

Simple Instance Face Detection

Name: 정의철	ID:201921166
Output Image (정답 출력 화면 캡처)	
	
<p>[1] Describe the main reason why the given framework (based on Cross-correlation) does not work correctly for detecting the instance face image. (주어진 코드가 동작하지 않는 이유 기술)</p> <p>Cross-correlation의 연산 결과는 $G = H \otimes F$ 이다. 따라서 $G[i,j]$의 값은 $[j - k \leq x \leq j + k]$, $[i - k \leq y \leq i + k]$구간에서 $F[y][x]$의 원소가 최댓값(255 또는 1)에 근접 하다면 큰 값을 가질 수 밖에 없다. 따라서 이번 과제에서 H는 얼굴 이미지이고, F는 단체 이미지라고 생각한다면, 가운데 여자의 흰 셔츠 어깨 부분이 전체 이미지에서 최댓값에 근접하기 때문에 해당 영역을 얼굴로 인식했다고 볼 수 있다.</p>	
<p>[2] Propose the proper method for fixing the given framework using Euclidean distance (ED). (ED를 이용한 개선 방향 설명)</p>	

Cross-correlation과는 다르게 ED를 사용하면 비슷할수록 값이 작아진다. 따라서

```
[y,x]=find(ret<threshold);
```

위 코드의 비교 연산을 '0 < ret & ret < threshold'와 같이 바꿔 주어야 할 것이다.

EuclideanDistance 함수는 $\sqrt{\sum_{i=1}^k \sum_{j=1}^k [H(i,j) - F(y+i, x+j)]^2}$ 식을 유도하여 코드를 완성하였다. ED는 Cross-Correlation과는 다르게, 거리를 측정할 수 있다. 따라서 F 이미지의 영역 대부분이 최댓값에 가깝더라도, Cross-Correlation 보다 더 좋은 성능을 보여줄 것이라고 생각한다.

[3] After modifying the code using ED, describe the performance changes according to the different threshold. (Threshold 변화에 따른 성능 변화 기술)

Cross-Correlation을 통해 계산할 경우, 얼굴 감지의 조건이 'threshold<ret'이지만 ED의 경우에는 그의 반대이다. 따라서 얼굴에 해당하는 부분의 조건을 '0 < ret & ret < threshold' 라고 했을 때, Threshold가 10으로 주어지면 H에 해당하는 submatrix 부분은 잘 찾아낸다. 하지만 Threshold가 13일 경우, 왼쪽 여자의 쇄골 부분도 얼굴로 감지 된다. 15이상으로 주어질 경우에는 왼쪽 여자의 얼굴도 감지되지만, 쇄골 주변부도 감지되는 단점이 있다. 그 이상의 값이 주어질 경우, 다른 얼굴도 찾아내지만 얼굴이 아닌 부분도 다수 검출된다. 얼굴이 아닌 부분의 ED 값이 12이상의 실수로 계산되기 때문에 threshold로 upper bound, lower bound를 지정하여도 얼굴만 검출되도록 할 수는 없는 것 같다. Threshold를 18로 고정하고, nms함수에 주어지는 실수 parameter를 0.05정도로 계산하면 그나마 나은 퍼포먼스를 보여준다.

따라서 Threshold가 높을수록 face detection에 대한 허용도가 높아져, 얼굴이 아닌 부분도 얼굴로 감지하기 때문에 적절한 threshold를 찾는 것이 성능을 높이는 방법일 것이다.

[4] Describe the procedure of NMS (non-maximum suppression) in this example. (얼굴 검출에서 NMS 사용 이유 및 동작 원리 간단 설명)

해당 과제에 주어진 코드에서, 얼굴 검출을 위해 Cross-Correlation이나 ED를 통해 ret 행렬을 계산한다. ret행렬은 특정 부분이 얼굴일 가능성 또는 거리 정보를 담고 있는데, 필연적으로 얼굴 주위에 많은 후보 값들이 생기게 된다. 그럴 경우 얼굴 주위에 여러 bounding box가 생겨서 제대로 된 성능을 내지 못한다. 따라서 얼굴 주위의 여러 후보군 들 중 하나만 남기고 나머지는 드랍하기 위해서 NMS를 사용한다.

NMS는 여러 후보군 들 중에서 가장 높은 가능성(또는 confidence)를 가진 후보를 선택하여, 이를 모든 후보군 들과 비교하는 과정을 거친다. 이 때, 가장 높은 가능성을 가진 후보와 영역(area)이 겹치는 정도 (overlap 변수)가 충분히 큰 후보는 후보군에서 제외하고, 이를 반복실행하여 최종적으로 후보 가능성이 높은 영역 들만 남게 되는 원리를 가진다.

이번 과제에 주어진 NMS 함수에서도 이와 똑같은 로직을 수행한다. 덕분에 최종 결과에서 우리는 redundant 한 bounding box가 제외된 최종 결과를 얻을 수 있다.

[5] Describe the fundamental limitation of this method and propose how to modify the current framework for detecting another face.

* Do not consider using deep learning and boosting theory. Do not consider adding more templates or face instances.

(딥러닝이나 부스팅 방법과 같은 얼굴 검출용 알고리즘을 사용하지 않고 현재 있는 코드에서 다른 얼굴을 검출할 수 있는 방법론을 제시하시오. 단, 등록 얼굴은 주어진 1개 faceimage.jpg로 가정하고 복수 얼굴을 등록하지 않고 검출 방법은 고민해 볼 것.)

Cross-Correlation이나 ED 방식 만을 사용하면 사람의 피부색이나 피부 컨디션 등에 취약할 수 있다. 또한 illumination 등에도 큰 영향을 받기 때문에 바람직한 알고리즘이라고 할 수 없다. 게다가 이번 과제에서는 커널 이미지의 사이즈 또는 orientation이 고정되었기 때문에, 인식 대상의 위치나 고개 각도가 조금만 달라도 인식하지 못할 가능성이 크다. 게다가 커널 이미지가 단순 사람 한 명을 crop 한 것에 불과하기에 사람의 표정이나 이목구비에 따라 인식 가능성이 크게 좌우된다는 단점도 있다.

이를 해결하기 위해서는 얼굴 인식에 필요한 정보만 남기고, 나머지 redundant 한 정보는 제거해야 할 것이다. 가령 사람의 피부색, 주름, 주근깨와 여드름 등이 해당한다. 이들 정보를 제거하면, blur 처리된 이미지의 edge에 해당하는 정보만 남게 된다. 이를 푸리에 변환하여, frequency domain에서 Cross-Correlation 연산 하는 방법을 구상해 보았다. 이를 식으로 표현하면 다음과 같다.

$$\text{ret}(y, x) = \text{FFT}^{-1}\{\text{FFT}(H) \otimes \text{FFT}(F[y - k, y + k : x - k, x + k])\}$$

위와 같은 연산을 진행하면 과제에서와 같이 ret 행렬을 얻을 수 있다. ret 행렬을 토대로 threshold 이상의 값들에 대해 NMS를 사용하면 얼굴에 해당하는 부분을 감지할 수 있을 것으로 예상된다.