

## 位相限定相関法に基づくステレオビジョンによる高精度頭部運動追従法

伊藤 敦郎<sup>†‡</sup> 酒井 修二<sup>‡</sup> 伊藤 康一<sup>‡</sup> 青木 孝文<sup>‡</sup> 鈴木 陽一<sup>†‡</sup> 坂本 修一<sup>†‡</sup>  
<sup>†</sup> 東北大学電気通信研究所 <sup>‡</sup> 東北大学大学院情報科学研究科

## 1 はじめに

近年、聴取者に音空間情報を提示する3次元聴覚ディスプレイの研究が盛んに行われている。とりわけ、聴取者の両外耳道入口における音圧を制御することでバーチャル音像を提示する伝達関数合成型聴覚ディスプレイは、小規模なシステムで臨場感の高い音空間を再現できることから注目されている。より高い臨場感を得るためには聴取者の頭部運動を反映させる必要があることから、頭部にセンサを装着して頭部運動をセンシングするシステムがこれまで多数提案されてきた(例えば[1]など)。しかし、より自然な聴取を行うためには、頭部にセンサを装着せずにセンシングすることが強く望まれる。そこで本稿では、位相限定相関法とパーティクルフィルタを用いて、動画像から高精度に聴取者の頭部運動を追従する新しい方法を提案する。パーティクルフィルタにおける運動モデルを位相限定相関法を用いて前後フレームの対応関係から定義し、時間軸方向の局所的なモデルを生成することで、高精度な運動追従を可能としている。

## 2 位相限定相関法

位相限定相関法(Phase-Only Correlation: POC)とは、画像の周波数スペクトルのうち、位相スペクトルのみに着目した画像マッチング手法であり、サブピクセルレベルの分解能で画像間の対応付けが可能である[2]。

画像  $g(n_1, n_2)$  が画像  $f(n_1, n_2)$  をサブピクセルレベルの微小量  $(\delta_1, \delta_2)$  だけ平行移動させた画像である場合を考える。このとき、 $f(n_1, n_2)$  及び  $g(n_1, n_2)$  の POC 関数は次式で与えられる。

$$r(n_1, n_2) \approx \frac{\alpha}{N_1 N_2} \frac{\sin(\pi(n_1 + \delta_1))}{\sin\left(\frac{\pi}{N_1}(n_1 + \delta_1)\right)} \frac{\sin(\pi(n_2 + \delta_2))}{\sin\left(\frac{\pi}{N_2}(n_2 + \delta_2)\right)} \quad (1)$$

ここで、 $\alpha$  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。 $\alpha$  を推定することで2つの信号の類似度を求めることができる。また、相関ピークの位置座標  $(\delta_1, \delta_2)$  を推定することで2つの信号の平行移動量を求めることができる。

High-accuracy head tracking using stereo vision based on the phase-only correlation method

<sup>†‡</sup>Atsuro ITO, Yōiti SUZUKI, Shuichi SAKAMOTO

<sup>‡</sup>Shuji SAKAI, Koichi ITO, Takafumi AOKI

<sup>†</sup>Research Institute of Electrical Communication, Tohoku University

<sup>‡</sup>Graduate School of Information Sciences, Tohoku University

## 3 パーティクルフィルタによる頭部運動追従

## 3.1 基本アルゴリズム

まず初期化部では、ステレオカメラから取得した顔のステレオ画像を元に、POCを用いて顔の3次元復元を行う。あらかじめ聴取者の顔の3次元モデルの測定と特徴点の選択をしておき、POCで得られた3次元点群とICPアルゴリズム[3]を用いてフィッティングする。その後、3次元モデル上の特徴点を画像上に投影することで、画像上の特徴点を取得し、同時にテンプレート画像の登録を行う。

初期化で取得した特徴点を用いて、追跡部においてパーティクルフィルタを適用することで頭部運動を逐次的に推定していく。時刻  $t-1$  における頭部姿勢ベクトルを  $p_{t-1} = (x, y, z, \phi, \theta, \psi)^T$ 、 $N$  個の仮説群とその尤度をそれぞれ  $s_{t-1}^{(i)}$ 、 $\pi_{t-1}^{(i)}$  ( $i = 1, \dots, N$ ) で表す。 $\phi, \theta, \psi$  はそれぞれ  $x$  軸、 $y$  軸、 $z$  軸周りの回転(ロール・ピッチ・ヨー)である。時刻  $t$  における暫定的な仮説群  $\hat{s}_t$  を、以下の式を用いて  $s_{t-1}^{(i)}$  から生成する。

$$\hat{s}_t = s_{t-1} + v_{t-1} + \omega \quad (2)$$

ここで、 $v_{t-1}$  は時刻  $t-1$  における頭部姿勢ベクトル  $p_{t-1}$  の速度ベクトルである。本稿における提案では、この速度ベクトルを逐次的に生成することで、ロバストな追跡を可能としている。 $\omega$  はランダム雑音であり、これにより仮説を拡散する。なお、このときのランダム雑音の分散の幅  $\sigma$  は経験的に設定している。続いて、推定した仮説  $\hat{s}_t$  に基づいてテンプレート画像を射影変換させ、その結果と時刻  $t$  における画像とで正規化相互相関(NCC)に基づくテンプレートマッチングを行い、尤度  $\pi_t^{(i)}$  を計算する。また  $\pi_t^{(i)}$  は合計して1になるように正規化する。最後に、尤度  $\pi_t^{(i)}$  が最大となる仮説を  $p_t$  とし、 $\pi_t^{(i)}$  の大きさに応じて新たな仮説  $s_t^{(i)}$  を取得する。この一連の流れを繰り返すことで、頭部運動を推定していく。

## 3.2 高精度頭部運動追従法の提案

本稿では、頭部運動を高精度に追従する方法として、POCによる運動モデルの逐次生成を提案する。時刻  $t-1$  における入力画像と時刻  $t$  における入力画像に対してPOCを適用し、前後フレームにおける画像の対応関係から復元した3次元点群  $M'$  を取得する。この  $M'$  と、

時刻  $t-1$  における 3 次元点群  $M_{t-1}$  との間における回転行列を  $R'$ 、並進ベクトルを  $t'$  とすると、速度ベクトル  $v$  は、並進の速度ベクトル  $v_t$  と回転の速度ベクトル  $v_R$  を用いて以下の式で得ることができる。

$$v = \begin{bmatrix} v_t \\ v_R \end{bmatrix} \quad (3)$$

$$v_t = R' t_{t-1} + t' - t_{t-1} \quad (4)$$

$$v_R = \mathcal{T}(R' R_{t-1}) - \mathcal{T}(R_{t-1}) \quad (5)$$

ここで、 $\mathcal{T}(R)$  は回転行列  $R$  のロール・ピッチ・ヨー表現である。 $R'$  及び  $t'$  は  $M'$  及び  $M_{t-1}$  から求まる共分散行列とその特異値分解から求めることができる [3]。以上の手法に基づき、運動モデルを逐次的に生成することで、突発的な運動に対してもロバストに追従できるアルゴリズムとなることが期待できる。

#### 4 提案法の精度評価

提案法に対して精度評価実験を実施した。運動モデルの逐次生成はせずに、フレーム間で等速度・等角速度運動をしていると仮定した運動モデルを従来法とし、真値に対する RMS 誤差で比較した。実験パラメータは、仮説数が 2000 個、回転行列の分散の幅が  $2^\circ$ 、並進ベクトルの分散の幅が 1.0 mm である。なお、純粋な追跡精度の定量的評価を行う必要があるため、ヒトの顔ではなくテクスチャを持つ球体の追跡により実験を行った。入力動画は、球体を回転テーブルに乗せ  $1^\circ$  間隔で回転させた画像列と、平行移動台に乗せ 1 mm 間隔で平行移動させた画像列とした。なおこのとき、本システムは約 7.2 fps で動作した。

図 1 は、追跡結果の軌跡の一例を示したものである。また、そのときの RMS 誤差を表 1 に示す。座標系はカメラに対して正面を  $x$  軸の正、左手方向を  $y$  軸の正、上方向を  $z$  軸の正と定義した。3 自由度の回転について、全ての回転角において提案法の RMS 誤差が従来法に対して改善されたことが確認できた。特に、左右の頭部回転が音空間知覚に重要であることが知られているため、ヨー方向の回転の高い推定精度が求められる。これについて、提案法は RMS 誤差が  $0.5^\circ$  以下となっている。音像移動の知覚検知限である最小弁別角度 (Minimum Audible Angle : MAA)[4] は正面方向でおよそ  $1^\circ$  であるため、十分な精度であると言える。平行移動については、 $x$  軸方向において従来法と比べ精度が大きく改善された一方で、 $y$  軸及び  $z$  軸方向は従来法とほぼ変わらない精度であった。しかし、追跡対象の運動速度を上げると、等速直線運動の仮定では対処しきれず、パーティクルの拡散範囲から実際の移動先が外れ、追跡に失敗することが多々見られた。これに

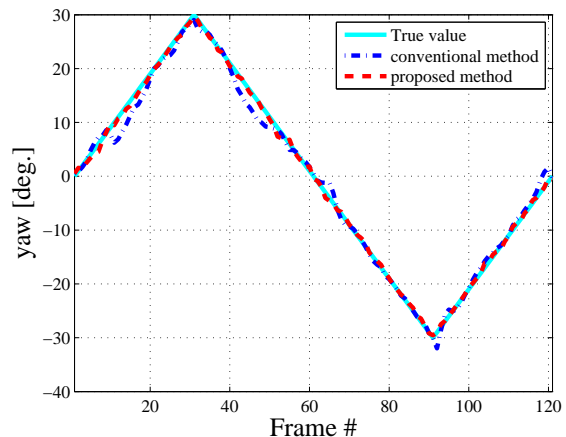


図 1: 追跡結果の軌跡の例 (ヨー回転)

表 1: 従来法と提案法の RMS 誤差

|     | $x$  | $y$  | $z$  | $\phi$ | $\theta$ | $\psi$ |
|-----|------|------|------|--------|----------|--------|
| 従来法 | 2.21 | 0.30 | 0.40 | 0.25   | 1.76     | 1.47   |
| 提案法 | 0.93 | 0.29 | 0.40 | 0.14   | 0.57     | 0.47   |

( $x, y, z$  [mm] and  $\phi, \theta, \psi$  [deg.] )

対し、提案法では問題なく追跡ができた。したがって、従来法に比べ提案法がよりロバストに物体の運動を追跡できることが実証された。

#### 5 まとめ

本稿では、パーティクルフィルタの運動モデルを位相限定相関法を用いて逐次的に生成することで、高精度な頭部運動追従法を提案した。その結果、特に水平回転運動について提案法は従来法よりも高精度に計測できることが示された。今後は聴覚ディスプレイと接続し、主観実験による音空間提示性能の評価を行う予定である。

#### 謝辞

本研究の一部は科研費基盤研究 A(24240016) 及び通研共同プロジェクト研究 (H24/A14) の助成による。

#### 参考文献

- [1] Yukio Iwaya and Yôiti Suzuki, "Rendering Method of Moving Sound Source with the Doppler Effect in Sound Space," *Applied Acoustics*, Vol. 68, No. 8, pp. 916-922, 2007.
- [2] K. Takita *et al.*, "A sub-pixel correspondence search for computer vision applications," *IEICE Trans. Fundamentals*, Vol. E87-A, No. 8, pp. 1913-1923, 2004.
- [3] 八木康史・斎藤英雄 編, "コンピュータビジョン最先端ガイド 3," アドコム・メディア株式会社, 2010.
- [4] A. W. Mills, "On the Minimum Audible Angle," *J. Acoust. Soc. Am.*, Vol. 30, No. 4, pp. 237-246, 1958.