

AAM Based HCI for an Intelligent Wheelchair

Pei Jia*, Huosheng Hu
Department of Computer Science, University of Essex
Colchester, Essex, UK, CO4 3SQ

ABSTRACT

This paper proposes an active appearance model (AAM) based human computer interface (HCI) to control an intelligent wheelchair (IW), namely, RoboChair. Adaboost is applied as the face detection module, while AAM is used to undertake face tracking task. Inverse compositional image alignment is implemented to fit the trained AAM. Subsequently, a straightforward method is carried out to estimate the face direction by comparing the current face shape to the template face shape, so that the intelligent wheelchair could be actuated just according to the head gestures. A well-designed decision making module is deployed to identify whether AAM is correctly tracking the face at the current frame. Results from simulation experiments show the robustness and veracity of Adaboost face detection, AAM face tracking, and face direction estimation presented. An image sequence of our local desktop simulation is provided to demonstrate its feasibility and reliability.

Keywords: AAM, inverse compositional image alignment, HCI, intelligent wheelchair

1. INTRODUCTION

Over the last twenty years, various electric-powered wheelchairs with diverse functionalities have been developed for the elderly and disabled people^{1,2}. However, most of these wheelchairs are still manually controlled by users and are not suitable for people with severe disabilities. This paper proposes a novel head gesture based HCI for hands-free control of a RoboChair, in which AAM based HCI is adopted to detect user face directions for the control purpose. Three basic modules are developed, namely face detection, face tracking, and face direction estimation.

The early version of our hands-free control^{3,4} was based on the integration of Adaboost face detection^{5,6,7} and Camshift object tracking⁸. Our RoboChair was controlled by four directions of head gesture only, i.e. turn left, turn right, speed up and slow down, corresponding to four face directions left, right, up and down. Meanwhile, the head gesture is constrained vertically, i.e., not too much deviation from the upright head posture is permitted. In this paper, AAM based HCI can roughly estimate the face turning angles even if the face is not upright to some extent. These face turning angles represent different speed demands to be sent to the RoboChair so that it could be controlled more smoothly.

According to the recent advances in face detection^{9,10,11}, Adaboost appearance based face detection is one of the most successful methods in terms of both accuracy and speed. On the other hand, regarding the face tracking and recognition, AAM has shown its excellent accuracy and robustness. Hence, in our current approach, Adaboost is applied as the face detection module, while AAM is used to undertake face tracking task. Inverse compositional image alignment is implemented to fit the trained AAM. Subsequently, a simple geometric method is adopted to estimate the face direction by comparing the current face shape to the template face shape. Then our RoboChair can be actuated according to the head turning angles.

AAM was originated from active shape model (ASM)^{12,13}. Unlike ASM that seeks to match the position of the model points, AAM tries to match both the shape and the texture representation of the object simultaneously. Some comparison between ASM and AAM can be found in¹⁴. The difficulty of AAM is how to fit an arbitrary image to the trained model template image. Cootes summarized several traditional AAM fitting methods in¹⁵, such as basic appearance model and direct appearance model¹⁶, etc. He also demonstrates that a small number of 2D view-based statistical appearance models can represent the face from a wide range of viewing angles. However, all these methods are comparably slow.

Recently, CMU proposed an inverse compositional image alignment algorithm for AAM fitting¹⁷. By pre-calculating most of the intermediate parameters based on delicate mathematical deduction for model optimization, their method is able to both run in real time and obtain excellent fitting results. It is reported in¹⁸ that this inverse compositional image alignment algorithm for 3D images AAM fitting could operate at around 286Hz, and it can deal with faces from wide

*jp4work@gmail.com; phone 44 1206 874092

ranges of viewing angles.

In this paper, the 2D inverse compositional image alignment algorithm is adopted as our face tracking module. In every tracking frame, AAM fitting is iterated until the current face is converged to the template trained face. With the fitted shape model parameters, a list of estimated shape points can be computed, which is able to describe face details. Unlike CMUs method presented in¹⁸, which is actually a combination of 2D and 3D AAM fitting algorithms, we ignore the 3D model fitting, and just calculate the face angles in a simple geometric way only based on the 2D fitted shape.

By defining a measurement between the face shape (the point list) of the current frame and that of the template, the face angles can be inferred approximately. Our RoboChair can be triggered by these estimated face angles in a linear way. That is, the speed of our RoboChair, including both forwarding speed and turning speeds, is linear correlated with the face angles, under the restriction of maximum speed limitations.

Since the user's face may not be quite upright sometimes, AAM can also be used to deal with this face rotation situation to some extent so that the face turning angles can be correctly tracked based on shape representation. To improve the robustness of the proposed HCI, a dedicated judgment module is developed in our system to perform a strict tracking verification by judging: 1) whether the face is out of view or not; 2) whether the face is too big or too small; 3) whether the AAM Delaunay triangulation is of a reasonable structure - every point in the shape model shouldn't be out of $\pm 3\delta$ of its trained position; etc. Results from simulation experiments show the robustness and real-time performance of Adaboost face detection, AAM face tracking, and face direction estimation presented.

2. AAM BACKGROUND

Our early version of Robochair is described in detail in^{3,4}, which explains Robochair system structure and our first HCI design for the Robochair. In order to obtain more control flexibility, the current version of our newly designed HCI is based on the integration of Adaboost face detection and AAM based face tracking. With both beautiful mathematical deductions and excellent performance for real-time face tracking, AAM has been one of the current hottest research topics in the area of computer vision^{15, 17}. Generally, AAM issue contains two stages: 1) "building" for constructing AAM models; 2) "fitting" for the real-time face tracking.

2.1 Building an AAM

AAM is a hybrid model for object representation, which tries to express both the object shape and the corresponding texture simultaneously. Generally speaking, an AAM contains a shape model and a texture model.

The shape of an AAM is defined as a vector of coordinates for v vertices, which actually makes up a mesh.

$$s = [x_1, y_1, x_2, y_2, \dots, x_v, y_v] \quad (1)$$

AAM shape model looks on an arbitrary face shape s as a base shape s_0 plus a linear combination of N shape vectors s_i .

$$s = s_0 + \sum_{i=1}^N p_i s_i \quad (2)$$

where p_i are the shape parameters and s_i are the orthonormal eigenvectors.

The texture of an AAM is defined as a vector of coordinate-related intensities over all the pixels inside the base mesh s_0 . AAM texture model looks on an arbitrary texture $A(x)$ as a base texture $A_0(x)$ plus a linear combination of M texture vectors $A_i(x)$.

$$A(x) = A_0(x) + \sum_{i=1}^M \lambda_i A_i(x) \quad \forall x \in s_0 \quad (3)$$

where $\mathbf{x} = (x, y)^T$ indicate the pixel coordinates, λ_i are the texture parameters and $A_i(\mathbf{x})$ are the orthonormal eigenvectors.

2.2 Fitting an AAM

The key of AAM issue nowadays is how to fit an arbitrary image to the trained model template image. Matthews and Baker summarized their method based on AAM model and Lucas-Kanade image alignment in detail in¹⁷. The AAM fitting method based on inverse compositional image alignment is implemented in our project, which is briefly reviewed in the following.

Generally speaking, AAM fitting issue could be summarized as an optimization problem to minimize the following L2 norm:

$$\left\| A_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i A_i(\mathbf{x}) - I(W(\mathbf{x}; \mathbf{p})) \right\|^2 \quad (4)$$

where $A_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i A_i(\mathbf{x})$ comes from (3); $\mathbf{p} = (p_1, p_2, \dots, p_N)^T$ is a vector of N shape parameters, refer to (2);

$W(\mathbf{x}; \mathbf{p})$ is the destination pixel coordinates warped from the source pixel \mathbf{x} by using the shape parameters \mathbf{p} on an arbitrary newly input image; and $I(W(\mathbf{x}; \mathbf{p}))$ is just the intensity (intensities) vector of all the warped pixels $W(\mathbf{x}; \mathbf{p})$ in this newly input image. The AAM fitting is just to minimize the above expression with respect to both $\mathbf{p} = (p_1, p_2, \dots, p_N)^T$ and $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)^T$. Apparently, (4) can be spanned as:

$$\begin{aligned} & \left\| A_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i A_i(\mathbf{x}) - I(W(\mathbf{x}; \mathbf{p})) \right\|_{\text{span}(A_i)^\perp}^2 + \left\| A_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i A_i(\mathbf{x}) - I(W(\mathbf{x}; \mathbf{p})) \right\|_{\text{span}(A_i)}^2 \\ &= \left\| A_0(\mathbf{x}) - I(W(\mathbf{x}; \mathbf{p})) \right\|_{\text{span}(A_i)^\perp}^2 + \left\| A_0(\mathbf{x}) + \sum_{i=1}^M \lambda_i A_i(\mathbf{x}) - I(W(\mathbf{x}; \mathbf{p})) \right\|_{\text{span}(A_i)}^2 \end{aligned} \quad (5)$$

Obviously, any vector projected (spanned) in the eigenvector space $\text{span}(A_i)$ can be seamlessly expressed by the linear combination of these eigenvectors A_i . So, no matter how is \mathbf{p} chosen, the second term in (5) could always be minimized to 0. Therefore,

$$\lambda_i = \sum_{\mathbf{x} \in S_0} A_i(\mathbf{x}) \cdot [I(W(\mathbf{x}; \mathbf{p})) - A_0(\mathbf{x})] \quad (6)$$

For the first term in (5), we will resort to inverse compositional image alignment method in the next subsection.

2.3 Inverse Compositional Image Alignment for AAM

The ordinary idea to minimize $\left\| A_0(\mathbf{x}) - I(W(\mathbf{x}; \mathbf{p})) \right\|_{\text{span}(A_i)^\perp}^2$ is to adopt traditional linear optimization method by linearly adjusting \mathbf{p} in each iteration: $\mathbf{p} \rightarrow \mathbf{p} + \Delta \mathbf{p}$. Compositional method brings a new thought by adjusting \mathbf{p} in a nesting way, so the optimization problem could be addressed as: $\left\| A_0(\mathbf{x}) - I(W(W(\mathbf{x}; \Delta \mathbf{p}); \mathbf{p})) \right\|_{\text{span}(A_i)^\perp}^2$. Inverse compositional image alignment method is just reversing the roles of the trained model template image $A_0(\mathbf{x})$ and the newly input image $I(W(W(\mathbf{x}; \Delta \mathbf{p}); \mathbf{p}))$ in compositional method, which results in

$$\left\| A_0(W(\mathbf{x}; \Delta \mathbf{p})) - I(W(\mathbf{x}; \mathbf{p})) \right\|_{\text{span}(A_i)^\perp}^2 \quad (7)$$

In order to get the extremum, after taking the first order Taylor series expansion in terms of Δp at $\Delta p=0$, the solution to Δp could be deducted as

$$\Delta p = H^{-1} \sum_x \left[\nabla A_0 \frac{\partial W}{\partial p} \Big|_{p=0} \right]_{span(A_i)^\perp}^T [I(W(x; p)) - A_0(x)] \quad (8)$$

$$H = \sum_x \left[\nabla A_0 \frac{\partial W}{\partial p} \Big|_{p=0} \right]_{span(A_i)^\perp}^T \left[\nabla A_0 \frac{\partial W}{\partial p} \Big|_{p=0} \right]_{span(A_i)^\perp} \quad (9)$$

where ∇A_0 is a 1×2 gradient vector for every pixel in the trained model template image; $\frac{\partial W}{\partial p} \Big|_{p=0}$ is a $2 \times N$ Jacobian matrix to express the relationship between the 2×1 warped point $W(x; p)$ and the $N \times 1$ AAM shape model parameters vector p , which is evaluated at $p=0$; $\nabla A_0 \frac{\partial W}{\partial p} \Big|_{p=0}$ is named steepest descent image, because it describes the shortest path to reach the minimum; H is the $N \times N$ Hessian matrix which can be pre-computed before the iterations due to its independence on p .

In brief, inverse compositional image alignment algorithm can be summarized in the following framework figure. The left block describes the items able to be pre-computed, and the right block describes the items to be calculated during the iterations.

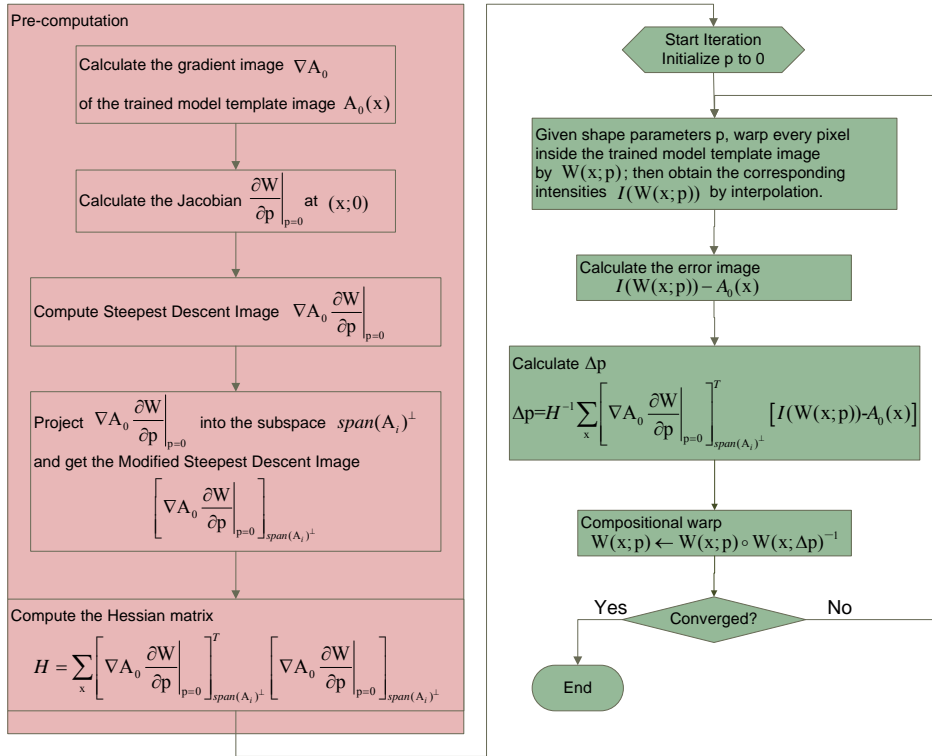


Fig. 1. Framework of inverse compositional image alignment algorithm for AAM fitting.

2.4 Implementations

In the above framework, how to compute the Jacobian matrix $\left. \frac{\partial W}{\partial \mathbf{p}} \right|_{\mathbf{p}=0}$ is particularly essential. It's not hard to see if this Jacobian matrix is computed, all the other items can be easily calculated subsequently. With limited space here, the solutions are given out directly in the following. Please refer to ¹⁹ for detailed deductions.

$$\begin{aligned} \left. \frac{\partial W_x}{\partial \mathbf{p}_i} \right|_{\mathbf{p}=0} &= \alpha s_i^{x_r} + \beta s_i^{x_s} + \gamma s_i^{x_t} \\ \left. \frac{\partial W_y}{\partial \mathbf{p}_i} \right|_{\mathbf{p}=0} &= \alpha s_i^{y_r} + \beta s_i^{y_s} + \gamma s_i^{y_t} \end{aligned} \quad (10)$$

If we look on the warp transform as a linear combination $\mathbf{x} = \alpha \mathbf{x}_r^0 + \beta \mathbf{x}_s^0 + \gamma \mathbf{x}_t^0$, where \mathbf{x}_r^0 , \mathbf{x}_s^0 , \mathbf{x}_t^0 are the three vertexes of the triangle containing the point \mathbf{x} , α , β , γ can be respectively deducted as:

$$\begin{aligned} \alpha &= \frac{x_s^0 y_t^0 - y_s^0 x_t^0 - x y_t^0 + y x_t^0 - y x_s^0 + x y_s^0}{x_s^0 y_t^0 - x_r^0 y_t^0 - x_s^0 y_r^0 - y_s^0 x_t^0 + y_r^0 x_t^0 + y_s^0 x_r^0} \\ \beta &= \frac{x y_t^0 - x_r^0 y_t^0 - x y_r^0 - y x_t^0 + y_r^0 x_t^0 + y x_r^0}{x_s^0 y_t^0 - x_r^0 y_t^0 - x_s^0 y_r^0 - y_s^0 x_t^0 + y_r^0 x_t^0 + y_s^0 x_r^0} \\ \gamma &= \frac{y x_s^0 - y_r^0 x_s^0 - y x_r^0 - x y_s^0 + x_r^0 y_s^0 + x y_r^0}{x_s^0 y_t^0 - x_r^0 y_t^0 - x_s^0 y_r^0 - y_s^0 x_t^0 + y_r^0 x_t^0 + y_s^0 x_r^0} \end{aligned} \quad (11)$$

Besides, $s_i^{x_r}$, $s_i^{x_s}$, $s_i^{x_t}$, $s_i^{y_r}$, $s_i^{y_s}$, $s_i^{y_t}$ denote the components of s_i (refer to (2)) corresponding to the positions $x_r, x_s, x_t, y_r, y_s, y_t$ respectively.

3. HEAD POSE ESTIMATION

Two head pose estimation methods are introduced in ²⁰ after a face has been successfully fitted by AAM. For the first method: after AAM fitting, a further refining detection for two outer corners of two eyes and out corners of the mouth is implemented to obtain a trapezoid. Simply comparing this trapezoid with the corresponding one built from the trained model template face, the head pose could be estimated roughly. The other method is to train AAM in a multi-view way. That is to say, for different face directions (normally, 5 face directions are used), different trained shape model vectors $\mathbf{p}_{\text{angles}[i]}$ could be obtained. By defining the similarity between the current fitting shape parameter vector \mathbf{p} and those trained shape model vectors $\mathbf{p}_{\text{angles}[i]}$, a rough angle could be calculated.

3.1 AAM Explorer

In order to analyze the relationship between the AAM shape parameters and the head poses, a standalone software "AAM Explorer" is developed for further investigation. We trained our AAM models by using the free FAME²¹ dataset, which contains 40 subjects; each subject has 6 snapshots with variant head poses and face expressions. Fortunately, we found that the first parameter of AAM model shape parameter vector \mathbf{p} just indicates the direction of the eigenvector with the biggest variance for the training face dataset. For FAME, the biggest variance among all the faces just comes from the left-right head pose changing, and the second biggest variance just comes from the up-bottom head pose changing as shown in the experimental results. Thus, our first proposed method is to directly tell the head pose by the

first two fitted AAM shape parameters. This method is of a straightforward idea, but it's not guaranteed that the training dataset has the above characters.

3.2 Simple Geometric Method

So, another simple thought is produced to compute the head poses by statistics. For a fitted image, we observed that the relative position between the center of the mouth MC and the center of the chin CC is changing due to the alteration of the head poses. The key is to analyze how much percentage of shift there is between the above two centers. By using this simple method, the head pose could be roughly calculated after the newly incoming face is fitted. In a mathematical expression, it is:

$$pose_x = F\left(\frac{MC-CC}{\text{face width}}\right); \quad pose_y = F\left(\frac{MC-CC}{\text{face height}}\right) \quad (12)$$

3.3 Trigger Intelligent Wheelchair by Head Pose

The intelligent wheelchair can just be triggered by the above calculated head pose in a linear way. To improve the robustness of the above proposed HCI, a dedicated judgment module is developed in our system to perform a strict tracking verification by judging: 1) whether the face is out of view or not; 2) whether the face is too big or too small; 3) whether the AAM Delaunay triangulation is of a reasonable structure - every point in the shape model shouldn't be out of $\pm 3\delta$ of its trained position; etc. We ignore the situation that the face is too profile to be fitted at our current stage of research. In such a case, no command will be sent to the intelligent wheelchair at this frame or time period.

4. EXPERIMENTAL RESULTS

In the following experiments, based on our AAM Explorer, by only changing the first two shape parameters, we try to analyze the relationship between the mouth center and the chin center in terms of different head poses. Afterwards, AAM fitting results for faces of different angles are given out, which shows that the inverse compositional image alignment method is a nice method for the face tracking task. Meanwhile, AAM shows its capability in telling the face details.

4.1 AAM Explorer Experiments

In the following five pair of images, pair (a) is the trained template image with the corresponding rebuilt shape back projected from the template model parameters without any alteration. Pairs (b) and (c) are the rebuilt shapes back projected from the modified template model parameters by only shifting the first shape parameters. Similarly, pairs (d) and (e) are the rebuilt shapes back projected from the modified template model parameters by only shifting the second shape parameters. Apparently, for these five pairs, the relative position between the center of mouth and the center of the chin differs from each other, which is able to tell the head poses. With limited space here and due to the simplicity of the above proposed geometric method, we omit the calculation results of the head gesture angles here.

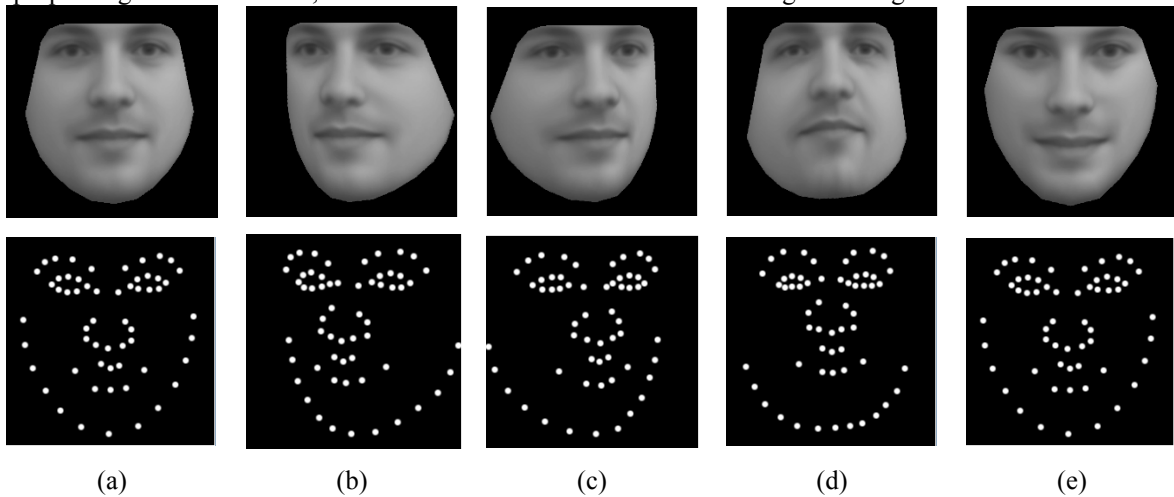


Fig. 2. AAM Explorer experiments to show the possibility of calculating head pose by (12)

4.2 AAM Fitting Experiments

In order to keep consistency, one sample subject in FAME dataset is chosen as the main role here in our experiments for AAM fitting. We carried out two experiments here. The first is to show AAM fitting is able to deal with half-profile faces after effective training; the second is to show a fitting procedure for a half-profile face.

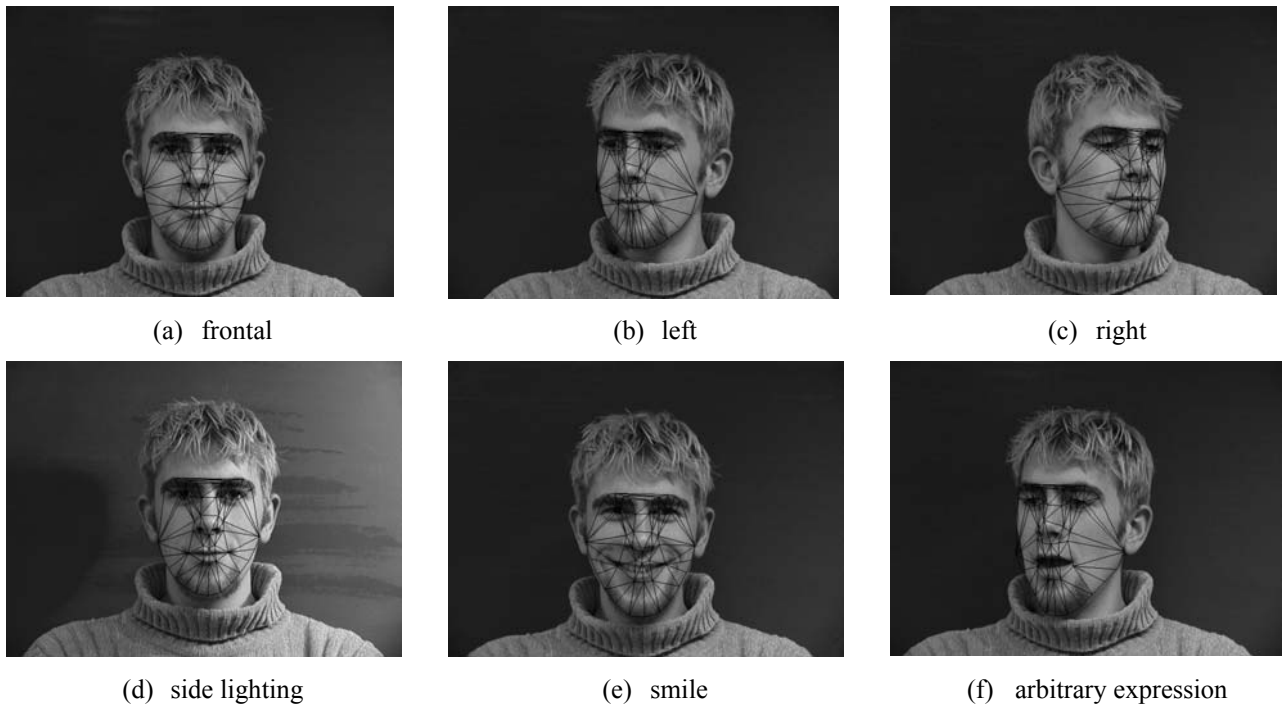


Fig. 3. AAM Fitting for one person with various head poses and face expressions



Fig. 4. AAM Fitting iterations

5. CONCLUSION AND FUTURE WORK

This paper describes a HCI for intelligent wheelchairs. AAM is able to provide face details. Based on these detailed information, a simple geometric method is proposed to calculate the head poses. This AAM based HCI enhanced mobility for the elderly and disabled people who have very restricted limb movements or severe handicap.

Our future research will be focused on two sides: 1) the multi-view head poses recognition for the totally profile faces, so that for the intelligent wheelchair, the user may have more flexible mobility; 2) Stabilize AAM fitting due to the time to time Delaunay Triangle topology alteration during the fitting process.

REFERENCES

1. R. Simpson, E. Lopresti, S. Hayashi, I. Nourbakhsh and D. Miller, "The smart wheelchair component system", *Journal of Rehabilitation Research and Development* 41(3B), pp. 429-442, 2004.
2. D. Ding and R. A. Cooper, "Electric powered wheelchairs", *IEEE Control Systems Magazine* 25, pp. 22-34, April 2005.
3. P. Jia and H. Hu, "Head gesture based control of an intelligent wheelchair", *The 11th Annual Conference of the Chinese Automation and Computing Society in the UK (CACSUUK)*, (Sheffield, UK), October 10 2005.
4. P. Jia, H. Hu, T. Lu, and K. Yuan, "Head gesture recognition for hands-free control of an intelligent wheelchair", *Journal of Industrial Robot* 34(1), 2007.
5. P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features", in *Proceedings of IEEE Computer Society's Computer Vision and Pattern Recognition*, 1, pp. 511-518, 2001.
6. R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection", in *Proceedings of IEEE International Conference on Image Processing*, 1, pp. 900-903, September 2002.
7. P. Viola and M. J. Jones, "Robust real-time face detection", *International Journal of Computer Vision* 57, pp. 137-154, May 2004.
8. G. Bradski, "Real time face and object tracking as a component of a perceptual user interface", in *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision*, pp. 214-219, (Princeton, NJ, USA), October 1998.
9. M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, pp. 34-58, January 2002.
10. M.-H. Yang, "Recent advances in face detection", in *Tutorial of IEEE Conference on Image Processing*, (Barcelona, Spain), September 13-17 2003.
11. M.-H. Yang, "Recent advances in face detection", in *Tutorial of IEEE Conference on Pattern Recognition*, (Cambridge, UK), August 23-26 2004.
12. T. Cootes and C. Taylor, "Active shape models - smart snakes", in *Proceedings of British Machine Vision Conference (BMVC 1992)*, pp. 266-275, Springer-Verlag, 1992.
13. T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models-their training and application", *Journal of Computer Vision and Image Understanding* 61, pp. 38-59, January 1995.
14. T. Cootes, G. Edwards, and C. Taylor, "Comparing active shape models with active appearance models", in *10th British Machine Vision Conference (BMVC 1999)*, pp. 173-182, (Nottingham, UK), September 1999.
15. T. Cootes and C. Taylor, "Statistical models of appearance for computer vision", Ongoing Draft of the Technical Report on Active Shape Models and Active Appearance Models, Imaging Science and Biomedical Engineering, University of Manchester, March 8 2004.
16. X. Hou, S. Li, and H. Zhang, "Direct appearance models", in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, pp. 828-833, 2001.
17. I. Matthews and S. Baker, "Active appearance models revisited", *International Journal of Computer Vision* 60, pp. 135-164, November 2004.
18. S. Baker, I. Matthews, J. Xiao, R. Gross, T. Kanade, and T. Ishikawa, "Real-time non-rigid driver head tracking for driver mental state estimation", in *11th World Congress on Intelligent Transportation Systems*, (Nagoya, Japan), October 2004.
19. P. Jia, "Active Appearance Model", *online open report at <http://www.visionopen.com/cv/aam.php>*, 2006.
20. K.-F. Kraiss, "Advanced Man-Machine Interaction – Fundamentals and Implementation", Springer, 2006
21. M. B. Stegmann, B. K. Ersböll, and R. Larsen. "FAME - a flexible appearance modelling environment", *IEEE Transactions on Medical Imaging*, 22(10):1319-1331, 2003