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

**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.

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.
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 $\times$ 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 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.

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 distinct patches.

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.
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:

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

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, although 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

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

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 its 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,
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 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

---

**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.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Run time</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC AP (sec/image)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD [1]</td>
<td>0.794 0.483</td>
<td>39.58</td>
<td></td>
</tr>
<tr>
<td>Approx-KNN [18]</td>
<td>0.767 0.467</td>
<td>1.63 24.28</td>
<td></td>
</tr>
<tr>
<td>PCA-Single-res</td>
<td>0.788 0.466</td>
<td>0.04 989.5</td>
<td></td>
</tr>
<tr>
<td>PCA-Multi-res</td>
<td>0.808 0.507</td>
<td>0.26 152.23</td>
<td></td>
</tr>
<tr>
<td>SED1 [3]</td>
<td>0.838 0.575</td>
<td>42.63</td>
<td></td>
</tr>
<tr>
<td>Approx-KNN [18]</td>
<td>0.826 0.6</td>
<td>1.59 26.84</td>
<td></td>
</tr>
<tr>
<td>PCA-Single-res</td>
<td>0.842 0.58</td>
<td>0.04 1068.37</td>
<td></td>
</tr>
<tr>
<td>PCA-Multi-res</td>
<td>0.849 0.602</td>
<td>0.3 142.1</td>
<td></td>
</tr>
<tr>
<td>MSRA [15]</td>
<td>0.850 0.628</td>
<td>39.64</td>
<td></td>
</tr>
<tr>
<td>Approx-KNN [18]</td>
<td>0.858 0.648</td>
<td>1.56 25.41</td>
<td></td>
</tr>
<tr>
<td>PCA-Single-res</td>
<td>0.85 0.619</td>
<td>0.03 1321.33</td>
<td></td>
</tr>
<tr>
<td>PCA-Multi-res</td>
<td>0.893 0.723</td>
<td>0.25 158.56</td>
<td></td>
</tr>
</tbody>
</table>

---

**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.
Our method offers comparable results to the best performing algorithm of the “Top-4” [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]: 1,000 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.

**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 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

<table>
<thead>
<tr>
<th>Rank</th>
<th>Datasets</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ours</td>
<td>Ours</td>
</tr>
<tr>
<td>2</td>
<td>CBS</td>
<td>CBS</td>
</tr>
<tr>
<td>3</td>
<td>SVO</td>
<td>SVO</td>
</tr>
<tr>
<td>4</td>
<td>CNTX</td>
<td>RC</td>
</tr>
<tr>
<td>5</td>
<td>RC</td>
<td>CNTX</td>
</tr>
</tbody>
</table>

Table 2. **Algorithm ranking**: Our method outperforms all other methods on all datasets as well as in the overall score.
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 Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI), Microsoft, Minerva funding, the Olleboff foundation and the Israel Science Foundation under Grant 1179/11.

References

Figure 12. **Qualitative comparison.** Salient 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.