一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

HCGシンポジウム2014
HUMAN COMMUNICATION GROUP SYMPOSIUM 2014
HCG2014-B-8-3

ポスター自動発表システムの実現に向けた 視線分布に基づくポスター対話の話題推定

吉本　廣雅†　　中村　裕一†

† 京都大学 学術情報メディアセンター
〒 606–8501 京都市左京区 吉田本町
E-mail: †yoshimoto@ccm.media.kyoto-u.ac.jp, ††yuichi@media.kyoto-u.ac.jp

あらまし　本稿では，ポスターを介した会話シーンを対象としたインタラクションのオンライン自動計測法と自動分析法を提案する．前者は体格・顔立ち，髪型・メガネの有無などの個人差の影響を受けにくい画像処理による頭部姿勢の３次元計測法であり，後者はポスター上のコンテンツとユーザの視線 (頭部方向) の関連から会話中の話題を自動で推定する方法である．本稿は手法の概要と実験結果について詳述する．また提案法の応用例としてポスターの自動発表システムを想定し，提案方法の有効性について議論する．
キーワード　多人数インタラクション，運動計測，視線分析，トピックモデル

Gaze-based Topic Estimation in Poster Conversation toward Automated Poster Presentation

Hiromasa YOSHIMOTO† and Yuichi NAKAMURA†

† Academic Center for Computing and Media Studies, Kyoto University Yoshidahonmachi，Sakyo，Kyoto，606–8501 Japan
E-mail: †yoshimoto@ccm.media.kyoto-u.ac.jp, ††yuichi@media.kyoto-u.ac.jp

Abstract　This paper proposes a couple of interaction sensing and interaction analysis methods, which focus on a multi-party conversation scene in front of a poster. The proposed sensing method is an image-based 3D motion capture that can cope with the individual variations of the appearances, such as stature, features, hairstyle, and other materials extraneous objects like glasses. The interaction analysis method is topic estimation that considers the relationship between the gaze and the contents on the poster. We evaluated the performance of the proposed methods through experiments. We also discusse the effectiveness in possible application of automated poster presentation.
Key words　Multi-Party Interaction, Motion sensing, Gaze analysis, and Topic model

1. はじめに

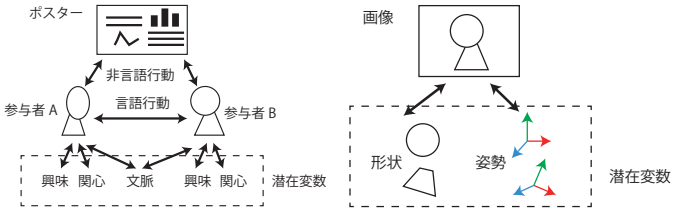
人間のコミュニケーション特性を理解し，さらにそれをオンラインで支援する計算機システムを実現するには，人間の行動をオンライン計測する手法と計測結果から行動の背後にある人間の心的状態を認識・理解する二つの手法が重要となる．

本研究は，**ポスター対話**を対象として，これら二つの手法を提案する．ここで，ポスター対話とは，学会のポスターセッションや掲示板のような状況で，不特定多数のユーザが次々とポスターや掲示板の前に立ち寄り，ポスターや掲示物の内容について自由に対話を行う場面を想定している．

ポスター対話では，言語的・非言語的行動を介したインタラクションが次々と発生している．具体的には，参与者–参与者間では発話や頷きが，参与者–ポスター間では注視などの行動が生じている．

ここで，行動と心的状態の関係を図 1i に示す．心的状態として興味や関心，対話の文脈を考えると，これらは外部からは直接観測できない潜在変数だと言える．そしてこれら潜在変数に応じて行動が表出する行動の生成モデルを考えることができる．ここで，この生成モデルには身体性などに起因する個人差が含まれている．そのため行動から心的状態を推定する処理には，身体性のような個人差をうまく扱いつつ，同時に文脈の変遷を推定するような，潜在変数の同時推定アルゴリズムが必要となる．

次に，人間の観測データと行動の関係を考える．人間の観測デバイスとしてカメラを利用する場合を考えると，これは画像からその人物の姿勢を推定する問題となる．画像と姿勢の関係を図 1ii に示す．この場合，人間の形状と姿勢が潜在変数であ



(i) 行動と文脈と興味・関心の関係　(ii) 撮影画像と姿勢と形状の関係

図 1: 観測と潜在変数の関係

り，それに応じて画像が撮影される生成モデルを考えることになる．ここで形状には体格・髪型・服装・メガネの有無などの個人差が含まれている．そのため画像から行動を計測する処理では，各個人の形状の個人差をうまく扱いつつ，同時に姿勢を推定するような，潜在変数の同時推定アルゴリズムが必要となる．

従来法の多くは，潜在変数の同時推定が実現できているとは言えない．たとえば心的状態の推定法としては，例えば中田らは非言語行動の出現パターンによる会話構造の抽出法を提案している [1]．しかし，これは個人差の影響を受けにくい普遍的なパターンを対象としたものであり，実際のインタラクション中に起こりうる全ての事象を網羅的に扱う手法であるとは言えない．また画像処理ベースの動作計測法としては，たとえば Active appearance model (AAM) を用いた顔向き検出方法がある [2]．これは，あらかじめ人物の詳細な形状モデルを用意し，それを画像に当てはめることで姿勢を推定する方法である．この方法は，形状モデルと実際の人物の形状とに不一致があると，処理が破綻してしまう欠点がある．そのため，実際のインタクション中に参加しうる様々な人物を頑健に計測できる手法であるとは言えない．

我々は，頑健で実用的な要素技術を実現するには，この個人差に対応できる枠組みが必要不可欠であると考える．そして本稿では，個人差に対応できる動作計測方法とその分析方法の二つを提案する．これはつまり，潜在変数の同時推定を行う手法である．

以下，本稿は 2. 節で動作計測法を，3. 節で分析方法の詳細を説明する．また実験では，のべ 24 名，720 分に渡る実際のポスター対話状況を用いて，提案方法の性能評価を行う．最後に，6. 節で実アプリケーションでの応用を前提とした考察を述べる．

2. 画像処理による インタラクションのオンライン３次元計測

図 2 にインタラクションのオンライン３次元計測法の概要を示す．本手法は，図 2a に示すように参与者を撮影した画像の時系列のみを入力として，図 2d に示すように各参与者の頭部と胴体の３次元姿勢の時系列を出力する．本手法の利点は，参与者の頭部や胴体に関する事前知識を極力用いない点にある．体格差や髪型や眼鏡の有無等に柔軟に対応できるように，本手法は参与者固有の外観や身体構造をオンラインで獲得する．それにより姿勢推定の頑健性や精度の向上も実現する．

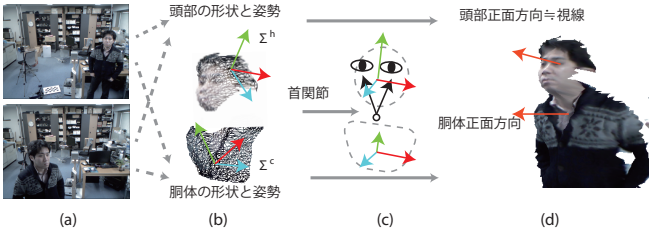


図 2: 画像処理による運動の３次元計測法の概要．(a) 入力画像列 (b) 形状と運動の同時推定 (c) 正面方向補正 (d) 計測結果

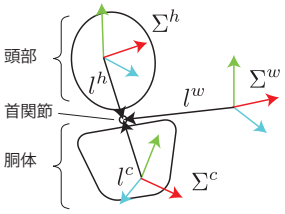


図 3: リンク構造への近似

2.1 頭部・胴体の形状と姿勢の同時推定

形状と姿勢の同時推定には我々が文献 [3, 4] で提案した手法を応用する．

ある人物 u の頭部を一つの剛体とみなし，その人物固有の**頭部形状**を S_u^h とする．S_u^h の実体は我々が文献 [4] で提案している cubistic 表現であり，具体的には，顔表面や髪型，眼鏡の形状に関する幾何情報と，その色情報が格納されている．次に，時刻 t での**頭部姿勢**を，並進と回転の 6 パラメータからなるベクトル $P_u^h(t)$ で表現する．

ここで時刻 t で u を RGB-D カメラで撮影して得た画像 $I_u(t)$ と形状 S_u^h と姿勢 $P_u^h(t)$ の幾何関係を考えると，前節で述べたように一枚の画像のみから，形状と姿勢を一意に求めることはできない．我々が文献 [4] で提案したアルゴリズムでは，撮影画像の時系列に対して初期姿勢 $P_u^h(t_0)$ を与えることができれば，初期姿勢を基準とした S_u^h と姿勢の時系列 $\{P_u^h(t)\}$ の対を求めることができる．また，胴体についても同様で，初期姿勢を与えることができれば，胴体形状 S_u^c と胴体姿勢 $P_u^c(t)$ の時系列を求めることができる．

具体的な手順は次のようになる．

（1） 初期頭部姿勢の計算．既存の正面顔検出アルゴリズム [5] を用いてユーザがカメラに正対した瞬間を検出し，RGB-D カメラの深度情報を用いて，その顔位置から初期頭部姿勢を計算する．

（2） 頭部形状と頭部姿勢の同時推定．初期頭部姿勢が得られ次第，文献 [4] に示したアルゴリズムにより，同時推定を開始する．

（3） 初期胴体姿勢の計算．初期頭部姿勢が得られ次第，頭部の下方に胴体があり，頭部と胴体は同じ方向を向いていると仮定し，初期頭部姿勢から初期胴体姿勢を計算する．

（4） 胴体形状と胴体姿勢の同時推定．初期胴体姿勢が得られ次第，胴体についても同時推定を開始する．

2.2 リンク構造への近似

オンラインで頑健な動作を要求される計算機システムでは，個々のモジュールの性能だけでなく，その処理の破綻を検出する枠組みも重要である．そこで形状と姿勢の同時推定処理の破綻を検出するために，個別に計算した頭部と胴体の運動を，頭部と胴体が首関節で接続されたリンク構造の運動として近似することを考える．これにより，その構造がリンク構造に近ければ，それは人間らしい正しい運動であると判断できる．一方，リンク構造とは似つかない構造であれば，処理は破綻しており，少なくともどちらかの片方の運動が正しく得られていないとみなし結果を破棄する．

リンク構造への近似は文献 [6] で提案されている手法を応用する．図 3 に概要を示す．この手法は，頭部，首関節，胴体からなるリンク構造を線形の式で近似する．

まず世界座標系を Σ^w，頭部のローカル座標系を Σ^h，胴体のローカル座標系を Σ^c とする．姿勢 P^h は Σ^h から Σ^w へ座標変換に対応している．次に Σ^w での首関節の位置をベクトル l^w，Σ^h での位置をベクトル l^h とすると，両ベクトルには次の関係が成り立つ．

$$\boldsymbol{l^w = R_{h2w} l^h + T_{h2w}} \tag{1}$$

ここで R_{h2w} は P^h の回転成分を 3×3 行列で表現したものであり，T_{h2w} は並進成分を 3 次の列ベクトルで表現したものである．q 同様に Σ^c での首関節の位置を l^c とすると，次の関係が成り立つ．

$$\boldsymbol{l^w = R_{c2w} l^c + T_{c2w}} \tag{2}$$

ここで R_{c2w}，T_{c2w} は P^c に対応する回転と並進である．最後に，式 1，2 を連立させると，次の 6 変数の線型方程式が得られる．

$$\left[(R_{h2w})(-R_{c2w})\right] \begin{bmatrix} l^h \\ l^c \end{bmatrix} = \left[-T_{h2w} + T_{c2w}\right], \tag{3}$$

ここで $[(R_{h2w})(-R_{c2w})]$ は 3×6 行列であり，右辺は 3 次の列ベクトルである．

式 3 から，P^c と P^h の対を複数サンプルできれば，最小自乗法により l^h と l^c の値が計算できる．その計算には特異値分解 (SVD) を用いる．なぜなら，SVD は数値的に安定であるだけでなく，SVD で得られる特異値を調べると，首関節の自由度や，リンク構造への近似の良さが評価できるためである．そして評価が悪い場合は，形状と姿勢の同時推定処理が破綻しているとみなし，一旦結果を破棄し，再度前節で述べた形状と姿勢の初期値から計算しなおす．

2.3 正面方向の補正

形状と姿勢の同時推定で得られる姿勢 P^h，P^c は初期姿勢を基準とした相対的な姿勢である．そのままでは顔や胴体の正面方向が一意に定まっておらず，分析等の処理で都合が悪い．そこで以下の手順で，形状 S^h，S^c を分析し，形状の正面方向を算出する．そしてその正面方向を基準とした方向へ姿勢データの補正を行なう．

姿勢データの補正処理は，元の形状を正面方向へ座標変換する 4×4 行列を M^h，M^c とすると，$P^h(t)$ を M^h で，P^c を M^c へそれぞれ座標変換する処理として実現できる．

M^c および M^h の計算手順は以下の通りである．剛体の 3 次元姿勢を決定するには，基底となる 3 つのベクトルを一意に定めれば良い．そこで以下の手順で，形状の正面方向に対応する 3 つの基底ベクトルを求める．まず現時点で利用できる情報の中では，最小自乗の結果であるという点で，首関節位置が比較的信頼できる情報である．また一般に人体の胴体と頭部の形状は左右対称であると仮定できる．そこで，首関節位置を起点として，形状の空間的な分布を調べ，分布が左右対称となるような基準面を算出する．次に，基準面上で形状の輪郭を求め，輪郭上の顔や胸板に対応する区間を直線で近似する．最後に，基準面の法線ベクトルと直線の方向ベクトルとの外積を求める．以上の計算で正面方向の姿勢に対応する 3 つの基底ベクトルを計算する．なお本研究では，外積で得たベクトルを剛体の z 軸，直線の方向ベクトルを y 軸，平面の法線ベクトルを x 軸と定める．M^c および M^h は，その回転成分はこれら 3 つの基底ベクトルを列ベクトルとして順に並べたもの，その並進成分は元の形状と変換後の形状の重心の移動量，として算出できる．

3. 視線分布に基づいたインタラクションの時空間解析

頭部と胴体の向きが得られれば，我々はそこから様々な情報を読み取ることができる．また先行研究により頭部方向は多くの場合視線方向と一致することが知られている．たとえば文献 [7] では会話中の動作を分析しており，視線と頭部方向のズレは 10 度程度であったことが報告されている．これらの知見に基づき，本稿では，頭部方向をそのまま視線方向とみなして分析を行なう．

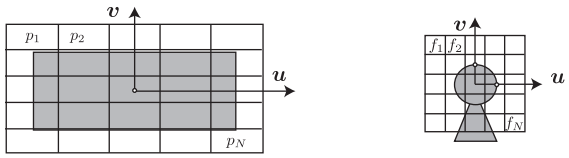
3.1 正規化ポスター空間

画像上でユーザの視線方向を分析しようとすると，各ユーザの立ち位置や，あるユーザからみた注視対象の方向など，さまざな要因を考慮した複雑な計算が必要となる．

一方我々が提案する運動の 3 次元計測方法を用いると，頭部姿勢からユーザの視線に対応する直線を求めることができ，その直線と他のユーザの顔や，ポスター平面との交点を計算するだけで，3 次元空間上の点として注視点が簡単に計算できる．

本節はこの注視点に基づいてインタラクション分析や比較を行なうことを考える．ここで，ユーザ間の立ち位置や距離などの影響を除外するために**正規化ポスター空間**を導入する．正規化ポスター空間は，各ユーザ毎に定義される空間である．対話の参与者を N 人とすると，正規化ポスター空間は，ポスターに対応する**正規化ポスター平面**と，他ユーザに対応する $(N-1)$ 枚の**正規化頭部平面**を基底とした，$2N$ 次元の空間となる．図 4 に概要を示す．

正規化ポスター平面は図 4i に示すように，原点がポスター中央に対応し，ポスターの水平方向に対応する u 軸と，ポスターの垂直方向に対応する v 軸で張る uv 平面として定義する．さらにポスター領域が $[-1, 1]$ の範囲に収まるようにスケール成分を補正する．



(i) ポスターと正規化ポスター平面　　(ii) 参与者と正規化頭部平面

図 4: 正規化ポスター空間

正規化頭部平面は図 4ii に示すように，原点が対象人物の顔位置に対応し，u 軸は対象人物の右方向，v 軸は頭頂方向に対応する．また頭部領域が $[-1,1]$ の範囲になるようにスケール成分を補正する．

Σ^w における注視点の 3 次元座標は，正規化ポスター空間上では $2N$ 次元のベクトルに変換される．この変換により，例えば 2 名のユーザから計算した 2 つの注視点の類似度は，正規化ポスター空間側に変換したベクトルのノルムとして簡単に計算できるようになる．

3.2　ポスター対話における視線分布

次に，正規化ポスター空間での，注視点の分布を考える．本稿では，図 4 に示すように，正規化ポスター平面と正規化頭部平面を格子状に分割し，注視点をこの格子単位で離散化して扱う．

具体的には，図 4 に示すように，各 uv 平面の $[-2,2]$ の範囲を D 等分した D^2 個の領域を考える．ここでポスター側の領域には順に $p_1, p_2, \ldots, p_{D^2}$ と記号を振る．同様に頭部側の領域には順に $f_1, f_2, \ldots, f_{D^2}$ と記号を振る．

これらの記号を用いると，注視点の時系列は対応する記号の列で表現できる．なお注視点が存在しない場合，またはいずれの領域にも含まれない場合は，記号 $null$ を与える．

3.3　トピックモデルを用いた視線分布パターンと話題遷移の同時推定

記号の列として離散化された視線と，潜在変数である興味や関心，話題の関係を考えると，そこには，興味や関心に応じて話題 (トピック) が決定され，その話題に応じて視線として記号が生成される，いわゆるトピックモデルを考えることができる．

図 5 にトピックモデルを示す．これは Latent Dirichlet Allocation (LDA) [8] と呼ばれるモデルである．LDA は主に文章解析等の目的において文書生成モデルとして広く利用されているが，本研究ではこれを視線の生成モデルとして用いる．

図 5 中の z はトピック，w は視線に対応する記号である．このモデルは，トピック z が定まると，そのトピックに応じて視線 w が生成されると考える．ここで θ はトピック混合比と呼ばれ，長さ N の注視動作 w が出力される過程でのトピック z の出現頻度を表している．α は θ のハイパーパラメータであり，本研究における文脈に対応する．また ϕ は z に対する w の生起確率であり，トピック z と興味や関心の対応関係を表している．β は ϕ のハイパーパラメータであり，本研究における個人の興味や関心に対応している．

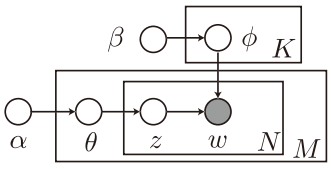


図 5: トピックモデル

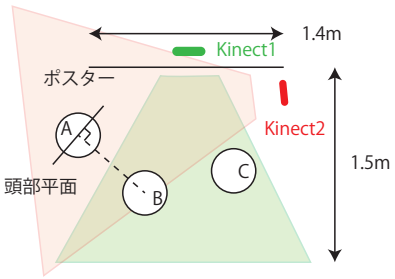


図 6: 実験環境の見取り図

このモデルでは，可観な変数は記号 w のみであり，それ以外の変数は全て潜在変数である．ここで Griffiths らが提案した手法を用いると，M 個の記号列 w に対して Gibbs sampling を行なうことで，各潜在変数を同時に推定できることが知られている [9]．つまり視線の記号列のみから，個人固有の視線の出現パターンと，その時のトピックの種類が，同時に計算できる．

4.　実験 1：ポスター対話の行動計測

4.1　実験環境とデータ収録

実験を始めるにあたり，まず実際のポスター対話を大量に収録した．一回の収録を 1 セッションとし，各セッションでは，発表者役で 1 名，聴衆役で 2 名，合計 3 名の被験者に参加を依頼した．図 6 に，ポスターと各被験者の配置状況の見取り図を示す．ポスターは 50 型の大型ディスプレイ上に静止画像として表示し，3 名の被験者にはディスプレイ前に立ち，指差しなどの自然な動作を交えながら，自由に対話を行なうよう指示した．

対話状況の撮影には Microsoft 社製の Kinect センサを 2 台利用した．配置状況は図 6 に示す通りである．また実験結果を解析するために，姿勢の真値を測定する 3 次元姿勢センサ，発話を記録するピンマイクも併用した．前者の 3 次元姿勢センサとしては POLHEMUS 社製 polhemus liberty を用いた．これは磁気式のセンサであり 240Hz で頭部と胴体の姿勢を計測できる．またピンマイクは SHURE 社製のピンマイクとワイヤレスシステム ULX を用いた．

以上の構成でポスター対話を合計 8 セッション収録した．一つのセッションは時間にして約 30 分あり，合計ではのべ 24 名分，720 分のデータを収録した．これにより，性別，体格，髪型，服装，眼鏡の有無など様々な個人差をもつ被験者達の，多種多様な行動の観測データを得た．

4.2　カメラキャリブレーションとポスタの位置計測

実験にあたり，ユーザやポスターの 3 次元位置を正確に計測するために，Kinect とポスターの配置関係を正確に計測する作業を行った．これは，実験環境を含む 3 次元空間を世界座標系 Σ^w とし，Σ^w で各 Kinect の幾何特性，両 Kinect の配置関係，およびポスター平面の 3 次元位置を計測する作業となる．以下，

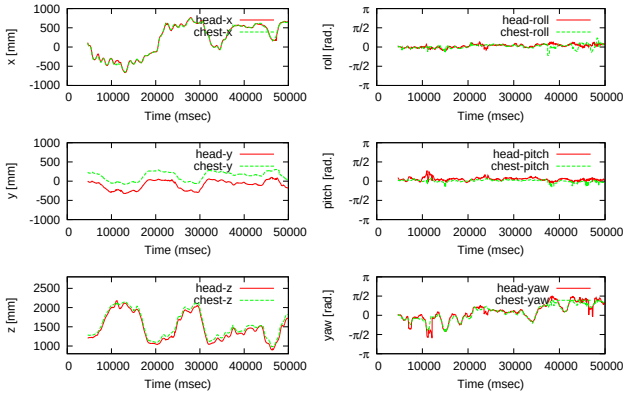


図 7: 頭部と胴体の運動計測例

順に作業概要を説明する．

まず各 Kinect については，文献 [10] の方法を利用した．これは Zhang が提案した chessboard を用いるカメラキャリブレーション法 [11] を応用したものである．

次に Σ^w 中に基準となる参照点を複数配置し，それを両 Kinect で同時撮影することで，Σ^w に対する Kinect の 3 次元位置を算出した．

最後に，ポスター平面上に参照点を複数配置し，その参照点を Kinect で計測し Σ^w 上でのポスター平面の 3 次元位置を計測した．なお，この際ポスター平面は Kinect1 からは全く撮影できず，Kienct2 からは部分的にしか撮影できない．そこで鏡を使用し，鏡の 3 次元位置と，鏡越しに計測した参照点の 3 次元位置から，ポスター平面の 3 次元位置を算出した．

以上の作業により，ポスターを含めた空間全体を 2 台の Kinect で 5mm 程度の分解能で計測できるようになった．この分解能は Kinect の深度情報の分解能を考慮して設定した値であり， Kinect とポスター平面の 3 次元位置の算出作業時は，RANSAC アルゴリズム [12] を用いて 5mm 以上の計測誤差は外れ値として除外するロバスト推定を行った．

4.3 運動の 3 次元計測実験

計測性能の指標としてリアルタイム性と計測可能範囲と計測精度の 3 点を確認した．

まずリアルタイム性に関しては，CPU として Intel 社の Core i5 3.2GHz を，GPU として nVidia 社の GeForce GTX 680 を搭載した PC を用いた場合で，毎秒約 30 frame のスループットでオンライン処理が実現できた．ただしこれは追跡対象人物が 1 名だけの場合である．例えば 3 名を同時追跡するためには，3 倍の計算機資源が必要となる．これは，GPU を 3 台同時利用するか，PC を 3 台利用した並列処理を行なうことで，実現可能であると考えられる．

次に計測可能範囲に関する計測結果を図 7 に示す．図の横軸は時刻であり，各グラフは各時刻での頭部と胴体の姿勢の並進成分 $(x,\ y,\ z)$ と回転成分 $(yaw,\ pitch,\ roll)$ である．このグラフが示すように，本手法は並進成分で $\pm 1m$ 程度の範囲を，回転成分で $\pm\frac{\pi}{2}$ 程度をカバーできている．これは，正面顔だけでなく横顔からも頭部方向を推定できることを意味している．

姿勢計測ソフトウェアの実装例としては，たとえば Microsoft 社が開発した Kinect SDK がある．これは AAM [2] をベースとしたソフトウェア実装であるが，Microsoft 社の API リファレンスによると，その計測可能範囲は *pitch* が ±10 度，*roll* は ±45 度，*yaw* は ±30 度程度しかない [13]．これに対して，提案手法は ±90 度というより広い範囲に対応しており，さらに頭部だけでなく胴体の向きも計測可能であるという 2 つの利点がある．

表 1: 計測精度

	並進 (mm)	方向 (度)
ユーザ B 側	11.2	4.7
ユーザ C 側	13.3	5.7

最後に姿勢の計測精度を計算した．結果を表 1 に示す．これは前述の収録データを用いて，磁気センサの値を真値として，それに対する提案法の誤差を計算している．ただし今回の実験環境では，ディスプレイ付近では磁場が乱れ，真値を正確に測定することができなかった．そのため誤差の評価は，ディスプレイ横に立つユーザ A は除外し，ユーザ B および C についてのみ行った．計測アルゴリズムとして評価すると，Kinect の深度情報の分解能が 5mm 程度であることを考慮すると，並進成分の 11mm 程度の誤差は十分に小さいと言える．特に，収録データには身長や髪型等様々な個人差があり，多種多様な方向の姿勢が含まれていることを考えると，この結果は提案法がこれらバリエーションを正しく扱うことができる頑健な手法であることを実証していると言える．

以上の結果から，本手法は，実環境下でも人間の姿勢を頑健に計測できる，実用的な手法であることを実証した．

5. 実験 2：ポスター対話の話題推定

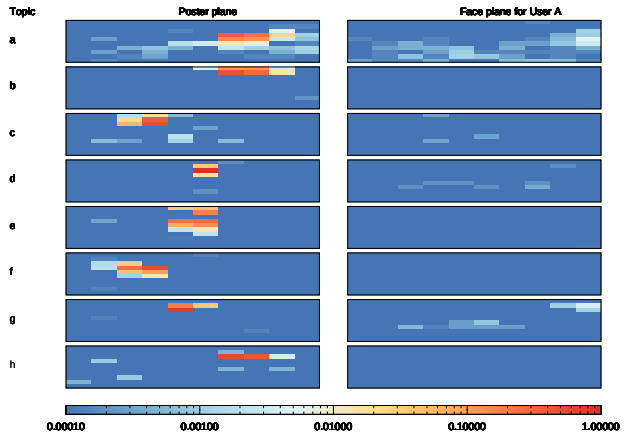
5.1 視線分布パターンと話題遷移の同時推定

4.1 節で収録したデータの各セッションには，一名の発表者がポスターのコンテンツに基づいて説明を行い，残りの 2 名は聴衆としてそれを聞きつつ適宜質問を行なう状況が記録されている．そこで本実験では，まず聴衆役の 2 名の被験者について，それぞれポスターと発表者に対する視線分布を個別に計算し，そこから視線分布パターンと話題遷移を推定した．

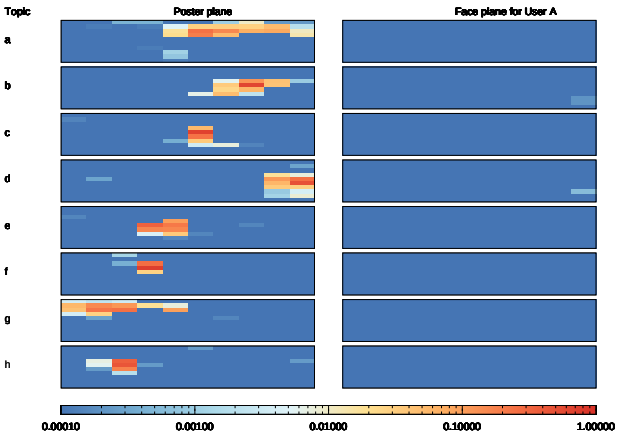
推定にあたり，まず正規化ポスター空間は平面の分割数を 10 として定義した．これにより各聴衆の注視点は，ポスター側と発表者側でそれぞれ 100 個さらに *null* を加えて合計 201 通りの記号へと変換した．

注視点は毎秒 30 回サンプリングできるため，記号列の長さも毎秒 30 となる．本実験ではこの記号列を平均 L 個となるポアソン分布でランダムに分割し，LDA による解析を行った．

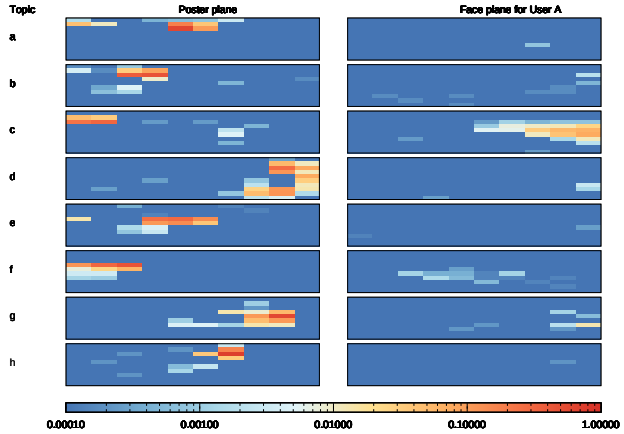
まず，セッション 1 のデータを，トピック数を 8，L を 300，つまり平均 10 秒程度で動作を分割し，LDA による解析を行った．結果を図 8 に示す．なお LDA の解析でもちいる Gibbs sampling は処理の反復回数が増加するにつれて，推定結果の質が単調増加する性質がある．予備実験により 4 万回の反復処理結果を解析したところ 8,000 回程度の反復で増加は頭打ちになることが確認できたため，以下の反復回数は 10,000 回に固



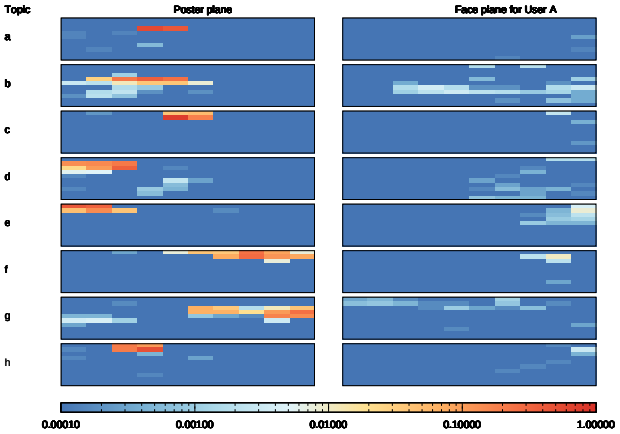
(i) セッション 1 のユーザ B の視点分布パターン



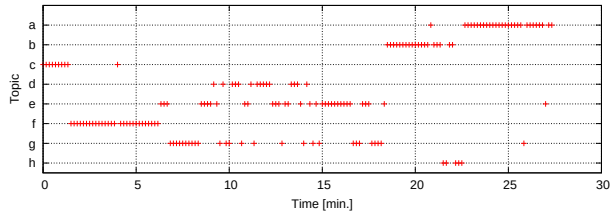
(i) セッション 5 のユーザ B の視点分布パターン



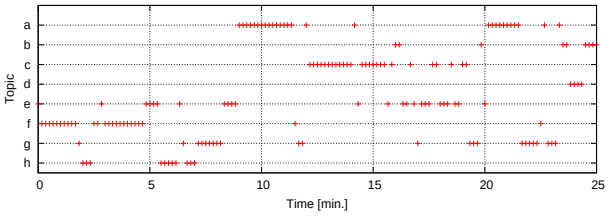
(ii) セッション 1 のユーザ C の視点分布パターン



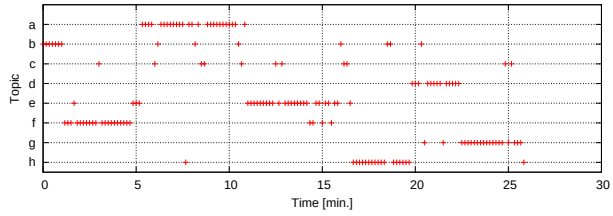
(ii) セッション 5 のユーザ C の視点分布パターン



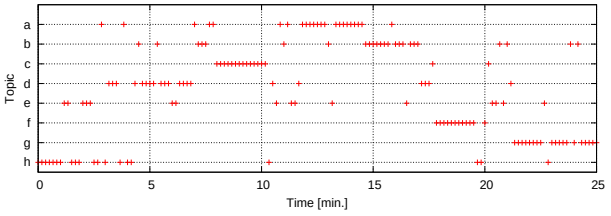
(iii) セッション 1 のユーザ B の話題遷移



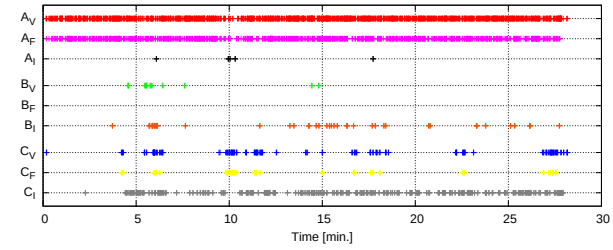
(iii) セッション 5 のユーザ B の話題遷移



(iv) セッション 1 のユーザ C の話題遷移

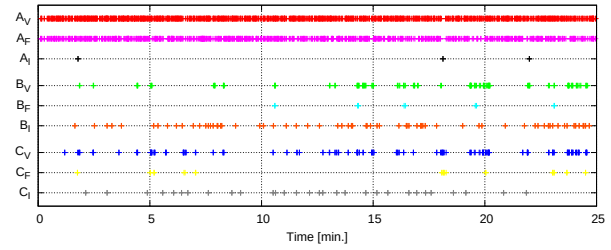


(iv) セッション 5 のユーザ C の話題遷移



(v) ユーザ A,B,C の音声情報 (発話 V/フィラー F/相槌 I)

図 8: セッション 1 の視線分布と話題推定結果



(v) ユーザ A,B,C の音声情報 (発話 V/フィラー F/相槌 I)

図 9: セッション 5 の視線分布と話題推定結果

定して実験を行った．

5.2 推定結果

図 8i に，ユーザ B について推定した視線分布のパターンを示す．これは ϕ を可視化したものである．具体的には，トピック毎に記号の生成確率を求め，その確率分布をポスター (左) および頭部平面 (右) 上にヒートマップとして可視化したものである．図 8iii は，この視線分布のパターンに対応する，トピックの時間遷移を可視化したものである．便宜上 8 個のトピックに a から h のラベルを付与し，以下図の説明を行なう．

まず図 8iii からセッション 1 のユーザ B は，最初にトピック c を選択していることが判る．ここで図 8i のトピック c を見ると，ユーザ B の注視点はポスターの左上に集中していることが判る．収録した音声データを解析すると，この時の対話状況は，発表者 A が自己紹介をしつつ，ポスター左上のコンテンツについて説明を開始した場面であった．

再び，図 8iii に戻ると，ユーザ B はトピック c の次に f を選択し，約 5 分間そのまま f を選択しつづけている．この時の視線はポスター左中央から下部に渡る領域に集中している．この時の対話状況は，発表者 A がポスター左中央付近を説明している場面であった．

同様のパターンはユーザ C からも抽出されている．図 8iv と 8ii を見ると，ユーザ C ではまずトピック b が選択され，それからトピック f が選択されている．ユーザ B とユーザ C を比較すると，視線分布とトピックの時間遷移のタイミングはほぼ一致している．

図 8iv を見ると，大局的にはユーザ C はトピックを b，f，a，e，h，d，e と切り替えていることが判る．この時の視線分布をみると，ポスター上では視線分布が，左上，左下，中央上，中央下，右上，右下と順に移動していることが判る．その分布と順序は，このセッションで用いたポスターのレイアウトと，その説明の順番と完全に一致している．

同様の傾向は，ユーザ B からも読み取れる．図 8iii の時間遷移をみると，ユーザ B は c，f とトピックを切り替え，その後 d，e，g のトピックを頻繁に切り替えながら，最後に b，a とトピックを選択している．ここで対応する視線分布をみると，d，e，g はポスター中央に対応しており，b はポスター右上，a は右下に対応している．

セッション全体でユーザ B とユーザ C を比較すると，両者とも視線分布とポスターのコンテンツが対応してる．またトピックの時間遷移のタイミングも概ね一致しており，特にセッション前半の 0 分から 5 分の区間と，後半の 20 分以降は一致していることがグラフからも読み取れる．

ここで重要な点が 2 つある．1 点目は，トピックの遷移タイミングが同期している点である．今回用いた推定アルゴリズムは，遷移タイミングの計算をユーザ B，C 毎に独立に行っており，B，C 間で情報の共有は行っていない．にもかかわらず，推定結果が同期したのは，対話状況における対話の文脈や雰囲気のような潜在変数が B，C 間で共有されており，それが LDA のトピック変数として上手く表現できたためであると考えられる．2 点目は，ϕ として抽出された視線分布のパターンがポスターのレイアウトと合致している点である．本実験ではポスターのコンテンツに関する情報を全く与えていない．にもかかわらず，視線分布がコンテンツのレイアウトと一致したのは，ユーザがコンテンツの意味的な構造を理解しており，その構造に応じた注視の切り替えを行っているためだと考えられる．

図 9 にセッション 5 での実験結果を示す．セッション 1 とセッション 5 は同じポスターを用いた収録データであるが，状況を若干変化させている．まず参与者は全て別人である．さらに聴衆役は皆学生であるが，発表者はセッション 1 では教員，セッション 5 では学生である．この変化により，対話状況には次のような変化が生じている．収録した音声データに対して，ユーザの発話区間 (V)，フィラー (F)，相槌 (I) の出現状況を手動でアノテーションした結果を図 8v と図 9v に示す．ユーザ A の発話区間 V_A を見るとわかるように，基本的に両セッションでは発表者 A が主に説明を行い，適宜聴衆 B，C が質問等をする流れになっている．ここで，両セッションを比較すると，発表者役を学生が担当したセッション 5 のほうが，聴衆の発話の頻度が高くなっている．これは，聴衆からの質問がより多く出たためである．そのため，セッション 5 では，ユーザ B のトピック e や，ユーザ C のトピック a などで，ポスターと発表者を交互に見ながら議論をする場面がトピックとして抽出されている．ただし全体的な傾向としては，セッション 1 と同様であり，トピックの遷移はユーザ間で同期している箇所があり，トピック毎の視線分布パターンはコンテンツのレイアウトと対応している．同様の傾向は，他の 6 セッションでも確認できた．紙面の都合上，詳細は割愛する．

以上の結果から，提案手法を用いた視線データの解析処理により，対話の文脈のような潜在変数をトピックとして，視線分布パターンを ϕ の確率分布として，同時計算できることが確認できた．

6. 考察：自動発表システムへの応用の検討

最後に実システムでの応用を視野に入れた考察を行う．提案法は，Kinect を介して記録した時系列データを，視線分布とトピックの対にクラスタリングしている．ここでトピックは対話中の文脈や場面に対応するパターンであり，視線分布は注視行動のパターンである．このクラスタリングの応用方法としては，次のようなものが考えられる．

6.1 視線分布パターンのオンライン検出

視線分布は確率分布である．あるユーザの注視点の時系列に対して，その軌跡上で確率を積分すると，視線分布とユーザの注視行動の一致度が計算できる．この一致度の大小を比較することで，複数の視線分布から一番近い視線分布が選択できる．ここで，視線分布の数を N，i 番目の視線分布に対する一致度を $\{score_i\}_{i=1}^{N}$ とすると，この選択処理をオンライン化したアルゴリズムは図 10 のようになる．この処理により逐次 $\{score_i\}$ を更新することで，一番似たパターンが特定できる．

6.2 収録音声・映像の自動分割

実験で示したようにトピックは対話状況の文脈に対応している．つまり収録音声や映像をトピック単位で分割するだけで，文脈単位の分割処理が実現できる．その処理過程では，アノ

```
foreach i in [1, N − 1] do
  | {score_i} = 0
end
while do
  | 新しいユーザの注視行動を計測し，注視点を計算する．
  | 注視点を正規化ポスター空間上での座標 g へ変換する．
  | foreach i in [1, N − 1] do
  |   | i 番目の視線分布について，点 g での確率 p を参照する．
  |   | {score_i} += p
  | end
end
```

図 10: 視線分布パターンのオンライン検出アルゴリズム

テーションのような煩雑な手作業は一切不要である．映像や音声の収録から，分割処理までが，すべて自動で行える．

本来，文脈や意味の理解は知的な処理が必要であり，そのような知的な処理を計算機で実現するのは非常に困難である．一方，本提案手法は，意味の理解は行わず，文脈と視線分布の組み合わせをパターンとして抽出している．そのため，自動分割されたデータは，質としてはあまり良いものであるとは言えない．しかし大量のデータが容易に収集できる点で，応用上は様々な利点があると言える．

たとえば自動分割の結果は意味的な分割の候補になると考えられる．自動分割の結果をインデックス的に利用すると，収録データの閲覧の手間が削減できると考えられる．

6.3 ポスターの自動発表システムへの応用

視線分布パターンのオンライン検出と，収録データの自動分割を組み合わせた実アプリケーション例の一つは，ポスターの自動発表システムである．

これは，あらかじめ発表者と聴衆のポスター対話を大量に収集分析しておき，それを自動分割することで，視線分布パターンとポスター対話のコンテンツの対としてデーターベース化しておく．そして発表者が不在の場合は，聴衆がポスターを注視した状況を自動検出し，その視線分布に似たコンテンツをデータベースから取り出し，聴衆へ提示する．

ここで，提示後の聴衆の注視パターンを分析することで，提示したコンテンツのふさわしさが評価できる．たとえば，提示したコンテンツを注視したり，コンテンツとポスターを交互に見る場合は，提示したコンテンツは聴衆の興味にあったものだと考えられる．一方，聴衆がコンテンツとは関係のない別の視線分布パターンに切り替わった場合は，提示したコンテンツはふさわしくないものであったと考えられる．このように本提案手法の応用例としては，例えば，コンテンツを自動生成しつつ，コンテンツの提示とその提示後の反応を繰り返すことで，より良いコンテンツを多数決的に選択する枠組みが考えられる．

このように，本提案手法は実環境で人間の振舞いに柔軟に対応できる計算機システムのための要素技術の一つとなりうると，我々は考えている．

7. おわりに

本稿では，個人差の影響を受けない頑健な頭部・胴体の３次元姿勢計測法と，視線方向の３次元解析に基づいた個人毎の視線分布パターンと話題遷移の同時推定法の２つについて提案を行った．そして，のべ 24 名，720 分の実データを用いた評価実験により，提案は実環境での様々な人物の動作を頑健に計測・分析できることを実証した．

提案法を用いると，大量の動作データを自動収集でき，またその分析も自動化できる．その際，煩雑な手動アノテーション作業は不要であり．また 6. で詳述したように，提案手法は人間の興味や関心に応じて自動動作するインタラクティブなアプリケーションを構築するための要素技術であると言える．

今後は，実アプリケーションの実証システムの構築や，アプリケーションの用途に応じたトピックモデルの改良などを行なう予定である．

謝辞 本研究の一部は，独立行政法人科学技術振興機構（JST）戦略的創造研究推進事業 (CREST)「マルチモーダルな場の認識に基づくセミナー・会議の多層的支援環境」の助成を受けて行った．

文　献

[1] 中田 篤志, 角 康之, and 西田 豊明. 非言語行動の出現パターンによる会話構造抽出. **電子情報通信学会論文誌** *D*, 94(1):113–1232, 2011.

[2] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681 –685, Jun. 2001.

[3] 吉本 廣雅 and 中村 祐一. ポスターセッションの分析のための不特定複数人物の頭部形状と姿勢のオンライン自動推定. In **電子情報通信学会技術研究報告** *(HCG* **シンポジウム** *2012)*, Dec. 2012.

[4] 吉本 廣雅 and 中村 祐一. 未知剛体の形状と姿勢の実時間同時推定のための cubistic 表現. **電子情報通信学会論文誌** *D*, 97(8):1218–1227, Aug. 2014.

[5] Paul Viola and Michael Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001.

[6] James F. O'Brien, Robert E. Bodenheimer, Gabriel J. Brostow, and Jessica K. Hodgins. Automatic joint parameter estimation from magnetic motion capture data. In *Proceedings of Graphics Interface 2000*, pages 53–60, May 2000.

[7] 矢野 正治, 中田 篤志, 福間 良平, 角 康之, and 西田 豊明. 非言語マルチモーダルデータを用いた会話構造の分析のた めの環境構築. In **情報処理学会研究報告．** *UBI,* **ユビキタスコンピューティ ング**, number 12 in 2009-UBI-22, 2009.

[8] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, Mar. 2003.

[9] Thomas L. Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235, 2004.

[10] Kurt Konolige and Patrick Mihelich. *Technical Description of Kinect Calibration*, 2013.

[11] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, November 2000.

[12] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, Jun. 1981.

[13] Microsoft. *Kinect for Windows SDK*, 2014.