Comparing Color Descriptors between Image Segments for Saliency Detection

Anurag Singh[1], Henry Chu[1,2] and Michael Pratt[3]
[1]*The Center for Advanced Computer Studies, University of Louisiana at Lafayette, Lafayette, LA, U.S.A.*
[2]*Informatics Research Institute, University of Louisiana at Lafayette, Lafayette, LA, U.S.A.*
[3]*W. H. Hall Department of Electrical & Computer Engineering, University of Louisiana at Lafayette, Lafayette, LA, U.S.A.*

Keywords: Visual Saliency, Superpixels, Earth Mover's Distance, Dominant Color Descriptors.

Abstract: Detecting salient regions in an image or video frame is an important step in early vision and image understanding. We present a visual saliency detection method by measuring the difference in color content of an image segment with that of its neighbors. We represent each segment with richer color descriptors in the form of a regional dominant color descriptor. The color difference between a pair of neighbors is found using the Earth Mover's Distance. The cost of moving color descriptors between neighboring segments robustly captures the difference between neighboring segments. We evaluate our method on standard datasets and compare it with other state-of-the-art methods to demonstrate that it has better true positive rate at a fixed false positive rate in detecting salient pixels relative to the ground truth. The proposed method uses local cues without being an edge highlighter, a common problem of local contrast-based methods.

1 INTRODUCTION

Visual saliency detection is a useful early step in many computer vision applications to focus attention to the most prominent object or region in an image or video. As such, a salient region is one that stands out, or acts as an outlier, with respect to other regions in its vicinity. The important considerations for a pixels or region to be salient are as follows. First, we consider the uniqueness in color space: color is one of the most important feature which differentiates between salient and non salient pixels; a salient group often stands out because it is different in color content compared to other groups. Secondly, visual fixation occurs in clusters (Treisman, 1982), (Ben-Av et al., 1992). The so-called superpixels are a suitable representation of clustering of pixels. Superpixels are mostly homogeneous in image content and often conform to edge definition in an image. Hence, superpixel segment (Felzenszwalb and Huttenlocher, 2004) can be chosen as the lowest level primitives. When so doing, it has the additional benefit of reducing the computational cost of comparisons drastically when compared to the cost using pixels or pixel-centered neighborhoods.

A number of methods have been proposed in the literature to detect image saliency (Liu et al., 2007),(Itti et al., 1998),(Harel et al., 2006),(Ma and Zhang, 2003),(Goferman et al., 2010),(Lin et al., 2013). The color contrast-based methods for saliency detection is done using either global comparison (Singh et al., 2014), (Cheng et al., 2011), (Perazzi et al., 2012) or local comparison (Goferman et al., 2010), (Liu et al., 2007). Since the eye fixation process of biological vision responses to local cues, the local comparison methods are de facto edge highlighters in that they highlight the region around the edge boundary (Fig 1).

In this paper we present a method which uses local contrast to highlight a whole salient segment. The saliency detection method is based on the difference of the salient region within its neighborhood, or the outlierness of the region. We summarize the color content of each segment by its dominant color descriptors (DCD). An advantage of using the DCDs is that DCDs of all image segments or regions are of the same size, irrespective of the number of pixels in the region. This DCD property facilitates the use of the Earth Mover's Distance to measure the difference of the region's DCD and those of its neighbors to determine the outlierness of a region.

The novelty of our method lies in exploiting the segment signatures of DCDs and the choice of metric, viz. the Earth Mover's Distance (EMD) to compute the color contrast. EMD was considered to be the

Singh, A., Chu, H. and Pratt, M.
Comparing Color Descriptors between Image Segments for Saliency Detection.
DOI: 10.5220/0005667705580565
In *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2016)*, pages 558-565
ISBN: 978-989-758-173-1

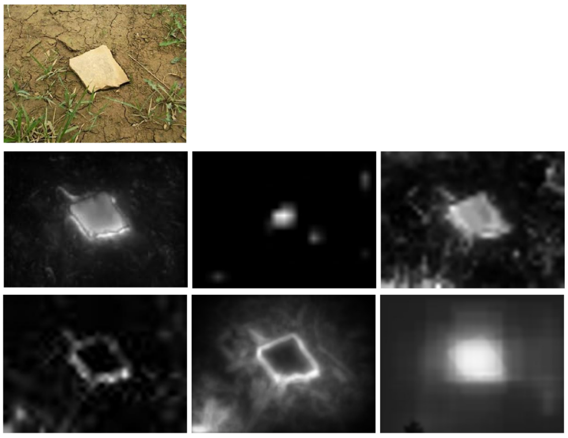


Figure 1: Examples of image saliency detection. *Top row:* Input image. *Middle row, left to right:* Methods SO (Liu et al., 2007), IT (Itti et al., 1998), GB (Harel et al., 2006); *Bottom row, left to right:* Methods MZ (Ma and Zhang, 2003), CA (Goferman et al., 2010), EMD proposed in the present paper. The saliency map's highlighted area varies depending on the method used.

best metric when comparing equal sized signatures, such as in the case of DCDs (Rubner et al., 2000a). EMD also enables us to make a weighted comparison between color feature vectors, so that color is used proportional to its importance. The contrast difference between segments is estimated by the amount of work needed to move the color from one segment to another. The more the amount of work is required for moving the contents, the more is the perpetual difference between segments.

When compared to the other color-contrast methods, the proposed algorithm uses only the neighboring segments to infer saliency and uses richer descriptors than, say, color averages. By highlighting whole salient objects, it overcomes problems associated with the local methods that act as edge highlighters (Perazzi et al., 2012).

2 SALIENCY DETECTION

We formulate the visual saliency as a measure of the amount of color difference between neighboring image segments. This formulation results in a graphical structure where each image segment is seen as the receiver of color from connected segments. In this section we discuss in detail the color description of each segment, a metric for comparison and a multi-resolution step for finding best segment. Fig. 2 shows the flow diagram for the saliency detection algorithm.

2.1 Superpixel Segmentation and Summarization

An input image is first divided into superpixel segments. Superpixels are usually homogeneous and conform to image edges, making them good primitives for comparison. Superpixels are of various sizes and to make true comparisons we need to use a normalized representation. Each superpixel's color information is summarized using the DCD.

The dominant color descriptor was introduced in the MPEG-7 standard (Manjunath et al., 2001) to provide a compact description of an image for applications such as content-based image retrieval from a collection of images of different sizes. We compute a DCD not for the entire image but for each superpixel segment. A descriptor for each superpixel therefore consists of a set of representative colors and their corresponding percentages in the superpixel region. DCDs provide a perceptually closer representation of the input image (Singh et al., 2014), hence they form a better descriptor model for comparison. The DCD for the kth image segment is given by $DCD_k = \{(c_{ki}, p_{ki}) : i = 1, \cdots, N_k\}$, where c_{ki} is a color in the CIE $L^*a^*b^*$ space, p_{ki} is a percentage of pixels in the superpixel represented by the corresponding color. The CIE $L^*a^*b^*$ color space is chosen, as it supports double opponency and is perceptually similar to the color scheme in the human visual cortex (Engel et al., 1997). There are different ways proposed in the literature to compare two DCDs. One such comparison method (Yang et al., 2008) is given as

$$D^2(DCD_1, DCD_2) = 1 - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \{(1 - (p_{1i} - p_{2j})) \times \min(p_{1i}, p_{2j}) a_{1i,2j}\} \tag{1}$$

where $a_{1i,2j}$ is the similarity between colors c_{1i} and c_{2j}, given by

$$a_{1i,2j} = \begin{cases} 1 - d_{1i,2j}/T_d, & d_{1i,2j} \leq T_d \\ 0, & otherwise \end{cases} \tag{2}$$

where $d_{1i,2j}$ is the Euclidean distance between colors c_{1i} and c_{2j}, and T_d is a threshold set between 0 and 1.

2.2 Saliency Computation

In this section we discuss in detail the steps followed to compute the saliency of a superpixel segment. As mentioned before, the color signature of each segment is represented by the clustered dominant colors.

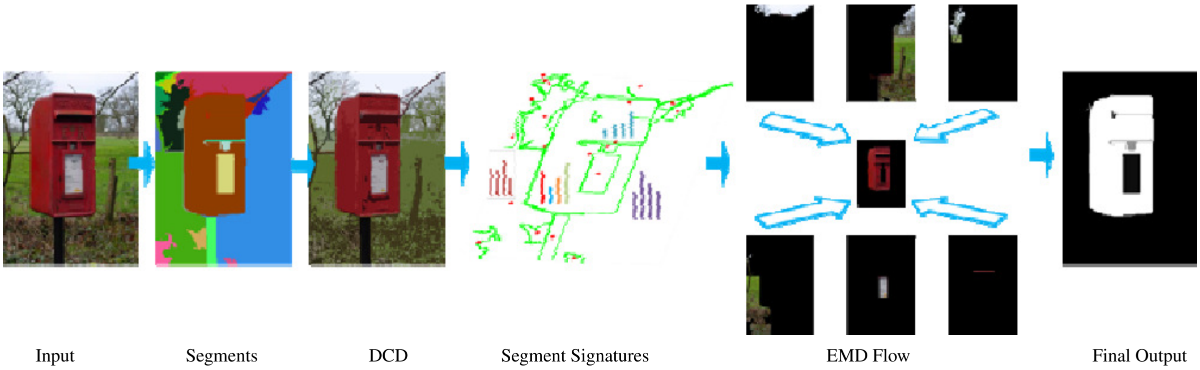


Figure 2: Flow diagram for the proposed local comparison-based saliency detection algorithm.

Algorithm 1: DCD computation.

Input: Pixels $\in$ Superpixel
Output: Dominant Color Desciptors
$RGB \rightarrow CIEL*a*b*$ transform;
$N = superpixel.size()$;
while $i \leq N$ **do**
  Read all pixels in a superpixel[i] ;
  Create 4 DCD cluster C_1, C_2, C_3, C_4;
  Cluster colors using K-means;
  for $j \leftarrow 1; j < 4$ **do**
    for $k \leftarrow j+1; j < 4$ **do**
      if $\|C_j - C_k\|_2 < \delta$ **then**
        Merge Clusters ;
      else
        Continue;
  Compute % of pixels for each DCD;

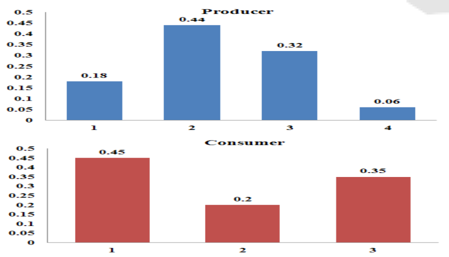


Figure 3: Image signature computed using dominant color descriptors.

The difference in DCD signatures is calculated using the Earth Mover's Distance (EMD). The EMD measures the minimum amount of work required to move the color contents of one superpixel segment to those of another superpixel segment. This metric gives us the amount of perceptual color difference between the superpixel segments.

The EMD finds its roots in a known solution of the transportation problem or Monge-Kantorovich problem. Suppose there are several suppliers and several consumers; the solution of the transportation problem is to find the least expensive flow of goods from the suppliers to the consumers.

For comparing visual information, the EMD (Rubner et al., 2000a) was introduced as a metric to compare two distributions in a transformed feature space, where a measure is computed as the amount of ground distance moved. EMD has been used as a metric for comparing the dissimilarity between two distributions of color and texture in image retrieval (Rubner et al., 2000b) and saliency detection (Lin et al., 2013).

The EMD is formalized as a linear programming problem (Rubner et al., 2000a). In our context, let P and Q be two segments represented by the dominant color descriptors as two signatures $DCD_P = ((c_{P1}, w_{P1}) \cdots (c_{Pm}, w_{Pm}))$ and $DCD_Q = ((c_{Q1}, w_{Q1}) \cdots (c_{Qn}, w_{Qn}))$, where c_{Pi} and c_{Qj} are color cluster representatives of P and Q, respectively, and w_{Pi} and w_{Qj} are the corresponding percentages of pixels as weight of the clusters. Let $D = [d_{ij}]$ be the cost matrix or the ground distance matrix for moving color between DCD_P and DCD_Q so that d_{ij} is the pairwise distance in the CIE $L*a*b*$ space between two colors, one from each DCD:

$$d_{ij} = ||c_{Pi} - c_{Qj}||. \tag{3}$$

The flow $F = [f_{ij}]$ between DCD_P and DCD_Q is minimized by

$$WORK(P,Q,F) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}. \tag{4}$$

The EMD is defined as the work done from Equation 4, normalized by the total flow and given as

$$EMD(P,Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \tag{5}$$

where total flow $\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}$ is given as $min(\sum_{i=1}^{m} w_{Pi}, \sum_{j=1}^{n} w_{Qj})$.

The EMD solution is subjected to the following constraints:

1. Supplies are moved only from P to Q one way; i.e.,
   $f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n$
2. Limits the amount of supplies that can be sent from clusters to the weights; i.e.,
   $\sum_{j=1}^{n} f_{ij} \leq w_{Pi}, 1 \leq i \leq m$
3. Limits the clusters in their weights; i.e.,
   $\sum_{i=1}^{m} f_{ij} \leq w_{Qi}, 1 \leq j \leq n$
4. Force to move maximum amount of supply; i.e.,
   $\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = min(\sum_{i=1}^{m} w_{Pi}, \sum_{j=1}^{n} w_{Qj})$

The major advantage of using EMD for superpixel comparison is that it captures the signature of superpixels better than an average color or position value. Another advantage of using EMD is that it matches similarities (or finds dissimilarities) in a way that is consistent with human perception. EMD is a true metric of comparing the distributions of similar masses, which is the case when the superpixels are approximately similarly sized. Our use of the EMD is similar to comparing histograms, where the amount of work done to move the bin contents from one histogram to another is a measure of the dissimilarity between the two histograms.

We note that EMD has been used to compare histograms in saliency detection (Lin et al., 2013). In that work, EMD was used to compute the center-surround difference in the framework of (Itti et al., 1998). The center-surround difference is computed directly from the color values of two image regions, so that the histograms have a large number of bins. Two histograms **a** and **b** have the same number of bins and the corresponding bins a_i and b_i represent the same range of colors, for all i. In our work the color information in each segment is summarized by a DCD. Each of the DCDs being compared may have up to 4 components and each component can represent a different color. In terms of EMD computation, the cost matrix D in (Lin et al., 2013) is large, computed once, and fixed for all pairs of histograms in an image whereas the matrix D in our work is much smaller (at most 4 by 4) and is computed for each pair of DCDs being compared.

Aggregating the Difference: The saliency of the Pth segment is computed by aggregating all the work done associated with moving colors from neighboring segments. The aggregated average saliency for segment P is given by

$$\hat{Sal}_P = \sum_{k=1}^{K} EMD(P, Q_k) \tag{6}$$

where K is the total number of neighbors, and the EMD between segment P and a neighboring segment Q_k is found using Equation 5.

Biasing for Center: There is an inherent central bias in visual saliency maps (Borji and Itti, 2013). The central bias is updated by penalizing the segments that are far from the center in the following way:

$$Sal_P = \hat{Sal}_P(1 - \delta) \tag{7}$$

where $\hat{Sal}_P$ is the saliency value for Segment P and δ is the normalized distance of the segment center from the center of the image.

2.3 Normalization Step

The human visual system suppresses the low level responses and focuses attention on stimuli with higher level responses. In this algorithm, the focus is desired to be on salient objects. A generic objectness measure is used, which is the probability of occurrence of an object in a window (Alexe et al., 2012). Sampling for object windows gives the notion of objectness (Sun and Ling, 2013), which ensures a higher probability value for the occurrence of an object. The normalized saliency map is given by

$$Sal_{norm} = \frac{1}{2}(Sal_{map} + Obj_{map}) \tag{8}$$

where Sal_{map} is the final saliency map formed by compositing the saliency values of the superpixel segments computed from Equation 7 and Obj_{map} is the objectness map.

2.3.1 Choosing the Best Segmented Image

The major limiting factor is the quality of the superpixel segmentation algorithm. If the segmentation algorithm divides the salient region into very small segments, the overall saliency algorithm enhances small segments found between salient and non-salient regions. In Fig. 4, an example of this is shown. To

Figure 4: Limitations due to non cohesive segmentation. Non-cohesive(*top row*) and cohesive (*bottom row*) segmentation.

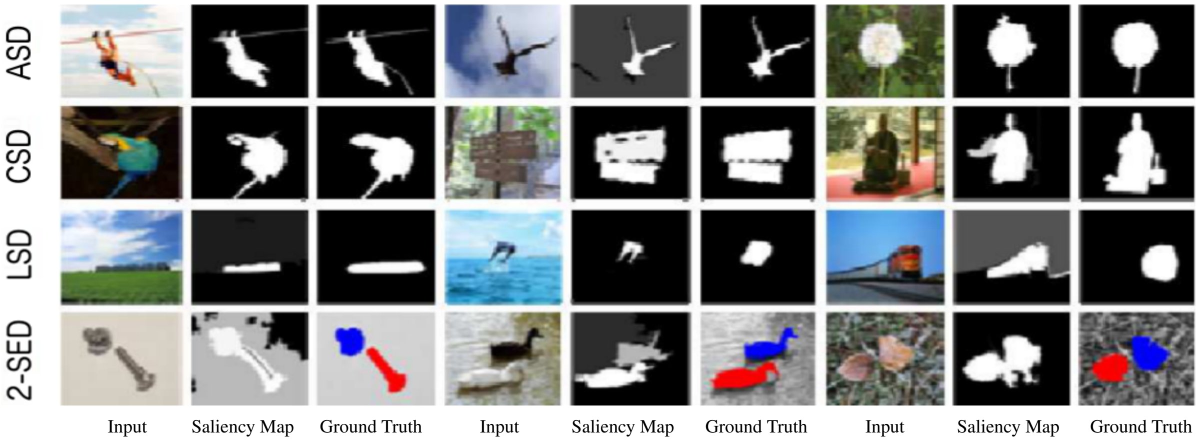


Figure 5: Saliency map results for four publicly available, standard datasets. Each row corresponds to a different dataset. On each row, we show three representative triplets of an input image, the saliency map, and the ground truth image.

overcome this, a measure for finding best segment (Bagon et al., 2008) is integrated into the algorithm. The goodness measure is based on the following conditions: i) every pixel belongs to a region, ii) every region is spatially connected, iii) regions are disjoint; and iv) all pixels in a region satisfy a specified similarity.

Initially, the input image is segmented into multiple images, each with superpixels at a particular resolution. The resolution that gives the the best goodness measure is chosen for further processing. From this resolution, the superpixels image with best cohesive segmentation is generated. The more cohesive the segments will result in better saliency detection.

Algorithm 2: EMD-based saliency detection.

> **Input**: Input Image
> **Output**: Saliency Map
> $RGB \rightarrow CIEL{*}a{*}b{*}$;
> $numRes \rightarrow 20$;
> **for** $j \leftarrow 0; j \leq numRes$ **do**
> $Res[j] \rightarrow goodSegmentScr()$;
> $BestSeg = \max(Res[j] \rightarrow goodSegmentScr())$;
> $N = BestSeg.size()$;
> **while** $i \leq N$ **do**
>   Compute DCD_i ;
>   Compute sp_i avgX,avgY;
>   Compute Image signature ;
>   Compute $neighList_i$;
> **while** $i \leq N$ **do**
> $sumSal = 0$;
>   **while** $k \leq N$ **do**
>   Compute cost matrix ;
>   Compute $emd_{ik} =$;
> $sumSal = sumSal + emdik$;
>   Aggregate total difference ;
> $\delta = ||sp_i \rightarrow avgX, avgY - img \rightarrow avgX, avgY||_2$;
>   Account for center bias $Sal_i(1 - \delta)$;

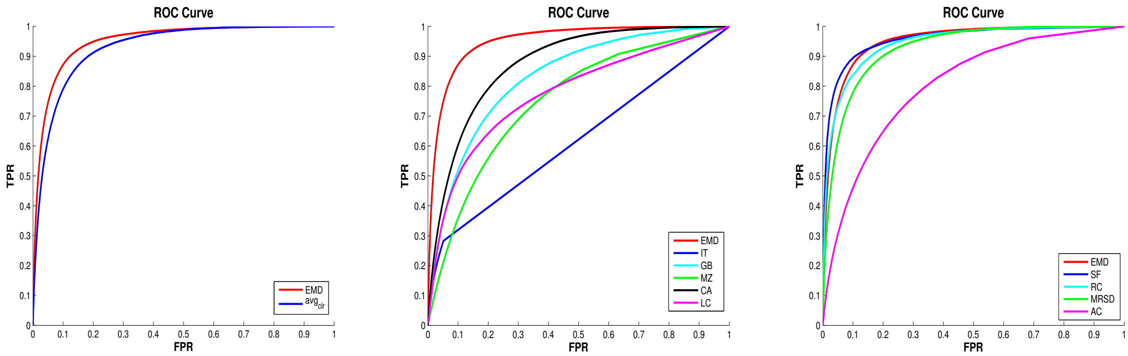
3 EXPERIMENTS

3.1 Implementations

The local comparisons based saliency detection algorithm was implemented using C++ and the OpenCV library, which has an implementation of EMD. For superpixel segmentation, the publicly available implementation given by Felzenszwalb et al(Felzenszwalb and Huttenlocher, 2004) was used. On average, the saliency computation took about 0.4 seconds on a single core, 3 GHz Intel i5 processor with 8 GB RAM.

Data Sets: For the experiments, the standard data sets ASD (Achanta et al., 2009), LSD (Li et al., 2013), 2-SED (Alpert et al., 2007) and complex-scene (CSD) (Yan et al., 2013) were used. ASD and CSD are two of the largest data sets publicly available with 1,000 images each. Each image in ASD has a single stand out object while each image in the CSD set has multiple stand out objects; each image in 2-SED has two salient objects and LSD has images with different categories.

Results: The results of testing the proposed method are shown in Fig. 5, in which each row shows the results of a dataset, organized as triplets of an input image, the result saliency map, and the ground truth. The saliency map should be visually closer to the ground truth image.



(a) (b) (c)

Figure 6: ROC curves of comparing the proposed method (“EMD”) with (a) the baseline average color method, (b) other local contrast methods, and (c) other global contrast methods.

3.2 Evaluation

We assess the performance of different methods as follows. For each input image, we obtain the saliency map from a given method and compare it to the ground truth. The ground truth map is either on (“1”) or off (“0”). Over a threshold value, when a ground truth “1”-pixel is labeled as “salient,” that pixel is considered a true positive. On the other hand, when a ground truth “0”-pixel is labeled as “salient,” it is considered as false positive. We average the performance over all images in a dataset. We plot the true positive rate against the false positive rate to obtain the ROC curve, which is used for quantitative evaluations. The area under the curve shows how well the saliency algorithm predicts against the ground truth, which is the human fixation data or the salient segment. The following two sets of quantitative evaluation are done.

Baseline Comparison: It shows that EMD is a better metric by comparing it to commonly used average color comparison. For consistency, the average color map is normalized with the objectness map. From Fig. 6a, it can be seen that the EMD-based metric works better.

Comparison to State-of-the-Art Methods: A comparison with local contrast-based saliency detection methods CA (Goferman et al., 2010), IT (Itti et al., 1998), GB (Harel et al., 2006), MZ (Ma and Zhang, 2003), LC (Zhai and Shah, 2006), and global contrast-based saliency detection methods AC (Achanta et al., 2009), MRSD (Singh et al., 2014), SF (Perazzi et al., 2012), RC (Cheng et al., 2011) on the data set ASD (Achanta et al., 2009) is done. It is worth noting that each of these methods is based on contrasting the color information, either locally or globally, to detect the salient regions. For this data set, each author has publicly available results. The ROC curve implementation given by (Borji and Itti, 2013) is used to compare the proposed method (“EMD”) to local methods and global methods, shown in Fig. 6b and Fig. 6c, respectively.

Visual Comparison: A visual comparison of the results is also presented to illustrate the performance of the proposed method (“EMD”) compared to other local contrast methods. (Fig. 7). The visual comparison shows that our method robustly highlights salient segments for various input images.

4 CONCLUDING REMARKS

A novel saliency detection algorithm by finding the amount of work done by moving dominant color descriptors from the neighboring superpixel segments is presented. This work overcame earlier problems of edge highlighting associated with local comparison-based methods. The EMD along with DCD captures the color difference better hence resulting in a better saliency map. Experimental results show that our results are more consistent with the human vision’s attention.

Our work shows a new promising direction of using EMD metric for local comparison-based saliency detection with quantitative results better than local comparison-based state-of-the-art methods. Ongoing work includes comparing the information from the objectness measure (Alexe et al., 2012) with other methods of extracting objects recently proposed (Li et al., 2014). Other ongoing work includes extending the method to compute saliency not just from a single frame but from a larger collection of images.

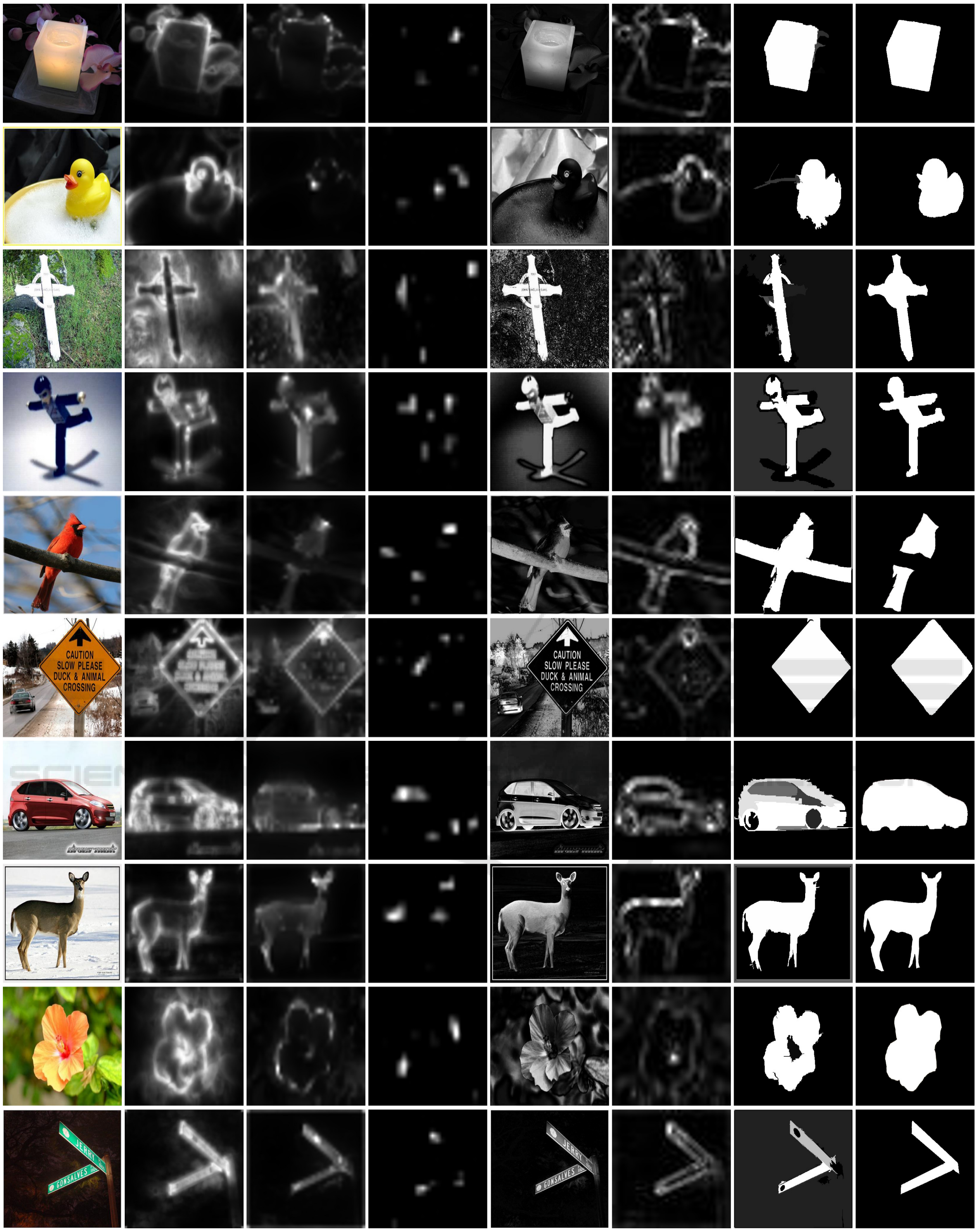


Figure 7: Visual comparison with other state-of-the-art-methods. In each row, from left to right are an input image, followed by results from methods CA (Goferman et al., 2010), GB (Harel et al., 2006), IT (Itti et al., 1998), LC (Cheng et al., 2011), MZ (Ma and Zhang, 2003), result from the proposed EMD-based saliency method ("EMD"), and the ground truth.

REFERENCES

Achanta, R., Hemami, S., Estrada, F., and Süsstrunk, S. (2009). Frequency-tuned salient region detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1597–1604.

Alexe, B., Deselaers, T., and Ferrari, V. (2012). Measuring the objectness of image windows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2189–2202.

Alpert, S., Galun, M., Basri, R., and Brandt, A. (2007). Image segmentation by probabilistic bottom-up aggregation and cue integration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Bagon, S., Boiman, O., and Irani, M. (2008). What is a good image segment? a unified approach to segment extraction. In *Proceedings of the 10th European Conference on Computer Vision: Part IV*, pages 30–44.

Ben-Av, M., Sagi, D., and Braun, J. (1992). Visual attention and perceptual grouping. *Perception and Psychophysics*, 52(3):277–294.

Borji, A. and Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207.

Cheng, M.-M., Zhang, G.-X., Mitra, N. J., Huang, X., and Hu, S.-M. (2011). Global contrast based salient region detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, pages 409–416.

Engel, S., Zhang, X., and Wandell, B. (1997). Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, 388(6637):68–71.

Felzenszwalb, P. and Huttenlocher, D. (2004). Efficient graph-based image segmentation. *International Journal Computer Vision*, 59(2):167–181.

Goferman, S., Zelnik-Manor, L., and Tal, A. (2010). Context-aware saliency detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2376–2383.

Harel, J., Koch, C., and Perona, P. (2006). Graph-based visual saliency. In *Advances in Neural Information Processing Systems*, pages 545–552. MIT Press.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.

Li, J., Levine, M., An, X., Xu, X., and He, H. (2013). Visual saliency based on scale-space analysis in the frequency domain. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(4):996–1010.

Li, Y., Hou, X., Koch, C., Rehg, J., and Yuille, A. (2014). The secrets of salient object segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4321–4328.

Lin, Y., Tang, Y., Fang, B., Shang, Z., Huang, Y., and Wang, S. (2013). A visual-attention model using earth mover's distance-based saliency measurement and nonlinear feature combination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):314–328.

Liu, T., Sun, J., Zheng, N.-N., Tang, X., and Shum, H.-Y. (2007). Learning to detect a salient object. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Ma, Y.-F. and Zhang, H.-J. (2003). Contrast-based image attention analysis by using fuzzy growing. In *ACM International Conference on Multimedia*, pages 374–381.

Manjunath, B. S., Ohm, J.-R., Vasudevan, V. V., and Yamada, A. (2001). Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715.

Perazzi, F., Krahenbuhl, P., Pritch, Y., and Hornung, A. (2012). Saliency filters: Contrast based filtering for salient region detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 733–740.

Rubner, Y., Tomasi, C., and Guibas, L. (2000a). The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121.

Rubner, Y., Tomasi, C., and Guibas, L. J. (2000b). The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121.

Singh, A., Chu, C., and Pratt, M. A. (2014). Multiresolution superpixels for visual saliency detection. In *IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing*, pages 1–8.

Sun, J. and Ling, H. (2013). Scale and object aware image thumbnailing. *International Journal of Computer Vision*, 104(2):135–153.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2):194.

Yan, Q., Xu, L., Shi, J., and Jia, J. (2013). Hierarchical saliency detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1155–1162.

Yang, N.-C., Chang, W.-H., Kuo, C.-M., and Li, T.-H. (2008). A fast mpeg-7 dominant color extraction with new similarity measure for image retrieval. *Journal of Visual Communication and Image Representation*, 19(2):92–105.

Zhai, Y. and Shah, M. (2006). Visual attention detection in video sequences using spatiotemporal cues. In *ACM International Conference on Multimedia*, pages 815–824.