

## 특징점 추적을 통한 다수 영상의 고속 스티칭 기법

\*박시영, \*\*김종호, \*\*\*유지상

\*광운대학교

\*pksiyoung@kw.ac.kr, \*\*jjong3927@kw.ac.kr, \*\*\*jsyoo@kw.ac.kr

### Fast stitching algorithm by using feature tracking

\*Siyoung Park \*\*Jongho Kim \*\*\*Jisang Yoo

\*Kwangwoon University

#### 요 약

본 논문에서는 비디오 영상을 입력 했을 때 특징점 추적을 통한 다수 영상의 고속 스티칭 기법을 제안한다. 빠른 속도로 특징점 추출을 위해서 FAST(Features from Accelerated Segment Test) 기법을 사용한다. 특징점 정합과정은 기존의 방법과는 다른 새로운 방법을 제안한다. Mean shift 를 통해 특징점이 포함된 영역을 추적하여 벡터(vector)를 구한다. 이 벡터를 사용하여 추출한 특징점들을 정합하는데 사용한다. 마지막으로 이상점(outlier)을 제거하기 위해 RANSAC(RANdom Sample Consensus) 기법을 사용한다. 입력된 두 영상의 호모그래피(homography) 변환 행렬을 구하여 하나의 파노라마 영상을 생성한다. 실험을 통해 제안하는 기법이 기존의 기법보다 속도가 향상되는 것을 확인하였다.

#### 1. 서론

파노라마 영상은 제공된 여러 개의 영상을 다양한 처리를 통해 같은 영역을 찾고 스티칭(stitching)하는 과정을 거쳐 생성되며 고해상도의 넓은 시야각을 제공한다. 그 응용 분야도 전시관 등에 활용될 뿐 아니라 스트리트 뷰 등으로 점차 확대되고 있다[1]. 근래에는 시청자와 함께 상호작용하면서 파노라마 AV 기술과 함께 인터페이스 기술을 결합하여 3D의 입체감, UHD의 사실감과는 또 다른 극대화된 현실감을 제공한다[2].

파노라마 서비스를 제공하기 위해서는 크게 파노라마 영상 획득 기술과 생성 기술 및 재생기술이 필요하다. 특히 한 대 이상의 카메라에서 획득된 다수의 영상으로부터 파노라마 영상을 생성 하기 위해서 영상간의 기하학 보정을 위한 스티칭 기술과 색상 보정을 위한 블렌딩(blending) 기술이 필요하다. 스티칭 기술은 중첩된 영상에서 동일한 영역을 찾는 작업이 필요하다. 또한 입력 받은 두 영상이 비슷한 영역을 포함하고 있더라도 그 영역의 크기가 다르고, 회전 및 이동 된 상태와 조명 변화도 고려해야 하기 때문에 매우 어려운 작업이다. 기존에는 SIFT(Scale Invariant Feature Transform) [3]나 SURF(Speeded Up Robust Feature) [4] 기법이 많이 이용되었다. 위의 기법은 디스크립터를 통해서 매칭 과정이 이루어지는데 이 때 걸리는 수행 시간이 많이 걸린다.

본 논문에서는 파노라마 영상을 생성하기 위하여 고속의 스티칭 기법을 새로이 제안하고 있다. 입력이 비디오 영상일 경우 생기는 문제를 해결하기 위하여 아

래와 같은 방법을 제안한다. 한 영상에서 특징점을 추출 할 수 있는 방법은 다양한데 그 중 FAST(Features from Accelerated Segment Test) 기법 [5]을 이용한다. FAST 는 SIFT 의 DOG 방법이나 SURF 의 Haar wavelet 방식보다 특징점을 추출하는 속도 측면에서 월등하며, 특징점으로서 중요한 요소 중에 하나인 높은 반복성(repeatability)을 가지고 있다. 이 부분에서 일차적으로 속도를 감소 시키고 기존에 사용했던 디스크립터 형성과 정합 부분을 대신하여, 비디오 영상의 경우 앞 뒤 프레임간에 움직임이 작고 조명 변화 측면에서도 많은 차이가 없기 때문에 mean shift 물체 추적 기법을 이용하여 특징점을 추적한다 [6]. 추적한 방향을 토대로 앞 뒤 영상의 전체적인 이동 방향을 예측하여 정합과정을 진행함으로써 상당한 시간을 줄일 수 있다. 이후 기존의 방식과 비슷하게 RANSAC 을 통해 이상점을 제거 [7]하고 호모그래피 변환 [8]을 통해서 기존의 방식보다 수행 시간을 단축 할 수 있다.

본 논문의 구성은 다음과 같다. 2 장에서는 입력으로 비디오 영상이 들어 올 경우 제안한 각각의 기법에 대한 설명과 전체적인 수행 방법에 대해 설명한다. 3 장에서는 SIFT 와 SURF 에 의해 생성되는 파노라마 기법에 대한 수행시간과 제안한 기법의 수행 시간을 비교하고 결과를 제시한다. 4 장에서는 결론을 내린다.

#### 2. 제안하는 고속의 영상 스티칭 기법

본 논문에서 제안하는 고속의 영상 스티칭(stitching)

기법의 흐름도는 그림 1 과 같다.

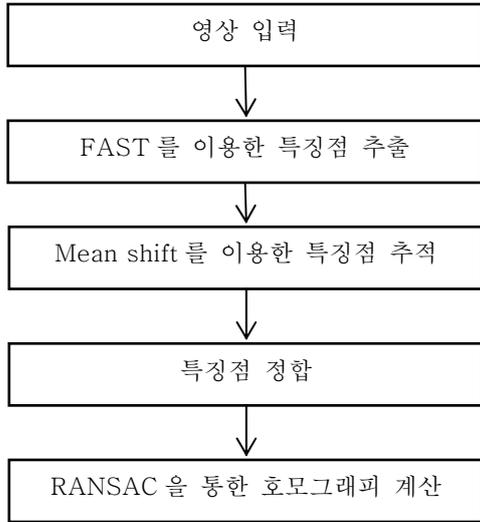


그림 1. 제안하는 고속 스티칭 기법의 흐름도

### 2-1. FAST 기법을 이용한 특징점 추출

FAST(features from Accelerated Segment Test) 기법은 기존의 특징점 추출 기법인 Harris corners[9], SIFT(Scale Invariant Feature Transform)의 DOG(Difference Of Gaussian), SURF(Speeded Up Robust Feature)의 Haar wavelet 과 SUSAN(Smallest Uni-Value Segment Assimilating Nucleus Test)[10] 기법 등과 특징점 추출 시간을 비교했을 때 상당히 빠르다는 장점이 있다. FAST 를 이용하여 특징점을 추출하는 방법은 다음과 같다. 그림 2 는 특징점을 결정하는 중앙 화소와 주변 화소의 관계를 보여주고 있다.

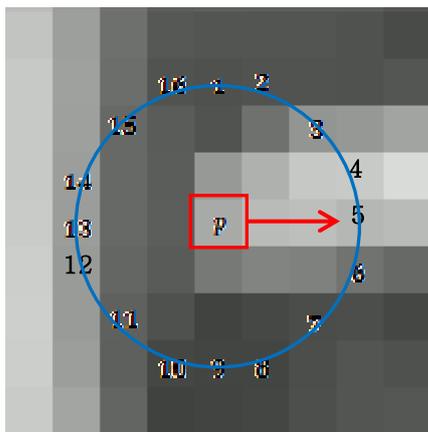


그림 2. 화소 p 를 중심으로 하는 원 위의 16 개의 화소

먼저 그림 2 와 같이 영상의 한 화소 p 를 중심으로 거리가 3 인 원을 형성한다. 형성된 원에 걸쳐있는 16 개의 화소들(그림 2 에서 원이 지나는 화소)과 중심 화소 p 의 밝기 값을 비교 한다. 16 개 화소의 밝기 값

과 p 화소의 밝기 값의 차이가 임계값(threshold) 이상인 값들이 연속적으로 N 개 이상 존재할 경우 중심의 p 화소를 특징점으로 정의한다. [5]에서는 N 이 9 일 경우 다음 프레임에서의 반복성(repeatability)이 가장 높고 임계값은 중심 화소 p 의 밝기 값에 따라 가변적으로 적용하여 밝기에 무관하게 특징점을 추출할 수 있도록 한다.

신속한 특징점 판별을 위해서 원을 구성하는 16 개 화소 중 N 개의 연속하는 화소가 서로 비슷한 밝기 값을 가져야 하는 특성을 고려해서 단순히 한 방향으로 검색하는 것이 아니라 가장 먼저 1 번 화소의 밝기 값을 비교하고 다음 9 번 화소, 5 번 화소 그리고 13 번의 화소의 밝기 값을 확인 한다. 특징점 조건을 만족하지 않으면 16 개의 화소 모두의 밝기 값을 비교하지 않기 때문에 속도를 개선할 수 있다. 특징점 조건을 만족하게 되면 시계방향으로 검색하면서 특징점 유무를 확인하게 된다.

### 2-2. Mean shift 를 이용한 특징점 추적

Mean shift 는 데이터 분포의 정점이나 무게 중심을 찾는 기법이다[13]. 이 기법은 영상에서 물체의 이동을 추적 하기 위해서 우선 물체가 포함된 영역에 대한 색상 히스토그램(histogram) 분포를 구한다. 그 다음 이전 영상에서 구한 히스토그램 정보를 가지고 다음 영상에서 이전 영역과 같은 크기로 모든 영역에 걸쳐서 히스토그램 정보가 가장 비슷한 영역을 찾는다.

제안한 특징점 정합 과정은 먼저 이전 영상에서 찾고자 하는 영역을 설정해 주는 작업이 필요하다. 특징점 추적이 가능하기 위해서는 설정한 영역이 이전 프레임과 다음 프레임에서 모두 존재 해야 되고, 알고리즘을 적용하기 전에는 영상의 이동 방향을 예측하지 못하기 때문에 영상의 중심에 가까운 영역일수록 영상의 이동방향과 상관없이 다음 프레임에 나올 확률이 높다. 따라서 특징점들 중에서 영상 중심에 가까운 특징점을 선택하여 이 특징점이 포함된 일정한 크기의 사각형 영역을 설정해 준다. 설정한 사각형 영역에 Mean shift 를 이용하여 다음 프레임에서 이동 영역을 추적한다.

### 2-3. 특징점 정합

입력으로 들어오는 비디오 영상에서 일반적으로 각 프레임 간의 움직임의 변화는 크지 않다. 프레임 간의 변화가 작기 때문에 영상의 전체적인 이동 방향을 같다고 볼 수 있다. 2-2 절에서 구한 영역의 이동 벡터를 통해 정합 과정을 진행한다.

### 2-4. RANSAC 을 통한 호모그래피 계산

두 영상을 하나의 영상으로 만들어 주기 위해서는 하나의 카메라 좌표계를 중심으로 호모그래피(homogr\_

aphy) 변환을 적용해야 한다. 호모그래피 행렬은 정합된 특징점 네 쌍을 가지고 구할 수 있다. 행렬을 구할 네 쌍의 특징점들을 선택하는 과정에서 이상점(outlier)이 스티칭을 하기 위한 특징점 쌍으로 이용될 경우 잘못된 호모그래피 행렬을 구할 수 있다. 이러한 이상점들을 제거하기 위하여 RANSAC 기법을 사용한다. RANSAC 기법은 특징점들의 쌍 중에서 무작위로 4 개의 쌍을 뽑아서 호모그래피 행렬을 생성한 후 이전 프레임을 호모그래피 변환한다. 호모그래피 변환에 의해 이전 영상 특징점들의 좌표가 변환된다. 변환된 좌표와 유클리디언 거리를 통해서 가장 차이가 작은 특징점 네 쌍을 찾아서 영상을 스티칭한다.

### 3. 실험 방법 및 결과

파노라마 영상을 생성하기 위해 본 논문에서는 1920x1080 해상도의 Samsung 사의 갤럭시 노트로 직접 획득한 비디오를 실험 영상으로 사용한다. 실험 영상은 각각 정수기와 건물이 포함된 영상이다. 실험 환경은 Microsoft 사의 Microsoft Visual Studio C++ 2010 과 OpenCV 2.4.8 라이브러리를 이용하여 구현하고, 3.40GHz 의 인텔 i5 쿼드코어 프로세서를 이용한다.

표 1 은 특징점 추출을 위해 제안한 FAST 와 기존에 주로 사용하는 SIFT 의 DOG 와 SURF 의 Haar wavelet 부분의 수행 시간을 비교하고 각 기법에 따라 추출된 특징점 개수를 비교한다. 기존의 DOG 와 Haar wavelet 를 사용하는 대신에 FAST 를 이용했을 때 수행 시간을 표 1 과 같이 줄일 수 있다.

표 1. 특징점 추출 수행 시간 및 특징점 개수 비교

방법	DOG (SIFT)	Haar wavelet (SURF)	제안하는 기법
정수기 (시간)	2071.9ms	135.8ms	3.23ms
정수기 (특징점 개수)	746	542	284
건물 (시간)	4553ms	176.81ms	5.179ms
건물 (특징점 개수)	3739	1995	1236

표 2 는 mean shift 이동 방향 검출을 이용한 정합 방법과 기존의 SIFT 와 SURF 를 통한 정합 시간의 비교 결과이다. SIFT 와 SURF 에서 사용하는 기법은 특징점 정합을 위해 디스크립터를 형성하고 이를 비교 요소로 사용하지만 제안한 방법은 입력이 비디오 영상일 경우 디스크립터를 사용하지 않고 이동방향을 기반으로 정합 과정을 수행하였기 때문에 표 2 와 같이 빠른 수행 시간의 결과를 얻을 수 있었다. 표 1 에서 같은 특징점 추출 기법이더라도 특징점 추출 개수가 다를 수 있다.

표 2. 특징점 정합 수행 시간 비교

Table 2. Comparison of the feature matching time

방법	SIFT	SURF	제안하는 기법
정수기 (시간)	355.7ms	112.71ms	14.89ms
건물 (시간)	1326.1ms	413.7ms	26.97ms

### 4. 결론

본 논문에서는 비디오 영상이 입력으로 들어 올 경우 기존에 사용하는 방법 대신 수행 시간을 줄이기 위한 새로운 스티칭 기법을 제안하였다. 제안하는 기법에서는 비디오 영상은 프레임간에 움직임이 많이 없다고 가정하였다. 입력으로 들어온 영상에 대해 FAST 기법을 통해 빠른 시간 안에 특징점을 추출 할 수 있었다. 기존에 사용하였던 디스크립터 생성과 이를 이용한 정합 방법보다 mean shift 를 이용한 특징점 영역의 이동방향을 추정하여 두 영상간의 특징점 정합 과정을 수행하여 기존의 방식보다 빠른 속도를 얻을 수 있었다.

### ACKNOWLEDGMENT

본 연구는 미래창조과학부의 정보통신방송연구개발사업의 일환으로 수행하였음. [라우드니스 기반의 방송음량 기술 및 실내 환경 소음의 스트레스 평가 기술 개발 (2014-044-055-022)]

### 참 고 문 헌

- [1] Y. J. Cho, J. M. Seok, S. Y. Lim, S. W. An, J. I Seo, and J. H. Chan, "Post-UHD Realistic media, high quality panoramic AV technology", *Electronics and Telecommunications Trends*, vol. 20, no. 3, pp. 33-46, June 2014.
- [2] Y. J. Lee, Y. J. Joe, M. S. Ki, S. Y. Lim, H. G. Lee and J. H. Cha, "고품질 휴먼융합형 파노라마 서비스", *The Korean Institute of Communications and Information Sciences*, vol. 28, no. 6, pp. 11-20, May 2011.
- [3] D. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, November 2004.
- [4] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, "Speeded-up robust feature", *Computer Vision and Image Understanding*, vol. 10, no. 3, pp. 346-359, June 2008.
- [5] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection", *9th European Conference on Computer Vision*, Graz, Austria, pp. 430-443, May 2006.
- [6] D. Comaniciu, V. Ramesh and P. Meer, "Real-time tracking of non-rigid objects using mean shift", *Proc. 2000 IEEE Conference Computer Vision and Patter Recognition*, vol. 2, pp. 142-149, June 2000.
- [7] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image

- analysis and automated cartography”, *Communications of the ACM*, vol. 24, no. 6, pp. 381-395, June 1981.
- [8] E. Dubrofsky, “Homography estimation”, UNIVERSIT OF BRITISH COLUMBIA, March 2009.
- [9] C. Harris and M. Stephens, “A combined corner and edge detector”, *proceedings of the 4<sup>th</sup> Alvey Vision Conference*, pp. 147-151, August 1988.
- [10] S. M. Smith and J. M. Brady, “Susan – a new approach to low level image processing“, *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45-78, May 1997.
- [12] D. Comaniciu and P. Meer, “Mean shift: a robust approach toward feature space analysis”, *IEEE Computer Society*, vol. 24, no. 5, pp. 603-619, May 2002.