

SIFT를 이용한 눈동자영역 추출

정재진, 황의성, 공재웅, 주동현, 김두영
 동아대학교 전기전자컴퓨터공학부

Eye Region Extraction Using SIFT

Jae-Jin Jung, Eui-Sung Hwang, Jae-Woong Gong, Dong-hyun Ju, Doo-young Kim
 School of Electrical, Electronic, and Computer Eng., Dong-A University

요 약

본 논문은 안면영상의 인증 요소로 사용되는 눈동자영역을 SIFT를 이용하여 추출해내는 방법을 제안하고 있다. 모델이 되는 눈동자영상과 추출 하고자 하는 입력영상의 SIFT 결과인 Keypoint descriptor를 이용하여 각각의 특징벡터를 구성하고 서로 정합한 후 두 특징 점들 사이에 affine transform이 존재하는지 판단하여 반수 이상에 대응하는 특징 점들에 대해 동일한 affine transform이 존재할 경우 눈동자영역이라고 판단하였다. 실험결과 학습과정이 없으므로 기존의 인식방법 보다 빠르게 영역을 추출하는 결과를 얻을 수 있었다.

I. 서 론

최근 휴먼 컴퓨터 인터페이스를 개발하기 위한 분야로서 얼굴, 눈, 입 등의 인식 기술에 대한 연구가 많이 진행되고 있다. 이 분야에서 이전에 행해진 많은 연구들은 얼굴영역의 추출에 중점을 두었기 때문에 추출된 얼굴영역의 확인을 위하여 눈이나 입의 위치만을 찾았다. 그러나 근래 들어 시간에 따른 눈동자나 그 형태 변화 및 움직임에 대한 정보를 추출하여 여러 분야에 응용하려는 연구가 활발히 진행되고 있다. 이중에서도 눈동자의 모양을 인식하거나 추적하는 연구는 상대방의 응시 방향을 분석하거나, 사용자와 컴퓨터간의 인터페이스에 사용되는 시선위치 추적(eye gaze detection)에 많이 적용되고 있다.[1] 이들 응용분야 중에는 객체의 움직임이나 위치를 빠른 시간 내에 검출을 요구하는 분야가 증가하고 있다.

객체의 위치 모양을 추출하는 방법에는 Kass 등이 제안한 능동 윤곽선 모델(Active Contour Model : ACM)이나 추출하고자 하는 형태를 PDM(Point Distribution Model)이라고 불리는 모델로 정의하여 이 모델과 비슷한 형태를 가지는 물체를 영상에서 찾는 방법인 능동 형태 모델(Active Shape Model : ASM)등이 있다. 이와 같은

방법들은 추출 대상 모델을 추정하기 위하여 많은 데이터를 필요로 하며 복잡한 학습단계를 거쳐야 객체의 추출 및 추적이 가능하다. 본 논문에서는 크기와 기울기 변화에도 좋은 결과를 얻을 수 있는 SIFT(Scale Invariant Feature Transform)를 모델과 입력영상에 적용한 후 Keypoint descriptor를 추출하여 각각의 특징 벡터를 구성, 정합한 후, 두 특징 점들 사이에 affine transform이 존재하는지 판단하여 반수 이상에 대응하는 특징 점들에 대해 동일한 affine transform이 존재할 경우 눈동자영역을 추출 하는 방법을 사용하였다.

II. 배경이론

1. SIFT(Scale Invariant Feature Transform)

SIFT는 1999년 Lowe에 의해 제안 되었으며 최근 인식 분야에 많이 적용되어 지고 있다. SIFT를 이용하여 Keypoint descriptor를 구하는 단계를 그림 1.에 나타내었다.

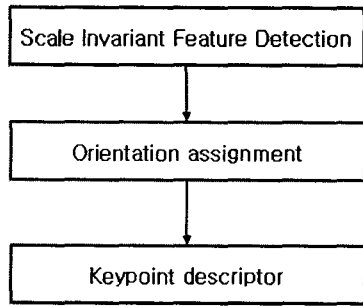


그림 1. SIFT 수행 과정

1) Scale Invariant Feature Detection

크기 불변 특징 점(Scale Invariant Feature)을 검출하기 위해서는 우선 크기 불변 공간(Scale-space)에서 가우시안 차분(Difference of Gaussian)을 계산한 후, 이웃한 8개의 픽셀들과 이웃한 두 크기(Scale)에서 극대와 극소가 되는 픽셀을 찾는다.

크기 공간은 아래의 그림 2와 같은 구조로 이루어져 있다.

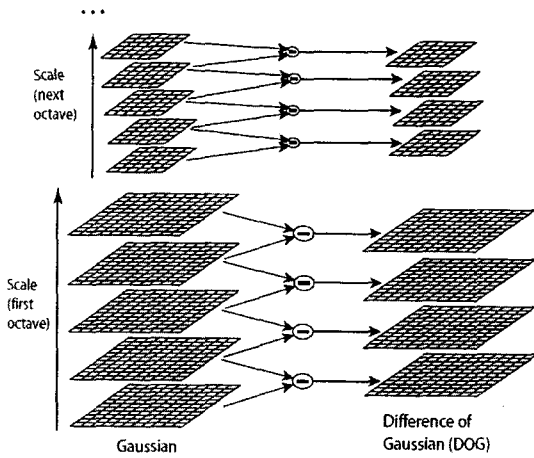


그림 2. 크기 공간의 생성

크기 불변 공간(Scale-space) 함수 $L(x, y, \sigma)$ 는 식(1)과 같이 입력영상 $I(x, y)$ 에 가변크기 가우시안 함수인 $G(x, y, \sigma)$ 를 컨볼루션 한 결과로 나타낼 수 있다.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

크기 불변 특징 점은 식(3)과 같이 계산되어 진다.

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (3)$$

$$= L(x, y, k\sigma) - L(x, y, \sigma)$$

극대와 극소는 그림 3과 같이 이웃한 8개의 픽셀들과 이웃한 두 개의 크기에서 DoG 값이 가장 크거나 DoG 값이 가장 작은 것을 의미한다.

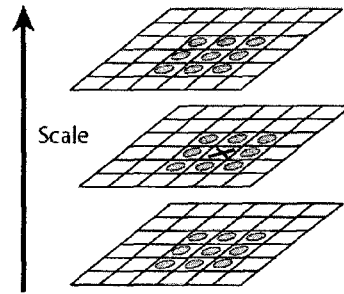


그림 3. 극대, 극소의 검출

실험결과 에지(edge) 근처에서 DoG 값이 극대, 극소가 발생하는 것을 볼 수 있었는데, 결과적으로 크기 불변 특징 점은 에지(edge) 근처에서 검출되는 것을 알 수 있었다.

2) Orientation assignment and Keypoint descriptor

검출된 특징 점의 주변 영역 $L(x, y)$ 에서 각각의 픽셀들의 gradient orientation인 $\theta(x, y)$ 와 magnitude $m(x, y)$ 를 구한다.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (4)$$

계산되어진 orientation을 36개의 각도로 양자화 하여 가장 빈도수가 높은 각도를 특징 점의 기준 orientation으로 결정하고 주변 영역 픽셀들의 orientation에서 기준 orientation을 뺀 새로운 orientation은 영상의 회전 변환에 불변인 canonical orientation이 되는데 이렇게 계산된 canonical orientation을 8개의 각도로 양자화 한다. 그림 4의 오른쪽과 같이 2x2의 영역에 대해 2x2개의 orientation histogram을 만들고 이를 특징 점의 특징 벡터인 Keypoint descriptor로 사용한다. 그림 4의 경우 각각의 특징 점마다 32 dimension의 특징 벡터를 얻게 된

다. 실제 실험에서는 4x4의 영역에 대해 8개의 각도로 양자화된 histogram를 사용하였으므로 각각의 특징 점마다 128 dimension의 특징 벡터를 구성 하였다.

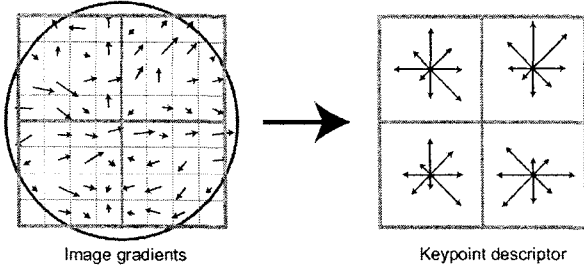


그림 4. Keypoint descriptor 생성

III. 눈동자영역 추출

아래 그림 5는 본 논문에서 사용한 눈동자 영역 추출 과정을 나타내는 전체 블록도이다.

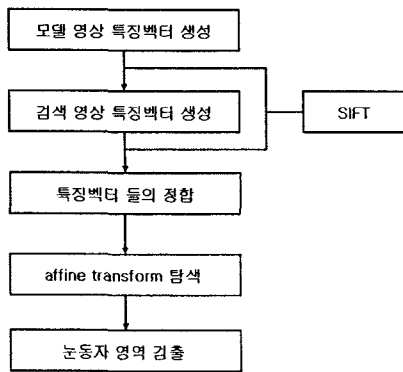


그림 5. 눈동자 영역 추출 전체 블록도

위 블록도에서 나타낸 것처럼 눈동자 모델영상과 검색 영상의 특징 점을 SIFT를 이용하여 추출하게 된다. 아래 그림 6은 추출한 특징 점을 나타내는 영상이다.

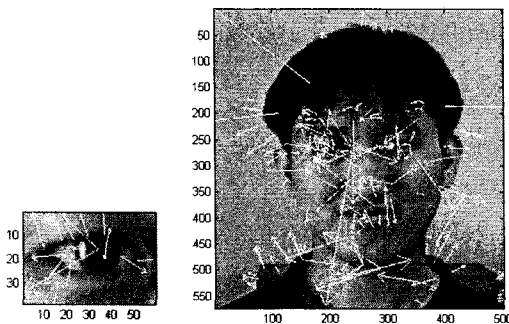


그림 6. 모델영상과 검색영상의 특징 점

추출된 특징 점들을 정합하는 것은 k-d 트리를 이용하

였는데 k-d 트리는 k dimension을 가지는 공간에서 입력으로 주어진 포인트에 인접한 포인트를 효과적으로 찾기 위한 검색 트리(search tree)이다. k-d 트리의 레벨은 순차적인 dimension에 따라 분리되며 트리를 구성하는 포인트의 수가 N개 이고 입력으로 주어진 포인트 주변의 일정 범위 안에 들어오는 n개의 포인트를 찾는다고 했을 때 걸리는 시간은 $O(\sqrt{N+n})$ 이 된다. 실험에는 모델 영상의 특징 벡터로 k-d 트리를 구성한 후, 검색영상에 대해 추출된 특징 벡터들과 가장 유사한 특징 벡터들을 찾는다. 이때 사용한 유사성의 척도는 euclidean distance 이고, 크기를 1로 정규화 하여 euclidean distance가 0.4보다 작은 경우에만 정합이 되었다고 판단하였다.

정합을 마친 후 특징 벡터들 사이에 affine transform이 존재하는지 판단하는데 모델영상 특징 벡터 $[x\ y]^T$ 의 검색영상의 특징 벡터 $[u\ v]^T$ 로 affine transform은 아래 식(5)와 같이 표현된다.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (5)$$

식(5)에서 $[t_x\ t_y]^T$ 은 전이값 이며 회전, 크기변화등은 파라미터 m_i 를 조절하여 나타내어지게 된다. 대응하는 특징 점들의 affine transform을 구하기 위해서 RANSAC(RANdom SAmples Consensus)을 사용하였다.

```

N:=Predefine iteration number
{P}:=The set of corresponding points
E(p):=residual of corresponding point p
<1>
i=0
prev_count=0
homography H
while i<= N
    Pick random four corresponding points in {P}.
    Calculate the affine transform Htemp using the four points.
    inlier_count = 0
    for all p in {P}
        if E(p)<Emax
            inlier_count = inlier_count +1
        end
        if inlier_count>prev_count
            prev_count = inlier_count
            H=Htemp
        end
    end
end
end
    
```

그림 7. RANSAC 알고리즘 pseudo code

최종적으로 계산되어진 inlier_count의 수가 전체 대응 점 수의 반을 넘을 경우 검색영상의 눈동자 영역이 된다. 그림 8은 검색영상에서 눈동자 영역을 검출한 결과이다.

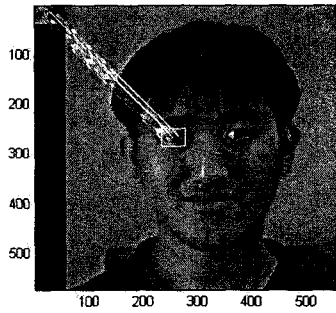
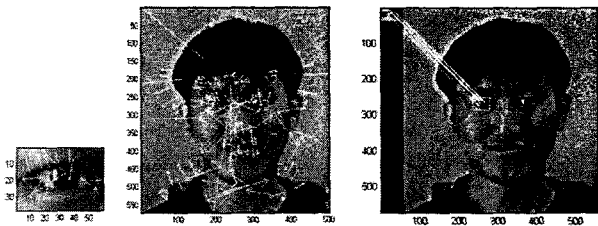


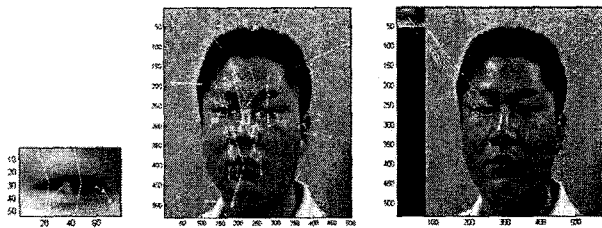
그림 8. 검색영상에 눈동자영역 추출 결과

IV. 실험 및 결과

SIFT를 사용하여 눈동자 영역을 추출하는 실험을 Raw 포맷의 안면 샘플 영상 20개를 대상으로 실시하였다. 아래 그림 9와 그림 10.은 실험 결과를 나타낸 그림이며, 각각의 a)는 모델의 특징 점 추출 결과 영상이고 b)는 검색 영상 특징 점 추출 결과, c)는 눈동자 영역 추출결과 영상이다.



(a) (b) (c)
그림 9. SIFT를 이용하여 눈동자 영역 추출 결과 I



(a) (b) (c)
그림 10. SIFT를 이용하여 눈동자 영역 추출 결과 II

V. 결 론

본 논문에서는 눈동자 영역 추출의 방법으로 SIFT의 Keypoint descriptor를 이용하였다. 우선 모델영상과 검색 영상의 Keypoint descriptor를 추출한 후 k-d 트리를 이

용하여 정합단계를 거치고 Keypoint descriptor간의 affine transform을 검색하여 최종적으로 눈동자영역을 추출하였다. SIFT의 경우 회전이나 크기변화 등에는 좋은 결과를 볼 수 있었지만 밝기나 형태의 변화에는 눈동자 영역을 추출하지 못하는 결과를 볼 수 있었다. 차 후 연구에서는 이러한 문제점들을 개선해 나갈 계획이다.

참 고 문 헌

- [1] Thomas. E., Preston W., Worthy N., Kelly C., Lisa A. Prey, "Human-Computer Interaction Using Eye Gaze Input" IEEE Trans. on Systems, Vol. 19, No. 6, pp. 1527-1534, 1998
- [2] David G. Lowe, "Object recognition from scale-invariant features", International Conference on Computer Vision, Corfu, Greece, pp. 1150-1157, 1999
- [3] David G. Lowe, "Distinctive image features from scale-invariant keypoints", preprint, submitted to International Journal of Computer Vision. 2004
- [4] Bhowmick, Bhattacharya, B.B, "Approximate fingerprint matching using kd-tree", International Conference on Pattern Recognition, pp. 544-547 Vol.1 2004
- [5] Hongsheng Zhang, Negahdaripour, "Improved temporal correspondences in stereo-vision by RANSAC" International Conference on Pattern Recognition, pp. 52-55 Vol.4 2004