Contrast Based Filtering for Salient Region Detection

Rohan Karnawat Ashwin Sudhir Vatsal Shah

Abstract

The algorithm we implemented consists of four basic steps. The method first decomposes a given image into compact, perceptually homogeneous elements that abstract unnecessary detail. Based on this abstraction, two measures of contrast are computed that rate the uniqueness and the spatial distribution of these elements. From the element contrast we then derive a saliency measure that produces a pixel-accurate saliency map which uniformly covers the objects of interest and consistently separates fore- and background.

1. Introduction

The computational identification of image elements that are likely to catch the attention of a human observer is a complex cross-disciplinary problem. Realistic, high-level models need to be founded on a combination of insights from neurosciences, biology, computer vision, and other fields. However, recent research has shown that computational models simulating low-level stimuli-driven attention are quite successful and represent useful tools in many application scenarios, including image segmentation , resizing and object detection .Results from perceptual research indicate that the most influential factor in low-level visual saliency is contrast

However, the definition of contrast in previous works is based on various different types of image features, including color variation of individual pixels, edges and gradients, spatial frequencies, structure and distribution of image patches, histograms, multi-scale descriptors, or combinations thereof. The significance of each individual feature often remains unclear, and as recent evaluations show even quite similar approaches may exhibit considerably varying performance.

2. Overview

1. **Abstraction.** The aim is to decompose the image into basic elements that preserve relevant structure, but abstract undesirable detail. Specifically, each element should locally abstract the image by clustering pixels with similar properties (like color) into perceptually homogeneous regions. Discontinuities between such regions, i.e., strong contours and edges in the image, should be preserved as boundaries between individual elements. Finally, constraints on shape and size should allow for compact, well localized elements. One approach to achieve this type of decomposition is an edge-preserving, localized over segmentation based on color. Thanks to this abstraction, contrast between whole image regions can be evaluated using just those elements. Furthermore, it is shown that the quality of saliency maps is extremely robust to the number of elements. The two measures for contrast can then be defined.

2. **Element uniqueness.** This first contrast measure implements the commonly employed assumption that image regions, which stand out from other regions in certain aspects, catch our attention and hence should be labeled more salient. We therefore evaluate how different each respective element is from all other elements constituting an image, essentially measuring the rarity of each element. In one form or another, this assumption has been the basis for most previous algorithms for contrast-based saliency. However, thanks to the abstraction, variation on the pixel level due to small scale textures or noise is rendered irrelevant, while discontinuities such as strong edges stay sharply localized.

3. **Element distribution.** While saliency implies uniqueness, the opposite might not always be true. Ideally colors belonging to the background will be distributed over the entire image exhibiting a high spatial variance, whereas foreground objects are generally more compact. The compactness and locality of image abstracting elements allows one to define a corresponding second measure, which renders unique elements more salient when they are grouped in a particular image region rather than evenly distributed over the whole image.

4. **Saliency assignment.** The two above contrast measures are defined on a per-element level. In a final step, the actual saliency values are assigned to the input image to get a pixel-accurate saliency map. Thanks to this step, this method can assign proper saliency val-

ues even to fine pixel-level detail that was excluded, on purpose, during the abstraction phase, but for which we still want a saliency estimate that conforms to the global saliency analysis.

3. Algorithm

3.1. Abstraction

For the image abstraction, an adaptation of SLIC superpixels is used to abstract the image into perceptually uniform regions. For image abstraction, we use K-means clustering in geodesic image distance in CIELab space. Geodesic image distance guarantees connectivity, while retaining locality, compactness and edge awareness.

3.2. Element uniqueness

Element uniqueness is generally defined as the rarity of a segment i given its position p_i and color in CIELab c_i compared to all other segments j:

$$U_i = \sum_{j=1}^{N} \|c_i - c_j\|^2 . \underbrace{w(p_i, p_j)}_{w_{ij}^{(p)}}$$

By introducing $w_{ij}^{(p)}$ we effectively combine global and local contrast estimation with control over the influence radius of the uniqueness operator. A local function $w_{ij}^{(p)}$ yields a local contrast term, which tends to overemphasize object boundaries in the saliency estimation , whereas $w_{ij}^{(p)} \approx 1$ yields a global uniqueness operator, which cannot represent sensitivity to local contrast variation.

3.3. Element distribution

Conceptually, we define the element distribution measure for a segment i using the spatial variance D_i of its color c_i, i.e., we measure its occurrence elsewhere in the image. As motivated before, low variance indicates a spatially compact object which should be considered more salient than spatially widely distributed elements. Hence we compute

$$D_i = \sum_{j=1}^{N} \|p_i - \mu_j\|^2 . \underbrace{w(c_i, c_j)}_{w_{ij}^{(c)}}$$

where $w_{ij}^{(c)}$ describes the similarity of color c_i and color c_j of segments i and j, respectively, p_j is again the position of segment j, and $\mu_i = \sum_{j=1}^{N} w_{ij}^{(c)} p_j$ defines the weighted mean position of color c_i.

3.4. Saliency assignment

We start by normalizing both uniqueness U_i and distribution D_i to the range[0..1]. We assume that both measures are independent, and hence we combine these terms as follows to compute a saliency value S_i for each element:

$$S_i = U_i.exp(-k.D_i)$$

zIn practice we found the distribution measure D_i to be of higher significance and discriminative power. Therefore, we use an exponential function in order to emphasize D_i. In all our experiments we use k=6 as the scaling factor for the exponential.

As the final step, we need to assign a final saliency value to each image pixel, which can be interpreted as an upsampling of the per-element saliency S_i However, naive upsampling by assigning S_i o every pixel contained in element i carries over all segmentation errors of the abstraction algorithm. Instead we adopt an idea proposed in the context of range image up-sampling and apply it to our frame- work. We define the saliency S_i of a pixel as a weighted linear combination of the saliency S_j of its surrounding image elements

4. Applications

4.1. Number Plate Extraction

We have used the saliency maps generated by this method to detect the number plate on cars based on the assumption that license plates involve contrast variations from the colour of the vehicle mass. After the saliency maps were generated we binarized them to get the mask for the number plate. Further optimisations can imporve the results obtained.

4.2. Image Blending

Once we have salient regions, the regions can be converted to a binary mask, which can then be used for blending using Laplacian and Gaussian pyramids of the source and target images.

Please refer to the images attached in the folder herewith.

References

[1] F. Perazzi, P. Krahenbuhl, Y. Pritch, A. Hornung. Saliency Filters : Contrast based filtering for salient region detection. CVPR, 2012

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk. SLIC Superpixels. Technical report, 2010.

[3] A. Criminisi, T. Sharp, C. Rother, and P. Prez. Geodesic image and video editing. ACM Trans. Graph., 29(5):134, 2010.