3D 모델과 Optical flow 를 이용한 실시간 얼굴 모션 추정

권오륜[1], 전준철[2]
경기대학교 정보과학부 컴퓨터그래픽스 연구실[1 2]
{kor5663[1], jcchun[2]}@kyonggi.ac.kr

Pose Estimation of Face Using 3D Model and Optical Flow in Real Time

Oh Ryun Kwon [1], Jun Chul Chun[2]
Computer Graphics Laboratory, Information Science Department
Kyonggi University [1 2]

요 약

HCI, 비전 기반 사용자 인터페이스 또는 제스쳐 인식과 같은 많은 분야에서 3 차원 얼굴 모션을 추정하는 것은 중요한 작업이다. 연속된 2 차원 이미지로부터 3 차원 모션을 추정하기 위한 방법으로는 크게 외형 기반 방법이나 모델을 이용하는 방법이 있다.

본 연구에서는 동영상으로부터 3 차원 실린더 모델과 Optical flow 를 이용하여 실시간으로 얼굴 모션을 추정하는 방법을 제안하고자 한다. 초기 프레임으로부터 얼굴의 피부색과 템플릿 매칭을 이용하여 얼굴 영역을 검출하고 검출된 얼굴 영역에 3 차원 실린더 모델을 투영하게 된다. 연속된 프레임으로부터 Lucas-Kanade 의 Optical flow 를 이용하여 얼굴 모션을 추정한다. 정확한 얼굴 모션 추정을 하기 위해 IRLS 방법을 이용하여 각 픽셀에 대한 가중치를 설정하게 된다. 또한, 동적 템플릿을 이용해 오랫동안 정확한 얼굴 모션 추정하는 방법을 제안한다.

Keyword : Pose estimation, 3D Model ,Optical flow, Real time

1. 서 론

3 차원 얼굴의 포즈를 추정하는 것은 비전 기반의 컨트롤, 인간과 컴퓨터간의 인터렉션과 감시 시스템 등 컴퓨터 비전 분야의 많은 애플리케이션을 위한 중요한 작업 중에 하나이다. 얼굴 인식이나 얼굴 표정의 분석의 문제점을 3 차원 얼굴 포즈 추정을 통해 수정된 이미지를 사용한다면 좀 더 이러한 문제들을 해결할 수 있다. 3 차원 얼굴의 위치와 방향을 결정하는 얼굴의 포즈 추정은 비전 기반의 사용자 인터페이스나 얼굴의 제스처 인식과 같은 것을 개발할 때 기본이 된다. 하지만 3 차원 얼굴 포즈를 추정하는 많은 애플리케이션의 문제점은 얼굴의 방향이나 크기 변화와 같은 얼굴 모션에 대한 정확한 추정하는 것이 어렵다는 것이다. 또한 대부분 실시간으로 추정하기에는 너무나 많은 연산시간을 필요로 한다는 문제점을 가지고 있다. 따라서 얼굴 모션에 대한 정확한 추정 방법과 좀 더 실시간으로 처리할 수 있는 방법에 대한 많은 연구가 진행 중이다.

이러한 3 차원 얼굴 포즈를 추정하기 위해 많은 방법들이 제안되고 있다. 이러한 방법으로는 크게 외형 기반 방법과 모델을 이용한 방법으로 나눌 수 있다. 먼저 외형 기반 방법으로는 연속된 이미지에서 얼굴의 특징점을 검출하여 검출된 특징점들간의 차이를 이용한 방법으로서 특정점을 정확히 검출하게 되면 좋은 결과를 나타낸다[1][2]. 하지만 특징점을 정확히 검출하지 못하는 경우에는 좋지 않은 결과를 나타내므로 특징점을 정확히 검출하는가에 따라 결과가 정해지게 된다. 모델을 이용하는 방법으로는 기하학적 모델을 이용하여 전체 얼굴 영역을 트래킹하는 방법으로서 특징점

의 차이를 이용한 방법보다 좀 더 좋은 결과를 얻을 수 있다. 얼굴 모션을 추정하기 위해 사용되는 모델로는 얼굴과 거의 유사한 모양의 3 차원 모델과 평면이나 실린더 또는 타원체와 같은 좀 더 단순한 모델이 있다. 얼굴과 거의 유사한 모델을 사용하는 경우에는 얼굴 영역과 모델간의 정확한 초기화가 이루어져야지만 얼굴 포즈 추정에 대한 오차가 적고 좋은 결과를 얻을 수 있다[3]. 하지만 정확한 초기화가 이루어지지 않는다면 얼굴 포즈 추정에 대한 오차가 점차 증가하게 되고 결과가 좋지 않다.

얼굴과 유사한 3 차원 얼굴 모델보다 단순한 모델을 사용하는 경우 종종 효과적이거나 초기화 때 생긴 오차를 줄일 수 있고 얼굴 포즈 추정에 대한 좋은 결과를 가질 수 있다. 먼저 평면을 이용한 얼굴의 모션을 추정하기 위해 2 차원 평면 또는 얼굴 텍스쳐를 사용한다[4]. 이러한 평면을 이용한 방법은 초기화 에러에 대해 별 영향을 받지 않고 얼굴의 포즈를 추정할 수 있다. 하지만 얼굴의 방향이 정면에서 많이 벗어나지 않을 때 좋은 결과를 얻을 수 있다. 다시 말해, 얼굴의 가려지는 부분이 생기면 좋은 결과를 기대할 수 없다. 따라서 이러한 평면 모델을 사용하는 방법의 단점을 해결하기 위해 3 차원 모델을 사용하는 방법이 있다. 단순한 3 차원 모델을 이용하는 방법으로서 실린더나 타원체를 이용한 방법은 좀 더 좋을 결과를 나타낸다. 또한 텍스쳐 매핑이 된 실린더를 이용하여 좀 더 빠르게 얼굴의 포즈를 추정하는 방법이 있다[5].

실질적으로 얼굴 모델과 얼굴 영상을 정확히 일치시키는 초기화가 거의 불가능하다. 따라서 본 논문에서는 실시간으로 얼굴 포즈를 추정하기 위해 작은 초기화 에러에 대해 영향을 받지 않는 3 차원 실린더 모델을 이용한 포즈 추정 방법을 제안한다. 초기 프레임으로부터 얼굴의 피부색과 템플릿 매칭을 이용하여 얼굴 영역을 검출하게 된다. 검출된 얼굴 영역에 3 차원 실린더 모델을 투영하게 된다. 이렇게 3 차원 실린더 모델을 투영한 후 연속된 프레임으로부터 Lucas-Kanade 의 Optical flow 를 이용하여 얼굴 영역의 밝기값 변화량을 측정하게 되고 측정된 변화량으로부터 얼굴 모션을 추정한다. Optical flow 는 밝기 값의 변화에 의해 측정되는 것이므로 주위 환경의 밝기가 급격하게 변하는 경우 좋지 않은 결과를 나타낸다. 따라서 좀 더 정확한 얼굴 모션을 추정하기 위해 최대-최소 정규화 방법을 이용하여 조명의 영향을 줄이고 IRLS(Iteratively re-weighted least squares) 방법을 이용하여 각 픽셀에 대한 가중치를 결정하게 되고 동적 템플릿을 이용해 오랫동안 정확한 얼굴 모션을 추정하는 방법에 대해 제안한다. 본론에서는 위와 같은 방법을 이용하여 얼굴 모션 추정 방법에 대해 설명한다. 마지막으로 실험 결과와 향후 연구 방안에 대해 논의하도록 한다.

2. 본 론

본 연구에서 제안하는 시스템은 크게 세가지 단계로 나눌 수 있다. 첫째, 초기 프레임에서 얼굴을 검출하는 단계로써 초기에 입력된 영상으로부터 얼굴의 고유한 피부색을 이용하여 얼굴의 후보영역을 결정하게 되고 결정된 후보영역에서 템플릿 매칭을 이용하여 얼굴을 검출하는 단계이다. 둘째, 검출된 얼굴로부터 얼굴의 포즈를 추정하는 단계로서 검출된 얼굴 영역에 3 차원 실린더 모델을 투영하게 되고 연속된 프레임 사이의 Optical flow 를 이용하여 밝기 값의 변화량을 측정하게 된다. 또한 IRLS 와 동적 템플릿을 이용하여 좀 더 정확한 얼굴의 포즈 추정을 한다. 마지막으로 추정된 얼굴 포즈 파라메터를 이용하여 3 차원 얼굴 모델에 적용하는 단계로서 실질적인 3 차원 얼굴 모델에 얼굴 포즈 파라메터를 적용하는 단계이다.

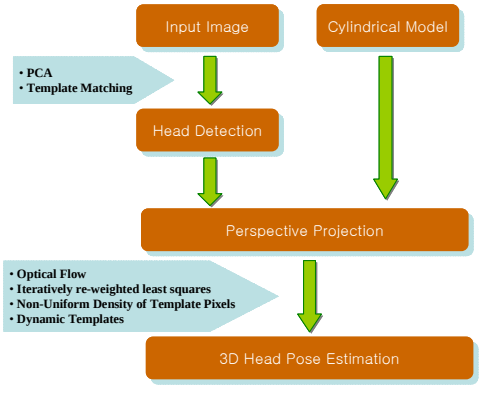


그림 1. 시스템 개략도

2-1. 얼굴 검출

얼굴 포즈를 추정하기 전 초기 프레임에서 얼굴을 검출한다. 본 연구에서는 얼굴을 검출하기 위해 얼굴의 피부색과 템플릿 매칭을 이용한다. 만약 얼굴의 피부색만으로 얼굴을 검출한다면 많은 문제점을 가지고 있다. 예를 들면, 입력 영상의 배경에 얼굴의 피부색과 유사한 색이 있다면 이 부분 또한 얼굴로 검출하게 된다. 따라서 피부색을 이용하여 얼굴의 후보 영역을 정하게 되고 후보 영역 안에서 얼굴을 검출하는 과정을 거치게 된다.

얼굴 검출을 위해 먼저 얼굴의 후보영역을 결정한다. 얼굴의 후보 영역을 결정하기 위해 얼굴의 피부색을 이용하게 된다. 피부색에 대한 색상 모델은 여러 가지가 있다. 여러 가지 종류의 색상 모델 중 적합한 색상 모델을 선택하기 위해서는 입력 영상의 특징을 잘 표현할 수 있어야 하며 분포가 일정해야만 한다. 그림 2 는 얼굴 검출에 많이 이용되는 색상 모델에 대한 색상 분포도이다.

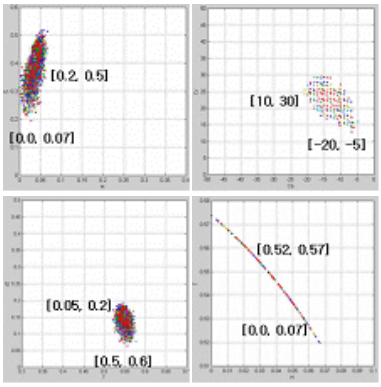


그림 2. 색상 분포도 (위 왼쪽부터 H-S, Cb-Cr, T-S, H-T)

그림 2 에서 볼 수 있듯이 HT 조합에 의한 모델에서 피부색 분포가 직선 형태로 밀집되어 있음을 확인할 수 있다. 또한, H 값은 임의의 영역에 대한 붉은색, 초록색, 노란색, 자줏빛과 같은 현저한 색상을 정의하는데 이용되며, 백색 조명과 같은 빛의 영향에 영향을 받지 않는 장점을 갖고 있다.

본 연구에서는 HT 색상 모델의 피부색 분포 형태가 직선 형태라는 특징을 이용한다[6]. HT 값에 의한 분포도를 각각의 최대값과 최소값을 지나는 직선의 방정식을 이용하여 피부색을 표현하고 직선과 입력 영상 사이의 거리를 이용해서 얼굴을 검출하게 된다(그림 3).

$$Distance(i,j) = \frac{\left| aH_{i,j} + bT_{i,j} + c \right|}{\sqrt{a^2 + b^2}} < \lambda$$

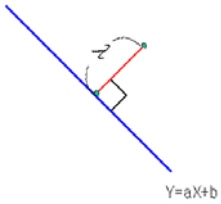


그림 3. 피부색 분포의 직선 형태 모델화

피부 영역 검출을 위한 거리 비교에서 임계값 (λ)은 0.003 을 이용한다. 이와 같은 방법으로 얼굴 후보 영역을 검출하게 되면 잡음이 발생하게 되는데 미디언 필터와 모폴로지 연산을 이용하여 발생된 잡음을 제거하게 된다. 잡음을 제거한 후 히스토그램을 이용하여 입력 영상에서 얼굴 후보 영역을 결정하게 된다.

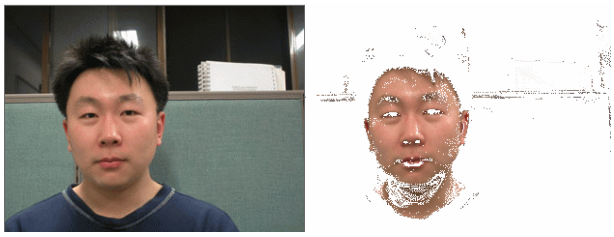


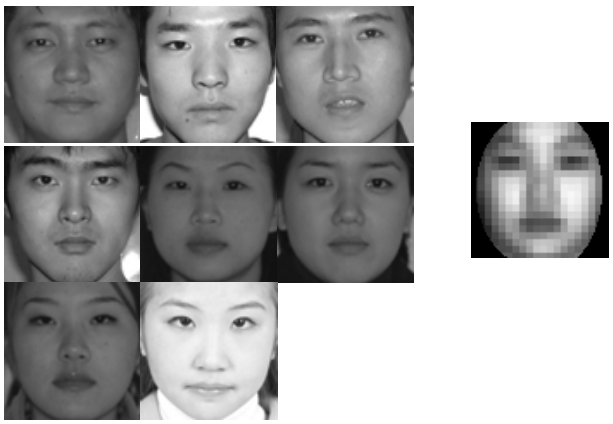
그림 4. 피부 영역 검출 결과

위와 같은 방법으로 얼굴의 후보영역을 검출한 후 얼굴 후보 영역 안에서 PCA 변환을 통해 템플릿과 비교할 영역을 고유 벡터로 구성하며 사전에 만들어진 템플릿과 비교 영역의 고유 벡터들 간의 템플릿 매칭을 이용함으로써 얼굴을 검출하게 된다. 먼저 얼굴 템플릿을 생성하기 위해서 훈련 영상을 80×80 의 크기로 만든 후 전처리 과정을 통한 후 평균 영상으로 만들어진다. 전처리 과정은 최대-최소 정규화 방법과 히스토그램 평활화를 이용한다. 먼저 최대-최소 정규화 방법은 조명 보정을 하기 위한 것으로 입력 영상을 새롭게 정의

된 데이터의 범위로 변형시켜주는 방법이다.

$$y = \left(\frac{y - \min_1}{\max_1 - \min_1} \right) (\max_2 - \min_2) + \min_2$$

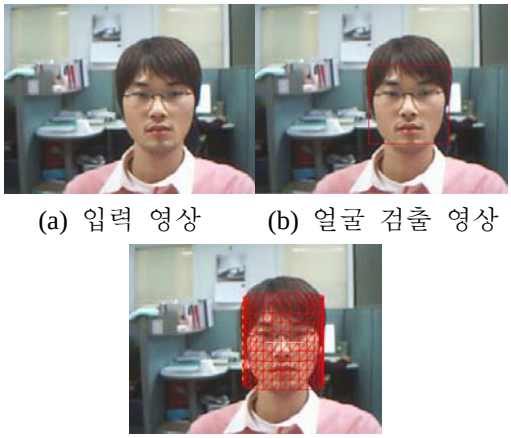
$\min_1$ 과 $\max_1$ 은 입력 영상의 최소, 최대 밝기 값이고 $\min_2$ 와 $\max_2$ 는 변형하고자 하는 데이터 범위의 최소, 최대 밝기 값이다. 최대-최소 정규화 방법을 통해 여러 영상의 밝기 값의 차이를 줄일 수 있는 효과를 가지고 있다. 히스토그램 평활화는 어두운 부분은 더욱 어둡고 밝은 부분은 더욱 밝게 만들어 주기 때문에, 얼굴의 명암차이에 의한 변화를 줄여주고 얼굴 요소간의 특징을 부각시킬 수 있다. 또한 모자익 영상을 생성하여 얼굴의 작은 기울어짐, 영상의 잡음 등에 대한 보정을 해줄 수 있기 때문에 4*4 모자익 영상을 생성한다. 그림 5 는 훈련 영상들과 훈련 영상들에 의해 만들어진 템플릿을 보여준다.



(a) 훈련 영상 (b) 템플릿 영상

그림 5. 템플릿 생성

이와 같이 템플릿을 생성한 후 템플릿 영상을 PCA 변환을 하게 되고 변환을 통해 템플릿의 고유벡터를 생성하게 된다. 검출된 얼굴 후보 영역 안에서 템플릿의 크기인 80×80 영역을 좌측상단에서부터 차례대로 선택하게 되며 선택된 영역을 최대-최소 정규화와 히스토그램 평활화 과정을 거친 후 PCA 변환을 하고 템플릿의 고유벡터와 선택된 영역의 고유벡터간의 유클리디언 거리를 구하여 가장 작은 영역을 얼굴로 검출하게 된다.



(a) 입력 영상 (b) 얼굴 검출 영상

(c) 실린더 모델 투영

그림 6. 얼굴 검출

2-2. 얼굴 포즈 추정

본 연구에서는 3 차원 실린더 모델을 이용하여 얼굴의 포즈를 추정한다. 얼굴 검출 과정을 통해 영상으로부터 얼굴을 검출하게 되면 검출된 영역에 3 차원 실린더 모델을 원근 투영한다. 3 차원 실린더 모델을 이용하여 포즈를 추정하기 위해 Lucas-Kanade 의 Optical flow 의 이론을 이용한다. 먼저 영상에서 픽셀 $\boldsymbol{u} = (u, v)$ 이라고 하면 시간 t 에서 영상은 $I(u,t)$ 로 표현한다. 시간 $t+1$ 에서 픽셀 $\boldsymbol{u}$ 는 $\boldsymbol{u}' = F(\boldsymbol{u}, \mu)$ 으로 이동한다. 여기서 μ 는 모션 파라메터 벡터이고 $F(\boldsymbol{u}, \mu)$ 는 파라메트릭 모션 모델이다. 만약 조명이 변하지 않는다고 가정하면 좌표는 이동하지만($\boldsymbol{u} \rightarrow \boldsymbol{u}'$), 영상의 밝기 값은 변하지 않기 때문에 다음과 같이 나타낼 수 있다.

$$I(F(\boldsymbol{u}, \mu), t+1) = I(\boldsymbol{u}, t)$$

모션 벡터 μ 를 계산하기 위한 방법 중 하나는 다음 식의 함수 값을 최소화함으로써 얻을 수 있다.

$$\min \; E(\mu) = \sum_{\boldsymbol{u} \in \Omega} \left(I(F(\boldsymbol{u}, \mu), t+1) - I(\boldsymbol{u}, t) \right)^2$$

여기서 Ω 는 시간 t 에서 측정하고자 하는 영역이고 모션 벡터를 얻기 위해서는 영역 Ω 안에 존재하는 픽셀만을 가지고 계산한다. 일반적으로 Optical flow 알고리즘 중 로컬 영역에 대한 Optical flow 를 계산하는 방법으로 Lucas-Kanade 의 방법이 있다. 또한 이 알고리즘은 정확하고 잡음에 강

인하다는 장점을 가지고 있다. 본 연구에서 이 알고리즘을 적용한다.

$$\mu = -\left(\sum_{\Omega}\left(I_u F_\mu\right)^T\left(I_u F_\mu\right)\right)^{-1}\sum_{\Omega}\left(I_t\left(I_u F_\mu\right)^T\right)$$

여기에서 I_u 와 I_t 는 각각 공간적 이미지 변화량과 시간적 이미지 변화량, F_μ 는 함수 F 에서 μ 에 대해 편미분을 한 것이다.

3 차원 실린더 모델을 원근 투영하고 투영된 모델을 이용하여 얼굴의 포즈를 추정하는 방법은 다음과 같다. 먼저 시간 t 에서 실린더 모델의 좌표 $X=[x,y,z,1]^T$ 는 $t+1$ 에서의 변환은 회전과 이동 변환을 적용한다.

$$X(t+1) = M\cdot X(t) = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}\cdot X(t)$$

여기서 $[\omega_x,\omega_y,\omega_z]$ 는 회전 행렬을 나타내고 $[t_x,t_y,t_z]$ 는 이동 행렬을 나타낸다.

시간 $t+1$ 에서 원근 투영에 의해 실린더 모델의 정점 X 가 이미지에 투영되는 픽셀 $\boldsymbol{u}$ 는 다음과 같다. 모션 파라메터 벡터 $\mu=[\omega_x\ \omega_y\ \omega_z\ t_x\ t_y\ t_z]$ 을 이용한 파라메트릭 모션 모델 $F(\boldsymbol{u},\mu)$ 이다.

$$\boldsymbol{u}(t+1) = \begin{bmatrix} x - y\omega_z + z\omega_y + t_x \\ x\omega_z + y - z\omega_x + t_y \end{bmatrix}\cdot\frac{f_L}{-x\omega_y + y\omega_x + z + t_z}$$

여기서 f_L 은 초점 거리를 나타낸다. 시간 t 에서 $\mu=0$ 일 때 F_μ 는 다음과 같다.

$$F_\mu|_{\mu=0} = \begin{bmatrix} -xy & x^2+z^2 & -yz & z & 0 & -x \\ -(y^2+z^2) & xy & xz & 0 & z & -y \end{bmatrix}\cdot\frac{f_L}{z^2}(t)$$

이와 같은 방법으로 얼굴의 포즈를 추정할 수 있지만 좀 더 정확한 포즈를 추정하기 위해 다음과 같은 방법을 이용한다. 첫째 모든 입력영상에 대한 조명 보정을 하기 위해 최대-최소 정규화 알고리즘을 적용한다. 이것은 Optical flow 가 영상의 밝기값을 이용하기 때문에 조명의 영향을 줄이는 것이 중요하기 때문이다. 둘째, IRLS(Iteratively re-weighted least squares) 알고리즘과 에지를 포함하는 픽셀에 대한 가중치 보정, 템플릿 픽셀에서 밀도가 동일하지 않은 것에 대한 보정방법을 사용한다. 다음 식은 가중치를 포함한 얼굴 포즈 추정을 하기 위한 식이다.

$$\mu = -\left(\sum_{\Omega}\left(w\left(I_u F_\mu\right)^T\left(I_u F_\mu\right)\right)\right)^{-1}\sum_{\Omega}\left(w\left(I_t\left(I_u F_\mu\right)^T\right)\right)$$

첫째, IRLS 는 잡음이나 가려짐과 같은 부분을 보정하기 위해 사용하는 값이다. IRLS 에 대한 가중치 w_I 를 구하는 방법 다음과 같다.

$$w_I = c_I\exp\left(-\left(I(u,t+1)-\hat{I}(u,t)\right)^2/2\sigma_I^2\right)$$

$$\sigma_I = 1.4826\cdot\underset{u\in\Omega}{median}\left|I(u,t+1)-\hat{I}(u,t)\right|$$

여기서 $\hat{I}$ 은 시간 t 에서 변화된 템플릿을 나타낸다.

둘째, 영상에서 에지는 유용한 정보가 될 수 있다. IRLS 에 대한 가중치를 사용하게 되면 이러한 에지에 대한 내용을 포함하지 못한다. 따라서 에지 정보에 대한 가중치 w_G 을 구하는 방법은 다음과 같다.

$$w_G = c_G\left(1-\exp\left(-\left(I_u(u,t+1)\right)^2/2\sigma_G^2\right)\right)$$

σ_G 는 128 값을 가진다.

셋째, 그림 7 과 같이 템플릿 픽셀에서 밀도가 동일하지 않은 것을 보정하기 위해 다음과 같은 식을 사용한다.

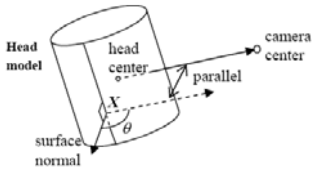


그림 7. 픽셀 밀도에 대한 가중치 w_D 에 사용되는 법선 벡터와 얼굴 중심에서 카메라 중심으로의 방향을 가지는 방향 벡터 사이의 각 θ

$$w_D = c_D\left(1-\min\left(\left|\theta(u)\right|,\pi/2\right)\cdot 2/\pi\right)^2$$

이와 같이 계산된 가중치는 포즈를 추정하는데 이용된다.

$$w = \left(w_I + w_G\right)\cdot w_D$$

하나의 템플릿을 사용하는 것은 오랜 시간 동안 정확한 포즈 추정을 할 수 없다. 왜냐하면, 주변 빛의 변화나 가려지는 현상이 일어나는 경우 하나

의 템플릿만으로 적용할 수 없다. 따라서 동적 템플릿을 이용한다. 각 프레임에서 얼굴의 포즈가 발견되면 얼굴 영역은 다음 프레임을 위해 템플릿을 사용된다. 가려지는 현상이 발생되었을 때 템플릿 영역에는 외곽선이 생기게 된다. 이러한 외곽선은 템플릿 영역에서 제거되어야만 한다. 제거 방법은 현재 템플릿과 이전 마지막 템플릿을 이용하여 제거한다.

$$\left|I(u,t)-\hat{I}(u,t)\right|>c\sigma_I$$

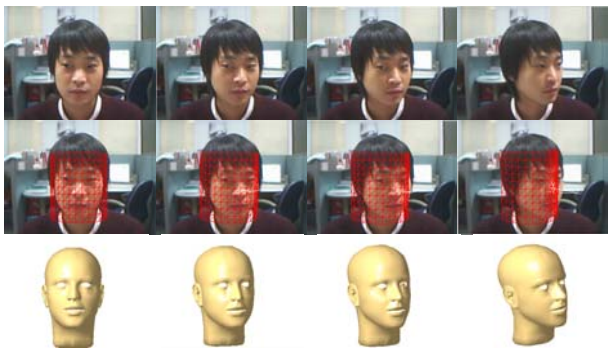
3. 결 론

본 연구에서는 3D 실린더 모델과 Optical flow 를 이용하여 얼굴 포즈를 추정하는 방법을 제시하였다. IRLS 알고리즘과 동적 템플릿을 이용하여 좀 더 정확한 포즈 추정을 할 수 있었다.

실험은 웹 캠을 이용하여 실시간으로 영상을 입력 받고 제안된 시스템을 통해 포즈를 추정한다. 얼굴 포즈 추정된 결과는 실제 3D 얼굴 모델에 적용하였다. 다시 말해, 입력 영상의 얼굴 포즈를 3D 얼굴 모델이 따라 하는 것이다.

그림 8. 실험에 사용된 3 차원 얼굴 모델

실험 결과는 다음과 같다.



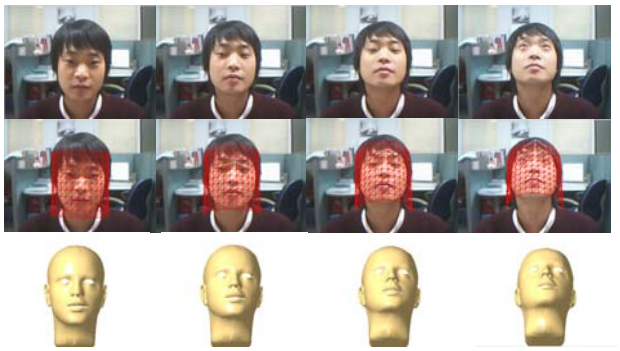


그림 9. 실험 결과

향후 연구 과제로는 3D 얼굴 모델에 좀 더 사실감을 주기 위해 텍스쳐 매핑이 이루어져야 한다. 또한 정면 얼굴이 아닐 경우 얼굴의 표정을 제어하기 위해 얼굴 포즈 추정과 연계할 수 있는 방안을 마련해야겠다.

참 고 문 헌

[1] Z.Liu and Z.Ahang, “Robust Head Motion Computation by Taking Advantage of Physical Properties”, HUMO2000, 2000

[2] T. Jebara and A.Pentland, “Parameterized Structure from Motion for 3D Adaptive Feedback Tracking of Faces”, CVPR97, 1997

[3] D. DeCarlo and D. Metaxas, “The Integration of Optical Flow and Deformable Models with Applications to Human Face Shape and Motion Estimation”, CVPR96, pp.231-238, 1996

[4] G.D. Hager and P.N. Belhumeur, “Efficient region tracking with parametric models of geometry and illumination”, PAMI, vol. 20, pp.1025`1039, 1998

[5] M.L. Cascia and S. Sclaroff, “Fast, Reliable Head Tracking under Varying Illumination”, CVPR99, pp.604-610, 1999

[6] Kyongpil Min, Junchul Chun and Goorack Park, ”A Nonparametric Skin Color Model for Face Detection from Color Images”, PDCAT 2004, pp.116-120, 2004