

# 하우스도르프 거리매칭과 캐리커처 효과를 이용한 얼굴표정 인식\*

박주상<sup>0</sup> 오경환  
한국전자통신연구원 서강대학교 컴퓨터학과  
kappa@etri.re.kr, kwol@ccs.sogang.ac.kr

## Facial Expression Recognition using Hausdorff Distance Matching and Caricatural Effect

Joo-Sang Park<sup>0</sup> Kyeong-Whan Oh  
Dept. of EC/CLAS, ETRI Dept. of Computer Science, Sogang University

### 요 약

기존의 얼굴표정 인식연구의 대부분은 얼굴영상에서 사전정보 획득과, 인식이 각각 별개로 수행되어, 전자의 결과가 후자를 보장하지 못하거나, 데이터와 계산 양의 과다, 그리고 인지과정이 사람과 다르다는 등의 문제가 있다. 이에 대해 하우스도르프 거리 매칭을 적용, 표정인식을 시도한다. 이는 전체적인 유사도를 측정하는 방법으로서 전체이론(Holistic theory)에 기반하여, '사람의 인지과정을 따른다. 그러나 축소된 데이터를 사용하므로, 이 방법의 인식결과가 부족한 경우, 영상위평을 적용하여 Brennan과 Garton이 제안한 캐리커처 효과를 이용한다. 이는 영상을 적절히 변형, 표정의 특징을 과장하고 잡음을 제거하여, 인식하기 쉬운, 분명한 표정을 생성하는 방법이다. 위 과정을 통해, 사람의 인지과정을 모사하고, 최소한의 데이터로써 사전정보 획득과정이 생략된, 입력영상으로부터 직접 표정을 인식하는 방법을 제안한다.

### 1. 서론

사람은 제스처와 여러 가지 얼굴 표정을 이용하여 의사를 전달하고 감정을 표현할 수 있다. 컴퓨터로 이러한 사람의 표정을 인식하기 위해 여러 방법들이 개발되었다. 그러나 기존 방법들은 여러 제약조건 하에서 수행되거나, 계산량이 많다. 조명, 포즈 등의 변화가 인식률을 크게 저하시키거나, 방법에 따라 수십 초 내지 수십 분의 시간이 소요되는 등의 단점이 있다.[7]또 인지과학 측면에서 볼 때, 인식방법이 인간과 아주 다르다는 공통의 문제가 있다. 사람은 단지 '훑듯 쳐다보기' 만으로도 인식이 가능하고, 주로 윤곽선으로 표정을 인식하는데, 이렇게 하면 인식에 필요한 데이터 양은 3~5% 정도에 불과하다.[5]

본 연구는 두 가지 문제로부터 출발한다. 첫째는 기존 방법이 인식을 위한 사전정보로서 특징점이나 기하학적 구조 정보를 따로 입력하거나 별도로 분리하여, 사전 정보 추출과 인식 연구가 별개로 진행되어, 전자의 결과가 후자의 성능을 보장하지 못하는 점이다. 둘째는 인식 방법이 인간과 크게 다르다는

점이다. 둘째는 인식 방법이 인간과 크게 다르다는 점이다. 인지과학 연구에 따르면 인간은 특징점 위치와 같은 구조 정보나 픽셀의 벡터같은 부분 정보가 아니라 전체(Holistic) 정보를 사용하여 인식한다.[1][5] 또 Takacs에 따르면 실제 인식에 필요한 데이터는 전체의 3~5%였다. 본 연구는 위 문제의 대안으로써 전체 이론(Holistic Theory)에 따라 두 영상의 전체 유사도를 측정하는 하우스도르프 거리 매칭(Hausdorff Distance Matching) 방법을 적용하고, 영상위평으로 표정을 강조한 캐리커처를 생성하여 인식하였다.

### 2. 본론

#### 2.1 하우스도르프 거리 매칭

하우스도르프 거리(Hausdorff Distance, 이하 HD)는 수학자 Felix Hausdorff에 의해 제안되어 '두 집합간의 전체 유사도'를 측정하는데 쓰인다. 이 거리 개념은 '한 집합으로부터 다른 한 집합의 가장 가까운 점까지의 거리'를 의미하며, 다음과 같이 정의되는 최대(Maximum)함수이다.[6]

\* 본 논문은 BK21 지능형 정보검색 에이전트 과제에의 일부로 수행된 결과임.

$$h(A, B) = \max_{a \in A} \{ \min_{b \in B} \{ d(a, b) \} \}$$

그러나, HD는 비대칭성으로 인해 대부분의 경우에  $h(A, B)$ 와  $h(B, A)$ 가 다르므로, 일반적인 정의는 다음과 같다.

$$H(A, B) = \max\{h(A, B), h(B, A)\}$$

이를 이용한 매칭 방법은 두 영상으로부터 검출된 에지를 이용, HD를 측정하여 기준영상이 시험영상(test image)에 존재하는지를 알아내는데 사용한다. 거리 값이 작을수록 더 잘 부합된다. 이 방법은 아래 예처럼, 잡영이 있거나, 부분적으로 지워지거나 또는 뒤틀린 영상에서도 좋은 결과를 보인다.

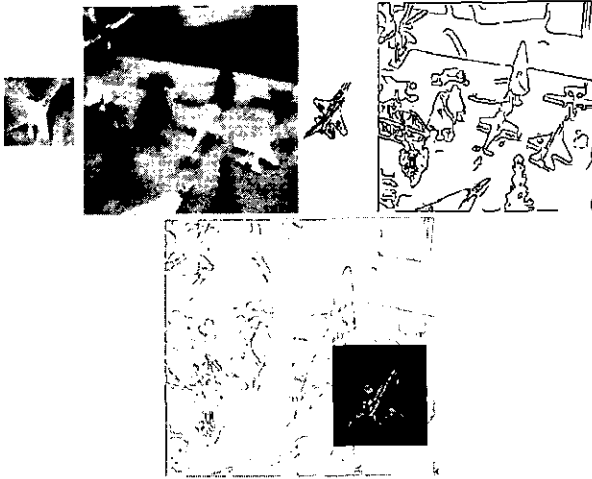


그림 1. 하우스도르프 거리 매칭 예[6]

### 2.2 캐리커처 효과

캐리커처(Caricature)는 16세기 말부터 시작되었으며, 정치 만평이나 특정인의 얼굴특징을 과장해서 그릴 때 쓰인다. 다음 그림은 캐리커처가 영상의 좌표변환에 근거함을 보여준다.[4]

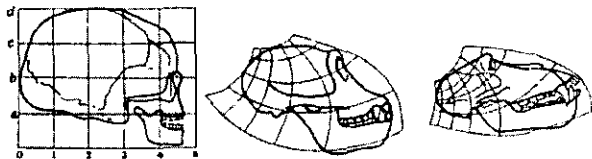


그림 2. 영상좌표변환과 캐리커처[4]

수작업 대신 컴퓨터를 이용한 자동화된 캐리커처 생성에 관한 연구는 1982년 Susan Elise Brennan의 '캐리커처 생성기(Caricature Generator)'이다. 그 후 Benson과 Perrett는 영상왜곡(Image Warping)을 통해 같은 연구를 하였다. 캐리커처는 가상의 평균얼굴과 비교하여 차이를 과장하고 불필요한 부분을 제거하여, 원 얼굴의 특징을 더욱 강조하므로 인식이 용이하다. 인식에 관한 기존 연구결과는 인간의 시각인식은 특징점이나 텍스처 또는 구조 정보를 이용한 것이 아님을 밝혀준다.

그렇기 때문에 본 연구는 인간의 인식방법을 모사한 전체적 이론(Holistic Theory)에 따른 인식을 구현하기 위해 하우스도르프 거리 매칭(Hausdorff Distance

Matching, 이하 HDM)을 적용하고, 캐리커처 효과를 이용하여 인식률을 보완하였다.

### 2.3 영상왜곡

영상왜곡은 특정 규칙에 따라 입력영상을 변환하는 것으로서, 고무종이 위에 그림을 그리듯이 여러 방향으로 영상을 확장할 수 있다. 여러 가지 영상왜곡 방법 중, 원근변환(Perspective Transformation)을 사용, 캐리커처 효과를 주어 표정을 강조하고 인식을 용이하게 하였다. 원근변환은 어파인(Affine) 변환의 모집합이며, 정방향 정합 함수식은 다음과 같다.[2]

$$x = \frac{a_{11}u + a_{21}v + a_{31}}{a_{13}u + a_{23}v + a_{33}} \quad y = \frac{a_{12}u + a_{22}v + a_{32}}{a_{13}u + a_{23}v + a_{33}} \quad (1)$$

계수의 값은 그 정합이 어파인(Affine)인가 아닌가에 따라 두 가지 다른 형식을 가져올 수 있다. 첫 번째 수행해야 할 작업은 여섯 개의 중간 값들을 계산하기 위한 것이다.

$$\Delta x_1 = x_1 - x_2 \quad \Delta x_2 = x_3 - x_2 \quad \Delta x_3 = x_0 - x_1 + x_2 - x_3 \quad (2)$$

$$\Delta y_1 = y_1 - y_2 \quad \Delta y_2 = y_3 - y_2 \quad \Delta y_3 = y_0 - y_1 + y_2 - y_3$$

만일  $\Delta x_3=0$ 이고,  $\Delta y_3=0$ 이면 xy 사변형은 평행 사변형이며, 그 정합은 어파인이다. 그 계수들은 다음 식에 의해 계산된다.

$$\begin{aligned} a_{11} &= x_1 - x_0 & a_{21} &= x_2 - x_1 & a_{31} &= x_0 \\ a_{12} &= y_1 - y_0 & a_{22} &= y_2 - y_1 & a_{32} &= y_0 \\ a_{13} &= 0 & a_{23} &= 0 & a_{33} &= 1 \end{aligned} \quad (3)$$

만일 정합이 어파인이 아니라면, 계수들은 다음 식에 의해 계산된다.

$$\begin{aligned} a_{11} &= x_1 - x_0 + a_{13}x_1 & a_{21} &= x_3 - x_0 + a_{23}x_3 & a_{31} &= x_0 \\ a_{12} &= y_1 - y_0 + a_{13}y_1 & a_{22} &= y_3 - y_0 + a_{23}y_3 & a_{32} &= y_0 \\ a_{13} &= (\Delta x_3 \Delta y_2 - \Delta x_2 \Delta y_3) / (\Delta x_1 \Delta y_2 - \Delta y_1 \Delta x_2) & & & & \\ a_{23} &= (\Delta x_1 \Delta y_2 - \Delta y_1 \Delta x_3) / (\Delta x_1 \Delta y_2 - \Delta y_1 \Delta x_2) & & & & \\ a_{33} &= 1 & & & & \end{aligned} \quad (4)$$

다음 그림은 원근변환의 예이다.



그림 3. 원근변환의 예

### 3. 실험

#### 3.1 실험개요

실험은 Yale 대학의 Yale Face Database(320×243, 8bit Grey Level)에서 90개 영상을 이용하였다. 먼저 얼굴표정정보를 유지하면서 처리할 데이터 양의 축소를 위해 Canny 에지 검출기를 이용하여 얼굴표정영상을 라인드로잉영상으로 바꾼다. 이렇게 만들어진 라인드로잉영상에 대해 하우스도르프 거리 매칭을 적용하여 1차로 표정 인식을 하고, 다시 얼굴특징을 강조하여 인식이 용이하도록 영상왜곡을 통해서 캐리커처 효과를 적

용한 영상을 이용하여 위의 과정을 반복하여 표정인식을 수행한다. 이는 표정간의 차이를 더 크게 하여 인식률을 높이기 위함이다.

### 3.2 라인 드로잉 영상과 기준 영상 생성

하우스도르프 거리 매칭을 위해, 캐니 에지 검출기로 라인 드로잉 영상을 생성하고,[3] 표정 인식에 오류를 일으키는 머리카락, 의복 등을 제거한다. 또한 적절한 임계값을 적용, 0과 255의 값만을 남겨 계산량을 감소시킨다.

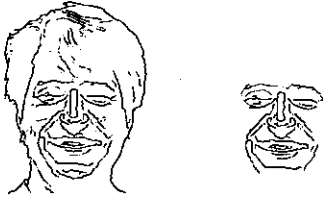


그림 4. 라인 드로잉 영상 생성

시험영상과 비교하여 표정을 결정하는 기준영상은 위와 같이 생성된 영상들을 병합한 후 형태연산자를 적용, 불필요한 부분을 제거하고 평균얼굴을 만들어 사용한다.

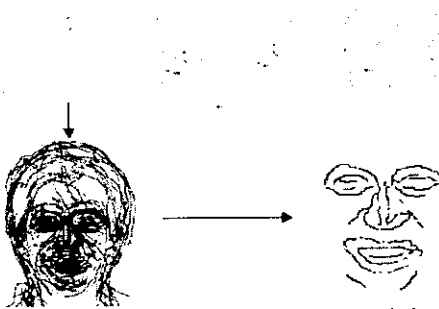


그림 5. 평균얼굴(Average Face) 생성

### 3.3 원근변환을 통한 캐리커처 효과

원근변환은 두 가지 방법으로 적용되었다. 첫째, 머리를 원점, 턱을 근점으로 한 원근변환. 둘째, 얼굴영상을 코를 중심으로 수평으로 분리, 중심으로부터 머리 방향과 턱 방향으로 각각 원근변환을 적용하고, 이를 다시 병합하였다. 왜냐하면 표정인식에 큰 영향을 미치는 요소가 입과 눈이므로, 각각을 따로 강조하기 위함이다. 첫째 방법으로만 원근변환을 할 경우, 머리 부분은 상대적으로 축소되고 턱 부분은 확대되므로, 입 모양만 강조되나, 둘째 방법은 눈과 입 각각을 강조한다.

### 3.4 실험결과

이상의 방법으로 실험을 수행한 결과는 다음 표와 같다. 1차 실험은 캐리커처 효과 없이 HDM을 적용했고, 2차 실험은 캐리커처 효과를 위해 머리에서 턱 방향으로 원근변환, 3차 실험은 눈과 입의 표정을 강조하기 위하여 얼굴영상을 수평으로 나누어 각각 원근변환하여 적용했다.

\*단위 : % (인식성공 개수 / 표정영상 15개)

	행복	슬픔	즐림	놀람	엇크	무표정	평균
실험1	66.7	60	53.3	60	46.7	70	58.9
실험2	86.7	80	60	73.3	60	73.3	72.22
*	29.9	33.3	12.6	22.2	28.5	4.7	22.6
실험3	93.3	80	66.7	86.7	73.3	86.7	81.12
**	7.6	0	11.1	18.3	22.2	18.3	12.3

표 1. 인식률

인식률은 표정이 가장 크게 나타나는 ‘행복’, ‘놀람’, ‘슬픔’의 순으로 나타났다. 라인 드로잉 영상에서 형태가 비슷한 ‘엇크’, ‘즐림’, ‘무표정’은 인식률이 낮다. ‘즐림’은 ‘무표정’으로, ‘엇크’는 ‘행복’으로 자주 오인되었다. 1차, 2차 실험에서 인식률의 증가율은 최저 4.7%, 최고 33.3%로 평균 22.6%, 2차, 3차 실험에서 인식률의 증가율은 평균 12.3%였다. 또 원 영상이 크기 320×243 픽셀과 화소 수 77,760인데 비해, 실험에서는 크기 110×110 픽셀, 화소 수 12,100으로 데이터양이 축소되었고, 이 실험영상에서 밝기 값 0으로써 선을 구성하는 실제 좌표 값은 대략 1,200 여 개였다. 그런데 선을 구성하는 1200 여 개의 점들은 서로 인접하므로 절반만 계산에 사용해도 전체 형태를 왜곡시키지 않으므로, 600 여 개의 화소만을 계산에 사용함으로써 데이터량을 7%로 감소시켰다.

### 4. 결론

본 연구는 사람의 인식방법을 모사한 전체이론에 따라 윤곽선을 추출하여 영상을 재구성하고, 전체적 유사도를 측정하는 하우스도르프 거리 매칭을 적용하여 인식을 수행하고, 이때 인식률 보안을 위해 얼굴표정영상에 영상워핑을 적용하여 캐리커처 효과를 주었다. 위의 결과는 이상의 방법으로 사전정보 획득과정 없이, 직접인식이 가능하며 데이터량과 계산량을 크게 줄일 수 있음을 보여준다. 그러나 기준영상, 세밀한 캐리커처 생성방법에 대한 추가연구와 얼굴영상 데이터베이스 표준규격 제정이 필요하다.

### 5. 참고 문헌

- [1] Shaogang Gong, J Mckenna, Alexandra Psarrou, "Dynamic Vision from Images to Face Recognition", Imperial College Press, pp.3-30, 81-84, 2000.
- [2] Randy Crane, "Simplified Approach to Image Processing", Prentice Hall PTR, 1997.
- [3] J.R.Parker, "Algorithms for Image Processing and Computer Vision", WILEY Computer Publishing, 1997.
- [4] Susan Elise Brennan, "Caricature Generator", Master's Thesis at M.I.T, 1982.
- [5] Barnabas Takacs, "Comparing Face Images using The Modified Hausdorff Distance", Pattern Recognition, Vol. 31, No. 12, pp.1873-1881, 1998.
- [6] Normand Grégoire, Mikael Bouillot, <http://jeff.cs.mcgill.ca/~godfried/teaching/cg-rojects/98/normand/main.html>
- [7] Jain, Halici, Hayachi, Lee, Tsutsui, "Intelligent Biometric Techniques in Finger Print and Face Recognition", CRC Press, pp.217-283,355-396,1999.