

Integer DCT 와 SVM 을 이용한 실시간 얼굴 검출

박현선, *김경수, *김희정, *정병희, *하명환, 김희을

한양대학교 전자통신전파공학과, *KBS 기술연구소

전화 : 02-2290-0561

Real Time Face Detection Using Integer DCT and SVM

Hyun-Sun Park, *Kyeong-Soo Kim, *Hee-Jung Kim, *Byung-Hee Jung, *Myung-Hwan Ha, Whoi-Yul Kim

Division of Electrical and Computer Engineering, Hanyang University

* Technical Research Institute, Korean Broadcasting System

E-mail : hspark@vision.hanyang.ac.kr

Abstract

The system for the real time face detection is described in this paper. For face verification, support vector machine (SVM) was utilized. Although SVM performs quit well, SVM has a drawback that the computational cost is high because all pixels in a mask are used as an input feature vector of SVM. To resolve this drawback, a method to reduce the dimension of feature vectors using the integer DCT was proposed. Also for the real time face detection applications, low-complexity methods for face candidate detection in a gray image were used. As a result, the accurate face detection was performed in real time.

I. 서론

얼굴 검출은 무인 감시 시스템, HCI (Human Computer Interaction) 시스템, 얼굴 인식 시스템 등에 이용되는 핵심 기술이다. 이러한 시스템 구현을 위해서는 실시간 얼굴 검출이 선행되어야 하며, 이를 위한 많은 방법들이 제안되고 있다.

얼굴 검출은 크게 특징 기반 검출, 영상 기반 검출로 나눌 수 있다[1]. 특징 기반 검출은 얼굴을 특징 지을 수 있는 눈, 코, 입 등을 먼저 검출하고 그 것들의 기하학적 관계를 분석하여 얼굴을 검출하는 방식이다. 영상 기반 검출은 마스크 영상을 신경망이나 SVM (Support

vector machine)과 같은 판별기 (classifier)를 이용하여 얼굴을 판단하는 방법이다. 특징 기반 얼굴 검출은 눈, 코, 입 등을 모두 검출해야 하기 때문에 해상도가 비교적 높아야 하며 알고리즘이 복잡한 단점이 있다. 영상 기반 얼굴 검출은 낮은 해상도에서도 얼굴을 검출할 수 있다는 장점이 있다. 하지만, 영상 기반 얼굴 검출은 마스크 영상의 모든 픽셀을 판별기의 입력 특징 벡터로 사용하고, 다양한 크기의 얼굴을 검출하기 위하여 영상의 크기를 변화시켜가며 전체 영상을 반복적으로 스캔하기 때문에 계산 시간이 오래 걸리는 단점이 있다.

본 논문에서는 그레이 영상에서 정확도가 높은 SVM 을 이용한 영상 기반 실시간 얼굴 검출 시스템을 제안한다. 실시간의 얼굴 검출 속도를 위하여 연산량이 매우 적은 얼굴 후보 영역 추출 방법과 SVM 에 입력되는 특징 벡터의 차원을 줄여서 SVM 의 계산 시간을 줄이는 방법을 제안한다.

II. 시스템 개요

전체 시스템은 그림 1과 같다. 입력된 영상은 그레이 영상으로 변환되고 얼굴 후보 영역을 검출한 다음 얼굴 후보로 판단된 영역에 대하여 얼굴을 검증한다. 각각의 알고리즘은 III 장과 IV 장에서 설명된다.

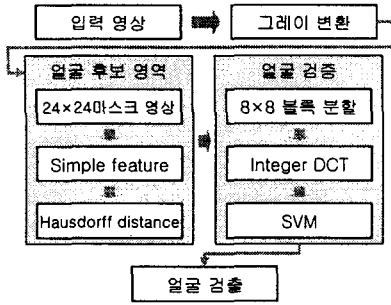


그림 1. 전체 시스템의 블록 다이어그램

III. 얼굴 후보 영역 검출

본 논문에서는 얼굴 검출을 위하여 24×24 크기의 마스크를 스캔하며 얼굴을 판별한다. 352×240 크기의 영상을 1.2 스케일로 축소해가고 마스크를 2 픽셀씩 이동하며 스캔하면 총 3340 개의 마스크 영상이 나온다. 그러나 영상에 존재하는 얼굴의 개수는 마스크 영상의 개수에 비해 매우 적다. 따라서 3340 개의 마스크 영상을 IV 장에서 제안할 방법으로 얼굴을 검증하는 작업은 매우 소모적이다. 이런 문제점을 극복하기 위하여 연산량이 매우 적은 알고리즘을 이용한 얼굴 후보 영역의 검출 방법이 필요하다.

얼굴 후보 영역을 검출하는 대표적인 방법으로 얼굴 색상을 이용하는 방법이 있다 [4][5]. 그러나 얼굴색상을 이용할 경우 조명이 매우 심하게 변하는 경우나 흑백 영상에 대하여서는 적용할 수 없는 단점이 있다. 본 논문에서는 이러한 단점을 피하기 위하여 그레이 영상에서 얼굴의 후보 영역을 검출하는 방법을 이용하였다.

3.1. Simple feature 를 이용한 얼굴 후보 영역 검출

Simple feature 는 얼굴에 존재하는 어둡고 밝은 영역의 특성을 이용하는 것이다 [7]. 예를 들어 그림 2와 같이 눈 주위는 어둡고 볼 주위는 밝은 특성을 이용하여 얼굴이기 위해서는 (영역 B의 픽셀 값의 합)-(영역 A의 픽셀 값의 합)이 임계치 이상이 되어야 하는 조건을 만들 수 있다.

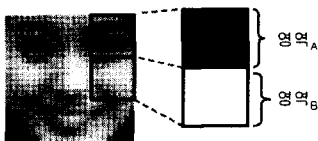


그림 2. Simple feature 의 예

본 논문의 실험에서 사용한 Simple feature 는 얼굴에서 변화가 적은 눈과 볼 영역의 밝기 값을 이용하였으며 그 모양은 그림 3과 같다.



그림 3. 실험에 사용된 Simple feature 들

3.2. Hausdorff distance 를 이용한 얼굴 후보 영역 검출

얼굴 후보 영역을 더욱 정확하게 검출하기 위하여 영상의 밝기값 뿐만 아니라 에지를 이용하였으며, 이를 위해서 Hausdorff distance (HD)를 사용하였다 [8]. HD 는 점으로 이루어진 두 집합 사이의 유사도를 나타낸다. 두 집합을 $A=\{a_1, a_2, \dots, a_m\}$, $B=\{b_1, b_2, \dots, b_n\}$ 라고 하면 HD 는 다음과 같이 정의 된다 [8].

$$HD(A, B) = \max(h_{mod}(A, B), h_{mod}(B, A)) \quad (1)$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (2)$$

$$h_{mod}(A, B) = \frac{h(A, B)}{|A|} \quad (3)$$

얼굴 후보 영역 검출의 경우 A 는 에지 템플릿, B 는 마스크 영상의 에지 영상이 된다. 본 논문에서는 에지 템플릿을 만들기 위하여 트레이닝 샘플을 에지 영상으로 만든 후 모든 에지 영상을 중첩한 다음, 중첩된 에지의 개수가 샘플 개수의 1/5 이상이면 1 로 그렇지 않으면 0 의 값을 갖게 하여 만들었다. 이때 에지를 구하기 위하여 소벨 연산을 이용하였다. 마스크 영상의 에지 또한 소벨 연산을 이용하여 만들었으며 에지의 크기가 임계치 이상이면 1 로 그렇지 않으면 0 으로 하여 HD 를 구하였다.

IV. 얼굴 검증

본 논문에서는 얼굴을 검증하기 위하여 SVM 을 이용하였다. SVM 은 판별 능력이 우수하기 때문에 패턴 인식에 다양하게 사용되고 있다 [2]. [2]에서는 SVM 을 이용하여 얼굴을 검출하기 위하여 19×19 크기의 마스크 영상을 283 차원의 특징 벡터로 만들어 SVM 의 입력으로 이용하였다 [6]. 본 논문에서는 특징 벡터의 차원을 줄여서 SVM 의 계산 시간을 빠르게 하기 위하여 Integer DCT 로 계산된 DCT 계수를 이용하였다.

4.1. SVM

인공 신경망 등의 전통적인 통계학적 학습 방법은 Empirical Risk Minimization 의 문제였는데 반해 SVM 은 Structural Risk Minimization 의 문제로서 일반화 능력이 매우 뛰어나다.

SVM 을 이용하여 두 개의 클래스를 분류하기 위한 학습은 다음과 같다. 우선 학습을 위한 N 개의 샘플을 $\{(x_i, d_i)\} \quad i=1,2,\dots,N$ 로 놓으면 결정 함수는 $f(x)=\text{sign}(wx+b)$ 가 된다. 여기서 $d_i \in \{-1,1\}$ 는 클래스를 나타내며, x_i 는 d_i 클래스에 속하는 샘플의 특징 벡터이다. 이 때 최적의 결정 평면은 다음 식으로 계산된다.

$$\begin{aligned} &\text{minimize: } \langle w \cdot w \rangle \\ &\text{subject to } d_i \langle w \cdot x_i + b \rangle \geq 1 \quad \forall i \end{aligned} \quad (4)$$

식 (4)는 Lagrange multiplier 를 다음 식으로 풀 수 있다.

$$\begin{aligned} &\text{Max } Q(\alpha) = \sum_{i=0}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) \\ &\text{Subject to the constraints} \\ &(1) \sum_{i=1}^N \alpha_i d_i = 0 \\ &(2) 0 \leq \alpha_i \leq C \text{ where } C \text{ is a positive parameter.} \end{aligned} \quad (5)$$

식 (5)는 Quadratic Programming 문제로 보고 풀 수 있으며, 커널 함수 $K(x_i, x_j)$ 를 이용하여 비선형적으로 패턴을 분류할 수 있다. 본 논문에서는 커널 함수로 2 차 polynomial 함수를 이용하였다.

식 (5)를 이용하면 결정 함수는 다음과 같은 형태를 갖는다.

$$f(x) = \text{sign}\left(\sum_i \alpha_i d_i K(x_i, x) + b\right) \quad (6)$$

4.2. Integer DCT

DCT (Discrete cosine transform)는 영상의 압축과 특징값 추출에 매우 유용하게 이용된다. DCT 는 최적의 압축인 KLT (Karhunen-Loeve transform) 압축과 흡사한 성능을 보인다. 이런 압축 능력을 이용하여 에너지가 많은 몇 개의 계수를 선택하면 효과적으로 영상의

특징값을 추출할 수 있다.

다음 식은 1 차원 DCT 를 나타낸다.

$$F(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos \left[\frac{(2x+1)u\pi}{2N} \right]$$

$$\text{이 때, } \alpha(u) = \begin{cases} \frac{1}{\sqrt{N}} & u=0 \\ \frac{2}{\sqrt{N}} & u=1,2,\dots,N-1 \end{cases} \quad (7)$$

DCT 는 압축 성능이 뛰어난 장점이 있지만 식 (7)에서 보이는 것처럼 많은 실수 연산이 필요하다. 이런 이유로 본 논문에서 제안한 실시간용 고속 얼굴 검출 시스템에는 사용할 수 없다.

Integer DCT 는 DCT 의 실수 연산을 근사화 하여 덧셈과 쉬프트 연산만을 이용하여 DCT 를 계산하는 매우 빠른 알고리즘이다 [6].

4.3. Integer DCT 와 SVM 을 이용한 얼굴 검출

Integer DCT 와 SVM 을 이용하여 얼굴을 검증하는 알고리즘은 다음과 같다. 우선 24×24 크기의 마스크 영상이 얼굴 검증을 위하여 입력되고 다시 이 영상은 그림 4와 같이 9 개의 8×8 크기의 블록으로 분할된다. 각각의 블록에 대하여 Integer DCT 를 수행하고 SVM 의 입력 특징 벡터를 만들기 위하여 DCT 계수의 DC 성분을 제외한 12 개의 저주파 DCT 계수를 지그재그 스캔을 이용하여 추출한다. 이 때 DC 성분을 제외한 이유는 조명 성분을 없애기 위해서다. 최종적으로 추출된 특징 벡터는 $9 \times 12 = 108$ 차원이 된다.

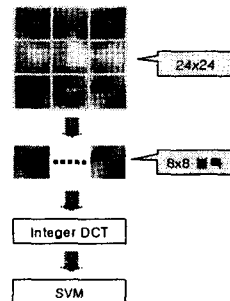


그림 4. 얼굴 검증 알고리즘

V. 실험 결과

실험에 쓰인 데이터는 352×240 크기의 방송용 비디오이며 PC 는 Pentium-IV 1.7 GHz 를 이용하였다.

SVM 학습을 위한 샘플은 Bootstrapping 알고리즘을 이용하여 구하였으며, 구해진 얼굴 샘플의 개수는 295 개이고 비 얼굴 샘플의 개수는 3736 개이다. 얼굴 샘플과 비 얼굴 샘플의 예는 그림 5과 같다.

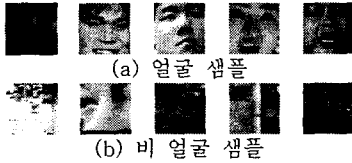


그림 5. SVM의 학습을 위한 샘플

그림 6의 (a)는 Simple feature 를 이용하여 구해진 얼굴 후보 영역이며 그림 6의 (b)는 HD 를 이용하여 검출된 얼굴 후보 영역이다. 그림 6의 (c)는 Simple feature 와 HD 를 직렬 연결하여 검출된 얼굴 후보 영역이다. Simple feature 와 HD 를 이용한 얼굴 후보 영역 검출의 성능이 우수함을 알 수 있다. 그림 6의 (d)는 최종적인 얼굴 검출 결과이다. 계산 시간은 평균 62 ms/f 으로 실시간 검출이 됨을 알 수 있었다.

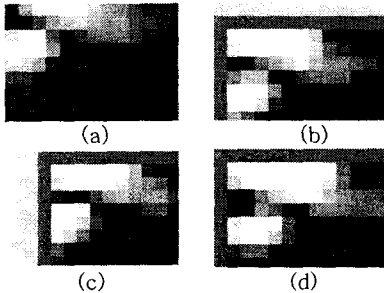


그림 6. 얼굴 검출 결과

그림 6의 (d)의 결과를 보면 한 얼굴에 검출 결과가 중첩된 것을 알 수 있다. 이를 해결하기 위하여 검출된 얼굴 영역의 면적이 75%이상 겹치게 되면 두 영역의 중심 위치를 평균하여 새로운 중심 위치로 설정하고, 두 영역 폭의 평균을 새로운 폭으로 설정하였다. 이 조건을 이용하여 중첩된 얼굴 영역을 병합한 최종 결과는 그림 7과 같다.



그림 7. 최종 얼굴 검출 결과

V. 결론 및 향후 과제

본 논문에서는 얼굴 검증을 위하여 판별 능력이 뛰어난 SVM 을 이용하였으며 SVM 의 입력 특징 벡터의 차원을 줄여서 계산 시간을 빠르게 하였다. 이 때 특징 값의 개수를 줄이기 위하여 연산량이 매우 적은 Integer DCT 를 이용하였다. 또한 계산 시간을 줄이기 위하여 그레이 영상의 밝기값과 에지 정보를 이용한 얼굴 후보 영역을 검출하였다. 실험 결과 조명이 심하게 변하는 영상과 배경이 복잡한 영상에서도 높은 검출률을 가지며 실시간으로 얼굴을 검출함을 알 수 있었다.

본 논문에서는 정면 얼굴에 대하여만 고려하였으나 앞으로 얼굴의 포즈 변화에 강한 얼굴 검출 알고리즘에 대한 연구가 필요하다.

References

- [1] Min-Hsuan Yang, David Kriegman, and Narendra Ahuja, " Detecting Faces in Images: A Survey," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, 2001.
- [2] H.-R. Byun and S.-W. Lee, " Applications of Support Vector Machines for Pattern Recognition: A Survey," Pattern Recognition with Support Vector Machines, Lecture Notes in Computer Science, vol. 2388, pp. 209-231, 2002.
- [3] Ying-Jui Chen, Soontorn Orintara, and Truong Nguyen, " Integer Discrete Cosine Transform (IntDCT)," Int. Conf. on Information, Communications and Signal Processing, Singapore, Dec. 1999.
- [4] K. Sobottka, I. Pitas. " Segmentation and tracking of faces in color images" , Int. Conf. on Automatic Face- and Gesture-Recognition, Killington, Vermont, USA, pp. 236-241, 1996.
- [5] Rein-Lien Hsu, Mohamed Abdel-Mottaleb and Anil K. Jain, " Face detection in color images," IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 696-706, 2002.
- [6] E. Osuna, R. Freund, and F. Girosi. " Training support vector machines: An application to face detection," Proceedings IEEE Conf. CVPR, pp. 130-136, 1997.
- [7] P. Viola and M. Jones. " Rapid Object Detection using a Boosted Cascade of Simple Features," Proceedings IEEE Conf. on CVPR 2001.
- [8] O. Jesorsky, K. Kirchberg and R. Frischholz, " Robust Face Detection Using the Hausdorff Distance," Int. Conf. on AVBPA, Halmstad, Sweden, Springer, pp. 90-95, 2001.