

다각도 영상 합성의 속도 향상을 위한 영상유사도 생성 알고리즘

김준식, 정광일, 황용완, 박필규, 박성환, 김규현
경희대학교

dnjemfhq@nate.com, kwangil503@gmail.com, wan2509@naver.com
gyu0715@nate.com, gocheenee@khu.ac.kr, kyuheonkim@khu.ac.kr

Video Similarity Generating Algorithm Improving the Speed of Various Multi-Angle Image Composition

Jun-sik Kim, Gwangil Jeong, Yongwan Hwang, Pilkyu Park, Seonghwan Park, Kyuheon Kim
Kyung Hee Univ.

요 약

현재 사용 되어지고 있는 대부분의 영상 합성 알고리즘들은 영상의 크기가 클수록 계산량이 많아지는 특징이 있다. 따라서 UHD 영상을 대상으로 영상 합성을 할 때 속도가 기존의 HD 영상에 비해 크게 느리다는 점과 특수 카메라 사용 혹은 카메라 위치 고정 등의 제한이 필요하다는 점의 단점들이 있다. 이에 본 논문에서는 스마트폰에서 콘텐츠를 촬영한 영상들을 이용하여 영상 분류, Map 생성, Rendering의 과정을 통해 현재 사용 되어지고 있는 영상 합성 알고리즘과 비교하여 보다 빠르게 영상 합성을 할 수 있는 알고리즘을 제안한다. 촬영 시 EXIF에 저장되는 GPS 정보를 이용하여 그룹화를 진행한 후 영상의 특징점 매칭을 통해 Map을 생성하는 것이 본 논문에서 제안하는 알고리즘이다. 이를 위해 본 논문에서는 FAST 특징점 알고리즘과 FREAK 기술자 알고리즘을 사용하였으며 Rendering을 통하여 해당 알고리즘을 검증 하였다.

1. 서론

최근 영상기술의 발달로 파노라마, VR 등 다양한 종류의 미디어들이 생산되고 있다. 또한, SNS의 발달로 인해 사용자들 간의 미디어 교류가 점점 활발해지고 있다. 이러한 환경에서 기존의 미디어를 이용 및 가공하여 새로운 형태의 미디어를 창조하는 것은 영상기술 분야의 주된 관심사이다. 사용자는 스마트폰을 이용하여 풍경, 건물, 스포츠 경기 등의 다양한 이벤트를 촬영한다. 일반적인 카메라를 이용하여 촬영할 때에는 카메라의 시야 각에 의해 촬영에 제한을 받게 된다. 인간의 시야각과 비교하여 비교적 좁은 각도의 시야 각을 갖기 때문에 사용자들 실제보다 현실감이 떨어지는 느낌을 갖기 쉽다. 이러한 문제점을 극복하기 위해 파노라마 영상 합성 프로그램[1] 등이 있지만 이를 사용하기 위해서는 여러 제한 조건이 요구된다. 따라서 본 논문에서는 사진 자체의 데이터만으로 콘텐츠를 합성 및 재현하는 방법을 제안하기로 하였다.

영상 내의 데이터를 이용하여 시야 각을 확장시키기 위해서는 우선 영상들 간의 연관성을 찾고 합성해야 한다. 영상들의 연관성을 찾을 때 사용되는 대표적인 알고리즘이 특징점 검출 알고리즘이다. 특징점 추출 알고리즘으로는 SIFT(Scale Invariant Feature Transform)[2], SURF(Speeded Up Robust Features)[3]등의 알고리즘이 있다. SIFT나 SURF 등의 알고리즘은 영상 Stitching에 널리 사용되는 알고리즘으로 뛰어난 정확도를 가진다는 장점을 가지고 있다. [4] 하지만 정확도가 뛰어난 만큼 처리시간이 비교적 길다는 단점이 있다. 이러한 과정을 거쳐 처리된 특징점을 각각 비교하여 매칭 후 출력하는 것이 기존의 영상

합성 알고리즘의 과정이다. 이러한 과정을 처리하는데 있어 특징점 추출에서 처리시간이 요구되며, 추출된 특징점을 매칭하는 과정에서도 처리시간이 요구되게 된다.

본 논문에서는 이러한 단점들을 보완하기 위한 방법을 영상 분류, Map 생성, Rendering의 세 단계를 거쳐 설명하였다. 우선 각 영상들의 경로를 저장한 후 처리시간의 가장 큰 부분을 차지하는 매칭 횟수를 줄이기 위해 GPS 기반의 그룹화 과정을 거치게 된다. 다음으로 특징점을 추출하는 처리시간을 줄이기 위해 영상을 Resize하고 처리시간 측면에서 다른 특징점 추출 알고리즘보다 비교적 성능이 우수한 FAST[5]를 이용한다. 또한 매칭 후에 다수의 특징점이 아닌, 하나의 특징점만으로 간단하게 영상 Map을 만드는 방법을 제안한다. 마지막으로 생성된 영상 Map을 Rendering하는 것으로 본 논문에서 제안한 방법의 성능을 검증한다.

2. 영상 분류

가. GPS Grouping

영상 분류의 가장 첫 번째 단계는 <그림 1>과 같이 영상을 입력 받아오는 것이다. 입력 받아온 영상을 구조체에 저장하며 이때 경로를 우선적으로 저장하고 영상의 EXIF(Exchangeable Image File Format)를 추출하여 GPS Latitude, GPS Longitude 값을 저장한다. 이후 영상들 간의 매칭 횟수를 줄이기 위해 같은 GPS를 갖고 있는 영상들을 묶어 그룹화한다.

가. 영상 Resize

<그림 4>에서 보여지는 바와 같이 경로를 통해 영상을 1:0.125 의 비율로 resize 하여 불러온다. <그림 5>를 통하여 SURF 특징점 검출 후 매칭 시 L2_norm 거리가 0.125 비율 resize 이후로 급증 하는 것을 확인할 수 있다. 이 결과를 통하여 0.125 비율로 resize 를 하는 것은 영상 처리 속도를 높이고 정확성을 크게 저하시키지 않는다는 것을 확인할 수 있다.

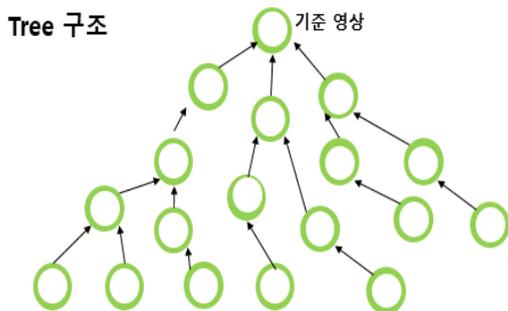


<그림 5>. SURF 결과값 비교

나. 특징점 및 기술자 추출

Resize 한 영상을 이용하여 특징점 및 기술자를 각각 FAST, FREAK (Fast Retina Keypoint) [6] 알고리즘을 이용하여 추출해 내고 저장한다. 이때 처리시간이 비교적 많이 소요되는 SIFT 와 SURF 를 사용하지 않고, 처리시간이 비교적 적게 걸리는 FAST 를 사용하여 처리시간을 단축한다. 또한 추출되는 특징점의 개수를 300 개로 제한하여 특징점 추출 처리시간을 감소시킨다. 이후 생성한 특징점과 기술자를 모두 저장하여 매칭에 사용한다.

다. List 내부 Map 생성



<그림 6>Tree 구조

본 논문에서는 저장한 특징점과 기술자를 이용하여 List 내부 영상들을 Brute-force [7] 알고리즘을 이용하여 매칭한다. 매칭 알고리즘은 매칭 하는 영상과 매칭 되는 영상으로 나누어 처리된다. 매칭 하는 영상은 매칭되는 영상들 중 HAMMING 거리의 기준을 넘는 영상이 있을 경우 매칭 된 특징점들 중 하나를 HAMMING 거리의 기준에 따라 선정한다. 이 후 그 특징점과 매칭된 자신의 특징점, 총 두가지를 저장한다. 이때 매칭 된 영상을 Target 영상이라고 부른다. 특징점을 저장하는

것은 각 영상 별 1 회만 실시하여 각각의 영상이 하나의 Target 영상을 갖도록 한다. 하지만 특징점을 저장한 이후에도 매칭 알고리즘은 계속 시행되며 매칭 되는 영상과의 매칭 알고리즘의 성공 횟수가 가장 많은 영상이 영상 Map 의 기준 영상이 된다. 이때 Map 은 Tree 구조로 표현이 된다. 앞서 말한 기준 영상의 Target 영상은 무시한 상태에서 기준 영상을 Target 영상으로 가지고 있는 영상들을 하위 노드로 배치한 뒤, 배치된 영상들을 Target 영상으로 가지고 있는 영상들을 배치, 이 과정을 매칭이 성공한 모든 영상들에 적용하면 <그림 6>과 같은 Tree 구조가 생성되게 된다.

라. 전체 Map 생성

<그림 4>에서 보여지는 것과 같이 각각 List 의 Tree 가 기준 영상을 이용하여 만들어진 후 이 기준 영상을 이용하여 다른 List 와 매칭 알고리즘을 실시한다. 이는 앞서 만들었던 Tree 구조를 여러 List 들을 대상으로 다시금 생성하기 위함이다. 이는 사용자가 한가지 콘텐츠를 기반으로 촬영을 실시 했기 때문에, 기준 영상들이 모두 한가지 콘텐츠를 포함하고 있기 때문에 가능하다. 이때 전체 Map 의 기준은 GPS 값의 무게중심 즉, 경도, 위도 분포의 무게 중심을 잡아서 기준을 설정한다. 우선 무게 중심을 구한 뒤 무게 중심과 가장 가까운 List 를 기준 List 로 설정하여 기준 List 부터 차례대로 panorama 를 출력하여 전체 합성 영상을 얻는다.

4. Rendering

본 논문에서는 분류, Map 생성, Rendering 의 과정으로 영상 합성 시 요구되는 제한 조건들을 만족시키며 합성 속도를 높이는 과정을 제안하였다. 첫번째로 영상 합성 시 받아오는 영상의 제한 조건을 줄이기 위해 영상 내부에 존재하는 GPS 정보를 이용해 분류를 실시하였다. 이 방법을 통해 결과적으로 매칭 정확도를 높이는 효과도 얻을 수 있었다. 두번째로 특정 콘텐츠를 기준으로 영상들의 위치관계를 구하기 위해 특징점 매칭 알고리즘을 이용하여 Tree 구조를 생성하였다. 마지막으로 Rendering 을 통해 Map 이 기존의 특징점 추출 알고리즘보다 비교적 적은 처리시간을 소요하면서도 현실감 있는 영상을 출력할 수 있음을 확인한다.



<그림 7> Tree 구조 Rendering

<그림 7>은 Tree 구조로 저장된 영상 Map 을 출력한 결과이다. 보여지는 바와 같이 Tree 구조를 이용하여 Map 생성을 하는 것 만으로도 영상합성이 이루어지는 것을 확인할 수 있다. 또한 영상의 매칭 정확도가 우수하다면 resize 비율에 무관하게 영상을 Rendering 할 수 있음을 알 수 있다. <그림 8>은 UHD 로 촬영된 147 장의 사진을 이용하여 생성한 전체 Map 을 Rendering 한 결과이다. 기준 List 가 콘텐츠의 중심에 있고, List 끼리의 기준 영상만으로 Map 생성이 이루어짐을 확인할 수 있다.



<그림 8> 전체 Map Rendering

5. 결론

현재 영상을 합성하는 방법 중 가장 대표적인 방법으로는 특징점 알고리즘을 사용하는 방법이 있다. 하지만 이러한 영상 합성 알고리즘은 많은 연산량으로 인하여 처리시간이 길기 때문에 고화질의 영상 혹은 비교적 연관성이 없는 영상들에 사용할 경우 처리시간이 길어진다는 단점이 있다. 따라서 본 논문에서는 사용자들에게 UHD 화질의 다각도 영상을 입력 받아 하나의 큰 영상으로 출력할 수 있도록 프로그램을 설계 및 구현하고 해당 프로그램의 성능을 검증하였다. 본 논문에서 제안한 GPS 정보를 이용한 분류는 스마트폰으로 촬영된 영상의 위치정보를 이용하여 분류함으로써 매칭의 정확도를 높이기 위하여 사용 되었다. Resize 와 FAST 알고리즘을 사용하는 방법, 추출하는 특징점의 개수를 제한하는 방법, 매칭에서 각각의 영상에서 한개의 특징점을 이용하여 Tree 구조를 작성하는 방법들은 처리시간을 최소화시키기 위하여 제안하였다. 단, 영상의 Map 을 생성하여 영상의 대략적인 위치를 파악하는 것 만으로는 완벽한 영상합성을 기대하기는 힘들기 때문에 영상이 겹치는 부위마다 Stitching 이 이루어져야 할 것이다.

* 본 논문은 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [B0126-15-1013, 퍼즐형 Ultra-wide viewing 공간 미디어 생성 및 소비 기술 개발]

참고문헌

- [1] Luo Juan, Gwun Oubong, "Surf applied in panorama image stitching", IPTA 2010
- [2] Lowe, D.G., "Distinctive image features from scale-invariant keypoints", IJCV 2004.
- [3] Bay, H., Tuytelaars, T., and Van Gool, L., "Surf: Speeded up robust features", in ECCV 2006
- [4] Zhen Hua, Yewei Li, and Jinjiang Li, "Image Stitch Algorithm Based on SIFT and MVSC", FSKD 2010
- [5] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection", in ECCV 2006
- [6] Alexandre Alahi, Raphael Ortiz, Pierre Vandergheynst, "FREAK: Fast Retina Keypoint", CVPR 2012

[7] Olof Enqvist, Fangyuan Jiang, Fredrik Kahl, "A brute-force algorithm for reconstructing a scene from two projections", CVPR 2011