

# Active Appearance Model 을 이용한 얼굴 추적 시스템

조경식<sup>1</sup>, 김용국<sup>2</sup>  
nopnse@naver.com<sup>1</sup>, ykim@sejong.ac.kr<sup>2</sup>

## Face Tracking System using Active Appearance Model

Kyoung Sic Cho<sup>1</sup>, Yong Guk Kim<sup>2</sup>  
Sejong University Department of Computer Engineering<sup>1,2</sup>

### 요약

얼굴 추적은 Vision base HCI 의 핵심인 얼굴인식, 표정인식 그리고 Gesture recognition 등의 다른 여러 기술을 지원하는 중요한 기술이다. 이런 얼굴 추적기술에는 영상(Image)의 Color 또는 Contour 등의 불변하는 특징들을 사용하거나 템플릿(template) 또는 형태(appearance)를 사용하는 방법 등이 있는데 이런 방법들은 조명환경이나 주위 배경등의 외부 환경에 민감하게 반응함으로써 다양한 환경에 사용할 수 없을 뿐더러 얼굴영상만을 정확하게 추출하기도 쉽지 않은 실정이다. 이에 본 논문에서는 deformable 한 model 을 사용하여 model 과 유사한 shape 과 appearance 를 찾아 내는 AAM(Active Appearance Model)을 사용하는 얼굴 추적 시스템을 제안하고자 한다. 제안된 시스템에는 기존의 Combined AAM 이 아닌 Independent AAM 을 사용하였고 또한 Fitting Algorithm 에 Inverse Compositional Image Alignment 를 사용하여 Fitting 속도를 향상 시켰다. AAM Model 을 만들기 위한 Train set 은 150 장의 4 가지 형태에 얼굴을 담고 있는 Gray-scale 영상을 사용 하였다. Shape Model 은 각 영상마다 직접 표기한 47 개의 Vertex 를 Trianglize 함으로서 생성되는 71 개의 Triangles 을 하나의 Mesh 로 구성하여 생성 하였고, Appearance Model 은 Shape 안쪽의 모든 픽셀을 사용해서 생성하였다. 시스템의 성능 평가는 Fitting 후 Shape 좌표의 정확도를 측정 함으로서 평가 하였다.

Keyword : Active Appearance Model, Inverse Compositional Image Alignment, Shape Model, appearance Model

### 1. 서론

얼굴 추적은 Vision base HCI 의 핵심인 얼굴인식, 표정인식 그리고 Gesture recognition 등의 다른 여러 기술을 지원하는 중요한 기술 이다.[1] 그 이유는 얼굴의 위치, 자세 그리고 표정은 사람의 상태를 나타내는 정보를 해석하고 추출하는 중요한 요소이기 때문이다. 얼굴 추적기술에는 영상의 Color 또는 Contour 등의 불변하는 특징들을 사용 하거나 템플릿 또는 형태(appearance)를 사용하는 방법 등이 있다. Color 나 Contour 를 이용하는 방법은

조명환경과 같은 외부 환경이 우수할 때는 좋은 성능을 발휘하나, 그렇지 못할 경우에는 성능이 급격하게 감소한다. 이러한 이유로 최근에는 형태 정보나 움직임 정보 등을 병합하여 얼굴 추적을 하고 있는데 이런 기술에는 본 논문의 AAMs 과 Morphable Model, Active Blobs 등이 있다.

본 논문에서는 컴퓨터에 Web-Cam 을 장착해서 획득한 실시간영상에 AAM(Active Appearance Model)을 적용하여 얼굴을 추적하는 얼굴추적 시스템을 제안하고자 한다. AAMs 는 Morphable Models 과 Active Blobs 의 유사한 개념으로 Visual phenomenon model 을 생성 한다.[3] AAMs 는 픽셀 값 범위에서 비선형 파라미터릭 모델(nonlinear

parametric model)을 생성하고, 입력영상과 AAMs 모델간의 최소 오차를 구하는 계수를 찾는다. 이 계수는 Fitting Algorithm 를 사용하여 찾게 되는데 이 Algorithms 에는 Lucas-Kanade Image Alignment, Compositional Image Alignment, Inverse Compositional Image Alignment 등을 사용한다[4]. AAMs 에서 사용하는 Fitting Algorithms 은 Inverse Compositional Image Alignment 인데 이 Fitting Algorithm 은 다른 Fitting Algorithms 보다 정확도는 조금 떨어지지만 Fitting 속도가 빠르기 때문에 다른 Algorithms 과 비교 했을 때 더 효율적이다[3].

본 논문의 구성은 다음과 같다. 2 장에서는 얼굴 추적 시스템을 설명하고, 3 장에서는 AAM 을 설명한다. 4 장에서는 AAMs 에서 사용하는 Inverse Compositional Image Alignment 를 설명하고, 5 장에서는 얼굴 추적 시스템의 성능 테스트 및 결과를 보여준다. 마지막 6 장에서는 얼굴 추적 시스템에 대한 결론 및 향후 과제로 끝을 맺는다.

## 2. 얼굴 추적 시스템

AAMs 얼굴 추적 시스템은 크게 모델 생성 컴포넌트, 얼굴 추적 컴포넌트 그리고 추적 결과 시각화 컴포넌트로 나뉘어 진다. (그림 1)은 전체 시스템 구성도 이다.

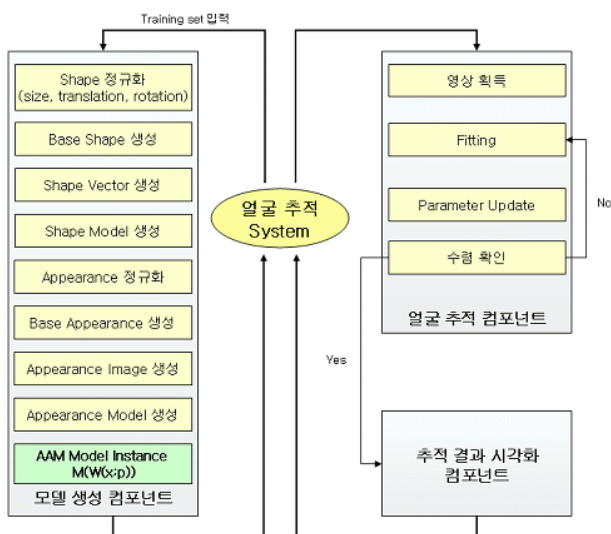


그림 1. AAM 얼굴 추적 System 전체 구성도

시스템 구성도의 좌측은 모델 생성 컴포넌트를 이고, 우측 상단은 얼굴 추적 컴포넌트이다. 그리고 우측 하단은 시각화 컴포넌트 이다. AAMs 얼굴 추적 시스템의 첫 단계는 모델을 생성하기 위해 구성된 Training Set 으로 AAMs 모델을 생성하는 것이고 두번째 단계는 웹 캠으로부터 영상을 입력 받아 AAMs 모델을 변화 시키며 얼굴을 추적하는 것이다. 얼굴이 잘 추적 되었으면 마지막 단계로 시각화 컴포넌트를 사용하여 시각화를 한다.

## 3. Active Appearance Model

AAMs 는 [5]에서 처음 제안 되었다. 대부분의 AAMs Application 이 얼굴을 모델링 하는데 사용 되었지만 다른 객체를 모델링할 때에도 매우 효과적이다[4][6][7]. AAMs 은 입력영상과 모델간의 오차를 가장 작게 해주는 파라미터를 찾아서 새로운 모델 인스턴스를 생성하는 것이 목적인데 하나의 모델은 다른 Application 에서도 사용 가능하다. 예를 들어 얼굴인식 Application 에서 사용한 모델은 포즈인식이나 표정인식에서도 사용 가능하다[8]. AAMs 에서 사용 하는 Fitting Algorithm 은 비선형(non-linear)으로서 Shape 과 Appearance 의 파라미터를 특정 임계값을 만족할 때까지 반복적으로 갱신하며 모델과 영상간의 오차를 줄여나간다. 예를 들어 현재 Shape 의 파라미터가 측정 되었으면 모델 좌표 프레임 위에 입력영상을 맞추고 현재 모델 인스턴스와 AAMs 이 Fitting 을 하는 영상사이에 오차영상(error image)을 구해 오차를 줄이며 파라미터를 최적화 하는 것이다.

### 3-1. Shape

AAMs 에서 Shape 은 mesh 를 구성하는 각 정점위치(Vertex location)들로 정의 하며, 정점은 영상의 특징이 되는 곳에 직접 표기해 놓은 Landmark 로 이루어진다. AAMs 의 Shape 은 선형 변환(linear variation)을 해야 하는데,

이것은 전체 Shape 의 평균인 base Shape  $s_0$  에  $n$  개의 Shape vector 가 조합된다는 뜻이며 식(1)로 나타낼 수 있다.

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

식(1)에서  $p_i$  는 Shape Parameter 의 계수(coefficients)이다. AAMs 에서는 Shape parameter 를 항상 선형으로 조합하기 때문에 Vector  $s_i$  는 Orthonormal 하다고 가정한다. 때문에 Vector  $s_i$  는 Training Image 의 Landmark 들에 좌표를 가지고 PCA 처리를 해서 생성해야 한다[9]. 일반적으로 Training Mesh 는 PCA 를 적용하기 전에 Procrustes analysis 를 사용하여 정규화 한다[3][10].

### 3-2. Appearance

AAMs 의 Appearance  $A(x)$  는 base shape  $s_0$  내에 놓여 있는 픽셀 이다. Appearance 는 Shape 과 마찬가지로 선형변환(linear variation)을 하기 때문에 식(2)에 보이는 바와 같이 Shape 과 유사하게 Training Image 에서 얻어낸 Base appearance  $A_0(x)$  에  $m$  개의 Appearance vector  $A_i(x)$  를 조합한다.

$$A(x) = A_0(x) + \sum_{i=1}^m \lambda_i A_i(x) \quad (2)$$

식(2)에서  $\lambda_i$  는 Appearance 파라미터이고  $A_i(x)$  는 Appearance vector 이다. Appearance 도 Shape 과 마찬가지로 선형변환을 한다고 했기 때문에 Appearance 도 또한 Orthonormal 하다고 가정해야 한다. 따라서 Appearance 도 PCA 처리를 해주어야 하는데 PCA 처리를 하기 위해서는 Appearance 의 벡터길이가 같아야 한다. 그러나 최초의 Appearance 는 Training image 의 영상에 표기된 Shape 내의 pixel 들이므로 Vector 의 길이가 다르다. 따라서 PCA 처리를 하기

전에 Piecewise affine warping 으로 Base shape  $s_0$  에 픽셀을 맞추어 Vector 의 길이를 같게 해주는 정규화를 해줘야 한다[10].

### 4. Inverse Compositional Image Alignment

이미지 정렬(Image alignment)은 입력영상 내에서 고정된 템플릿 영상의 위치를 찾는 것을 말한다. 이번 장에서는 최근 가장 많이 사용하는 이미지 정렬 알고리즘인 Inverse Compositional Image Alignment 를 소개 한다. Inverse Compositional Image Alignment 는 앞서 제안된 이미지 정렬 알고리즘과 달리 템플릿과 입력영상의 역할을 역으로 놓은 것이다. 따라서 incremental warp 을  $I(W(x; p))$  에 관해 계산 하는 것이 아니라  $A_0(x)$  에 관해 계산한다는 것이다. Inverse Compositional Algorithm 이 앞서 제안된 다른 이미지 정렬 알고리즘 보다 효과적이라는 것은 [4][12]에서 증명 되었다. 다음 식(3)은 Inverse Compositional Algorithm 을 축소해 놓은 것이다.

$$\sum_x [I(W(x; p)) - A_0(W(x; \Delta p))]^2 \quad (3)$$

위 식(3)을 선형화 하기 위해 테일러 시리즈를 적용하면 다음과 같은 식(4)를 유도할 수 있고,

$$\sum_x [I(W(x; p)) - A_0(W(x; 0)) - \nabla A_0 \frac{\partial W}{\partial P} \Delta p]^2 \quad (4)$$

위 식(4)에서  $W(x; 0)$  이 Identity warp 이라고 가정하면 최소 자승법으로  $\Delta p$  를 구할 수 있는데 이것은 다음의 식(5)와 같이 표현 된다.

$$\Delta p = H^{-1} \sum_x [\nabla A_0 \frac{\partial W}{\partial P} \Delta p]^T [I(W(x; p)) - A_0(W(x; 0))] \quad (5)$$

위 식에서  $H$  는 Hessian matrix 를 말하고, 다음의 식(6)으로 구할 수 있다.

$$H = \sum_x [\nabla A_0 \frac{\partial W}{\partial P} \Delta p]^T [\nabla A_0 \frac{\partial W}{\partial P} \Delta p] (6)$$

## 5. 구현 및 결과

이번 장에서는 얼굴 추적 시스템의 모델을 만들기 위해 사용된 Training set 에 관하여 설명하고 시스템의 추적 정확도를 알아보기 위한 Test set 을 구성하여 평가한 결과를 제시한다.

### 5-1. Training set

AAM model 을 생성하기 위해 사용한 Training set 은 웹캠을 사용하여 직접 촬영한 Gray Scale 의 RAW file 영상 150 장에 47 개의 Landmark 를 직접 표기하여 구성 하였다. (그림 1)은 Landmark 가 표기된 Training 영상을 보여 준다.

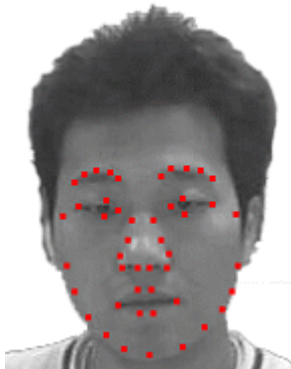


그림 1. Training 영상

표기된 각각의 Landmark 는 주변의 Landmark 와 연결되어 총 71 개의 Triangle 을 만들어 내고 이 Triangle 이 모여서 다시 하나의 mesh 가 생성된다. mesh 를 생성하는 이유는 AAMs 의 핵심인 Shape 구성과 영상의 Appearance 를 획득하기 위함이다. (그림 2)는 (그림 1)의 각각의 Landmark 를 연결해서 생성된 mesh 영상이다. 150 장의 이미지를 사용 하였으므로 mesh 도 150 개가 생성되는데 이 mesh 들은 Shape training set 이 되며 AAM Shape Model 을 만들 때 사용 된다. Appearance training set 은 각 이미지의 mesh 안에 놓여 있는 픽셀들을 Shape Model 을

만드는 과정 중에 생성되는 base mesh 로 Piecewise Affine Warp 하여 생성된다.



그림 2. Mesh 영상

(그림 13)는 Appearance training set 을 보여준다.



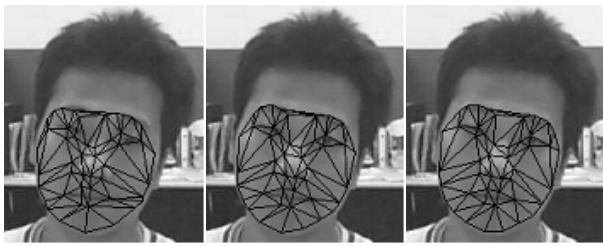
그림 13. Appearance training set

base mesh 에 warp 을 하는 이유는 Shape training set 의 mesh 에 크기가 다르기 때문에 warp 을 하지 않을 경우 Appearance 를 구성하는 픽셀의 수가 다르게 되기 때문에 Appearance model 을 만들 수가 없기 때문이다.

### 5-2. 실험 결과

이번 절에서는 구현된 시스템의 정확도를 평가하고 얼굴 추적 시스템의 성능에 큰 영향을 주는 적정 Iteration 횟수를 알아보기 위한 각 Iteration 별 수렴 정도를 확인한다. 평가 방법으로는 AAM model 을 생성할 때 사용 했던 Training set 중 4 가지 형태(얼굴의 Pose 변화, 입의 변화)의 얼굴영상을 각각 10 장씩 선택하여 40 장으로 4 개의 그룹으로 만들어 Test set 을 구성했고 이것을 Warping 하여 결과로 나온 Shape 좌표와 Training set 을 구성할 때 직접

찍었던 Landmark 의 좌표 사이의 오차를 L2 norm 으로 계산하여 평가하였다.

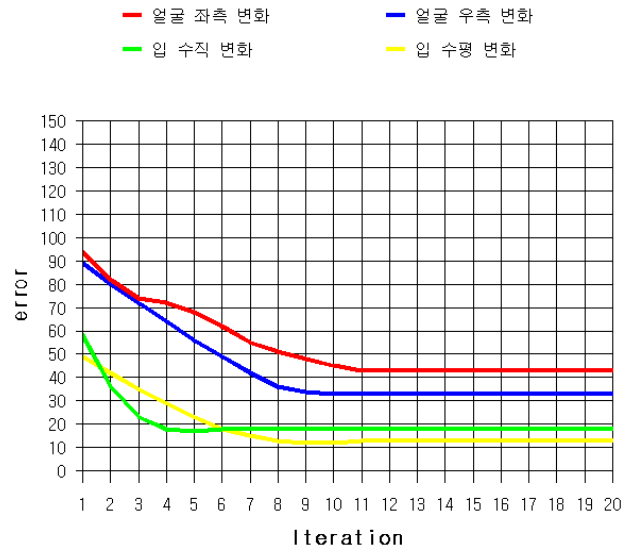


Iteration 3      Iteration 6      Iteration 10  
 error : 71.10    error : 18.59    error : 17.91

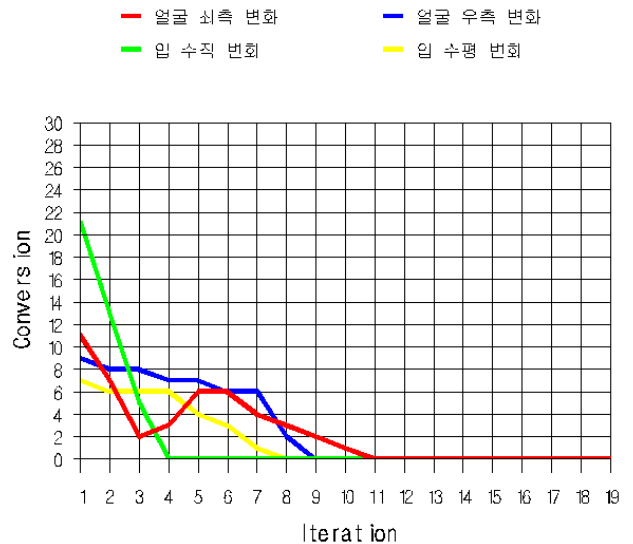
그림 3. Iteration 별 오차

(그림 3)은 평가를 위해 사용된 영상 중 한 영상에 대한 각 Iteration 별 Warping 수렴 결과와 오차를 보여주고 있다. 오차 계산은 얼굴의 형태에 따른 Warping 의 성능을 평가하기 위하여 각 그룹 별로 정확도를 확인할 수 있도록 하였고, 또한 Iteration 횟수를 20 회로 충분히 하여 시스템의 적정 Iteration 횟수를 알아 볼 수 있도록 하였다.

(그래프 1)은 시스템의 Warping 성능을 평가하기 위한 것으로서 동일 그룹의 Iteration 별 오차를 보여주고 있다. (그래프 1)을 보면 평균 7~8 회 에서부터 Iteration 별 오차가 40 대 이하로 떨어짐을 알 수 있다. 앞서 하나의 Shape 에 총 47 개의 정점을 사용 한다고 하였다. 이것으로 보았을 때 하나의 정점당 평균오차가 1 이하가 된다는 것을 알 수 있고 또한 7~8 회 정도의 Iteration 이면 Warping 이 적절하게 이루어 진다는 것도 알 수 있다. (그래프 2)는 시스템의 적정 Iteration 횟수를 알아보기 위한 자료로써 동일 그룹의 Iteration 별 수렴결과를 보여주고 있다. 보는 바와 같이 (그래프 2)에서도 (그래프 1)과 마찬가지로 7~8 회의 Iteration 에서 부터 Warping 수렴이 급격하게 작아짐을 볼 수 있는데 이것으로 미루어 보아 시스템의 최대 Iteration 횟수를 10 회 정도로 결정해도 별 무리가 없다는 것을 알 수 있다.



그래프 1. 그룹의 Iteration 별 오차



그래프 2. 그룹의 Iteration 별 수렴

## 6. 결론 및 향후 계획

제안된 얼굴 추적 시스템은 기존의 얼굴 추적 시스템들이 주로 사용한 기술의 단점을 보완할 수 있는 AAMs 을 사용하여 어두운 조명이나 밝은 조명 또는 얼굴 뒤편의 복잡한 배경등의 외부 환경에 크게 영향을 받지 않고, 얼굴의 위치, 자세와 함께 얼굴표정을 분석하는데 중요한 역할을 하는 입의 모양 변화에도 강건하게 대응하여 얼굴을 추적할 수 있는 얼굴 추적 시스템이다. 또한 기존의 얼굴 추적 시스템들이 정확한 얼굴이 아닌 얼굴영역 추적을 했다면 제안된 얼굴 추적 시스템은 deformable 한 model 을 연속적으로 갱신하며 정확한 얼굴을

추적할 수 있도록 하였다. 이 의미는 얼굴 추적 시스템이 지원하는 얼굴인식, 표정인식, चेस्चे인식등의 시스템에 얼굴의 여러 가지 정보를 정확하게 제공 함으로서 인식률을 높일 수 있게 한다는 것이다. 앞의 5 장에서 Iteration 별 오차와 수렴 정도를 실험을 통하여 알아 봄으로서 시스템 성능이 우수하다는 것을 입증 하였다. 그러나 본 시스템의 치명적 약점은 시스템 구동 최초에 입력되는 영상의 얼굴 영역이 Base shape 내에 어느 정도 포함되어야 한다는 것이다. 이에 향후 계획으로 최초 입력영상의 얼굴 영역에 Base shape 을 자동으로 맞추어주는 기술 연구가 필요하다.

## 7. 참고 문헌

[1] M. Turk, C. Hu, R. Feris "TLA Based Face Tracking", <http://ilab.cs.ucsb.edu/projects/turk/Turk%20et%20a1%202002.pdf>

[2] R. J. Qian, M. I Sezan, K. E. Matthews "A Robust Real-Time Face Tracking Algorithm", Proc IEEE, 1998.

[3] I. Matthews, S. Baker "Active Appearance Models revisited" International Journal of Computer vision, pages 135-164, 2004.

[4] S. Baker, I. Matthews "Lucas-Kanade 20 Years on: A Unifying Framework Part1", International Journal of Computer Vision, pages 221-255, 2004.

[5] G. J. Edwards, C. J. Taylor and T. F. Cootes "Interpreting Face Images using Active Appearance Models", In Proc. International Conference on Automatic Face and Gesture Recognition, pages 300-305, June 1998.

[6] M. Jones and T. Poggio. "Multidimensional morphable models" In Proceedings of the IEEE International Conference on Computer Vision, pages 683-688, 1998.

[7] S. Sclaroff and J. Isidoro. Active Blobs: "region-based, deformable appearance models" Computer vision and Image Understanding, 89(2/3):197-255, Feb. 2003.

[8] A. Lanitis, C. J. Taylor, and T. F. Cootes. "Automatic interpretation and coding of face images using flexible models" IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):742-756, 1997.

[9] T. F. Cootes, G. J. Edwards, and C. J. Taylor. "Active appearance models" IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6):681-685, June 2001.

[10] M. B. Stegmann and D. D. Gomez "A Brief Introduction to Statistical Shape Analysis" Online Lecture notes available from <http://www2.imm.dtu.dk/~aam/>, 2002.

[11] B. Lucas and T. Kanade. "An iterative image registration technique with an application to stereo vision" In Proceedings of the International Joint Conference on Artificial Intelligence, pages 674-679, 1981.

[12] S. Baker and I. Matthews. "Equivalence and efficiency of image alignment algorithms" In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 1090-1097, 2001.

[13] S. Z. Li and A. K. Jain, "Handbook of Face Recognition", pages 65-88, 2004

[14] T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham. "Active Shape Models - Their Training and Application" COMPUTER VISION AND IMAGE UNDERSTANDING, pages 38-59, 1995

[15] A. U. Batur and M. H. Hayes. "Adaptive Active Appearance Models" IEEE Transactions on Image Processing, 2005