

Matched Filter를 이용한 얼굴 영상 검출

황인택⁰, 신영숙^{*}, 최광남

중앙대학교 컴퓨터공학부

*조선대학교 전자정보통신공학부

Facial Image Detection with Matched Filters

In Teck WhangO, Young-Suk Shin*, Kwang Nam Choi

School of Computer Science & Engineering, Chung-Ang University

*Division of Electronics & Information Communication Engineering, Chosun University

요 약

이 논문은 Matched Filter 기술을 사용해 다양한 영상 중 얼굴 영상을 분류하는 연구에 대해서 기술한다. 기본 목표는 얼굴 영상과 그 외의 영상을 분류할 수 있는 효과적인 필터를 개발하는 것이다. 이런 Matched Filter는 Fourier 역변환을 사용해 훈련영상(Training Image)으로부터 얻을 수 있다. 실험평가는 Yale대학의 얼굴 데이터베이스의 얼굴영상과 다양한 형태를 보이는 임의의 영상으로 평가한다. 우리는 여기서 얼굴 영상을 분류하기 위한 방법 중 한 가지로서 Matched Filter를 이용할 수 있음을 확인할 수 있다.

1. 소 개

얼굴검출(Face Detection)에서의 초기 연구들은 간단한 경험적(Heuristic) 및 인체 측정학적(Anthropometric) 기법들이 사용되었던 1970년대의 초로 되돌아간다 [1]. 이들 기법들은 단순한 배경의 정면 얼굴과 같은 가정들 때문에 상당히 유통성이 없었다. 이 후 1990년대까지 정체된 채로 있다가 과거 십년동안 얼굴 검출을 확장시키는 많은 연구가 이루어져 왔다. 특히, 움직임(motion), 컬러(color) 그리고 일반화된 정보(generalized information)를 사용하는 더 강인한 분할체제들(segmentation schemes)이 제시되어 왔다 [2][3][4][5]. 또한, 통계학적인 방법 및 신경망의 사용은 카메라로부터 다른 거리들에서 혼잡한 장면들로부터 얼굴들을 검출하는 것을 가능하게 하였다 [6][7].

또한, 얼굴 인식에서는 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 [8]. 이것은 Cootes, Taylor, Cooper 그리고 Graham에 의해서 점을 이용한 분산모델로서 Parametric Shape Template들의 사용을 포함으로서 더 발전시켰다 [9]. 손으로 한 것보다 Shape Model은 Training 집합의 공분산구조를 통해 얻을 수 있다. 특정 기반 인식방법은 Turk와 Pentland의 Eigenface 아이디어와 Rao와 Ballard의 Iconic 표현을 예로 들 수 있다 [10][11]. Turk와 Pentland의 아이디어는 특정한 얼굴 영상에서 얼굴 영상을 나눌 수 있는 중요한 컴포넌트들을 적용시키는 것이다. Rao와 Ballard는 무표정한 얼굴을 인식하기 위해 고정된 자연수 기반함수들을 사용했다 [12]. 이 함수들은 다양한 크기의 인간얼굴의 Iconic 표현을 만드는 것에 사용된다.

두 가지 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 없앰으로서 이 연구를 완성시켰다. 이 연구는 얼굴의 중요한 부분의 분석으로 얼굴의 특징을 뽑고, 템플릿과 대

용되는 면과의 유효한 통계적인 거리사용에 있다.

이 조합방법론은 얼굴의 음기부분으로 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 채택되었고, 특징은 'Jet'이라 명명된 고차원의 Feature-Vector로 구성한 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 최소화함으로서 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-Matching의 관련속성 처리로 얼굴 인식을 실현하기 위한 새로운 방법의 이용에 초점을 맞추어 연구하고 있다 [13]. 이 외에 얼굴 방향성 인식에 관한 연구도 활발히 진행되고 있다 [14][15].

이 논문은 Matched-Filter를 이용한 영상분류 시스템의 첫 번째 단계인 필터를 설계하고, 이것을 이용한 얼굴 영상 분류에 대해서 기술한다. Matched Filter 기술을 이용해 다양한 영상을 중에서 원하는 얼굴 영상을 추출해내는 것이 목표다. 그리고, Matched Filter를 이용한 영상분류 시스템 구현의 가능성과 그 효율성을 타진해 본다.

.

2. 필터 설계

훈련영상에 대응되는 회선필터를 이끌어 내기 위해 영상공간에서의 회선과 Fourier공간에서의 곱 사이의 이중성을 이용한다. High-convolution 반응을 나타내기 원하는 지역을 손으로 구분지어 놓은 영상의 집합으로부터 시작한다. 실험에서 특정부분은 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 필터

를 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.

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

(3)

이 연구는 Yale대학의 얼굴 데이터베이스에 기반한다. 이 데이터베이스는 실험대상의 얼굴에 각각 조명의 변화와 시선의 방향에 따른 영상이 있다. 얼굴은 시범적으로 정면을 바라보는 얼굴 10개를 선정하였으며, 이것을 기반으로 얼굴을 찾아내는 필터의 개발을 목표로 한다. 영상은 훈련과 평가집합의 데이터로 나눌 수 있다. 훈련집합에서는 얼굴영상 10개를 대상으로 하며, 평가집합에는 이외에 50개의 얼굴영상과 임의의 비얼굴 영상 50개를 포함한다.

얼굴 분류를 위한 필터는 손으로 한번 표시하고서, 그 다음 작업은 필터를 통한 결과 영상의 반응값을 측정하는 것이다. 이를 위해 분류하고자 하는 얼굴 영상의 평균영상을 구하고, 그 평균영상의 공통점을 분류를 위한 필터에 의해 가장 크게 반응하는 위치에 표시되어진다. 이 절차는 찾고자하는 특징과 같은 크기를 가진 가우시안 필터로서 미리 블러링을 함으로서 안정되게 할 수 있다.

3. 실험과 평가

이번 실험의 목적은 Matched Filter 기술이 다양한 영상들 중에서 원하는 특정 영상을 분류할 수 있는지를 검증해서 영상분류 시스템의 구현이 가능한지를 타진해 보고 그 효율성이 어느 정도인지를 평가해 보는 것이다.

얼굴 영상 분류 필터를 만들기 위한 훈련영상을 [그림 1]에서 보여주고 있다. 전처리를 마친 이들의 평균 영상을 구하고, 평균영상에서 분류하고자 하는 얼굴 영상의 특징을 가장 잘 표현해 주는 부분을 선정하여 필터를 추출해 낸다. 이번 실험에서는 경험적인 과정을 거쳐 코와 입술의 사이를 중심으로 하는 두 부분을 포함하는 영역으로 선정하였다. 그 예를 [그림 2]에서 보여주고 있다. 추출한 필터는 훈련영상을 포함하는 평가집합을 대상으로 적용시키게 된다. 평가집합의 예를 [그림 3]에서 보여준다.

다음으로 평가집합의 영상을 대상으로 추출한 Matched

Filter를 적용시켜 영상 분류를 위한 데이터를 추출하고, 이것을 Matched Filter 추출단계에서 지정해 놓은 윈도우와 회선(Convolution)을 시키며 Bhattacharyya Distance를 계산하고, 이를 값 중 최대값을 영상 분류를 위한 평가값으로 사용한다.

확률밀도함수 $f_A(x)$ 와 $f_B(x)$ 사이의 Bhattacharyya Distance의 수식은 다음과 같다.

$$B = \int \sqrt{f_A(x)} \sqrt{f_B(x)} dx$$

평가집합 전체에 관하여 위와 같은 단계를 거쳐서 평가값을 모두 획득 후, FAR(False Acceptance Rate)과 FRR(False Reject Rate) 그래프를 이용해 적절한 임계값(Threshold)과 Matched Filter를 이용한 영상분류 시스템의 효율성을 측정한다.

[그림 4]는 위의 단계를 실행한 결과를 보여준다. x축의 임계값은 Bhattacharyya Distance 수식에서 $f_A(x)=f_B(x)$ 인 경우를 100%로 가정할 경우의 퍼센티지를 보여주고 있으며, y축은 그에 따른 비율을 보여준다. [그림 4]에서 보는 바와 같이 Matched Filter를 통한 적절한 임계값은 80이며, FAR, FRR의 최저치가 12%임을 알 수 있다.

4. 결 론

이 논문에서는 영상을 분류하기 위한 다양한 방법 중 한가지로서 Matched Filter 기술을 사용하였다. 이 실험의 결과로 다양한 영상 가운데 우리가 원하는 특정 영상들을 90%에 가까운 분류가 가능하다는 것을 확인했다. 이로서 영상분류를 위한 방법 중 한가지로서 Matched Filter기술을 적용할 수 있음을 알 수 있다.

이번 실험에서는 훈련영상과 평가영상의 개수가 상당히 제한적이었다는 점에서 조금 더 발전 시킬 필요성이 있으며, 향후 얼굴영상의 크기와 방향에 좀 더 견고한 필터를 개발할 필요가 있다.



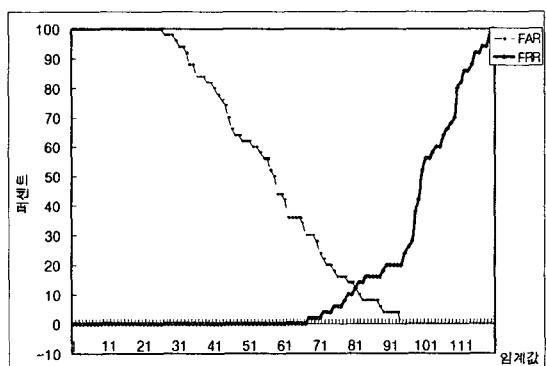
[그림 1] 훈련 영상들(Training Images).



[그림 2] 평균영상과 Matched Filter를 통한 영상.



[그림 3] 얼굴 영상과 비얼굴 영상이 혼합된 평가 영상의 예.



[그림 4] 평가집합의 영상 분류 결과.

[참고문헌]

- [1] T. Sakai, M. Nagao, and T. Kanade, Computer Analysis and Classification of Photographs of Human Faces, in Proc. First USA-Japan Computer Conference, pages 3-7. 1972.
- [2] F. Lethon and M. Lievin, Lip Motion Automatic Detection, in *Scandinavian Conference on Image Analysis, Lappeenranta*, Finland, 1997.
- [3] J. L. Crowley and F. Berard, Multi-model tracking of faces for video communications, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition, Puerto Rico*, Jun. 1997.
- [4] L. C. De Silva, K. Aizawa, and M. Hatori, Detection and tracking of facial features by using a facial feature model and deformable circular template, *IEICE Trans. Inform. Systems*E78-D(9), pages 1195-1207. 1995.
- [5] D. Reisfeld, H. Wolfson, and Y. Yeshurun, Context-free attentional operators: The generalized symmetry transform, *Int. J. Comput. Vision*14, pages 119-130. 1995.
- [6] A. J. Colmenarez and T. S. Huang, Face detection with information-based maximum discrimination, in IEEE Proc. of Int. Conf. on *Computer Vision and Pattern Recognition*, 6 1997.
- [7] H. A. Rowley, S. Baluja, and T. Kanade, Neural network-based face detection, *IEEE Trans. Pattern Anal. Mach. Intell.*20, pages 23-38. January 1998.
- [8] 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.
- [9] Cootes, T., Taylor, C., Cooper, D., and Graham, J. . *Active shape models: their training and application. Computer Vision and Image Understanding*, 61:38-59. 1995.
- [10] Turk, M. and Pentland, A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:7186. 1991.
- [11] Rao, R. and Ballard, D. An active vision architecture based on iconic representations. *Artificial Intelligence Journal (Special Issue on Vision)*, pages 461-505. 1995.
- [12] 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.
- [13] 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.
- [14] Kwang Nam. Choi, P. Worthington and E.R. Hancock. *Estimating Facial Pose using Shape-from-Shading*. In Pattern Recognition Letters, pages 533-548. 2002.
- [15] Kwang Nam. Choi, M. Carcassoni, and E.R. Hancock, *Recovering Facial Pose using the EM Algorithm*. In Pattern Recognition, pages 73-93. 2002.