

# 수정 윈도우를 이용한 얼굴 특징점의 추적

김정선, 조남익

서울대학교 전기컴퓨터공학부

## Facial Feature Tracking With Modified Windows

Jung Sung Kim, Nam Ik Cho

School of Electrical Engineering, Seoul National University, Seoul 151-742, Korea

E-mail : mirr@ispl.snu.ac.kr, nicho@snu.ac.kr

### 요약

본 논문은 얼굴의 특징점 추적을 위하여 얼굴의 회전 변환과 크기변환을 고려한 BMA(Block matching algorithm)을 이용한 방법을 제안한다. 우선 얼굴의 크기 변화를 구하기 위하여 얼굴 영역을 분리하여 그 면적을 구한다. 이 면적을 이전 프레임에서 구한 얼굴 영역의 면적과 비교하여 크기 비례를 계산한다. 다음으로 각 특징점을 중심으로 하는 8방위 영역의 화소들로 집합을 설정한다. 집합을 설정할 때에는 얼굴의 크기 변화를 고려하여 영역 내 화소들을 포함하는 양을 수정한다. 그리고 새로운 영상에서 화소 집합간의 거리가 가장 작은 화소를 새로운 특징점으로 지정한다. 이때, 회전 변환을 고려하여 화소 집합의 순서를 순차적으로 바꿔 집합 간 거리를 산출한다. 제안하는 방법은 회전과 크기 변환에 강인한 특성을 보일 뿐 아니라, 단순한 움직임 예측 방법인 BMA보다도 쉽고 빠르게 계산된다.

## 1 서론

얼굴 특징점의 추적은 얼굴 애니메이션이나 화상 회의 등과 같이 얼굴의 움직임을 주요 내용으로 하는 응용 기술에서 매우 중요한 역할을 한다. 이러한 응용에서는 얼굴의 주변 환경의 변화가 거의 없으므로 영상을 전체적으로 부호화하는 것보다 중요시되는 얼굴의 움직임만을 부호화하는 것이 훨씬 효율적이다. 이를 위하여 MPEG-4에서는 얼굴 특징점 전송을 위하여 FAP(Facial animation parameter)와 FDP(Facial define parameter)를 표준화하였다 [1, 2]. 이를 통하여 입력 영상에서 얼굴의 특징점들만이 부호화되어 전송되어 출력에서 쉽게 재구성된다. 눈, 코, 입 등 얼굴의 특징적인 부분을 추출하는 기존의 방법은 주로 윤곽선이나 색도를 기반으로 하고 있다 [3, 4, 5]. 최근에는 이를 위하여 ASM(Active shape model)이나 PCA(Principal component analysis) 등이 사용되고 있다. [6, 7, 8, 9, 11]. 그러나 윤곽선에 기반한 특징점 추출은 사람의 얼굴 특성에 최적화되지 못하는 점이 많다. 사람의 얼굴은 인공적으로 만들어진 얼굴 모델과는 다르게 명확한 윤곽선으로 특징적인 부분을 구분할 수 없다. 눈썹의 끝이나 입술의 경계는 윤곽선만으로 결정하기에는 불확실한 면이 있다. 또한 ASM이나 PCA는 상당한 양의 계산을 필요로 한다. 특징점을 추적하는 전형적인 방법은 BMA(Block matching algorithm)이다. MPEG이나 H.26x와 같은 비디오 압축 표준에서 움직임 예측 방법으로 사용되고 있는 BMA는 계산 방법이 간단하고 평행 이동된 움직임에 강인한 성질을 보인다. 하지만 지정한 영역과 똑같은 영역을 찾는 BMA의 특성 상 회전이나 크기 변환이 일어났을 경우 성능이 현저하게 떨어진다. 따라서 본 논문에서는 BMA의 특성을 이용하는 것과 동시에 계산량을 줄이고 크기와 회전 변환에 강인한 특징점 추적 방법을 제안한다. 이를 위하여 우선 얼굴의 크기 변화를 측정하기 위한 얼굴 영

역을 분리한다. 분리된 얼굴 영역을 이전 영상의 얼굴 영역과 크기를 비교하여 변화량을 구한다. 각 특징점에서는 특징점을 중심으로 하는 8방위에서 화소의 휘도와 색도 값을 취하여 새로운 집합을 만든다. 이때 값을 선택하는 범위는 먼저 구한 영역의 크기 변화에 따라 조정한다. 새롭게 만들어진 집합은 이전 영상에서 만들어진 집합과 거리를 비교하여 가장 작은 차이를 보이는 지점을 새로운 특징점으로 선택한다. 집합 간의 거리를 비교할 때에는 회전에 대하여 고려하기 위해 8방위에서 취한 화소들의 값을 순차적으로 교환하여 집합을 수정하면서 최소의 거리를 구한다. 제안하는 방법은 기본적으로 BMA에 기반하므로 매 영상마다 특징점을 추출하는 방법으로 특징점을 추적하는 것에 비해 훨씬 빠르고, 특징점 주위의 영역을 고정된 BMA보다 얼굴의 특징점에 따라 유연하게 특성화시킬 수 있다. 또한 BMA보다 계산량이 훨씬 줄어들 뿐만 아니라 휘도에만 의존하지 않고 색도까지 고려할 수 있다. 그리고 회전과 크기 변화에 따른 수정이 어려운 BMA에 비하여 약간 추가적인 계산으로 회전과 크기 변화까지 고려에 넣을 수 있다. 본 논문에서는 2.1에서 얼굴 영역을 추출법에 대하여는 하고 2.2에서 화소 집합을 구성하는 방법에 대하여 기술한다. 2.3에서는 각 화소 집합을 크기 변화에 따라 수정하고 서로 비교하는 방법에 대하여 기술한다. 실험 결과는 3장에서 보이고 마지막으로 4장에서 마무리한다.

## 2 알고리즘

본 논문에서 제안하는 알고리즘은 크게 세 가지 부분으로 나뉜다. 우선 영상 간의 얼굴 영역 크기 변화를 측정하기 위한 얼굴 영역 추출 부분이 있다. 다음으로는 특징점을 중심으로 하여 화소 집합을 만드는 부분과 만들어진 화소 집합을 조정하면서 거리 차이를 구하는 부분이 있다.

### 2.1. 얼굴 영역의 추출과 보정

사람의 얼굴 내부의 구성 요소들, 즉 눈, 눈썹, 코, 입은 일반적으로 같은 비율로 커지거나 작아진다. 특히 구성 요소들은 얼굴 크기의 변화와 같은 비율로 변화한다. 따라서 얼굴 영역의 크기를 먼저 구하여 이전 영상의 얼굴 영역 크기와 비교하면 전체적인 크기 변화를 구할 수 있다. 이것은 각 구성 요소들의 크기 변화에 그대로 활용이 가능하다. 얼굴 검색과 인식의 기본 단계로 사용되는 얼굴 영역의 추출에는 여러 가지 방법이 사용된다. 색도를 이용한 기본적인 방법 외에도 PCA를 이용하거나 템플릿을 이용하는 방법 등 여러 가지 복잡한 방법들이 많다. 그러나 본 논문은 얼굴의 특징점 추적을 목적으로 하고 있고, 기본적으로 영상에 얼굴 영역이 분명히 존재하며 중심부 대다수 지역을 차지한다고 가정하므로 따라서 본 논문에서는 복잡하고 정밀한 방법보다 색도를 이용하여 얼굴 영역

을 추출하는 빠른 방법을 선택한다. 색상은 휘도에 비하여 조명의 영향이 약하고 얼굴 영역을 배경에서 확실하게 구분하는 특징이므로 휘도에 따른 윤곽선을 이용한 얼굴 영역 구분보다 여러 가지 경우에 있어 강인한 성능을 보인다. 얼굴 영역의 분리는 미리 결정되어 있는 얼굴의 색도 모델을 기준으로 한다. 데이터 베이스에서 미리 구한 색상 정보를 기준으로 문턱치 이하에 포함되는 색상 정보를 갖는 화소는 일단 얼굴 영역에 포함시킨다. 즉, 영상에서  $(x, y)$  좌표에 위치한 화소의 값을  $P(x, y)$ 라고 했을 때, 다음의 식을 만족하면  $P(x, y)$ 는 얼굴 영역으로 구분한다.

$$\sqrt{(C_b(x, y) - CentreC_b)^2 + (C_r(x, y) - CentreC_r)^2} < Threshold \quad (1)$$

여기서  $C_b(x, y)$ 와  $C_r(x, y)$ 는 각각  $P(x, y)$ 의  $C_b$ 와  $C_r$  정보이고  $CentreC_b$ ,  $CentreC_r$ 은 모델에서 결정된 얼굴 색상의  $C_b$ 와  $C_r$ 을 의미한다. 입술 영역은 인위적 처리가 없는 한 얼굴 영역의 색도 범위 내에 포함된다. 코와 콧구멍 역시 휘도 차이가 고려되지 않기 때문에 영역 안으로 흡수된다. 눈썹과 눈의 일부도 얼굴 영역으로 포함되지만 색상 차이가 선명한 눈썹 머리 부분과 눈동자 부분은 포함되지 않는다. 또한 완전히 깨끗한 배경을 전제로 하는 것이 아니기 때문에 배경 부분에서도 일부 색상 영역에 포함되는 부분이 생긴다. 얼굴의 면적은 얼굴 전체 면적을 기준으로 계산되므로 얼굴 내부에 있는 빈 공간, 즉 눈썹과 눈이 만든 부분은 얼굴 영역으로 포함시켜야 한다. 또한 얼굴 외부에서 얼굴로 인식되는 일부 구간은 제거해야 한다. 그러기 위해서 erosion과 dilation 기법을 이용한다. erosion과 dilation을 이용하면 배경에 조금씩 나타나는 영역은 사라지고 얼굴 내부의 빈 공간도 채워진다. 또한 커와 목, 이마선 등의 영향으로 불규칙적이었던 외곽선 역시 깨끗하게 정리된다. 얼굴 면적은 표준화된 영역에 포함되는 화소의 개수로 결정한다.

## 2.2. 화소 집합의 구성

비디오 부호화에서 사용하는 BMA는 현재 영상의 일정 영역과 가장 유사한 영역을 이전 영상에서 찾는다. 얼굴 특징점 추적에 BMA를 이용한다면 특징점을 중심으로 하는 영역을 기준으로 유사한 영역을 다음 영상에서 찾아야 한다. BMA에서 기준으로 하는 영역, 즉 윈도우의 모양은 대부분의 경우 고정적인 직사각형이다. 따라서 영상에 회전이나 크기 변화가 생기면 이전의 윈도우와 정확히 일치하는 영역을 찾을 수 없다. 이러한 점을 보완하기 위하여 비디오 압축에서는 여러 가지 방법이 제안되고 있다. 예를 들어 HMA(hexagonal matching algorithm) [10]는 삼각형 윈도우와 3 개의 삼각형 윈도우를 연결한 6각형 윈도우를 사용하여 관계 변환(Affine transform)을 수행하여 윈도우의 크기와 각도를 유연하게 조정한다. 그러나 이 경우 관계 변환에 사용될 각도와 크기를 구하기 위하여 작은 영역의 BMA를 여러번 수행해야 하므로 계산량이 기존의 BMA에 비해 현저하게 증가한다. 또한 얼굴의 특징점은 작은 지역에 밀집되어 있으므로 HMA에서처럼 각도와 크기를 구하기 위한 BMA를 다시 사용하는 것은 효율적이지 못하다. 일반적인 경우 BMA는 영상의 휘도 정보에 대해서만 적용된다. BMA는 여러 가지 추적 기법에 비하여 계산량이 적은 편이지만 색상 정보에까지 이용할 경우 계산량은 3배로 증가하게 된다. 사람의 시각이 색상보다 휘도에 민감하기 때문에 많은 영상 정보들이 휘도에 보다 많은 비중을 두지만, 얼굴과 같이 윤곽선에서의 휘도차가 명확하지 않고 휘도 변화가 전체적으로 적을 경우에는 유사도 측정에 있어 색상정보도 함께 활용하는 것이 보다 정확한 결과를 기대할 수 있다. 본 논문에서는 이와 같은 점을 고려하여 얼굴의 크기 변화와 회전에 유연하게 대처하고 휘도뿐만 아니라 색상 정보까지 포함하여 특징점을 나타낼 수 있는 집합을 제안한다.

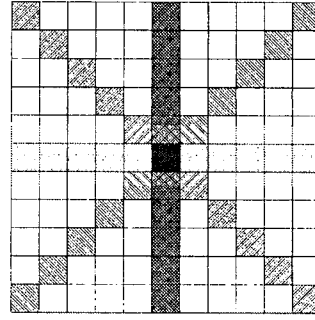


그림 1: 화소 집합의 구성

새로운 집합은 특징점을 기준으로 하여 8방위의 화소 값을 원소로 갖는다. 즉 그림 1와 같이 특징점을 포함하는 수직 방향의 화소와 수평방향의 화소, 대각선 방향의 화소가 순차적으로 화소 집합에 포함된다. 또한 각 화소의 휘도뿐만 아니라 색상 정보 역시 같은 식으로 집합에 포함시킨다. 따라서 한 개의 화소 집합은 다음과 같이 구성된다.

$$PixelSet = \{Y_1, Y_2, \dots, Y_N, C_{b1}, C_{b2}, \dots, C_{bN}, C_{r1}, C_{r2}, \dots, C_{rN}\} \quad (2)$$

여기서 N은 특징점 상하 혹은 좌우로 집합 내에 포함될 화소 개수를 의미한다. N이 클수록 특징점에서 멀리 떨어진 화소까지 고려할 수 있다. 하나의 특징점에 대하여 위와 같은 화소 집합은 각 방향에 대하여 1개씩 총 4개가 생긴다.

## 2.3. 비교

비디오 부호화에서 사용되는 움직임 예측 등과 같이 BMA의 기본적인 유사도 측정 방법은 영역간의 차이의 총합(Sum of Absolute difference)이다. 현재 영상의 일정 영역 안에 포함된 화소들과 이전 영상의 검색 영역 내에서 같은 크기의 영역 안에 포함된 화소 간의 1:1 차이를 구하여 전체 영역에 걸쳐 합산한다. 이 값이 가장 작은 영역이 현재 영상의 영역과 일치하는 부분으로 결정된다.

$$SAD = \begin{aligned} & \alpha \cdot |Cur(S_{ver}) - Next(S_{ver})| + \\ & \beta \cdot |Cur(S_{hor}) - Next(S_{hor})| + \\ & \gamma \cdot |Cur(S_{diag1}) - Next(S_{diag1})| + \\ & \delta \cdot |Cur(S_{diag2}) - Next(S_{diag2})| \end{aligned} \quad (3)$$

$\alpha, \beta, \gamma, \delta : weights$   
 $\alpha + \beta + \gamma + \delta = 1$

이 때  $Cur(\cdot)$ 과  $Next(\cdot)$ 은 각각 현재 영상과 다음 영상에서의 값을 의미하고,  $S_{ver}$ ,  $S_{hor}$ ,  $S_{diag1}$ ,  $S_{diag2}$ 는 각각 수직 방향, 수평 방향, 원대각선 방향, 오른대각선 방향의 화소 집합을 말한다. 이와 같은 식으로, 화소 집합의 원소 간의 차이를 총합하여 그 값이 가장 작은 화소를 새로운 특징점으로 선택한다. 즉, 다음 영상의 검색 영역 내의 화소들마다 화소 집합을 만들어 그것을 현재 영상의 특징점의 화소 집합과 비교한다. 이 때, 각 특징점의 특성에 따라 4 방향에 다른 가중치를 부여한다. 수평데이터가 중요한 경우, 수평 데이터를 포함하는 화소 집합을 수직 데이터를 포함하는 화소 집합보다 큰 가중치를 두어 수평 데이터가 새로운 특징점을 결정하는 데 보다 큰 영향을 미치도록 조정한다. eq. (3)의 SAD가 가장 작은 값이 새로운 특징점이 된다.

만약 2.1의 과정을 통해 크기 변화가 감지되어 그 변화율이 계산되었다면, 화소 집합의 크기를 그에 따라 수정한다. 즉,

다음 영상의 얼굴이 현재 영상보다 작아진다면 그 비율대로 화소 집합의 원소를 다운샘플링한다. 반면 현재 영상보다 얼굴의 크기가 커진다면 원소들을 보간하여 집합의 크기를 증가시킨다. 이러한 경우에는 다음 영상에서 화소 집합을 구할 때에도 집합의 크기를 비율에 따라 현재 영상의 경우보다 크게 하거나 작게 한다. 다시 말해서, 2-2에서 말한 N의 값은 면적비에 따라 커지거나 작아진다. 이 때  $S_{mod}(i)$ 는 새롭게 구성되는 화소 집합의  $i$ 번째 요소를 나타낸다.  $A_{next}$ 와  $A_{cur}$ 는 각각 다음 영상에서의 얼굴 영역 면적과 현재 영상에서의 얼굴 영역 면적이다. 얼굴 영역이 작아지거나 커졌을 때는 다음과 같은 식을 이용하여 새로운 화소 집합을 구한다.

$$S_{mod}(i) = S(i \times \frac{A_{next}}{A_{cur}}) \quad (4)$$

큰 영역에서 BMA는 일반적으로 회전 변환에 대하여 강인하지 못하다. 회전 각도가 커질수록 BMA는 현재 지정되어 있는 영역과 유사한 영역을 찾지 못한다. 그런데 얼굴 영상과 같은 경우 2,3초 간격으로 샘플링된 영상이라 할 지라도 큰 회전이 일어나지 않는다. 또한 SAD를 기준으로 하는 유사도 측정법은 구해진 값들 중 최소의 값을 가지는 지점을 선택하므로 45도 이상의 큰 회전이 일어나지 않는 이상 적정 수준에서 벗어나지 않는 추적을 계속할 수 있다. 본 논문에서는 화소 집합의 회전은 미리 결정하지 않고 비교 시점에서 화소 집합들을 회전시켜 SAD를 측정하여, 회전시키지 않았을 때의 SAD와 비교해서 회전 여부를 결정하고 현재의 화소 집합들을 회전시킨다. 화소 집합은 특징점을 기준으로 하여 그림 1처럼 구성된다. 영상이 회전하게 되었다면 그림 1은 그림 2처럼 변형된다. 따라서 화소 집합을 회전시키기 위해서는 기존의 화소 집합의 순서를 바꿔주면 된다. 이것은 좌우로 45도의 회전을 의미한다. 이러한 회전 변환을 통하여 45도 이상의 회전 변화에 적절하게 대응할 수 있다.

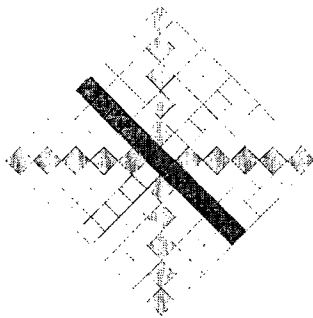


그림 2: 화소 집합의 회전

$$\begin{aligned} Srot_{ver} &= S_{diag1} \\ Srot_{hor} &= S_{diag2} \\ Srot_{diag1} &= S_{ver} \\ Srot_{diag2} &= S_{hor} \end{aligned} \quad (5)$$

여기서  $Srot_{(.)}$ 은 회전된 화소 집합을 의미한다.  $S_{diag1}$ 과  $S_{diag2}$ 는 왼쪽 회전인 경우 각각 원대각선 방향 화소 집합, 오른쪽 대각선 방향 화소 집합이며 오른쪽 회전인 경우 반대가 된다.

### 3 실험 결과

얼굴의 특징점은 눈, 코, 입과 같이 얼굴의 중요부분의 각 끝점과 최상위점, 최하위점을 의미한다. 본 논문에서는 양 눈에서

4개, 코에서 4개, 입술에서 6개 등 총 18개의 특징점을 지정하였다. 입술에서는 입의 움직임을 추적하기 위하여 양끝과 상하점에 양 입술 사이의 지점을 추가한다. 특징점의 추적을 위한 초기 위치 설정은 경험적으로 결정하거나 ASM을 이용하여 설정한다. 입술을 제외하고는 얼굴의 특징점들은, 특히 화상 회의와 같은 경우, 초당 변화가 심하지 않다. 그러므로 초당 5장 이상의 영상처리 속도를 보인다면, 1초에 최대 30장의 영상을 처리해야만 하는 실시간 구현에서 중간 단계는 보간법을 사용하여 충분히 활용될 수 있다. 본 논문에서 제안하는 방법을 이용하면, SIF(320x240) 영상에 대하여 얼굴 영역 분리를 포함하여 초당 6장 정도의 영상을 처리할 수 있다.

### 4 결론

본 논문에서는 얼굴의 특징점을 추적하는 방법에 대해서 제안했다. 제안된 방법은 BMA를 응용하여 특징점을 나타내는 새로운 집합을 구성하여 회전과 크기 변화에 강인하도록 수정하였다. 얼굴 영역을 미리 검출하여 크기 변화를 측정함으로써 크기 변화에 강인하도록 하였고 새롭게 구성된 화소 집합을 45도 간격으로 회전시켜 비교함으로써 회전 변화에 대응하도록 하였다. 뿐만 아니라 화소 집합은 그 수가 BMA에서 특징점을 포함하는 영역의 크기보다 훨씬 작아 수행속도가 BMA보다도 빠르고 계산의 복잡도가 다른 추적 방법에 비해 적다. 이를 통하여 실시간 구현 역시 가능하다.

### 5 References

- [1] ISO/IEC JTC1/SC29/WG11 N1666pub MPEG-4 SNHC VM, April 1997
- [2] Jorn Ostermann, "Animation of Synthetic Faces in MPEG-4" *Computer Animation 98. Proceedings*, pp 78-112, 1998.
- [3] Chang, T.C., Huang, T.S., Novak, C. "Facial Feature Extraction from Color Images" *Pattern Recognition - Conference B: Computer Vision & Image Processing*, vol. 2, 1994 *Proceedings of the 12th IAPR International Conference on*, vol. 2, pp 39-43, 1994.
- [4] Wu, H.; Yokoyama, T.; Pramadhanto, D.; Yachida, M. "Face and Facial Feature Extraction from Color Image" *Automatic Face and Gesture Recognition. Proceedings of the Second International Conference on* pp 345-350, 1996
- [5] Yokoyama, T.; Wu, H.; Yachida, M. "Automatic detection of facial feature points and contours" *Robot and Human Communication. 5th IEEE International Workshop on*, pp 335-340, 1996.
- [6] Jordao, L.; Perrone, M.; Costeira, J.P.; Santos-Victor, J. "Active Face and Feature Tracking" *Image Analysis and Processing, 1999. Proceedings. International Conference on*, pp 572-576, 1999.
- [7] Cootes, T.F.; Taylor, C.J. "Statistical Models of Appearance for Computer Vision" *Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, Manchester M13 9PT, United Kingdom*, 1999.
- [8] Kirby, M.; Sirovich, L. "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, pp. 103-108, 1990.

- [9] Yueta, P.C.; Dai, D.Q.; Feng, G.C. "Wavelet-based PCA for Human Face Recognition" *Image Analysis and Interpretation, 1998 IEEE Southwest Symposium on* , pp. 223-228, 1998.
- [10] Nakaya, Yuicro; Harashima, Hiroshi "Motion Compensation Based on Spatial Transformations" *IEEE Transactions on Circuit and System for Video Technology*, vol 4, No. 3, pp. 339-355, 1994.
- [11] Liew, Alan W.C.; Sum, K.L.; Leung, S.H; Lau, W.H. "Fuzzy Segmentation of lip image using cluster analysis" *Proceeding of EuroSpeech '99*, vol 1, pp. 335-338, 1999.