

Gabor 특징과 FloatBoost 학습을 이용한 고속 표정인식

정광희, 민부천, 김중규
성균관대학교 전자전기컴퓨터공학과
e-mail : daedooya@skku.edu

Fast Facial Expression Recognition Using Gabor Feature and FloatBoost Learning Algorithm

Kwang-Hee Jung, Bu-Cheon Min, Joong-Kyu Kim
Dept of Electrical and Computer Engineering, Sung Kyun Kwan University

요 약

본 논문은 Gabor 특징과 FloatBoost 학습을 이용한 효과적인 표정 인식 방법을 제안한다. 제안된 방법은 FloatBoost 알고리즘에 의해 각 표정 단위로 최적의 Gabor 특징을 학습하고 입력 영상으로부터 학습된 각 표정에 대한 Gabor 특징들의 조합을 비교하여 표정을 효과적으로 인식한다. 실험을 통해 제안된 방법의 성능 및 효율성의 우수함을 확인 할 수 있다.

1. 서론

표정인식은 최근 얼굴 인식과 더불어 많은 관심을 받고 있는 생체 인식 기술 중의 하나이다. 얼굴 표정은 사람의 내적상태를 반영하기 때문에 사회 심리학에서는 표정이 인간들의 의사소통에 있어 가장 중요한 모달리티라는 점을 나타내고 있다. 때문에 컴퓨터가 사람의 표정을 인식하여 사람의 현재 심리상태에 맞는 인간중심의 컴퓨팅 환경을 제공 하는 것이 가능하다면 더 편리한 생활의 영유가 가능해 질 것이다. 따라서 본 논문에서는 AdaBoost 알고리즘의 문제점을 개선한 FloatBoost 알고리즘을 이용하여 표정인식에 적합한 특징들을 학습하고 학습된 특징들을 이용한 표정 인식 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 FloatBoost 알고리즘의 기반이 되는 AdaBoost 알고리즘과 Haar-like 특징의 문제점을 설명한다. 3장과 4장에서는 이러한 문제점을 개선하기 위해 사용되는 Gabor 특징과 FloatBoost 알고리즘에 대하여 설명하고 5장에서는 이를 이용한 인식 과정에 대해 설명한다. 이 후 6장에서 실험 결과를 보여주고 7장에서 결론 및 향후 연구 과제를 제시한다.

2. AdaBoost 알고리즘과 Haar-like 특징

Viola와 Jones에 의해 제안된 AdaBoost 알고리즘은 약한 분류기의 선형 조합으로 강한 분류기를 구성하는 학습 알고리즘으로 간단하지만 효과적인 고차원의 비선형 분류기 학습 방법이다[1]. 보통 AdaBoost 알고리즘은 얼굴 검출에 주로 이용 되는데 이때 사용되는 특징은 누적 영상 표현법(integral image representation)을 이용한 Haar-like 특징이다. Haar-like 특징은 인접한 직사각형 영역들의 픽셀 값들의 합의 차로 단순하면서도 계산하기에 용이하다. 보통 얼굴검출 시스템에서 적용하고 있는 Haar-like 특징들은 모두 같은 모양, 같은 면적을 가진 수평, 수직으로 인접한 직사각형으로 구성되어 있다. 이렇게 계산된 각 특징 값들은 AdaBoost 알고리즘에 효과적으로 적용할 수 있다. AdaBoost 알고리즘은 정해진 라운드동안 반복적으로 약한 학습기를 호출하여 약한 분류기를 선정하고 분류성능에 맞게 가중치를 부여한다. 이렇게 선정된 약한 분류기의 선형 조합으로 강한 분류기를 구성하여 간단하지만 효과적인 얼굴인식을 가능하게 한다.

하지만 기존의 AdaBoost 알고리즘은 훈련영상에서 한계의 지수 함수를 최소화하도록 학습되어지는데

이러한 방법은 이론과 수학적 분석에 편리함을 주는 대신 실제 적용에서는 훈련영상에서 잘못된 분류기가 선정될 경우에는 전체적인 에러 율에 악영향을 미치게 된다. 또한 여기에 사용되는 Haar-like 특징은 단순하면서도 계산이 용이한 장점이 있지만 표정의 특징을 나타내기에는 부적절한 방법이다. 이러한 이유로 본 논문에서는 기존 알고리즘의 한계점을 극복하기 위해 표정의 특징 표현에 적합한 Gabor 특징과 학습시의 에러 율의 최소화를 가능하게 하는 FloatBoost 알고리즘을 이용한 표정인식 방법을 제안한다.

3. Gabor 특징

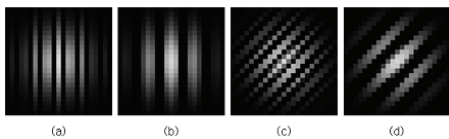
Gabor 필터에 기반을 둔 Gabor 특징은 최근 30년간 영상 처리에서 유용한 여러 가지 특성 때문에 많이 사용되어 왔다. V. Kyrki와 J. K. Kamarainen에 의해 제안된 Gabor 특징은 방향성분, 공간 주파수, 그리고 공간 지역성을 효율적으로 추출한다[2]. Gabor 특징은 인간의 시각 신경 계통 특성에 대응하는 실험적 필터에 대한 접근으로 볼 수 있으며 이러한 생물학적 연관성과 계산적 특성 때문에 영상 인식에 널리 사용되고 있다. Gabor 특징의 가장 주요한 특성중의 하나는 영상의 조명, 회전, 크기 그리고 이동에 민감하지 않다는 것이다. 이러한 특성은 많은 연구를 통해 확인되어 왔는데 특히, Jingfu Ye와 Yongzhao Zhan에 의해 Gabor 특징이 조명의 변화와 개개인의 특성에 민감하지 않다는 것이 실험을 통해 확인되었다[3]. 이러한 특성은 Gabor 특징이 표정인식에 적합한 특징임을 말해 준다.

공간 영역에서 2D Gabor 필터는 정현평면파에 의해 변조된 가우시안 커널로써 다음과 같이 표현된다.

$$g_{\Pi(f,\theta)}(x,y) = w(x,y)s(x,y) = \exp(-(x'^2 + y'^2)) \times \exp(j2\pi f x') \quad (1)$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$

여기서 f 는 정현평면파의 중심 주파수이고, θ 는 가우시안 평면파의 반시계방향 회전각이다.



(그림 1) Gabor 필터 (a) $\Pi(0.4,0)$ (b) $\Pi(0.2,0)$ (c) $\Pi(0.4,\pi/4)$ (d) $\Pi(0.2,\pi/4)$

Gabor 특징은 다음과 같이 Gabor 필터와 입력 영상과의 컨볼루션을 통해 얻어진다.

$$O_{\Pi(f,\theta)}(x,y) = I(x,y) * g_{\Pi(f,\theta)}(x,y) \quad (2)$$

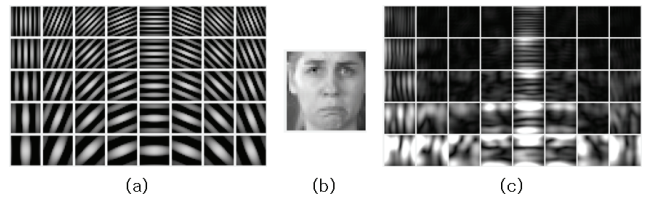
Gabor 필터의 중심 주파수와 회전각의 수는 다양하게 설정 할 수 있지만 본 연구에서는 5개의 중심 주파수와 8개의 회전각으로 구성된 40개의 Gabor 필터가 이용된다.

$$f_u = \frac{f_{max}}{\sqrt{2^u}}, \theta_v = \frac{v}{8}\pi, \text{ where, } u=0..4, v=0..7 \quad (3)$$

이렇게 구성된 40개의 Gabor 필터와 입력 영상의 컨볼루션의 결과 Gabor 특징은 다음과 같이 구성된다.

$$S = \{O_{u,v}(x,y) : u \in \{0, \dots, 4\}, v \in \{0, \dots, 7\}\} \quad (4)$$

여기서 $O_{u,v}(x,y) = |I(x,y) * g_{\Pi(f_u,\theta_v)}(x,y)|$ 이다.



(그림 2) Gabor 필터와 컨볼루션 결과

(a) 40개의 Gabor 필터 (b) 입력 영상

(c) 입력영상에 각 Gabor 필터가 적용된 결과 영상 위의 그림 2는 40개의 Gabor 필터와 입력 얼굴 영상과의 컨볼루션 결과를 보여준다. 36×36 픽셀 크기의 얼굴 영상을 통해 $36 \times 36 \times 5 \times 8 = 51,840$ 개의 Gabor 특징을 얻을 수 있으며 이 특징들은 모두 다음 장에 소개될 FloatBoost 학습 알고리즘의 훈련을 위한 약한 분류기의 집합을 형성한다.

4. FloatBoost 학습

Stan Z. Li와 ZhenQiu Zhang등에 의해 제안된 FloatBoost 알고리즘은 에러 율을 최소화하기 위해 AdaBoost 학습 과정의 각 라운드 이후 backtrack 방법을 통해 직접적으로 에러 율을 최소화하게 된다 [4]. backtrack 방법은 에러 율에 악영향을 주는 약한 분류기를 찾을 수 있도록 해주는데, 이러한 학습 방법을 통해 FloatBoost 알고리즘은 기존의 AdaBoost 알고리즘보다 적은 수의 약한 분류기를 통해 훈련 영상 집합과 테스트 영상 집합에서 더 낮은 에러 율을 얻을 수 있게 해준다.

FloatBoost 알고리즘의 입력이 되는 훈련 영상 집합은 $S_m = \{(x_1, y_1), \dots, (x_N, y_N)\}$ 으로 표현되며 각 x_i 에 대해 $y_i = 1$ 일 경우에 positive 영상을, $y_i = -1$ 일 경우에는 negative 영상을 나타낸다. FloatBoost 알고리즘은 AdaBoost 알고리즘처럼 $M = 1, \dots, M_{max}$ 라운드 동안 반복적으로 약한 학습기를 호출하여 훈

런 영상 집합에 대해서 가장 작은 에러를 가지는 약한 분류기를 선정하고 분류성능에 맞게 가중치(α_m)를 부여한다. 다음 식은 강한분류기와 약한 분류기의 표현이다.

$$H_{M(x)} = \sum_{m=1}^M h_m(x) \quad (5)$$

$$h_m(x) = \begin{cases} \alpha_m & \text{if } p_m f_m(x) < p_m \theta_m \\ -\alpha_m & \text{otherwise} \end{cases} \quad (6)$$

여기서 식(6)의 $h_m(x)$ 는 약한 분류기로 본 표정인 식시스템에서는 앞서 설명한 Gabor 특징에 의해 구성되며 실수 값을 가진다. $H(x) = \text{sign}[H_M(x)]$ 을 통해 부호가 +이면 positive, -이면 negative 로 인식을 하고, $|H_M(x)|$ 의 값은 입력 영상의 해당 표정에 대한 신뢰도를 나타낸다. 입력 영상에 대해 가능한 모든 약한 분류기에 대해 m번째 분류기를 $h_m(x)$ 라고 하면 임계값 θ_m 와 극값 p_m 을 이용해 Gabor 특징 값인 $f_m(x)$ 가 일정 값보다 크거나 작을 경우 positive 라고 판단하게 된다. 본 연구에서는 부호 및 임계값을 훈련 영상 집합에 대한 Gabor 특징 값의 가우시안 분포를 가정함으로써 구한다.

여기서 FloatBoost 알고리즘의 backtrack 방법을 적용하기 위해 먼저 각 라운드에서의 최소 에러를 정의하는 것이 요구된다. M 라운드에서의 최소 에러를 ϵ_M^{\min} 로 정의하면 ϵ_M^{\min} 은 Missing Error Rate와 False Alarm Rate의 평균으로 계산되어진다.

M개의 약한 분류기가 선택된 M번째 라운드를 가정하면 선택된 약한 분류기 중에서 한 개의 약한 분류기를 뺀 M-1개의 약한 분류기들로 이루어진 M개의 조합이 존재한다. 각 조합의 에러 율 중 가장 작은 에러 율이 ϵ_{M-1}^{\min} 보다 작다면 악영향을 주는 약한 분류기를 제거하고 M-1 라운드로 돌아가 다시 악영향을 주는 약한 분류기를 선택을 하고 악영향을 주는 약한 분류기가 없다면 다음 라운드로 진행하게 된다.

이러한 backtrack 방법은 각 라운드에서 최적의 약한 분류기를 선택하지만 M 라운드까지 선택 되어진 약한 분류기의 조합이 M-1 라운드까지 선택 되어진 약한 분류기의 조합보다 반드시 최적의 인식 성능을 보이지 않는 AdaBoost 알고리즘의 문제점을 효과적으로 해결하기 위한 방법으로 전체 인식 율에 악영향을 주는 약한 분류기를 제거함으로써 각 라운드가 진행함에 따라 에러 율이 감소하고 결과적으로 인식 성능이 향상되는 것을 보장하게 된다.

아래 그림 3은 Stan Z. Li와 ZhenQiu Zhang에 의해 제시된 일반화된 FloatBoost 알고리즘을 요약한 것이다.

```

0. (Input)
(1) Training examples  $Z = \{(x_1, y_1), \dots, (x_N, y_N)\}$ ,
    where  $N = a + b$ : of which a examples have
     $y_i = +1$  and b examples have  $y_i = -1$ ;
(2) The maximum number  $M_{\max}$  of weak classifiers;
(3) The error rate  $\epsilon(\mathcal{H}_M)$ , and the acceptance threshold  $\epsilon^*$ .
1. (Initialization)
(1)  $w_i^{(0)} = \frac{1}{2a}$  for those examples with  $y_i = +1$  or
     $w_i^{(0)} = \frac{1}{2b}$  for those examples with  $y_i = -1$ ;
(2)  $\epsilon_M^{\min} = \text{max-value}$  (for  $m=1, \dots, M_{\max}$ ),
     $M = 0, h_0 = \{ \}$ .
2. (Forward Inclusion)
(1)  $M \leftarrow M+1$ ;
(2) Choose  $h_M$  according to Eq. 1;
(3) Update  $w_i^{(M)} \leftarrow \exp[-y_i H_M(x_i)]$ , and normalize to  $\sum_i w_i^{(M)} = 1$ ;
(4)  $\mathcal{H}_M = \mathcal{H}_{M-1} \cup \{h_M\}$ : If  $\epsilon_M^{\min} > \epsilon(\mathcal{H}_M)$ , then  $\epsilon_M^{\min} = \epsilon(\mathcal{H}_M)$ ;
3. (Conditional Exclusion)
(1)  $h' = \text{argmin}_{h \in \mathcal{H}_M} \epsilon(\mathcal{H}_M - h)$ ;
(2) If  $\epsilon(\mathcal{H}_M - h') < \epsilon_{M-1}^{\min}$ , then
    (a)  $\mathcal{H}_{M-1} = \mathcal{H}_M - h'$ 
         $\epsilon_{M-1}^{\min} = \epsilon(\mathcal{H}_M - h')$ ;  $M = M-1$ ;
    (b)  $\mathcal{H}_M = \sum_{h \in \mathcal{H}_M} h$ ;
    (c) goto 3.(1);
(3) else
    (a) if  $M = M_{\max}$  or  $\epsilon(\mathcal{H}_M) < \epsilon^*$ , then goto 4;
    (b)  $w_i^{(M)} \leftarrow \exp[-y_i H_M(x_i)]$ ; goto 2.(1);
4. (Output)
 $H(x) = \text{sign}[\sum_{h \in \mathcal{H}_M} h(x)]$ 
    
```

(그림 3) FloatBoost 알고리즘

5. 인식 과정

훈련 과정에서는 모든 Gabor 특징을 구하기 위해 훈련 영상에 40개의 Gabor 필터와 콘볼루션 연산을 수행하지만 인식 과정에서는 계산량을 줄이기 위해 훈련을 통해 선택된 M개의 Gabor 특징의 값만을 계산하여 인식에 사용하게 된다.

훈련을 통해 M개의 약한 분류기 $h_m(x)$ 가 선택되어 인식 하는 과정은 다음 식과 같이 계산된다.

$$H_{M(x)} = \frac{\sum_{m=1}^M h_m(x)}{\sum_{m=1}^M \alpha_m} \quad \text{where} \quad (7)$$

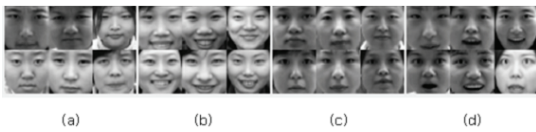
$$h_m(x) = \begin{cases} \alpha_m & \text{if } p_m f_m(x) < p_m \theta_m \\ 0 & \text{otherwise} \end{cases}$$

위와 같이 정규화 된 $H_M(x)$ 의 값이 0.5보다 크면 positive 영상으로 인식을 하게 되고 1에 가까울수록 신뢰도가 크다고 판단한다. 그리고 각각의 표정에 대한 독립적인 훈련 결과를 입력 영상에 대해 병렬로 적용하게 된다. 이 때 복 수개의 표정 인식기에 대해 positive 영상으로 인식이 되는 경우가 발생할

수 있는데 이러한 경우 $H_M(x)$ 의 값을 비교해 가장 큰 값을 가지는 표정으로 인식하게 된다.

6. 실험 결과

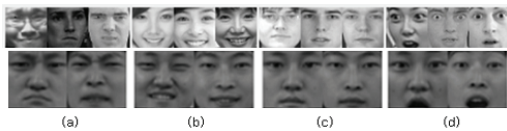
제안된 방법은 Windows XP의 운영체제를 갖춘 펜티엄4 1.6GHz CPU의 데스크톱에서 Visual C++ 6.0으로 구현되었고 훈련 영상 집합은 CAS-PEAL Facial Image Database를 사용하여 4가지의 각 표정을 짓는 560장의 positive 영상과 다른 표정을 짓는 1,680장의 negative 영상으로 구성하였다. 그림 4는 각 표정에 대한 훈련영상집합의 일부를 보여준다.



(그림 4) 훈련 영상

(a) 찡그림 (b) 웃음 (c) 무표정 (d) 놀람

또한 웹에서 획득한 여러 영상들로부터 각각의 표정에 해당하는 50개의 얼굴 영상을 수집하여 웹 영상 집합을 구성하였고, 자체적으로 연구원 5명에 대하여 표정마다 5장의 영상을 각각 구성하여 각 표정마다 25장의 영상을 제작 했다. 이때 영상은 36×36 픽셀 크기로 고정하여 실험하였다. 그림 5는 각 영상 집합의 일부를 보여주고 표 1은 각 영상집합에서 네 개의 각 표정에 대한 인식률을 보여준다.

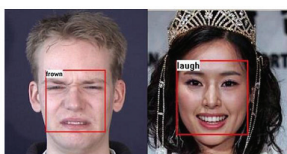


(그림 5) 웹 영상 및 자체 제작 영상
(a) 찡그림 (b) 웃음 (c) 무표정 (d) 놀람

	찡그림	웃음	무표정	놀람
Web 영상 집합	90%	96%	94%	94%
자체 제작 영상 집합	80%	92%	92%	96%
훈련 영상 집합	92%	98%	92%	94%

(표 1) 각 영상 집합에 대한 인식결과

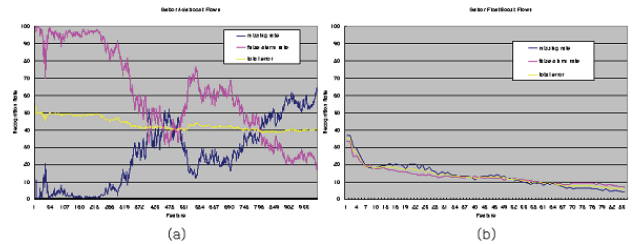
다음 그림 6은 정지 영상 기반 실험 결과를 나타낸다.



(그림 6) 정지영상기반 표정인식 결과

마지막으로 그림 7을 보면 AdaBoost 알고리즘에 비해 FloatBoost 알고리즘이 인식률에 악영향을 주

는 약한 분류기를 제거 하면서 라운드를 진행하기 때문에 전체 에러 율의 감소 효과를 얻는 것을 확인할 수 있다.



(그림 7) (a) AdaBoost (b) FloatBoost의 라운드별 에러 율

7. 결론

본 논문에서는 FloatBoost 알고리즘과 Gabor 특징을 이용한 빠르고 효과적인 표정인식 방법을 제안하였다. 제안된 방법은 기존의 AdaBoost 알고리즘과 Haar-like특징의 단점을 개선하여 에러 율 감소와 처리 속도 향상 측면에서의 개선이 이루어 졌다. 하지만 실험 결과에서 보면 다른 표정에 비해 찡그린 표정이 잘 인식이 되지 않는 것을 알 수 있다. 이는 웃음과 놀람이 보편적으로 공통된 특성을 잘 지니고 있는 반면에 찡그림은 나타나는 형태가 매우 다양하며 36×36 픽셀 크기의 영상에서 뚜렷하게 표현되지 않기 때문인 것으로 분석된다. 따라서 추후 찡그림에 대한 오류를 줄이는 연구가 요구된다.

참고문헌

[1] Paul Viola and Michael Jones. "Robust real-time object detection" Second International Workshop on Statistical Learning and Computational Theories of Vision Modeling, Learning, Computing and Sampling, July 2001.
 [2] Chengjun Liu and Harry Wechsler. "A Gabor Feature Classifier for Face Recognition" IEEE Transactions On Computer Vision, vol.2 , 270~275, 2001.
 [3] Yongzhao Zhan, Jingfu Ye, Dejiao Niu, Peng Cao "Facial Expression Recognition Based on Gabor Wavelet Transformation and Elastic Templates Matching" Int. J. Image Graphics 6(1): 125-138, 2006.
 [4] Li, S.Z. and Zhenqiu Zhang. "FloatBoost learning and statistical face detection" IEEE Transactions on Pattern Analysis and Machine Intelligence. vol.26, no.9 September 2004.