2016 년 가을 포스텍 컴퓨터공학과 과제연구 최종보고서

Re-composition of 360-degree Images into Human-like Photos

학 번: 20140996
이 름: 황일환
연구 지도교수: 이승용

---

* 아래 사항들에 대해 20 페이지 이내로 명료하게 서술한 뒤, 지도교수의 검토를 거쳐 pdf 형태로 제출합니다. 이를 지키지 않을 시 감점입니다.

연구 목적

본 연구는 360 도 이미지에서 흥미로운 부분을 찾아내고 그 부분을 실제 사람이 찍은 것 같은 구도로 보기 좋게 composite 하는 것이 목적이다.

연구 배경

360 도 카메라의 등장과 함께 촬영이라는 행위는 새로운 국면을 맞이하게 되었다. 360 도 카메라는 모든 방향의 정보를 담을 수 있기 때문에 찍는 순간에 어디를 찍을지 고민하지 않아도 된다. 일단 원하는 순간을 촬영한 후 나중에 원하는 부분을 편집해서 잘라내면 되기 때문이다.

하지만 360 도 영상은 그 특성상 영상을 보고 어디를 취해야 할지 명확하지 않다. 사람의 인지는 한계가 있기 때문에 공간과 시간을 모두 고려하면서 어디가 가장 의미 있는 부분인지 생각하는 것은 굉장히 힘들다. 특히 영상 소스가 굉장히 많을 때는 더욱 그렇다.

그렇기 때문에 자동으로 360 도 영상을 편집할 수 있는 기능이 필요하게 된다. 본 연구에서는 영상 자체를 편집하는 기능은 일단 후속 연구로 두고 먼저 한 장의 사진에서 자연스러운 composition 을 자동으로 만들어내는 것에 초점을 두었다.

연구 방법

- 일반적인 saliency-detection 알고리즘을 360 도 이미지에 적용하는 방법에 대해 연구
- 실제 사진을 찍을 때 필요한 composition 기술에 대한 분석
- 각 composition 과 주어진 input scenery 상황의 matching 방법 연구

연구 결과 및 평가

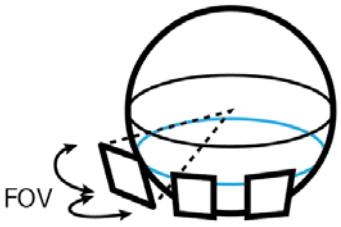
360 도 이미지 위에서의 Saliency-detection

기존의 연구들은 모두 평범한 2 차원 이미지 위에서의 saliency detection 만을 다루고 있다. 하지만 360 도 이미지의 경우 왜곡이 있기 때문에 이를 그대로 적용할 수 없다. 이 섹션에서는 360 도 이미지 위에서 왜곡을 고려해 saliency detection 을 하는 방법을 소개한다.

Saliency detection 에는 <What Makes a Patch Distinct?>[i] 에 소개된 알고리즘을 참고하였다. 원문에서 제시한 방법은 다음과 같다.

1. 이미지 상의 patch 들을 dense 하게 sample 하여 patch space 위에서 PCA 를 풀고 PCA 와의 거리를 그 픽셀의 saliency 로 할당한다.
2. <Global Contrast Based Salient Region Detection>[ii]에 소개된 방법으로 color salient map 을 구한다.
3. 위 두 과정에서 구한 saliency 를 곱하여 최종 saliency 를 만들어낸다.

구면 위에서의 patch



구면 위에서 patch 를 dense 하게 .sample 하기 위해 sliding window approach 를 사용한다. 360 도 이미지를 어떤 작은 FOV 를 가지는 화면에 투사한 것을 하나의 patch 로 사용하고, 이를 φ를

바꾸어 가며 $\vec{\theta}$방향으로 여러 개 샘플 한다. 이 때 sample density 를 uniform 하게 유지해야 하므로 어떤 한 φ에 대해 생기는 원을 따라 만들어지는 샘플 수는 다음과 같다.

$$\mathrm{N}(\varphi) = \frac{\mathrm{N}(0)}{\cos\varphi}$$

위와 같이 patch 를 sample 하면 uniform 에 근사 하는 density 를 얻을 수 있다. 실제 예제에선 $\mathrm{FOV} = 5°$인 patch 를 2.5°마다 sample 하였다.

구면 위에서의 segmentation

<Global Contrast Based Salient Region Detection>은 이미지의 색 히스토그램을 이용해 어떤 한 색상이 얼마나 다른 색상과 멀리 떨어져 있는지를 saliency 로 사용한다. 이 saliency 를 계산할 때 한 픽셀 단위로 하는 것이 아니라 super pixel 을 만들어서 그 super pixel 의 히스토그램끼리 비교하게 되는데, 이 super pixel 을 만드는 알고리즘을 360 도 이미지에 알맞게 바꿔주는 작업이 필요하다. 이 방법은 <Efficient graph-based image segmentation>[iii]에 소개되어있는 그래프 기반 알고리즘이다. 이 방법은 각 픽셀을 vertex 로 취급하여 이웃 픽셀을 edge 로 연결하고 그 그래프에서 MST 를 찾는 것으로 segmentation 을 만들어 낸다. 여기서 vertex 가 많이 연결될수록 다른 것을 연결하기가 힘들어지는데, 여기서 픽셀에 weight 를 줌으로써 실제 픽셀의 크기를 반영하여 360 도 이미지에서도 사용이 가능한 알고리즘으로 바꾸었다. 각 vertex v마다 weight w_v가 있을 때, $|C|$를 다음과 같이 재정의한다.

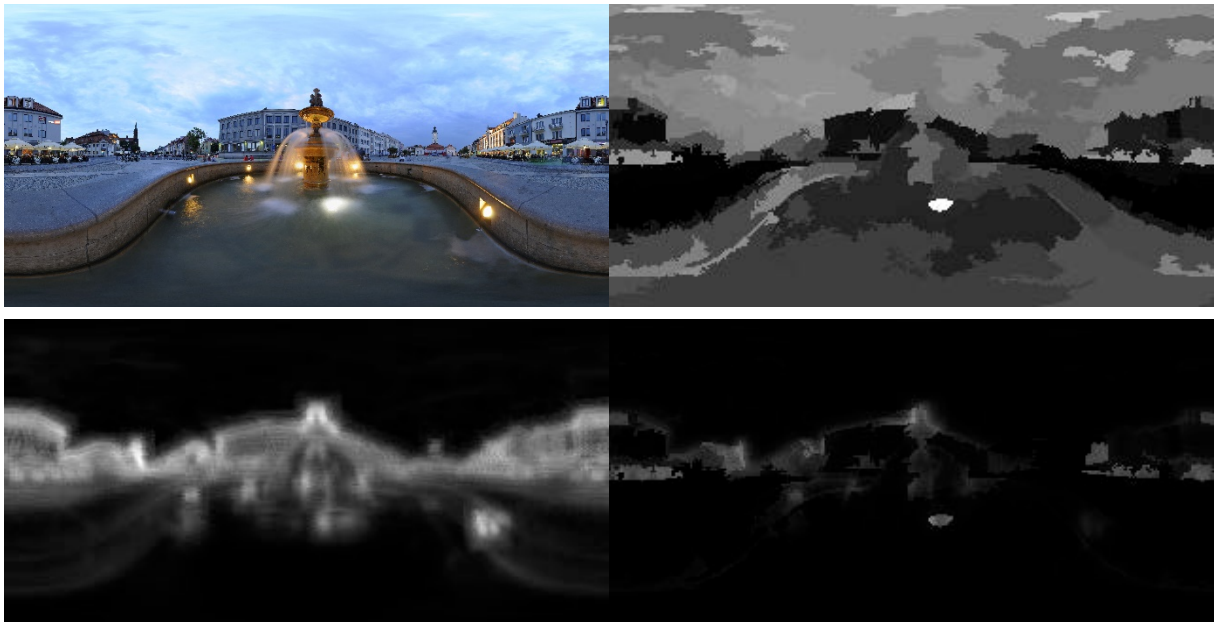
$$|C| = \sum_{v \in C} w_v$$

w_v가 실제 픽셀의 넓이를 나타내야 하므로 구면상에서 (θ, φ)위에 있는 픽셀의 vertex v의 weight w_v는 다음과 같다.

$$w_v = \frac{1}{\cos\varphi}$$

이 weight 값은 히스토그램을 만들 때에도 사용된다. 히스토그램에 기여하는 정도가 실제 픽셀의 넓이 만큼이어야 하기 때문이다.

Saliency detection 결과



▲원본 이미지 [iv] ▲Global Contrast Saliency
▲Patch Distinction Saliency ▲두 saliency map 을 곱한 것

실제 사진의 composition 기술 분석

3 분할 법칙

사진을 찍을 때 매우 일반적으로 쓰이는 기법 중 하나는 바로 '3 분할 법칙'이다. 이는 수평선, 피사체, 그 혹은 그 외의 중요한 것들이 화면의 3 분의 1 지점에 와야 안정감을 얻을 수 있다는 법칙으로, 이 법칙만 지켜져도 상당히 높은 수준의 사진을 얻을 수 있다.

화면 구성

사진을 찍을 때 화면 상의 물체의 배치들을 이용해서 다양한 구도를 만들어낼 수 있다. 배경과 피사체의 구성은 다음과 같은 것들이 있다.

▲삼각형 구도 [v] ▲3 분할 직선 구도 [vi]
▲3 분할 지점에 물체가 있는 구도 [vii] ▲원형 구도 [viii]

FOV

사진을 찍을 때 사용되는 FOV 의 범위는 한정되어있다. 광각렌즈는 104°정도, 그리고 망원렌즈는 26°정도의 FOV 를 갖는다. 사람의 눈과 가장 흡사한 FOV 는 47°이다.

Input scenery 에 맞는 composition 찾기

Composition 은 2 차원으로 나열된 픽셀의 집합이다. 주어진 360 도 이미지에서 re-project 하여 만든 '적절한' composition 은 다음과 같은 특성을 만족해야 한다.

1. Composition 그 자체로 3 분할 법칙이나 화면 구성을 이룬다.
2. Composition 의 중요한 부분에 salient 한 region 이 많이 포함된다.

여기서 주어진 360 도 이미지는 이미 수평선에 맞춰서 촬영되었다고 가정한다. 즉, $\varphi = 0$은 수평선이다. 이 가정으로부터 다른 가정을 만들어볼 수 있다.

1. 수평선과 멀어질수록 중요하지 않은 물체이다.
2. 실제 사람이 사진을 찍을 때는 고개를 위 아래로 많이 움직이지 않는다.

이 두 가정은 아래 value function 에서 등장할 것이다.

Composition 의 form value 측정

Form value 란 composition 이 얼마나 사용자가 요구하는 구도에 맞는지를 판단하는 척도이다. Composition 에 구도가 있다는 것은 composition 을 커다란 구역으로 나눌 수 있고 그 구역들 사이에 기하학적 관계가 있음을 뜻한다. Form $F = \{F_1, F_2, \dots, F_n\}$ 을 composition C 의 partition 이라고 하자. F_k의 히스토그램이 $H_k = \{h_1, h_2, \dots, h_m\}$라면 form F에 대한 composition C의 form value 는 다음과 같이 정의된다.

$$form_F(C) = \sum_{1 \le i < j \le k} \frac{Dist(H_i, H_j)}{k+1} - \sum_{1 \le i \le k} \frac{Dist(H_i, H_i)}{k}$$

여기서 $Dist(H_i, H_i)$는 두 히스토그램의 거리로, 다음과 같이 계산한다. 실제 개발에서 color space 는 Lu*v* color space 를 사용했고, norm 은 L2-norm 을 사용했다.

$$Dist(H, G) = \sum_{i,j \in color(H) \cup color(G)} |Color_i - Color_j| \cdot h_i \cdot g_j$$

위 두 식을 통해 form 이 제공하는 대로 composition 을 나눌 경우 각 구역 사이의 contrast 는 최대가 되고 구역 내에서의 contrast 는 최소가 되어야 높은 form value 를 얻을 수 있음을 알 수 있다.

Composition 의 saliency value 측정

Saliency value 란 composition 의 중요한 부분에 얼마나 중요한 정보를 담고 있느냐를 판단하는 척도이다. 이는 아까 사용된 form 에서 가장 첫번 째 원소, 즉, 중요한 영역이라고 가정하는 F_1이 saliency map 과 얼마나 overlap 되는지를 측정한다. 수식은 다음과 같다.

$$sal_F(C) = \sum_{p \in F_1} I_{sal}(p)$$

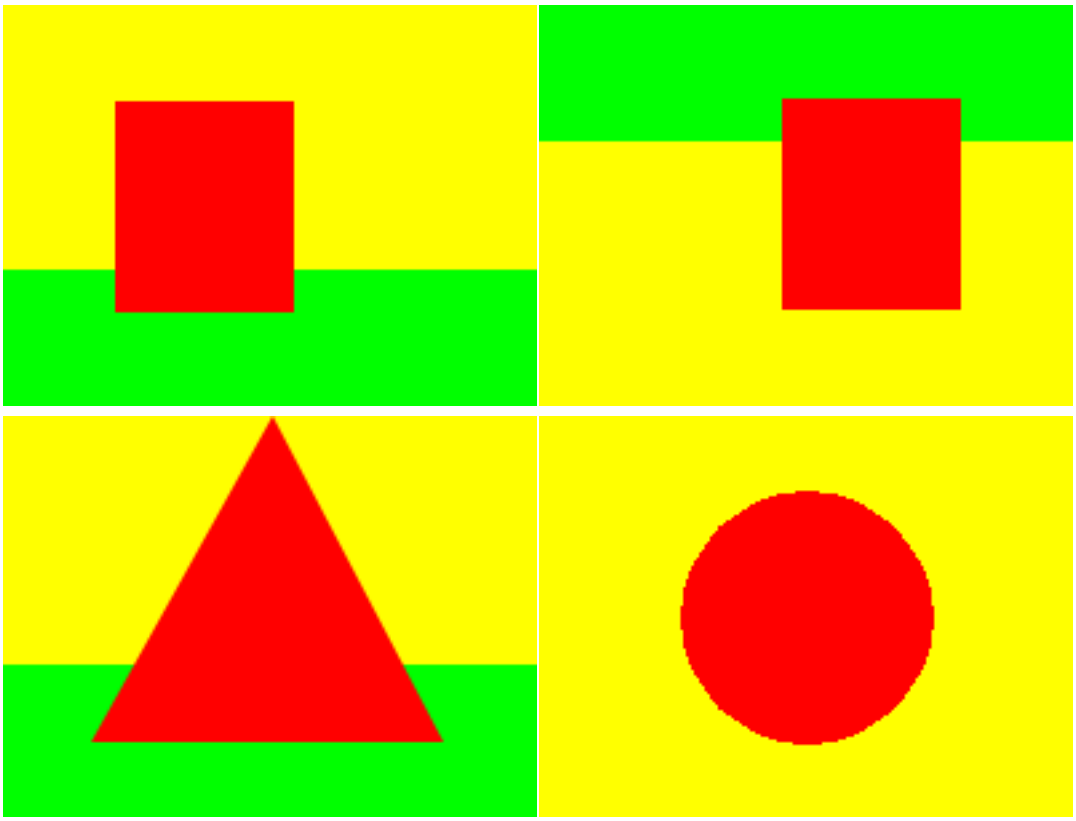
Composition 의 value

어떤 주어진 Form 에 대해서 composition 의 value 는 위에서 얻은 두 식의 곱에 φ값에 대한 함수를 곱한 것으로 정해진다.

$$value_F(C) = form_F(C) \cdot sal_F(C) \cdot (\cos \varphi)$$

우리의 목표는 가장 높은 $value_F(C)$를 주는 composition C를 찾는 것이다.

Finding Maxima: Sliding window approach

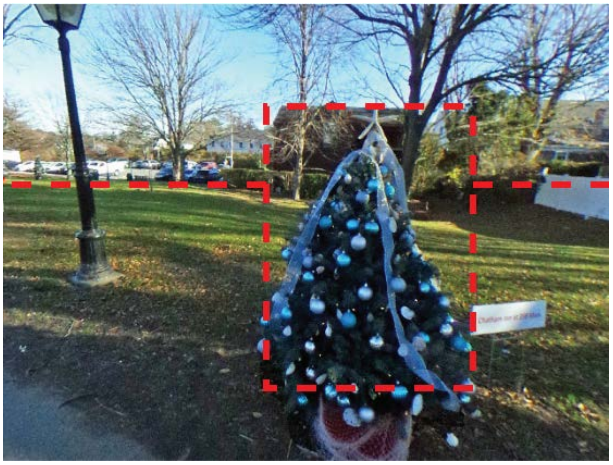


▲3 분할 지점에 물체가 있는 구도 ▲3 분할 지점에 물체가 있는 구도(반전)
▲삼각형 구도 ▲수평선이 없는 원형 구도
빨간 영역이 F_1이다.

Form 에 맞는 composition 을 찾기 위해 sliding window approach 를 사용하였다. 위에서 patch 를 sample 할 때와 마찬가지로 φ 의 값에 따라 sample 수를 다르게 하여 uniform 하게 sample 되도록 하였다. 한 위치에서 각각 다른 다섯개의 FOV 로 여러 번 sample 하였다. FOV 는 실제 카메라에 쓰이는 값 대로 $28°, 47°, 63°, 84°, 104°$로, 광각부터 망원까지 커버하였다. φ가

큰 곳에서는 $\cos\varphi$ 값이 매우 작아 value function 이 낮을 것이 분명하므로 아예 sample 하지 않았다.

Sample 은 18°마다 하여 다소 coarse 하게 하였고, 대신 sample 후 non-maxima suppression 을 가한 뒤 남은 composition 중 가장 value function 이 높은 3 개를 골라 상하좌우로 조금씩 움직이거나 FOV 를 조금씩 늘리고 줄이면서 local maxima 에 도달하도록 하였다.



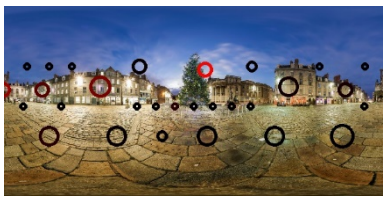
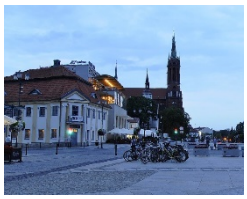
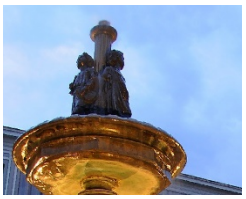
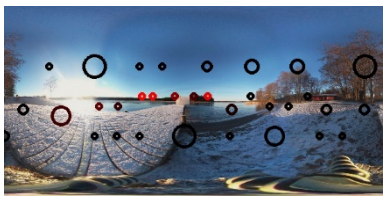
▲원본 이미지

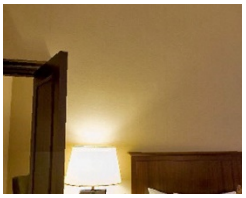
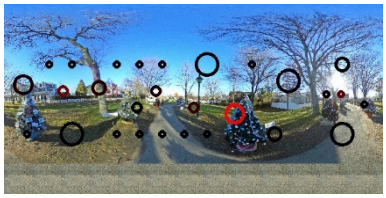
▲결과 composition

두 번째 Form 에 맞추어서 composition 이 만들어진 것을 확인할 수 있다. 하얀 경계선으로 나누어진 영역을 보면 F_1 영역은 짙은 녹색, 위쪽 영역은 하늘색, 아래쪽 영역은 녹색이 대부분을 이루고, 이들이 강한 대비를 이룬다.

실행 결과

아래 실행 결과는 3 분할 form 과 그것의 뒤집힌 형태를 같이 이용해서 만들어낸 composition 이다. 두 form 을 이용할 때에는 각 form 마다 value function 을 구하여 높은 쪽을 취하면 된다.

원본 사진 및 detection	1 순위	2 순위	3 순위
			
			
			
			
			

원본 사진 및 detection	1 순위	2 순위	3 순위
			
			

평가

심미성

실제로 ground truth 가 존재하는 것이 아니므로 정량적인 평가는 할 수 없으나, 구도가 잘 갖추어진 좋은 composition 이 나오는 것을 확인할 수 있었다. 특히 두 번째 사진의 2 순위나 4 번째 사진의 1 순위 composition 의 경우 사람이 의도하고 찍은 것이라고 해도 손색이 없을 정도의 결과를 얻을 수 있었다. 하지만 form 형태의 특성상 비슷한 구도가 많이 나오게 되므로 더 많은 form 을 사용하여 만든다면 더 좋은 결과를 얻을 수 있을 것이다. 또한 saliency detector 의 특성상 광원에 높은 saliency 가 몰리게 되어 2 번째 사진의 3 순위 composition 처럼 별로 중요하지 않은 곳에 composition 이 생기는 것을 볼 수 있는데, saliency detector 를 바꾸어서 해결해야 할 것이다.

연산속도 및 정확도

연산속도는 매 스텝마다 히스토그램(n=40)을 만들어 평가하는 방법이라 매우 느리다. 또한 sliding window 기법을 기반으로 작동하기 때문에 속도도 느릴 뿐더러 global optima 라는 보장을 할 수 없다. OpenCV 를 다루는 것이 미숙해서 생긴 overhead 도 상당하므로 좀 더 다듬으면 더욱 빠르게 연산할 수 있을 것이다.

토론 및 전망

연구 초기에 기대했던 것처럼 액션 카메라에 탑재되기 위해서는 몇 가지 더 필요한 요소가 있다. 첫 번째는 시간에 대한 고려, 그리고 두 번째는 객체 인식이다. 현재의 알고리즘은 중요한 부분을 눈에 보기 좋은 구도로 담는 것이 전부이지만, 동영상에서 활용되려면 앞으로 구도가 어떤 식으로 영향을 받을 것이고 구도의 주체(F1 영역의 물체)가 어떻게 움직일지도 고려해야 할 것이다. 나중에 영상이 될 때엔 사진 기술 뿐 아니라 영상 촬영 기술도 분석하여 같이 고려하여 만들어야 할 것이다.

참고 문헌

[i] Margolin, R., Tal, A., & Zelnik-Manor, L. (2013). What makes a patch distinct?. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1139-1146).

[ii] Cheng, M. M., Mitra, N. J., Huang, X., Torr, P. H., & Hu, S. M. (2015). Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(3), 569-582.

[iii] Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, *59*(2), 167-181.

[iv] https://farm4.staticflickr.com/3919/15209793877_8fd36f1341_o.jpg

[v] http://photovil.hani.co.kr/files/attach/images/82/261/019/80873_33503.jpg

[vi] http://cfile28.uf.tistory.com/image/261981385482478E1FDE36

[vii] http://pse-mendelejew.de/en; by Alchemist-hp

[viii] http://chulsa.kr/files/attach/images/6220430/531/624/018/05bc7041edc0837ce3464c6288175743.jpg