

# Psych 221 Final Project

## Saliency Detection

Zixuan Wang

### Abstract

This paper presents a saliency detection algorithm, which aims at detecting the image regions that represent the scene. The algorithm is based on three psychological evidences. By considering the local and global information and visual saliency organization rule, we extract the saliency map from the input image. The experimental results indicate robust saliency detection of our algorithm.

### 1. Introduction

Given an image, human can detect salient regions from the image extraordinarily fast and reliable. The salient regions in the image may contain foreground, parts of the background, interesting patterns and so on. Figure 1 shows a typical saliency detection task, where we want to extract areas contain the leaf.

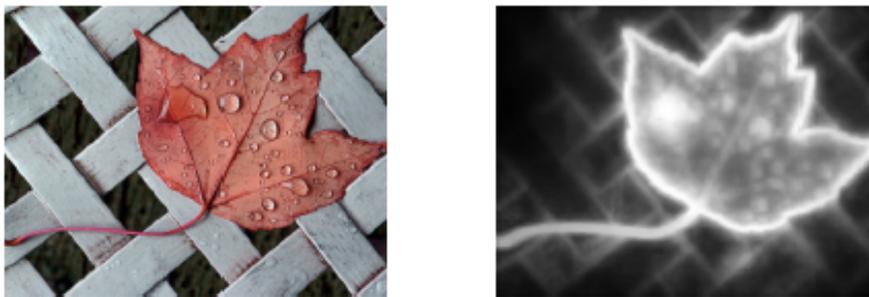


Figure 1. Saliency detection

Saliency detection is useful in many high level image-processing tasks including image segmentation and object recognition. It can be used as the pre-processing step to reduce the search space. One important application of saliency detection is image retargeting, in which we would like to keep the salient regions the same but remove pixels, which are not salient. Figure 2 shows an example of image retargeting. If we know the people is the saliency in the image, when we change the aspect ration, we keep the salient patches the same, but remove non-salient patches.

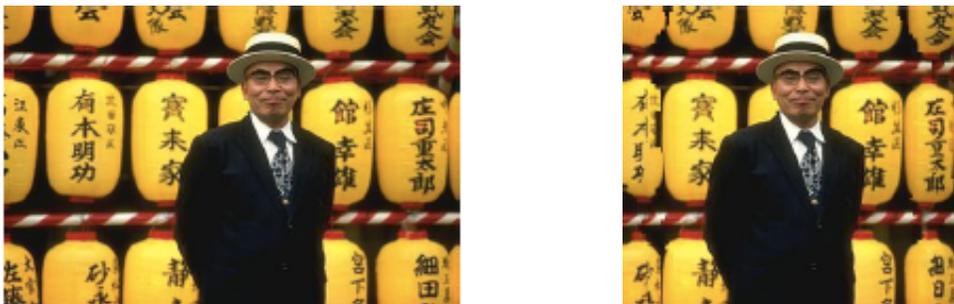


Figure 2. Image retargeting

Hou[1] proposed a simple algorithm to detect saliency in the spectral domain. Recently Goferman[2] take into account the context of the image to include useful background information in the salient regions. Our algorithm is similar to Gorferman's work [2] but we further consider the pattern information of the patch and experiments show our algorithm can achieve better results. Section 2 shows the algorithm details and in section 3 shows experiment result and the last section is the conclusion and future work.

## 2. Approach

In this project, our goal is to explore saliency detection algorithm based on psychological evidences:

- 1) Local low-level consideration: areas that have distinct colors or pattern should obtain high saliency.
- 2) Global consideration: frequently occurring features should be suppressed.
- 3) Visual organization rule: the salient pixels should be grouped together, and not spread all over the image.

### 2.1 Salient patches

A pixel is salient if its appearance is unique. But we should not look at an isolated pixel, but at its surrounding patch so we first divide the image into several patches. Each patch has  $8 \times 8$  pixels. From psychological evidences, a patch is salient if its colors or pattern is unique. We use  $L^*a^*b^*$  color space to measure the color difference between two patches because the distance in  $L^*a^*b^*$  color space reflects the visual difference. To measure the difference of patterns, we take discrete cosine transform (DCT) of each patch. Because human eyes are sensitive to low frequency component but insensitive to high frequency, we compute the weighted distance in the frequency domain, where we use the quantization table in JPEG compression to give higher weights to low frequency components and lower weights to high frequency components. We define the color distance as follows:

$$d_{color}(p_i, p_j) = \alpha \sum_{u=0}^7 \sum_{v=0}^7 |p_i(u, v) - p_j(u, v)|$$

where  $\alpha$  is the factor that normalizes the color distance between any two patches in the range of  $[0, 1]$ . We define the frequency distance as follows:

$$d_{freq}(p_i, p_j) = \sum_{u=0}^7 \sum_{v=0}^7 \frac{1}{q(u, v)} |P_i(u, v) - P_j(u, v)|$$

where  $q(u, v)$  is the entry in the following quantization table.  $P_i(u, v)$  and  $P_j(u, v)$  represent the DCT of patch  $p_i$  and  $p_j$ .

16	11	12	14	12	10	16	14
13	14	18	17	16	19	24	40
26	24	22	22	24	49	35	37
29	40	58	51	61	60	57	51
56	55	64	72	92	78	64	68
87	69	55	56	80	109	81	87
95	98	103	104	103	62	77	113
121	112	100	120	92	101	103	99

Table 1. Quantization table

We also consider the distance between two patches since non-salient patches are likely to have many similar patches both near and far away in the image. This is in contrast to salient patches, which tend to be grouped together. This implies that a patch  $p_i$  is salient when the patches similar to it are nearby, and it is less salient when the resembling patches are not close. We define  $d_{position}(p_i, p_j)$  as the Euclidean distance between the centers of patches  $p_i$  and  $p_j$ , normalized by the larger image dimension.

## 2.2 Saliency score

After computing the color distance, frequency distance and positional distance, we can define the saliency score for each patch. To combine these distances, we use the following formula to define the distance between patch  $p_i$  and  $p_j$

$$d(p_i, p_j) = \frac{c_1 d_{color}(p_i, p_j) d_{freq}(p_i, p_j)}{1 + c_2 d_{position}(p_i, p_j)}$$

For each patch, we compute these three distances to all other patches and select  $K$  most similar patches and its saliency score is defined as:

$$S_i = 1 - \exp\left(-\frac{1}{K} \sum_{k=1}^K d(p_i, p_j)\right)$$

In the experiment, we choose  $c_1=0.25$ ,  $c_2=3$ ,  $K=64$ .

## 2.3 Saliency map

The saliency map is a gray level image, where light regions correspond to salient areas and dark. We initialize the saliency map to be a black background and for each patch, we set it to be its saliency score and normalize the whole image to the range [0, 255]. After that, we use a  $7 \times 7$  Gaussian kernel to smooth the saliency map.

### 3. Experimental results

In the following experiments, we compare our method with the one does not consider the frequency distance. The left image is the original image, the middle image is the saliency map only using color distance and positional distance and the right image is the result of our algorithm.

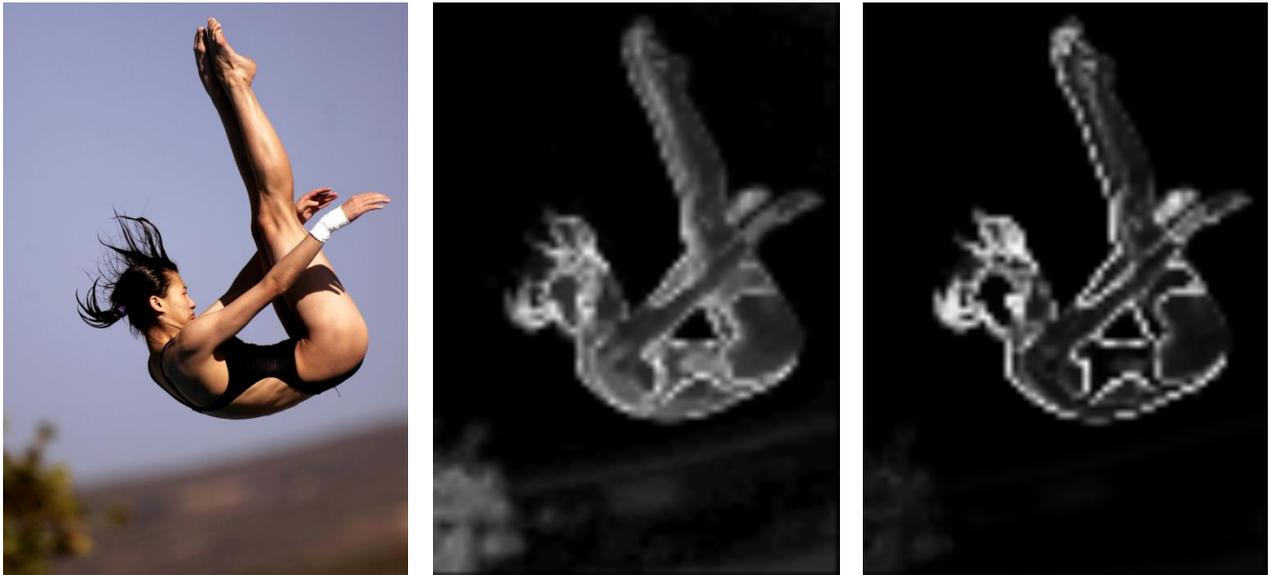


Figure 3. The resolution of the image is  $707 \times 500$ . Our algorithm takes 115 seconds to compute the saliency map on a 2.33GHz PC.

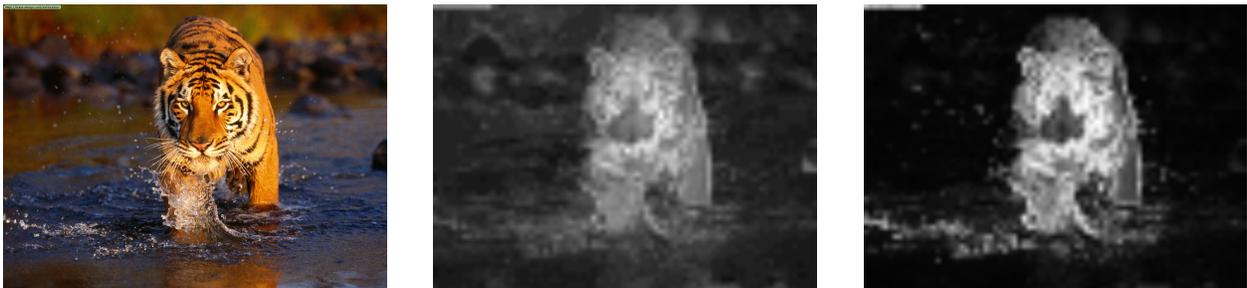


Figure 4. The resolution of the image is  $640 \times 480$ . Our algorithm takes 82 seconds to compute the saliency map on a 2.33GHz PC.



Figure 5. The resolution of the image is  $507 \times 335$ . Our algorithm takes 25 seconds to compute the saliency map on a 2.33GHz PC.

From the experiments, we can see that our algorithm suppresses non-salient areas and thus has better results compared with only considering color and positional distances.

#### **4. Conclusion and future work**

This paper proposes an algorithm to detect saliency in the image based on psychological evidences. Experiments show that this algorithm has satisfactory results. The current algorithm is not scale invariant, we would like to build the image pyramid and apply the saliency detection algorithm to images in different levels and combine the results.

#### **Reference:**

- [1] Xiaodi Hou and Liqing Zhang, "Saliency Detection: A Spectral Residual Approach", CVPR, 2007
- [2] Stas Goferman, Lihi Zelnik-Manor, "Context-Aware Saliency Detection", CVPR, 2010