

외형 기반의 얼굴 검출을 위한 학습 데이터의 최적화

이재훈⁰, 조병모, 이필규
인하대학교 전자계산공학과
{lune⁰, chobm77}@im.inha.ac.kr, pkrhee@inha.ac.kr

Training Set Optimize for Face Detection by Appearance-based Model

Jae-Hoon Lee, Byoung-Mo Cho, Phill-Kyu Rhee
Dept. of Computer science&engineering Inha-Univ.

요약

얼굴 검출은 하나의 영상으로부터 얼굴 존재 유무를 판단하고 그 위치와 방향, 크기 등을 알아내는 기술로 정의된다. 그러나 영상 내의 특정 위치에 대한 얼굴 여부의 판단은 여러 가지 환경 변화와 매우 다양한 종류의 얼굴로 인해 정확하고 빠른 검출이 어렵다. 따라서 본 논문에서는 얼굴여부를 판단하기 위한 학습 데이터를 최적화하여 일반적인 외형기반의 알고리즘에 적용할 수 있는 방법을 제안한다.

제안된 방법은 영상에 대한 기본적인 전처리부터 입력으로 사용될 데이터의 추출에 이르기까지 최대한의 환경변화를 고려함으로써 실제 적용 시 정확하고 빠른 판단이 가능하도록 하였다. 영상의 전처리로는 조명의 보상과 히스토그램 평활화가 사용되었고, 입력으로 사용하기 위한 학습 데이터의 정렬과 영상 샘플링 방법이 제안되었다.

얼굴 여부의 판단 실험은 각각 역전파 신경망, 마할라노비스 거리를 사용하여 영상의 얼굴 여부를 판단하고, 성공률을 측정하였다. 실험 결과 최적화 방법을 적용했을 때 적용하기 전보다 높은 성능의 성공률을 보였다.

1. 서론

얼굴 정보를 인식하고 분석하는 시스템에서 가장 근본적으로 해결되어야 할 문제는 주어진 영상에서 얼굴 영역의 위치를 추정하는 것이고, 이를 위해서는 최소의 하위 영역에서 해당영역의 이미지가 얼굴인지 아닌지 정확히 판단할 수 있어야 한다. 얼굴 여부의 판단은 일반적으로 미리 정의되거나 학습되어진 데이터에 의존하므로, 이를 위한 학습 데이터의 최적화는 필수적이라고 할 수 있다.

본 논문에서 제안하는 학습 데이터의 최적화 방법은 디지털화 된 임의의 영상으로부터 학습 영역을 효율적으로 추출해내고 이것에 대한 적절한 전처리와 샘플링을 통해 각종 외부환경요인의 영향을 최소화할 수 있도록 한다.

학습 데이터는 최초 같은 크기와 같은 각도로 정렬되어 학습영역이 추출되어지며, 추출된 영상에 대해 조명 보정, 히스토그램 평활화 작업을 거치게 된다. 이렇게 전처리가 끝난 영상은 마지막으로 환경변화에 대한 영향을 최소화 하기 위해 변화가 가장 작은 부분을 기준으로 샘플링 되어 신경망의 입력으로 사용되기 위한 하나의 일차원 벡터로 변환되게 된다.

본 논문의 구성은 다음과 같다. 2장에서는 외형 기반의 얼굴 검출에 대한 기존 연구방향을 기술하고 3장에서는 영상의 전처리, 특징 추출에 대한 기본 개념을 설명한다. 4장에서는 앞서의 개념을 사용한 학습 데이터의 최적화 방법을 기술하고, 5장은 제안된 방법을 사용하여 만들어진 학습데이터를 이용한 실험 결과를, 6장에서는 실험 결과를 바탕으로 결론 및 향후 연구과제에 대해 기술한다.

2. 관련 연구

외형 기반 모델(Appearance-based Model)은 형태(shape)와 조적상태

(texture)를 바탕으로 하여 대상의 컬러나 밝기값을 패턴으로 추출하는 것이다.[2] 일반적인 외형기반의 패턴 인식 시스템은 여기에서 추출한 패턴을 그대로 입력으로 하여 신경망(Neural-Network), SVM(Support Vector Machine) 또는 단순히 통계적 방법을 이용하거나 기타 적절한 알고리즘으로 그 역할을 수행한다.[3][4][5]

3. 영상의 전처리 및 특징 추출

3.1. 조명 보정

같은 영상 내에서의 영상의 배경조명의 분포를 조사하고, 이 분포로부터 도출된 조명 변화도(illumination gradient)를 이용하면 조명의 영향을 보상할 수 있다.



$$I(x, y) = \frac{\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} I(x+i, y+j)}{(2n+1)^2} \quad (1)$$

$$G(x, y) = \max(I(x, y) - I(x, y), 0) \quad (2)$$

$I'(x, y)$: 배경조명 모델링 함수[7]

$I(x, y)$: 원 영상

$G(x, y)$: 개선 영상

영상에서 식 (2)를 거치기 전의 중앙값에서 거친 후의 중앙값을 뺀 값으로 지역적 밝기를 알 수 있고, 이 값을 이용하여 원 영상에서부터 감산을 한 후, 각 픽셀에 대해 원 영상 전체의 평균을 합하면 개선된 영상을 구할 수 있다.

3. 2. 히스토그램 평활화

히스토그램 평활화는 영상의 히스토그램이 평탄하게 되도록 명암값의 재분배를 통해 농도를 조절하는 것이다.

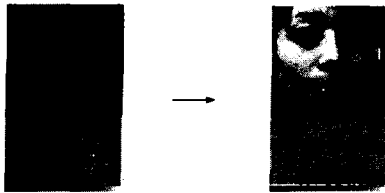


그림 2. 히스토그램 평활화

3. 3. 특징 추출

모든 외형기반의 인식에 있어서 가장 중요한 것은 원 영상에서부터 어떠한 부분을 어떻게 추출할 것인가 하는 것이다.

본 논문의 초점이 되는 특징들은 얼굴의 표정이나 다른 얼굴과의 차를 최소로 하는 부분을 찾기 위해 총 2000개의 얼굴영상(100명, 한명당 20장)에 대해 32개의 특징점을 수작업으로 만들어 다음장에 설명된 정렬방법을 사용한 후에 특징점들의 표준편차를 구하여 그 값이 상대적으로 가장 적은 3개의 점을 사용하였다. 변화가 가장 적을것으로 기대되는 특징점은 3개로, 두 눈의 중심과 입의 중심이 그것이다.

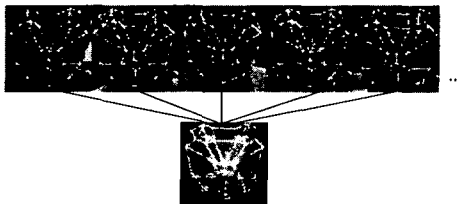


그림 3. 최소 변화영역의 도출

4. 학습 데이터의 최적화

4. 1. 학습 데이터의 정렬

얼굴영역을 학습데이터로 정하는데 있어서 정렬은 매우 중요한 요소이다. 이에 대해 Henry A. Rowley는 두 눈과 입의 중심점을 기준으로 하여 표준이 되는 평균 좌표를 만든 후 기존 영상을 이에 맞추어 변형하는 방법을 사용하였다.[3]



그림 4. 영상의 변형

그러나, 이 방법은 원 영상의 상하, 좌우 비율에 대한 변화가 일어나게 되므로 그 신뢰성을 보장할 수 없게 된다.

따라서 본 논문에서는 영상의 정렬과 추출에 대해 다음과 같은 방법을 제안한다.

1. 양쪽눈의 높이가 똑같이 되도록 학습데이터의 회전변환을 수행한다.
2. 양쪽눈의 중심 좌표간 거리가 모두 똑같이 되도록 학습데이터의 크기를 가로, 세로 비율을 유지한 상태로 조절한다.
3. 모든 데이터의 눈과 입사이의 중심좌표에 대한 평균을 구한다.
4. 3에서 구한 평균 중심좌표를 기준으로 눈의 거리까지를 반지름으로 하는 원을 그리고, 이 원의 외곽을 따라 영상을 다시 추출해낸다.



그림 5. 영상의 추출

이렇게 추출된 영상은 앞 장에서 설명한 조명 보정의 알고리즘과 히스토그램 평활화를 통하여 얼굴의 특징정보를 유지하면서도 외부의 환경 변화요인을 최소화하게 된다.

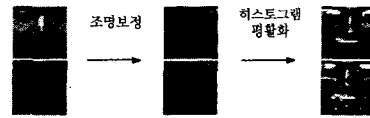


그림 6. 영상의 전처리

4. 2. 학습 데이터의 샘플링

일반적으로 영상의 최종 샘플링은 영상의 좌측 상단에서 우측 하단의 방향으로 1차원 벡터화 하거나[4], 영상을 단위 셀(Cell)로 나누어 벡터화를 시도하는 방법[3], 또는 각 특징에 대한 관계를 통계적으로 산출하는 방법[5]들이 있다.

그러나, 얼굴의 영상은 개인에 따라 특정 위치에 대해서는 변화가 심할 수 있다. 일반적으로 얼굴의 영역에서 변화가 거의 없는 부분은 눈과 코 부분인데 이부분은 개인차가 있는 경우를 고려하더라도 크게 차이가 나지 않으므로 최종 입력을 위한 샘플링의 대상에 가장 적합한 부분이 될 수 있다.

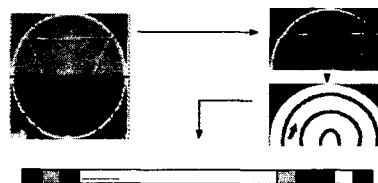


그림 7. 반원형 샘플링.

따라서, 본 논문에서의 학습데이터 최종 샘플링은 두 눈과 입의 중심을 기준으로 한 상단의 반원을 1차원 벡터화 하여 이루어진다.

5. 실험결과

본 논문의 실험으로 사용된 데이터는 얼굴 영상의 경우, 20x20의 크기를 기준으로 하여 정렬되어진 총 3000여개의 데이터중 1000개는 학습 데이터로, 2000개는 테스트 데이터로 사용하였다. 이중 1000개의 데이터는 University of Essex에서 보유하고 있는 데이터이다[8] 또한 이 데이터를 사용한 학습의 성능을 검증하기 위해 각각 마할라노비스 거리 기반의 검증, 역전파 신경망기반의 검증을 사용하였으며, Negative Data는 얼굴이 포함되지 않은 이미지 1000개를 무작위 추출하였다.

비교 실험을 위해 학습데이터 제작시 기초적인 방법을 적용할 경우 히스토그램 평활화와 Henry A. Lowley가 제시한 정렬방법[3] 및 단순 1차원 벡터화 방법을 사용하였고, 제시된 방법에 대해서는 본 논문에서 사용된 모든 과정을 적용하였다.

5. 1. 마할라노비스 거리 기반의 얼굴검출 검증

마할라노비스 거리(Squared Mahalanobis Distance)[6]는 임의의 모델 m 과 임의의 패턴 x 의 유사도를 측정하는 유용한 방법들 중에 하나로 사용된다. 식 (2)에 마할라노비스 거리의 식을 나타냈다.

$$r^2 = (x - m_x)' C_x^{-1} (x - m_x) \quad (3)$$

마할라노비스 거리를 사용하여 표 1과 표 2에 판단 성공률의 실험 결과를 보였다. 표 1은 얼굴에 대한 실험 결과이고 표 2는 비얼굴에 대한 결과이다.

표 1. 마할라노비스 거리기반의 얼굴 판단 성공률

93.1 %	88.3 %	90.7 %
98.2 %	94.4 %	96.3 %

표 2. 마할라노비스 거리기반의 비얼굴 판단 성공률

89.4%	106 / 1000
91.4%	86 / 1000

5. 2. 역전파 신경망 기반의 얼굴검출 검증

역전파 신경망은 Rumelhart등에 의해 구체화된 알고리즘으로, 직감적인 이해가 가능하며 비교적 쉬운 프로그래밍 때문에 주로 패턴인식, 제어 등에서 많이 사용된다.[1]

역전파 신경망의 학습을 위해 별도로 무작위 추출된 1000개의 Negative data를 사용하였으며 그 외의 데이터는 앞 절의 마할라노비스에서 사용된 것과 동일한 것을 사용하였다.

표 3. 역전파 신경망 기반의 얼굴 판단 성공률

93.5 %	91.1 %	92.3 %
95.7 %	93.8 %	94.75 %

표 4. 역전파 신경망 기반의 비얼굴 판단 성공률

90.1 %	99 / 1000
93.2 %	68 / 1000

실험에서 알 수 있듯이 제안된 방법에 의한 학습 데이터 작성시 효율적인 성능을 보이는 것을 알 수 있다. 또한 역전파 신경망의 경우는 학습데이터의 증대를 통해 실제 알고리즘의 적용에서 보다 높은 정확도를 기대할 수 있다.

6. 결론 및 향후 연구방향

본 논문에서는 얼굴 검출을 위한 학습데이터의 작성에 대해 가장 효과적이라고 생각할 수 있는 방법을 제안하였다. 여러 가지 환경의 변화요인을 고려하고 실제 적용에서 환경 변화에 대해 안정적인 성능을 보이며, 최대한 성능을 높이는 방법으로 방향을 모색했다. 그러나 단순히 얼굴영상에 대한 판단만으로는 최종적으로 요구되는 얼굴 영역의 검출 및 인식을 해낼 수는 없다. 따라서 이에 대한 추가적인 연구와 실험이 지속적으로 행해져야 할 것이다.

7. 참고 문헌

[1] Joey Rogers, "Object-Oriented Neural Networks in C++", Academic Press, 1997
 [2] T.F Cootes, C.J.Taylor. "Statistical Models of Appearance for Computer Vision", Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, Manchester M13 9PT, United Kingdom, September 1999.
 [3] Henry A. RowLey, Shumeet Baluja and Takeo Kanade, "Neural Network-Based Face Detection", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.20, No.1, Jan 1998.
 [4] M. Propp and A. Samal, "Artificial neural network architectures for human face detection", Intell. Engrn. Syst. Artif. Neural Networks, pp. 535-540, 1992
 [5] M. C. Burl, T. K. Leung and P. Perona, "Face Localization via Shape Statistics", International Workshop on Automatic Face- and Gesture-Recognition, pp. 154-159, 1995.
 [6] Richard O. Duda, Peter E. Hart, David G. Stock, Pattern Classification 2nd Edition, Wiley-Interscience 2001.
 [7] Ramesh Jain, Rangachar Kasturi, Brian G. Schunck. "Machine Vision", pp.118~121, McGraw-Hill, 1995
 [8] Vision Group of Essex University Face Database, "Description of Labor Spacek's Collection of Facial Images", <http://cswww.essex.ac.uk/allfaces>