**고급 소프트웨어 실습 보고서**

**9 주차**

**4반**

**20131540**

**김준호**

**숙제 (1) 투영에 사용되는 eigenvector의 개수를 나타내는 c 값을 {50, 100, 200, 300}와 같이 변화시키면서 각각의 인식률을 구해보자. 상위 몇 개의 eigenvector를 제거하고도 인식률을 구해본다.**

1) c의 값을 {50, 100, 200, 300}으로 변화시킬 때 각각의 인식률

(상위 eigenvector들을 보존시킴)

예상되는 결과 :

비교하는 특정vector의 개수가 많아질수록 인식률이 높아질 것으로 예상된다. 상식적으로 두 얼굴을 비교하는데 비교하는 특성이 많을수록 인식률이 좋을 것이기 때문이다. (예를 들어 눈의 크기만 비교한다면 같은 사람인지 인식할 확률이 크게 떨어질 것이다. 하지만 눈의 크기, 눈의 위치, 코의 크기, 코의 위치 등을 비교한다면 인식률이 증가할 것이다. 이러한 비교는 다음과 같이 각 벡터의 거리계산으로 한다.)



나타난 결과 :

실제로 c값을 변화시키면서 나타나는 인식률을 보면 다음과 같다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C | 50 | 100 | 200 | 300 |
| 인식률 | 68% | 74% | 76% | 74% |

c의 값이 50, 100, 200으로 증가할 때 예상과 같이 인식률이 증가하는 것을 볼 수 있다. 즉 비교하는 벡터의 개수가 증가할수록 인식률이 증가한다. 하지만 C가 200에서 300으로 증가하자 인식률이 떨어진다. 왜 이런 일이 발생하는지 생각해보았다.

직관적으로 생각하면, 얼굴을 비교하기 위해 사용하는 벡터가 어느 정도의 임계 값을 넘으면 뒤로 갈수록(즉 분산이 작은 벡터일수록) 얼굴을 비교하는데 불필요한 벡터가 존재함을 알 수 있다. 위의 테스트에선 그 임계 값이 200개와 300개 사이에 있는 것을 알 수 있다.

2) 상위 eigenvector를 제거한 인식률

예상되는 결과 :

상위에 존재하는 eigenvector일수록, 즉 분산이 큰 eigenvector일수록 얼굴 비교에 critical한 영향을 미칠 것 같다.(분산이 크다는 것은 그 특성이 사람 얼굴마다 다양하게 갖을 수 있는 특성이라는 것이다.) 따라서 상위 eigenvector를 제거하면 제거할수록 인식률이 떨어질 것으로 예상된다.

나타난 결과 :

2-1)C = 50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 제거한 수 | 1 | 2 | 5 | 10 | 20 |
| 인식률 | 70% | 70% | 64% | 62% | 70% |

2-2)C = 100

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 제거한 수 | 1 | 2 | 5 | 10 | 20 |
| 인식률 | 74% | 74% | 74% | 72% | 74% |

2-3)C = 200

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 제거한 수 | 1 | 2 | 5 | 10 | 20 |
| 인식률 | 76% | 76% | 76% | 76% | 74% |

2-4)C = 400

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 제거한 수 | 1 | 2 | 5 | 10 | 20 |
| 인식률 | 74% | 74% | 76% | 74% | 74% |

1. 번의 결과에서 가장 큰 인식률을 갖은 C = 200일 때의 상황을 보자.

제거한 eigenvector가 많아질수록 인식률이 떨어질 줄 알았는데 제거한 수가 1개부터 10개까지의 인식률은 76%로 같고, 20개를 제거하자 인식률이 떨어진 것을 확인할 수 있었다. C = 400일때에는 상위 vector를 2개에서 5개로 더 제거했는데 오히려 인식률이 증가한 것을 확인할 수 있었다. 왜 이런 경우가 발생하는지 생각해보았다. 우선, 상위 eigenvector들의 각 분산이 별 차이가 없고, 그러한 eigenvector들이 상위에 많이 존재한다면, 다시 말해 상위 eigenvector들의 분산(각 특성의 분산이 아니다.)이 크지 않다면 상위 몇 개를 제거한다고 해도 인식률에 큰 영향을 주지 않을 수 있을 것 같다. 물론 C가 작으면, 상위 몇 개를 제거하는 것이 critical할 수 있다. (C가 50일 때 제거한 수가 2에서 5로 증가할 때 인식률이 크게 떨어진 것을 확인할 수 있다.) 두 번째 경우는, 앞서 하위 eigenvector들이 얼굴 비교에 사용할 필요가 없음을 확인하였는데, 반대로 상위 eigenvector들이 얼굴 비교에 사용할 필요가 없을 수 있다는 것이다. 따라서, 각 특성의 분산이 크거나 작음이 그 특성이 얼굴비교에 효과적인지 아닌지를 판단할 수 없을 것 같다.