

이미지 합성 편의성을 높이는 메타데이터 구조 및 플랫폼 설계

박정석, 김동호

서울과학기술대학교

whiter1234@naver.com, dongho.kim@snut.ac.kr

Design of Metadata Structure and Platform to Increase Image Compositeness Convenience

Jungsuk Park and Dong Ho Kim

Seoul National University of Science and Technology

요 약

기존의 이미지 분리 및 합성 과정은 전문 프로그램의 도움을 받아 이미지로부터 사물이나 환경을 분리하거나 합성하여 이루어져 왔으며 생산되는 이미지의 양에서 일반 사용자가 생성하는 이미지의 비중이 높음에도 상대적으로 적은 수의 인원만이 기존 이미지를 분리하고 합성할 수 있었다. 본 논문에서는 다량의 기존 이미지 내 요소를 손쉽게 분리, 합성하여 새로운 콘텐츠를 제작할 수 있도록 하는 메타데이터 구조와 이를 활용하여 이미지 합성에 대한 사용자 편의성을 높이는 플랫폼을 제안한다. 이는 object segmentation 을 기반으로 이미지의 각 요소를 분리하고 계층화 하여 이루어지며 이미지 합성에 대한 접근성을 높이고 분리된 이미지의 속성을 메타데이터로 함께 표기하여 다량의 기존 이미지에서 필요로 하는 이미지 요소를 빠르게 찾을 수 있도록 한다. 또한 분리된 이미지 요소의 속성을 구체화하기 위해 사용할 수 있는 방법들에 대해 논의한다. 결과적으로 위 제안은 기존 이미지 요소를 분리, 합성하기 위해 필요한 장벽을 낮추는 역할을 수행하여 더 많은 사용자들이 이미지 합성에 참여할 수 있게 할 것으로 기대된다.

1. 서론

멀티미디어 시대가 지속됨에 따라 이미지와 영상을 포함한 수많은 미디어들이 생성 및 저장되고 있다. 글로벌 시장조사 기관인 IDC(International Data Corporation)에 따르면 연간 데이터 생산량은 2025 년 163 ZB 규모가 될 것으로 전망된다.[1] 이렇게 생산 및 저장된 미디어는 그 자체만으로도 가치를 지니지만, 대부분의 이미지나 영상은 그 순간에만 가치를 지니고 이후에는 저장매체에 잠들어 있게 된다. 그러나 이미지 요소 분리·합성 과정을 거치면 새로운 미디어를 생성하는 용도로도 사용할 수 있다.

기존의 이미지 요소 분리·합성 과정은 적합한 이미지를 탐색하고 포토샵 등의 툴을 사용하여 이미지 요소를 분리한 뒤 기존 이미지에 합성하는 절차로 이루어져 있다. 이런 기술을 지원하는 툴을 보유하지 않은 일반 사용자들은 이미지 요소 분리·합성에 접근하기 어려우므로 비교적 소수의 사용자에 의해 이미지 분리·합성이 이루어지고 있다. 따라서 전체 이미지 중 실제로 탐색되어 사용하게 되는 이미지는 일부분에 불과하다. 이미지 툴 등의 진입장벽 없이 일반 사용자가 쉽게 참여할 수 있는 이미지 합성의 예시로 이미지에 필터를 적용하는 사례가 있다. 필터는 이미지만이 아니라 실시간으로 적용 가능하다는 특징을 갖고 있으며 SNS 등의 매체에서 기존과 다른 가치와 맥락을 갖는 이미지를 생성하는 일에 기여하고 있다. 이미지

합성에서 필터가 갖는 장점은 특별한 지식 없이도 적용 가능하며 시각적으로 즉시 확인할 수 있다는 점이다.

이를 이미지 분리·합성에 적용하여 보면 기존 이미지 요소의 분리 과정을 단축하고 편의성을 높임으로써 다수의 일반 사용자들이 이미지 합성에 참여할 수 있도록 할 수 있을 것이다. 본 논문에서는 이미지 분리 과정의 편의성을 높이기 위해 이미지 요소를 분할 저장하여 이미지 합성에 대한 접근성을 개선하는 방안을 제안한다. 그러나 만약 이미지 요소를 일괄적으로, 또는 기계적으로 분리하여 저장하면 합성이 어렵거나 불가능한 이미지가 생성되므로 무가치한 이미지 생성에 자원을 낭비하게 된다. 따라서 이미지 요소의 분리 저장만이 아니라 완전한 이미지 요소를 저장하기 위한 연구 또한 필요하다.

따라서 본 논문의 구성은 일반 사용자들의 이미지 합성 편의성을 높이기 위해 이미지 요소의 합성에 대한 편의성을 높이는 메타데이터 구조의 필요성 및 제안과 이에 수반될 수 있는 불완전한 이미지 요소 발생 등의 한계에 대한 논의, 결론과 완전한 이미지 요소를 분리 저장하기 위해 앞으로 할 연구에 대하여 논의한다.

2. 메타데이터 구조의 필요성

사진의 저장이 쉬워지면서 수많은 이미지가 아무런 역할도 없이 저장되어 있는 경우가 많다. 이렇게 잠들어 있는 이미지들을 활용하여 새로운 가치를 창출할 수 있게 하는 것이 이미지 분리·합성이다. 예를 들어 다른 이미지에 있는 인물, 사물을 현재 이미지에 합성하거나, 반대로 현재 배경을 다른 사진의 배경으로 합성하여 원하는 장소에서 촬영한 것처럼 바꿀 수 있으며 이는 AR, VR 환경에서도 동일하게 적용할 수 있다.

AR, VR 환경은 컴퓨터에 상당한 부하를 주는 환경이므로 기존 이미지를 분해하는 일까지 자원을 투입하기 어렵다. 설령 자원을 투입할 여력이 있더라도 합성하기를 원하는 사물이나 환경을 찾기 위해 검색된 이미지 전체를 두고 하나하나 검토한 다음 이미지 요소를 분리하는 작업이 필요하게 된다. 따라서 합성하려는 이미지 요소를 포함한 전체 이미지가 아니라 이미지 요소만 검색할 수 있다면 적은 연산부하로 쉽게 이미지 합성을 진행할 수 있을 것이다.

수많은 연구자들의 노력으로 이미지 요소를 탐지해 내는 object segmentation [2][3][4], object detection [5][6][7] 기술이 다양한 서비스에서 활용되고 있음에도 기존의 이미지 분리·합성 과정은 사용자가 툴을 이용하여 이미지 내 요소를 분리한 뒤 대상 이미지에 덧씌우는 방식으로 진행되었으며, 툴을 사용해야

한다는 점이 진입장벽이 되어 기존 이미지의 요소를 분리·합성하는 일에 걸림돌이 되었다. 이미지 합성 편의성을 높이는 메타데이터 구조는 이미지 요소 분리·합성 과정 중 이미지 요소 분리 과정을 단축하는 역할을 수행한다. 이미지 내에서 다른 이미지에 합성 가능한 잠재성이 있는 요소들을 미리 분리하여 전체 이미지 메타데이터의 하위 계층에 포함함으로써 추가적인 작업 없이 합성에 필요한 이미지 요소를 불러올 수 있게 한다. 이는 사용자가 플랫폼에서 합성하길 원하는 이미지 요소(사물, 환경 등)를 검색했을 때 이미지 전체를 불러와 사용자의 검토 및 조작을 기다릴 필요 없이 분리된 이미지들만 읽어올 수 있으므로 이미지 요소를 분리하는 작업 없이 바로 합성을 시도할 수 있다. 이미지를 분리하는 과정이 단축됨으로써 사용자가 어려움 없이 분리된 기존 이미지 요소를 합성하여 새로운 맥락과 의미를 갖는 이미지를 만들어 낼 수 있을 것이다.

3. 메타데이터 구조 제안

이미지 합성 편의성을 높이는 메타데이터 구조는 이미지를 parent - child 이미지로 분리하는 것에서 시작한다. parent 이미지는 원본 이미지를 의미하며 child 이미지는 원본 이미지 내의 이미지 요소를 분리한 이미지이다. Object segmentation 알고리즘으로 탐지된 이미지 요소의 경계를 기준으로 이미지를 분할한 예시가 그림 1에 나타나 있다.

그림 1. Parent 이미지 (좌), Child 이미지 (우)



그림 1 에 사용된 COCO Dataset[8] 은 328k 개 이미지 내의 사물을 object detection, object segmentation 을 적용하여 91 종류의 카테고리 분류하고 detection 과 segmentation 좌표, 이미지를 자연어로 묘사한 이미지 캡셔닝(captioning)[9] 데이터, 이미지 내 사람의 관절 위치에 대한 추정 데이터 등을 포함한 자료이며 그림 1 의 이미지 요소 경계 및 카테고리 정보는 COCO Dataset 에 기록된 정보에서 획득하였다. 분할된 이미지는 원본 이미지 메타데이터의 child 항목에 추가되며 카테고리, 원본 이미지에서의 좌표, 원본 이미지의 id 를 상속받은 child id 정보를 포함한다. 그림 1 에 대해 생성된 메타데이터 예시가 그림 2 에 나타나 있다.

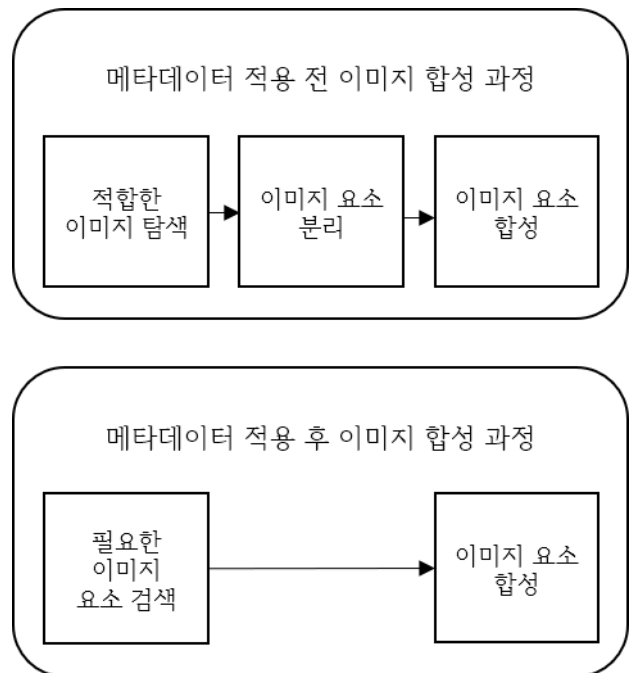
그림 2-1 의 parent 이미지 카테고리 정보는 원본 이미지 내의 모든 사물에 대한 대분류 및 소분류를 포함하고 있으며 child 정보는 원본 이미지에서 분리된 child 이미지에 대한 정보가 순차적으로 기록되어 있다. 그림 2-2 의 child 이미지 정보는 해당 이미지 요소의 대분류, 소분류와 원본 이미지상에서 이미지 요소의 좌표, parent 이미지의 id 를 상속한 child id 로 이루어져 있다. 이 중 원본 이미지상에서의 좌표는 저장공간 문제를 해결하기 위한 요소로, 원본 이미지의 사용률이 떨어진다면 원본 이미지를 보관하지 않고 필요할 때 메타데이터를 기반으로 재구성하기 위한 정보이다.

그림 2. 메타데이터 예시

2-1 Parent 이미지의 카테고리 정보
<pre> 'categories': [{'label': 'oven', 'refrigerator'}, 'super_category': ['appliance']] </pre>
2-2 Parent 이미지에 속한 child 이미지 정보
<pre> 'child': [{'child_categories': [{'label': 'refrigerator', 'super_category': ['appliance']}, 'child_coor': [245.05, 185.23, 165.76, 356.0], 'child_id': '000000000802-01'], {'child_categories': [{'label': 'oven', 'super_category': ['appliance']}, 'child_coor': [33.08, 289.08, 126.56, 231.55], 'child_id': '000000000802-02'], </pre>
2-3 Parent 이미지 id 정보
<pre> 'image_id': 000000000802 </pre>

그림 3 은 이미지 합성 편의성을 높이는 메타데이터 구조 적용 전, 후의 이미지 분리·합성 과정 비교이다. 기존에는 필요한 이미지 요소를 포함하는 전체 이미지를 검색한 뒤 이미지 툴을 사용하여 이미지 요소를 분리하고 기존 이미지에 합성하는 절차가 필요했다면, 메타데이터 구조를 적용하고 나서는 필요한 이미지 요소를 검색한 뒤 바로 기존 이미지에 합성하는 것으로 이미지 분리·합성 과정이 끝나게 된다. 만약 검색한 이미지 요소를 포함하는 parent 이미지들이 필요한 경우 parent 이미지의 메타데이터를 기반으로 검색하여 제공하거나 해당 parent 에 속한 child 이미지들을 결합하여 원본 이미지를 복원하여 대응한다.

그림 3. 이미지 합성 편의성을 높이는 메타데이터 구조 적용 전, 후



4. 논의

이미지 합성 편의성을 높이는 메타데이터 구조 및 플랫폼은 이미지 요소를 분리 저장하여 기존 이미지를 쉽게 응용할 수 있지만, 불완전한 이미지를 저장할 수 있다는 문제를 안고 있다. 예를 들어 1. 이미지 내 사물이 겹쳐 있는 경우, 혹은 2. 이미지 내의 배경을 child 이미지로 저장하는 경우에 발생할 수 있다. 예시 1 의 경우 상대적으로 뒤에 있는 물체는 앞선 물체에 의해 가려지므로 가려진 이미지 요소를 현재 이미지에 합성했을 때

부자연스러운 결과를 낳는다. 예시 2 의 경우에도 배경에 있던 사물 등의 이미지 요소가 제거된 이미지를 사용하게 되므로 마찬가지로 자연스러운 합성이 어렵다. 이 문제에 대한 해결 방안으로 서로 겹쳐 있는 물체를 하나의 그룹으로 취급하여 저장하는 것을 떠올릴 수 있다. 문제는, 예시 2 의 경우 해당 방법을 적용할 수 없다는 점이다. 어떤 이미지 요소가 일부 잘려 있어도 사용 가능한지, 사용 불가능 한지에 대한 기준을 마련하여야 불완전한 이미지 요소 저장으로 인한 손실을 피할 수 있을 것이다.

5. 결론

이미지 합성 편의성을 높이는 메타데이터 구조 및 플랫폼은 이미지 합성에 필요한 요구조건을 완화하며 기존 이미지 요소를 활용하여 새로운 이미지를 생성하는 서비스에 활용될 수 있다. 그럼에도 겹쳐 있거나 비교적 가치가 떨어지는 이미지 요소를 다루는 부분에 있어서 한계가 있다. 이러한 문제를 해결하고 이미지 요소의 속성을 메타데이터에 반영하고자 차후 이미지 캡셔닝(captioning)을 기반으로 한 child 이미지 계층화를 진행하고자 한다. 이미지 캡셔닝은 이미지 전체를 설명하는 자연어 문장을 생성하는 기술로 각 이미지 요소의 속성을 묘사하는 것에 도움을 줄 수 있을 뿐 아니라 그룹으로 묶인 이미지 요소의 관계를 묘사하여 복합적인 이미지를 사용자가 찾기 쉽게 할 수 있을 것으로 기대된다. 또한 이미지 캡셔닝의 요소인 attention mechanism 은 인간의 인지기능을 모방한 기술로 이미지의 핵심 요소를 더 중요하게 다루고 중요도가 떨어지는 사물을 무시하는 문장을 생성하도록 돕는다[10]. 이를 응용하여 이미지에서 너무 작거나 파편화 되어 사용하기 어려운 이미지 요소를 제거하는 것에 사용할 수 있을 것으로 보인다.

5. Acknowledgement

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-00994, 이용 환경을 반영하는 자율적 VR·AR 콘텐츠 생성 기술개발)

6. 참고문헌

[1] Data Age 2025: The Evolution of Data to Life-Critical (IDC, 2017.4)

- [2] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.
- [3] Bolya, Daniel, et al. "Yolact: Real-time instance segmentation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- [4] Wang, Xinlong, et al. "Solo: Segmenting objects by locations." European Conference on Computer Vision. Springer, Cham, 2020.
- [5] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." arXiv preprint arXiv:1506.01497 (2015).
- [6] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.
- [7] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [8] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014.
- [9] Chen, Xinlei, et al. "Microsoft coco captions: Data collection and evaluation server." arXiv preprint arXiv:1504.00325 (2015).
- [10] Wang, Haoran, Yue Zhang, and Xiaosheng Yu. "An Overview of Image Caption Generation Methods." Computational intelligence and neuroscience 2020 (2020).