

視聴覚環境とのインタラクションのための視線追跡

森山 剛^{*1} 金出武雄^{*2} Jeffrey F. Cohn^{*3}

Gaze Tracking for Interactive Control of Hypermedia System

Tsuyoshi Moriyama^{*1} Takeo Kanade^{*2} Jeffrey F. Cohn^{*3}

Abstract: This paper describes an automated gaze tracking method that enables interactive control of audiovisual system. The method utilizes a meticulously detailed eye region model that parameterizes both the structure and the motion of an eye region. We train Active Appearance Models (AAMs) for parameterizing the variance of individuality in structure of eye region. Assuming the initial frame contains only neutral faces, the system projects the given eye region image onto the low dimensional subspace spanned by the AAM and retrieves the structure parameters of the nearest training sample and initializes the eye region model using them. Once the eye region model is individualized in terms of the structure parameters, the system tracks the motion parameters frame by frame for the rest of the image sequence. Experimental results show the proposed method achieves accurate tracking of a comprehensive set of eye motions.

1. はじめに

顔画像列に対して視線方向を追跡するためには、視線以外の要因をも検出するために、目領域の表情筋の運動全体を追跡する必要がある。視線の追跡を行う問題を困難にしているのは図1に示すような目領域画像の多様性である。これは主に個性に起因する静的な成分と、目領域の表情筋の運動に起因する動的な成分とに分けられると考えられる。従来、モデルに基づいた目の解析では、まぶたと瞳に対してそれぞれ放物線と円をモデルとして使うのが代表的であったが、一重（ひとえ）や二重等の個性を吸収するのに十分ではなかった¹⁾。主成分分析等を用いた統計モデルに基づいた目の解析では、個人差は吸収できるものの、学習データの制約から目の各部位の運動を詳細に記述することはできなかった²⁾。

そこで筆者らは、目領域画像の多様性を表現する高精細な目領域モデル、及びそのモデルを用いた目とその周辺の詳細な動き解析システムを提案した³⁾。本目領域モデルは、目領域の静的な構造と表情筋の運動による動きを、それぞれ構造パラメータ、運動

パラメータによって記述し、頭部運動追跡システムと組み合わせることにより、目領域の動きをロバストに追跡することに成功した。

表情を伴う目は、それが平静の見え（元々の構造）であるか、表情（運動）を伴った見えであるかが区別できないため、初期の実装では、画像列の第一フレームの目領域を平静である（いかなる表情筋の緊張をも伴わない）と仮定し、構造モデルの初期化を手動で行っていた。

本論文では、Active Appearance Model (AAM)⁴⁾を用いて、平静な目領域の見えを低次元で表現し、その低次元空間において入力に最も近い学習サンプルを求め、その学習サンプルにあらかじめ付与された構造パラメータを入力目領域の構造パラメータとして自動的に決定する手法を用い、視線追跡を自動で行う方法を提案する。

2. 目領域モデル

図2に目領域モデルを示す³⁾。本モデルは、複数のレイヤに目の各部品に対応した部品モデルを描き、それら全てを重畳して一つの目領域を作る。部品モデルとして、一番下のレイヤから白目 (sclera)、

^{*1} 東京工芸大学工学部メディア画像学科助教 ^{*2} カーネギーメロン大学ロボティクス研究所ワイタカー記念全学教授

^{*3} ピッツバーグ大学心理学科教授 E-mail: moriyama@mega.t.kougei.ac.jp

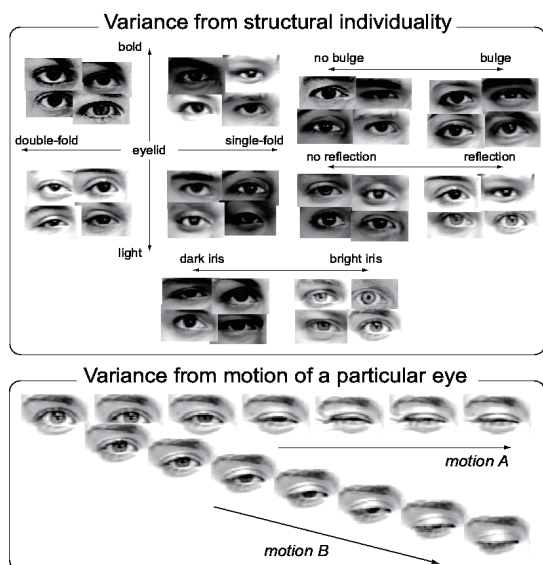
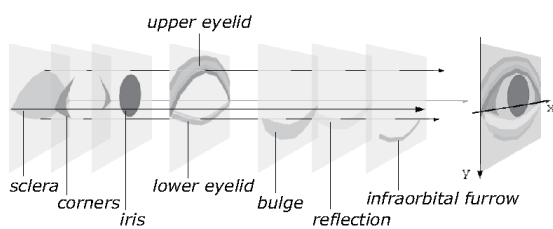


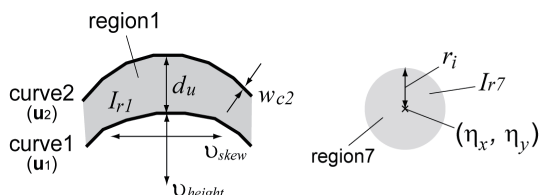
図1 目領域画像の多様性

左右コーナ (corners), 瞳 (iris), 上下眼瞼 (まぶた; upper eyelid and lower eyelid), 目下部のふくらみ (bulge), ふくらみ上の明るい領域 (reflection), そして, ふくらみ下部のしわ (infraorbital furrow) を持つ. それぞれ矩形のレイヤ内において, 部品の描かれる画素は不透明なある明度を持ち, 描かれない画素は透明である. そしてレイヤが重畳される際には, 目領域モデル上のある2次元座標について, 下位レイヤに部品の描かれた画素は, 上位レイヤでその画素に描画が起こる場合はその描画によって隠され, 全ての上位レイヤで描画が起こらない場合は目領域モデル上に現れる. 例えば, 上下眼瞼レイヤは上下眼瞼を表す二つの曲線から成り, 曲線の間の領域 (眼瞼裂) が透明であるのに対して, 上眼瞼の上側は二重 (ふたえ) を含む構造で塗りつぶされ, 下眼瞼の下側も肌色で塗りつぶされる. これが瞳レイヤの上位レイヤとして重畳されると, 瞳として描画された円盤領域のうち, 上下眼瞼に挟まれた部分のみが目領域モデルに現れ, その他は隠される.

本目領域モデルは, 目領域の詳細な構造を表す6個の構造パラメータ $\mathbf{s} = \{d_u, f, d_b, d_r, I_r, r_i\}$ と, 目領域の詳細な運動を表す5個の運動パラメータ $\mathbf{m} = \{v_{\text{height}}, v_{\text{skew}}, \lambda_{\text{height}}, \eta_x, \eta_y\}$ ($v_{\text{height}}, v_{\text{skew}}$ は上眼瞼の上下動と左右の歪み, λ_{height} は下眼瞼の上下動, η_x, η_y は瞳の中心位置) によって制御される. 以下に, 目領域の各部位の構造とそれらのモデルについて述べる.

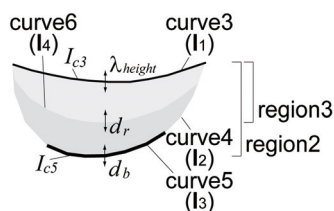


(a) 各部品をレイヤに分けて描画



(b) 上眼瞼のモデル

(c) 瞳のモデル



(d) 下眼瞼のモデル

図2 マルチレイヤ構造の目領域モデル

2.1. 上眼瞼の構造とそのモデル

上眼瞼 (upper eyelid) は, 眼瞼裂 (palpebral fissure) の上部の皮膚であり, 眼輪筋 (orbicularis oculi muscle) が皮下脂肪で覆われた構造をしている. 眼輪筋は, 上から眼窩部 (orbital orbicularis), 隔膜前部 (preseptal orbicularis), 瞼板前部 (pretarsal orbicularis) の三つの部分に分かれ, さらに瞼板前部が上部と下部に分かれる. 開瞼 (目を開くこと) の際には, 瞼板前部の上部と下部の境界で折りたたまれて筋 (crease, 上眼瞼溝; sulcus suprapalpebralis) ができる.

この上眼瞼溝 (二重のしわ) は, 上眼瞼挙筋の腱膜 (上眼瞼を引き上げる腱; levator aponeurosis) と眼窩隔膜 (orbital septum) の結合する位置によってその上下の位置が決まる. その位置が低い場合, 東アジア人に特徴的な一重まぶた (もしくは奥二重: 東アジア人全体のおよそ 50%) になり, 位置が高い場合は幅の広い二重まぶたになる. 上眼瞼溝の深さは, 上眼瞼下部の瞼板 (superior tarsal plate) と皮膚の間の繊維性の結合強度と, 結合に関与しない上部

の皮膚と皮下脂肪の量によって決まる。結合強度がより強く、かつ皮膚と皮下脂肪の量が多い場合、上眼瞼挙筋 (levator muscle) が上眼瞼を上引き上げる際、下部の皮膚、眼輪筋及び瞼板が一つの塊としてより高く引き上げられ、より多くの皮膚がその上にせり出して、上眼瞼溝が深く“濃い”二重となる。

すなわち、上眼瞼の見えの多様性を与える構造には、上眼瞼溝の位置と濃さの二つがあると考えられる。著者らの目領域モデル³⁾では、上眼瞼の下辺と上眼瞼溝を共に折れ線分でモデル化した上で (curve1 及び curve2)、両折れ線分間の距離 (d_0) としわの濃さ (f) の二つのパラメータを上眼瞼の構造情報とする。パラメータ f は、両折れ線分で囲まれた領域 region1 の明るさ I_{r1} と curve2 の太さ w_{c2} を、 $I_{r1}=I_{r1}^{\text{brightest}} \cdot \beta_1 \cdot f$ 及び $w_{c2}=\beta_2 \cdot f + w_{c2}^{\text{thickest}}$ (β_1, β_2 は定数) のように与える (図 2(b))。

2.2. 下眼瞼の構造とそのモデル

下眼瞼 (lower eyelid) は、下眼瞼上辺の瞼板 (inferior tarsal plate) が、目の両端点の腱 (内端点: medial canthal tendon, 外端点: lateral canthal tendon) を支点にして、眼瞼裂の下部を覆う皮膚を吊り下げたハンモックのようになっており、その瞼板は、眼窩隔膜 (orbital septum) によって眼窩 (眼球を入れる頭蓋骨のくぼみ) の縁 (orbital rim) につながっている。皮下の眼輪筋の長さが収縮して皮膚が盛り上がったたり、加齢によって弛んだりすると、下眼瞼に膨らみが生じ、眼窩下溝 (頬の上にある斜めのしわ, infraorbital furrow) がより深くなる。また、膨らみの上下幅には生まれつきの個人差もある。

すなわち、下眼瞼の見えの多様性を与える構造には、この膨らみの幅や高さ (奥行き方向) があると考えられる。著者らのモデルでは、下眼瞼の上辺と眼窩下溝を共に折れ線分でモデル化した上で (curve3 及び curve4)、両線分間の距離 (d_b) をパラメータとしている (両折れ線分に挟まれた領域を region2)。また、照明がこの膨らみに反射して周辺より明るくなる場合も考慮し、その領域の下辺を折れ線分でモデル化し (curve6)、curve6 と curve3 の距離 (d_r) も考慮する (これらに囲まれる領域を region3)。膨らみの幅 d_b と照明反射による明るい領域の幅 d_r の二つのパラメータを下眼瞼の構造情報とする (図 2(d))。

2.3. 眼瞼裂の構造とそのモデル

眼瞼裂 (palpebral fissure) は、上眼瞼と下眼瞼に挟まれた開口部である。眼瞼裂から見える部位は、大きく二つの部分、すなわち虹彩 (iris) 及び強膜 (白目; sclera) に分けられ、強膜の鼻側の内端点には涙器による赤褐色の領域がある。虹彩の大きさは、眼球の大きさと共に年齢差及び個人差がある。虹彩の中心には、黒色の空洞である瞳孔 (pupil) があり、網膜 (retina) に届く光の量によって大きさが変化する。虹彩の色は、色素の量によって異なり、色素が最も多い場合は茶になり、最も少ない場合は青になる。また色素が少ない場合には、虹彩の模様がより明瞭に見える。虹彩には、角膜 (cornea) 及び水晶体 (lens) の反射像 (ブルキン像) が映る。

すなわち、眼瞼裂の見えの多様性を与える構造には、主に虹彩の大きさと内部の模様の二つの要因があると考えられる。著者らのモデルでは、虹彩内部の模様については考慮せず、虹彩を一つの明度値を持つ円盤としてモデル化した上で、その大きさ (r_i : 半径) と明度値 (I_{r7}) の二つのパラメータを眼瞼裂の構造情報とする (図 2(c))。

3. 目領域モデルを用いた視線追跡

目画像解析の対象は、人物の表情表出を記録した画像列である。通常、表情は自発的な頭部運動を伴うため、顔画像の変動は剛体運動 (頭部運動) と顔表面の非剛体運動 (表情) が合成されたものになっている。そこで、頭部運動の追跡を行い、全ての顔領域画像を特定の頭部姿勢へ投影することにより、ほぼ表情のみによる変動を抽出する。抽出された表情のみの変動を含む顔領域画像に対して、本目領域モデルを用いた目領域画像解析を行う。図 3 に示すように、初期フレームで目領域モデルを目領域画像に位置合わせし、構造パラメータ s を推定して目領域モデルを個人化する。次に頭部運動正規化画像列全てにわたって運動パラメータ m_i を追跡する。

3.1. 頭部運動追跡

3D円筒モデルを用いて頭部運動の追跡を行う³⁾。初期フレームで円筒姿勢と顔特徴点 (目の端点を含む56点) を手入力し、残りのフレームは自動で頭部追跡を行う。

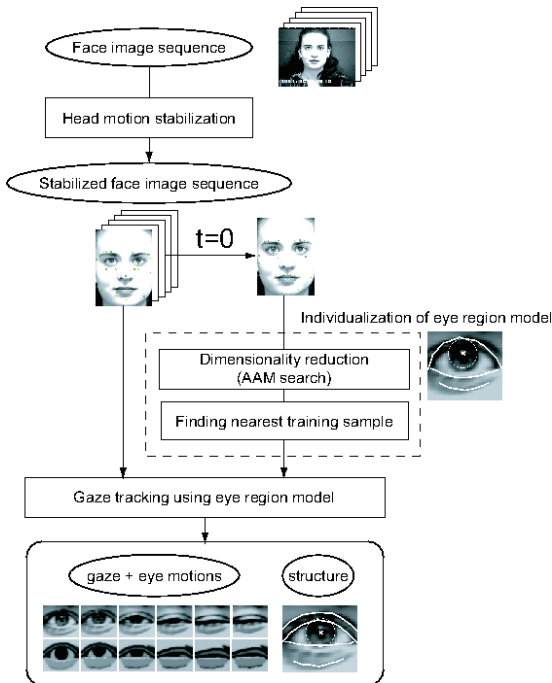


図3 目領域モデルを用いた視線追跡

3.2. 目領域モデルの個人化

第一フレームの目領域画像から目領域モデルの構造パラメータを推定する。本手法では、直接画像から推定する代わりに、あらかじめ構造パラメータが既知の学習サンプル（平静の目領域画像）のうち、入力目領域画像に最も似たものを探し、その学習サンプルに付与された構造パラメータを用いて目領域モデルを個人化する。またこの時、画像同士を比較する代わりに、学習サンプルに手入力した特徴点を基に AAM を学習しておき、AAM による低次元空間に投影して高速に比較を行う。AAM とは、形状とテクスチャを、低次元の主成分の線形和で表現する手法である。本手法では、目領域を図4の3つの小領域に分けて AAM を学習することによって、各部位の詳細な構造が AAM の次元に反映されるようにしている。AAM 探索では、まゆげと目の両端点計4点を固定して行う制約付き AAM⁴⁾を用いる。

3.3. 目の運動追跡

目の運動追跡は、既に求まっている構造パラメータ s と、運動パラメータ m_t の初期フレームの値を

(a) 上眼瞼（まゆげの下端5点と上眼瞼下端8点の計13点）

(b) 眼瞼裂（上眼瞼下端8点と下眼瞼上端11点の計19点）

(c) 下眼瞼（下眼瞼上端11点とそれに平行な点11点の計22点）

図4 各小領域における特徴点と三角パッチ

基に、残りのすべてのフレームについて、常に前フレームの値を初期値として現フレームの値を最急勾配法で推定する。各フレーム t における運動パラメータ m_t の推定は、次の(1)式の評価関数 D を最小化することで実現する。 p_t は目領域モデル座標系から目領域画像座標系への変換 W のパラメータであり、ここでは平行移動のみを含む。

$$D = \sum [T(x; m_t + \delta m_t) - I(W(x; p_t + \delta p_t))]^2 \quad (1)$$

(1)式をテイラー展開して、 δm_t 及び δp_t に関する連立方程式を解くことによりこれらを求め、次の(2)式で解を更新する。

$$m_t \leftarrow m_t + \delta m_t, \quad p_t \leftarrow p_t + \delta p_t \quad (2)$$

いずれかのパラメータの更新量が、あらかじめ定めた閾値を下回ったら解の更新を止め、次フレームに進む操作を繰り返すことで、全てのフレームについて運動パラメータ m_t を推定する。

4. 実験

提案した手法による目領域画像解析を、2つの顔表情データベース（Cohn-Kanade AU-coded Facial Expression Image Database⁵⁾ 及び Ekman-Hager Facial

表 1 多様な上眼瞼に対する追跡結果の例

(a) 一重	(b) 二重	(c) まつげの濃い場合	(d) 奥二重

表 2 多様な眼瞼裂に対する追跡結果の例

(a) 明るい瞳	(b) 暗い瞳

表 3 多様な下眼瞼に対する追跡結果の例

(a) ふくらみ	(b) ふくらみに反射

Action Exemplars⁶⁾）に対して行った。両データベースには併せて 576 画像列（C-K 490，E-H 86），9530 フレームが含まれ，被験者は 118 名で性別，民族，年齢が多岐に渡っている。

実験結果は，図 1 に示した目領域画像の多様性要因のそれぞれに関して，追跡結果を目視で評価すると共に，画像解像度，先鋭度，目領域モデルの詳細度，目領域モデル個人化の精度について定量評価を行った。AAM を用いた自動個人化手法については，手動の方法と動き追跡結果を比較した。

4.1. 頭部運動追跡結果

頭部運動のカラー画像列に対し回転及び並進 6 自由度全てについて追跡を行い，合成画像及び実画像を用いた実験において，yaw 角及び pitch 角がそれ

ぞれ最大 75 度及び 40 度まで変動した場合でも，誤差 3 度以内で頭部追跡を行うことができた。

4.2. 動き追跡の例

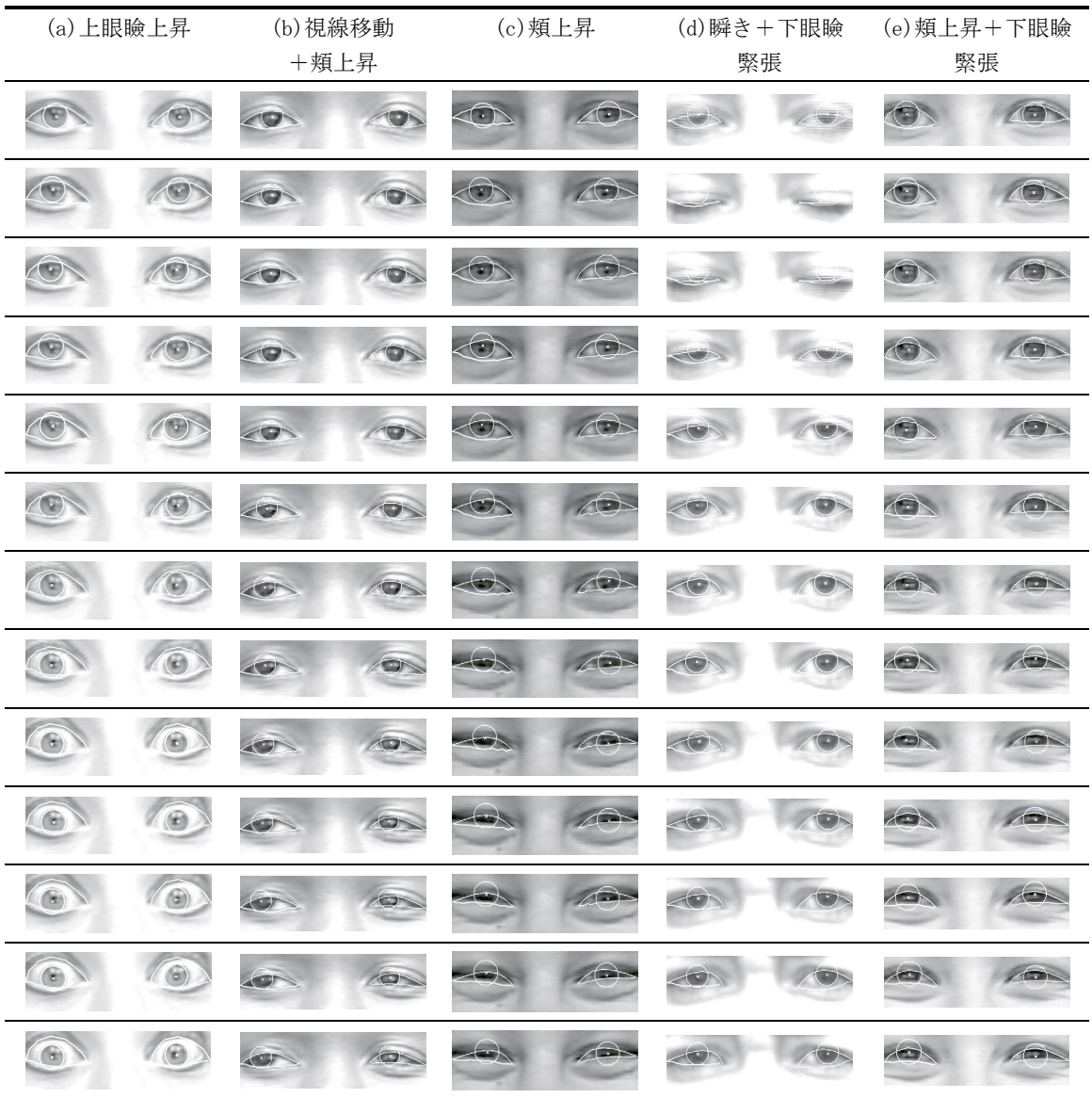
4.2.1. 多様な上眼瞼に対する結果

従来の手法で最も起こりやすかった追跡誤りは，二重の場合に，上眼瞼の二重のしわを眼瞼裂との境界線と誤って追跡してしまう場合であった。本手法では表 1 に示すように，一重，二重，まつげの濃い場合，奥二重の場合でも正しく追跡できている。

4.2.2. 多様な眼瞼裂（瞳）に対する結果

明度の高い瞳は，目領域の他の場所（例えば目頭）にある影と誤追跡されやすいが，表 2 に示すように正しく追跡されている。

表 4 多様な運動に対する追跡結果の例



4.2.3. 多様な下眼瞼に対する結果

下眼瞼の（眼瞼裂との境界線の）追跡は、ふくらみ下の斜めのしわと誤ることが多く、特に下眼瞼のふくらみが明るい場合、ふくらみと斜めのしわで作るパターンと、白目と下まぶたで作るパターンが酷似し、追跡が困難であった。表3に示すように、本モデルでは目領域全体を精細にモデル化することで、このような追跡誤りなく正しく追跡できている。

4.2.4. 多様な動きに対する結果

従来最も誤りやすかったのは、瞬き等によって上下まぶたが接近した時や、まぶたの形状の変形が大きく単純なモデルで表現できない時であった。本モデルでは、画素ドメインではなくモデルパラメータドメインで追跡を行っているため、表4に示すように、上下まぶたが近接しても正しく追跡がされている。また上まぶたの水平歪みのモデル化により、まぶた形状の変形も十分に表現できている。表4の結果は FACS⁷⁾で定義されている AU (Action Units)のうち、(a)AU5, (b)AU6+62, (c)AU6, (d)AU45 と AU7,

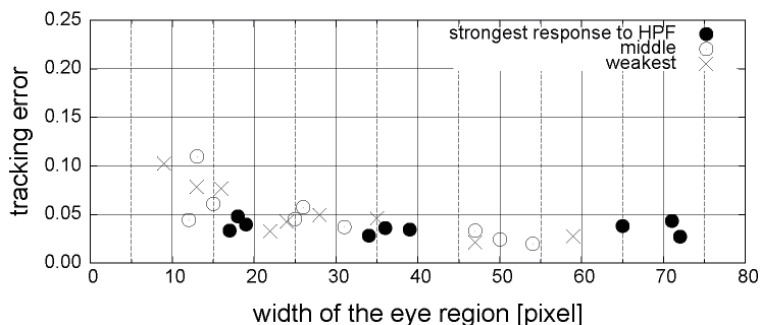


図 5 入力画像のサイズに対する感度（上眼瞼の例，下眼瞼及び眼瞼裂もほぼ同様の結果）

表 5 異なる解像度に対する計算時間（Pentium M, 1.6GHz, 768MB RAM, Windows XP, 10 回の平均）

		元解像度 (59 画素)	半分 (28 画素)	四分の一 (16 画素)
目領域画像				
目領域モデル (元のスケールで個人化)				
計算時間	目領域モデル生成	13.1ms	6.6ms	5.6ms
	入力画像のモデル上への投影	2.2ms	0.4ms	0.1ms
	LK 法によるモデルの更新	134.0ms	77.9ms	61.3ms
	計	149.3ms	84.9ms	67.0ms

(e)AU6+7 に該当し，目領域で起こり得る動きについて，いずれも正しく追跡された。

4.3. 定量評価

目の輪郭上の点を手入力したものと，システムの追跡結果とのずれを評価した。

4.3.1. 入力画像のサイズと先鋭度に対する感度

入力画像のサイズが小さい場合や入力画像がぼけている場合，目領域の構造を決定するのに十分な情報が得られない可能性がある。この時の本手法の性能を評価するために，データベースから HPF (High Pass Filter) の応答の最大のもの，最小のもの，それらの中間のものを取り出し，さらに画素を間引いてサイズを 3 段階に変えた画像を生成した。これ

らに対する追跡結果を図 5 及び表 5 に示す（目領域の両端点を元に切り出しているため，画像サイズは半分，四分の一丁度にならない）。その結果，目領域画像の幅が 15 画素程度までは安定して追跡を行えることがわかった。

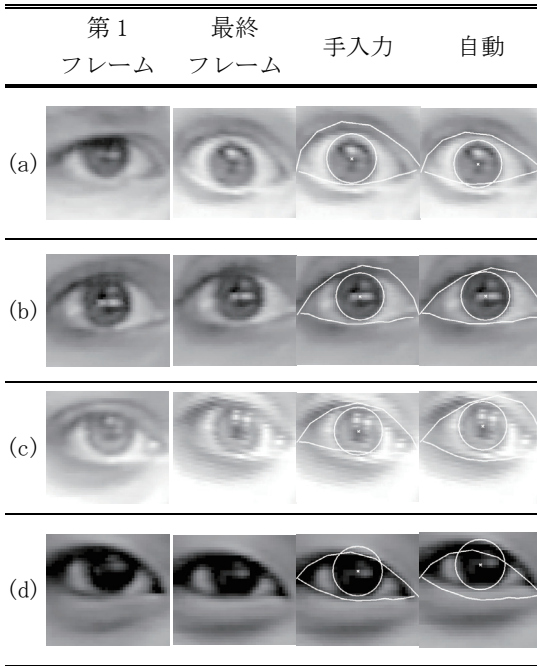
4.3.2. 目領域モデルの詳細度の効果

上眼瞼の二重のモデルなしでは二重の追跡を失敗し，下眼瞼のふくらのモデルなしでは下眼瞼の追跡を失敗した。全てのパーツが多様な目領域の動き追跡に必要であることが示された³⁾。

4.3.3. 目領域モデルの初期化精度に対する感度

目領域モデルを個人化し，パラメータ値を一つずつ増減させて，追跡結果の安定性を調べた。その結

表6 手入力による目領域モデルの個人化と
自動個人化による追跡結果の比較



果、初期化の精度に敏感であること、モデルパラメータが互いに相関を持つことが示された³⁾。

4.4. AAM を用いた目領域モデルの自動個人化による追跡結果

生成された AAM は、小領域ごとにおよそ 20 次元であった。表 6 に示すように、手動で目領域モデルを個人化した場合と、ほぼ変わらない精度で動き追跡を行うことができた。表 6(c) で上眼瞼がずれているのは、目を見開くと同時に上眼瞼が次第に明るくなっているのに対して、本目領域モデルがこれを表現するパラメータを持たないためである。表 6(d) で上眼瞼がずれているのは、学習サンプルに一重まぶたの被験者が含まれなかったためである。

5. まとめ

視聴覚環境における、自動視線追跡を用いた対話型インタフェースを目指して、目領域の動き解析を行う手法を提案した。目領域の構造の個人差や動きの多様性を表現できるモデルを開発し、頭部運動追跡システムと組み合わせて、目領域の動きを追跡す

ることに成功した。様々な目領域の構造や動きについて追跡結果を評価すると共に、目領域モデルを個人化するために、目領域の構造を自動的に取得する拡張を行った。今後は、自動の方法について定量評価を行うと共に、実際のマルチメディア視聴環境でリアルタイムかつオンラインで動作するシステムへと拡張する。

謝辞

本研究は、米国国立精神保健研究所補助金 R01 MH51435 の支援により行った。

参考文献

- 1) Tian, Y. and Kanade, T. and Cohn, J. F., "Eye-state detection by local regional information", Proceedings of the International Conference on Multimodal User Interface, pp.143-150, October, 2000.
- 2) Blanz, V. and Vetter, T., "A Morphable Model For The Synthesis Of 3D Faces", SIGGRAPH '99, pp.187-194, 1999.
- 3) Tsuyoshi Moriyama, Takeo Kanade, Jing Xiao, Jeffrey F. Cohn, "Passive Eye Monitoring", chapter 2, pp.17-40, Springer, March 2008.
- 4) T. F. Cootes, C. J. Taylor, "Constrained active appearance models", IEEE International Conference on Computer Vision, Vol.1, pp. 748-754, July 2001.
- 5) Kanade, T. and Cohn, J. F. and Tian, Y., "Comprehensive database for facial expression analysis", IEEE Face and Gesture '00, pp.46-53, 2000.
- 6) Ekman, P. and Hagar, J. and Methvin, C. H. and Irwin, W., "Ekman-Hagar Facial Action Exemplars", (unpublished data), Human Interaction Laboratory, Univ. of California, San Francisco.
- 7) Ekman, p. and Friesen, W., "Facial Action Coding System", Consulting Psychologists Press, Palo Alto, CA, 1978.