

이미지 검색을 활용한 자가촬영 사진의 배경 합성

윤웅직^o, 윤성의⁺

^o KAIST 웹사이언스공학 전공

⁺ KAIST 전산학부

woongjick.youn@kaist.ac.kr, sungeui@gmail.com

요 약

본 연구에서는 기존의 사진 합성방식과 다르게 대용량 이미지 데이터셋을 기반으로 자가촬영 사진의 주변 배경을 확장해 나가며 합성하는 방식을 제안하였다. 근사 근접 이웃 기반의 프러덕트 양자화를 이용해 이미지 기술자를 양자화하여 유사한 사진을 검색하였고, 위치 정보를 활용하기 위한 블록 형식의 격자로 기술자에 가중치를 주었다. 이러한 방식은 색상뿐만 아니라 사진의 질감 또한 고려하여 원본 사진과 유사한 사진의 검색이 가능하고, 유사한 사진에 포아송 이미지 에디팅을 사용하여 반복적으로 주변 사진을 합성하여 결과물을 생성한다. 이를 통하여 이미지 검색 후 합성을 사용하여 영상 정보를 얻을 수 없는 주변 영역으로 확장시켜 나갈 수 있음을 확인하였다. 또한 사용자는 기존 원본 사진보다 몇 배 확장 할 지만 결정하면 기존의 사진 합성 방식처럼 합성할 영역을 특정하지 않아도 새로운 사진을 생성하는 장점이 있다.

1. 서론

최근 스마트기기 보급의 증가로 Flickr, Instagram, Facebook 같은 다양한 소셜네트워크 서비스 (SNS) 의 이용자가 급증하고 있다. 이들은 스마트기기로 촬영한 사진을 소셜네트워크 서비스에 실시간으로 업로드 하며 전세계를 연결하는 온라인 네트워크를 형성한다. 이용자들은 음식, 풍경 같은 사진 뿐만 아니라, 주로 자신의 컴퓨터에 보관하던 자가촬영 사진 (selfie) 같은 개인적인 사진을 온라인에 업로드 하는 추세이다. 또한 셀카봉이나 어안렌즈 같은 자가촬영을 돕는 기기들이 출시됨에 따라 언제 어디서든 간편하게 촬영이 가능해졌다.

이러한 자가촬영 사진의 인기를 반영하듯 얼마 전 호주 관광청은 일본인 여행객들을 대상으로 원거리 자가촬영을 위한 'Giga selfie' 서비스를 제공하였다. 이 서비스는 약 110m 떨어진 초고화질 카메라가 이용자의 사진을 찍어주는 것인데, 주변 풍경 뿐만 아니라 확대하여 이용자의 얼굴을 확인할 수 있다. 일반적인 자가촬영으로는 찍을 수 없는 사진을 손쉽게 제공하여 온라인에서 큰 화제가 되었다. 기존의 자가촬영 사진에 다른 배경 사진을 합성하여 단순히 큰 이미지를 만들어 내는 것이 기술적으로 가능하긴 하지만 여러 제약이 따른다.

영상처리분야에서 사진 합성은 오랜 기간 연구되어 온 분야이다. 이전에는 주로 사진 속의 특정 영역을 제거하고 다른 사진으로 대체하는 방식을

이용한 사진합성 연구가 이루어졌다 [1, 2, 3, 4]. 하지만, 최근에는 소셜네트워크 서비스를 이용한 사진 데이터의 수집이 용이해짐에 따라, 대용량 데이터를 활용한 내용기반의 사진 합성 연구가 많이 이루어지고 있다 [5, 6]. 내용기반의 사진 합성을 위해서는 각 사진에서 색조나 계조 등의 색상정보를 추출하고 기계학습 기술을 활용한 대용량 이미지 데이터의 검색이 필수적이다. 본 논문에서는 대용량 이미지 데이터에서 근사 근접 이웃 검색 (Approximate nearest neighbor search)을 기반으로 이미지 검색을 활용하여 자가촬영 사진의 주변 배경을 합성하는 방법을 제안하고자 한다.

2. 관련 연구

그래프 컷 알고리즘[1]은 이미지 분할에서 널리 쓰이는 기술이다. 이 알고리즘은 이미지내의 모든 화소를 연결한 그래프로 구성된 에너지 함수를 최소화하여 분할하는 방식인데, 두 사진을 합성 할 때 계산된 비용을 최소화 하여 이질감 없는 결과물을 생성하는데 이용할 수 있다. 이와 함께 포아송 이미지 에디팅[2]을 주로 사용하는데, 이 알고리즘은 두 사진을 합성 할 때 겹치는 영역의 색상이나 밝기 차이로 인한 이질감을 없애기 위하여 두 영역의 계조를 고려하여 합성을 하는 방식이다. 이외에도 합성하고자 하는 영역 주변의 구조 정보를 활용하여 비슷한 구조의 패치를 합성하는 방식의 연구도

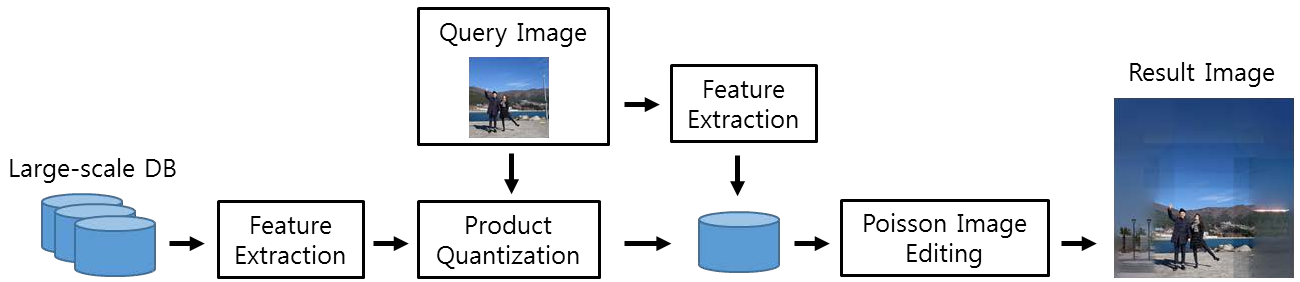


그림 1. 제안한 사진 합성 방식의 순서도

최근까지 진행되어 왔다 [3, 4, 6].

[5]에서 제안한 풍경 합성 방식은 여러 대의 머신을 이용하여 소셜네트워크 서비스 Flickr 를 통해 얻은 약 230 만장의 사진에서 GIST[7] 기술자를 추출하여 방대한 이미지 데이터를 수집하였다. 그 중 합성할 사진과 비슷한 200 장의 사진을 근접 질의 기반으로 검색하고, 이들을 그래프 컷과 포아송 이미지 에디팅 알고리즘을 이용하여 합성하였다. 이를 통하여 기존의 다른 합성 알고리즘처럼 복잡한 수식을 사용하지 않고도 방대한 이미지 데이터를 기반으로 자연스러운 사진을 생성해 낼 수 있음을 확인하였다. 하지만, 사진 검색을 위해 GIST 기술자만을 사용하기 때문에 비슷한 질감을 찾는 데 실패하는 경우가 발생 하고, 사용자가 합성하고자 하는 영역을 직접 설정하여야 하는 문제가 있다.

최근 대용량 이미지 데이터에서는 풍경 합성 방식[5]에서 사용한 근접 이웃 검색 (Nearest Neighbor search) 기법 대신 근사 근접 이웃 검색 방식이 주로 사용되고 있다. 이 방식은 가장 가까운 거리의 사진을 정확하게 찾는 근접 이웃 검색 방식보다 정확도가 떨어지지만, 근접 이웃을 추정하는 방식으로 더 빠르게 검색이 가능하다. 근사 근접 이웃 검색 방식은 대표적으로 해싱 (Hashing) 을 통한 이진 코드화 방식 [8, 9]과 양자화 (Quantization) [10, 11]를 통한 압축 방식이 연구되어 왔다. 그 중 프러덕트 양자화 (Product quantization)[10]는 고차원 벡터를 일정한 블록으로 차원을 구분하여 K-평균 군집 (K-means clustering) 알고리즘으로 양자화하여 데이터를 압축하는 기술이다.

3. 색상정보를 활용한 자가촬영 사진의 배경 합성 방법

3.1 장에서는 근사 근접 이웃 기반의 프러덕트 양자화를 이용하여 유사한 사진을 검색하는 방법에 대해 설명하고 3.2 장에서는 원본 사진의 주변 배경을 합성하는 방법에 대해 설명한다.

3.1 프러덕트 양자화를 이용한 사진 검색

본 논문에서는 [12]에서 제공하는 MIRFlickr-

25000 데이터셋을 사용하였는데, 이 데이터셋은 Flickr 에 업로드 된 25,000 장의 다양한 사진으로 이루어져 있다. 데이터셋에서 원본 사진과 확연히 다른 사진을 제외하기 위해서, 모든 사진에서 추출한 GIST 기술자와 원본 사진으로부터 추출한 GIST 기술자를 비교하였다. 2 단계의 재 검색 방식은 전체 GIST 기술자에서 검색하지 않고 일부를 제외하여도 검색 성능이 크게 떨어지지 않는다는 것이 기존 연구에 의해 알려져 있다 [13]. 본 논문에서는 이와 비슷한 방법으로 양자화하여 빠르게 검색하고자 하였다.

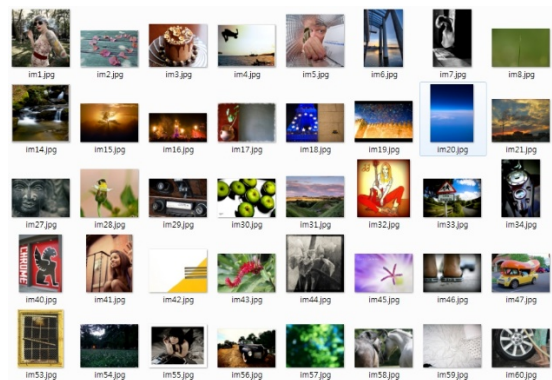


그림 3. MIRFlickr-25000 데이터셋의 예시

우선, 25,000 장의 모든 사진에서 각각의 GIST 기술자를 추출하고, 기술자의 차원을 8 개의 블록으로 구분하고 각각의 블록을 K-평균 군집 알고리즘으로 양자화하여 프러덕트 양자화를 학습하였다. 이 프러덕트 양자화에 모든 사진의 기술자를 할당해 양자화한 데이터셋과 원본 사진의 GIST 기술자를 비교하여 가장 근접한 5,000 장의 사진으로 데이터셋을 재구성 하였다. 이러한 근사 근접 이웃 기반의 검색은 정확도가 조금 떨어지는 문제가 있지만, 근접 질의보다 더 빠르게 검색이 가능하다는 장점이 있다. 원본 사진 주변에 합성을 하기 위한 사진을 검색하기 때문에 완전히 동일한 사진보다는 비슷하게 추정된 사진이 배경을 채우기에 더 좋을 것이라고 판단하였고, 이로 인하여 계산시간의 이득을 얻을 수 있었다.

3.2 패치 기반의 부분 영역 합성

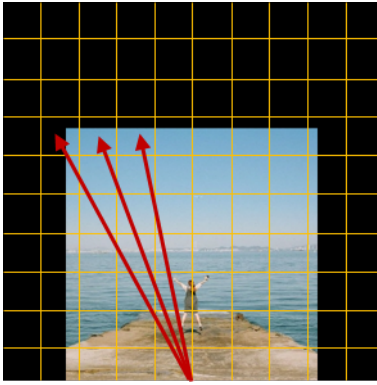


그림 2. 패치영역 탐색 방법

원본 사진의 주변 배경 합성을 위해 사진의 상과 좌, 우 영역을 넓힌 결과 사진을 우선 생성한다. 그리고 결과 사진을 가로와 세로로 나누어서 격자를 생성한다. 이후 격자로 생성된 각각의 패치와 원본 사진의 아래쪽 중점까지 거리를 기준으로 합성할 패치의 우선순위를 정한다. 중점과 가까운 패치일수록 겹치는 영역이 넓고, 가장 비슷한 사진의 검색이 가능하므로 이를 통해 최대한 원본 사진과 겹치는 패치 영역을 우선적으로 합성하고자 하였다.

각각의 패치영역은 [5]의 방식처럼 8 x 8 블록 형식의 격자로 나누어서 사진의 위치 정보를 활용하여 겹치는 영역에는 1의 가중치를 주고, 겹치지 않는 영역에는 0의 가중치를 주어서 겹치는 영역에서만 비교를 통한 검색을 하고자 하였다. GIST 기술자뿐만 아니라 질감 비교를 위하여 HSV 색 공간의 히스토그램을 추출하여 3개의 채널 (H, S, V)에 대하여 LBP (Local Binary Patterns) [14] 기술자를 추출하여 사용하였다. LBP 기술자는 흑백 사진에서 각 픽셀의 주변 영역의 상대적인 밝기차이를 변환하여 생성하는데, 이는 GIST 기술자만을 이용하여 검색 시에 질감을 충분히 고려치 못하는 문제점을 보완하기 위하여 사용하였다. 두 사진의 GIST와 3개의 LBP 기술자들 각각의 유클리드 거리를 계산하였다. 두 사진의 거리는 다음과 같이 정의 한다.

$$d_g = \|g_i - g_j\| \quad (1)$$

$$d_h = \|h_i - h_j\| \quad (2)$$

$$d_s = \|s_i - s_j\| \quad (3)$$

$$d_v = \|v_i - v_j\| \quad (4)$$

$$dist = w \cdot d_g + w \cdot d_h + w \cdot d_s + w \cdot d_v \quad (5)$$

d_g 는 두 GIST 기술자 간의 거리이고 d_h, d_s, d_v 는 HSV 각 채널의 두 LBP 기술자 간의 거리이다. w 는 64 칸으로 이루어진 패치영역의 블록에 대한 가중치이다. 가장 유사한 사진과 원본 사진의 경계를 기준으로 일정 픽셀 반경의 영역에서 포아송 이미지 에디팅을 적용하여 두 사진을 합성하였다. 본 논문에서는 두 사진의 색상이 다를 경우에 경계부

분뿐만 아니라 주변 영역의 색상을 자연스럽게 혼합하는 방식[15]을 사용하였다. 따라서 반복되는 합성에도 색상이 부자연스러운 부분을 최대한 없애고자 하였다.

4. 실험 결과 및 분석

본 논문에서 제안한 자가촬영 사진의 배경합성 방식은 기존의 사진 합성과는 다른 점이 많아 기존 연구와 비교가 힘든 점이 있다. 우선 기존의 연구들은 원본 사진의 특정 영역에 유사한 다른 한 장의 사진으로 그 부분만 메우는 방식이었지만, 본 연구는 전혀 정보가 존재하지 않는 원본 사진 외부의 영역에 수많은 사진을 조합하여 확장시키는 방식을 사용하였다. 따라서 본 연구는 Adobe사의 Photoshop CC에서 제공하는 내용 인식 채우기 (Content-aware fill) 기법과 비교하였다.

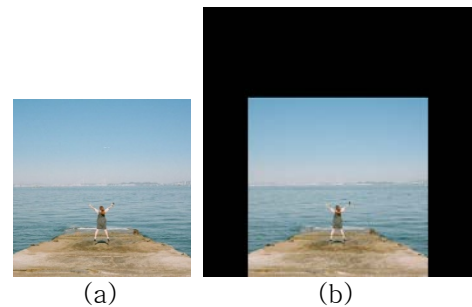
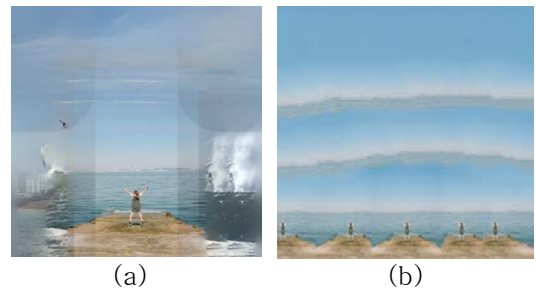


그림 3. 원본 사진(a)과 확장할 영역 사진(b)



(c)

(d)

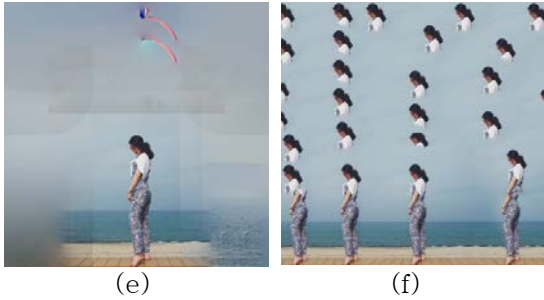


그림 4. 제안한 방식 (a, c, d)과 Photoshop CC의 내용 인식 채우기 (b, d, f) 결과물 비교

Photoshop의 내용 인식 채우기 기법은 원본 사진 속의 영역을 그대로 복사해서 나머지 부분을 채워나가는 방식을 사용한다. 이러한 방식은 풍경 사진 같은 단조로운 패턴에서는 뛰어나지만, 자가촬영 사진의 경우 사람이 복사되어 채워지는 문제가 발생한다. 본 연구의 결과물은 주변 영역에 원본 사진을 복사하지 않았기 때문에 그러한 문제가 발생하지 않았다. 또한 기존의 사진 합성 방식처럼 사용자가 직접 합성할 영역을 특정하지 않아도 자동으로 합성할 영역을 결정하여 확장해 나가는 장점을 확인할 수 있었다. 본 연구에서 제안한 방식의 결과물에서는 원본과 상이한 몇 개의 사진이 합성되었지만 더 큰 데이터셋을 사용한다면 더 나은 결과물을 낼 수 있을 것으로 기대한다. MIRFlickr-25000 데이터셋을 분석한 결과, 바다가 태그된 사진이 25,000 장 중에 약 300 장밖에 되지 않았는데, 이로 인하여 그림 4. (a)에서 바다 영역의 합성 결과가 만족스럽지 않은 것으로 평가한다. 이후에 MIRFlickr-1M 같은 백 만장의 데이터셋을 사용하면 더 좋은 결과물을 얻을 수 있을 것이다.

5. 결론

본 논문에서는 대용량 이미지 데이터에서 근사 근접 이웃 검색을 기반으로 색상정보를 활용하여 자가촬영 사진의 주변 배경을 합성하는 방법을 제안하였다. 근사 근접 이웃 기반의 프러덕트 양자화를 이용하여 GIST 기술자를 양자화 하여 원본 사진과 근접한 사진으로 재구성 하였다. 그리고 주변이 확장된 원본 사진에서 가로와 세로로 나눈 격자 형태의 패치를 생성하여, 원본 사진과 겹치는 영역이 넓은 패치부터 순차적으로 합성하고자 하였다. 패치 영역은 8 x 8 블록형식의 격자로 블록의 위치 영역을 활용하고자 하였고, 각각의 블록에 겹치는 영역에 대한 가중치를 주었다. GIST 기술자와 HSV 색 공간의 LBP 기술자를 함께 사용하여 색상뿐만 아니라 사진의 질감도 고려하고자 하였다. 이렇게 검색된 유사한 사진과 원본 사진의 경계 주변에 포아송 이미지 에디팅을 적용하여 두 사진을 합성하였다. 합성할 영역을 특정해야 하는 이전까지의 사진 합성방식과 비교하여 사용자가 합성하는 영역을

고려하지 않아도 되는 장점을 볼 수 있었다. 본 연구의 결과물과 Adobe사의 Photoshop CC에서 제공하는 내용 채우기 기법과 비교하였는데, 본 연구의 결과물에서는 데이터셋에 유사한 사진이 적게 있었음에도 불구하고 주변 영역을 효과적으로 채워 나갈 수 있었다. 따라서 더 큰 데이터셋에서 제안한 방식을 적용한다면 더 나은 결과물이 나올 것으로 예상된다.

감사의 글

리뷰를 도와주신 이윤석, 조재형, 김동혁님께 감사드립니다. 본 연구는 NRF-2013R1A1A2058052와 미래창조과학부 및 정보통신기술연구원진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [R0126-15-1108, 대용량 이미지 검색과 시제품 렌더링을 위한 근접질의 SW 개발]

참고문헌

- [1] V. Kwatra, et al., "Graphcut textures: image and video synthesis using graph cuts", ACM Transactions on Graphics (ToG), 22, 3, 2003.
- [2] P. Pérez, et al., "Poisson image editing", ACM Transactions on Graphics (TOG), 22, 3, 2003.
- [3] J. Sun, et al., "Image completion with structure propagation", ACM Transactions on Graphics (ToG), 24, 3, 2005.
- [4] N. Komodakis, and G. Tziritas, "Image Completion Using Global Optimization", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006.
- [5] J. Hays, and A. A. Efros, "Scene completion using millions of photographs", ACM Transactions on Graphics (TOG), 26, 3, 4, 2007.
- [6] J.B. Huang, et al., "Image completion using planar structure guidance", ACM Transactions on Graphics (TOG) 33, 4, 129, 2014.
- [7] A. Oliva, and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope", International journal of computer vision, 42.3, 145-175, 2001.
- [8] Y. Weiss, et al., "Spectral hashing", Advances in neural information processing systems (NIPS), 2008.
- [9] J.-P. Heo, et al., "Spherical hashing", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.
- [10] H. Jegou, et al., "Product quantization for nearest neighbor search", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 33.1, 117-128, 2011.
- [11] J.-P. Heo, et al., "Distance encoded product quantization." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014
- [12] M. J. Huiskes, and M. S. Lew, "The MIR Flickr Retrieval Evaluation", ACM International Conference on Multimedia Information Retrieval (MIR), 2008.

- [13] M. Douze, et al., "Evaluation of gist descriptors for web-scale image search", Proceedings of the ACM International Conference on Image and Video Retrieval, 2009.
- [14] T. Ojala, et al., "A comparative study of texture measures with classification based on featured distributions", Pattern recognition, 29.1, 51-59, 1996.
- [15] M. Tanaka, et al., "Seamless image cloning by a closed form solution of a modified poisson problem", SIGGRAPH Asia, 2012.