A Weighted Sparse Coding Framework for Saliency Detection

Nianyi Li, Bilin Sun, Jingyi Yu
Department of Computer and Information Sciences, University of Delaware, DE, USA.

There is an emerging interest on using high-dimensional datasets beyond 2D images in saliency detection. Examples include 3D data based on stereo matching and Kinect sensors and more recently 4D light field data. However, these techniques adopt very different solution frameworks, in both type of features and procedures on using them.

In this paper, we present a universal saliency detection framework for handling heterogenous types of input data. We set out to build saliency/non-saliency dictionaries using data-specific features. Specifically, we first select a group of potential foreground superpixels to build the saliency dictionary. We then prune the outliers and test on the remaining super-pixels to iteratively refine the dictionaries. A major advantage of our technique is that it provides a universal framework for all different types. The only variation to the algorithm is input features: for 2D images, we use color, texture and focusness characteristics; for stereo data, we add depth/disparity cues; and for the 4D light field data, we add focusness cues on focus stack. Comprehensive experiments on a broad range of datasets (MSRA-1000 and SOD for 2D, SSB [4] for 3D, and the light field saliency dataset[2] for 4D) show that our technique outperforms state-of-the-art solutions.

Our approach is based on building saliency/non-saliency dictionaries, which are built for superpixels in reference image I, *i.e.*, image used for generate disparity map in stereo pair data and the all-focus image in light filed data.

For each pixel in the reference image, we set out to associate with a feature vector. As mentioned above, we utilize different feature descriptors for 2D, 3D and 4D imagery data. From the feature vectors of all pixels, we generate two feature matrices for all super-pixel. The first scheme is through averaging per-pixel feature vectors within the superpixel. Our second scheme computes the histogram over three color channels. We use F^A and F^H to representing the resulting feature matrix respectively.

From F^A and F^H, we develop a sparse coding framework: saliency superpixels correspond to the ones that yield to low reconstruction error from the saliency dictionary. We use the error measure to refine the foreground superpixels and to identify foreground saliency ones.

For saliency detection, we adopt the weighted sparse coding scheme[1]:

$$\boldsymbol{\alpha_i} = \arg\min_{\boldsymbol{\alpha_i}} \|\boldsymbol{f_i} - \boldsymbol{D\alpha_i}\|_2^2 + \lambda \|\boldsymbol{diag(\omega_i)} \cdot \boldsymbol{\alpha_i}\|_1 \quad (1)$$

where the jth value of ω_i is the penalty for using the jth member in template D to encode f_i. The goal is to find a sparse code α_i that can achieve the maximum/minimum reconstruction error. Notice that large ω_i will suppress nonzero entries α_i and force the solution α to concentrate on indices where ω_i is small. Therefore,if the f_i is similar to some template in D, the penalty ω_i should be small and vice versa.

We use $\omega_{r_i}^D$ to represent the weight/penalty for superpixel r_i, where $D \in \{A,H\}$. $\omega_{r_i}^D = [g(r_i,D_1), g(r_i,D_2) \dots g(r_i,D_K)]^T$ is a vector that computes the similarity between superpixel r_i (in feature matrix F^D) to all the members in template D and

$$g(r_i,D_j) = e^{\|F_{r_i}^D - D_j\|} \quad (2)$$

Next, we use (A, $\omega_{r_i}^A$) and (H, $\omega_{r_i}^H$) as input to Eqn. 1 to generate to sparsely coded dictionary $\alpha_{r_i}^A$ and $\alpha_{r_i}^H$ respectively. We then compute the reconstruction error $\varepsilon_{r_i}^A$ and $\varepsilon_{r_i}^H$ for each r_i:

$$\varepsilon_{r_i}^D = \|F_{r_i}^D - D\alpha_{r_i}^D\|_2^2 \quad (3)$$

The saliency function $Sal^D(r_i)$ relates to the dictionary's type (saliency or non-saliency). For non-saliency dictionary, it will assign high values to superpixels of a high $\varepsilon_{r_i}^D$ value. Similarly, for saliency dictionary, $Sal^D(r_i)$ will assign high value to superpixels with low $\varepsilon_{r_i}^D$.

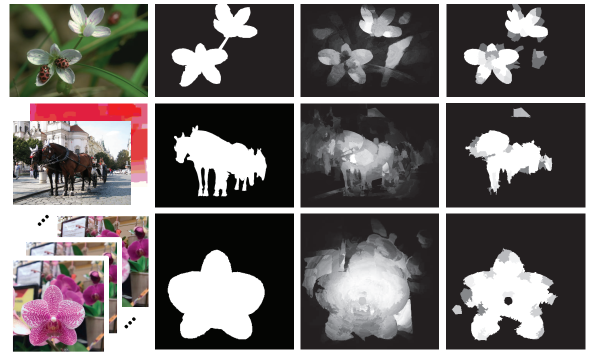

Figure 1: Our method vs. the latest feature-matrix-based DSR algorithm [3] on different data inputs. From top to bottom: we show results on 2D images, 3D stereo data, and 4D light field data.

Finally, we combine $Sal^A(r_i)$ and $Sal^H(r_i)$ to get the saliency value for r_i:

$$Sal(r_i) = Sal^A(r_i) + Sal^H(r_i) \quad (4)$$

We define saliency dictionary as a set of superpixels $S = \{r_{s_1}, r_{s_2}, \dots r_{s_k}\}$ which are regarded as the potential saliency regions and will be refined through our framework. To get the initial saliency dictionary, we compute a non-saliency dictionary by [5] to reconstruct the reference image, and patches with high reconstruction error are selected saliency dictionary S^0. We start with using S^0 as input to the weighted sparse framework. At each iteration, we will refine the saliency dictionary using the estimated saliency map. The algorithm terminates when there is no change to the saliency dictionary.

In conclusion, we have presented a novel saliency detection algorithm that is applicable to 2D image data, 3D stero/depth data, and 4D light field data without modifying the processing pipeline. We show that two types of feature descriptors are complimentary to each other for handling variational types of texture/color scene compositions. Comprehensive experiments have shown that it outperforms previous tailored solutions for different data types.

[1] Jian Huang, Shuangge Ma, and Cun hui Zhang. Adaptive lasso for sparse highdimensional regression. Technical report, University of Iowa, 2006.

[2] Nianyi Li, Jinwei Ye, Yu Ji, Haibin Ling, and Jingyi Yu. Saliency detection on light field. In *CVPR*, June 2014.

[3] Xiaohui Li, Huchuan Lu, Lihe Zhang, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via dense and sparse reconstruction. In *ICCV*, 2013.

[4] Yuzhen Niu, Yujie Geng, Xueqing Li, and Feng Liu. Leveraging stereopsis for saliency analysis. In *CVPR*, 2012.

[5] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun. Saliency optimization from robust background detection. In *CVPR*. IEEE, 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.