

Optical flow를 이용한 얼굴요소 및 얼굴의 움직임 추정값에 따른 3차원 얼굴모델의 움직임 합성

박도영, 변혜란
연세대학교 컴퓨터과학과

Motions syntheses in 3D facial model using features and motion parameters estimated through optical flow

Doyoung Park, Hyeran Byun
Dept of Computer Science, Yonsei University

요 약

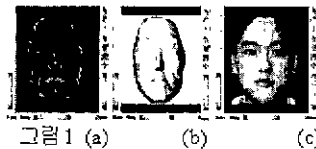
동영상에서 얼굴의 움직임을 이해하는 것은 인간과 컴퓨터간의 상호작용을 이루는 분야에서 중요한 문제이다. 본 논문에서는 2차원 동영상에서 얼굴요소 및 얼굴의 움직임을 추정하기 위해 optical flow를 통해 매개변수화된 움직임 벡터를 추출한다. 그리고나서, 이를 소수의 매개변수들의 조합으로 만들어 얼굴의 움직임에 대한 정보는 묘사할 수 있게 하였다. 매개변수화된 움직임 벡터는 얼굴 및 얼굴 요소의 특징에 따라 다른 벡터 모델을 사용한다. 2차원 동영상에서 매개변수화된 움직임 벡터는 매 프레임마다 갱신되어 각 프레임에서 얼굴 및 얼굴 요소의 위치를 파악한다. 또한, 갱신된 벡터의 매개변수 조합으로 만들어 확인된 움직임에 대한 정보가 3차원 얼굴모델에 전달되며 3차원 얼굴 모델의 단위행위(Action Unit)와 연결 되어 2차원 동영상에서의 얼굴 움직임을 합성할 수 있게 하였다.

1. 서론

최근, 얼굴 및 얼굴 표정의 추적 및 이해는 컴퓨터 시각과 인간과 컴퓨터간의 상호작용을 이루는 분야에서 그 관심이 증가되고 있으며 활발히 연구가 진행중이다. 이러한 문제와 관련된 초창기의 연구는 Optical flow에 기반한 Azarbayejani와 Pentland[1]에 의한 것과 기법 뎀폴릿에 기반한 Darral et al.[2]등이 있다.

본 연구에서는 매개변수화된 움직임 모델(벡터)를 이용하여 위 두 가지 방식의 Hybrid방식을 사용한다. 이러한 매개변수화된 움직임 모델은 소수의 매개변수 조합으로 얼굴 및 얼굴요소의 강직한 움직임 뿐만 아니라 비강직성을 가지는 움직임도 묘사해 낼 수가 있다. 본 연구에 쓰인 매개변수화된 움직임 벡터는 얼굴 및 얼굴 요소의 특징에 따라 다른 모델이 사용되었다. 동영상의 매 프레임이 진행되면서 각 모델의 파라미터값이 Optical flow를 이용하여 측정되고 이 값에 따라 각 얼굴 요소 및 얼굴영역을 추적하며 움직임 정보가 결정된다. 그러나, 본 연구에서는 첫 프레임에서 각 영역을 위치시키는 것[3,4]은 관심을 두지 않는다. 본 연구는 2차원 동영상에서 얼굴요소 및 얼굴 추적뿐만 아니라 3차원으로 그 개념을 확장한다. 위에서 측정된 파라미터 값이 3차원 얼굴모델에 전달되어 얼굴모델의 단위행위(Action Unit)과 연결되어 2차원 동영상의 움직임을 3차원에서 합성하는 것이다. 이를 위해 사용된 얼굴모델은 그림1과 같다. 3차원 얼굴

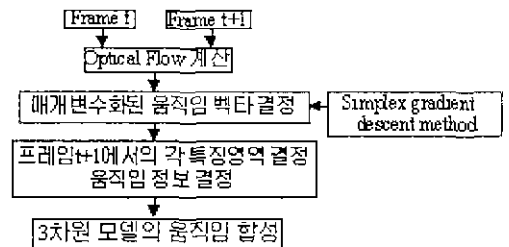
모델의 단위행위들은 P.Ekman과 W.V.Friesen이 제안한 FACS(Facial Action Coding System)[5]에 근거한 46개의 기본동작으로 구성되어 있다. 이후, 다음 2장에서는 시스템구성을 소개하고 3장은 실험결과, 4장에서는 결론을 내린다.



(a)특정인의 얼굴에 적용된 3차원 얼굴 mesh 모델
(b)텍스처가 인더리지 많은 영역 모델
(c)특정인의 텍스처가 입력된 3차원 얼굴 모델

2. 시스템 구성

2.1. 전체 시스템 개요



2.2. 매개변수화된 움직임 모델(벡터)

얼굴요소 및 얼굴의 움직임에 사용된 매개변수화된 움직임 모델은 다음과 같다.

얼굴(1)[6], 눈(2)[7], 눈썹미, 입(3)[8]

$$u(x,y)=a_0+a_1x+a_2y+pa_x^2+pa_y^2, v(x,y)=a_3+a_4x+a_5y+pa_x^2+pa_y^2 \quad (1)$$

$$u(x,y)=a_0+a_1x+a_2y, v(x,y)=a_3+a_4x+a_5y \quad (2)$$

$$u(x,y)=a_0+a_1x+a_2y, v(x,y)=a_3+a_4x+a_5y+ca_x^2 \quad (3)$$

여기서 $u(x)=[u(x,v),v(x,y)]^T$ 는 2차원 영상의 한 점 $x=(x,y)$ 에서 수평방향 및 수직방향의 움직임의 요소이다

이러한 파라미터 a_i 들은 다음과 같은 움직임에 대한 정보를 가지고 있다[9].

$$\text{수평방향으로의 움직임} = a_0, \quad (4)$$

$$\text{수직방향으로의 움직임} = a_3, \quad (5)$$

$$\text{발산성(Divergence)} = a_1 + a_5 = (u_x + v_y), \quad (6)$$

$$\text{회전성(Curl)} = -a_2 + a_4 = -(u_y - v_x), \quad (7)$$

$$\text{변형성(Deformation)} = a_1 - a_5 = (u_x - v_y) \quad (8)$$

이를 도시한 것이 아래의 그림2이다.

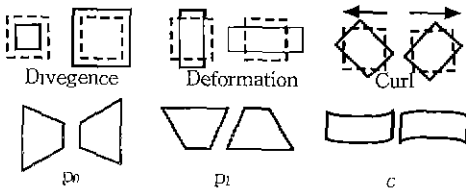


그림 2

그림2에서 직선은 변형된 영역을 의미하며, 각 정보들의 왼쪽편은 값이 줄어들 경우이며 오른쪽편은 커질 경우이다.

2.3. Optical flow 계산[10]

선통적으로 Optical flow를 계산하기 위해서는 세가지의 접근방식-영상 밝기의 미분값에 기반한(gradient-based) 접근방식, 연속된 두 영상의 국부적인 영상 밝기 분포가 유지된다는 가정을 이용한(correlation-based) 접근방식, 시간상에서 뿐 아니라 공간상에서의 주파수성질을 이용한(spatiotemporal energy-based) 접근방식-이 사용되고 있다.

본 논문에서의 접근방식은 이 세가지의 접근방식을 종합한 성격에 갖고 있다 이것은 두 단계의 과정을 거쳐 Optical flow를 계산하게 된다

이에 대한 절차(framework)는 다음과 같다.

1 각 영상에 대해 band-pass 필터링을 수행한다.

ii 현재 영상에서 각 점을 중심으로 $(2n-1) \times (2n+1)$ 의 윈도우를 형성한다 또한 다음 영상에서 일치하는 점을 중심으로 $(2N+1) \times (2N+1)$ 의 탐색영역을 설정한다. (9)식을 사용해 탐색영역상에서 여러분포를 계산하고 이것을 (10)식을 이용해 반응분포로 변환한다.

$$E_i = \sum_{x=-n}^n \sum_{y=-n}^n (I_1(x-t, y-v) - I_2(x+u-t, y+v-j))^2, -N \leq u, v \leq +N \quad (9)$$

$$R_i(u,v) = e^{-k \cdot E_i} \quad (10)$$

iii (11),(12)식을 이용해 1단계 Optical flow $U_{cc}=(u_{cc}, v_{cc})$ 의 분

산행렬 S_{cc} 를 계산한다.

$$u_{cc} = \frac{\sum_u \sum_v R_i(u,v)u}{\sum_u \sum_v R_i(u,v)}, v_{cc} = \frac{\sum_u \sum_v R_i(u,v)v}{\sum_u \sum_v R_i(u,v)} \quad (11)$$

$$S_{cc} = \begin{pmatrix} \frac{\sum_u \sum_v R_i(u,v)(u-u_{cc})^2}{\sum_u \sum_v R_i(u,v)} & \frac{\sum_u \sum_v R_i(u,v)(u-u_{cc})(v-v_{cc})}{\sum_u \sum_v R_i(u,v)} \\ \frac{\sum_u \sum_v R_i(u,v)(u-u_{cc})(v-v_{cc})}{\sum_u \sum_v R_i(u,v)} & \frac{\sum_u \sum_v R_i(u,v)(v-v_{cc})^2}{\sum_u \sum_v R_i(u,v)} \end{pmatrix} \quad (12)$$

iv. 1차적으로 계산된 Optical flow상에서 갱신시킬 점을 중심으로 $(2w+1) \times (2w+1)$ 의 가우시안 마스크를 씌어 값을 변화시킨후 (13),(14)식을 이용하여 평균 U_n 와 분산 행렬 S_n 을 계산한다

$$u_n = \frac{\sum_u \sum_v R_n(u,v)u}{\sum_u \sum_v R_n(u,v)}, v_n = \frac{\sum_u \sum_v R_n(u,v)v}{\sum_u \sum_v R_n(u,v)} \quad (13)$$

$$S_n = \begin{pmatrix} \frac{\sum_u \sum_v R_n(u,v)(u-u_n)^2}{\sum_u \sum_v R_n(u,v)} & \frac{\sum_u \sum_v R_n(u,v)(u-u_n)(v-v_n)}{\sum_u \sum_v R_n(u,v)} \\ \frac{\sum_u \sum_v R_n(u,v)(u-u_n)(v-v_n)}{\sum_u \sum_v R_n(u,v)} & \frac{\sum_u \sum_v R_n(u,v)(v-v_n)^2}{\sum_u \sum_v R_n(u,v)} \end{pmatrix} \quad (14)$$

v (15)식을 사용하여 갱신시킬 점의 속도를 갱신한다.

$$U^{t+1} = (S_n^{-1} S_{t+1}^{-1})^{-1} (S_n^{-1} U_n + S_{t+1}^{-1} U_{t+1}), U^{t+1} = U_{cc} \quad (15)$$

vi. 두 프레임간의 변화가 충분히 작을때까지 iv, v 과정을 반복한다

2.4. 최적의 움직임 벡터 값 설정

매개변수화된 움직임 벡터의 표기상 편리함을 위해 각각의 움직임 모델을 다음과 같이 정의한다

$$X(x) = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & x^2 & 0 \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 & y^2 \end{bmatrix} \quad (16)$$

$$A = [a_0 \ a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ 0 \ 0 \ 0]^T \quad (17)$$

$$P = [a_0 \ a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ p_1 \ p_2 \ 0]^T \quad (18)$$

$$C = [a_0 \ a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ 0 \ 0 \ c]^T \quad (19)$$

그러면, $X(x)A$ 는 눈 영역을 나타내는 움직임 벡터가 된다 좀더 자세히 하자면 $X(x)A$ 은 왼쪽 눈영역, $X(x)A_r$ 은 오른쪽 눈영역 움직임 벡터이다 마찬가지로 $X(x)C_m$, $X(x)C_l$, $X(x)C_r$ 들은 각각, 입, 왼쪽 눈썹미, 오른쪽 눈썹미의 움직임 벡터이다

R 을 얼굴요소 혹은 얼굴 영역이라고 하자 그러면, C_{mn} 은 입영역이 된다. 현재 프레임에서의 한 특징영역에 대한 초기의 움직임 벡터를 c_i 라고 할 때 목표는 실제의 Optical flow를 가장 잘 고려하는 후보 움직임 벡터 c_i 를 찾는 것이다 본 연구에서 사용한 측정법(식20)은 Optical flow와 움직임 벡터에 의한 flow사이의 최소화승기법(mean squared error)이다. 위 2.2절에서 계산된 영상에서의 Optical flow를 벡터 $[u' \ v']^T$ 로 움직임 벡터에 의한 flow를 $[u \ v]^T$ 로 정의하면,

$$e_k = (u-u')^2 + (v-v')^2, E(c_n, c_i) = \frac{1}{n} \sum_{j=1}^n e_{kj}, V(x,y) \in R \quad (20)$$

n 은 영역에 속하는 점들의 개수이다

(20)식을 이용하여 실제의 Optical flow를 가장 잘 고려하는 것은

(21)식을 만족하는 최적의 매개변수 벡터 c_i^* 이다

$$c_i^* = \arg(\min_j E(c_n, c_j)) \quad (21)$$

E 값을 최소로 하면서 수개의 파라미터값을 정확히 구해내는 것은 불가능한 일이므로 국부적 최소치를 찾고자 한다 이러한 최소치는 c_i 시작점, E 를 에리함수로 하여 simplex gradient descent[11]방법을 이용하였다.

2.5. 얼굴요소 및 얼굴 영역 예측, 변환

두 프레임 사이에서 측정된 얼굴요소 및 얼굴에 대한 매개변수화된 움직임 벡터값은 움직임 정보를 암시할 뿐 아니라 다음 프레임에서 특정 영역의 위치를 예측하게 된다. 눈영역이나 얼굴영역은 네 개의 꼭지점에 의해 표현될 수 있는 간단한 사변형이다. 따라서 네 꼭지점(x)에 $X(x)P+x$, $X(x)AI+x$, $X(x)Ar-x$ 을 적용하여 영역을 변화시킨다 그러나 입이나 눈썹미 부분은 곡률변수(식(3)의 'n')로 인해 이미지 마스크를 사용하여 변화시켰다

2.6. 움직임 정보에 의한 3차원 얼굴 모델의 변환

2.1절에서는 움직임 벡터를 구성하는 몇 개의 변수의 조합으로 나타낼 수 있는 움직임 정보를 제시하였다. 이러한 움직임 정보가 3차원 얼굴모델에 전달 되어 각 정보와 관련된 단위행위들이 작동된다 또한, 단위행위들의 움직임량 역시 이 정보로부터 유추되어 진다.

이들간의 상호 관계가 다음과 같다

입영역에서 주된 움직임 정보는 곡률정보와 변형성 정보이며 이것과 그 움직임임에 따라 단위행위 14, 15, 18, 20, 24, 25를 조절하였다. 눈썹미 영역에서의 주된 움직임 정보는 a_3 이며 이것에 따라 단위행위 1, 2, 4등이 조절된다. 눈영역에서는 매개변수 a_3 와 변형성정보이며 이에 따라 단위행위 5, 41등이 조절된다.

3. 실험결과

동영상에서 Optical flow정보에 따른 얼굴요소의 2차원적 추적과 3차원 얼굴모델에 전달된 움직임 정보로서 움직임을 보인 3차원 얼굴 모델의 합성결과외의 모습이 그림 3에 나타나 있다. 50여개의 프레임들 시키면서 매 6번째 프레임마다 추적된 모습이 맨 위쪽 부분은 동영상의 원래 영상을 보여주고 있으며, 가운데 부분에서는 해당 영상에서 측정된 얼굴요소영역과 함께 간접적인 움직임량을 짐작하게 한다 맨 아래부분은 특정인의 얼굴에 결합된 3차원 얼굴 모델이 전달된 움직임 정보에 따라 해당 단위행위(Action Unit)들을 적용한 후의 모습을 보여주고 있다 다음 프레임의 특징 영역 추출은 현재 프레임의 특징영역과 비교하여 이루어진다 본 연구에서는 프레임이 진행되면서 추출된 특징영역이 예상하는 것보다 조금씩 넓어지게 되어 필요없는 부분까지 포함함으로써 아주 세밀한 영역 추출과 움직임을 찾는 데 어려움이 따른다. 따라서, 3차원 모델이 전체 형상의 움직임을 합성하나 세부적인 움직임을 합성하는 데는 정확성이 떨어진다. 이는 좀더 정확한 Optical flow를 계산함으로써 개선될 수 있을 것이다

4. 결론 및 향후 연구

본 논문에서는 산출된 Optical flow값에 따라 각 얼굴요소 및 얼굴영역의 움직임을 가장 잘 묘사하는 매개변수화된 움직임 벡터값을 갖는다 이 벡터값에 따라 얼굴요소 및 얼굴영역의 위치를 예측할 수 있을 뿐 아니라 움직임을 해석할 수 있다. 따라서 이러한 정보들로 3차원 얼굴모델이 2차원 영상에서의 움직임을 그 단위행위의 조합으로 합성할 수 있게 한다. 이러한 정보에 대한 보다 깊은 이해는 얼굴 움직임 추적뿐 아니라 얼굴표정인식 시스템 구축도 가능하게 할 것이다 따라서 앞으로 이러한 정보를 깊이 분석하여 얼굴 움직임을 인식

할 수 있게 한다면 인간과 컴퓨터간의 상호성을 더욱 향상시킬 수 있을 것으로 기대한다

본 논문에서는 얼굴의 움직임을 평면 모델을 사용하여 추적하게 하였다. 그러나, 인간의 얼굴이 3차원인점을 감안하면 평면모델이 아닌 3차원 모델-이를테면, 생성된 3차원 mesh 모델 자체-을 이용한다면 훨씬 좋은 성능을 가져올 것이다. 따라서, 이에 대한 연구도 기대 된다.

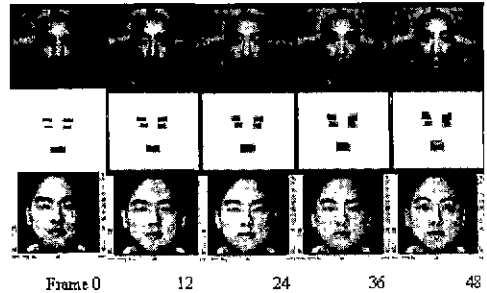


그림 3

참고문헌

- [1] I.Essa and A.Pentland, "A vision system for observing and extracting facial action parameters", In Proceedings of Computer Vision and Pattern Recognition, Seattle, WA, pp 76-83, 1992.
- [2] T.Darrell, B.Moghaddam, and A.Pentland, "Active face tracking and pose estimation in an interactive room", Submitted to Computer Vision and Pattern Recognition Conference, Nov, 1995
- [3] 이경희, 변혜란, 정찬섭, "스네이크를 이용한 얼굴요소의 윤곽선 추출 한국정보과학회 춘계학술대회 논문집, 1998
- [4] 고재필, 변혜란, 정찬섭, "주성분 분석에 기반한 객체학습 방법을 통한 얼굴영역 추출" 한국정보과학회 춘계학술대회 논문집, 1998
- [5] P.Elman and W.V.Friesen, Facial Action Coding System Palo Alto, CA: Consulting Psychologist Press, 1978
- [6] A.M.Waxman, B.Kamgar-Parsi and M.Subbarano, "Closeform solutions to image flow equations", In Proc int Conf on Computer Vision, London, England, pp.12-24, 1987
- [7] J.J.Koenderink and A.J.van Doorn, "Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer", Optica Acta, 22(9):773-791, 1975
- [8] M.J.Black and Yacob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion", In Proceedings of the International Conference on Computer Vision, pages 374-381, IEEE Computer Society, 1995.
- [9] R.Cipolla and A.Blake, "Surface orientation and time to contact from image divergence and deformation", In Proc ECCV-92, G.Sndini, Springer-Verlag, vol 588 of LNCS-Series, pp187-202, 1992.
- [10] Ajit Singh, Optic Flow Computation, IEEE Computer Society Press, 1991.
- [11] William H.Press, Brian P.Flannery, Saul A. Teukolsky, and Willam T.Vetterling. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, 1988.