12-2010

# Exploiting intensity inhomogeneity to extract textured objects from natural scenes

Jundi DING

Jialie SHEN
*Singapore Management University*, jlshen@smu.edu.sg

Hwee Hwa PANG
*Singapore Management University*, hhpang@smu.edu.sg

Songcan CHEN

Jingyu YANG

**DOI:** https://doi.org/10.1007/978-3-642-12297-2_1

Published in Computer vision: ACCV 2009: 9th Asian Conference on Computer Vision, Xi'an, September 23-27, pp. 1-10
https://doi.org/10.1007/978-3-642-12297-2_1

# Exploiting Intensity Inhomogeneity to Extract Textured Objects from Natural Scenes

Jundi Ding[1,2,⋆], Jialie Shen[2], HweeHwa Pang[2], Songcan Chen[3], and Jingyu Yang[1]

[1] Nanjing University of Science and Technology, China
[2] Singapore Management University, School of Information Systems
[3] Nanjing University of Aeronautics and Astronautics, China
dingjundi@nuaa.edu.cn, {jlshen,hhpang}@smu.edu.sg,
s.chen@nuaa.edu.cn, yangjy@mail.njust.edu.cn

**Abstract.** Extracting textured objects from natural scenes is a challenging task in computer vision. The main difficulties arise from the intrinsic randomness of natural textures and the high-semblance between the objects and the background. In this paper, we approach the extraction problem with a seeded region-growing framework that purely exploits the statistical properties of intensity inhomogeneity. The pixels in the interior of potential textured regions are first found as texture seeds in an unsupervised manner. The labels of the texture seeds are then propagated through their respective inhomogeneous neighborhoods, to eventually cover the different texture regions in the image. Extensive experiments on a large variety of natural images confirm that our framework is able to extract accurately the salient regions occupied by textured objects, without any complicated cue integration and specific priors about objects of interest.

## 1 Introduction

Extracting salient textured objects in natural scenes has long been a central but tantalizing problem in computer vision. Unlike mosaic texture, natural textures tend to be more random. The texture appearance of an object of interest, e.g. the stripes/blocky-fur of a zebra/wild cat (see the top two rows in Fig.1), may even vary greatly in scale, shape, size and orientation. Textural properties like roughness, linearity, density, directionality, frequency and phase all seem to be far too rudimentary to characterize the plausible regularities behind complex natural textures [1,2,3]. Moreover, the background tends to show a high degree of resemblance in appearance to the contained objects in many situations. The two images in the bottom rows of Fig.1 illustrate such examples. In the square patch marked on each image, the pixels come from both the ground/riffle background and the lizard/otter object (zoomed in second panel). The local differences among them are however very hard to detect in the respective original image. The two factors jointly explain why existing methods based solely

⋆ Supported by National Natural Science Foundation of China of Grant No. 60632050.

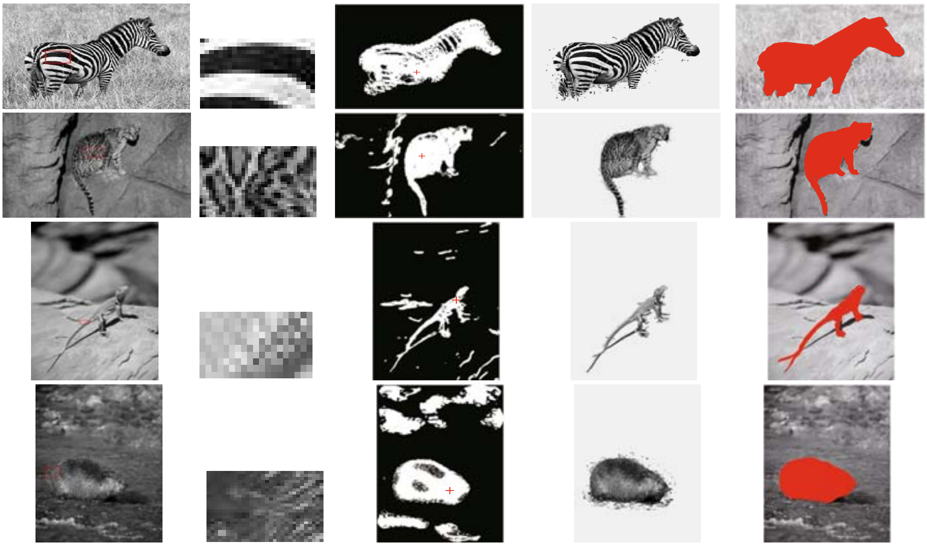

**Fig. 1.** The success of our INP on a set of challenging images. Our work is to aggregate the inhomogeneous pixels around the textured seeds shown in the third panel (in white). Beginning with an arbitrary object-seed (marked by '+'), our INP accurately extracts each desired object (fourth panel) in one piece which is consistent with human segmentation (fifth panel).

on texture homogeneity seldom achieve satisfactory results in natural texture segmentation.

To this end, recent years have seen a surge of interests in this field in two directions: cue integration [4,5,6,7] and interactive or semi-supervised segmentation [8,9,10,11]. In the former algorithms, multiple cues including texture are utilized to reach a combined similarity measure for image segmentation. Each cue handled by a separate module is to assess the coherence of nearby pixels or regions with respect to that cue. Note that each module typically comes with its own set of parameters. Careful assignment of these parameter values is a non-trivial job, which critically influences the segmentation results in many cases [5,11]. The ultimate goal of the latter methodology is to extract the desired objects with some useful prior knowledge about the textures, edges, contours, shapes, curvatures or motions of objects. Different priors have a preference towards different types of task-driven segmentations. Such image prior is usually incorporated into the segmentation process in three ways: (i) being "seeds" specified by users in an initialization step [8]; (ii) being a regularization term formulated into a meaningful energy function [9]; (iii) serving as top-down cues globally blended with a bottom-up segmentation process [10,11,5]. Appropriate prior knowledge is beneficial to a good segmentation, but the challenge of automatically obtaining the prior knowledge for a variety of natural images still lies ahead.

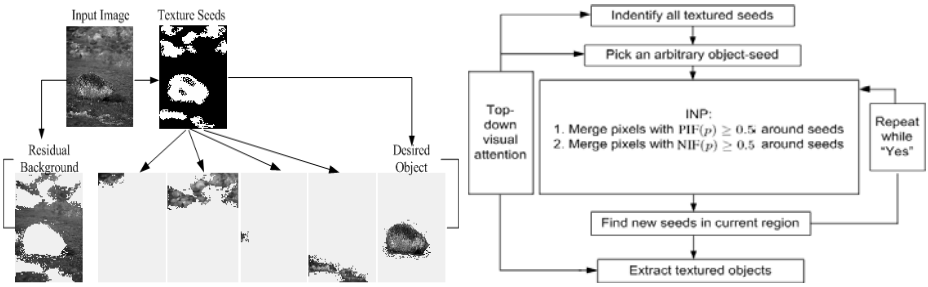


**Fig. 2.** A general schematic framework (left) and flowchart (right) of INP

In this paper, we focus on a different strategy which exploits solely the statistical inhomogeneity in intensity within the image to segment. In practice, almost all textures involve spatially significant fluctuations or variations in pixel intensity in a low-level perspective [2,12]. That is, the pixels in a textured region do not simply adhere to a piecewise-constant or piecewise-smooth homogeneity in intensity. For example, zebras are easily recognizable by their very black and white striped body. As such, we here address the problem of textured object extraction head on via the intensity inhomogeneity. We believe it to be an intrinsic property of just about any texture in the natural world [2].

Our approach, called *Inhomogeneous Neighborhood Propagation* (INP), is designed to work with a seeded region growing strategy. It is to aggregate the nearby inhomogeneous pixels all together in a bottom-up merge manner. Figure 2 (left) depicts a general schematic framework of our INP. The framework consists of three primary steps: (i) identifying all the inhomogeneous pixels which with high probability are in the interior of potential textured regions, and thereby perform the important role of textured seeds; (ii) propagating the labels of the texture seeds through their respective inhomogeneous neighborhoods by a sensible principle that specifies an equivalence relation over the set of textured seeds; (iii) extracting the desired objects according to human vision from the formed saliently textured regions that are covered by adjacent inhomogeneous pixels in the image. Here it is worthwhile to highlight two aspects:

– INP often identifies many background pixels as textured seeds (see the white areas in the third panel of Fig.1, especially the bottom three cases). The reason is that intensity discontinuities may also be caused by grassy or foliaged clutter, surface markings, occlusions, shadows and reflections. All of them are common in the background of natural images. We have made no effort to simplify the image to segment, so the unbiased statistics of the image are well preserved (including disturbances in the background). In such situations, the background is usually fragmented into pieces by INP, see an example in Fig.2 (left).

– INP is robust to the order of the initial seed selection as it virtually yields a *partition* of the set of textured seeds in mathematics. This means, with respect to its two parameters, INP maps every identified seed pixel into one and only one equivalence class. Namely, INP defines an equivalence relation (ER) over a non-empty, finite set of textured seeds. The properties of an ER (i.e. reflexivity, symmetry and transitivity) ensure that the segmented results are invariant under different seed selection orders in INP for a fixed parameter setting.

By virtue of this quality, among the identified textured seeds, we can concentrate on only the object seeds irrespective of those in the background. Specifically, a top-down visual attention is integrated to position an object-seed with a '+' mark as illustrated in the third panel in Fig.1. It allows our INP to grow only the region around the selected object-seed to cover the desired object. In our implementation, each target object in the image is accurately extracted in such a low-cost shortcut (4th panel in Fig.1). In this way, our INP requires only two stages: textured seeds identification and object-seed labels propagation. Section 2 details the two stages as well as the related key concepts. The algorithmic analysis about parameter sensitivity and computational efficiency is discussed in Section 3. Experiment results and evaluations reported in Section 4 confirm the effectiveness of INP in a variety of natural images. Finally, a conclusion is given in Section 5.

## 2 Our Method: INP

In general, an image $\mathbf{I}$ is a pair$(\mathcal{I}, I)$, consisting of a finite set of pixels $\mathcal{I}$ in a grid space $Z^2$ and a mapping $I$ that assigns each pixel $p = (p_x, p_y) \in \mathcal{I}$ with an intensity value $I(p)$ in some arbitrary value space. A textured region here is just described as a function of spatial variations in pixel intensities. In what follows, the work is thus all related to the local intensity contrasts between pixels.

### 2.1 Textured Seeds Identification

Consider the square neighborhood $N(p)$ of each pixel $p$, for a given threshold $\varepsilon \geq 0$, there should be pixels in the sets

$$\Omega(p) = \{q \in N(p) : |I(p) - I(q)| > \varepsilon\} \tag{1}$$

$$\Omega'(p) = \{q \in N(p) : |I(p) - I(q)| \leq \varepsilon\} \tag{2}$$

where $N(p)=\{q\in \mathcal{I}:|p_x\text{-}q_x|\leq k,|p_y\text{-}q_y|\leq k\}$, $k\geq 1$ and $k\in Z$. Since $\Omega_p \bigcup \Omega'_p = N(p)$, it is straightforward for us to define a *pixel inhomogeneity factor* (PIF) as follows:

$$\mathrm{PIF}(p) = \frac{|\Omega(p)|}{|N(p)|} \tag{3}$$

where $|\cdot|$ denotes the cardinality of a set, i.e. the number of elements in the set. This value within [0,1] will be quite discrepant for different pixels. It is obvious that $\mathrm{PIF}(p)<0.5$ when $|\Omega(p)|<|\Omega'(p)|$. In such a situation, the intensity variations between $p$ and most of its adjacent pixels are low. With high probability, they belong to a smooth region [13]. In contrast, $\mathrm{PIF}(p)\geq 0.5$ when $|\Omega(p)|\geq|\Omega'(p)|$. It implies that the majority of pixels around $p$ have intensity values much larger or smaller than that of $p$. In that case, $p$ usually lies in some inhomogeneous image region, such as object contour or boundary [13]. It is thus reasonable to score the intensity inhomogeneity of pixels by PIF($p$)$\geq$0.5.

To ensure that the pixels indeed originated from textured objects, we further highlight the other important aspect of a potential textured pixel $p$, i.e, most of its neighboring pixels should also have inhomogeneous intensities. In this respect, a *neighborhood inhomogeneity factor* (NIF) is put forward in the following:

$$\mathrm{NIF}(p) = \frac{|\mathrm{InNeb}(p)|}{|N(p)|} \tag{4}$$

where InNeb($p$)=$\{q \in N(p)$: PIF($q$)$\geq$ 0.5, $p \in \mathcal{I}\}$. It represents the set of inhomogeneous neighbors of an arbitrary pixel $p$ in the image. Putting the two terms together, the set of seed pixels for growing the desired textured regions is defined as below:

$$\mathrm{SEED} = \{p : \mathrm{PIF}(p) \geq 0.5, \mathrm{NIF}(p) \geq 0.5, p \in \mathcal{I}\} \tag{5}$$

### 2.2 Inhomogeneous Neighborhood Propagation

Algorithmically, our INP belongs to the family of region growing and merging techniques. This old but popular technique has been revived in the last few years due to its native hierarchy configuration and ease of implementation [10,7,11,5,13]. In region growing, pixels being elementary regions are gradually merged to produce larger and larger regions in a sequence of iterative steps. From a probabilistic viewpoint, a demanding statistical test has to be done to give a merging predicate and an order in merging [10,11,5].

A recent work in [13] turns around to first find the most representative "seed" pixels and then define an equivalence relation on the seed set. Each region of interest in the image is hence associated with an equivalence class. In set theory, it ensures the separability of an arbitrary image, as well as the robustness to the selection order of initial seed pixels. To achieve that, the authors in [13] have come up with the segmentation criterion of $\varepsilon$-neighbor coherence. Based on this idea, we specify a principle of neighbor inhomogeneity for texture segmentation.

*For an arbitrary seed* $p \in$ SEED *in a texture region, its neighbor* $q$ *satisfying* PIF($q$)$\geq$0.5 *or* NIF($q$)$\geq$0.5 *should belong to the same textured region as* $p$.

It is obvious that this principle depicts a "transitive relationship" among the seed pixels. That is, assume the pixels $p, q, t \in$ SEED, if $t \in N(q)$ and $q \in N(p)$, $t$ is grouped into the same region as $q$ while $q$ is grouped into the same region as $p$. In such a way, $t$ is also grouped into the same region as $p$. Further, like the $\varepsilon$-neighbor coherence criterion in [13], our principle also specifies an equivalence relation on the set of texture seeds.

**Equivalence Relation.** For any two seed pixels '$p \sim q$' if $p, q$ satisfy either of the two conditions: 1) $p \in N(q)$; 2) there exists a finite number of seed pixels $p_1, p_2, \cdots, p_n$ such that $p \in N(p_1)$, $p_k \in N(p_{k+1})$, $k$=1, $\cdots$, $n$-1, $p_n \in N(q)$.

It is easy to prove the three properties: reflexive ('$p \sim p$'), symmetric ('$p \sim q$' implies '$q \sim p$') and transitive. This principle implies that our INP can start from an arbitrary textured seed to propagate the label through all its inhomogeneous neighbors. Moreover, according to the analysis detailed above, the inhomogeneous pixels involved in the same propagating chain would, with high probability, delineate a single textured object.

**Propagation Termination.** Such an equivalence relation can partition the set of texture seeds into several equivalence classes. The number of equivalence classes just determines the number of interesting regions. Note that each ultimate textured region contains the texture seeds in an equivalence class and some non-seed texture pixels besides. The presence of these non-seed texture pixels is responsible for the termination of the label propagation. In other words, the growth of a region will stop when there is no new textured seed in this region. Figure 2 (right) summarizes the flowchart of INP for object extraction with an arbitrarily picked object-seed.

## 3 Algorithm Analysis of INP

**Parameter Sensitivity.** INP involves two parameters $k$ and $\varepsilon$. On one hand, $k$ determines the size of the local neighborhood of each pixel. In the common case, an optimal $k$ could be chosen in a range of 5-12. However, a "huge" close-shot object usually needs a larger $k$ ($\geq$12); while a "little" long-shot object requires a smaller $k$ ($\leq$5). On the other hand, with respect to a given neighborhood size $k$, one can figure out some meaningful statistics in intensity such as $\mathrm{Mean}(k)_p$ and $\mathrm{Ave}(k)$. They are respectively formulated in Eq.6:

$$\mathrm{Mean}(k)_p = \frac{\sum_{q \in N(p)} \mid I(p) - I(q) \mid}{\mid N(p) \mid}, \mathrm{Ave}(k) = \frac{\sum_{p \in \mathbf{I}} (\mathrm{Mean}(k)_p)}{\mid \mathbf{I} \mid} \tag{6}$$

By definition, $\mathrm{Mean}(k)_p$ exposes the mean difference in intensity within the neighborhood of each pixel $p$; and $\mathrm{Ave}(k)$ is the average value of all $\mathrm{Mean}(k)_p$, which reflects the global variation in intensity in the image.

In addition, the threshold of intensity contrast (see Eq.1 or Eq.2) $\varepsilon$ characterizes the degree of inhomogeneity or homogeneity in intensity between pairwise neighboring pixels. For a central pixel $p$, if $\varepsilon$ is larger than the mean intensity difference in its neighborhood $\mathrm{Mean}(k)_p$, most of its neighbors will be in the set $\Omega'(p)$ instead of $\Omega(p)$. From the discussion detailed above, $p$ will be not an inhomogeneous pixel. Otherwise, if $\varepsilon$ is smaller than $\mathrm{Mean}(k)_p$, most of its neighbors will appear in the set $\Omega(p)$ and thereby $p$ becomes an inhomogeneous pixel.

However, it is impossible to select a proper $\varepsilon$ with regard to $\mathrm{Mean}(k)_p$ which varies with different pixels. A good candidate for $\varepsilon$ is $\mathrm{Ave}(k)$, which is invariant for a given $k$. In practice, the value of $\varepsilon$ in our experiments fluctuates around $\mathrm{Ave}(k)$. When intensities of the foreground pixels (e.g. a zebra roaming the grassland) vary sharply, $\varepsilon$ is selected to be a little smaller than $\mathrm{Ave}(k)$. If the intensities of the background pixels (e.g. a clutter background with a flying bird)

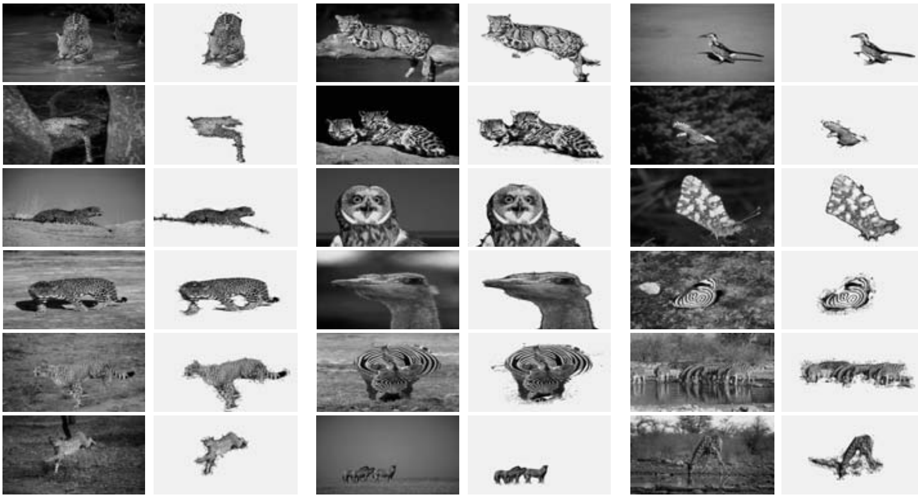

**Fig. 3.** Experimental results on Corel images

vary significantly, $\varepsilon$ is set to be a little larger than Ave($k$). An adaptive $\varepsilon$ has been observed in the range [Ave($k$)–10, Ave($k$)+10] in our experiments.

**Computational Efficiency.** Recall that INP performs object extraction following the general flowchart shown in Fig.2 (right). It is easy to see that the most time-consuming work is the identification of textured seeds. It requires the calculation of $\Omega_p$ for each pixel. Because of the properties of ER, it is not necessary to compute $N_p$ and $\Omega_p$ of the pixels which are in the first and last $k$ rows/columns of the image $\mathbf{I}(w, h)$, where $w$ and $h$ are the width and height of the image respectively. Let $N = (w - 2k) * (h - 2k), M = (2k + 1)^2 - 1, k \in Z$, the running time of calculating all $\Omega_p$ is O($MN$). When $k$ is not very large ($\leq 12$), it takes nearly O($N$) in proportion to the size of the image $\mathbf{N}$. In addition, the recursive propagation procedure for covering all object pixels takes less than O($\mathbf{N}$) as only those pixels around the selected object-seed are scanned once. Besides, the automatic selection of $\varepsilon$ requires computing Mean$(k)_p$ and Ave($k$). It takes O($MN$) like the calculation of $\Omega_p$. Overall, our INP is efficient with a computational complexity of O($MN$) that is nearly linear in the size of the image.

## 4 Experiments and Evaluation Results

We have conducted extensive experiments and comparisons to evaluate the performance of our INP. We first test the qualitative effectiveness of INP on a large number of natural images that contain a variety of challenging textures. All of the sample images are readily available from the Corel image library [13]. For a further quantitative evaluation, we apply our INP to all the 100 gray level images in an open database compiled by Alpert et al. recently in [5]. The F-measure is used to assess the consistency of our results with the ground-truth segmentations in the database.

**Qualitative Results on The Corel Dataset.** The Corel dataset is commonly used in computer vision. It contains 30,000 images covering a wide range of subject matters. For our extraction task, we are interested in those images that include distinct physical objects in the natural world, particularly the ones that are of animals in natural scenes. The animal furs by nature exhibit a variety of challenging textures. Figure 1 shows several representative samples, where our extraction results (fourth panel) are in marked agreement with human segmentations (red, fifth panel). The salient objects of interest together with the long but thin bodies, legs or tails are all segmented in one piece, even if the animals are camouflaged against their backgrounds due to the shadows, illuminations and reflections. Figure 3 further illustrates our results on a set of challenging images. The leopards in the first panel are in different poses (crouching, sitting, eating, standing, walking, running, etc.) in different cluttered backgrounds. A few "Leopard" images among them occur quite often in the texture segmentation literature [11]. These methods have had to integrate many cues of intensities, contours, shapes and motions in order to produce satisfactory results. It is unclear whether they are robust to the variations in poses, shadows, shapes and motions in our experiments. Exploiting the naive intensity inhomogeneity, our INP succeeds in these difficult "Leopard" images. The integrity of each leopard object is well preserved. Moreover, the "Leopard" objects extracted by INP are consistent with the human semantic perception. Other results are also presented on the images with the animal tiger, butterfly, birds, zebra or giraffe. They exhibit a rich diversity of texture appearances in randomness and irregularity. Despite these difficulties, our INP still yields good figure-ground separation. These extracted salient regions can be useful for content-based or object-based image retrieval, indexing and classification in multimedia analysis.

**Quantitative Evaluation of Consistency.** A quantitative evaluation of the results produced by segmentation algorithms is challenging, since it is difficult to come up with canonical test sets providing ground truth segmentations. Recently, Alpert et al. has compiled a new database containing 100 gray images along with ground truth segmentations [5]. To avoid potential ambiguities, the selected images clearly depict one object in the background. Each image is segmented manually by three different people. A pixel is declared as foreground only when it was marked as foreground by at least two people. For an objective evaluation, we have applied our INP to all the 100 images. Some results are shown in Fig.4. Note that the salient regions here represent more generic textures in the natural scenes. Visually, our results are very consistent with the human-driven segmentations (in red color) on the same image. To clarify this point, we use the F-measure to assess its consistency quantitatively [5]. The amount of fragmentation is determined simply by the number of segments needed to cover the foreground object. Table 1 presents the F-measure scores of our results on

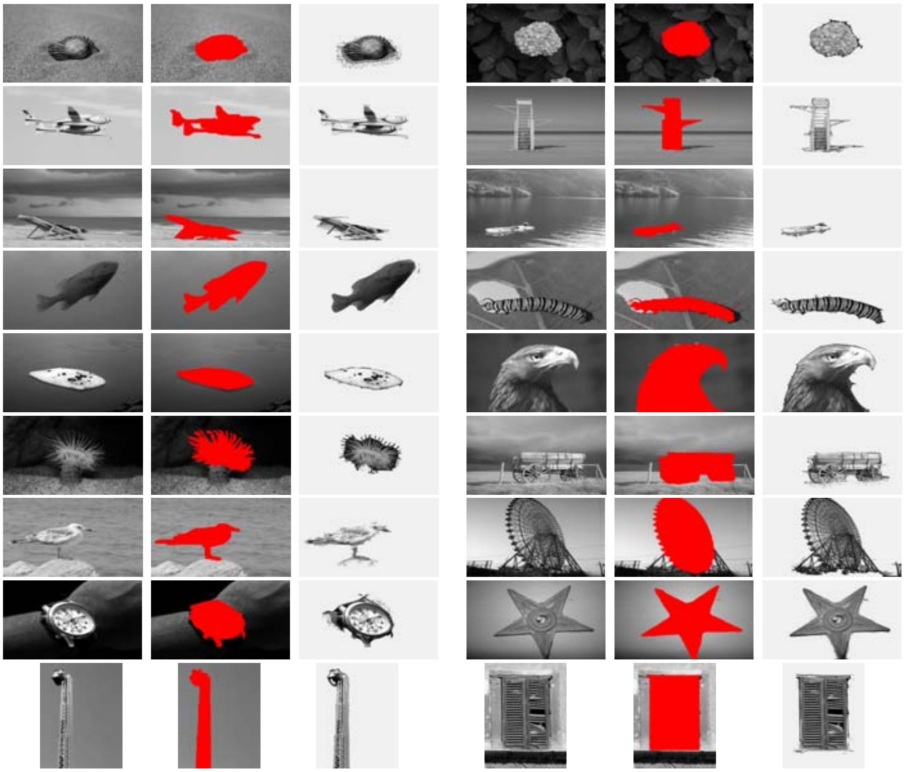

**Fig. 4.** Experimental comparisons on a new image database from [5]

**Table 1.** Our F-measure Score on Test Images from The Dataset in [5]

| Image | F-measure Score | Image | F-measure Score |
|---|---|---|---|
| IMG_2577 | 0.83878 | nitpix_P1280114 | 0.9671 |
| 0677845-R1-067-32_a | 0.94971 | 110016671724 | 0.8092 |
| aaa | 0.8092 | boy-float-lake | 0.9534 |
| Bream_In_Basin | 0.96807 | caterpiller | 0.91923 |
| DSC04575 | 0.93302 | DSC_0959 | 0.9815 |
| tendrils | 0.80518 | DSCF0034 | 0.93466 |
| DSCF0459 | 0.85897 | osaka060102_DYJSN071 | 0.84795 |
| PIC1092515922117 | 0.92676 | PIC7227 | 0.96282 |
| PIC1080629574 | 0.94435 | windowCN_0078 | 0.96394 |

the test images in Fig.4. The large F-measures (some even approximate to the maximum 1) achieved by INP is another evidence of its effectiveness[1].

[1] The averaged F-measure score of our INP is $0.82 \pm 0.027$ that is competitive with the highest one $0.86 \pm 0.012$ reported in Table 1 in [5].

## 5 Summary

In this paper, we present a novel approach called INP to extract textured objects in natural images by exploiting intensity inhomogeneity. Along with a top-down visual attention, INP works by aggregating neighboring inhomogeneous pixels together within a seeded region growing framework. It requires no complicated computations on multi-cue integration or specific priors about the objects of interest. Both theoretical analysis and experiment results confirm that our INP is **easy** to interpret and implement, **efficient** in computational cost and **effective** for textured object extraction in a variety of natural images.

## References


1. Julesz, B.: Textons, the elements of texture perception and their interactions. Nature 290, 91–97 (1981)
2. Chen, C., Pau, L.: Texture analysis. In: Wang, P.S.P. (ed.) The Handbook of Pattern Recognition and Computer Vision, 2nd edn., ch. 2.1, pp. 207–248. World Scientific Publishing Co., Singapore (1998)
3. Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. International Journal of Computer Vision 62(1-2), 61–81 (2005)
4. Deng, Y., Manjunath, B.S.: Unsupervised segmentation of color-texture regions in images and video. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(8), 800–810 (2001)
5. Alpert, S., Basri, M., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: International Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
6. Liang, K., Tjahjadi, T.: Adaptive scale fixing for multiscale texture segmentation. IEEE Transactions on Image Processing 15(1), 249–256 (2006)
7. Xia, Y., Feng, D., Zhao, R.: Morphology-based multifractal estimation for texture segmentation. IEEE Transactions on Image Processing 15(3), 614–623 (2006)
8. Xiang, S., Nie, F., Zhang, C.: Texture segmentation: An interactive framework based on adaptive features and transductive learning. In: Narayanan, P.J., Nayar, S.K., Shum, H.-Y. (eds.) ACCV 2006. LNCS, vol. 3851, pp. 216–225. Springer, Heidelberg (2006)
9. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for supervised texture segmentation. International Journal of Computer Vision 46(3), 223–247 (2002)
10. Nock, R., Nielsen, F.: Statistical region merging. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(11), 1452–1458 (2004)
11. Sharon, E., Galun, M., Sharon, D., Basri, R., Brandt, A.: Hierarchy and adaptivity in segmenting visual scenes. Nature 442(7104), 810–813 (2006)
12. Suyash, P., Awate, T.T., Whitaker, R.T.: Unsupervised texture segmentation with nonparametric neighborhood statistics. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 494–507. Springer, Heidelberg (2006)
13. Ding, J., Ma, R., Chen, S.: A scale-based coherence connected tree algorithm for image segmentation. IEEE Transactions on Image Processing 17(2), 204–216 (2008)