

Matched Filter를 이용한 얼굴 특징점 위치추출

황인택[○], 최광남[○]
중앙대학교 컴퓨터공학부
{sitdown[○], knchoi}@cau.ac.kr

Estimating Facial Feature Position with Matched Filters

In Teck Whang[○], Kwang Nam Choi[○]
School of Computer Science & Engineering, Chung-Ang University
요 약

이 논문은 Matched Filter 기술을 사용해 얼굴 특징점 위치를 추출하는 연구에 대해서 기술한다. 기본 목표는 얼굴의 서로 다른 8개(양쪽 눈과 눈썹, 머리선, 코, 입, 턱)의 부분을 구분할 수 있는 필터들을 개발하는 것이다. 이런 Matched Filter는 Fourier 역변환을 사용해 훈련영상(Training Image)으로부터 얻을 수 있다. 실험평가는 베른대학의 얼굴 데이터베이스에 근거한다. 우리는 여기서 다양한 얼굴의 방향성에 효과적으로 적용할 수 있도록 하는 훈련 영상 자료가 무엇인지 알 수 있다. 그리고 안경을 썼을 때 얼굴을 인식할 수 있는 가장 좋은 방법도 알아본다.

1. 소개

대체적으로, 얼굴 인식은 템플릿(Template)기반 혹은 특징(Feature)기반 방법을 사용하고 있다. 템플릿기반 방법에서, 얼굴을 여러 조각으로 나눈 것의 상호위치와 그 면에서 얼굴식별정보를 얻는다. 반면 특징기반 방법은 영상명암에서 식별정보의 구조를 정의한다. 템플릿기반 방법을 사용한 사람은 얼굴모습의 다양함을 설명할 수 있는 직접 수정 가능한 Shape Model을 개발한 Yuille, Hallinan과 Cohen이다 [1]. 이것은 Cootes, Taylor, Cooper 그리고 Graham에 의해서 점을 이용한 분산모델로서 Parametric Shape Template들의 사용을 포함으로서 더 발전시켰다 [2]. 손으로 한 것보다 Shape Model은 Training 집합의 공분산구조를 통해 얻을 수 있다. 특징 기반 인식방법은 Turk와 Pentland의 Eigenface 아이디어와 Rao와 Ballard의 Iconic 표현을 예로 들 수 있다 [3][4]. Turk과 Pentland의 아이디어는 특정한 얼굴 영상에서 얼굴 영상을 나눌 수 있는 중요한 컴포넌트들을 적용시키는 것이다. Rao와 Ballard는 무표정한 얼굴을 인식하기 위해 고정된 자연수 기반함수들을 사용했다 [5]. 이 함수들은 다양한 크기의 인간얼굴의 Iconic 표현을 만드는 것에 사용된다.

두가지 방법들 모두 유용한 결과를 얻었지만, Brunelli와 Poggio는 얼굴인식의 가장 좋은 성능은 두 가지의 방법을 조합해서 사용하는 것임을 알아내었다. 기초모델은 눈, 코, 입, 턱 등의 구분되는 특징의 파라미터와 상호위치를 이용했다. Craw 등은 Shape-Normalisation의 결함을 없앴으로서 이 연구를 완성시켰다. 이 연구는 얼굴의 중요한 부분의 분석으로 얼굴의 특징을 뽑고, 템플릿과 대

응되는 면과의 유효한 통계적인 거리사용에 있다.

이 조합방법론은 얼굴의 용기부분으로 얼굴의 특징을 분석하는 방법 대신, 템플릿 레벨에서 그 곳의 구조를 이용한 것을 볼 수 있다. 이 모델은 영상그래프를 사용한 Von der Malsburg 등에 의해 명시적으로 채택되었다. 특징은 'Jet'이라 명명된 고차원의 Feature-Vector로 구성된 것이다. 기본 인식 체계는 Bunch-Graph이다. 이것은 훈련영상 한개의 샘플에서 'Jet'의 평균값을 속성으로 가지고 특징점을 노드(node)로 하는 그래프들의 집합이다. 모델과 데이터그래프 사이에 매칭에너지(matching energy)를 최소화함으로써 인식을 촉진시킨다.

비록 특징기반과 템플릿기반 방법의 조합 개념이 우수해 보이지만, 본질적으로 절차적으로 이루어진다는 한계성이 지적되었다. 최근의 논문들은 그래프의 일치과정(Matching Process)과 얼굴구분, 양쪽의 통계적인 모델을 개발하고 있다. 또한, Graph-Matching[6]의 관련속성 처리로 얼굴 인식을 실현하기 위한 새로운 방법의 이용에 초점을 맞추어 연구하고 있다. 이 논문은 이 절차의 첫 번째 단계를 기술하고, Matched-Filter를 이용한 얼굴 특징의 구분 및 위치파악에 대해서 다룬다.

이 논문의 목적은 첫 번째로 그들의 그래프구조를 가질 수 있는 클래스(눈, 코, 턱 등)에 따른 일반적인 얼굴 특징의 위치파악에 사용되는 필터를 개발하는 것이다. 두 번째로 시선방향에 따른 각각의 모습에서 구분할 수 있는 필터를 개발하는 것이다.

2. 필터 설계

훈련영상에 대응되는 회선필터를 이끌어 내기 위해 영상공간에서의 회선과 Fourier공간에서의 곱 사이의 이중

성을 이용한다. High-convolution 반응을 나타내기 원하는 지역을 손으로 구분지어 놓은 영상의 집합으로부터 시작한다. 실험에서 특징부분은 하이비트(high-bit)로 나머지 배경은 로우비트(low-bit)로 표시한다. 훈련데이터의 형태는 경계부분이 고주파요소와 연관되어져 있어서 Fourier 영역에 적당하지 않다. 이 문제의 해결방법으로 가우시안 스무딩커널(Gaussian Smoothing Kernel)을 이용해서 흐리게 만들어준다. 다시 말해, 훈련영상도 저주파수 통과필터로 전처리를 수행한다. 블러링 크기는 특징들의 크기를 고려하여 같은 크기의 순서로 한다.

만약 I_n 이 원영상이고, O_n 이 블러링된 영상을 나타낸다면, 우리는 훈련집합에서 입출력 영상과 가장 잘 매치되는 회선필터 C_n 을 찾을 수 있다. 다시 말해, '*'를 공간영역에서의 회선 연산 표시로 하고, 다음 식에 대해 필터를 찾을 수 있다.

$$O_n = I_n * C_n \quad (1)$$

부등식의 항등성질을 이용해, 식의 양쪽에 Fourier 변환을 한다.

$$F(O_n) = F(I_n) \times F(C_n) \quad (2)$$

회선필터 C_n 은 다음에 따르는 방법으로 역Fourier변환을 함으로서 얻을 수 있다.

$$C_n = F^{-1}\left(\frac{F(O_n)}{F(I_n)}\right) \quad (3)$$

이 연구는 빠른 대학의 얼굴 데이터베이스에 기반한다. 이 데이터베이스는 20개의 다른 실험대상의 얼굴에 다섯가지 포즈를 가지고 있다. 정면을 바라볼 때, 위, 아래, 그리고 왼쪽과 오른쪽의 다섯가지 포즈가 있다. 사람은 유럽인과 동양인 그리고 안경의 착용유무로 혼합되어 있다. 우리는 여덟가지 두드러진 얼굴의 특징(양 눈과 머리, 양 눈썹, 코, 입, 턱)을 인식하는 필터를 개발하는 것을 목표로 한다. 사람에 따라 인종별로 나누고, 안경착용에 따라 나누어서 네 가지의 클래스로 세분화 할 수 있다. 또한 우리는 포즈에 따라서 다섯 가지로 다시 구분할 수 있다. 클래스에서 나누고 난 후, 훈련과 평가집합의 데이터로 나눌 수 있다. 포즈의 경우, 훈련영상 중에서 어느 시선의 영상이 가장 효과적인 지를 연구해왔다. 훈련데이터의 각각의 영상에 대해 우리는 직접 8군데의 특징을 손으로 표시한다. 얼굴의 여덟 개의 특징 위치 지정 필터는 훈련집합에서 각 주제(subject)에 대해 다섯가지 포즈 전체의 평균이다.

얼굴특징추출에 대한 필터는 손으로 한번 하고서, 그 다음 작업은 그와 비슷한 특징의 위치를 파악하는 것이다. 우리는 두가지 단계절차로 만들 수 있다. 순서대로 각 영상은 여덟 개의 특징을 특성화시킬 수 있는 필터로

분리되어진 영상이다. 그 특징들은 각 필터에 의해 가장 크게 반응하는 위치에 표시되어진다. 이 절차는 찾고자 하는 특징과 같은 크기를 가진 가우시안 필터로서 미리 블러링을 함으로서 안정되게 할 수 있다.

3. 실험과 평가

우리 실험의 목적은 두 가지이다. 첫 번째로 Matched Filter 기술이 여러 가지 모습과 포즈를 가진 하나의 샘플에 관해 전체적으로 효과적인 것을 보여주는 것이다. 두 번째 목적은 훈련영상의 선택, 인종그룹 별 그리고 안경에 따라서 특징타입 별로 특징을 추출해내는 정도를 평가하는 것이다.

얼굴데이터베이스의 특징위치선정 절차의 예를 설명하기 위해, [그림 1]은 각각 다른 얼굴 특징에 대해 여덟 개의 필터로 계산되어진 영상을 보여준다. 표시된 점은 훈련데이터의 특징의 중앙을 나타낸다. 훈련집합은 정면, 왼쪽, 약간 아래를 보는 모습의 유럽인과 아시아인을 포함하고, 안경을 쓴 모습도 포함한다. 특징 위치선정 필터는 다른 포즈와 다른 주제에 관한 전체의 평균이다. [그림 2] 처럼 더 많은 양의 종류로 입증되어진다. 시점이 다른 훈련, 평가영상의 조합에 따른 결과를 볼 수 있다. 이것으로 정면과 오른쪽, 그리고 약간 아래쪽 영상에서 가장 적은 에러를 낸 것을 확실히 알 수 있다. 또 훈련영상의 수에 따라 에러가 감소하는 것을 알 수 있다.

안경착용에 따른 인식정도도 연구하였다. [그림 3]은 훈련영상에서 안경착용 예를 포함했을 때의 결과를 비교해주고 있다. 안경착용과 미착용의 각각의 영상들은 유럽인과 아시아인으로 구성된 평가 영상에서 부가적 평가 자료로 사용된다. 그 그림에서 '+'점은 훈련영상에서 안경을 착용했을때의 결과를 보여주고, 'o'점은 미착용시의 결과를 보여준다. 'x' 점은 두가지를 혼합했을 때의 결과를 보여준다. 두가지의 경우를 혼합했을 때 가장 좋은 필터를 얻을 수 있다. 안경 미착용시에는 대칭축을 가지는 코, 입술, 턱 등의 위치선정시에 더 좋은결과를 얻을 수 있다. [그림 2]와 [그림 3]을 비교하면, 안경착용 영상의 조합이 훈련포즈중에서 에러율이 높은 것을 선택하는 것보다 에러가 적다는 것을 알 수 있다. [그림 4]는 안경착용유무의 혼합된 집합영상을 이용한 필터로 얻은 영상의 결과를 보여주는 [그림 3]의 그래프에서 사용되어진 몇가지 자세한 예를 보여준다.

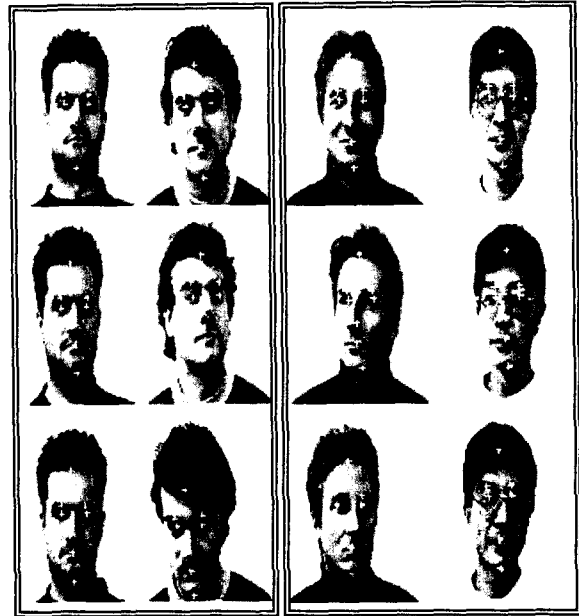
4. 결론

이 논문에서는 얼굴 특징점을 찾기 위한 방법으로 Matched Filter의 기술을 사용하였다. 얼굴의 방향과 사

람의 수에 따른 특징화된 필터를 살펴보았으며, 얼굴의 다양한 방향에 따라 그에 맞는 효과적인 Matched Filter를 적용시킬 수 있었다. 훈련과 평가를 위한 얼굴 영상의 집합은 안경 착용여부와 동·서양인을 혼합하여 선택하였다. 스무명의 사람에 대해 각각 다섯가지의 다른 시선방향을 가진 데이터 베이스로 우리는 훈련영상에 정면 방향의 영상이 포함될 때, 작은 오류를 낸다는 결론을 얻을 수 있었다.

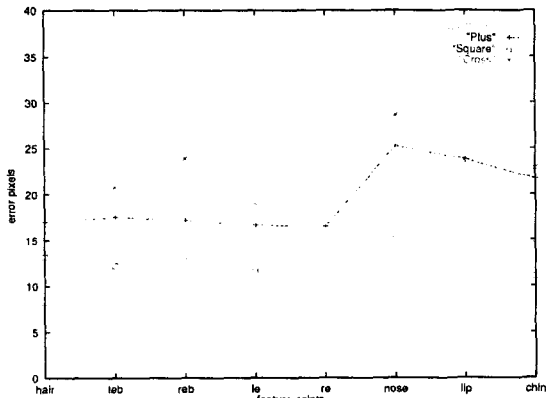
참 고

[1] A.L. Yuille, D.S. Cohen, and P.W. Hallinan. Feature extraction from faces using deformable templates. *Int. J. Comput. Vision*, 8:99-112, 1992.
 [2] Cootes, T., Taylor, C., Cooper, D., and Graham, J. . *Active shape modelstheir training and application. Computer Vision and Image Understanding*, 61:38-59. 1995.
 [3] Turk, M. and Pentland, A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:7186. 1991.
 [4] Rao, R. and Ballard, D. An active vision architecture based on iconic representations. *Artificial Intelligence Journal (Special Issue on Vision)*, pages 461-505. 1995.
 [5] Rao, R. and Ballard, D. Natural basis functions and topographic memory for face recognition. Proc. of the International Joint Conference on Artificial Intelligence (IJCAI), pages 10-17. 1995.
 [6] R.C. Wilson and E.R. Hancock. Gauging relational consistency and correcting structural errors. *IEEE Computer Vision and Pattern Recognition Conference*, pages 47-54, 1996.

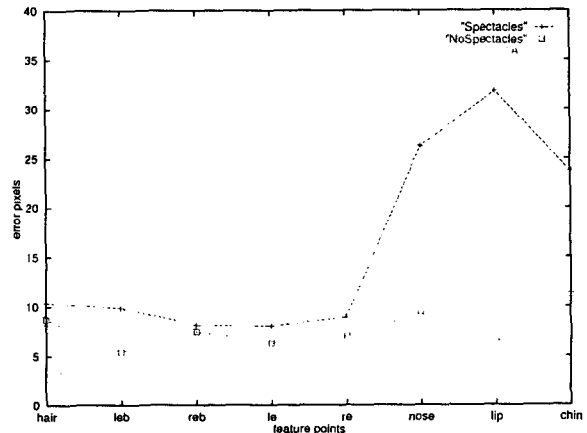


[그림 1] 훈련영상.

[그림 4] 혼합된 훈련영상을 이용한 필터로 찾은 얼굴 특징점.



[그림 2] 훈련영상의 시선방향에 따른 위치선정의 오차효과;
 '◇'(전방, 오른쪽, 아래),
 '+' (왼쪽, 위), '□'(전방, 왼쪽, 위),
 'x'(오른쪽, 아래)



[그림 3] 안경착용에 따른 오차의 효과;
 '+'점은 훈련영상에 안경 쓴 영상을 포함했을 경우,
 '□'점은 안경 쓴 영상을 제외했을 경우.
 '◇'점은 두가지를 혼합한 훈련영상을 사용했을 경우.