

速度ピークの相関を用いた頷きと横振り検出

山田 貢己[†] 山口 修[†]

映像中から顔領域と顔中心座標（鼻の近傍）を検出し、顔中心座標の速さが極大となる時刻の速度ベクトル（速度ピーク）を記録し、速度ピーク間の自己相関に基づいて頭部ジェスチャ検出を行う新しいアルゴリズムを提案する。速度ピーク相関は、頭部ジェスチャにおいて観測される往復/反復運動を効果的に検出することができる。頭部の頷きと横振りを検出するシステムを PC 上に構築し、動作を確認し、アルゴリズムの有効性を示した。

Nod and Shake Detection using Autocorrelation at Velocity Peaks

Miki Yamada[†] Osamu Yamaguchi[†]

This paper proposes a new algorithm for head gesture detection using correlation at velocity vector peaks. The velocity peak correlations effectively detect reciprocating/repetitive motions that are observed in head gestures. We built a head nod and shake detection system and confirmed the effectiveness of the algorithm.

1 まえがき

頷きや首の横振りなどの頭部ジェスチャを自動検出することができれば、さまざまな分野で応用が可能である。例えば、無意識の頷きを検出することができれば、通信回線を経由した遠隔コミュニケーションや手話の認識の際の非言語的情報獲得方法として利用できる。意識的な頷きの検出は、電化製品のリモコンによらない遠隔操作、コンソール端末の操作性向上、キーボードやマウスに不慣れたユーザーへの入力支援、ゲーム機器の入力インタフェース、運転中のドライバーへの機器操作支援等に使える可能性がある。

従来の頷き検出技術の多くは、隠れマルコフモデル (HMM) などの状態モデルを用いている [1-7]。Morimoto らは、顔領域のオブティカルフローを計算し、動的ベクトル量子化を用いて 7 種類の動作記号 (UP, DOWN, LEFT, RIGHT, IN, OUT, REST) に分類し、記号の系列と HMM を用いて頭部ジェス

チャ (yes, no, maybe, hello) を認識している [1]。

Zelinsky と Heinzmann は、atomic actions (up, down, etc.) の組み合わせで “Yes” と “No” のジェスチャを記述する。そして、画像中の顔領域の動きベクトルを計算し、ベクトル状態が静止→下降→上昇のように変化したときに頷きを認識する [2, 3]。

Kapoor と Picard は、瞳の追跡のために赤外線 LED を用いた赤外線高感度カメラを用い、離散 HMM により頭部の頷きを検出する [4]。

Ng と De Silva は、状態入力として画像の invariant moment を用い、5 状態 HMM を用いて頷き検出を行う [5]。

Davis と Vaks は、finite state machine model を用いて首振り運動を検出する [6]。2 次元の運動のヒストグラムのピークの向きを量子化し、Down, Up, Left, Right, NULL に分類する。状態が Down → UP → $\tilde{U}p$ と変化したときに頷きを検出する。

状態モデルや学習を用いない頭部ジェスチャ認識アルゴリズムも研究されている [8, 9]。Wu らは、連続 DP 法を用いて入力画像時系列を標準パターンと比較して頭部ジェスチャを認識する。標準パターン

[†] (株) 東芝 研究開発センター
Corporate R&D Center, Toshiba Corp.

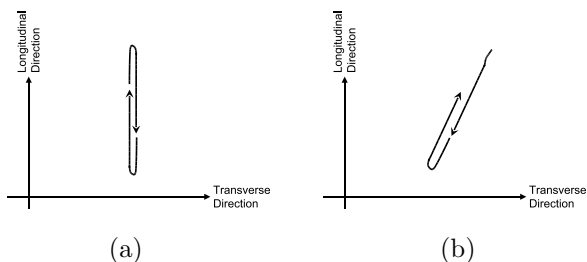


図1 標準的な往復運動と斜めの往復運動の軌跡.

は各ユーザー毎に構築される [8]. Kawato と Ohya は、眉間を検出してその点の最大座標値に基づいて頭部ジェスチャを検出している [9].

本研究報告は、顔中心座標の速度ベクトルピーク間の相関を用いる、新しい頭部ジェスチャ検出アルゴリズムを提案する. このアルゴリズムは、下降運動や上昇運動を独立に検出したり判定することはせず、下降運動の速度ベクトルと、これに続いて生じる上昇運動の速度ベクトルとが成す角度（或いは、内積）が平角（180度）に近いかどうかを見ることにより頭部の頷き/横振りを検出する. 以下で詳細を述べる.

我々は、身体の重い部位の運動から起こるジェスチャは往復/反復運動（図1）から成ると予想する. なぜならば、それらは最も単純な運動であり、人間にとって容易に行えるからである. 人体の頭部は指や手よりも重い為、典型的な頭部ジェスチャである頷きと横振りは往復/反復運動から構成されると考えられる. ここで、往復運動は、往きと帰りが逆向きとなる運動であり、両者の運動のベクトルが互いに逆向きで、成す角度が平角（180度）に近い運動であると解釈する. さらに、これら比較すべき2つのベクトルは、顔中心座標の速度ベクトルの（長さの）極大値を与える点の速度ベクトル（これを速度ベクトルピーク、または、速度ピークと呼ぶことにする）を採用することにする. 速度ピークによって往きの動きと帰りの動きを代表させることには以下で述べる様々な利点がある.

頭部ジェスチャにおけるもう一つの特徴的な運動である反復運動は、往復運動を連続して行ったときに観測される運動である. 速度ピークをリストアップし、一つ置き速度ピークを比較したとき、それら2つベクトルが平行（0度）に近いときに反復運動で

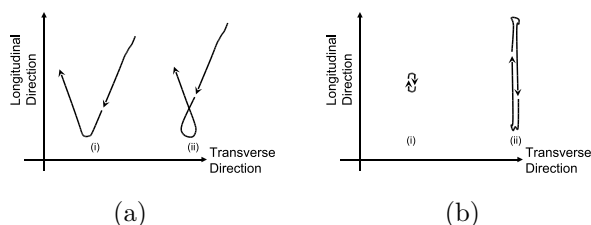


図2 (a) 従来手法では「頷き」として誤検出し易い非往復運動の軌跡. (b) 小さい往復運動と大きい往復運動の軌跡（頷きの場合）.

あると定義する. 顔中心座標の速度ピークの観測によって、往復運動と反復運動を検出することで、頭部の頷きと横振りを判定するというのが提案手法の要点である.

頷きと横振りは非常に単純な動きであるがゆえに、それを精度良く検出することは難しい. 普段の人間の動作において、頭部は頷き以外の上下動（図2(a))を頻繁に行っている. 従って、実用に耐え得る頷き検出システムは、本当の頷き（図1）とそれ以外の上下動を区別しなければならない. さらに、人間は頷き動作の前後に頭部を静止させるとは限らないため、頷きの前後の静止を検出するアルゴリズムはこれが原因で頷き検出の精度を低下させてしまう. 本手法は、これらの問題を解決する. 図1(b)のような斜めの頷きと、図2(a)のようなV字形の運動は、頭部の下降運動と上昇運動を独立して判定する手法 [1-4, 6, 7] では区別できない. 斜めの頷きは、画像を用いる場合のカメラ位置の変化や、人間の姿勢の微妙な変化によってしばしばそれを検出しなければいけない状況が生じ得る. 本手法は、上昇と下降の相対的なベクトルの向きを調べるため、斜めの頷きとV字形運動とを区別することができる. さらに、静止状態を判定に利用していないため、静止しなかったことが原因で頷きが検出されなかったという失敗が起こらない. インタフェースとして「意識的な頷き」を利用する場合でも、頭部の静止を待たずに、頷き動作が完了する前（速度ピークを通過した瞬間）に頷きを検出できることはメリットとなる.

本手法は速度ピークのみを用いているのだが、そこは最も動きが安定している点である（図2(b)). なぜならば、センサからの入力には対象の動きとは無関係で大きさが一定のノイズ要因（顔特徴点の位置

の誤差など)があり、動きが大きくなると動き情報に対して相対的にノイズは小さくなるからである。また、この手法はジェスチャ全体を通じて少数個の点のみを考慮し、速いジェスチャと遅いジェスチャを原理的に同等に検出する。時系列パターンのマッチングを用いた従来手法では、検出精度が時間方向の伸び縮みの影響を受けるのに対し、本手法は影響を受けない。

本手法が利用可能な座標値は、顔領域の中心座標(2次元または3次元)の他に、顔特徴点の位置や顔向き角度など、頭部運動を表す何らかの座標値であればよい。さらに、この報告では主として画像処理によるジェスチャ認識を想定しているが、画像以外のセンサ(例えば、磁気位置センサ)から得られる頭部の座標情報を取得して、その速度ピーク間の相関を利用したジェスチャ検出を行うことも可能である。

2 速度ピーク相関を用いたジェスチャ検出アルゴリズム

2.1 座標の速度ピーク間の相関

観測された顔中心座標に関して、その速さの極大点 \mathbf{v}_i^{\max} をリストアップし、連続する2個の極大点の速度ベクトル間余弦($\cos \theta$)を計算する。顔中心座標は、形状抽出とパターン照合の組み合わせに基づく顔検出手法[10]によって求める。第 f 画像フレームの2次元速度ベクトル \mathbf{v}_f は次式で計算される:

$$\begin{aligned} \mathbf{v}_f &= \begin{bmatrix} v_x(f) \\ v_y(f) \end{bmatrix}, \\ v_x(f) &= \frac{x_{f+1} - x_f}{t_{f+1} - t_f}, \\ v_y(f) &= \frac{y_{f+1} - y_f}{t_{f+1} - t_f}. \end{aligned} \quad (1)$$

ただし、 $[x_f, y_f]^T$ は第 f フレームの顔中心座標、 t_f は第 f フレームの時刻である。速度ベクトル \mathbf{v}_f は次式の不等式が成り立つとき、速度ピークと呼ぶことにする:

$$\begin{aligned} |\mathbf{v}_f| &> |\mathbf{v}_{f-1}|, \\ |\mathbf{v}_f| &> |\mathbf{v}_{f+1}|, \\ |\mathbf{v}_f| &> \theta_{\text{peak}}. \end{aligned} \quad (2)$$

ここで、 $|\mathbf{v}|$ はベクトル \mathbf{v} の長さを表し、 θ_{peak} は頭部ジェスチャに対する最小速度を表す閾値である。速

度ピークには通し番号を振って、 \mathbf{v}_i^{\max} , ($i = 1, \dots$) と表す。

頭部ジェスチャの構成要素として以下で「往復運動」(図1,2(b))を定義する。連続する速度ピーク間の相関 C_i^{ser} を

$$C_i^{\text{ser}} = \frac{\mathbf{v}_i^{\max} \cdot \mathbf{v}_{i-1}^{\max}}{|\mathbf{v}_i^{\max}| |\mathbf{v}_{i-1}^{\max}|} \quad (3)$$

で表すことにすれば、往復運動が行われたときは2つのベクトル \mathbf{v}_i^{\max} と \mathbf{v}_{i-1}^{\max} が互いに逆向きになるので、 C_i^{ser} の値が -1 に近づくと考えられる。上式において $\mathbf{v} \cdot \mathbf{w}$ はベクトル \mathbf{v} と \mathbf{w} の内積を表す。従って、 i 番目の速度ピークが

$$C_i^{\text{ser}} < \theta_{\text{reci}}, \quad (4)$$

$$-1 < \theta_{\text{reci}} < 1 \quad (5)$$

を満たしたとき、往復運動が行われたと判定することができる。ここで、 θ_{reci} は頭部ジェスチャにおける往復運動の検出閾値である。 C_i^{ser} の値は、完全な往復運動のときに -1 に等しくなる。

さらに、頭部ジェスチャの構成要素として「反復運動」を定義する。一つおきの速度ピーク間の相関 C_i^{alt} を

$$C_i^{\text{alt}} = \frac{\mathbf{v}_i^{\max} \cdot \mathbf{v}_{i-2}^{\max}}{|\mathbf{v}_i^{\max}| |\mathbf{v}_{i-2}^{\max}|} \quad (6)$$

で表すことにすれば、反復運動が行われたときは2つのベクトル \mathbf{v}_i^{\max} と \mathbf{v}_{i-2}^{\max} が互いに平行になるので、 C_i^{alt} の値が 1 に近づくと考えられる。従って、 i 番目の速度ピークが

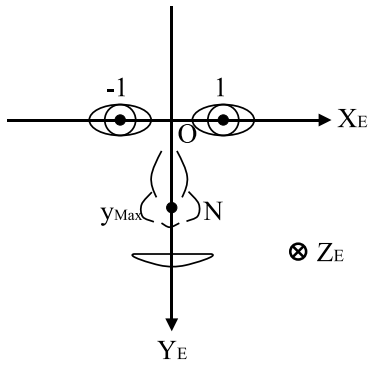
$$C_i^{\text{alt}} > \theta_{\text{repe}}, \quad (7)$$

$$-1 < \theta_{\text{repe}} < 1 \quad (8)$$

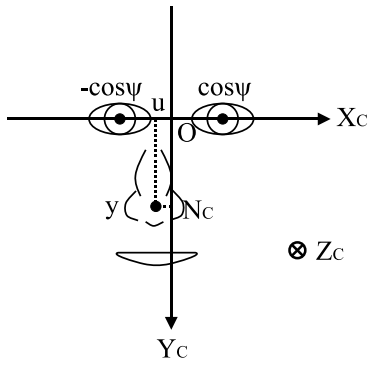
を満たしたとき、反復運動が行われたと判定することができる。ここで、 θ_{repe} は反復運動に対する閾値であり、 C_i^{alt} の値は、完全な反復運動のときに 1 に等しくなる。

頭部姿勢(顔向き)の回転角度もまた、顔中心座標と同様に、頭部ジェスチャ検出のために利用できる。この場合、顔向きの角速度ベクトルが速度ベクトルと同様に扱われ、角速度ピークが往復/反復運動を検出するために用いられる。

顔向きの縦方向の角度 θ と横方向の角度 ψ を両目と鼻の3点の位置から計算することを考える。最



(a)



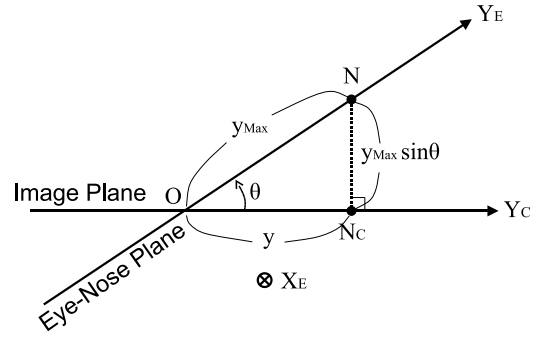
(b)

図3 (a) 目鼻平面座標における顔特徴点の配置. 両目と鼻の位置で正規化されている. (b) カメラ座標における顔特徴点の配置.

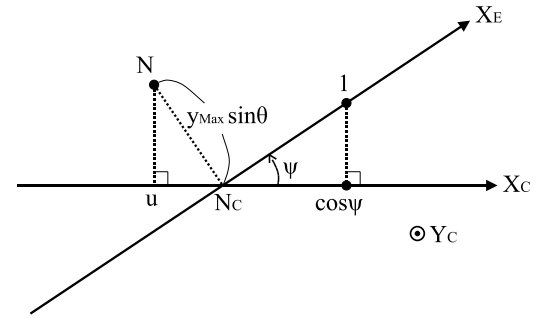
初に図 3(a) のように, 3 点を通る平面 (目鼻平面と呼ぶ) を考え, 両目の座標を $(-1, 0, 0)^T$ と $(1, 0, 0)^T$, 鼻の座標を $(0, y_{\max}, 0)^T$ と正規化して設定し, カメラ座標も最初はこの目鼻平面座標に一致させる. 次に, 目鼻平面を X 軸の回りに $-\theta$ だけ回転させる (図 4(a)), さらに Y_c 軸の回りに ψ だけ回転させる (図 4(b)). このとき, カメラ座標で見た目鼻の配置は図 3(b) のようになる. このとき, 顔向き角度 (θ, ψ) とカメラ座標における鼻の位置 (u, y) との関係は,

$$\begin{aligned} y &= \frac{y_{\max} \cos \theta}{\cos \psi}, \\ u &= -y_{\max} \sin \theta \tan \psi \end{aligned} \quad (9)$$

となる. ただし, u と y は, カメラ座標における両目の位置が $(-1, 0, 0)^T$ と $(1, 0, 0)^T$ になるように



(a)



(b)

図4 目鼻平面座標とカメラ座標との関係. (a) 最初は画像平面と目鼻平面が一致しており, まず目鼻平面を X 軸の回りに $-\theta$ だけ回転させる. (b) 次に, 目鼻平面を Y_c 軸の回りに ψ だけ回転させる.

再度正規化している. 上式を変形すると,

$$\theta = \arcsin \left[\sqrt{\frac{-b + \sqrt{b^2 + 4u^2}}{2y_{\max}^2}} \right], \quad (10)$$

$$b = y^2 - y_{\max}^2 + u^2, \quad (11)$$

$$\psi = -\arctan \left(\frac{u}{y_{\max} \sin \theta} \right) \quad (12)$$

となり, この式を用いることにより, 目鼻の 3 点の画像座標 (カメラ座標) から顔向き角度 (θ, ψ) を計算することができる.

頭部ジェスチャの検出精度は連続した往復/反復運動を検出することによって向上させることができる. これは, 人間の傾き/横振り動作は繰り返されることによる. 実用的には, 顔中心座標と顔向き角度の両方を用い, 連続した往復/反復運動の検出を行うことで性能向上がはかれる (図 5).

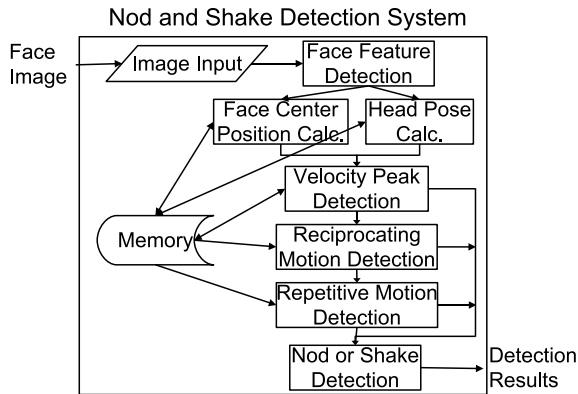


図5 頷きと横振り検出システム.



図7 顔と顔特徴点の検出の例.

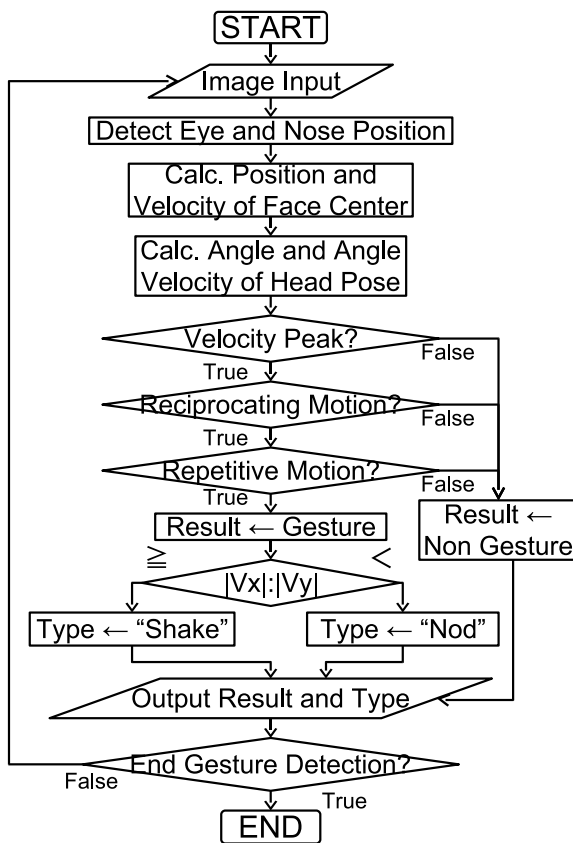


図6 頷きと横振り検出アルゴリズムの流れ図.

3 頷きと横振り検出システム

図5, 6はそれぞれ頷きと横振り検出システムのブロック図と流れ図である.

システムは、顔特徴点検出部、顔中心及び顔向き計算部、速度ピーク検出部、往復/反復運動検出部、頷き/横振り検出部から成る.

頭部ジェスチャ検出のアルゴリズムの全体を図6

に示す. まず最初に、各画像フレームから顔領域を検出する. 顔と顔領域検出アルゴリズムは、形状抽出とパターン照合の組み合わせに基づく方法 [10] を用いる (図7). この手法によって取得された目と鼻孔の位置は以下の処理で利用される.

顔中心の速度と顔向きの角速度は特徴点の位置から計算される. 速度ピークと往復/反復運動が検出されると、システムは頭部ジェスチャが検出されたと判定するが、ジェスチャの種類はこの段階ではまだ判っていない.

頭部ジェスチャが検出されたら、速度ベクトルの向きが調べられてジェスチャの種類が決定される. 頷きか横振りかを判定する前に頭部ジェスチャを検出することは本手法の特徴の一つである.

このシステムは Pentium4 3GHz PC と画像キャプチャボード (PCI) を使用して、画像キャプチャと表示を含めて1秒間に20フレームを処理することができる.

4 実験

2つの画像系列に対する頭部ジェスチャ検出の結果を図8, 9に示す. 処理画像は、横320画素、縦240画素で30フレーム/秒のモノクロ画像である. この実験はオフラインにて行った. 図において、頷きと横振りが検出された時刻を1段目にそれぞれ“●”と“▲”で示した. 横軸はフレーム番号を表す. 2段目に顔中心位置を示し、様々な頭部運動が行われていることも明示した. 3段目には顔向き角度を示した. 速度ベクトルの長さや角速度ベクトルの長さは4段目に示されている. 最も下の2段には、式(3), (6)

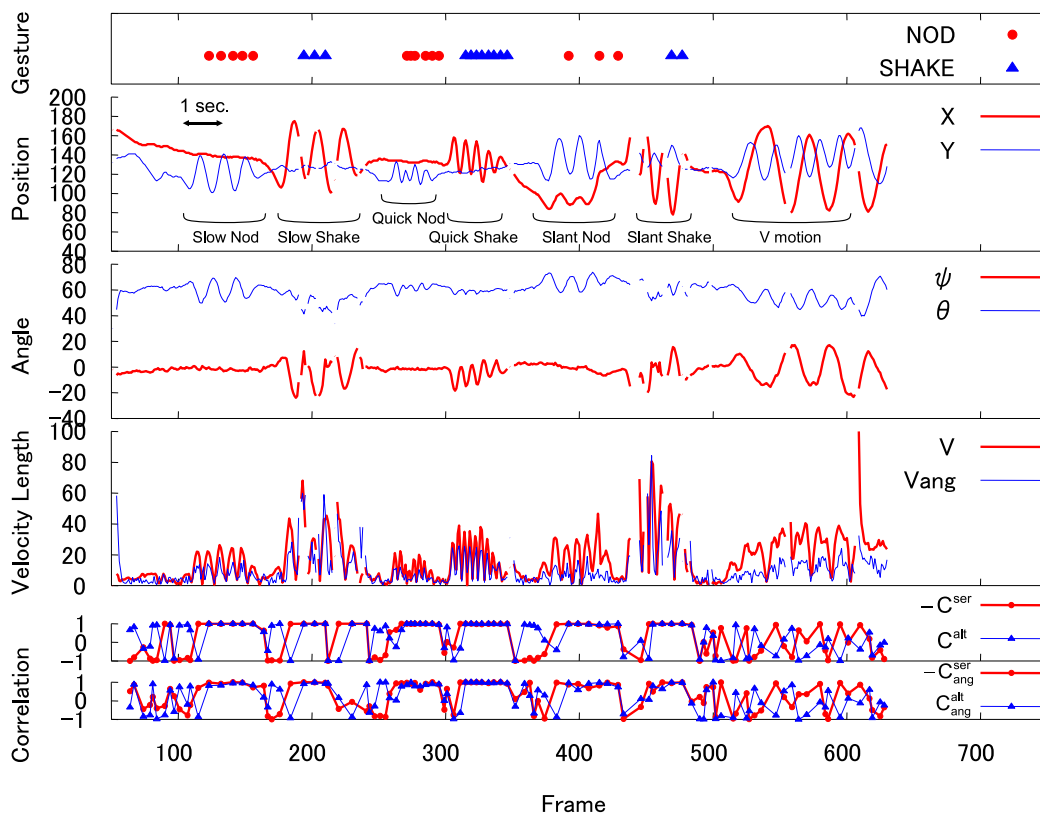


図8 様々なスピードと向きに対する頭部ジェスチャ検出の結果。頷きと横振りが検出された時刻をそれぞれ“●”と“▲”で示した。横軸はフレーム番号を表す。2段目には顔中心位置座標を示し、様々な頭部運動が行われた区間を図示した。3段目には顔向き角度を示した。速度ベクトルの長さや角速度ベクトルの長さは4段目に示されている。最も下の2段には、式(3)、(6)で定義された相関値 $-C_i^{ser}$, C_i^{alt} , $-C_{ang,i}^{ser}$, $C_{ang,i}^{alt}$ を示す(全てが1に近ければジェスチャが検出されるように符号を調整)。添え字“ang”は角速度に関する値を示す。頭部ジェスチャは4種類全ての相関値のグラフの値が1に近い場合に検出される。2段目から4段目までの全ての曲線において、線が途切れている部分は顔検出が失敗していることを示す。

で定義された相関値 $-C_i^{ser}$, C_i^{alt} , $-C_{ang,i}^{ser}$, $C_{ang,i}^{alt}$ を示した。添え字“ang”は角速度に関する値を示す。

頭部ジェスチャは4種類全ての相関値が1に近い場合に検出される。全ての曲線において、線が途切れている部分は顔検出が失敗していることを示す。V字運動(V-shaped motion)、円運動(circle-shaped motion)、はずむ運動(bouncing)、平行移動(transverse shift)は、4つの相関値のうち少なくともどれか一つが小さい値であるため、頭部ジェスチャとしては検出されない。

顔検出は、横振り動作の際、(特に横向き角度が大きいところで)ときおり未検出となる。しかしながら、本手法は速度ピークのみを用いており、通常は顔向き角度が小さいところに運動の速度の極大が観測

されるため、この場合の顔未検出の悪影響を回避できる。大きく横を向くような大きな横振り動作の検出に対しても本手法は有効であると考えられるのである。

5 むすび

速度ベクトルのピーク間の相関を用いて頭部ジェスチャ検出を行う新しいアルゴリズムを提案した。速度ピーク相関は頭部ジェスチャにおいて観測される往復/反復運動を効果的に検出する。画像から顔を検出し、頭部の頷きと横振りを検出するシステムを構築し、アルゴリズムの有効性を確認した。

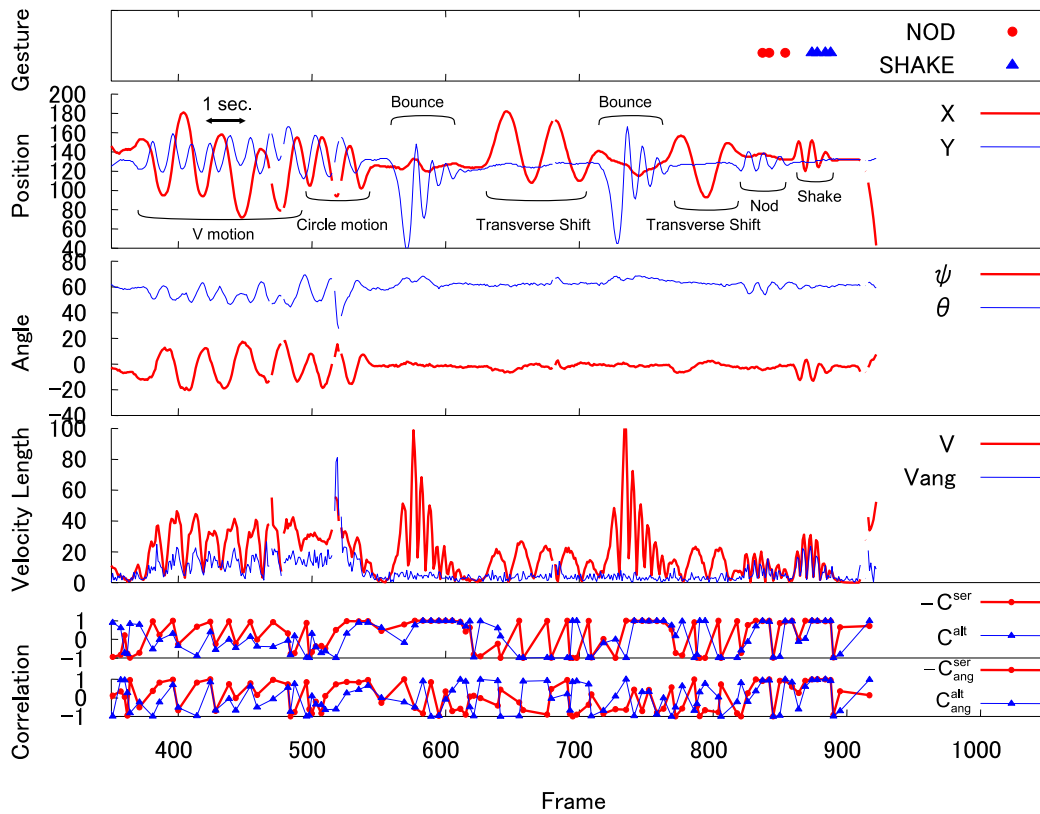


図9 図8と同じ検出パラメタによる様々な頭部運動に対する頭部ジェスチャ検出の結果。V字運動 (V-shaped motion), 円運動 (circle-shaped motion), はずむ運動 (bouncing), 平行移動 (transverse shift) は、4つの相関値のグラフのうち少なくともどれか一つが小さい値であるため、頭部ジェスチャとしては検出されない。

参考文献

- [1] Morimoto, C., Yacoub, Y. and Davis, L.: Recognition of Head Gestures using Hidden Markov Models, in *Proc. 13th Int'l. Conf. on Patt. Recog. (ICPR)*, Vol. III, pp. C-461, Vienna, Austria (1996).
- [2] Zelinsky, A. and Heinzmann, J.: Real-Time Visual Recognition of Facial Gestures for Human-Computer Interaction, in *Proc. 2nd Int'l. Conf. Automatic Face and Gesture Recognition (FG'96)*, pp. 351-356, Killington, Vermont (1996).
- [3] Zelinsky, A. and Heinzmann, J.: A Novel Visual Interface for Human-Robot Communication, *Advanced Robotics*, Vol. 11, No. 8, pp. 827-852 (1998).
- [4] Kapoor, A. and Picard, R. W.: A Real-Time Head Nod and Shake Detector, in *Proc. 2001 Workshop on Perceptive User Interfaces (PUI2001)*, Orlando, Florida (2001).
- [5] Ng, P. C. and De Silva, L. C.: Head Gestures Recognition, in *Proc. IEEE Int'l. Conf. on Image Processing (ICIP-01)*, Vol. 3, pp. 266-269, Thessaloniki, Greece (2001).
- [6] Davis, J. W. and Vaks, S.: A Perceptual User Interface for Recognizing Head Gesture Acknowledgements, in *Proc. 2001 Workshop on Perceptive User Interfaces (PUI2001)*, Orlando, Florida (2001).
- [7] McGlaun, G., Althoff, F., Lang, M. and Rigoll, G.: Robust Video-Based Recognition of Dynamic Head Gestures in Various Domains — Comparing A Rule-Based

- and A Stochastic Approach, in *Proc. 5th Int'l. Workshop on Gesture and Sign Language based Human-Computer Interaction (GW2003)*, Genova, Italia (2003).
- [8] Wu, H., Shioyama, T. and Kobayashi, H.: Spotting Recognition of Head Gestures from Color Image Series, in *Proc. 14th Int'l. Conf. on Patt. Recog. (ICPR)*, Vol. I, pp. 83-85, Brisbane, Australia (1998).
- [9] Kawato, S. and Ohya, J.: Real-Time Detection of Nodding and Head-Shaking by Directly Detecting and Tracking the "Between-Eyes", in *Proc. Int'l. Conf. on Automatic Face and Gesture Recognition (FG2000)*, Vol. IV, pp. 40-45, Grenoble, France (2000).
- [10] Yamaguchi, O. and Fukui, K.: "Smartface"-A Robust Face Recognition System under Varying Facial Pose and Expression, *IEICE Trans. Inf. & Syst.*, Vol. E86-D, No. 1, pp. 37-44 (2003).