

What Makes a Patch Distinct?

Ran Margolin
Technion
Haifa, Israel

margolin@tx.technion.ac.il

Ayellet Tal
Technion
Haifa, Israel

ayellet@ee.technion.ac.il

Lihi Zelnik-Manor
Technion
Haifa, Israel

lihi@ee.technion.ac.il

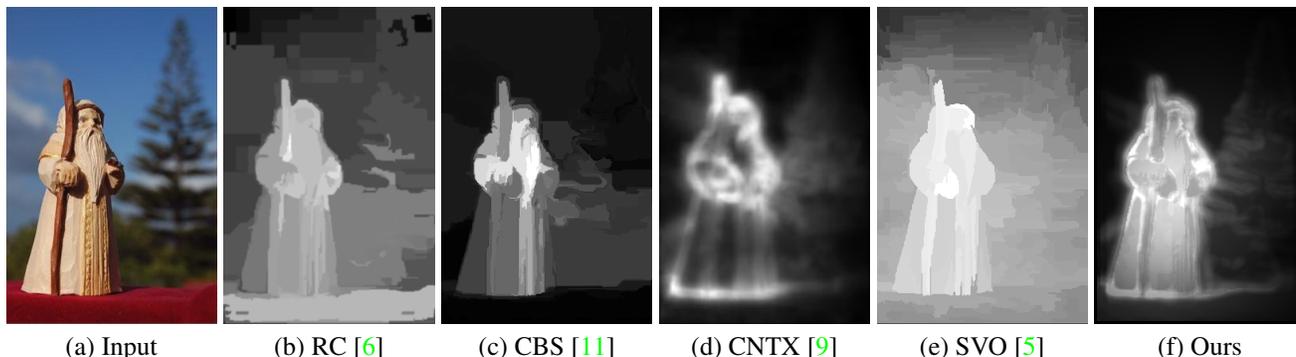


Figure 1. **Salient object detection:** This figure compares our results to those of the “Top-4” algorithms according to [4]. (b) [6] consider only color distinctness, hence, erroneously detect the red surface as salient. (c) [11] rely on shape priors and thus, detect only the beard and arm. (d) [9] search for unique patches, hence, detect mostly the outline of the statue. (e) [5] add an objectness measure to [9]. Their result is fuzzy due to the objects in the background (tree and clouds). (f) Our algorithm accurately detects the entire statue, excluding all background pixels, by considering both color and pattern distinctness.

Abstract

What makes an object salient? Most previous work assert that distinctness is the dominating factor. The difference between the various algorithms is in the way they compute distinctness. Some focus on the patterns, others on the colors, and several add high-level cues and priors. We propose a simple, yet powerful, algorithm that integrates these three factors. Our key contribution is a novel and fast approach to compute pattern distinctness. We rely on the inner statistics of the patches in the image for identifying unique patterns. We provide an extensive evaluation and show that our approach outperforms all state-of-the-art methods on the five most commonly-used datasets.

1. Introduction

The detection of the most salient region of an image has numerous applications, including object detection and recognition [13], image compression [10], video summarization [16], and photo collage [8], to name a few. Therefore, it is not surprising that much work has been done on

saliency detection. Different aspects of distinctness have been examined before. Some algorithms look for regions of distinct color [6, 11]. As shown in Figure 1(b) this is insufficient, as some regions of distinct color may be non-salient. Other algorithms [5, 9] detect distinct patterns, such as the boundaries between an object and the background. As illustrated in Figure 1(d), this could lead to missing homogeneous regions of the salient object.

In this paper, we introduce a new algorithm for salient object detection, which solves the above problems. It integrates pattern and color distinctness in a unique manner. Our key idea is that the analysis of the inner statistics of patches in the image provides acute insight on the distinctness of regions. A popular and efficient method to reveal the internal structure of the data is Principal Component Analysis (PCA). It finds the components that best explain the variance in the data. Therefore, we propose to use PCA to represent the set of patches of an image and use this representation to determine distinctness. This is in contrast to previous approaches that compared each patch to its k -nearest neighbors [9, 5], without taking into account the internal statistics of all the other image patches.

We test our method on the recently-published benchmark of Borji et al. [4]. This benchmark consists of five well-known datasets of natural images, with one or more salient objects. In [4], many algorithms are compared on these datasets and the “Top-4” algorithms, which outshine all others, are identified. We show that our algorithm outperforms all “Top-4” algorithms on all the data-sets of the benchmark. Furthermore, our method is computationally efficient.

The rest of this paper is organized as follows. We begin by describing our approach, which consists of three steps: pattern distinctness detection (Section 2.1), color distinctness detection (Section 2.2) and finally incorporating priors on human preferences and image organization (Section 2.3). We then proceed to evaluate our method both quantitatively and qualitatively in Section 3.

2. Proposed approach

The guiding principle of our approach is that a salient object consists of pixels whose local neighborhood (region or patch) is distinctive in both color and pattern. As illustrated in Figure 2, integrating pattern and color distinctness is essential for handling complex images. Pattern distinctness is determined by considering the internal statistics of the patches in the image. A pixel is deemed salient if the pattern of its surrounding patch cannot be explained well by other image patches. We further consider the color uniqueness of the pixel’s local neighborhood. Finally, we incorporate known priors on image organization. In what follows, we elaborate on each of these steps.

2.1. Pattern Distinctness

The common solution to measure pattern distinctness is based on comparing each image patch to all other image patches [5, 9, 19]. A patch that is different from all other image patches, is considered salient. While this solution works nicely in many cases, it overlooks the correlation between pixels and hence errs in some cases. Furthermore, this solution is inefficient as it requires numerous patch-to-patch distance calculations.

Instead, by analyzing the properties of patches of natural images, we make several observations that improve detection accuracy via a fast and simple solution. Our first observation is that the non-distinct patches of a natural image are mostly concentrated in the high-dimensional space, while distinct patches are more scattered.

This phenomenon is evident from the plots in Figure 3 that were obtained as follows. For each one of 100 images, randomly selected from the ASD data-set [1], we first extract all 9×9 patches and compute the average patch. We then calculate the distance between every patch and the average patch and normalize by the maximum distance. Since the data-set is available with a labeled ground-truth, we analyze separately distinct and non-distinct regions. The solid

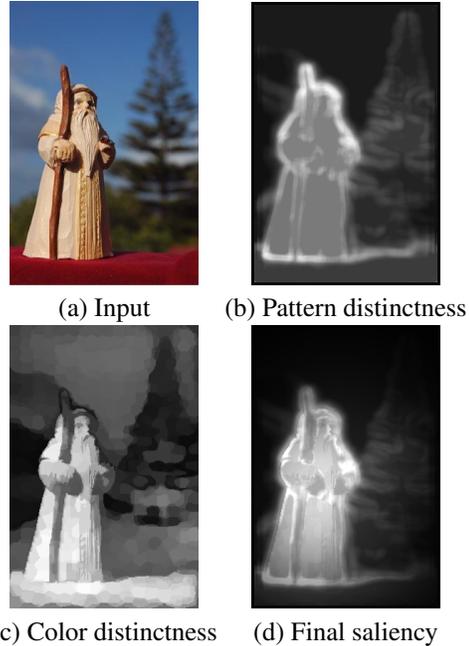


Figure 2. **System overview:** Our pattern distinctness (b), captures the unique textures on the statue, but also part of the tree in the background. Our color distinctness (c), detects the statue fully, but also the red podium and part of the sky. In the final result (d), only the statue is maintained, as it is the only part detected by both.

lines in Figure 3 show the cumulative histograms of the distances between non-distinct patches and the average patch. The dashed lines represent statistics of distinct patches only. As can be seen, non-distinct patches are much more concentrated around the average patch than distinct patches. For example, using the L_1 metric, 60% of the non-distinct patches are within a distance of 0.1, while only less than 20% of the distinct patches are within this distance.

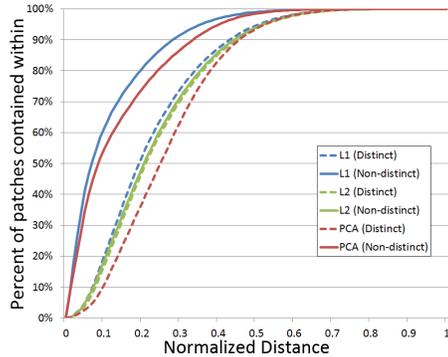


Figure 3. **Scatter distinguishes between distinct and non-distinct patches:** This figure presents the cumulative histograms of the distances between distinct (dashed lines) and non-distinct (solid lines) patches to the average patch. Both L_1 and PCA approaches show that non-distinct patches are significantly more concentrated around the average patch than non-distinct patches.

The plots of Figure 3 suggest that one could possibly identify the distinct patches by measuring the distance to the average patch. In particular, we use the *average patch* p_A under the L_1 norm:

$$p_A = \frac{1}{N} \sum_{x=1}^N p_x. \quad (1)$$

An image patch p_x is considered distinct if it is dissimilar to the average patch p_A .

Note that computing the distance between every patch and the average patch bares some conceptual resemblance to the common approach of [5, 9, 19]. They try to measure the isolation of a patch in patch-space by computing the distance to its k -nearest neighbors. Instead, we propose a significantly more efficient solution, as all patches are compared to a single patch p_A .

At a first thought, this simple idea might not make much sense. Suppose that a certain patch appears in two different images. These two images could have the same average patch, thus the distance of the patch to the average would be equal. However, the saliency of this patch should be totally different, when the images have different patch distributions. This is illustrated in Figure 4. In this figure the patch p_x (marked in red) should be considered as salient in image Im_2 and non-salient in image Im_1 . Yet, the Euclidean distance between p_x and the average patch p_A (dashed purple line) is the same for both images. Were we to rely on this distance to determine distinctness we would likely fail. This behavior is also one of the downfalls of the k -nearest patches approach. As can be seen in Figure 4, the patch p_x has the same k -nearest patches in both images (contained within the dashed red circle) and hence will be assigned the same level of distinctness by [5, 9].

So, how come it works? Using either L_2 or L_1 to measure distances between patches ignores the internal statistics of the image patches. The reason patch p_x should be considered as distinct in image Im_2 is that it is inconsistent with the other patches of image Im_2 . The statistics of patches in each image are different, as evident from the distributions of the patches in Figure 4. This is overlooked by the conventional distance metrics.

Our second observation is that the distance to the average patch should consider the patch distribution in the image. We realize this observation by computing the principal components, thus capturing the dominant variations among patches. We then consider a patch distinct if the path connecting it to the average patch, *along the principal components*, is long. For each patch we march along the principal components towards the average patch and compute the accumulated length of this path.

Mathematically, this boils down to calculating the L_1 norm of p_x in PCA coordinates. Thus, pattern distinctness

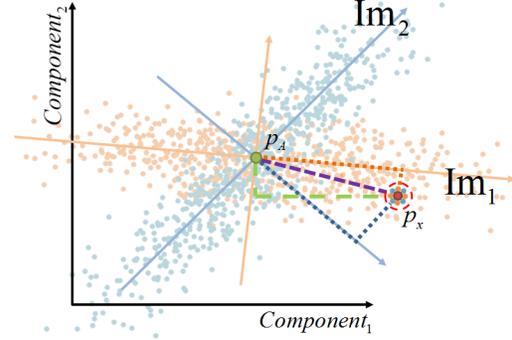


Figure 4. **Saliency should depend on patch distribution:** Im_1 and Im_2 represent two different images whose principal components are marked by the solid lines. The images share the average patch p_A . The patch p_x is highly probable in the distribution of Im_1 and hence should not be considered distinct in Im_1 , while the same patch is less probable in image Im_2 and hence should be considered distinct in Im_2 . The L_2 distance (purple line) and L_1 distance (green line) between p_x and p_A are oblivious to the image distributions and therefore will assign the same level of distinctness to p_x in both images. Instead, computing the length of the paths between p_x and p_A , along the principal components of each image, takes under consideration the distribution of patches in each image. The path for image Im_2 (dashed blue line) is longer than the path for image Im_1 (dashed orange line), correctly corresponding to the distinctness level of p_x in each image.

$P(p_x)$ is defined as:

$$P(p_x) = \|\tilde{p}_x\|_1, \quad (2)$$

where \tilde{p}_x is p_x 's coordinates in the PCA coordinate system.

As shown in Figure 4, the path from p_x to p_A along the principal components of image Im_2 (marked in blue) is much longer than the path along the principal components of image Im_1 (marked in orange). Hence, the patch p_x will be considered more salient in image Im_2 than in image Im_1 .

Figure 5 provides further visualization of the proposed pattern distinctness measure. In this image, the drawings on the wall are salient because they contain unique patterns, compared to the building's facade. The path along the principal components, between the average patch and a patch on the drawings, contains meaningful patterns from the image.

Implementation details: To disregard lighting effects we a-priori subtract from each patch their mean value. To detect distinct regions regardless of their size, we compute the pattern distinctness of Eq. (2) in three resolutions: 100%, 50% and 25% and average them. Finally, we apply morphological operations to fill holes in the pattern map [20].

Computational efficiency: A major benefit of using the approach described above is its computational efficiency,

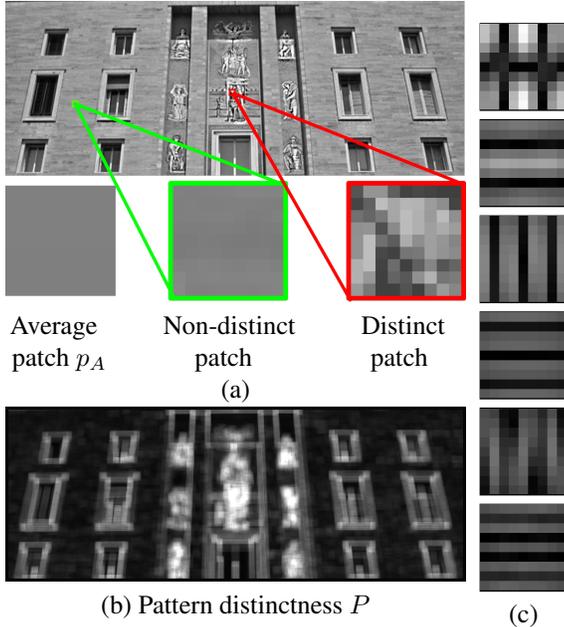


Figure 5. **The principal components:** (a) An image with its average patch and samples of a non-distinct and a distinct patch. (b) Our pattern distinctness P . (c) The absolute value of the top six principal components, added to the “red” patch along the PCA path to p_A . It can be seen that the path from the “red” patch to p_A adds patterns that can be found in the image.

in comparison to the k -nearest patches approach. To compute the PCA, we use only patches that contain patterns and ignore homogeneous patches. To quickly reject homogeneous regions we compute SLIC super-pixels [2] and keep the 25% with highest variance. We then take all the patches from within these super-pixels and use them to compute the PCA.

We compare our accuracy and run-times against those of the k -nearest neighbours approach, using both accurate and approximate search [18]. The evaluation was performed on three well known datasets [1, 3, 15]. Table 1 summarizes our results. To measure accuracy, we report the Area-Under-the-Curve (AUC) and Average Precision (AP) (the higher the better). In addition, we compare run-times. Our approach is more accurate than both Exact-KNN and

Method	Accuracy		Run time (sec/image)	Speedup
	AUC	AP		
ASD [1]				
Exact-KNN	0.794	0.483	39.58	1
Approx-KNN [18]	0.767	0.467	1.63	24.28
PCA-Single-res	0.788	0.466	0.04	989.5
PCA-Multi-res	0.808	0.507	0.26	152.23
SEDI [3]				
Exact-KNN	0.838	0.575	42.63	1
Approx-KNN [18]	0.826	0.6	1.59	26.84
PCA-Single-res	0.842	0.58	0.04	1068.37
PCA-Multi-res	0.849	0.602	0.3	142.1
MSRA [15]				
Exact-KNN	0.855	0.628	39.64	1
Approx-KNN [18]	0.858	0.648	1.56	25.41
PCA-Single-res	0.85	0.619	0.03	1321.33
PCA-Multi-res	0.893	0.723	0.25	158.56

Table 1. **Accuracy and run-time of pattern distinctness:** Our PCA-based approach offers an incredible speedup over the KNN methods, together with an improvement in accuracy. The method was tested on images of a maximal dimension of 150 pixels (excluding the multi-resolution PCA), on a Pentium 2.5GHz CPU, 4GB RAM.

Approximate-KNN, while being significantly faster than both.

A benefit of a faster solution is enabling analysis of images at higher resolutions. This is crucial for some images, as illustrated in Figure 6. Computing pattern distinctness of the input image leads to mediocre detection results for both KNN approaches (Figure 6(b),(c)) as well as for single resolution PCA (Figure 6(d)). By using multiple resolutions, our PCA approach leads to much finer results, while still being orders of magnitude faster than the KNN approaches.

2.2. Color Distinctness

While pattern distinctness identifies the unique patterns in the image, it is not sufficient for all images. This is illustrated in Figure 7(a), where the golden statue is salient only due to its unique color. Much like previous approaches [6, 11], we adopt a two step solution for detecting regions of distinct color. We first segment the image

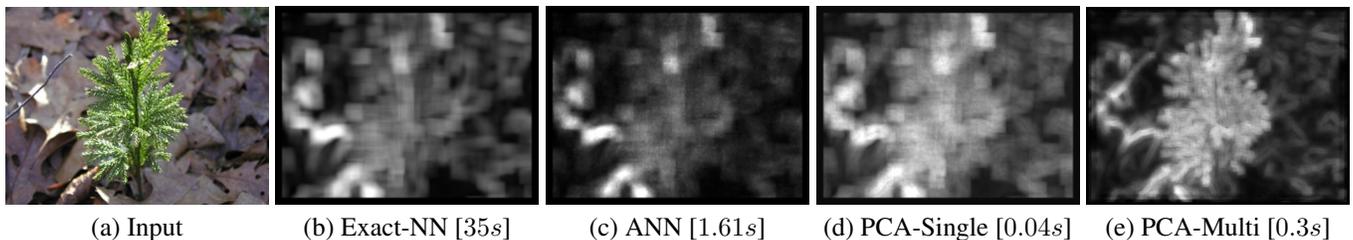


Figure 6. **Processing at high resolution results in higher accuracy:** Thanks to the efficiency of our PCA approach, we are able to process images at multiple higher resolutions leading to improved accuracy, while maintaining significantly lower run-times.



(a) Input

(b) Color distinctness

Figure 7. **Color distinctness:** Color is a crucial cue in image saliency. In this particular image, due solely to color distinctness, the golden statue catches our attention.

into regions and then determine which regions are distinct in color.

The first step is solved by using the SLIC superpixels [2], already computed in Section 2.1 to construct the PCA basis. We solve the second step by defining the color distinctness of a region as the sum of L_2 distances from all other regions in CIE LAB color-space. Given M regions, the color distinctness of region r_x is computed by:

$$C(r_x) = \sum_{i=1}^M \|r_x - r_i\|_2. \quad (3)$$

This calculation is efficient due to the relatively small number of SLIC regions in most images. For further robustness, we compute color distinctness at three resolutions: 100%, 50% and 25% and average them.

Figure 7(b) demonstrates a result of our color distinctness. The golden statue was properly detected, however, also a meaningless dark gap between the statues was detected as distinct in color.

2.3. Putting it all together

We seek regions that are salient in both color and pattern. Therefore, to integrate color and pattern distinctness we simply take the product of the two:

$$D(p_x) = P(p_x) \cdot C(p_x). \quad (4)$$

This map is normalized to the range $[0, 1]$.

To further refine our results, we next incorporate known priors on image organization. First, we note that the salient pixels tend to be grouped together into clusters, as they typically correspond to real objects in the scene. Furthermore, as was shown by [7, 12, 14], people have a tendency to place the subject of the photograph near the center of the image.

To take these observations under consideration, we do the following. We start by detecting the clusters of distinct pixels by iteratively thresholding the distinctness map $D(p_x)$ using 10 regularly spaced thresholds between 0 and 1. We compute the center-of-mass of each threshold result and place a Gaussian with $\sigma = 10000$ at its location. We associate with each of these Gaussians an importance weight,

corresponding to its threshold value. In addition, to accommodate for the center prior, we further add a Gaussian at the center of the image with an associated weight of 5. We then generate a weight map $G(p_x)$ that is the weighted sum of all the Gaussians.

Our final saliency map $S(p_x)$ is a simple product between the distinctness map and the Gaussian weight map:

$$S(p_x) = G(p_x) \cdot D(p_x). \quad (5)$$

We present a few examples of our saliency detection in Figure 8. We note that none of the three considerations: pattern, color or organization (Figures 8(b,c,e)), suffices to achieve a good detection. The pattern distinctness suffers from non-salient distinct patterns, such as the fish drawings on the blue wall (top row). The color distinctness may capture background colors, such as the sky in the penguin road sign (bottom row). The organization map offers a fuzzy map. Yet, by combining the three maps, a high quality detection is achieved (f).

3. Empirical evaluation

To evaluate our approach, we compare it to the state-of-the-art according to the benchmark proposed just recently in [4]. This benchmark suggests five well accepted datasets:

1. **MSRA [15]:** 5,000 images labeled by nine users. Salient objects were marked by a bounding box.
2. **ASD [1]:** 1000 images from the MSRA dataset, for which a more refined manually-segmented ground-truth was created.
3. **SED1 [3]:** 100 images of a single salient object annotated manually by three users.
4. **SED2 [3]:** 100 images of two salient objects annotated manually by three users.
5. **SOD [17]:** 300 images taken from the Berkeley Segmentation Dataset for which seven users selected the boundaries of the salient objects.

According to [4], the “Top-4” highest scoring salient object detection algorithms are: SVO [5], CR [6], CNTX [9], and CBS [11]. Therefore, we compare our results to theirs.

Accuracy: Figure 9 shows the Area-under-the-curve scores for each of the datasets and an overall score of the combined performance over all of the datasets. Unlike the “Top-4” approaches, which perform well on a single dataset and less so on others, our approach significantly outperforms all other methods on all of the datasets (Table 2).

To further evaluate our method, we test it on the dataset of Judd et al. [12]. This dataset is aimed at gaze-prediction, which differs from our task of salient object detection. Still, we show in Figure 10 that our method offers comparable results to the best performing algorithm of the “Top-4” [5].

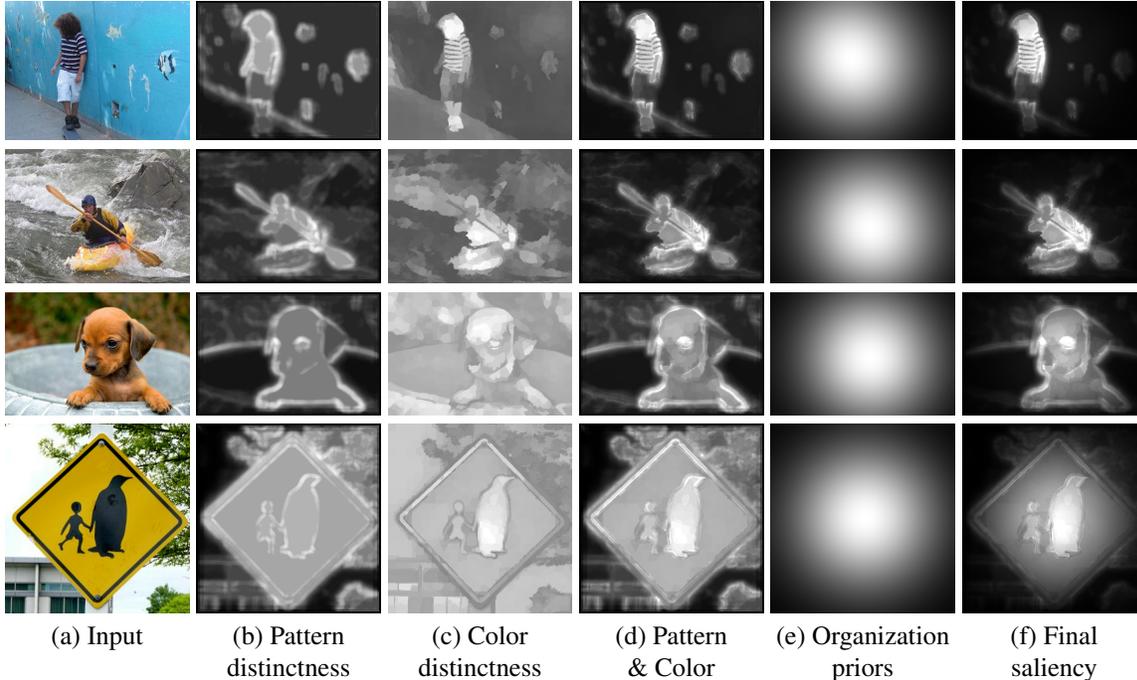


Figure 8. **Combining the three considerations is essential:** Given an input image (a), we compute for each pixel its pattern distinctness (b) and its color distinctness (c). The two distinctness maps are combined (d) and then integrated with priors of image organization (e), to obtain our final saliency results in (f). As can be seen, the final saliency maps are more accurate than each of the components.

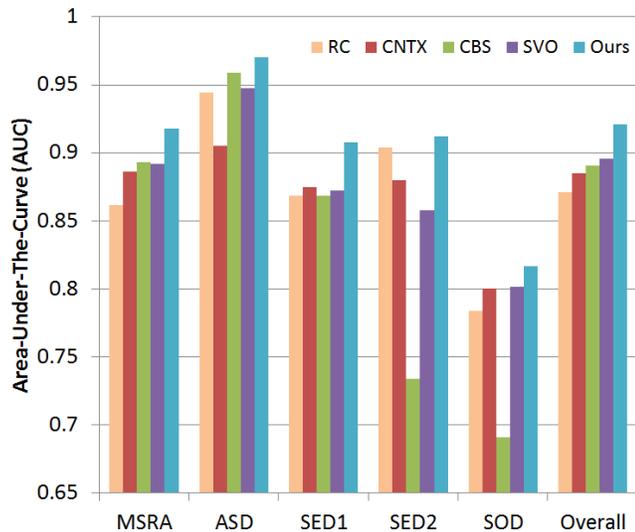


Figure 9. **Detection accuracy:** We present Area-Under-the-Curve (AUC) scores of the “Top-4” algorithms [4] and ours on five well known datasets. Our approach outperforms all other algorithms on all the datasets and in the overall score.

Run-time: Typically, more accurate results are achieved at the cost of a longer run-time. However, this is not our case, as we achieve the most accurate results, while maintaining low run-times, as demonstrated in Figure 11. In particular, the fastest algorithm among the “Top-4” is RC [6], but it is ranked lowest in Table 2. The most accurate algorithm among the “Top-4” is SVO [5], but its running times

Rank	Datasets					Overall
	MSRA	ASD	SED1	SED2	SOD	
1	Ours	Ours	Ours	Ours	Ours	Ours
2	CBS	CBS	CNTX	RC	SVO	SVO
3	SVO	SVO	SVO	CNTX	CNTX	CBS
4	CNTX	RC	CBS	SVO	RC	CNTX
5	RC	CNTX	RC	CBS	CBS	RC

Table 2. **Algorithm ranking:** Our method outperforms all other methods on all datasets as well as in the overall score.

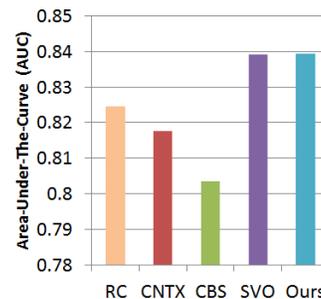


Figure 10. **Gaze Prediction:** Our approach offers comparable results on the gaze-prediction dataset of Judd et. al [12] to that of the top scoring method, SVO [5].

are significantly longer than others (over one minute per image). Such long processing time could render it inapplicable for some applications. Our method, on the other hand, provides even higher accuracy than SVO, while maintaining a

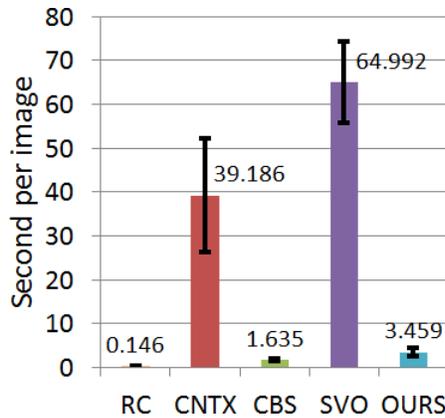


Figure 11. **Run-time:** Our method has a good average run-time per image, compared to the state-of-the-art techniques, while achieving higher accuracy. The reported run-times were computed on the SED1 dataset [3], on a Pentium 2.5GHz CPU, 4GB RAM.

reasonable run-time of ~ 3.5 seconds per image.

Qualitative evaluation: Figure 12 presents a qualitative comparison of our method with the current state-of-the-art. It can be seen that while SVO [5] detects the salient regions, parts of the background are erroneously detected as salient. By relying solely on color, RC [6] can mistakenly focus on distinct background colors, e.g., the shadow of the animal is captured instead of the animal itself. Conversely, CNTX [9] relies mostly on patterns, hence, it detects the outlines of the flower and the cat, while missing their interior. The CBS method [11] relies on shape priors and therefore often detects only parts of the salient objects (e.g., the flower) or convex background regions (e.g., the water of the harbor). Our method integrates color and pattern distinctness, and hence captures both the outline, as well as the inner pixels of the salient objects. We do not make any assumptions on the shape of the salient regions, hence, we can handle convex as well as concave shapes.

4. Conclusion

Let's go back to the title of this paper and ask ourselves what makes a patch distinct. In this paper we have shown that the statistics of patches in the image plays a central role in identifying the salient patches. We made use of the patch distribution for computing pattern distinctness via PCA.

We have shown that we outperform the state-of-art results, while not sacrificing too much run-time. This is done by combining our novel pattern distinctness estimation with standard techniques for color uniqueness and organization priors.

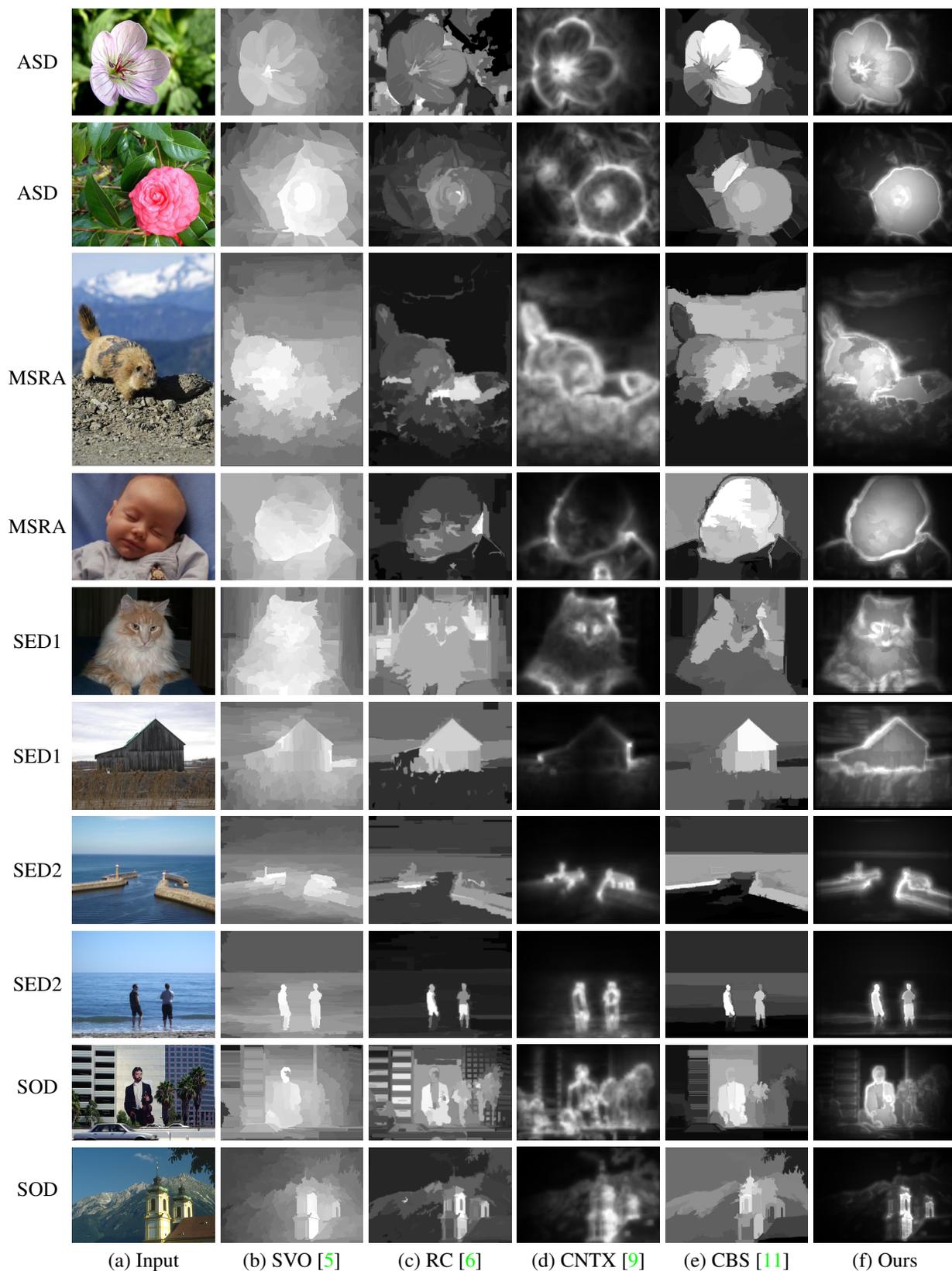
A drawback of our algorithm is not using high-level cues, such as face detection or object recognition. This can be easily addressed, by adding off-the-shelf recognition

tools.

Acknowledgements: This research was supported in part by the the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI), Microsoft, Minerva funding, the Ollendorff foundation and the Israel Science Foundation under Grant 1179/11.

References

- [1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604, 2009. 2, 4, 5
- [2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels. *Technical Report 149300 EPFL*, (June), 2010. 4, 5
- [3] S. Alpert, M. Galun, R. Basri, and A. Brandt. Image segmentation by probabilistic bottom-up aggregation and cue integration. In *CVPR*, pages 1–8, June 2007. 4, 5, 7
- [4] A. Borji, D. Sihite, and L. Itti. Salient object detection: A benchmark. In *ECCV*, pages 414–429, 2012. 1, 2, 5, 6, 8
- [5] K. Chang, T. Liu, H. Chen, and S. Lai. Fusing generic objectness and visual saliency for salient object detection. In *ICCV*, pages 914–921, 2011. 1, 2, 3, 5, 6, 7, 8
- [6] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Global contrast based salient region detection. In *CVPR*, pages 409–416, 2011. 1, 4, 5, 6, 7, 8
- [7] F. Durand, T. Judd, F. Durand, A. Torralba, et al. A benchmark of computational models of saliency to predict human fixations. Technical report, MIT, 2012. 5
- [8] S. Goferman, A. Tal, and L. Zelnik-Manor. Puzzle-like collage. *Computer Graphics Forum*, 29:459–468, 2010. 1
- [9] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *CVPR*, pages 2376–2383, 2010. 1, 2, 3, 5, 7, 8
- [10] L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, 13(10):1304–1318, 2004. 1
- [11] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li. Automatic salient object segmentation based on context and shape prior. In *BMVC*, page 7, 2012. 1, 4, 5, 7, 8
- [12] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *ICCV*, pages 2106–2113, 2009. 5, 6
- [13] C. Kanan and G. Cottrell. Robust classification of objects, faces, and flowers using natural image statistics. In *CVPR*, pages 2472–2479, 2010. 1
- [14] T. Liu, S. Slotnick, J. Serences, and S. Yantis. Cortical mechanisms of feature-based attentional control. *Cerebral Cortex*, 13(12):1334–1343, 2003. 5
- [15] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum. Learning to detect a salient object. *PAMI*, pages 353–367, 2010. 4, 5
- [16] Y. Ma, X. Hua, L. Lu, and H. Zhang. A generic framework of user attention model and its application in video summarization. *IEEE Transactions on Multimedia*, 7(5):907–919, 2005. 1
- [17] V. Movahedi and J. Elder. Design and perceptual validation of performance measures for salient object segmentation. In *CVPRW*, pages 49–56, 2010. 5
- [18] M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISSAPP*, pages 331–340. INSTICC Press, 2009. 4
- [19] H. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. *Journal of Vision*, 9(12), 2009. 2, 3
- [20] P. Soille. *Morphological image analysis: principles and applications*. Springer-Verlag New York, Inc., 2003. 3



(a) Input (b) SVO [5] (c) RC [6] (d) CNTX [9] (e) CBS [11] (f) Ours

Figure 12. **Qualitative comparison.** Saliient object detection results on ten example images, two from each dataset in the benchmark of [4]. It can be seen that our results are consistently more accurate than those of other methods.