

원근 움직임 모델을 이용한 특징 공간 상에서의 효율적인 얼굴 영역 추적

최송하, 이성환

고려대학교 컴퓨터학과/인공시각연구센터
E-mail: {shchoi, swlee}@image.korea.ac.kr

Efficient face tracking using perspective motion model in feature space

Song-Ha Choi and Seong-Whan Lee

Dept. of Computer Science and Engineering/Center for Artificial Vision Research,
Korea University

요약

본 논문에서는 입력 영상열에서 얼굴 영역을 추출하고, 영역 내 특징점들의 움직임 벡터를 원근 움직임 모델에 정합하여 얼굴 영역을 추적하는 새로운 방법을 제안한다. 제안된 방법은 계층적 형판 정합을 이용하여 얼굴 영역을 추출하고, 해당 영역에서 DoG 반응의 극부최대치를 찾아 특징점을 구한다. 그리고 최소제곱추정기법을 이용하여 각 특징점에서 일어난 움직임 벡터를 원근 모델에 정합한다. 제안된 방법은 선별된 특징점에서 움직임 벡터를 계산함으로써 연산량을 줄일 수 있었고, 원근 움직임 모델을 이용함으로써 잡영에 강인한 특성을 보인다.

1. 서론

최근 급속히 증가한 동영상 매체는 개인용 컴퓨터에서도 많이 사용될 정도로 대중화가 되었다. 이러한 동영상의 활용에 있어서 가장 관심과 연구의 대상이 되는 것은 바로 화상 회의, 보안 감시, 농업 자동화, 의료영상, 시각기반 제어 등에서 많이 이용되는 영역 추적이다.

영역 추적의 어려움 중 하나는 시간에 따른 대상 영역(target region)의 변화 가능성이다. 대상 영역은 크기, 형태, 각도 등의 연속적인 변화를 거치기 때문에 종종 대상 영역의 정보가 종종 손실된다. 그러므로 이러한 문제를 해결하기 위해서는, 대상 영역의 변화를 수용할 수 있는 움직임 모델과 이러한 움직임 모델을 이용하는데 사용되는 효율적인 알고리즘을 필요로 하게 된다.

기존의 영역 추적 방법은 크게 밀집 대응관계(dense correspondence)를 이용한 방법과 Snake 등에 기반한 윤곽선 추적 방법이 있다[1]. 밀집 대응관계 방법은 영역 내의 움직임 정보를 적절히 필터링하여 추적하고자 하는 방법이고, 윤곽선 추적 방법은 대상 영역의 윤곽선을 추출한 뒤 정의된 에너지 함수를 최소화하여 윤곽선의 움직임을 추적하는 방법이다. 밀집 대응관계의 방법은 사용할 수 있는 정보량에 비해 계산량이 지나치게 많아지게 되고 이에 비해 윤곽선 추적 방법은 연산량이 적으나 잡영 등으로 인하여 에너지 함수가 물체의 변형을 수용하지 못할 경우 이에 대한 보정이 곤란하게 된다.

움직임 모델을 이용한 기존의 영역 추적 방법에서는 주로 아핀(affine) 움직임 모델을 사용했다. Hager와 Belhumeur[2]는 아핀 모델과 광류를 이용하여 조영에 무관한 영역 추적 방법을 제안하였고, Black과 Jepson[3] 및 Torre, Gong, McKenna[1]는 아핀 모델을 이용하여 고유

벡터 공간 상에서 영역을 추적하였다. 원근 움직임 모델은 아핀 움직임 모델의 상위 집합으로서 3차원 공간 상에서의 선형 변환이 2차원으로 투영되어 나타나는 모든 변형을 수용한다.



그림 1. 아핀 변형과 원근 변형의 예

본 논문에서는 대상 영역의 다양한 변화를 수용할 수 있는 움직임 모델 정합 방법 및 대상 영역 내에서 특징점을 선별하는 방법을 제안한다. 사용된 움직임 모델은 원근 모델로서, 최근 많이 사용되고 있는 아핀 모델보다 일반적인 변형을 수용할 수 있다. 또한 3차원 실세계의 움직임을 표현할 수 있는 가장 기본적인면서도 확장성 있는 모델이다. 또한, 대상 영역에서 특징점에 해당하는 움직임 벡터를 고려하기 때문에 움직임 정보의 잡영은 이러한 일련의 과정에서 제거되고 수정된다. 제안된 모델 정합 방법은 오차의 제곱합을 최소화하는 방법에 기반하고 있다. 그림 2는 제안된 방법의 개괄적 흐름을 보여준다.

2. 원근 움직임 추적 방법

본 장에서는 입력 영상열에서 얼굴을 자동으로 추출하고 원근 움직임 모델에 정합하기 위한 과정을 설명한다. 1절에서는 얼굴 영역 검출 방법을, 2절에서는 얼굴 영역에서 특징점을 추출하는 방법을, 3절에서는 움직임 벡터를

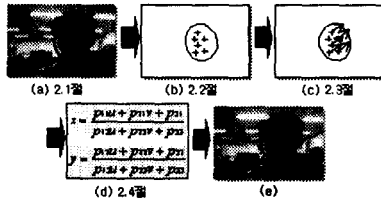


그림 2. 전체 흐름도: (a) 얼굴 영역 검출 (b) 특징점 추출 (c) 움직임 벡터 추출 (d) 원근 계수 추출 (e) 영역 갱신
구하는 방법을, 마지막 4절에서는 특징점에서 얻어진 움직임 벡터를 원근 모델에 정합하는 방법에 대해 설명한다.

2.1 얼굴 영역 검출

일반 영상에서 얼굴을 신속하게 검출하기 위해서 본 논문에서는 계층적 형판 정합을 이용하였다. 계층적 형판 정합은 입력 영상을 다해상도의 피라미드 표현으로 바꾼 뒤 각 계층의 영상에 대해 얼굴 형판을 영상 전체에서 정합하여 MSE(Minimum Squared Error)를 찾는다. 각 계층의 정규화된 MSE 값 중 최소의 위치가 최종적인 얼굴 영역이 된다.

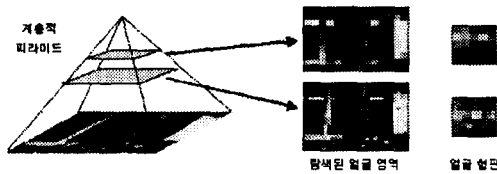


그림 3. 계층적 형판 정합의 과정

2.2 특징점 추출

찾아진 얼굴 영역에서 특징점을 추출하기 위해 우선 입력 영상에 DoG 함수의 필터링을 적용한다. DoG 함수는 기존의 연구에서도 보여주듯이 수용장과 유사한 반응값을 보여주기 때문에 생물학적 기반의 시각 연구에서 많이 사용되는 함수이다[6]. 같은 이유로 Gabor wavelet도 많이 사용되지만 본 논문에서는 방향성 정보는 그리 중요하지 않으며 DoG가 좀더 단순하기 때문에 DoG 함수를 채택하였다. DoG 함수는 다음과 같이 정의된다. 여기서 σ_1, σ_2 는 두 가우시안 함수의 분산이다. 이 매개변수를 조정하여 DoG 반응의 폭과 높이를 조정할 수 있다.

$$DoG(x, y) = \frac{e^{-\frac{(x^2+y^2)}{2\sigma_1^2}}}{2\pi\sigma_1^2} - \frac{e^{-\frac{(x^2+y^2)}{2\sigma_2^2}}}{2\pi\sigma_2^2} \quad (1)$$

이제 얻어진 DoG 반응 중에서 국부 최대점을 구한다. 국부 최대점은 전 영역에서 나타난 반응값 중 가장 높은 값을 갖는 점으로서 현재의 관심 영역이다. 주의 이동을 위해서 현재의 관심 영역에 자기 억제 함수를 적용한다. 본 논문에서는 주의 이동 경로를 필요로 하지 않기 때문에 주의 회복 과정은 생략하였다. 자기 억제 함수는 다음과 같이 정의된다. 단, α, β, A, B 는 상수이다.

$$M(\vec{x}_0, \vec{x}') = \left(1 + \alpha \exp\left(\frac{|\vec{x}_0 - \vec{x}'|^2}{A^2}\right)\right) \left(\beta \exp\left(\frac{|\vec{x}_0 - \vec{x}'|^2}{B^2}\right) - 1\right) \quad (2)$$



그림 4. 특징점 추출 과정

2.3 움직임 벡터 추출

앞서 구한 특징점 각각의 위치에서 움직임 벡터를 얻기 위해 본 논문에서는 상관관계 기반의 움직임 벡터 추출 방법을 이용한다. 이 벡터는 2차원으로서 특징점 주위의 일정 영역을 하나의 셀 단위로 정의하고, 해당 위치에서 일정 범위 내의 픽셀값 차이를 최소로 만족하는 위치까지의 좌표 차이를 움직임 벡터로 규정한다. 국부 최소값에 빠지는 것을 방지하기 위해 전역 탐색(Full search) 방법을 사용했다.

2.4 원근 움직임 모델의 계수 추출

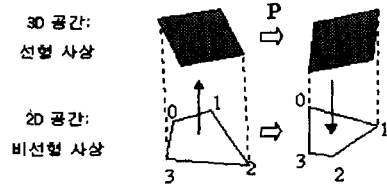


그림 5. 원근 움직임 모델의 선형화 과정

움직임 벡터들을 최적으로 만족하는 움직임 모델은 여러 가지가 있다. 그 중 가장 흔히 쓰이는 모델은 어파인 모델인데, 이는 2차원의 가장 일반적인 모든 변환을 수용할 뿐 아니라 모델의 계수가 선형적으로 얻어지기 때문이다. 이에 비해 원근 모델은 3차원 상의 선형 변환을 2차원으로 투영한 모든 변환을 수용할 수 있기 때문에 좀더 일반적이면서, 어파인 모델과의 연산량 차이는 그리 크지 않다. 그러나 원근 모델은 변환 자체가 비선형적이므로 계수의 연산이 복잡해진다. 본 장에서는 원근 모델의 계수를 3차원으로 변환하여 선형적으로 해결하는 방법을 설명한다. 다음은 원근 변환의 2차원 표현이다.

$$x = \frac{au + bv + c}{gu + hv + 1}, y = \frac{du + ev + f}{gu + hv + 1} \quad (3)$$

여기서 원근 변환을 3차원 변환으로 바꾸면 다음과 같이 표현할 수 있다.

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = P \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}, P = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{pmatrix} \quad (4)$$

이제 선형으로 변환된 원근 모델에서 최소제곱추정기법을 적용하면 모델 계수의 최적화는 다음 식 F를 최소화하는 문제로 재정의된다.

$$F = \sum_{i=1}^n \left\| \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} - \begin{pmatrix} \tilde{x}_i \\ \tilde{y}_i \\ \tilde{z}_i \end{pmatrix} \right\|^2$$

$$= \sum_{i=1}^n \left\| P \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} - \begin{pmatrix} \tilde{x}'_i \\ \tilde{y}'_i \\ \tilde{z}'_i \end{pmatrix} \right\|^2 \quad (5)$$

$$= \sum_{i=1}^n \{ (au_i + bv_i + c - \tilde{x}'_i)^2 + (du_i + ev_i + f - \tilde{y}'_i)^2 + (gu_i + hv_i + 1 - \tilde{z}'_i)^2 \} \quad (6)$$

$$= \sum_{i=1}^n \{ (au_i + bv_i + c - \tilde{x}_i \tilde{z}'_i)^2 + (du_i + ev_i + f - \tilde{y}_i \tilde{z}'_i)^2 + (gu_i + hv_i + 1 - \tilde{z}'_i)^2 \} \quad (7)$$

$$= \sum_{i=1}^n \{ (au_i + bv_i + c - \tilde{x}_i(gu + hv + 1))^2 + (du_i + ev_i + f - \tilde{y}_i(gu + hv + 1))^2 \} \quad (8)$$

여기서 z'_i 의 값은 원근 모델의 함수 $z'_i = gu + hv + 1$ 에서 얻어진다. g 와 h 는 식 3에서 입력 u, v 와 출력 x, y 로 얻어지는 계수이다. 위 식의 해를 구하기 위해 F 를 원근 모델의 계수 각각에 대해 편미분을 행한다. 여기서 얻어진 식들의 해는 위 식의 해와 동치이다.

$$\left(\frac{\partial F}{\partial a}, \frac{\partial F}{\partial b}, \dots, \frac{\partial F}{\partial h} \right) = 0 \quad (9)$$

3. 실험 결과 및 분석

3.1 실험 환경

본 논문에서 제안한 영역 추적 방법은 IBM 호환 333Mhz Pentium PC 상에서 실험하였다. 실험에 사용된 동영상 데이터는 얼굴이 등장하는 표준 데이터인 Claire 데이터를 사용하였다.

3.2 결과 영상

그림 6은 실험된 결과 영상을 보여준다. 원근 움직임 모델을 이용한 추적 결과는 얼굴의 회전 및 표정 변화에도 견고함을 보여준다.

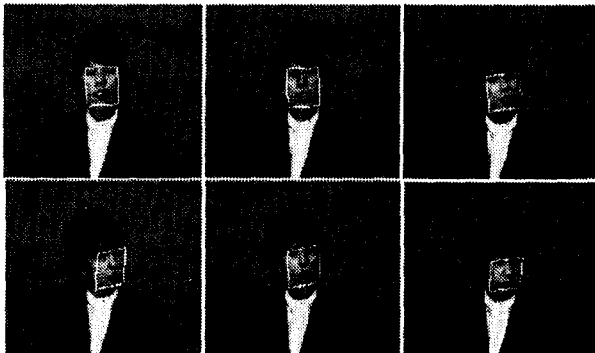


그림 6. 얼굴 영역 추적 결과

3.3 실험 분석 및 향후 연구 방향

본 논문에서는 DoG 필터를 이용한 특징점 추출과 원근 움직임 모델을 이용한 영역 추적 방법을 제안하였다. 입력 영상에 대하여 계층적 평면 정합을 이용해 얼굴 영역을

검출하였고, 검출된 얼굴 영역에 대하여 DoG 반응의 극부 최대점을 구하여 특징점으로 삼아 각 위치에서 움직임 벡터를 추출하였다. 최종적으로 얻어진 움직임 벡터의 집합을 실제 세계에서 3차원적으로 움직이는 모델의 선형적 움직임에 정합하기 위해 최소제곱추정기법을 이용하여 원근 모델의 계수를 구한다.

제안된 방법은 입력 영상에서 얼굴 추출 및 움직임 모델 정합에 이르는 추적 모형을 제시하였으며, 원근 모델을 선형화하여 2차원 평면 상에 투영된 움직임을 3차원으로 정합하는 과정의 이해와 구현이 쉬운 방법이다.

특징점 추출 과정은 선택적 주의 집중 모델에서 사용하는 주의 이동 경로 모형을 따랐기 때문에 계산 시간이 다소 소요된다. DoG 및 Gabor wavelet 반응값의 극부 최대점을 구할 때 주의 억제/회복 모델을 사용하지 않는다면 좀더 속도를 개선시킬 수 있는 여지가 있을 것이다. 또한, 본 논문에서는 원근 모델의 계수값을 일반적인 Gauss-Jordan 소거법을 이용하여 구했으나 대칭적으로 나타나는 계수 행렬의 특성을 최적화하는 알고리즘을 사용하면 좀더 적은 연산으로 움직임 모델의 계수를 구할 수 있을 것으로 생각된다.

제안된 방법은 주어진 3차원 목적 영역이 평면이라고 가정하였으므로 굴곡이 심한 물체는 추적하기 곤란하며, 가려짐이나 변형이 심한 사물에 대한 추적에 대한 연구가 요구된다. 따라서, 앞으로 3차원 메쉬 모형의 복잡한 관절 움직임 추적 문제에 대한 연구가 계속되어야 할 것이다.

감사의 말씀

본 연구는 부분적으로 과학기술부 지원 창의적연구진흥 사업의 연구비 지원을 받았음.

참고 문헌

- [1] F. de la Torre, S. Gong and S. McKenna, "View alignment with dynamically updated affine tracking," *Proc. Int'l Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, pp. 14-16, Apr. 1998.
- [2] G. D. Hager and P. N. Belhumeur, "Efficient Region Tracking with Parametric Models of Geometry and Illumination," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 10, pp. 1025-1039, Oct. 1998.
- [3] M. J. Black and A. D. Jepson, "Eigen tracking: Robust matching and tracking of articulated objects using view-based representation," *Int'l Journal of Computer Vision*, Vol. 26, No. 1, pp. 63-84, 1998.
- [4] A. K. Chhabra, "Real-time computation of optical flow along contours of significant intensity change," *Real-time Imaging*, Vol. 3, No. 2, pp. 87-99, Apr. 1997
- [5] G. Farin, *NURB: Curves and surfaces*, A. K. Peters, pp. 23-36, 1995.
- [6] D. Marr, *Vision*, W. H. Freeman and Company, pp. 61-64, 1982.
- [7] 장석우, 이근수, 최형일, "어파인 모델을 이용한 카메라의 동작 추출," *정보과학회논문지(B)*, 제 26권 제 8호, pp. 1000-1009, 1999년 8월.