

박지성과 유해진의 이미지 분류

김새봄(nosnova623@gmail.com)

오우창(woochang.oh90@gmail.com)

지승배(seungbae.ji@gmail.com)

Why?

외국인은 알아보기 힘든 박지성과 유해진, 컴퓨터도 헛갈릴까?



맨유 팬 페이지에 올라온
박지성 스캔들 제보



이태원 클럽에서 외국인들이
박지성으로 오해

Random forest와 SVM으로 박지성과 유해진의 이미지를 분류

How?

사진 데이터를 어떻게 수집할 것인가?



Train data 각각 500장, Test data 각각 15장

구글 이미지 검색 및 예능, 영화, 인터뷰 등의 영상 캡처로 수집

How?

사진 데이터의 전처리는 어떻게 할 것인가?



이미지 캡처



100 x 100

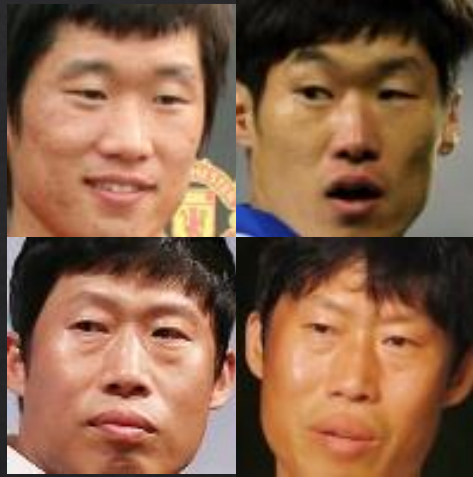


70 x 100

분석에 필요한 부분만을 담기 위해 사진 전처리 진행

How?

사진 데이터를 어떻게 분석에 이용할 것인가?



인물 이미지에서
얼굴 부분 추출

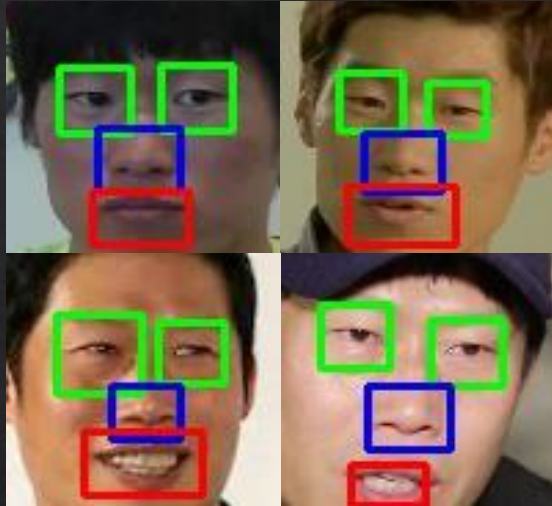


얼굴 부분을 추출한
이미지의 흑백 전환

변환한 이미지만으로 이미지 인식 성능을 확인

How?

사진 데이터를 어떻게 분석에 이용할 것인가?



OPEN CV를 이용한
눈코입 인식

눈간 거리
미간 코 거리
입의 폭
코의 폭
코볼의 크기

얼굴의 특징을
수치화하여 변수로 활용

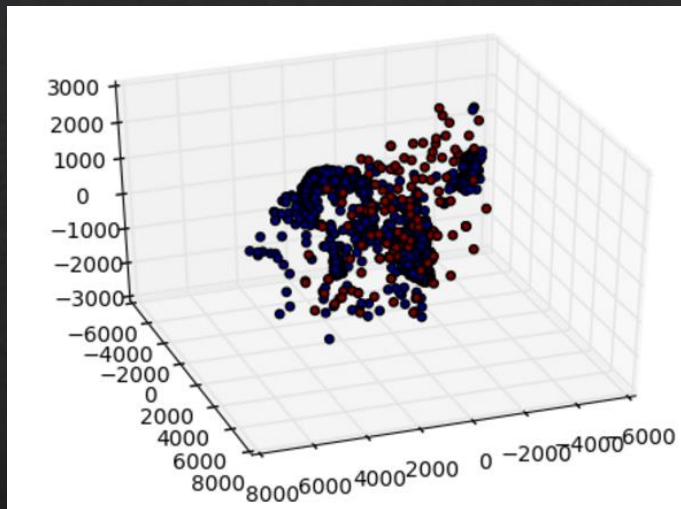
영상처리 한 이미지로 추출한 얼굴의 특징으로 분류 성능을 확인

How?

사진 데이터를 어떻게 분석에 이용할 것인가?



인물 이미지에서
얼굴 부분 추출



PCA를 통한
차원 축소



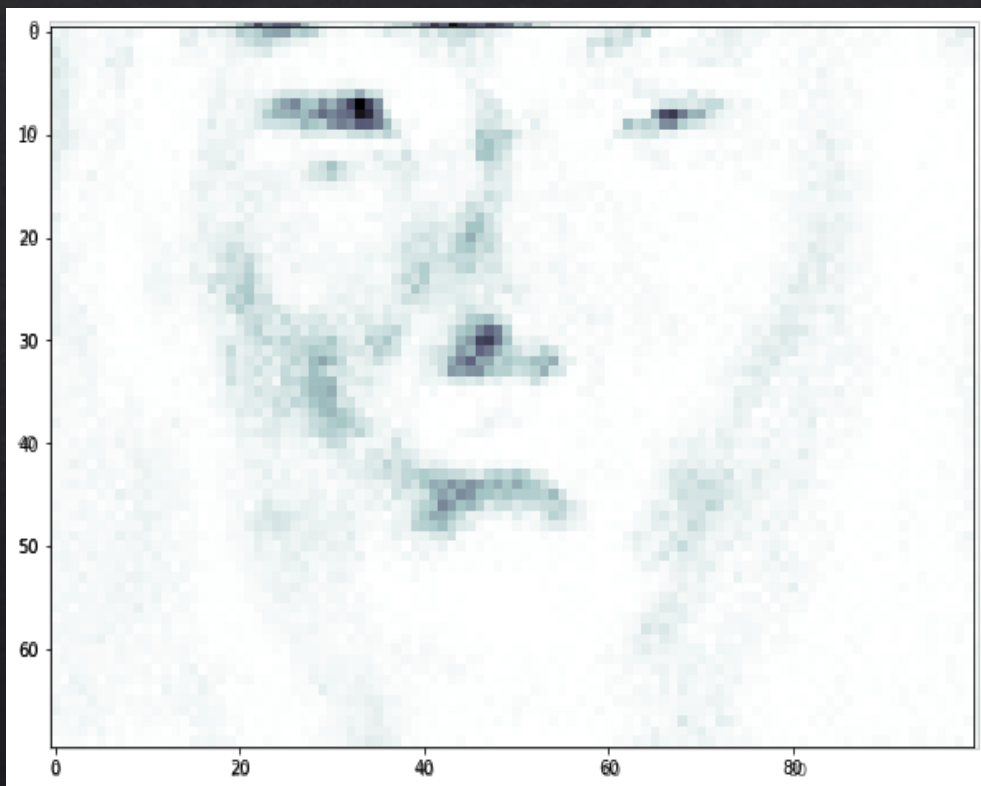
```
array([[ -1927.93851083, -1508.11067897, -659.2340294 ],  
       [ -1782.64675741, -1421.66494617, -670.77038445],  
       [   606.28427219, -617.71762618, -42.48399241],  
       ...,  
       [    75.59078641,  1318.25978759,   51.18416155],  
       [ -1530.28095314,  3293.23295304,   747.60915363],  
       [   975.013564 ,  1790.69821493,  1276.76620169]])
```

PCA로 얻은
변수 사용

컴퓨터가 스스로 추출한 얼굴의 특징으로 두 사람의 닮은 정도를 확인

Result!

Original 이미지를 통한 Random Forest

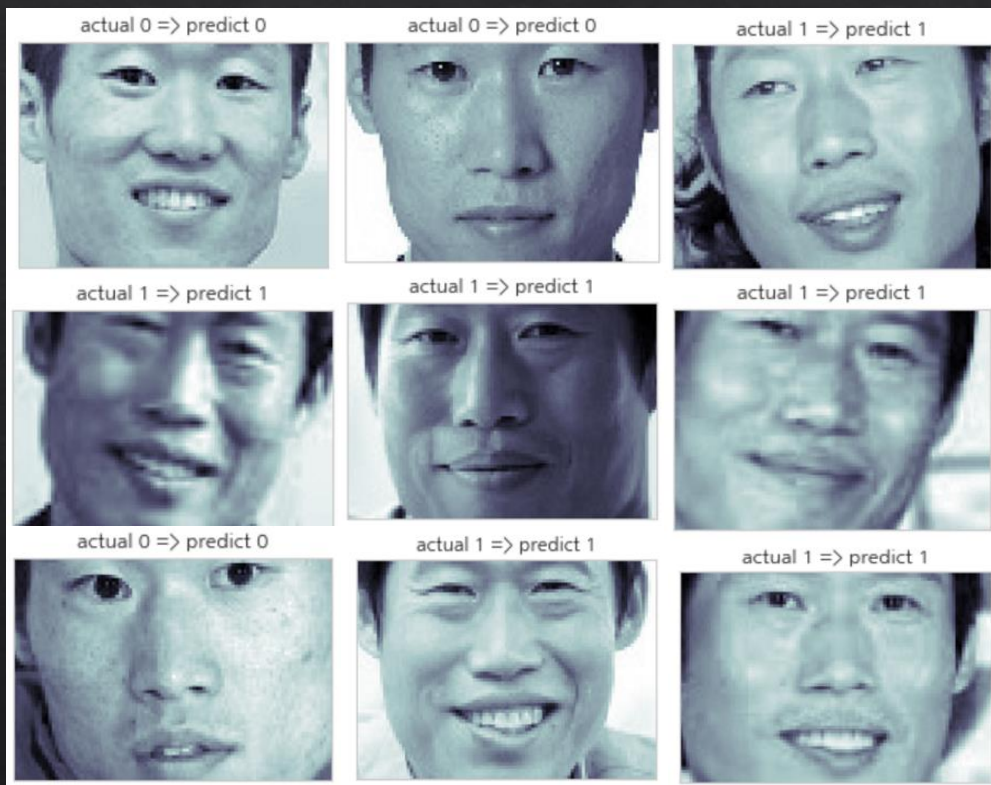


	precision	recall	f1-score	support
0.0	1.00	0.27	0.42	15
1.0	0.58	1.00	0.73	15
avg / total	0.79	0.63	0.58	30

Random forest로 분석한 feature importance 시각화

Result!

Original 이미지로 Support Vector Machine

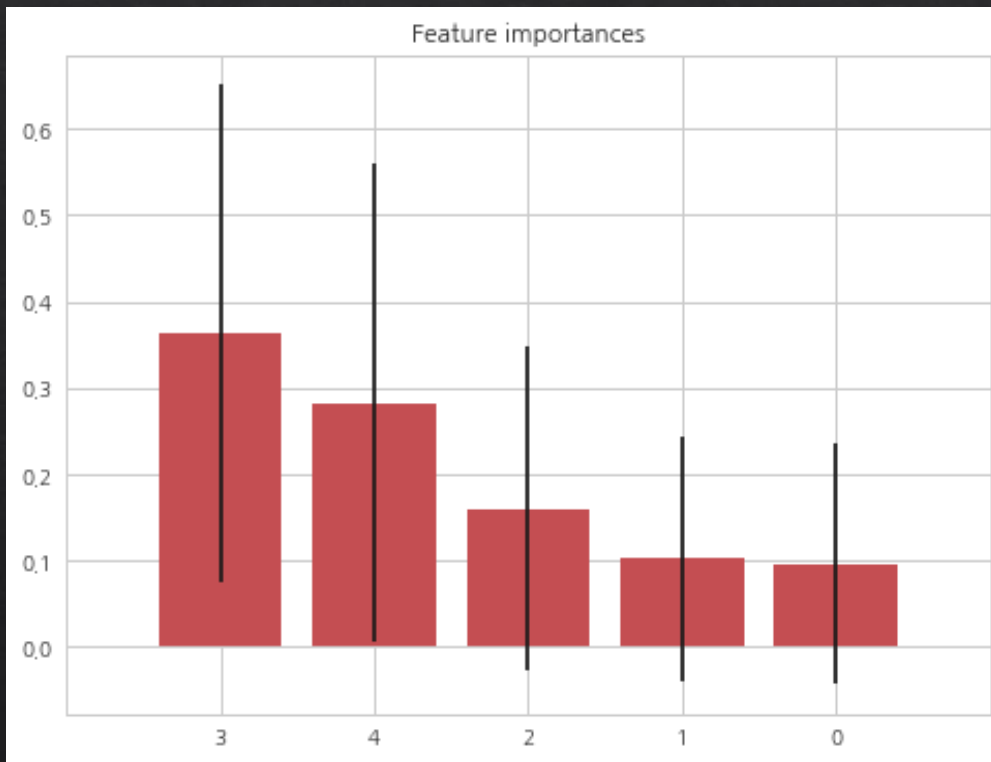


	precision	recall	f1-score	support
0.0	1.00	0.80	0.89	15
1.0	0.83	1.00	0.91	15
avg / total	0.92	0.90	0.90	30

높은 정확도를 보이는 SVM 분류 결과

Result!

영상처리로 얻은 변수를 통해 Random Forest



코 폭(f3): 0.362667

입 폭(f4): 0.281908

콧볼 폭(f2): 0.159335

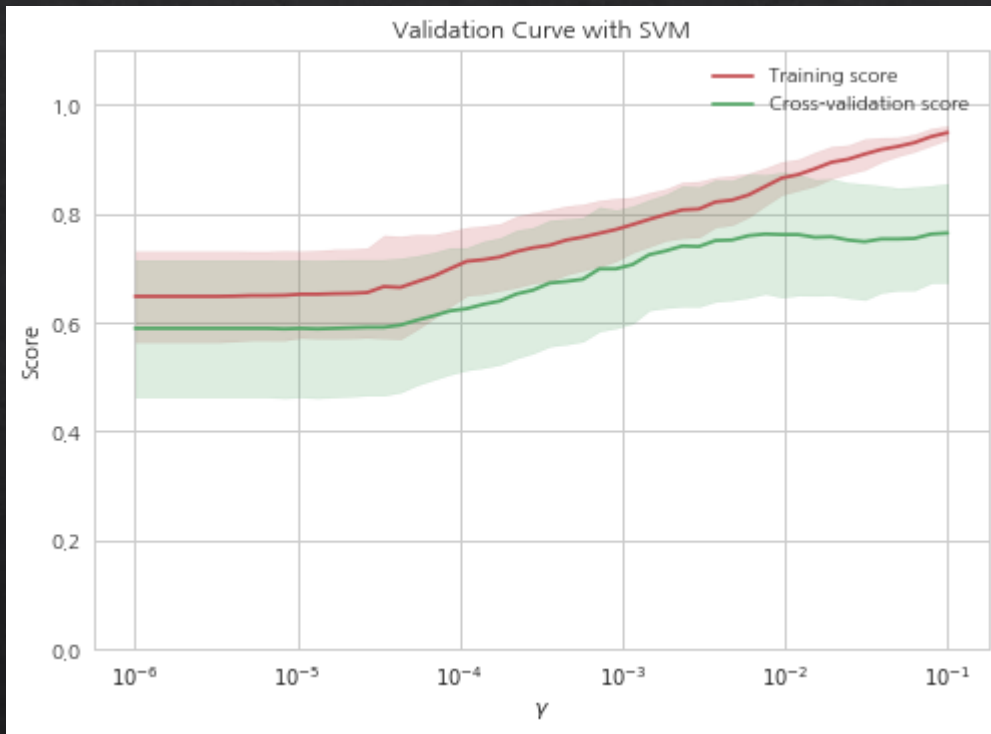
미간-코 거리(f1): 0.101429

눈간 거리(f0): 0.094662

영상처리로 추출한 변수를 random forest로 분석한 feature importance 결과

Result!

영상처리로 얻은 변수를 통해 Support Vector Machine



	precision	recall	f1-score	support
0	0.70	0.47	0.56	15
1	0.60	0.80	0.69	15
avg / total	0.65	0.63	0.62	30

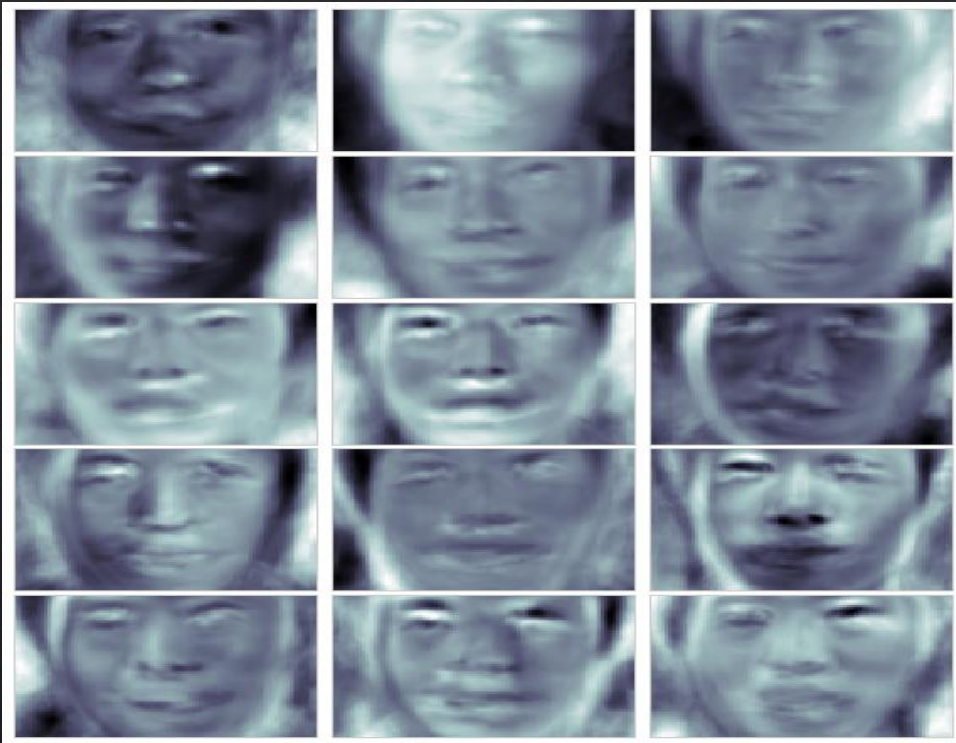
	precision	recall	f1-score	support
0	0.94	0.68	0.79	500
1	0.75	0.96	0.84	500
avg / total	0.84	0.82	0.81	1000

```
0.714  
{'clf__gamma': 0.10000000000000001, 'clf__c': 0.10000000000000001, 'clf__kernel': 'rbf'}
```

grid search로 SVM kernel의 종류와 파라미터 값을 확인

Result!

PCA 된 이미지로 Support Vector Machine



eigenface

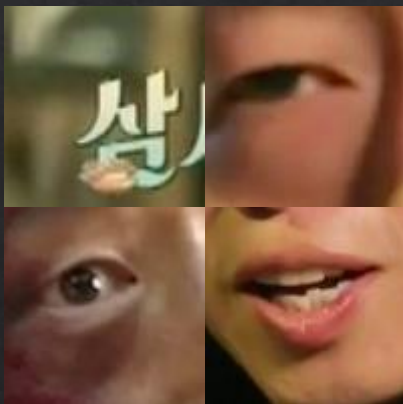
	precision	recall	f1-score	support
0.0	0.76	0.96	0.85	500
1.0	0.94	0.70	0.80	500
avg / total	0.85	0.83	0.83	1000

	precision	recall	f1-score	support
0.0	0.40	0.53	0.46	15
1.0	0.30	0.20	0.24	15
avg / total	0.35	0.37	0.35	30

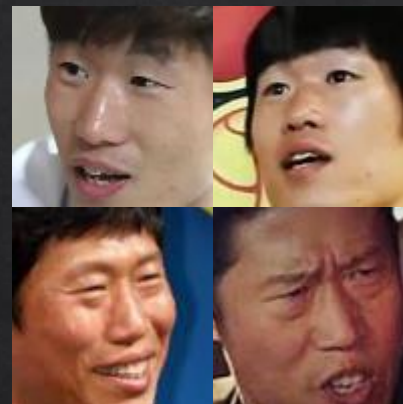
PCA를 통한 박지성과 유해진의 유사도 확인

discussion

한계점 1 – 이미지 추출 및 눈, 코, 입 인식의 어려움



얼굴 인식 및
이미지 추출의 어려움



측면 사진의 유의미한
데이터 추출 어려움

양질의 결과를 위해선 많은 데이터와 데이터의 품질이 중요하다

discussion

한계점 2 – 고차원 데이터 해석의 한계

Random forest의 결과는 모방이 가능했지만,
SVM에서는 특징의 해석에 어려움을 겪음

고차원 데이터 해석을 위한 수치화 및 시각화 노력의 필요

discussion

한계점 3 – 기술적인 한계로 더 자세한 분석에 어려움

광대에 대한 특이점이 보였지만,
분석에 필요한 수치적 도출에 한계
측면 사진을 분석에 활용하기 어려움

앞으로 관련 연구가 더욱 필요할 것으로 보임

Conclusion

R1.

Random forest를 통해
두 사람의 외모적 특징을 시각화 할 수 있다!
(눈, 코, 입, 광대 등의 차이점 발견)

Conclusion

R2.

**PCA로 박지성과 유해진의 사진을 확인한 결과
두 사람은 닮았다는 구체적인 증명 가능**

유해진이 박지성보다 특징을 더 가지고 있음

Conclusion

R3.

**박지성과 유해진의 정면 사진을 이용할 경우
SVM 만으로도 90% 이상 구별이 가능하다!**

Conclusion

R4.

박지성과 유해진을 구별하는 중요한 특징은
코의 형태, 미간과 코의 거리,
눈과 눈의 거리, 입의 형태 순이다!

감사합니다