

복잡한 배경에서 신경망을 이용한 얼굴인식
한준희^{*} · 남기환^{*} · 박호식^{*} · 이영식^{*} · 정연길^{*} · 나상동^{**} · 배철수^{*}
^{*}관동대학교 · ^{**}조선대학교

Face Recognition on complex backgrounds using Neural Network

Jun-hee Han^{*} · Kee-hwan Nam^{*} · Ho-sik Park^{*} · Young-sik Lee^{*} · Yeon-gil Jung^{*} · Sang-dong Ra^{**} · Cheol-soo Bae^{*}
^{*}Kwandong University · ^{**}Chosun University

E-mail : hanjh123@empal.com

요 약

복잡한 배경을 지닌 이미지에서 얼굴을 검출하기란 매우 어려운 일이다. 본 논문에서는 신경망 모델을 기반으로 한 제한생성모델(CGM: Constrained Generative Model)을 제안한다. 학습 과정의 목표라 할 수 있는 생성은 신경망 모델이 입력 데이터를 발생시킬 확률을 계산하도록 하는 것이고, 계산하는데 걸리는 시간을 줄이기 위해서 고속 탐지 알고리즘을 제안한다. 얼굴 측면 검출과 오 인식의 수를 줄이기 위해서 조건을 혼합한 신경망을 사용하였고 반증으로 인한 제한을 뚫으로써 모델의 측정 품질을 증가시켰다. 본 논문에서 제안한 검출 알고리즘이 0. ~60. 사이에서는 90%정도의 검출율을 나타냄을 알 수 있었다.

ABSTRACT

Detecting faces in images with complex backgrounds is a difficult task. Our approach, which obtains state of the art results, is based on a generative neural network model: the Constrained Generative Model (CGM). To detect side view faces and to decrease the number of false alarms, a conditional mixture of networks is used. To decrease the computational time cost, a fast search algorithm is proposed. The level of performance reached, in terms of detection accuracy and processing time, allows to apply this detector to a real word application: the indexation of face images on the Web.

키워드

CGM, MLP

1. 서 론

이미지에서 얼굴을 검출하기 위해서는 이미지의 평면(x,y)과 크기 또는 배율(z)을 찾아야 한다. 얼굴 이미지는 타원형안에 눈, 입, 코 같은 특징들의 집합으로 볼 수 있으며, 확실한 모델로 사용할 수 있다. 특징을 추출하기 위한 간단하고 빠른 방법은 코를 기준으로 수평축상의 눈과 입을 찾기 위하여 수직축 상의 예지

이미지 검출이나 이미지 투영법과[1], 가버 필터[2], 타원형 검출[3], 결정 트리[4], 신경망[5]등 여러 가지의 다른 방법들이 특징을 추출하기 위해 사용된다. 확률 또는 통계 모델은 각 서브 윈도우의 픽셀 강도를 분석하기 위하여 사용된다. 이 모델은 여러 가지 다른 방법들과 생성되는데 신경망[6], 주성분분석[7], Kullback 거리측정 방법[8]등이 대표적이다.

이런 모델들의 장점은 픽셀 강도 분석을 바로 수행하는 방법과 비교해 볼 때 특징 추출 알고리즘의 속도

와 유사도 측정면에서 뛰어나다는 것이다. 본 논문에서는 제안한 것은 표준화 된 영상 처리를 기반으로 한 단순한 처리와 통계적 분석을 기반으로 한 정교한 처리 방법이다. 본 논문의 구성은 다음과 같다. 2장에서 모션 필터, 컬러 필터, 전 네트워크 필터, 신경망 필터의 조합으로 방향에 따라 얼굴 검출 능력을 향상시키기 위한 방법을 기술하며 3장에서 빠른 얼굴 검출 알고리즘에 대해 기술하고, 4장에서 제안한 얼굴 검출기의 수행 능력을 비교 분석하였으며, 5장에서 결론을 맺었다.

II. 얼굴 검출기

x 는 서브 윈도우로 이미지로부터 15×20 픽셀 크기를 추출하여 얼굴영역 ($x \in V$) 또는 비 얼굴영역 ($x \in N$)으로 구별하며 얼굴 검출기는 단순하고 빠르지만 정확도는 떨어지고 복잡하고 느리지만 정확도가 높은 4개의 필터들로 이루어져 있으며 구조는 계층적이다.(그림. 1) 이 구조의 장점으로 첫 번째 필터는 빠르면서 계산 시간 비용을 줄이는 것이지만 단점으로 검출율 감소의 요소가 있습니다. 이 문제를 해결하기 위하여 처음 3개의 필터는 매우 높은 검출율을 가져야 합니다.

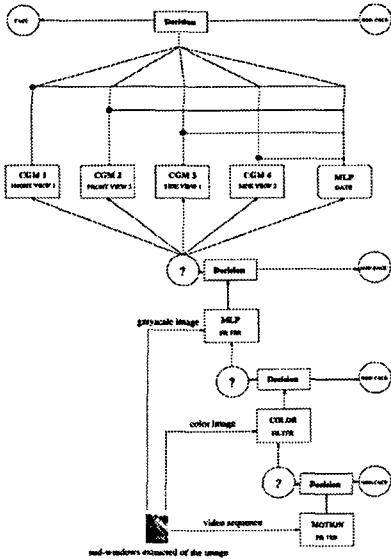


그림 1. 4개의 필터를 조합한 얼굴 검출기

2.1 가설 제거

사람의 얼굴은 말하고 숨쉬며 눈을 깜박이고 하는 등 대부분의 시간에 항상 움직이고 있고, 간단한 문턱치화 된 영상 차이로 구성된 모션 필터가 비디오 시퀀스에서 작동되고 있다고 가정하자. 문턱치와 비디오 시퀀스에 따라서, 자동 프레임을 이용한 실험은 일반적으로 가설의 90%를 허용치 않는다. 칼라 정보가 이용될 때는 픽셀 테이블로

구성되고, 얼굴 이미지의 수집에서 직접 고른 칼라 필터가 적용된다. 작은 피부 픽셀로 이루어진 서브 윈도우는 배경 서브 윈도우로 간주된다.

서브 윈도우의 총 갯수의 대략 40%로 대응되는 다른 픽셀들은 신경망 전 필터로 계산되어진다. 전 네트워크(prenetwork)는 단일 다층 퍼셉트론(MLP)이다[9]. 이것은 추출된 서브 윈도우의 크기에 대응하는 300입력과 20개의 은둔 뉴런, 총 6041개의 가중 값에 대한 하나의 출력(face/nonface)을 갖게 된다. 전 네트워크는 표준 역전파 알고리즘을 사용하여 훈련시킨다. 얼굴 훈련 셋은 8000개의 정면 영상과 측면 영상으로 구성된다.

그림 2는 15×20 픽셀 서브 윈도우에서 얼굴 중앙을 나타낸 것이다. 얼굴과 대응하지 않는 15×20 픽셀 서브 윈도우로 이루어진 50,000개의 비 얼굴 영상은 다음 절에서 기술할 반복 알고리즘을 이용하여 수집된다.



그림 2. 0. - 20. 와 20. - 40. 의 얼굴영상

서브 윈도우는 히스토그램 균등화와 평활화에 의해서 향상된다. 이때 서브 윈도우는 평균 얼굴의 템플릿으로 정규화된다. 획득된 전 네트워크는 매우 빠른 검출율(99%이상)을 갖는 비교적 작고 고속의 네트워크이며 고속의 오류 경고율(1% 정도)을 갖는다. 약한 오류 경고율로 인하여 단독으로 이용되지 않는 이 네트워크는 가설의 93%보다 더 많이 제외시키는 필터로 이용된다.

2.2 CGM의 조합

방향에 따른 얼굴 검출 능력을 높이기 위하여 확장하기 위하여 몇 개의 CGM들을 조합시켰다. 출력 f_i 는 각 추정 f_i 의 출력들의 평균입니다.

$$f(x) = \frac{1}{N} \sum_i f_i(x)$$

N 은 추정의 수이고 CGM i 의 출력은 $f_i(x) = P_i(y=1|x)$ 이다. y 가 이진 변수이기 때문에 $f_i(x) = E_i[y|x]$ 를 가질 수 있다. 임의의 변수 θ 는 2개의 부분 집합으로 나누어집니다.

1. ε_1 , 정면 얼굴 집합($\theta=1$)
2. ε_2 , 측면 얼굴 집합($\theta=2$)

임의 변수 θ 값을 알고 있을 때 각 모듈은 서브 윈도우의 확률을 평가합니다. CGM j 게이트 네트워크의 출력은 다음과 같다.

$$f_w(x, j) = \overline{P}(\theta = j|x)$$

W 는 게이트 네트워크의 가중치이다. 그러면 모

들러 시스템의 출력의 다음과 같다.

$$\bar{P}(y=1|x) = \sum_{j=1}^N (f_w(x, j)) \bar{P}(y=1|x, \theta = j)$$

얼굴 서브 윈도우에서 응답하는 임의 변수 $y=1$ 이면, N 은 추정된 수이고 CGMj의 출력이다.

게이트 네트워크의 훈련 위상동안 비용 함수는 다음과 같다.

$$C_w = \sum_{x_i, y_i} [f_w(x_i, j) \bar{P}(y=1|x_i, \theta = j) - y_i]^2$$

예를 들어 만일 2가지의 추정이 사용되면 다음과 같이 4가지로 정의된다.

- F 정면 얼굴 집합
- P 변형된 얼굴 집합이고 $F \cap P = \emptyset$
- 얼굴 집합은 $V = F \cup P$
- N 은 비얼굴 집합이며 $V \cap N = \emptyset$

결국 $P(x \in V|x)$ 를 평가하는 것이며 추정 범칙은 각각 계산합니다.

- $P(x \in F|x \in F \cup N, x) (CGM1(x))$
- $P(x \in P|x \in P \cup N, x) (CGM2(x))$

게이트 네트워크 $f_w(x)$ 는 다음과 같은 비용함수를 사용하여 실험 이미지를 얼굴로 확률적으로 계산되어진다.

$$C_w = \sum_{x_i \in V} ([f_w(x_i) CGM1(x) + (1 - f_w(x_i)) CGM2(x) - y_i]^2 + \sum_{x_i \in V} (f_w(x_i) - 0.5)^2)$$

III. 검색 알고리즘

검출기는 고정된 크기 15x20 픽셀의 서브 윈도우에 얼굴을 위치시킵니다. 다른 크기의 얼굴을 검출하기 위하여, 원본 이미지의 서브 샘플링을 만들어 수행합니다. 철저한 검색은 서브 윈도우의 매우 큰 수를 평가하기 위한 것입니다. 모든 서브 샘플링 이미지에 모든 서브 윈도우는 테스트해야 합니다. 처음 2개의 필터(모션 필터와 컬러 필터)의 목표는 처리 시간의 매우 작은량을 사용하면서 가설을 삭제하는 것입니다. 그러므로 그레이 스케일 이미지들에게는 이 필터들은 사용할 수 없습니다. 단 남아있는 필터는 선 네트워크 필터이며 6,041 가중치로 구성되어 있습니다. 각 추출된 서브 윈도우에, 6,041번의 곱셈과 6,041번의 가산은 수행되어야 합니다. 계산 시간 비용을 감소시키기 위하여, 단일 다층 퍼셉트론은 전 네트워크[10]와 같이 사용될 수 있습니다. 얼굴의 가능한 위치를 결정하고 다음에 더 큰 네트워크는 정밀한 위치를 찾도록 사용됩니다. Ben-Yacoub[11]에 의해서 개발된 또 다른 접근 방법은 Fourier 공간에서 이미지의 푸리에 변환을 계산하는 것입니다. 얼굴 검출 처리의 계산 시간 비용을 감소시키기 위하여, 분석한 서브 윈도우의 수를 감소시키는 것입니다. 얼굴 검출기는 매우 선택적[12]입니다. 배경 서브 윈도우에

평균 출력은 얼굴 서브 윈도우의 평균 출력과 비교하여 낮습니다. 이런 특성들로 얼굴 검출 처리 성능을 향상시키기 위해서는 다음과 같다.

● 그리드의 교점에 대응하는 화소는 각 크기별로 검출하며 필터(동작 필터, 색상 필터, 신경망 필터 그리고 모듈 시스템)의 의하여 균일하게 분산된다.

● 최초 임계값보다 최종 모듈의 출력이 커지게 되면 국부적 총체적 검색이 $G(x)$ 의 지점에서 수행 된다.

● 두번째 임계값보다 커지게 되면 $G(x)$ 에서 국부적인 총체적 검색 지점은 각 크기별로 보조창에 대응되고 V_c 에 저장된다.

● V_c 의 보조창의 다른 지점과 크기 사이에서 겹쳐서 제거되거나 추가된 부분을 얼굴의 위치를 정하는 과정을 수행한다.

컬러 이미지에서 얼굴의 상세한 검색, [15x20, 150x200] 픽셀 범위 이내의 크기, 500,000 테스트를 필요로 합니다. 빠른 검색 알고리즘의 사용은 25,000번의 테스트 수를 감소시킵니다. 18,600개의 가설은 선 network 필터에 의해서 삭제된 것에 의해서 버려집니다 그리고 그 다음에 모듈러 시스템은 600개의 보조 윈도우를 평가합니다.

IV. 실험 결과

서로 다른 모델과 모델 조합사이에 검색을 하여 비교하였으며, 검출율, 오류 경고율, 그리고 처리 시간에 대한 검색 알고리즘 영향력을 분석하였다. 얼굴 데이터 베이스는 8,000개의 다양한 얼굴 예를 포함시켰다. 데이터 베이스는 서로 다른 각도 [0, 20,], [20, 40,], [40, 60,], [60, 80,] 해당하면서 같은 크기의 4개의 집합으로 나누었다. 5개의 CGM으로 구성되며 처음부터 네 개 (CGM1, CGM2, CGM3, CGM4)는 각 방향에 해당하고, 마지막(CGM5)은 모든 얼굴 데이터 베이스로 사용합니다. 각 얼굴 집합의 75%은 훈련을 위해서 사용되고 그리고 남아있는 25% 얼굴은 모델로 선택할 것이다. 학습 처리 동안에, 비 얼굴 예들은 100가지의 배경 이미지의 집합으로 모아 집니다. 모델을 선택하기 위하여, 100개의 배경 이미지 2번째 집합은 오류 경고율을 평가하기 위하여 사용한다.

CMU의 테스트 집합 A를 사용하였습니다. 그리고 다양한 품질의 42가지의 이미지를 포함하였습니다. 반례(GM)없이 한 오류 검출 비율은 너무 커서 얼굴 검출기로서 사용하지 못합니다. CGM(CGM5)에 의해 수행된 얼굴 이미지들의 확률 분포는 one SWN에 의해 얻어진 것보다 많이 정밀합니다.(표. 1)

우리들의 얼굴 검출기는 측면 얼굴도 검출할 수 있습니다. 모든 알고리즘들(상세 검색, 빠른 검색, 매우 빠른 검색)의 검출 비율은 60도까지 90%(표. 2)의 정도입니다.

90%이상의 검출율을 가졌다.

표 1. 테스트 집합 A CMU 결과

Model	Detection Rate	False Alarm Rate
GM	84%	1000 10^{-6}
CGM1	77%	5.43 10^{-6}
CGM3	85%	6.3 10^{-6}
CGM5	85%	1.36 10^{-6}
one SWN(Rowley95)	84%	8.13 10^{-6}
Ensemble	74%	0.71 10^{-6}
Conditional ensemble	82%	0.77 10^{-6}
Conditional mixture	87%	1.15 10^{-6}

표 2. 방향에 의한 결과

Orientation	Algorithm1	Algorithm2	Algorithm3
0	99%	99%	99%
10	97%	97%	97%
20	95%	95%	95%
30	99%	97%	95%
40	96%	96%	95%
50	92%	91%	90%
60	89%	88%	87%
70	66%	64%	62%
80	45%	44%	42%
90	23%	23%	22%

V. 결 론

본 논문에서는 새로운 신경망 모델인 제한 생성 모델(CGM)을 제안하였고, 반증된 예들의 작은 집합을 사용하여 얼굴의 정확한 특징을 검출합니다.

이 알고리즘의 단점은 모델의 크기이지만, 단점을 해결하기 위한 방법으로 선 처리 필터와 빠른 검색 알고리즘을 사용하였다. 이 검출기는 정면 영상뿐 아니라 측면 영상도 검출하게 되고, 오류 경고율은 5.10^{-6} 에 존재하며 고속 알고리즘을 사용하였다. 검출율과 오류 경고율을 향상시키기 위해서 처리 시간의 증가와 관계없이 더 많은 평가를 수행하였다. 본 논문에서 제안한 알고리즘이 60. 까지

참고문헌

- [1] T. Kanade, "Picture processin by Computer Complex and Recognition of Human Faces", technical report, *Dept. of Computer Science*, Kyoto Univ., 1973.
- [2] B.S. Manjunath, R. Chellappa, and C. von der Malsburg, "A Feature-Based Approach to Face Recognition", *Proc. IEEE Computer Soc. Conf. Computer Vision*, pp. 373-378, 1992.
- [3] R. Seguir, "Human Faces Detection and Tracking in video Sequence", *Proc. Seventh portugese on Pattern Recognition*, 1995.
- [4] J.Huang, S. Gutta, and H. Wechsler, "Detection of uman Faces Using Decision Trees", *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, PP. 248-252, 1996.
- [5] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, Oct. 1993.
- [6] E. Viennet, F. Fogelman-Soulie, "Scene segmentation Using Multiresolution Analysis and MLP", *Artical Neural Networks*, vol. 2, pp.1599-1602, 1992.
- [7] M. Turk and A. Pentland, "Eigenfaces for Recognition", *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
- [8] A.J. Colmenarez and T.S. Huang, *Face Detection and Recognition*, pp. 174-185. Springer-Verlag, 1998.
- [9] R. Vaillant, C. Monrocq, and Y. Le Cun, "Original Approach for the Localistion of Objects in Images", *IEEE Proc. Visual Image Signal Peocess*, vol. 141, pp. 572-577, Aug. 1994.
- [10] H. Rowley, S. Baluja, and T. Kanade, "Neural Network base Face Detection", *IEEE Trans. Pattern Analysis Machine Intelligence*, 1998.
- [11] S. Ben-Yacoub, "Fast Object Detection Using MLP and FFT", *technical report, IDIAP, Martigny, Switzerland*, 1997.
- [12] T. Yamaguchi, M.Tominage, K. Mrakami, H. Koshimizu, "Regeneration of facial image eye-contacting with partner on TV Conference environment", *IEEE International Conference on Systems, Man, and Cybernetics*, vol.2, pp.1169-1174, 2000.