얼굴 이미지 검색을 위한 Product Quantization 기반의 깊은 신경망 피쳐 매칭

*장영균 **이석희 ***조남익

서울대학교

*kyun0914@ispl.snu.ac.kr **seokheel@ispl.snu.ac.kr ***nicho@snu.ac.kr

Pedestrian Detection using YOLO and Tracking

*Jang, Young Kyun, **Lee, Seok Hee ***Cho, Nam Ik

Seoul National University

요약

최근 딥 러닝을 이용한 방법들이 이미지 분류에서 뛰어난 성능을 보임에 따라, 컴퓨터 비전의 중요한 문제 중 하나인 이미지 검색에도 이를 활용하고 있다. 특히, 이미지 검색에 사용할 수 있는 이미지 기술자 (Image descriptor)를 깊은 신경망 구조의 일부분인 Fully-connected layer에서 추출하여 사용하는 방법들이 제시되고 있고, 이를 위해 알맞은 목적함수를 설계하여 깊은 신경망을 학습하는 것이 중요해지고 있다. 딥 러닝을 통해 얻은 이미지 기술자는 실수형 데이터로서, 한 장의 이미지를 수치화하여 표현하는 데 많은 메모리를 소모하게 된다. 이를 보완하기 위해 이미지 기술자를 작은 용량의 이진코드로 mapping 하는 해싱 (hashing) 이라는 과정이 필수적이나 이에 따른 한계점이 발생한다. 본 연구에서는 실수형 데이터가 갖는 거리 계산에서의 이점과 이진코드의 장점을 동시에 살릴 수 있는 Product Quantization 방식의 이미지 검색 방법을 이용하여 한계점을 극복하였다. 우리는 제안한 방법을 얼굴 이미지 데이터 셋에 실험하였고 기존 방식보다 뛰어난 성능을 보이는 것을 확인할 수 있었다.

1. 서론

멀티미디어와 소셜 미디어가 발전함에 따라, 매일 수많은 이미지가 새롭게 생성되고 있다. 이 중 얼굴 이미지 차지하는 비중이 상당히 높으며, 다른 이미지들과는 달리 얼굴 이미지가 갖는 고유한 특징 때문에 얼굴 이미지 검색은 어려운 문제로 자리잡고 있다. 효율적인 이미지 검색을 위해 이미지에서 이미지를 대표할 수 있는 짧은 코드를 추출할 필요가 있는데, 대표적으로 SIFT (Scale Invariant Feature Transfrom) [1]나 SURF (Speed-Up Robust Features) [2]와 같은 이미지 기술자를 사용할 수 있다. 그러나 이러한 이미지 기술자들은 Grayscale 이미지에서 추출한 것이어서 RGB각 채널의 고유한 특징을 반영하지 못한다는 단점이 있고, 깊은 신경망의 지도학습에 사용되는 이미지의 레이블 값을 활용하지 않는다는 한계를 가진다. 따라서 본 연구에서는 이미지의 레이블 값을 활용하여 깊은 신경망을 학습하고, 학습된 신경망의 Fully-connected layer 부분에서 이미지 기술자를 추출하는 방식을 사용하였다.

일반적인 깊은 신경망의 Fully-connected layer를 사용하여 이미지 기술자를 생성하면, 이미지를 잘 구별하면서 동시에 길이도 짧은 이미지기술자를 얻을 수 있으나, 이미지의 수가 늘어나면 이를 활용하여 만든 검색 데이터베이스마저 메모리의 크기가 수십 기가바이트를 넘을 수 있다. 이에 대한 해결책으로 여러 방법이 제안되었는데, 우리는 그 중 실수형 이미지 기술자를 일정한 크기의 세그먼트로 나눈 뒤, 대표 값으로 이루어진 룩업 테이블에서 가장 거리가 가까운 값에 대응시켜 이진 코드로 전환하는 Product Quantization 방식을 활용하였다.

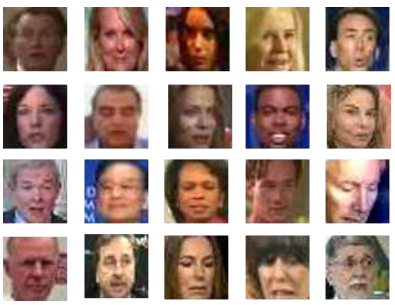

그림 1 YouTube Faces 데이터 셋

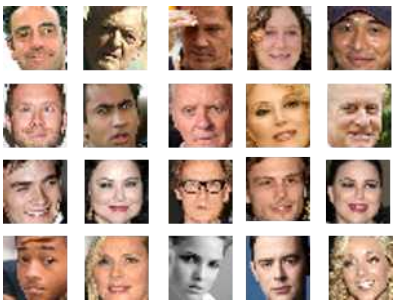

그림 2 FaceScrub 데이터 셋

그림 1과 2는 각각 본 연구에서 활용한 두 가지 얼굴 이미지 데이터 셋의 예시이다. 동일한 사람의 얼굴에 같은 레이블이 할당되어 있고, YouTube Faces 데이터 셋의 경우 1,595명의 사람에 대해 총 63,000장을 트레이닝과 검색 데이터 베이스를 만드는데 사용하였다.

FaceScrub 데이터 셋의 경우 530명의 사람에 대해 총 63,717장을 트레이닝과 검색 데이터 베이스를 만드는데 사용하였다. 테스트에는 두 데이터 셋 모두 한 사람당 5장의 이미지를 사용하였다. 이미지 검색의 결과로는 테스트 이미지와 데이터 베이스 내부의 이미지들 사이의 거리 계산을 통해 얻은 Ranked-list를 출력한다.

본 논문에서는 얼굴 이미지 검색을 효과적으로 할 수 있는 방법을 제안하고자 한다. 먼저, 이미지의 레이블을 활용하여 분류를 위한 깊은 신경망을 학습하고, Fully-connected layer에서 분류성능이 뛰어난 이미지 기술자를 추출한다. 이를 빠른 이미지 검색에 활용하기 위해 Product Quantiztaion 방식을 응용하여 이미지 기술자를 이진 코드로 mapping한다. 이진 코드로 mapping된 이미지 기술자를 데이터 베이스에 저장해두고, 테스트용 이미지와의 거리 계산을 통해 Ranked-list 결과를 얻는다. YouTube Faces와 FaceScrub 데이터 셋에 대해 제안한 방법을 실험해본 결과, 기존의 방법들에 비해 뛰어난 성능을 보이는 것을 확인할 수 있었다.

2. Product Quantization을 활용한 이미지 검색

최근 뛰어난 분류성능을 보이는 다양한 형태의 깊은 신경망 구조가 제안되었는데, 그중 구조를 쉽게 변환시킬 수 있는 VGG-16 [3] 수정하여 사용하였다. 얼굴 이미지는 scale에 따라 구별되는 특징을 가지고 있기에 마지막 3개의 Fully-connected layer를 수정하여 첫 번째 Global Average Pooling으로 대체하여 multi-scale에 대처할 수 있도록 하였고, 나머지 2개의 Fully-connected layer들은 고정된 크기의 이미지 기술자를 생성할 수 있도록 수정하였다. 간략한 구조는 아래의 그림3과 같다.

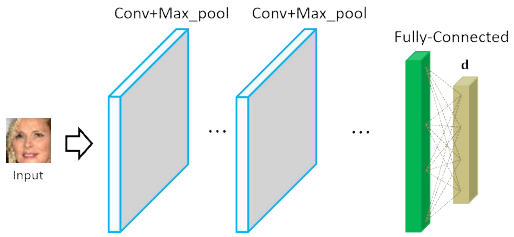

그림 3 깊은 신경망 구조

이렇게 생성된 이미지 기술자: **d**를 이진 코드로 mapping 하기 위해 Product Quantization 기법을 활용하였고, 이를 위한 Quantization 테이블을 구성하였다. 테이블은 K개 열과 M개의 행으로 이루어진 행렬의 형태로, K개의 코드북과 M개의 코드워드로 이루어진 구조를 가지고 있다. 코드워드의 길이와 K의 곱은 이미지 기술자의 길이와 동일하다. 테이블과 이미지 기술자 간의 매칭을 위해 이미지 기술자를 코드워드의 길이로 잘라 이미지 기술자를 K개의 세그먼트로 만들고 i번째 세그먼트와 i번째 코드북 내의 코드워드 간에 유클리디안 거리를 계산한다. 거리가 가장 가까운 코드워드의 색인 값을 이진 코드로 변환하여 이미지 기술자를 이진 코드의 조합으로 mapping할 수 있게 된다. 이에 대한 모식도는 아래의 그림 4와 같다.

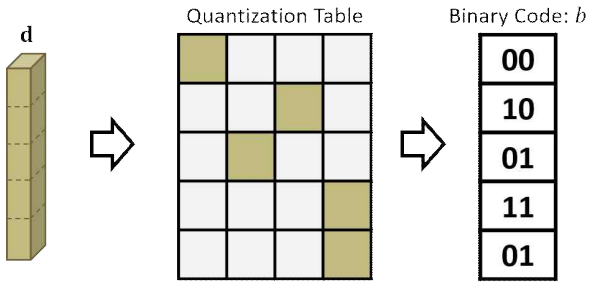

그림 4 이진 코드 mapping 과정

Quantization 테이블은 [4]에서 제안된 Soft Quantization방식을 적용하여 학습하였다. Quantization된 이미지 기술자가 cross-entropy 목적함수를 활용하여 이미지 레이블에 따라 구분이 잘되도록 깊은 신경망을 학습하였고, 테이블 또한 이에 영향을 받아 학습된다. 이에 따른 결과로 이미지 기술자가 각 레이블의 정보를 효과적으로 담아낼 수 있게 되었고, Quantization 테이블은 이미지 분포의 대푯값 들의 집합으로 구성된다. 그리고 각 코드북은 이미지 기술자의 세그먼트들의 서브 스페이스를 나타내게 되고, 이미지 기술자 스페이스를 서브 스페이스들의 곱집합의 형태로 표현하게 되어 단일 스페이스로 구성된 것보다 다양한 대푯값 들을 담아낼 수 있게 된다.

이진 코드를 직접 이미지 검색을 위한 계산에 사용하면, 검색 속도에는 실수형 데이터를 통한 계산보다 이점이 있을 수 있으나, 이미지 기술자 간의 거리의 표현력에는 제한이 생기게 된다. 따라서, 이진 코드와 Quantization 테이블을 동시에 데이터 베이스에 가지고 있으면서, 이진 코드는 Quantization 테이블의 코드워드를 부르는데 사용한다. 실수형 값으로 이루어진 코드워드와 쿼리 이미지 (테스트 이미지)에서 추출된 실수형 이미지 기술자의 세그먼트 사이의 거리 계산을 통해 최종 Ranked-list를 얻을 수 있게 된다.

3. 실험 결과

[5]에서 제안된 얼굴 이미지 검색 프로토콜에 따라, 이미지에서 추출한 이미지 기술자를 12/24/36/48 비트의 이진 코드에 mapping하여 실험을 진행하였다. 제안된 방법의 성능 평가 방식으로는 이미지 검색 분야에서 많이 쓰이는 mean Average Precision (mAP)을 사용하였다. 이는 연관된 (relevant, 같은 레이블) 이미지가 검색된 경우 정답으로 간주하며, 데이터 베이스의 모든 이미지에 대해 결과를 계산하는 방식으로 이미지 검색 성능을 객관적으로 드러낼 수 있는 지표이다.

비교 실험을 위해 기존의 얼굴 이미지 검색 방법들, DHCQ: Deep Hashing based on Classification and Quantization errors [5], DDH: Discriminative Deep Hashing [6], DDQH: Discriminative Deep Quantization Hashing [7]을 실험에 사용하였다. 그 결과로 우리가 제안한 방식이 얼굴 이미지 검색 성능에 있어서 모든 길이의 비트 수에 있어서 가장 뛰어난 성능을 보인다는 것을 확인할 수 있었다.

방법	mAP			
	12-bit	24-bit	36-bit	48-bit
DHCQ	0.2511	0.4872	0.6809	0.7592
DDH	0.4025	0.8224	0.8556	0.9011

DDQH	0.6332	0.9652	0.9801	0.9834
Proposed	**0.9853**	**0.9912**	**0.9930**	**0.9942**

표1. YouTube Faces 데이터 셋에 대한 검색 성능

방법	mAP			
	12-bit	24-bit	36-bit	48-bit
DHCQ	0.1866	0.2201	0.2368	0.2816
DDH	0.0652	0.1203	0.1462	0.1898
DDQH	0.1186	0.2673	0.3512	0.4625
Proposed	**0.6858**	**0.7345**	**0.7929**	**0.8321**

표2. FaceScrub 데이터 셋에 대한 검색 성능

4. 결론

본 논문에서는 효과적이고 빠른 얼굴 이미지 검색 위한 방법으로 Product Quantization 방식의 이진 코드 mapping 과정을 깊은 신경망 구조에 제안하였다. 그 결과 실수형의 이미지 기술자의 특징을 활용하여 이미지 간의 거리를 계산할 수 있었고, 기존의 방식들에 비해 더 좋은 검색 성능을 보여주었다. 추가적으로, 목적함수로 cross entropy가 아닌 다른 Metric learning 기법을 활용하면 더 높은 얼굴 이미지 검색 성능을 얻을 수 있을 것으로 기대된다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT연구센터육성지원사업의 연구결과로 수행되었음
(IITP-2018-2016-0-00288)

참고문헌

[1] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." , IEEE Transactions on pattern analysis and machine intelligence 24.7, 2002

[2] H. Bay et al., “Speeded-up robust features (SURF),” Computer Vision and Image Understanding, 2008.

[3] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015

[4] T. Yu, J. Yuan, C. Fang, and H. Jin. Product quantization network for fast image retrieval. In ECCV, 2018..

[5] J. Tang, Z. Li, and X. Zhu. Supervised deep hashing for scalable face image retrieval. Pattern Recognition, 75:25 - 32, 2018.

[6] Lin, J., Li, Z., Tang, J.: Discriminative deep hashing for scalable face image retrieval. In IJCAI, (2017)

[7] J. Tang, J. Lin, Z. Li, and J. Yang. Discriminative deep quantization hashing for face image retrieval. IEEE Transactions on Neural Networks and Learning Systems, 2018.