

조명얼굴 영상을 위한 협력적 지역 능동표현 모델

(Collaborative Local Active Appearance Models for Illuminated Face Images)

양 준 영 [†]
(Junyoung Yang)

고 재 필 ^{**}
(Jaepil Ko)

변 혜 란 ^{***}
(Hyeran Byun)

요 약 얼굴영상 공간에서 얼굴영상들은 조명이나 포즈에 의해 비선형적 분포를 갖는다. 이들을 선형 모델에 기반을 둔 AAM으로 모델링 하는 것은 한계가 있다. 본 논문에서는 얼굴영상에 대한 몇 개의 군집이 주어졌다고 가정하고, 각 군집 별로 지역적인 AAM 모델을 구축하여 정합과정 중에 적합한 모델이 선택되도록 한다. 정합과정에서 발생하는 모델변경에 따른 모델간의 정합 인자 갱신의 문제는 인자 공간에서 모델간의 선형 관계를 미리 학습하여 해결한다. 심각한 정합 실패에 따른 잘못된 모델 선택을 줄이기 위해 점진적으로 모델변경이 이루어지도록 한다. 실험에서는 제안하는 방법을 Yale-B 조명얼굴 영상에 적용하여 모델을 생성하고 기존 방법과 정합 성능을 비교한다. 제안 방법은 심각한 그림자가 발생하는 강도 높은 조명얼굴 영상에서 성공적인 정합 결과를 보여주었다.

키워드 : 지역 Active Appearance Models, 얼굴조명보정, 얼굴정렬, 얼굴인식

Abstract In the face space, face images due to illumination and pose variations have a nonlinear distribution. Active Appearance Models (AAM) based on the linear model have limits to the nonlinear distribution of face images. In this paper, we assume that a few clusters of face images are given; we build local AAMs according to the clusters of face images, and then select a proper AAM model during the fitting phase. To solve the problem of updating fitting parameters among the models due to the model changing, we propose to build in advance relationships among the clusters in the parameter space from the training images. In addition, we suggest a gradual model changing to reduce improper model selections due to serious fitting failures. In our experiment, we apply the proposed model to Yale Face Database B and compare it with the previous method. The proposed method demonstrated successful fitting results with strongly illuminated face images of deep shadows.

Key words : Local Active Appearance Models, Face Illumination Compensation, Face Alignment, Face Recognition

· 이 논문은 2005년 교육인적자원부의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (KRF-2005-202-D00450)

[†] 학생회원 : 연세대학교 컴퓨터과학과
gundid@yonsei.ac.kr

^{**} 정 회 원 : 금오공과대학교 컴퓨터공학부 교수
nonezero@kumoh.ac.kr
(Corresponding author)

^{***} 종신회원 : 연세대학교 컴퓨터과학과 교수
hrbyun@yonsei.ac.kr

논문접수 : 2009년 7월 15일

심사완료 : 2009년 8월 15일

Copyright©2009 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 소프트웨어 및 응용 제36권 제10호(2009.10)

1. 서 론

Active Shape Models(ASM)[1] 및 Active Appearance Models(AAM)[2]은 얼굴영상 및 의료영상 분야에 성공적으로 적용되어 왔다. AAM은 학습영상의 모양을 나타내는 좌표집합과 이들로 정의된 밝기패턴(texture)에 대한 주성분 분석을 통해 외관(appearance) 변화를 모델링 한다. AAM은 몇 개의 인자를 통해 영상을 합성해 낼 수 있고 이러한 특성을 이용해 이전에 보이지 않았던 영상에 외관모델을 정합시킨다. 얼굴영상의 경우 정합결과로 얻어진 인자는 얼굴인식, 얼굴포즈 추정 및 추적, 표정 합성, 얼굴 정렬 등에 사용되어왔다. 정합속도 및 강인성 측면에서의 정합성능은 이러한 응

용에서 중요한 요소이다.

초기 AAM은 빠른 정합속도를 보장하기 위해 인자 갱신을 위한 변화도(gradient) 행렬을 매번 계산하지 않고 학습단계에서 미리 계산된 고정된 행렬을 사용하였다. 정합 최적화 문제에 Inverse-Compositional[3] 인자 갱신방법을 적용하여 정합의 매 단계마다 새롭게 계산해야 할 항을 대폭 줄이는 방법과[4], 동영상에서 3차원 정보를 추출하고 이를 활용한 방법이 발표되었다[5].

강인한 정합을 위해, Adaptive AAM[6]은 고정된 변화도 행렬을 포기하고 정합단계에서 이를 적응적으로 갱신하는 방법을 제안하였다. 정보를 추가하는 방식에서는 스테레오 카메라를 이용한 깊이 정보[7]를 활용하였다. 모양벡터 자체에 정보를 추가하는 방식에서는 각도 요소[8]나 모양벡터의 가시성 여부를 나타내는 가중치 요소[9,10]를 활용하였다. 그러나, 이러한 방법들은 [8, 9]를 제외하고는 모두 학습샘플의 분포를 가우시안으로 가정하였다.

실제 응용에서는 수작업에 의한 모양좌표 설정의 비-일관성, 포즈변화, 조명변화, 가려짐 등에 의해 학습샘플은 비선형분포를 따른다. 즉, 가우시안을 가정할 수 없다. 모양좌표의 비-일관성 문제를 해결하기 위한 방법으로 자동 모델 생성방법이 제안되었으며[10], 전역적인 외관 변화를 모델링하는 주성분 분석대신 지역적인 변화에 대처할 수 있도록 각각 독립성분 분석과 웨이블릿 분석을 도입하였다[11,12]. 학습샘플의 비선형적인 분포를 모델링하기 위해 믹스춰 모델에 기반한 방법도 발표되었다[12]. 모양변화의 비선형성을 모델링 하기 위해 Kernel-PCA를 적용하기도 하였다[8]. 그러나, 이 방법은 특정 커널에 맞추어진 Pre-Image[13] 추정을 필요로 하는 단점이 있다. 가우시안 혼합모델과 베이지안 혼합모델을 기반으로 하는 방법도[14] 발표되었으나, 여러 개의 지역적 모델을 융합된 형태로 간주하는 이들 혼합모델은 무효한 외관을 합성할 수 있다는 지적을 받았다[15]. 비선형 분포를 모델링 하는 또 다른 방법으로는 입력공간을 여러 개의 부분 공간으로 나누어 각각을 선형으로 모델링 하는 것이다. 이러한 방식으로 비선형 주성분 분석을 제안하여 효과적으로 비선형 분포를 모델링 하는 방법이 발표되었다[16]. 같은 맥락에서 [17]은 포즈 별로 독립적인 모델을 생성하고 목표영상에 적합한 모델을 선택하여 정합할 수 있도록 하였다.

본 논문에서는 [17]에서 동기를 얻어 얼굴영상의 조명 변화에 대처하기 위해 조명 별로 독립적인 모델을 생성하고 정합단계에서 점진적인 모델 선택방법을 제안한다. 제안하는 방법은 입력공간에서 지역적으로는 가우시안임을 가정하고 부분 그룹별로 AAM을 구축한다. 그리고, 정합의 매 반복마다 목표 밝기패턴에 적절한 AAM

모델을 선택함으로써 목표 밝기패턴에 적절한(plausible) 인자를 찾을 수 있도록 한다. 이것은 하나의 모델에서 완전히 정합이 이루어진 후 다른 모델을 선택하는 것이 아니라 정합과정 중에 모델변경이 일어날 수 있음을 의미한다. 이때, 매 반복마다 모델이 변경됨으로 인해 발생하는 정합 인자 갱신의 문제는 부분 그룹간의 관계를 미리 학습하여 해결한다.

유사한 영상들로 구성된 N 개의 부분 그룹이 주어지면 각 부분 그룹에 대한 N 개의 AAM 모델을 만들 수 있다. 모든 그룹간의 선형변환을 고려하면 부분 그룹이 많아질수록 요구되는 변환행렬 개수도 기하급수적으로 늘어난다. 또 다른 문제로는 심각한 정합 오류가 발생하는 경우 목표 밝기패턴에 적절한 모델이 선택되지 않을 수도 있다는 점이다. 본 논문에서는 정합과정 중에 급격한 부분 공간간의 이동은 없다고 가정하고 이웃 공간에 한정하여 모델변경이 점진적으로 이루어질 수 있도록 한다. 예를 들어, 좌측 조명영상을 정합하는 과정에서 갑자기 우측 조명영상 모델로 변경될 수 없다고 가정한다.

2장에서는 고유공간을 이용한 부분 그룹간의 선형관계 수립을 설명하고, 3장에서는 AAM을 간단하게 요약한다. 4장에서는 제안 방법을 설명하고, 5장에서는 조명얼굴 영상에 대해 모델 생성과정과 정합 실험결과를 보인다. 6장에서는 결론을 맺는다.

2. 고유공간기반의 선형조명 변환

다양한 조명을 받은 얼굴영상 $\mathbf{x} \in R^N$ 에 대한 집합이 주어졌을 때 고차원의 얼굴영상 공간은 조명의 위치에 따라 K 개의 부분그룹으로 구분될 수 있다고 가정하자. 부분그룹 i 에 속한 영상들을 열 벡터로 하여 행렬 $X_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iM})$ 라 하자. 조명이 i 에서 j 로 위치가 이동하면 이들의 관계는 선형으로 설명할 수 있다. 그러나, 조명의 위치가 선형으로 변환되었다고 해도 조명그룹 X_i 와 X_j 는 선형관계를 가정하기 어렵다.

$$X_j = T_{ji}X_i + E \quad (1)$$

여기서 에러 항 $E = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_M] \in R^{N \times M}$ 이고 각각의 에러 벡터 $\|\mathbf{e}_i\| \gg \theta_{\max}$ 이다. 이때 θ_{\max} 는 충분히 큰 값을 나타낸다. 그러나, 고차원의 영상공간을 저차원의 인자 공간으로 변환하면 상황은 달라질 수 있다.

$$A = \Phi^T X \quad (2)$$

$$A_j = M_{ji}A_i + E \quad (3)$$

주성분 분석을 통해 형성된 고유공간의 차원이 충분히 작은 경우 즉, 고유벡터로 구성된 Φ 의 컬럼 수가 적은 경우, 식 (3)에서 에러 항 $E = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_M]$ 의 모든 요소 크기는 매우 작다.

$$\|e_1\| \approx \|e_2\| \approx \dots \approx \|e_M\| \ll \theta_{\min} \quad (4)$$

즉, 각 에러 벡터 $\|e_i\| \approx 0$ 을 가정할 수 있다. 따라서, 식 (3)에서 에러 항을 제거하면 고유공간에서는 서로 다른 조명 그룹간의 선형관계를 가정할 수 있다.

$$A_j \approx M_{ji} A_i \quad (5)$$

여기서, M_{ji} 은 최소자승법을 적용하면 다음과 같이 계산할 수 있다.

$$M_{ji} = A_j A_i^T (A_i A_i^T)^{-1} \quad (6)$$

이때, 동일인에 대한 계수끼리 선형관계를 형성해 주고 신원정보를 담을 수 있도록 충분히 높은 차원을 유지한다.

그림 1은 각기 다른 조명 그룹에 속한 영상에 대하여, 고차원의 영상공간 및 저차원의 고유공간에서 각각 선형을 가정하고 최소자승법으로 변환행렬을 구한 후 변환한 결과를 보여준다. 고차원의 영상공간에서 변환한 영상은 얼굴의 형태조차 알아볼 수 없으나, 저차원의 고유공간에서의 변환한 영상은 신원정보까지 유지되는 것을 볼 수 있다.

제안 방법을 이용하면 입력된 얼굴영상이 어느 조명 그룹인지를 알면 고유공간에서 조명 그룹간의 선형관계를 이용하여 다른 조명의 얼굴영상을 합성할 수 있다.

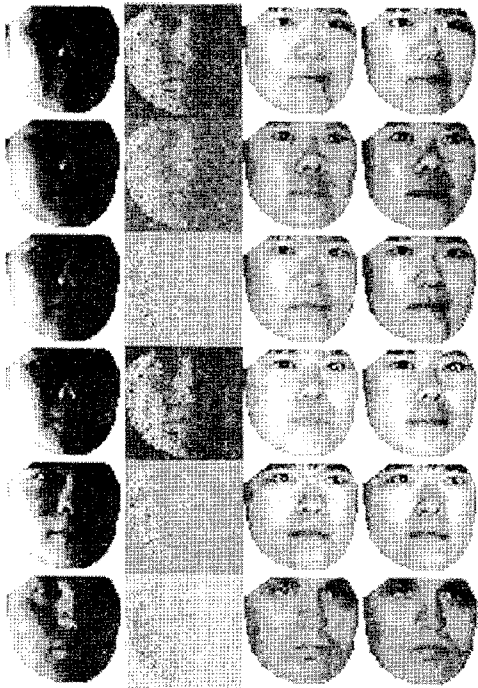


그림 1 영상공간과 고유공간에서의 조명 그룹간 선형변환 결과. 왼쪽에서 오른쪽으로 차례대로 입력영상, 고차원 영상공간에서 변환한 영상, 저차원 고유공간에서 변환한 영상, 변환대상 원본 영상

3. Active Appearance Models

본 장에서는 [6]의 식을 빌려 기본 AAM을 간단히 설명한다. l 개의 좌표 값으로 이루어진 영상 좌표계에서의 모양벡터 $s \in \mathbb{R}^{2 \times l}$ m 개에 대한 학습집합 $\{s_i\}_{i=1}^m$ 이 주어졌다고 하자. 이들에 대해 Procrustes Analysis[1]를 적용하여 정규화된 모양벡터 집합 $\{x_i\}_{i=1}^m$ 을 구하고, 다시 이들에 대해 주성분 분석을 적용하면 합성된 모양 벡터 x_s 는 다음과 같이 표현된다.

$$x_s = \bar{x} + E_x c_x \quad (7)$$

여기서, \bar{x} 는 평균모양 벡터이고, E_x 은 모양벡터 집합에 대한 공분산 행렬로부터 구한 고유벡터들을 열로 하는 행렬이다. c_x 는 모양 인자 벡터이다. 정규화 좌표계에서 합성 모양벡터 x_s 를 영상 좌표계로 변환하기 위해서는 어파인(affine) 변환 $S_g(\cdot)$ 을 이용한다. 여기서, g_s 는 어파인 변환을 위한 인자 벡터이다.

$$s = S_g(x_s) \quad (8)$$

밝기패턴 변화를 모델링하기 위해서는 s 에 의한 위평 함수 W_s 를 이용하여 각각의 학습 밝기패턴 Γ 를 평균 모양으로 위평하여 평균모양에 대한 밝기패턴[2] 집합 $\{\Gamma_w^i\}_{i=1}^m$ 을 얻는다. 이에 대하여 다시 밝기패턴 정규화 함수 $T_g(\cdot)$ 를 적용하여 $\{t_i\}_{i=1}^m$ 를 구하고, 여기에 주성분 분석을 적용하면 다음과 같이 t_s 를 합성할 수 있다. 여기서, g_t 는 밝기패턴 정규화 인자 벡터이다.

$$t_s = \bar{t} + E_t c_t \quad (9)$$

여기서, \bar{t} 는 평균 밝기패턴 벡터이고, E_t 은 $\{t_i\}_{i=1}^m$ 에 대한 공분산 행렬로부터 구한 고유벡터들을 열로 하는 행렬이다. c_t 는 밝기패턴 인자 벡터이다.

마지막으로, 외관변화에 대한 모델링은 모양과 밝기패턴 인자 벡터를 연결한 $a = (\Lambda c_x c_t)^T$ 에 대한 학습집합 $\{a_i\}_{i=1}^m$ 에 대해 다시 한번 주성분 분석을 통해 수행된다. 여기서, Λ 은 모양과 밝기패턴 인자간의 단위 차이를 보정하는 대각행렬이다. 합성된 외관인자 a_s 는 다음과 같이 표현된다.

$$a_s = E_a c_a \quad (10)$$

여기서, E_a 은 $\{a_i\}_{i=1}^m$ 에 대한 공분산 행렬로부터 구한 고유벡터들을 열로 하는 행렬이다. c_a 는 외관인자 벡터이다. 합성된 모양과 밝기패턴은 다음과 같이 외관인자 c_a 로 표현될 수 있다.

$$\begin{aligned} \mathbf{x}_s &= \bar{\mathbf{x}} + \mathbf{Q}_x \mathbf{c}_a \\ \mathbf{t}_s &= \bar{\mathbf{t}} + \mathbf{Q}_t \mathbf{c}_a \end{aligned} \quad (11)$$

여기서 $\mathbf{Q}_x = \mathbf{E}_x[\Lambda^{-1} | \mathbf{0}] \mathbf{E}_a$ 이고 $\mathbf{Q}_t = \mathbf{E}_t[\mathbf{0} | \mathbf{I}] \mathbf{E}_a$ 이다.

정규화 좌표계에서의 영상은 지금까지 소개된 모든 인자를 표현한 인자 벡터 $\mathbf{p} = (\mathbf{g}_s, \mathbf{g}_t, \mathbf{c}_a)^T$ 로 표현될 수 있다. 이제, \mathbf{P} 를 변화시킴으로써 학습되었던 영상과 유사한 영상을 합성할 수 있다.

입력영상에 대해 최적의 \mathbf{P} 를 찾는 과정이 AAM 정합이다. 목표 영상 Γ 가 주어지면 초기 인자 \mathbf{P} 로부터 합성된 모양벡터 \mathbf{x}_s 로부터 영상 좌표계에서의 모양벡터 \mathbf{s} 를 구한 후, 이를 다시 정규화 좌표계로 워핑하여 정규화 목표 밝기패턴 \mathbf{t}_n 을 다음과 같이 얻는다.

$$\mathbf{t}_n = \mathbf{T}_{\mathbf{g}_t}^{-1}(\mathbf{W}_s(\Gamma)) \quad (12)$$

그러면, \mathbf{t}_n 과 \mathbf{P} 로부터 합성된 밝기패턴 \mathbf{t}_s 와의 차이는 인자 벡터 \mathbf{P} 에 대한 식으로 표현된다.

$$\mathbf{r}(\mathbf{p}) = \mathbf{t}_n - \mathbf{t}_s \quad (13)$$

정합의 목표는 식 (13)을 최소화하는 \mathbf{P} 를 찾는 것으로서 비용함수를 $\|\mathbf{r}(\mathbf{p})\|^2$ 로 두어 다음과 같이 계산한다. 먼저, $\mathbf{r}(\mathbf{p})$ 을 1차 테일러 확장한 $\mathbf{r}(\mathbf{p} + \delta\mathbf{p}) \cong \mathbf{r}(\mathbf{p}) + (\partial\mathbf{r}/\partial\mathbf{p})\delta\mathbf{p}$ 으로 대치한다. 그러면, 정합의 목표는 $\|\mathbf{r}(\mathbf{p} + \delta\mathbf{p})\|^2$ 를 최소화 하는 $\delta\mathbf{p}$ 를 찾는 것으로 바뀐다. $\delta\mathbf{p}$ 는 다음과 같이 계산된다.

$$\begin{aligned} \delta\mathbf{p} &= -\mathbf{R}\mathbf{r}(\mathbf{p}), \\ \mathbf{R} &= \left(\frac{\partial\mathbf{r}^T}{\partial\mathbf{p}} \frac{\partial\mathbf{r}}{\partial\mathbf{p}} \right)^{-1} \frac{\partial\mathbf{r}^T}{\partial\mathbf{p}} \end{aligned} \quad (14)$$

여기서 최적의 인자 \mathbf{P} 는 $\mathbf{P} + \delta\mathbf{p}$ 를 반복하여 얻어진다. 이때, 매 반복마다 변화도 행렬 \mathbf{R} 의 계산이 요구되는데, [2]은 인자 변화 $\delta\mathbf{p}$ 와 영상변화 $\delta\mathbf{r}$ 간의 관계를 선형으로 간주하고 학습영상으로부터 미리 구하여 실시간 정합을 달성 하였다.

4. 지역 Active Appearance Models

영상의 분포가 여러 개의 군집을 이루는 경우, AAM 은 선형변환에 기반을 두고 있기 때문에 모델 하나만으로는 성공적인 정합을 할 수 없다. 본 논문에서는 각 군집에 대응되는 부분 그룹별로 AAM 모델을 구축하고 정합의 매 반복마다 목표 밝기패턴에 적절한 AAM 모델을 선택할 수 있도록 한다.

이러한 접근법에서 해결해야 할 문제는 다음과 같다.

첫째 정합의 매 반복마다 AAM 모델 변경이 이루어질 수 있기 때문에 모델 변경에 따른 모델간 외관인자 호환 문제가 발생한다. 둘째, 심각한 정합 오류가 발생하는 경우 목표 밝기패턴에 적절한 모델이 선택되지 않을 수도 있다. 예를 들어, 왼쪽 조명그룹의 모델에서 갑자기 오른쪽 조명그룹의 모델로 변경될 수 있다.

인자 호환문제는 2장에서 설명한 고유공간에서의 부분 그룹간 선형관계를 수립함으로써 해결한다. 정합과정 중에 급격한 부분 공간 이동은 없다고 가정하고 이웃 공간에 한정하여 모델변경이 이루어질 수 있도록 한다. 즉, 좌측 조명영상을 정합하는 과정에서 갑자기 우측 조명영상 모델로 변경되는 경우는 심각한 정합실패를 제외하고는 있을 수 없다고 가정한다. 이것은 적절한 모델 선택의 문제를 전체 N 개의 모델 중에서 하나를 선택하는 N 클래스 분류 문제로 간주하지 않고 현 모델의 이웃 모델 중에서 선택하는 M 클래스 분류 문제로 간주하는 것이다. 이렇게 함으로써 정합의 매 반복을 통해 점진적으로 최적 해를 구할 수 있다. 5장에서 설명하는 얼굴조명 그룹에서는 M 을 3으로 설정하였다.

N 개의 부분그룹이 주어졌다고 하자. 부분그룹 i 에 속한 학습집합에 대하여 AAM_i 모델을 생성하면 합성된 모양 \mathbf{x}_s^i 및 밝기패턴 \mathbf{t}_s^i 는 식 (11)의 표식을 빌리면 다음과 같이 외관인자 \mathbf{c}_a^i 로 표현된다.

$$\begin{aligned} \mathbf{x}_s^i &= \bar{\mathbf{x}} + \mathbf{Q}_x^i \mathbf{c}_a^i \\ \mathbf{t}_s^i &= \bar{\mathbf{t}} + \mathbf{Q}_t^i \mathbf{c}_a^i \end{aligned} \quad (15)$$

정합 과정에서 모델 변경이 이루어지면 새로운 모델의 인자를 현재 모델의 인자로부터 추정해야 하는데, 이를 위해 미리 그룹간의 선형관계를 구축하여 둔다. 문제는 모든 그룹간의 선형변환을 고려하면 $N(N-1)/2$ 개의 변환이 필요하므로 부분 그룹이 많아질수록 요구되는 변환의 개수는 기하급수적으로 늘어난다는 점이다. 다음과 같은 방법을 통해 이 문제를 해결할 수 있다.

평균그룹: N 개 부분그룹의 평균을 구하고 이 그룹을 통해서 서로 다른 두 부분그룹간의 변환이 이루어지도록 한다. 이렇게 함으로써, 변환의 개수를 N 개로 줄일 수 있다. 그림 2는 8개의 부분 그룹에 대해 평균 그룹을 도입하였을 때와 그렇지 않을 때의 변환의 수를 보여준다(조명그룹에서는 정면조명 그룹을 평균그룹으로 활용한다).

이웃그룹: 부분그룹들의 분포가 그림 3과 같이 곡면 경로를 따르고 정합과정에서 변경될 수 있는 그룹간의 범위를 두 이웃으로 한정하면 변환행렬의 개수를 $N-1$ 개로 줄일 수 있다. 조명 및 포즈변화 얼굴영상을 2차원으로 표현한 MLE[18]의 실험결과로부터 얼굴 조명 그룹들 간의 이웃 그룹은 충분히 가정할 수 있다.

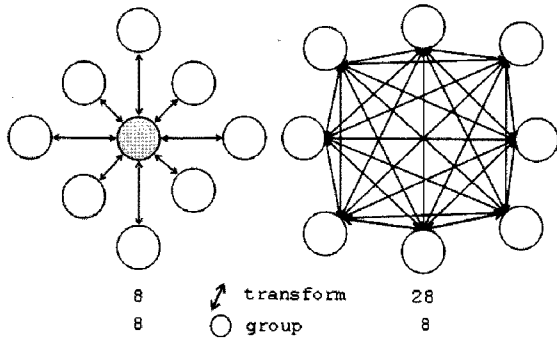


그림 2 평균그룹과 전체그룹 연결방식에 따른 변환의 수 비교

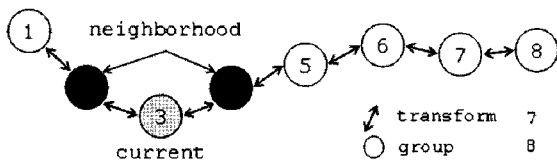


그림 3 이웃그룹 연결방식에 따른 변환의 수

두 방법은 각각 다음과 같은 단점이 있다. 평균그룹 방법은 한 그룹에서 다른 그룹으로 변환할 때 평균그룹을 거쳐 간접적으로 변경되기 때문에 두 번째 경우처럼 직접 변경되는 경우보다 변환 여려가 더 크다. 이웃그룹 방법은 모델선택의 범위를 두 이웃만으로 한정함으로써 수렴 속도가 느리다. 본 논문에서는 다음과 같은 절충안을 제시한다.

먼저, 부분그룹 i 에 속한 l 개의 샘플들에 대한 외관인자 벡터 \mathbf{c}_a^i 로 이루어진 행렬 \mathbf{A}_i 를 다음과 같이 정의하자.

$$\mathbf{A}_i = [\mathbf{c}_{a_1}^i, \mathbf{c}_{a_2}^i, \dots, \mathbf{c}_{a_l}^i] \quad (16)$$

여기서 \mathbf{c}_a^i 는 이미 주성분 분석을 통해 차원이 충분히 줄어든 상태이므로 부분그룹 i 와 부분그룹 j 는 식 (5)와 같이 선형관계를 가정할 수 있으며 이들 부분그룹 간 변환 M_{ij} 는 식 (6)을 통해 계산할 수 있다.

부분 그룹별로 AAM 모델과 각 부분 그룹간의 선형 변환 행렬 M 이 주어졌다고 하자. 먼저 정합의 매 반복 과정에서 이루어지는 모델 변경은 다음과 같다.

현재 정합된 모델이 AAM_i 라면 이 모델의 정보를 식 (12)에 적용하여 목표 밝기패턴 \mathbf{t}_n^i 을 생성한다. 현재 모델과 이웃 관계에 있는 모델의 평균을 $\bar{\mathbf{t}}_{k \in G}$ 라 하면 \mathbf{t}_n^i 과 $\bar{\mathbf{t}}_k$, 그리고 AAM_i 의 정규직교 행렬 Φ_i 을 이용해 \mathbf{t}_n^i 과 $\bar{\mathbf{t}}_k$ 의 거리를 계산한다. 여기서, $G = \{i-1, i, i+1\}$ 는 현재 모델과 이웃 모델의 색인 집합이다. 거리계산은

식 (17)과 같이 DFFS[19]를 사용하여 이웃 그룹간의 거리를 각각 계산한다. 이들 중 거리가 최소가 되는 모델의 색인을 k^* 라고 하자.

$$DFFS(\mathbf{t}_n^i, \bar{\mathbf{t}}_k, \Phi_i) = \|\mathbf{t}_n^i - \bar{\mathbf{t}}_k - \Phi_i \Phi_i^T (\mathbf{t}_n^i - \bar{\mathbf{t}}_k)\| \quad (17)$$

$$k^* = \underset{k}{\operatorname{arg\,min}} DFFS(\mathbf{t}_n^i, \bar{\mathbf{t}}_k, \Phi_i) \quad (18)$$

만약 $i+1$ 모델이 선택되었다면, 즉 k^* 가 $i+1$ 라면 $G = \{i, i+1, i+2\}$ 로 놓고 다시 DFFS를 계산하여 최적의 모델 k^* 을 선택한다. 더 이상 모델 변경이 이루어지지 않을 때까지 이를 반복한다. 만약 최종적으로 j 모델이 선택되었다면 현재의 외관인자 \mathbf{c}_a^j 를 \mathbf{c}_a^i 와 M_{ji} 를 통해 다음과 같이 계산한다.

$$\mathbf{c}_a^j = M_{ji} \mathbf{c}_a^i \quad (19)$$

$$M_{ji} = M_{jm} M_{mi} \quad (20)$$

여기서, M_{mi} 는 부분그룹 i 에서 평균그룹 m 으로의 변환이고 M_{jm} 는 평균그룹 m 에서 부분그룹 j 로의 변환을 나타낸다. 이러한 방식을 통해 모델이 여러 번 변경되는 동안 한번만 인자를 갱신하므로 수렴 속도를 빠르게 할 수 있다.

새로운 모델에서의 외관인자 \mathbf{c}_a^j 를 추정하였다면 이를 이용해 선택된 모델 AAM_j 에서 다시 정합을 수행한다. 수렴 임계치에 도달할 때까지 이 과정을 반복한다. 제안하는 정합 알고리즘은 그림 4에 제시한다.

5. 얼굴영상에 대한 모델 생성 및 실험결과

5.1 모델생성

모델생성과 제안방법의 성능을 평가하기 위하여 Yale-B[20]를 선택하였다. 본 데이터베이스는 10명의 사람에 대하여 9개의 포즈와 64개의 조명 조건으로부터 촬영된 총 5760장의 영상으로 구성되어 있다. 사람, 포즈, 조명 별로 파일명이 지정되어 있다. 예를 들어, 파일명 "YaleB_P02+035E-10"에서, P02는 두 번째 사람임을, +035E는 조명의 방위각을, -10은 조명의 높이를 나타낸다. 본 논문에서는 64개의 조명조건 중 45개를 선택하고 여기서 다시 각 조명 그룹별로 74장의 영상을 선별하여 총 3330장의 영상을 수집하였다. 그림 5는 수집한 영상의 일부를 보여준다.

이렇게 수집한 각 영상에 대하여 수작업으로 64개의 얼굴좌표를 설정한다. 설정한 얼굴좌표의 위치는 그림 6에 제시한다.

1. 목표 영상 Γ 가 주어진다. 부분그룹의 색인을 $i \leftarrow 0$ 으로 설정한다. 그리고, 현재 AAM_i 모델의 인자 벡터를 \mathbf{p}_c 라고 하고 이로부터 외관인자 \mathbf{c}_a^i 를 얻는다.
2. 정합을 통해 얻은 \mathbf{c}_a^i 를 이용해 목표 영상을 정규화 좌표계로 워핑하여 목표 밝기패턴 \mathbf{t}_n^i 을 얻는다.
3. 정규화 좌표계에서 밝기패턴 \mathbf{t}_n^i 를 합성한다.
4. 에러 $\mathbf{r} = \mathbf{t}_n^i - \mathbf{t}_n^j$ 를 계산한다.
5. 부분그룹 i 의 변화도 행렬 \mathbf{R}_i 를 이용해 $\delta \mathbf{p}_c = -\mathbf{R}_i \mathbf{r}$ 를 계산한다.
6. 인자를 $\mathbf{p}_c \leftarrow \mathbf{p}_c + \delta \mathbf{p}_c$ 으로 갱신한다. 그리고, 새로운 에러 \mathbf{r}_{new} 를 계산한다.
7. 에러가 임계치 보다 작으면 종료하고 그렇지 않으면 최적 모델을 선정하기 위해 다음 단계를 수행한다.
8. 현재 모델을 기준으로 이웃모델을 선정하기 위해 $k \leftarrow i$ 로 설정한다.
9. 모델변경 대상이 되는 모델을 $AAM_{k-1}, AAM_k, AAM_{k+1}$ 으로 한정한다.
10. AAM_k 모델과 이웃 모델 AAM_{k-1}, AAM_{k+1} 의 평균 밝기패턴 $\bar{\mathbf{t}}_{k-1}, \bar{\mathbf{t}}_k, \bar{\mathbf{t}}_{k+1}$ 그리고 \mathbf{t}_n^i 을 이용해 (17)식의 DFFS를 각각 계산하고, (18)식을 이용해 최적의 모델 k^* 를 선정한다.
11. 만약 $k^* \neq k$, 즉 현재 모델이 최적모델이 아니라면 $k \leftarrow k^*$ 설정하고 단계 9로 돌아간다. 그렇지 않으면 모델변경과 이에 따른 파라미터 갱신을 처리한다.
12. $j \leftarrow k^*$ 설정하여 k^* 를 최종 변경대상이 되는 모델로 선정한다.
13. (20)식을 이용하여 AAM_i 와 AAM_j 의 관계를 나타내는 M_{ji} 를 계산한다.
14. M_{ji} 을 이용하여 \mathbf{c}_a^i 로 부터 \mathbf{c}_a^j 를 계산한다.
15. $\mathbf{c}_a^i \leftarrow \mathbf{c}_a^j, i \leftarrow j$ 로 갱신하고 단계 2로 돌아간다.

그림 4 제안 알고리즘

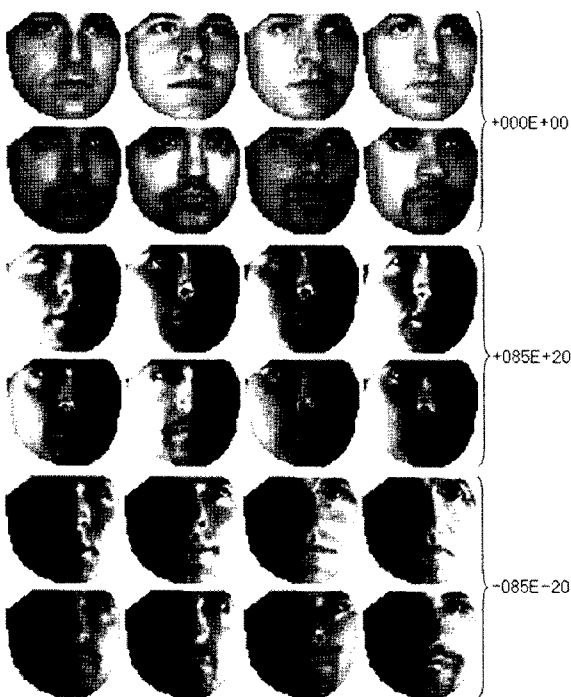


그림 5 Yale-B 영상의 그룹별 영상의 예

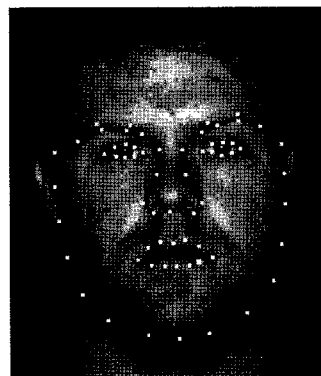


그림 6 64개의 얼굴좌표 위치

주어진 데이터를 정리해 보면 다음과 같다. 하나의 영상은 64개의 얼굴좌표로 이루어지므로 모양벡터의 \mathbf{s} 는 128차원이다. 74장의 영상이 하나의 조명그룹에 포함되므로 조명그룹 A_i 는 $\{\mathbf{s}_i\}_{i=1}^{74}$ 의 각 원소를 열로 하여 $A_i \in \{A_i\}_{i=0}^{44}$ 와 같이 45개를 생성한다.

각기 다른 조명그룹의 얼굴영상은 조명 조건만 달리

하여 동시에 촬영하였기 때문에 얼굴의 위치는 모두 같다. 따라서, 모양모델은 A_0 에 해당하는 "+000E+00" 그룹의 얼굴영상 74장을 이용하여 한번만 생성한다. 주성분 분석결과 모양 인자 c_x 의 차원은 10차원으로 하였다. 이는 전체 변화량의 90%에 해당한다. 그림 7은 상위 3개 차원에서의 모양 변화를 보여준다.

밝기패턴 모델링은 각 그룹별로 이루어진다. 각 그룹에 대한 모델생성 과정은 동일하다. 어떤 그룹의 모양벡터 집합 $\{s_i\}_{i=1}^{74}$ 의 모양벡터의 평균을 구하고, [23]에서 제시한 어파인 워핑을 이용하여 평균모양에 대한 밝기패턴 집합 $\{\Gamma_w^i\}_{i=1}^{74}$ 를 얻어 낸다. 여기서, 각 Γ_w^i 는 2478개의 픽셀 값으로 구성하였다. 그러면, $\{\Gamma_w^i\}_{i=1}^{74}$ 의 각 원소를 열로 하는 2478×74 차원의 행렬을 만들어 밝기패턴 변화를 분석하는데 사용한다. 주성분 분석결과, 밝기패턴 인자 c_b 의 차원은 20차원으로 하였다. 이는 각 그룹별 변화량이 모두 90%이상 이 되도록 하는 차원이다.

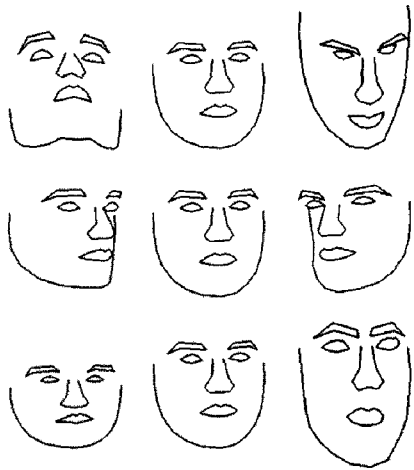


그림 7 상위 3개 차원에서의 모양변화에 예

외관 모델링도 각 그룹별로 이루어진다. 모양 인자와 밝기패턴 인자를 연결하여 30차원의 외관 학습집합 $\{a_i\}_{i=1}^{74}$ 을 구성한다. 이 집합의 각 원소를 열로 하는 30×74 차원의 행렬에 대해 다시 주성분 분석을 적용하여 외관 인자 c_a 를 얻는다. 이것도 역시 20차원으로 설정하였다.

실시간 AAM 정합을 위해서는 식 (14)의 외관 인자 변화와 밝기패턴 변화에 대한 변화도 행렬 R 을 미리 계산해 둘 필요가 있다. 이 역시 조명 그룹별로 모두 구한다. 각 조명 그룹별 R_i 는 해당 그룹에 속한 74장의 학습영상으로부터 계산한다. 외관 인자 변화 범위를 외관 인자 분산의 ± 1 배 사이로 하여 예러 영상을 구하기

를 5번 반복하고 이들의 평균을 계산한다.

마지막으로, 조명 그룹간의 선형변환 행렬 M 의 계산은 밝기패턴 인자 20차원의 c_b 를 사용한다. 평균그룹을 A_0 선정하여, 나머지 44개의 그룹 A_i 에서 A_0 로 변환하는 M_{0i} , $1 \leq i \leq 44$ 와 A_0 에서 A_i 로 변환하는 M_{i0} , $1 \leq i \leq 44$ 를 식 (6)를 적용하여 계산한다.

5.2 실험결과

주어진 실험 데이터베이스는 10명에 대해 획득한 영상들이다. 정합성능을 평가하기 위해, 10명 중 한 사람을 선택하여 이 사람에게 해당되는 영상을 테스트 집합으로, 나머지 사람에게 해당되는 영상을 학습 집합으로 사용하였다. 그리고, 이 과정을 각 사람에 대해 10번 반복하여 평균을 계산하였다.

성능평가는 초기 모델의 모양벡터 s 에 대해 다음과 같은 변화를 주어 정합한 후 정합결과 x_s 과의 정합에러를 측정한다. 먼저, 이동변화는 x축과 y축으로 각각 -5~+5픽셀 범위 안에 4번, 크기변화는 0.95~0.8 범위 안에 4번, 각도변화는 -5~범위 안에 4번 주었다. 정합에러는 단순히 대응되는 얼굴좌표 간의 유클리드 거리 $\|s - x_s\|$ 로 계산하였다. 그림 8 및 표 1은 3330장의 영상을 모두 사용하여 학습한 기존의 AAM[2]과 제안방법의 정합성능을 비교한다.

그림 8 및 표 1에 따르면 모델의 초기위치에 대한 이동, 회전, 크기 변화에 대해 제안방법의 에러가 전체 공간을 모두 학습하는 기존의 방법보다 월등히 낮은 것을

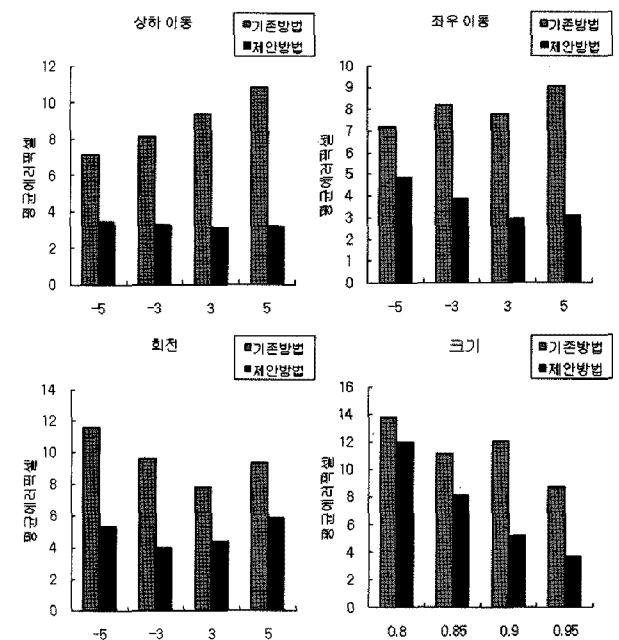


그림 8 모델 초기 위치의 이동, 회전, 크기 변화에 따른 정합에러

표 1 모델 초기 위치의 이동, 회전, 크기 변화에 따른 정합에러

상하이동		-5	-3	3	5
기존방법	평균	7.16	8.16	9.39	10.82
제안방법	평균	3.46	3.26	3.13	3.17
좌우이동		-5	-3	3	5
기존방법	평균	7.16	8.16	7.78	9.02
제안방법	평균	4.85	3.87	2.95	3.06
회전		-5	-3	3	5
기존방법	평균	11.55	9.53	7.80	9.29
제안방법	평균	5.28	3.93	4.30	5.85
크기		0.8	0.85	0.9	0.95
기존방법	평균	13.81	12.05	11.16	8.73
제안방법	평균	11.93	8.13	5.17	3.67

알 수 있다. 기존방법과 제안방법의 수행시간은 펜티엄 PC 2G Hz에서 각각 0.4초와 1.4초로 제안방법의 수행 시간이 약 3배 정도 증가되었다.

그림 9는 제안방법의 정합 예를 보여준다. 심각한 그림자가 발생하는 강도 높은 조명얼굴 영상에도 불구하고 성공적인 정합을 보여준다.



그림 9 정합 결과. 왼쪽은 입력영상, 오른쪽은 정합합성 영상

6. 결론

본 논문에서는 샘플공간에서의 비선형적 분포를 모델링하기 위해 부분 그룹별로 지역적인 AAM 모델을 구축하였다. 정합단계에서는 정합수렴과정 중에 모델변경이 이루어 질 수 있도록 하였으며, 이를 위해 부분 그룹간의 관계를 AAM의 인자 공간에서 선형으로 모델링하고 이웃모델로 한정된 점진적 모델선택 방법을 제시하였다. 제안방법을 Yale-B 조명얼굴 영상에 적용하여 지역적으로 모델링 하는 방법의 우수성을 보였다. 제안방법은 조명이 일정하지 않은 상황에서 얼굴의 조명보정, 정렬, 인식 등에 적용할 수 있을 것으로 기대한다. 그러나, 제안방법은 부분 그룹별로 군집이 되어 있는 데이터를 필요로 하며 기존의 방법에 비해 정합속도 저하가 있다.

참고 문헌

- [1] T. F. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models—their training and their applications," *Computer Vision and Image Understanding*, vol.61, no.1, pp.38-59, 1995.
- [2] T. F. Cootes, D. J. Edwards, and C. J. Taylor, "Active Appearance Models," *IEEE Trans. on PAMI*, vol.23, no.6, pp.681-685, 2001.
- [3] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *Int'l Journal of Computer Vision*, vol.56, no.3, pp.221-255, 2004.
- [4] I. Matthews and S. Baker, "Active Appearance Models Revisited," *Int'l Journal of Computer Vision*, vol.60, no.2, pp.135-164, 2004.
- [5] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-Time Combined 2D+3D Active Appearance Models," *Proc. of IEEE Int'l Conf. on CVPR*, vol.2, pp.535-542, 2004.
- [6] A. U. Batur and M. H. Hayes, "Adaptive Active Appearance Models," *IEEE Trans. on Image Processing*, vol.14, no.11, pp.1707-1720, 2005.
- [7] J. Liebelt, J. Xiao, and Jie Yang, "Robust AAM Fitting by Fusion of Images and Disparity Data," *Proc. of the IEEE Int'l Conf. on CVPR*, vol.2, pp. 2483-2490, 2006.
- [8] S. Gong, A. Psarrou, and S. Romdhani, "Corresponding dynamic appearances," *Image and Vision Computing*, vol.20, pp.307-318, 2002.
- [9] Y. Zhou, W. Zhang, X. Tang, and H. Shum, "A Bayesian Mixture Model for Multi-view Face Alignment," *Proc. of IEEE Int'l Conf. on CVPR*, vol.2, pp.741-746, 2005.
- [10] S. Baker and I. Matthews, "Automatic Construction of Active Appearance Models as an Image Coding Problem," *IEEE Trans. on PAMI*, vol.26, no.10, pp.1380-1384, 2004.

- [11] C. Hu, R. Feris, and M. Turk, "Real-time View-based Face Alignment using Active Wavelet Networks," *Proc. of IEEE Int'l Workshop on Analysis and Modeling of Faces and Gestures*, pp. 215-221, 2003.
- [12] M. Uzumcii, A. F. Frangi, M. Sonka, J. H. Reiber, and B.P. Lelieveldt, "ICA vs. PCA Active Appearance Models: Application to Cardiac MR Segmentation," *LNCS 2878*, pp.451-458, 2003.
- [13] S. Mika, B. Scholkopf, A. Smola, G. Ratsch, K. Muller, M. Scholz, and G. Ratsch, "Kernel-PCA and De-Noising in Feature Spaces," *Advances in Neural Information Processing Systems*, vol.11, pp.536-542, 1999.
- [14] C. M. Christoudias and T. Darrell, "On Modeling Nonlinear Shape-and-Texture Appearance Manifolds," *Proc. of the IEEE Int'l Conf. on CVPR*, vol.2, pp. 1067-1074, 2005.
- [15] C. Butakoff and A. Frangi, "A Framework for Weighted Fusion of Multiple Statistical Models of Shape and Appearance," *IEEE Trans. on PAMI*, vol.28, no.11, 2006.
- [16] R. Bowden, T.A. Mitchell, and M. Sahardi, "Cluster based non-linear principal component analysis," *IEE Electronics Letters*, vol.33, no.22, pp. 1858-1859, 1997.
- [17] T.F. Cootes, G.V. Wheeler, K.N. Walker, and C.J. Taylor, "View-based active appearance models," *Image and Vision Computing*, vol.20, pp.657-664, 2002.
- [18] Z. Zhang, and J. Wang, "MLLE: Modified Locally Linear Embedding Using Multiple Weights," *Proc. of Int'l NIPS*, vol.19, pp.1618-1625, 2006.
- [19] K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Trans. on PAMI*, vol.20, no.1, pp.39-51, 1998.
- [20] Yale Face Database B: <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>.

변 해 란

정보과학회논문지 : 소프트웨어 및 응용
제 36 권 제 8 호 참조



양 준 영

2005년 2월 연세대학교 컴퓨터과학과 석사. 2005년 3월~현재 연세대학교 컴퓨터과학과 박사과정. 관심분야는 패턴인식, 영상처리



고 재 필

2004년 2월 연세대학교 컴퓨터과학과 박사. 2004년 3월~현재 금오공과대학교 컴퓨터공학부 조교수. 관심분야는 패턴인식