

시각적 주목도를 이용한 정지영상의 움직임 생성 기법

Automatic Still Shot Movement Generation Based on Visual Attention

이 상 민

Sangmin Lee

성균관대학 컴퓨터공학과

Dept. of Computer Engineering

Sungkyunkwan University

krsm1@skku.edu

이 성 길

Sungkil Lee

성균관대학 컴퓨터공학과

Dept. of Computer Engineering

Sungkyunkwan University

sungkil@skku.edu

요약문

정지 영상에 동적인 효과를 주기 위해 panning 이나 zoom 과 같은 카메라의 움직임을 영상에 더하기도 한다. 이러한 카메라의 움직임은 보통 디자이너가 수동으로 제작하는 것이 보편적이거나, 다수의 이미지에 대한 처리가 어렵고 디자이너의 안목에 의존하게 된다. 본 논문은 이러한 정지 영상의 움직임을 자동으로 생성하는 기법을 제안한다. 정지 영상을 움직이기 위한 중요한 지점들은 시각적 주목도를 기반으로 찾아지며, 주목도의 정도에 따라 움직임의 경로를 설정하여 정지영상을 움직일 수 있다. 본 기법은 10 사람이 정지 영상을 보는 것과 유사한 자연스러운 움직임을 빠르게 자동으로 생성할 수 있다.

ABSTRACT

Static images are often animated as camera moves with panning and zooming to give dynamic impression on them. In general, such camera movements are manually created by designers, which are often costly and relies heavily on designer's insights. This paper presents a technique to automatically generate the movements of still shots. Important spots in images are found using visual saliency, and their saliency levels are used to configure camera paths. Our technique can quickly generate natural movements, similar to how people see static images.

주제어

Visual attention, Still shot, Image browsing

서론

다큐멘터리의 인트로, 여행지 소개 영상 또는 기념 영상 등에서 연속적인 still shot 을 이용하여 영상을 제작하는 경우가 많다. 정적인 영상에 동적이고 보다 극적인 효과를 주기 위해 화면 간에 fade-in/out 효과를 주거나 다음 화면이 이전 영상을 밀어내는 등의 다양한 방법이 존재한다. 그 중 카메라의 움직임과 흡사하게 이미지를 움직여 (panning 과 zoom) 영상을 제작하기도 한다.

이러한 움직임이 포함된 영상은 사람의 눈길이 많이 가는 부분을 정하고 해당 부분을 확대 및 축소하거나 경로를 설정하여 정지된 이미지의 연속에서 움직이는 영상으로 렌더링하게 된다. 그러나 영상을 제작하기 위해 준비한 많은 수의 이미지 속에서, 극적 효과를 위해 직접 스케일을 조정하거나 경로를 설정하는 작업은 많은 노력이 필요하며 자동화하기 어렵다.

본 논문은 시각적으로 중요한 지점들을 자동으로 찾고, 이를 바탕으로 카메라의 움직임 경로를 설정하는 방법을 제시한다. 실제 사람의 시각체계의 원리와 유사하게 이미지에서 눈에 잘 띄는 부분을 찾기 위해 visual saliency map[1]을 이용한다. Saliency map 에서 다수의 중요한 지점을 순서대로 찾기 위해 현재 가장 높은 saliency 지점을 가리는 inhibition of return 방법을 사용한다. 이러한 중요한 지점들은 애니메이션을 위한 기준점들로서 사용된다.

관련연구

시각적 관심에 영향을 주는 요소를 찾는 방법은 크게 두 가지로, 이미지의 특징 기반 bottom-up 접근 방식(saliency 기반)과 사용자의 의도를 반영한 top-down 접근 방식이 있다. 본 논문에서는 사용자의 의도를 모르는 상태에서, 이미지의 특징만을 이용하여 카메라 경로를 이용하므로, visual saliency 에 대한 간략한 리뷰를 한다.

Itti 는 입력된 이미지에 대해 bottom-up 접근을 제안했다[1]. 필터링을 통하여 color, intensity, orientation 3 가지의 map 으로 분류하고 이를 이용해 feature map 을 생성하였다. 구해진 Saliency map 은 사람 시각체계의 자극과 반응에 기반을 두고 계산되며 이미지에서 눈길을 끄는 부분이 어느 부분인지 알 수 있게 하였다.

Liu 는 모바일 장치에서 크기가 큰 사진을 볼 때 작은 화면의 한계점을 개선하기 위하여 saliency 를 이용하고 그에 따라 알맞게 화면을 자동적으로 움직임으로써, 사용자가 직접 사진을 확대하고 움직임 필요가 없도록 하였다[2]. Ip 과 Varshney 는 크기가 큰 사진 중 특히 풍경사진에서 시각적으로 흥미를 끌 수 있는 부분으로 이미지를 탐색하는 방법을 제안하였다[3].

SALIENCY 포인트를 선택하기 위한 알고리즘

본 논문에서는 주어진 이미지에 대해서 saliency 를 측정하고, 보간 기법 중 Catmull-Rom spline 을 사용해 사용자가 원하는 개수의 saliency 포인트 중 가장 높은 saliency 값을 가지고 있는 포인트부터 차례대로 자연스럽게 화면이 이동하는 방법을 제안한다.

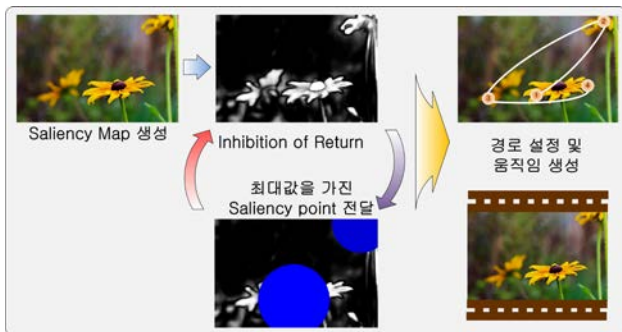


그림 1. 알고리즘 개요

- Saliency Map 생성

Saliency map을 생성하기 위한 기본적인 접근은 [1]에서 제시한 방법론을 기초에 둔다. 이 방법은 input 이미지를 3가지 채널로 분리시킨 후, 각 채널에 대해 down sampling을 이용하여 가장 낮은 이미지 피라미드 수준(level)인 원본 이미지부터 가장 높은 수준인 하나의 픽셀을 가진 이미지까지 이미지 피라미드를 생성시킨다. 이 때 이미지 피라미드의 수준은 이미지의 blur 정도와 관계 있으므로, center(c) 와 surround(s)의 수준을 $c \in \{2, 3, 4\}$, $s = c + \delta$, $\delta \in \{3, 4\}$ 인 경우로 6개의 (c, s) 짝을 정의한다. (c, s) 짝에서 center-

surround difference는 coarser 이미지를 finer 한 이미지로 up scaling 하고 서로의 차이를 통해 각 채널의 feature map들을 구할 수 있으며, 이 feature map들의 결합을 통해 saliency map을 생성할 수 있다. 본 논문에서는 feature map을 생성하는데 있어 위의 접근방식과는 상이하게 [4]에서 제안한 방법과 같이 CIE L*a*b* 색 공간에서의 3D Euclidean distance를 사용하여 이 거리를 유일한 feature로 사용하고, 이를 통해 나온 6개의 feature map의 결합을 통해 saliency map을 생성한다. 이러한 접근은 Lab distance를 유일한 특징으로 사용하므로 이미지의 색, 밝기, 방향의 3가지 feature map을 고려하는 [1]의 방식과 달리, 선형결합의 계수를 구할 필요가 없어 편리하다. 그림 2는 위에서 설명한 방법을 이용해 생성한 saliency map을 예시한다 (밝은 부분이 높은 saliency를 나타낸다).



그림 2. 입력 이미지와 생성된 Saliency Map

- Saliency의 최대값 탐색

Saliency map에서 가장 눈에 띄는 점을 찾기 위해 maximum mipmap[5]을 이용한다. Saliency map을 이용하여 새로운 텍스처의 rgb 채널에는 xyz좌표를, alpha 채널에는 해당 픽셀의 saliency를 저장한 후, 이를 이용하여 maximum mipmap을 실시하며, 이때 고수준 이미지의 픽셀은 1 수준 낮은 이미지의 saliency 값을 저장한다. 이 저장되는 saliency 값은 이미지에서 salient 범위가 크고 작은 부분 중 큰 부분이 눈에 더 잘 될 것이므로 maximum mipmap시 주변의 saliency 값의 평균에 가중치를 두어 저장한다. Maximum mipmap이 완료되면 1x1 크기의 이미지가 생성되고, 이 최고수준 이미지의 해당 픽셀은 전체 saliency map에서 최대값을 가진 점의 좌표 값과 saliency 값을 가지게 된다.

- Inhibition of Return

Saliency map에서 saliency 값이 높게 나오는 부분의 집합을 이용하여 움직임을 만들기 위해서는 최대값 탐색을 이용하여 여러 개의 좌표 값이 필요하다. 이를 위해 최대값 탐색 시 이전 최대값이었던 부분을 선택하지 않기 위하여 [1]과

같이 inhibition of return 방법을 사용한다. 다만 본 논문에서는 [1]에서 제시한 전체 이미지 크기의 1/6 크기의 원을 그려 saliency map을 가리는 방법을 택하기보다는 계산된 최대 평균 saliency 값을 이용하여 원의 크기에 가중치를 두어 saliency map을 가리는 방법을 택하였다. 그림 3은 saliency map에서 산출된 saliency 값에 따라 적절한 크기의 원을 이용한 inhibition of return을 구현의 예를 보인다.

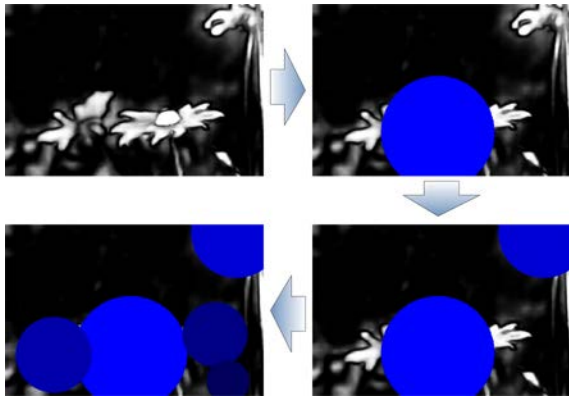


그림 3. 포인트 개수에 따른 Inhibition of Return

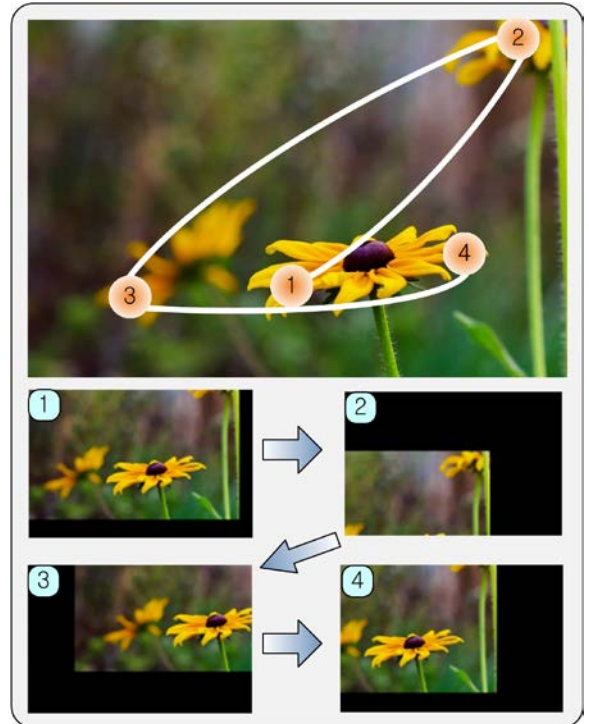


그림 4. Saliency Map 기반 포인트와 경로 및 실행 영상 □

결과 및 토론

그림 4는 주어진 이미지에서 가장 높은 4개의 saliency 포인트를 Catmull-Rom spline에 따라 생성한 카메라의 움직임 곡선을 보인다.

그림에서 보듯이, saliency map에서 얻은 점들을 이용한 카메라의 움직임은 가장 선명하게 나온 꽃으로부터 나머지 2개 꽃을 거쳐 돌아오며, 이 움직임은 곡선을 통해 자연스럽게 이동한다. 따라서 사용자는 직접 경로를 설정하지 않고도 사람의 시선을 따라가는 카메라 움직임을 얻을 수 있어, 자연스럽게 효율적인 영상생성이 가능하다. Saliency map으로부터 얻어진 곡선과 실제 사람이 사진을 보았을 때 시선의 움직임을 비교하여 사람들이 하는 결과와 생성된 결과가 유사한지에 대한 사용자 실험이 필요하다.

사사의 글

이 논문은 미래창조과학부의 재원으로 한국연구재단의 지원을 받아 <실감교류 인체감응솔루션> 글로벌 프론티어사업, 중견연구자지원사업으로 수행된 연구임 (2012M3A6A3055695, 2012R1A2A2A01045719)

참고 문헌

1. Itti, L., Koch, C., & Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence* (1998), 20(11), 1254-1259.
2. Liu, H., Xie, X., Ma, W. Y., and Zhang, H. J. Automatic browsing of large pictures on mobile devices. In *Proceedings of the eleventh ACM international conference on Multimedia*, (2003), pp. 148-155.
3. Ip, C. Y., and Varshney, A. Saliency-assisted navigation of very large landscape images. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12), (2011), 1737-1746.
4. Kim, M., Lee, S., and Choi, S. Saliency-Driven Real-Time Video-to-Tactile Translation, (2013).
5. Tevs, A., Ihrke, I. and Seidel, H. P. Maximum mipmaps for fast, accurate, and scalable dynamic height field rendering. *Proc. ACM Symp. 13D* (2008), 183-190.