

3次元 Active Appearance Model を用いた手形状認識

山下 亮[†] 滝口 哲也^{††} 有木 康雄^{††}

[†] 神戸大学大学院システム情報学研究科 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学自然科学系先端融合研究環 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: [†]ryo@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

あらまし 本研究では、高機能TVなどに対するジェスチャー入力として3次元モデルを使用した、複雑な手の形状認識の手法を提案する。従来のジェスチャー認識では、カメラに対して正面に手を向ける必要があり、任意の手の傾きに対応できない問題点がある。そこで、3D Active Appearance Models を使用することで、あらゆる方位にも対応できる手の形状追跡を実現する。高精度な距離画像センサー Kinect を用いて、対象のRGB画像と深度情報を取得し、モデルの学習及びテストを行った。複数の3D-AAMを使用することにより、複雑な指の形状を、方向の変化に対して頑健に認識することができた。

キーワード ハンドジェスチャー認識, ユーザインターフェース, 3次元モデル, Active Appearance Models, 形状認識

1. はじめに

近年、複雑な動作を入力できる高機能テレビの普及や、不特定多数のユーザが利用する公共施設の電子案内表示板の登場により、コントローラを必要としないジェスチャー入力技術への期待が高まっている [1]。Microsoft 社の [Kinect for Xbox 360] では、ゲームだけでなく手をマウス代わりに使用し、ポインタを操作することや、動画や音楽を操作することも可能である。このように、コントローラが不要になる利点は多くある。手が自由になり、現実世界と同じ動きで直観的にデバイスを操作することが可能になる。ボタン操作を覚える必要もなく、コントローラの電池切れの心配や、紛失の恐れもなくなる。

しかし、高精度な距離画像センサを利用して、やはりコントローラを使用するジェスチャー入力技術と比較すると、認識の際に様々な制約が生じる。認識性能の制約とは、主に、認識可能範囲が限られることと、手首から先の細かな動きといった認識しにくい動作があるということである。そのため、Kinect による操作では、手を動かしてポインタを操作し、所望するアイコンの上でポインタをしばらく静止することで「決定」の操作を行っている。この方法では、「決定」を行うのにある程度の時間を必要とし、また、手を数秒間制止させるという決定法では操作ミスを誘発しやすいという問題もある。

コストの面では光源不要なカメラによるアピアランス情報を利用した画像認識の方が安価に済み、手首から先の指先の形状を認識することができる。しかし、この方法では手をカメラに正面にかざす必要があり、ユーザへの負担につながる。

この点から、本論文では、指先の形状によるジェスチャーを、深度情報とアピアランス情報の両方を考慮することで認識する手法を提案する。距離画像センサによ

り取得した深度情報とアピアランス情報により、手の3次元モデルを複数学習することで、手首から先の複雑な形状を方位の変化等に対して頑健に認識する。

本論文は次のように構成されている。2. で提案手法の概要について述べ、3. で提案手法の詳細について述べる。4. で提案手法に対する有効性を確認する実験を行い、最後に5. で本論文をまとめる。

2. 提案手法の概要

図1は提案システムの流れを示している。まず学習段階では、距離画像センサ Kinect により取得された深度情報とRGB画像から複数個の3次元モデルを学習する。本研究では、3次元モデルとして、Active Appearance Models(AAM) [2] [3] [4] に3次元情報を付加した3D Active Appearance Models(3D-AAM) [5] [6] [7] を使用する。認識段階では、kinect からの入力画像に対して、それぞれの3次元モデルでフィッティングを行う。最終的な認識結果としてモデルと対象との誤差値が最小となる形状を出力する。

3. 提案手法の詳細

3.1 AAM による手の特徴点抽出

本節では、AAM を用いた手の特徴点抽出について述べる。

3.1.1 AAM の概要

AAM とは Cootes らによって提案された手法で、主に顔の特徴点抽出に用いられている。対象を形状 (shape) とテクスチャ (appearance) に分け、それぞれを主成分分析 (Principal Component Analysis: PCA) [8] で次元圧縮することにより、少ないパラメータで対象の形状の変化とテクスチャの変化を表現できるようにしたモデルである。対象の方向の変化に対して、頑健かつ高速に特徴点

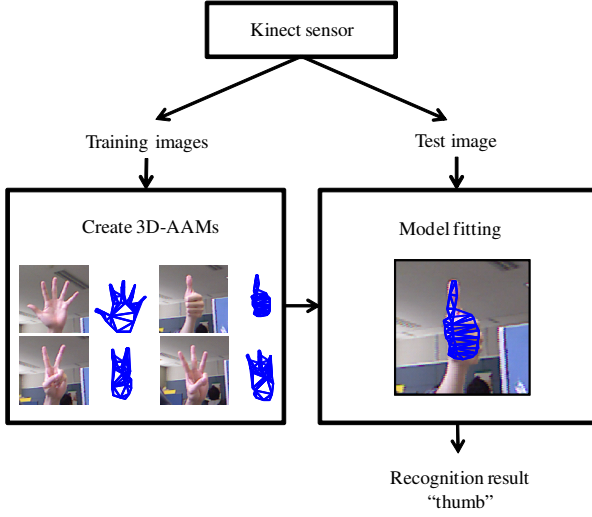


図1 System flow

を抽出することが可能であり、形状とテクスチャの情報を低次元の Combined パラメータで表現することができる。提案手法では shape を 3 次元に拡張した 3D-AAM を手に適用することで、特徴点を抽出する。

3.1.2 AAM の構築

対象となる物体に対して、図. 2 のように任意に特徴点を与える。特徴点となりうる点とは、人間が手動で与えるときに一位に場所を指定できるような輝度値の差の大きい場所となる。対象となる物体の形状を shape と呼

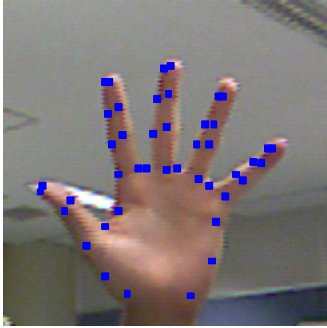


図2 Hand feature points

び、shape 内のテクスチャを appearance と呼ぶ。shape を変形させたとき、appearance は、図3のような三角形領域をそれぞれアフィン変換を用いて変形させることで求められる。特徴点から三角形領域の生成は、ドロネー図を作成することにより行う。i 番目の特徴点座標を $\{x_i, y_i\}$ ($i \leq n$) とすると全特徴点を並べたベクトル \mathbf{x} は、

$$\mathbf{x} = (x_1, y_1, \dots, x_n, y_n)^T \quad (1)$$

と表される。この x を shape ベクトルと呼ぶ。全学習データに対して手動で shape ベクトルを求め、それらに対する平均 shape ベクトル \bar{x} とする。 \mathbf{x} は、 \bar{x} からの偏差を PCA して得られる固有ベクトル行列 P_s を用いて、

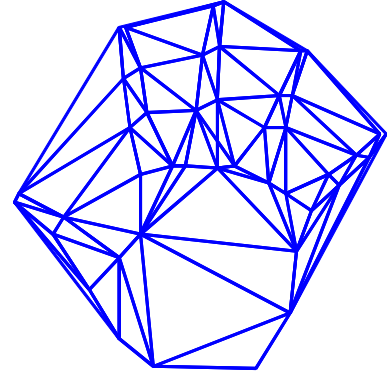


図3 Shape

$$\mathbf{x} = \bar{x} + P_s b_s \quad (2)$$

と表すことができる。 b_s はパラメータベクトルであり、shape パラメータと呼ばれる。

次に、appearance モデルの構築法について述べる。図3でしめしたような shape 内のテクスチャを、三角領域ごとに平均 shape の形状に変形することによって、正規化する。正規化されたテクスチャの輝度値を並べたベクトルを \mathbf{g} とし、appearance ベクトルと呼ぶ。appearance ベクトル \mathbf{g} も式 (2) と同様に、それらを平均した平均 appearance ベクトル \bar{g} からの偏差を PCA し、得られる固有ベクトル行列 P_g を用いて、

$$\mathbf{g} = \bar{g} + P_g b_g \quad (3)$$

と表すことができる。 b_g はパラメータベクトルであり、appearance パラメータと呼ばれる。パラメータベクトル b_g 、 b_s は平均からの変化を表すパラメータであり、これらを変動させることで shape と appearance を変化させることができる。また、shape と appearance に相関があることから、 b_s と b_g をさらに PCA することで、式 (4)、(5) のように表現できる。

$$\mathbf{b} = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (x - \bar{x}) \\ P_g^T (g - \bar{g}) \end{pmatrix} = Qc \quad (4)$$

$$\mathbf{Q} = \begin{pmatrix} Q_s \\ Q_g \end{pmatrix} \quad (5)$$

ここで、 W_s は shape ベクトルと appearance ベクトルの単位の違いを正規化する行列、 Q は固有ベクトル行列、 c は shape と appearance の両方を制御するパラメータで combined パラメータと呼ばれる。 c を用いて x 、 g を表現すると式 (6)、(7) のようになる。

$$\mathbf{x}(c) = \bar{x} + P_s W_s^{-1} Q_s c \quad (6)$$

$$\mathbf{g}(c) = \bar{g} + P_g Q_g c \quad (7)$$

このようにして、パラメータベクトル c を制御することによって、shape と appearance を同時に扱い、対象の変化を表現することが可能となる。

3.2 Pose パラメータ

前項により、対象の形状変化とテクスチャ変化を、combined パラメータ c のみで制御できるようになったが、これらは対象内部の局所的な変化であり、対象が画像中のどこに、どんなサイズで、どんな向きで存在するか、という大域的な変化に関しては考慮していない。そこで、AAM 全体を変化させるような、大域的な姿勢変動に関するパラメータを考慮する。本論文では、それを pose パラメータ q と呼ぶ。 q は、

$$q = [\text{roll} \ \text{scale} \ \text{trans}_x \ \text{trans}_y] \quad (8)$$

と表現される。ただし、 roll はモデルの画像平面に対する回転角度、 scale はモデルのサイズ、 $\text{trans}_x, \text{trans}_y$ は、それぞれ、 x, y 軸方向のモデルの平行移動量を表す。

3.3 AAM の探索

AAM の探索とは、AAM を combined パラメータ c と pose パラメータ q で局所的、大域的に変化させて、対象画像を生成し、それと入力画像を比較して、誤差が最小となるような c と q を見つけることである。具体的には、ある c' と q' に対して、shape x を c' から得られる shape パラメータ b_s' と pose パラメータ q' により変形する関数を $W(x; q', b_s')$ とし、入力画像 Img と、shape x を与えたときに、shape 内の輝度値を求める関数を $I(Img, x)$ とすると、誤差値 E は、

$$E = [(\bar{g} + P_g Q_g c') - I(Img, W(x; q', b_s'))]^2 \quad (9)$$

で表される。この E を最小とするような、最適なパラメータ c^*, q^* を最急降下法によって求める [10]。ここで、探索パラメータを $p = (c', q')$ とまとめると、誤差値 E は

$$E(p) = \|r(p)\|^2 \quad (10)$$

と表される。 p を Δp 動かした際の $E(p + \Delta p)$ を最小にすることを考え、 r 周りのテイラー展開を行うと、

$$r(p + \Delta p) = r(p) + G\Delta p + O(\Delta p) \quad (11)$$

ただし、

$$G = \frac{\partial}{\partial p} r(p) \quad (12)$$

である。 $O(\Delta p)$ は十分に小さいので、結局、

$$E(p + \Delta p) = \|r(p) + G\Delta p\|^2 \quad (13)$$

を最小にする p を求めればよいことになる。また、 $\|r(p) + G\Delta p\|^2$ を展開すると、

$$\begin{aligned} \|r(p) + G\Delta p\|^2 &= r(p)^T r(p) + r(p)^T G\Delta p \\ &\quad + \Delta p^T G^T r(p) + \Delta p^T G^T G\Delta p \end{aligned} \quad (14)$$

となり、これを Δp について偏微分すると、

$$\begin{aligned} \frac{\partial}{\partial \Delta p} \|r(p) + G\Delta p\|^2 &= (r(p)^T G)^T + G^T r(p) \\ &\quad + (G^T G + G^T G)\Delta p \\ &= 2G^T r(p) + 2G^T G\Delta p \end{aligned} \quad (15)$$

となる。式 (15) = 0 とすると、

$$\begin{aligned} G^T r(p) + G^T G\Delta p &= 0 \\ G^T G\Delta p &= -G^T r(p) \\ \Delta p &= -(G^T G)^{-1} G^T r(p) \end{aligned} \quad (16)$$

ベクトル Δp は探索空間における有望な更新値を与えてくれる。これにより、新たな誤差評価関数を考えると、

$$p' = p + \Delta p, \quad E' = E(p') \quad (17)$$

となる。もし、 $E' < E$ の時、 $p' \rightarrow p$ に更新し、収束するまで続ける。 $E' > E$ の時は更新幅を小さくする (0.5, 0.25)。改善が見られない時は収束したと定義する。

3.4 3D-AAM への拡張

本節では、2次元の AAM を3次元の AAM に拡張する手法 [5] [6] [7] について述べる。

3.4.1 3D-AAM の概要

学習画像に様々な方位の変動を含む対象画像を含めることにより、様々な方位の変形が可能な AAM を作成できる。例えば、手を左右上下に傾けた画像を用意することで、対象の方位の変動にも対応した AAM を作成することが原理的には可能である。しかし、AAM の中に多数の変動成分を入れると、PCA で次元圧縮した時に変動を表現しきれなくなるため、対象の特徴点抽出精度が下がることが考えられる。そもそも、対象の方位の変動というのは、AAM 自体の幾何学的な変形によって表現することができるので、3次元の対象の形状と、正面のテクスチャがあれば、任意の方向のモデルを生成することができる。従って 3D-AAM は、提案システムに有効であると考えられる。

3.4.2 深度画像センサーによる RGBD 画像取得

3D-AAM を作成する際に、まず、対象の3次元の形状の取得が必要となる。3次元の形状を取得する方法には、3次元スキャナを用いる方法や、ステレオカメラによる手法などがあるが、本研究では Microsoft 社の Kinect を使用する。このデバイスでは、赤外線を用いた距離センサーに RGB カメラを組み合わせることで、特殊なマーカーなしでモーションキャプチャを行うことを実現している。このデバイスで使用されている距離センサーは、特殊な赤外線パターンを照射し、それを解析することで距離画像の生成が可能である。距離画像センサーから対象の3次元空間内での特徴点座標を取得することができるので、式 (1) は、以下のように変更される。

$$\mathbf{x} = (x_1, y_1, z_1, \dots, x_n, y_n, z_n)^T \quad (18)$$

3.4.3 Pose パラメータの拡張

Pose パラメータ (8) を 3次元に拡張すると、横方向の角度変化 yaw と傾き方向の角度変化 $pitch$ を加えて、以下ようになる。

$$q = [yaw \ pitch \ roll \ scale \ trans_x \ trans_y] \quad (19)$$

各パラメータの動きの例は図4のようになる。この6次

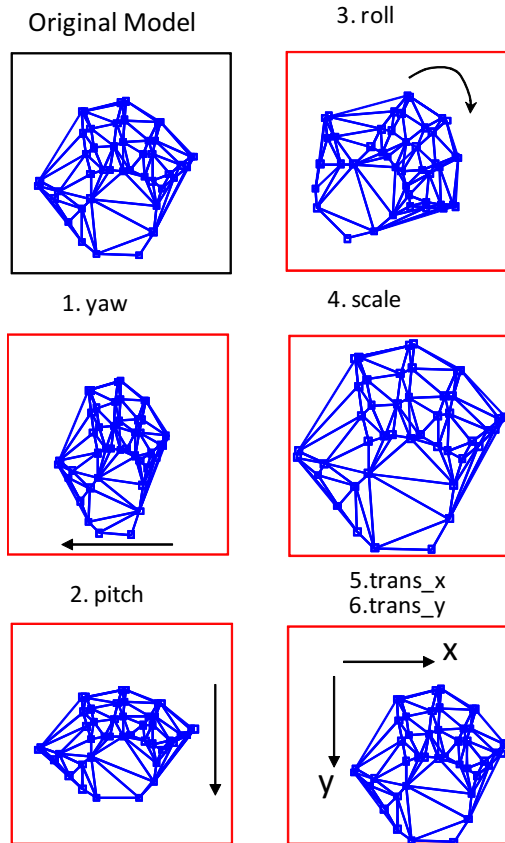


図4 3D pose parameter

元の pose パラメータにより、AAM をあらゆる方向・角度・位置へ動かすことが可能となる。また、この pose パラメータを用いて shape を変化させるには、変化前の shape を x_b 、変換後の shape を x_a とすると、

$$x_a = Trans \cdot Scale \cdot RotZ \cdot RotY \cdot RotX \cdot x_b \quad (20)$$

で求めることができる。ただし、各行列の内容は図5の通りである。

3.4.4 3D-AAM の探索

3D-AAM では、入力は 2次元画像であるのに対して、モデルは 3次元であるので、入力画像とモデルの誤差値を求める際には、3次元のモデルを 2次元に射影する必要がある。すなわち、3次元の shape を 2次元に射影する関数を $P(x)$ とすると、式 (9) は、

$$E = [(\bar{g} + P_g Q_g c') - I(Img, P(W(x; q', b_s')))]^2 \quad (21)$$

と拡張される。最適パラメータの推定は 2D のときと同様に最急降下法によって行う。

$$Trans = \begin{pmatrix} 1 & 0 & 0 & trans_x \\ 0 & 1 & 0 & trans_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$Scale = \begin{pmatrix} scale & 0 & 0 & 0 \\ 0 & scale & 0 & 0 \\ 0 & 0 & scale & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$RotX = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(yaw * \pi / 180) & -\sin(yaw * \pi / 180) & 0 \\ 0 & \sin(yaw * \pi / 180) & \cos(yaw * \pi / 180) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$RotY = \begin{pmatrix} \cos(pitch * \pi / 180) & 0 & \sin(pitch * \pi / 180) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(pitch * \pi / 180) & 0 & \cos(pitch * \pi / 180) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$RotZ = \begin{pmatrix} \cos(roll * \pi / 180) & -\sin(roll * \pi / 180) & 0 & 0 \\ \sin(roll * \pi / 180) & \cos(roll * \pi / 180) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

図5 3D rotation matrix

4. 実験

本研究では複数の 3D-AAM による探索を行い、各モデルごとに正規化した誤差値 (21) を比較し、最も値の小さいものを最終的な出力とする。

4.1 実験条件

今回実験では使用した 4 種類の 3次元モデルを図6に示す。それぞれの特徴点数は 43 点, 34 点, 34 点, 38 点となっている。学習に用いた距離画像は各モデルにつき 4 枚で、事前に背景差分を行いモデルに含まれている背景部分は除外した。学習・テスト共に同一人物のデータを使用しており、テストに使用したデータは 4 種類の指の形状を様々な方位に傾けたデータセットとなっている。

4.2 実験結果

形状ごとの認識結果の Confusion matrix を図7に示す。実験の結果, shape1, shape2 に関しては高い認識率を示すことができたが, shape3 と shape4 にはお互いに誤認識がみられた。これは, shape3 と shape4 が構造的に似ていること、また、3D-AAM の探索の際にモデル全体が対象内に含まれてしまうと、結果として式 (21) の値が低くなってしまふことが原因と考えられる。

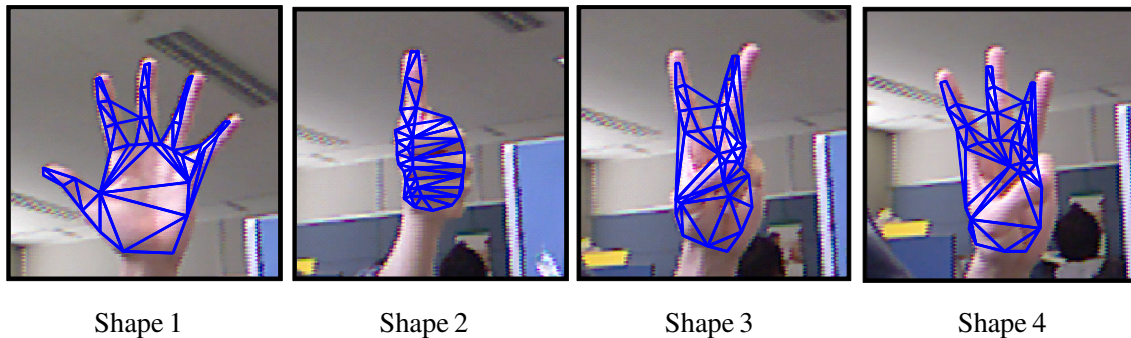


図 6 Shape pattern

	Shape 1	Shape 2	Shape 3	Shape 4
Shape 1	1.000	0.000	0.000	0.000
Shape 2	0.000	1.000	0.000	0.000
Shape 3	0.000	0.011	0.966	0.023
Shape 4	0.000	0.000	0.327	0.673

図 7 Confusion matrix

- synthesis of 3D faces,” SIGGRAPH 1999
- [8] 石井健一郎, 上田修功, 前田英作, 村瀬洋共著, “ わかりやすいパターン認識,” オーム社, 1998
- [9] 平岡和幸, 堀玄共著, “プログラミングのための線形代数,” オーム社, 2004
- [10] Mikkel B. Stegmann, “The AAM-API: An Open Source Active Appearance Model Implementation,” Medical Image Computing and Computer-Assisted Intervention 2003

5. ま と め

本論文では、複数の3次元モデルを使用することにより、指先の複雑な構造を認識する手法を提案した。入力デバイスに距離画像センサを用いることで、対象の深度情報とRGB情報を取得しモデルを学習した。3次元モデルを適用することで様々な方位の変化に頑健な形状認識が実現できた。今後は、入力画像の3次元情報も考慮することでより高精度な形状認識を行い、ハンドジェスチャー認識を実現していきたい。

文 献

- [1] 根津禎, “Kinectに見るジェスチャー入力の可能性”, 日経エレクトロニクス, pp. 73 - 85, 2010-12-27.
- [2] T.F. Cootes, G.J. Edwards, and C.J. Taylor, “Active appearance models,” In Burkhardt and Neumann, editors, Computer Vision - ECCV '98 Vol. II, Freiburg, Germany, 1998
- [3] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, “Active Shape Models - Their Training and Application,” Computer Vision and Image Understanding, Vol. 6, No. 1, pp. 38-59 1995
- [4] T.F.Cootes, K Walker, C.J.Taylor, “View-based Active Appearance Models”, Image and Vision Computing 20(2002), pp.657-664, 2002.
- [5] Jing Xiao, Simon Baker, Iain Matthews, and Takeo Kanade, “Real-Time Combined 2D+3D Active Appearance Models,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 535 - 542, 2004
- [6] Mingcai Zhou, Yangsheng Wang, and Xiangsheng Huang, “Real-time 3D Face and Facial action tracking using Extended 2D+3D AAMs,” Proceedings of the IEEE Conference on International Conference on Pattern Recognition, pp. 3963-3966, 2010
- [7] V. Blanz and T. Vetter, “A morphable model for the