCV に実装されている顔検出機能を用いて利用者の顔を検出する [12]。ここで、誤検出を除去するため、検出された顔領域に対して肌色検出を行う。顔領域のうち、肌色の占める割合が小さい場合は誤検出とみなす。肌色検出には HSV 色空間を用いる。HSV 色空間とは色相 (H)、彩度 (S)、明度 (V) の三つの成分からなる色空間であり、明度を除外することで、明るさの変化に対してロバ

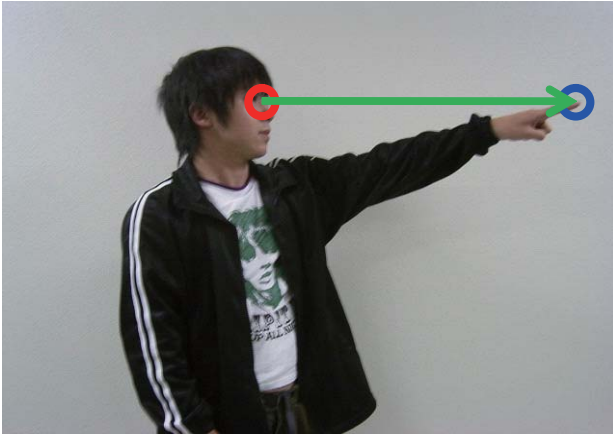


図 4: 指差し方向の定義

ストな肌色検出が可能となる．RGB 色空間から HSV 色空間へ変換は次式により行われる．

$$H = \begin{cases} 60 \times \frac{G-B}{M-m} + 0 & \text{if } M = R \\ 60 \times \frac{B-R}{M-m} + 120 & \text{if } M = G \\ 60 \times \frac{R-G}{M-m} + 240 & \text{if } M = B \end{cases} \quad (6)$$

$$S = \frac{M-m}{M} \quad (7)$$

$$V = M \quad (8)$$

ただし， $M = \max(R, G, B)$ ， $m = \min(R, G, B)$ とする．今回は色相のみを利用し，ほぼ赤色から黄色までに属する領域を肌色として検出した．

また，1 フレーム前で顔が検出されている場合，現在のフレームでは 1 フレーム前で得られた顔領域付近のみに顔検出処理を行うだけで十分である．これにより画像全体に顔検出を行う場合に比べ，高速に検出を行うことができる．実際には顔の未検出が起きた場合を想定して，一定フレームに一回画像全体に対して顔検出処理を行う必要がある．今回は，5 フレームに一回画像全体に顔検出処理を行う．

検出された顔領域に対して，図 5 で示す位置を視点と定義し，利用者の視点の 2 次元座標を求める．得られた視点の 2 次元座標から，前章で述べた POC に基づくステレオ画像の対応付けにより，視点の 3 次元座標を求める．

3.1.2 カルマンフィルタの適用

カルマンフィルタは誤差のある観測値から予測と更新の二つの手続きを行い，真の値を推定するのに用いられる [13]．また，ばらつきの大きい観測値に対して，平滑化を行うためにも利用される．ここでは，フレーム間での視点の位置のばらつきを抑えるために 3.1.1 で得られた視点の 3 次元座標に対して，カルマンフィルタを適用する．

3.2 3次元復元

各利用者の指先を検出するために 3 次元復元を行う．ここで，指先は利用者の付近にあると考えられるため，

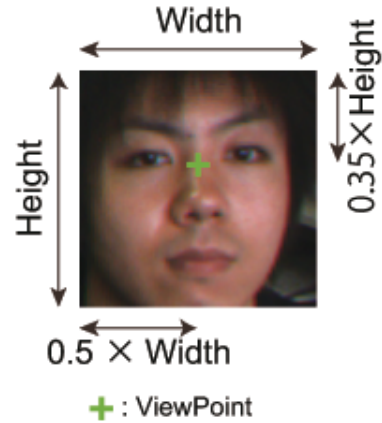


図 5: 視点座標の定義

画像全体を 3 次元復元する必要はなく，利用者の視点とあらかじめ求めておいたディスプレイ四隅の 3 次元座標を利用してある程度限定された領域に対してのみ復元を行えばよい．まず，ディスプレイ四隅と 3.1 で求めた視点の 3 次元座標から図 2 で定義されるような六面体を定義する．指先は，この六面体の中に必ず存在するため，それぞれの利用者に対してこの六面体を作成し，図 6 のようにこれらをカメラから得られた画像上に投影する．3 次元復元は，図 6 で示される領域の内側のみ行えばよい．

また，指先は肌色であるため，肌色検出を行うことで，さらに復元領域を限定できる．ここでの肌色検出は顔検出のときと同様である．通常の復元領域と肌色以外の画素を除去した復元領域を図 7 に示す．このような処理を行うことで，限定された領域に対して高速かつ確実に指先の 3 次元復元を行うことができる．

一方，1 フレーム前で指先を検出している場合，指先の位置はフレーム間でほとんど変わらないと考えられる．そのため，1 フレーム前に指先が検出された場合，図 8 で示されるように，その 1 フレーム前の指先の 2 次元座標付近のみを 3 次元復元するだけで指先は検出される．それにより，さらに計算量を大幅に減らすことが可能である．実際には，指先の未検出が起きた場合を想定して，一定フレームに一回投影された六面体領域に対して 3 次元復元を行う必要がある．今回は，5 フレームに一回六面体領域に対して復元を行う．

3.3 手の抽出

復元された 3 次元点群には，六面体を 2 次元画像に投影して復元領域を限定する都合上，復元される 3 次元点群には実際には六面体内に含まれていない点も存在している．そのため，3.2 で復元された 3 次元点群から，六面体内にある点だけを抽出する．さらに，抽出された点群には顔の点群なども含まれているので，視点から 20 ~ 80cm の距離にある点のみを抽出する．最終的に，図 9 に示すような手の 3 次元点群のみが抽出される．

3.4 外れ点の除去

3.3 で抽出された 3 次元点群には外れ点が含まれている可能性があるため，それを除去する必要がある．外れ点の付近には 3 次元点群が存在しない，もしくは少数しかないと考えられる．そのため，以下の処理を全



(a)



(b)

図 7: 肌色情報の利用による復元領域の削減:(a) 削減なし, (b) 削減あり



図 6: 六面体の利用による復元領域の削減



図 8: 1 フレーム前の指先座標の利用による復元領域の削減

ての残った 3 次元点に対して行うことにより外れ点を除去する。

1. 注目する 3 次元点と、他の 3 次元点との距離を測定する。
2. 注目点を中心とする半径 r の球を作成し、球の中に入った点数 n を調べる。今回 $r = 25$ とする。
3. すべての点に対して、1, 2 の処理を行う。
4. n が閾値 th 以下なら、注目点を外れ点とみなし除去する。今回 $th=20$ とした。

以上の処理により、外れ点を除去することができる。

3.5 指差し位置推定

上記の処理の後、3 次元点が残っている場合は指差しが行われていると判定する。複数の点が残っていた場合、視点から最も遠い点を指先とみなす。一方、3 次元点が残っていなかった場合、指差しは行われていな



図 9: 抽出された手

いと判定する。指差しが行われている場合、指差し位置は、視点と指先を通る直線とディスプレイ平面上の交点として求める。また、視点と同様に、指先位置のばらつきを抑えるために指先の 3 次元座標に対してカルマンフィルタを適用している。

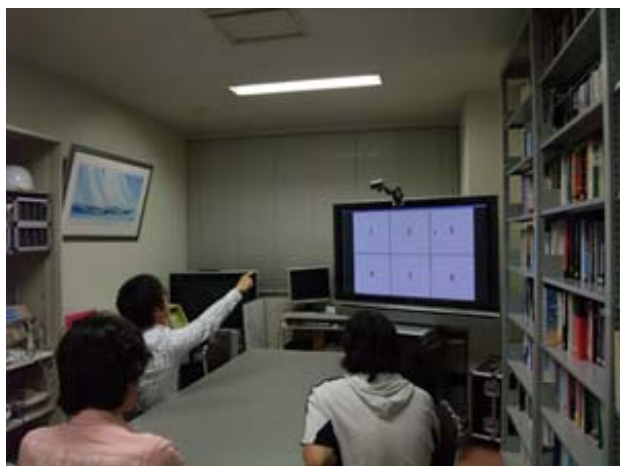


図 10: 実験の様子

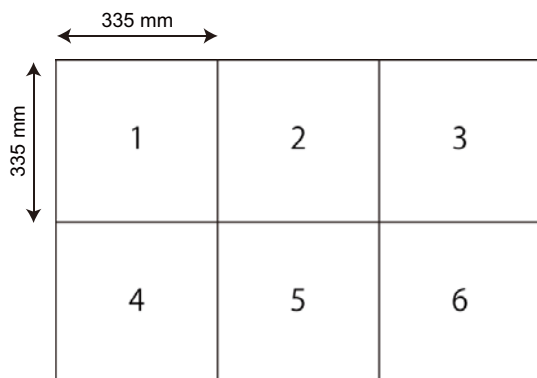


図 11: ディスプレイに表示させる図

4. 実験・考察

本章では、提案する指差し認識アルゴリズムを評価する。実験の様子を図 10 に示す。実験では、図 11 のような数字が描かれた表をディスプレイに表示させている。評価には全部で 4 シーンの動画を用いる。シーン 1 とシーン 2 は、カメラの画角内に一人の被験者のみが存在し、ディスプレイの指定された箇所を 5 秒間指す。シーン 3 とシーン 4 は、画角内に 3 人の被験者が存在し、そのうち一人がディスプレイの指定された箇所を 5 秒間指す。評価は、指先の認識率と処理速度、指差し位置の精度で与える。ステレオカメラは、カラーとモノクロにより構成されている。カメラの基線長は 50mm とし、画像サイズは 960×1280pixel である。提案手法の実装は C++ を用いて行い、使用したマシンスペックは OS: Windows XP, CPU: Core i7 2.93GHz, メモリ: 3GB である。

4.1 指差し認識率に関する評価

指差し認識率に関する実験結果を表 1 に示す。指差し認識率は、被験者が指差しを行っているフレームのうち、指先が正しく認識されているフレームの割合で評価する。表 1 より全てのシーンで高い認識率を得られており、高い精度で指差し認識が行われていることがわかる。指差し認識が行われているフレームの例を

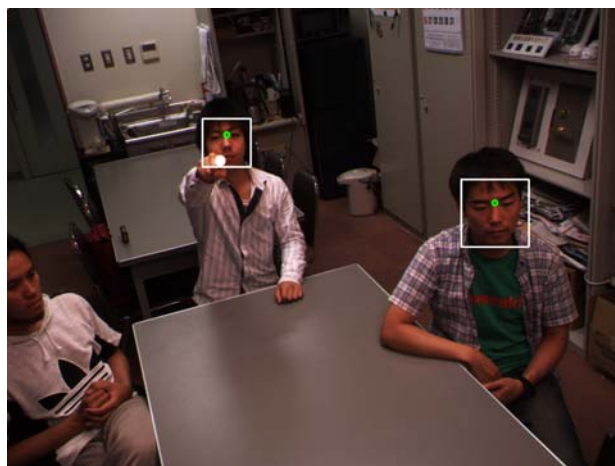


図 12: 認識結果

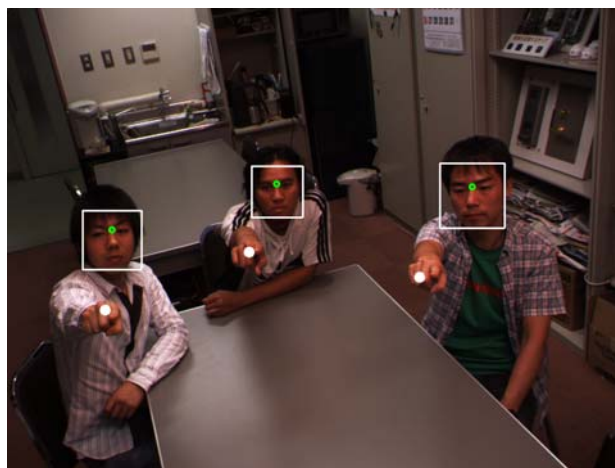


図 13: 多人数での認識結果

図 12 に示す。図 12 は検出された顔領域と視点、指先の位置を示している。また、図 13 からわかるように、多人数でも問題なく各個人の指差しを認識することができる。指差しが認識されない原因としては、顔の検出に失敗していることが挙げられる。顔が検出されない原因としては、顔が少し傾いていたり、指差しの際に、カメラと顔の間に手が重なってしまうことなどがある。この問題を解決するには、顔検出の際に用いられる学習データの中に、顔が傾いているデータや顔に手が重なったデータなどを含める方法などが考えられる。

4.2 処理速度に関する評価

処理時間の結果を表 2 に示す。表 2 より、提案手法では顔検出の際の探索領域の限定や、被験者とディスプレイの位置関係および色情報を利用した 3 次元復元領域の削減により 10fps ~ 20fps を達成している。そのため、リアルタイムでの指差し認識も十分可能である。

4.3 指差し位置の精度に関する評価

指差し位置の精度に関する結果を表 3 に示す。指差し位置の精度は指差しが認識されているフレームのうち、推定されたディスプレイ上の指差し位置が指定した数字のマスの中に入っているフレームの割合で評価

表 2: 処理ごとの計算時間 [ms]

シーン	視点の検出	3次元復元	手の抽出	外れ点除去	指差し位置推定	合計
シーン 1	21.9	24.5	0.3	0.6	1.1	48.4 (20.7 fps)
シーン 2	20.8	24.2	0.2	0.6	1.4	47.2 (21.2 fps)
シーン 3	32.7	45.2	0.4	1.7	2.0	83.0 (12.0 fps)
シーン 4	34.2	46.4	0.5	1.3	1.9	84.3 (11.9 fps)

表 1: 指差し認識率

シーン	認識率 [%]
シーン 1	100.0
シーン 2	96.9
シーン 3	88.9
シーン 4	87.3
平均	93.3

表 3: 正答率

シーン	正答率 [%]
シーン 1	100.0
シーン 2	93.7
シーン 3	100.0
シーン 4	87.5
平均	95.3

する。表 3 から、全てのシーンにおいて、指定した位置と推定された指差し位置がほぼ一致していることがわかる。位置のずれの原因としては、カメラのキャリブレーション誤差やディスプレイ四隅の 3 次元座標の誤差、指先の 3 次元復元がうまくできていないことなどが挙げられる。しかし、提案手法による指差し認識システムではレーザーポインタのようにディスプレイ上に指差し位置が表示されるため、それが意図した箇所からずれていたとしても利用者が容易にフィードバックをかけて調整することができる。そのため、指差し位置推定の多少のずれは大きな問題にならないと考えられる。

5. むすび

本論文では、ディスプレイを用いた会議などを想定した指差しジェスチャの認識手法を提案した。提案手法では、基線長を短くすることで小型化したステレオカメラをディスプレイの上に取り付けるだけで利用者がディスプレイのどこを指しているのかを認識することができる。ステレオ画像の対応付けに位相限定相関法を利用することで、高精度な 3 次元情報を取得し正確に指差しの位置を推定することが可能である。また、提案手法ではディスプレイと利用者の位置関係や色情報などを利用することで画像中の計測領域を削減し、処理の大幅な高速化を達成している。本論文では、実際にディスプレイとステレオカメラを組み合わせた指差し認識システムを構築し、実験を通して利用者が複数の場合でも提案手法によって高精度かつリアルタイム (10 fps 以上) に指差しを認識できることを示した。

参考文献

- [1] M. S. Lee, D. Weinshall, E. Cohen-Solal, A. Colmenarez, and D. Lyons. A computer vision system for on-screen item selection by finger pointing. *Proc. of IEEE Computer Society Conference on Computer vision and Patter Recognition (CVPR)*, Vol. 1, pp. 1026–1033, 2001.
- [2] R. Kehl and L. Van Gool. Real-time pointing gesture recognition for an immersive environment. *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 577–582, May 2004.
- [3] K. Nickel, E. Seeman, and R. Stiefelhagen. 3d-tracking of head and hands for pointing gesture recognition in a human-robot interaction scenario. *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 565–570, May 2004.
- [4] Y. Yamamoto, I. Yoda, and K. Sakaue. Arm-pointing gesture interface using surrounded stereo cameras system. *Proc. of International Conference on Pattern Recognition (ICPR)*, pp. 965–970, 2004.
- [5] A. Wu, M. Shah, and N. Da Vitoria Lobo. A virtual 3D blackboard: 3D finger tracking using a single camera. *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 536–543, 2000.
- [6] D. A. Bowman, E. Kruijff, J. Joseph, and I. Poupyrev. *3D User Interfaces Theory and Practice*, Addison. Addison-Wesley, 2005.
- [7] MESA Imaging AG. Swiss ranger sr4000. <http://www.mesa-imaging.ch/prodview4k.php>.
- [8] Panasonic Electric Works Co. Ltd. 距離画像センサ. <http://denko.panasonic.biz/Ebox/kyorigazou/index.html>.
- [9] 3DV Systems. Z-cam. <http://www.3dvsystems.com/technology/technology.html>.
- [10] K. Takita, M. A. Muquit, T. Aoki, and T. Higuchi. A sub-pixel correspondence search

technique for computer vision applications. *IE-ICE Transactions.Fundamentals*, Vol. E87-A, No. 8, pp. 1913–1923, Aug 2004.

- [11] 柴原琢磨, 沼徳仁, 長嶋聖, 青木孝文, 中島寛, 小林孝次. 一次元位相限定相関法に基づくステレオ画像の高精度サブピクセル対応付け手法. 電気情報通信学会論文誌 D, Vol. J91-D, No. 9, pp. 2343–2356, September 2008.
- [12] Open source computer vision library. <http://sourceforge.net/projects/opencvlibrary/>.
- [13] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, Vol. 82, pp. 35–45, 1960.