Saliency-Based Boundary Object Detection in Naturally Complex Scenes

Kohji Kamejima
Faculty of Information Science and Technology, Osaka Institute of Technology
1-79-1 Kitayama, Hirakata 573-0196 JAPAN
e-mail: kamejima@is.oit.ac.jp
October 5, 2011

Abstract: A stochastic scheme is presented for cooperative detection of landmark objects distributed in roadway boundaries. By indexing chromatic diversity within a locally Gaussian color space, saliency patterns are extracted with respect to the as-is primary system. Through saccadic scan of the saliency patterns, boundary objects are successively articulated into a system of fractal attractors consistent with the ground-object structure. As the result, the fractal model is indicated within the perspective of the naturally complex scenes.

Keywords: Object Detection; Saliency-based Approach; Simulated Saccade; Naturally Complex Scene

Table of Contents

1 Introductory Remarks 2
2 Chromatic Complexity Generator 3
3 Stochastic Saliency Index 4
4 Simulated Saccade on Saliency Index 5
5 Multi-Fractal Articulation 7
6 Discussions 8
7 Concluding Remarks 9

1Proceedings of the 20th International Symposium on Robot and Human Interactive Communication (IEEE-RoMan11), July 31–August 2, Atlanta, Georgia, pp.407–412.
1 Introductory Remarks

Annotating large scale geographical information based on landmark objects plays a crucial role for over-the-horizon cooperation of maneuvering processes; bottom-up aggregation of object images from the multitude of robotic probes yields a computational basis for dynamic monitoring and cooperative optimization of transportation systems; anticipative localization and guidance along resulted trajectory makes it possible for individual robotic vehicle to a priori focus on landmarks in future scenes as well. Due to the diversity and mutual independence of the viewer specific intentions, it is essential for such over-the-horizon processes to maintain continuous interaction with human’s inherent perception.

Despite the diversity of view points and observation conditions, natural scenes can afford to exhibit environment specific landmarks to be identified within individual maneuvering context. To control the focus to such a landmark object, perception processes should generate the array of image features and activate ‘feature integration’ schemes for visualizing ‘saliency patterns’ in the perspective of the complex scenes [9]. In the early stage of perception, it is known that the cue to object detection is extracted via the multi-scale Gaussian filter [17], [18] to be implemented by inherent distributed parameter system [10]. Such ‘fast’ scale information has been exploited for the ground-object separation as shown in Fig. 1; we have the representation of a connected open space in terms of a closed graph spanning a fractal attractor [13]. In many natural scenes, the open space is confined by the distribution of boundary objects; the same optic array is accepted by a more sophisticated system where spectral diversity of the incoming light is stochastically factorized into the trichromatic system [3]. Despite the loss of the depth information in the color perception process, it is easy for inherent vision to associate ‘matted’ images with 3D objects [16].

In conventional feature integration scheme, observed images are assumed to be structured in terms of preassigned aspect of image features. In many natural scenes, the image features should support the multitude of viewer specific decision makings. However, the maintenance of the consistency is still an open problem in the feature representation as inter-process support. Adding to it, practical scene images exhibit superfluous image features redundant for individual decision process. In such a situation, it is well known that the contextual analysis of a priori fixed image features easily incurs combinatorial explosion.

To cope with such a computational difficulty, we introduce a preestablished restriction of image features to be focused in naturally complex scenes. As the results of the evolution in the really existing world, human’s vision system is equipped with not-yet-explicated information processing mechanism for understanding the scenes filled with friendly or undesirable neighbors [20]. To maintain the multitude of mutual coexistence in resulted environment, the neighbors substantiate warning signs and/or informative designs to be accepted as ‘saliency patterns’. As intentional participants of the co-evolution process, inherent vision should be ‘preset’ to chromatic saliency [4]: scene specific spectral complexity discriminating object patterns in the perspective of the scenes. Through the co-evolution process, the inherent vision has developed an attention control mechanism within the surroundings [19] on the premise that imminent decision making should be evoked by ‘light speed transformation’ of optic array [7] and universal preference to a class of fractal patterns [6], [8]. Such physical-geometric structure underlying the naturally complex scenes makes it possible to implement a cooperative decision scheme with inherent vision; it has been demonstrated that the mental structure of an ‘intelligent vehicle’ can be developed through intentional articulation of ambient light [2].

As a part of the really existing world, naturally complex scenes should present the entire participants with ‘readable image’ of the saliency sign/design. For accepting the context free representation, in this
paper, we extend the scope of the perception to ‘subconscious’ aspect of image features: the fluctuation of local scale shift and, in particular, the diversity of subtle chromatic scattering as invariant representation spanning apparent diversity of scene images. By indexing the chromatic diversity in a probabilistic color space, in what follows, a stochastic dynamics is introduced for nondeterministically scanning saliency patterns. To detect the boundary objects, a system of fractal attractors is designed on the saliency patterns and visualized within the generic ground-object structure underlying the scene images.

2 Chromatic Complexity Generator

To cooperate with inherent perception, the object detection process should be sensitive to the random distribution of ‘saliency colors’ in uncontrolled ambient light as indicated in Fig. 1. For simulating such a detection process, let the chromatic diversity of the incoming light be identified within a color space coordinated by tricolor primaries $R, G, B$. In this color space, the information conveyed by the spectral distribution is described in terms of the subjective weight with respect to the primary $f^{RGB}(\cdot) \in RGB$, i.e.,

$$
\begin{align*}
\chi_{\rho}(\cdot) = f^{RGB}(\cdot) / |f^{RGB}|.
\end{align*}
$$

Define $\phi_\omega = \chi_{RGB}(\cdot) / |\chi_{RGB}|$. By identifying the totality of the chromatic information $\phi_\omega$ with the positive part of a unit sphere $\Phi^+$, we can induce the following measure:

$$
\begin{align*}
g_\alpha (\phi|\phi_\omega) = \frac{1}{2\pi\alpha} \exp \left[ - \frac{|\phi - \phi_\omega|^2}{2\alpha} \right],
\end{align*}
$$

for $\phi \in \Phi^+$; following experimental studies using various types of roadway scene images including natural objects, the diversity scale $\alpha$ should be adjusted to $1/10 \sim 1/100$ [12]. For sufficiently small deviation $|\phi - \phi_\omega|$, the measure $g_\alpha (\phi|\phi_\omega)$ approximates the Gaussian distribution on tangential space at $\phi_\omega$.

Noticing that the inherent perception of the natural colors is essentially regenerated by linear combination of the trichromatic primaries [5], let the index $\phi_\omega$ be projected into the following planar representation of the color space:

$$
\begin{align*}
\Gamma \ni \gamma = e^{RGB} \phi_\omega, \quad e^{RGB} = [e^R, e^G, e^B],
\end{align*}
$$

with primary vector $e(\cdot) = [\cos \theta(\cdot), \sin \theta(\cdot)]^T$ assigned at

$$
\begin{align*}
\theta_R = \pi/2, \quad \theta_G(\theta_B) = \theta_R (-\pi/3).
\end{align*}
$$

Consider the fractal model generating the chromatic complexity for supporting the inherent saliency impression within the locally Gaussian color space. To this end, let a set of samples $s = \{ \phi_i, i = 1, 2, \ldots \}$ with size $||s||$ be collected in a scene image to generate the information $\varphi_\rho (\gamma|s)$ satisfying

$$
\begin{align*}
\frac{1}{2} \Delta \varphi_\rho (\gamma|s) + \rho [\chi_s - \varphi_\rho (\gamma|s)] = 0,
\end{align*}
$$

where $\chi_s$ denotes the aggregation of Dirac’s delta measure distributed on the set $\Gamma_s = \{ \gamma(\phi_\omega) \mid \phi_\omega \in s \}$. Assume that the distribution $\Gamma_s$ is identified with a degenerated version of fractal attractor $\Xi_s$. Suppose that the attractor is generated via the iterated function system [1] and consider the inverse problem: the estimation of a set of fixed points $\tilde{\Pi} = \{ \tilde{\pi} \}$ essentially governing the fractal dynamics [11]. By adjusting the parameter $\rho$ to the complexity of the iterated function system, we can evaluate the probability for capturing the not-yet-identified attractor $\Xi_s$ in terms of the solution $\varphi_\rho (\omega|s)$.

Noticing that the fixed point must be allocated on the Laplacian-Gaussian boundary $\partial^2 \Xi_s$ with respect to $\varphi_\rho (\gamma|s)$, we can identify an $as-is$ primary with the fixed point $\tilde{\Pi}$. The identification process is divided into the following three steps. First, a possible fixed point $\tilde{\gamma}_0$ is located on the Laplacian-Gaussian boundary $\partial^2 \Xi_s$ and expanded via the following successive scheme:

$$
\begin{align*}
\tilde{\Gamma}_t \equiv \tilde{\Gamma}_0 \cup \partial \tilde{\Gamma}_t, \quad \tilde{\Gamma}_0 = \{ \tilde{\gamma}_0 \},
\end{align*}
$$

where the increment is selected with respect as follows

$$
\begin{align*}
d\tilde{\Gamma}_t = \{ \partial \tilde{\gamma}_f \mid \forall \partial \tilde{\gamma} : \tilde{\gamma} (\partial \tilde{\gamma}_f, \tilde{\Gamma}_t) \geq \tilde{\gamma} (\partial \tilde{\gamma}, \tilde{\Gamma}_t) \},
\end{align*}
$$

where $\gamma_{\tilde{\gamma}} \in \partial^2 \Xi_s - \tilde{\Gamma}_f$, $\partial \tilde{\gamma} \in \partial^2 \Xi_s - \tilde{\Gamma}_f$.
Saliency-Based Boundary Object Detection in Naturally Complex Scenes

with respect to \( \eta (\gamma, \Lambda) = \min_{\lambda \in \Lambda} |\gamma - \lambda| \). Next, a subset \( \{ \hat{\gamma}_k \} \) satisfying the following conditions is selected as an estimate of the vertices

\[
\forall m, k : \quad \theta_{mk} - \theta_{nk} < \pi,
\]

\[
\gamma_{\ell} - \hat{\gamma}_k = |\gamma_{\ell} - \hat{\gamma}_k| e^{i(\varphi_{\ell} + \theta_{\ell})}, \quad \gamma_{\ell} \in \hat{\Gamma}, \quad \hat{\gamma}_k = |\hat{\gamma}_k| e^{i\theta_k}.
\]

Finally, the distribution of \( \hat{\Gamma} \) is expanded along the following repulsive force:

\[
d_{\hat{\gamma}_k} = \sum_{\hat{\gamma}_j \in \hat{\Gamma}} \left( \hat{\gamma}_k - \hat{\gamma}_j \right) g_{\alpha}(\phi_k, \phi_j),
\]

within the possible coloring circle \( |\hat{\gamma}_k| \leq 1 \); in this equation, a the repulsive force is induced in the color space to separate the vertices \( \{ \hat{\gamma}_k \} \) towards a set of \( \text{as-is} \) primaries \( \Pi \).

The scheme (4) combined with (5) yields a set of fixed points to be associated with a set of contraction mapping for regenerating the distribution \( \Gamma_s \). By adding the dynamics (6), we have a design of the iterated function system for the restoration of the attractor \( \Xi_s \); the control parameter of the fractal dynamics is given by \( \hat{\pi}_i = \tilde{\pi}_i + \pi_i, \text{RGB} \) where

\[
\tilde{\pi}_i = \frac{2}{3} (e_{\text{RGB}})^T \hat{\gamma}_i, \quad 3\hat{\pi}_i^2 + 2\hat{\pi}_i^T e_{\text{RGB}}, \hat{\pi}_i + |\hat{\pi}_i|^2 = 1.
\]

### 3 Stochastic Saliency Index

The chromatic diversity arising in the scene images can be indexed within the framework of Kolmogorov’s complexity [15]. To this end, the probability for the pixelwise selection of a primary \( \hat{\pi} \in \Pi \) is evaluated as follows:

\[
p(\omega|\hat{\pi}_i) = \frac{g_{\alpha}(\phi_\omega|\hat{\pi}_i)}{\sum_{\hat{\pi}_i \in \Pi} g_{\alpha}(\phi_\omega|\hat{\pi}_i)}.
\]

By evaluating the complexity arising in the ‘primary mixing” process in terms of the associated Shannon’s entropy

\[
\hat{H}_\omega = - \sum_{\hat{\pi}_i \in \Pi} p(\omega|\hat{\pi}_i) \log p(\omega|\hat{\pi}_i),
\]

we can induce the following distribution

\[
\hat{\psi}_\omega = e^{-\hat{H}_\omega} / \int_{\Omega} e^{-\hat{H}_\omega} d\omega,
\]

as an index of pixelwise saliency. Noticing that the ‘negative entropy' is available as a global evaluation of predictability in the locally Gaussian color space, we can exploit the saliency index for discriminating ‘easy-to-predict' patterns in the naturally complex scene.

The implication of the chromatic complexity generator is demonstrated in Figs. 2 and 3. The distribution of the samples \( s \) in the scene image (Fig. 1) is indicated in Fig. 2(a) where the distribution of

---

**Figure 2:** Chromatic Complexity Generation
Saliency-Based Boundary Object Detection in Naturally Complex Scenes

the chromatic information $\phi_\omega$ is confined by RGB-primary. In (b) the scene specific palette is displayed with the distribution of possible fixed points; $\hat{\gamma}_r, \hat{\gamma}_g, \hat{\gamma}_b$ to be associated with the trichromatic primary and scene specific $\hat{\gamma}_{rg}, \hat{\gamma}_{gb}$; These fixed points are separated to yield the as-is primaries $\hat{\pi} = \{ \hat{\pi}_r, \hat{\pi}_g, \hat{\pi}_b \}$. By using the estimate $\hat{\pi}$, we have a design of saliency filter to enhance landmarks sign/design as indicated in Fig 3; the degeneration of possible saliency colors is restored via spectrum shift and/or split towards the as-is primary $\hat{\pi}$. Thus, we can confine object image on the $\hat{\psi}_\omega$-distribution.

4 Simulated Saccade on Saliency Index

For scanning landmark objects in complex background, we induce a probability distribution based on the saliency index $\hat{\psi}_\omega$. Let a pixel in a saliency pattern be located at $m_\omega$ in the image plane $\Omega$ and consider 2D Gaussian distribution $g_\sigma(\omega|m_\omega), \omega \in \Omega$ conditioned by $m_\omega$. Noticing the following generation mechanism

$$g_\sigma(\omega|m_\omega) - \delta_{m_\omega} \sim \frac{\sigma}{2} \Delta g_\sigma(\omega|m_\omega),$$

with Dirac’s delta measure $\delta_{m_\omega}$, we can evaluate the probability for capturing a Brownian motion with origin $m_\omega$ at some time $t \leq \sqrt{\sigma}$ in terms of $g_\sigma(\omega|m_\omega)$. This evaluation can be extended to the saccadic scan of distributed targets generating joint invariant measure $\hat{\psi}_\omega$; in this case, the capturing probability $\varphi_\sigma(\omega|\hat{\Pi})$ is generated as the solution to the following partial differential equation:

$$\frac{\sigma}{2} \Delta \varphi_\sigma(\omega|\hat{\Pi}) + \left[ \hat{\psi}_\omega - \varphi_\sigma(\omega|\hat{\Pi}) \right] = 0,$$

(8)

where $\sigma$ is adjusted to the scale of object images.

Suppose that an a priori estimate of the object is given as a pixel $\omega_m$ in the saliency pattern. By evaluating the deviation from the nearest delta measure in terms of the deviation factor given by

$$n_\omega^{\hat{\Pi}} = \sqrt{-2\sigma \log \left[ \frac{\varphi_\sigma(\omega|\hat{\Pi})}{\max_{\omega \in \Omega} \varphi_\sigma(\omega|\hat{\Pi})} \right]},$$

we have the following model update scheme:

$$\omega^m_{t+dt} = \omega^m_t + d\omega^m_t,$$

(9)

$$d\omega^m = 2n_\omega^{\hat{\Pi}} \frac{\nabla \varphi_\sigma(\omega^m_t|\hat{\Pi})}{|\nabla \varphi_\sigma(\omega^m_t|\hat{\Pi})|}.$$

By iterating the update process (9), we have a saliency sensitive scanning mechanism in the scene image.

The effectiveness of the scanning mechanism is verified through simulation studies. A part of the simulation results are shown in Fig. 4 where the capturing probability $\varphi_\sigma(\omega|\hat{\Pi})$ is generated with the excitation term $\hat{\psi}_\omega$ consisting of two delta measures designated by target ($\mathcal{M}_T$) and distractive objects.
Saliency-Based Boundary Object Detection in Naturally Complex Scenes

Figure 4: Simulated Saccade

(a) Saliency Sensitive

(b) Random Walk

Figure 5: Saliency Pattern Capturing

(\mathcal{M}_D); the initial estimate \( \omega^m_0 \) is shifted by the random vector \( d\omega^m \) successively. As demonstrated in Fig. 4, the update scheme (9) generates explorative motion towards not-yet-identified targets.

By adjusting the scale \( \sigma \) in (8) to the resolution for object detection, we have the capturing probability \( \varphi_\sigma \left( \omega|\hat{\Pi} \right) \) associated with the fractal attractor on a saliency pattern. This implies that saliency patterns can be efficiently scanned via the simulated saccade process. Figure 5 shows examples of experimental results; scanning processes by using the dynamics (9) and random sampling are illustrated in (a) and (b), respectively. Since the information \( \varphi_\sigma \left( \omega|\hat{\Pi} \right) \) is rapidly diffused in the entire image plane, more than 84% of the sampling steps were well-oriented towards the nearest delta measure. As demonstrated in (a), the scanning trajectory controlled by the deviation factor \( n^m_n \) is well concentrated on the saliency patterns; in this case, the scanning mechanism (9) restored 79% of the maximum value of the distribution...
Saliency-Based Boundary Object Detection in Naturally Complex Scenes

Figure 6: Multi-Fractal Coding

\[ \varphi_\sigma (\omega | \Pi) \] through 200 iterations; this restoration rate is improved more than 12% relative to random scan.

In figure 5, the local maxima of the distribution \( \varphi_\sigma (\omega | \Pi) \) with respect to the estimate \( \hat{\pi}_i \) are marked by circles on the \( \hat{\psi}_\omega \)-distribution and associated palette view. By using this palette-probability association, we can design an efficient focusing mechanism in practical scene images. The focusing dynamics (9) yields a nondeterministic trajectory scanning the entire distribution of chromatic diversity. Along the trajectory, the vision system generate a sequence of saliency patterns to be labeled with respect to the \( \hat{\pi}_i \) palette \( \hat{\Pi} \). The representative pixel of each palette color is indicated on the scene image \( f_{\text{RGB}} \) with distributed emphasis \( \hat{\psi}_\omega \). Thus, we have a cluster of invariant measures to be articulated into a multi-fractal system [1].

5 Multi-Fractal Articulation

Through the boundary object detection, we have an \textit{in-situ} instantiation of the generic ground-object structure. To this end, we apply fractal articulation [14] to \( \varphi_\sigma (\omega | \Pi) \)-distribution; the problem is to design a set of contraction mappings \( \mu_i : \Omega \mapsto \Omega \) of the following form:

\[
\mu_i(\omega) = \frac{1}{2} [\omega + \omega_{\mu_i}],
\]

(10)
generating the attractor \( \Xi \) satisfying the self-similarity

\[
\Xi = \bigcup_{\mu_i \in \nu} \mu_i(\Xi),
\]

within a connected region of the \( \hat{\psi}_\omega f_{\text{RGB}} \) image. Equation (10) implies that the attractor \( \Xi \) is controlled by the fixed point \( \omega_{\mu_i} \) to be allocated on the saliency pattern.

As demonstrated in Fig. 5, we can select a saliency pattern \( \Lambda \) on the \( \hat{\psi}_\omega f_{\text{RGB}} \)-image. By applying the fractal articulation process to the pattern \( \Lambda \), we have a fractal model \( \nu \) for labeling landmark image in complex background as shown in Fig. 6 where associated attractor \( \Xi \) is generated on the object image selected as \( \Lambda \). Through the fractal articulation process, the expansion of the attractor is restricted in terms of the chromatic diversity of the samples \( s \) on the set \( \Xi \). This implies that the saliency pattern should be supported by a chromatically connected area of an object.

Furthermore, the fractal model is verified to be consistent with the ground-object structure as shown in Fig. 7; the distribution of multi-scale information is extracted from the scene image (Fig. 1) as displayed in this figure. The fractal model \( \nu \) is applied to the local maxima of the multi-scale image to yield a sufficient invariant links with respect to \( \nu \). This implies that there exists a perpendicular plane supporting the fractal attractor within the perspective of the roadway area. Thus, the fractal model \( \nu \) is perceptually consistent with the perspective of the scene; the chromatically connected area \( \Xi \) specified by \( \nu \) can be substantiated by a 3D object as well as the the connected open space in Fig. 1. As the result, a maneuvering context is computationally organized in terms of multi-fractal system for regenerating the saliency pattern and the roadway area; figure 8 illustrates that the detected sign pattern, to be associated with the post office via the human’s understanding, is localized in the left boundary of the open space.
Figure 7: Finite Invariance Test on Multi-Scale Information

Figure 8: Contextual Visualization (post office)

Table 1: Complexity Reduction via $\tilde{\psi}_\omega$-filtering

| scene                        | $dS_G$   | $dS_{\tilde{G}}$ | $||\hat{\Pi}||$ |
|------------------------------|----------|------------------|-----------------|
| shopping street              | 0.131549 | < 1.597940       | 4               |
| post office                  | 0.200029 | < 1.349640       | 5               |
| street view at night         | 0.148731 | < 2.718602       | 3               |
| industrial park              | 0.103450 | < 2.198932       | 3               |

The multi-fractal articulation scheme has been applied to the detection of landmark objects in various naturally complex scenes as shown in Fig. 9 and 10. In these experiments, saliency patterns are extracted with respect to shifted primary $\hat{\Pi} = \{ \hat{\pi}_r, \hat{\pi}_g, \hat{\pi}_b \}$ as indicated in (a); the chromatic- and geometric-connectedness of the associated fractal codes were verified through the perceptual consistency analysis and visualized within the context of the ground-object structure as illustrated in (b). These experimental results demonstrate that the simulated saccade scheme is effective for activating the fractal coding in even ill-conditioned scenes where landmark objects are observed as non-dominant patterns; a warning color of vehicles sometimes yields minor image in attractive distractions as displayed in Fig. 9; the post is rather ‘low-keyed’ object in a night view as shown in Fig. 10.

6 Discussions

The essential part of the $\tilde{\psi}_\omega$-filtering is formalized by one dimensional successive manipulation (4)+(5)+(6) of finite features $\tilde{\Gamma}^f$. Noting that the features are extracted via parallel distributed system (3), we can implement the object detection scheme as an efficient computation mechanism equipped with locally parallel processors.

In the current implementation, the fractal modeling process is activated in response to human’s selection of a saliency pattern. The significance of the $\tilde{\psi}_\omega$-filtering in the reduction of the mental load arising in such a computational task was estimated by using various types of scene images. The results
are summarized in Table 1 where the complexity of the pattern selection is evaluated in terms of relative entropy $dS(\cdot) = S_\emptyset - S(\cdot)$; $S_\emptyset$ and $S_G$ designate the Shannon’s entropy with respect to the uniform distribution and the gray level distribution $f_\omega$, respectively; $S_\pi$ stands for the following entropy

\[
S_\pi = -\int_\Omega \hat{\psi}_\omega \log (\hat{\psi}_\omega) d\omega.
\]

The $\hat{\psi}_\omega$-filter concentrate the information distributed in the image plane to reduce the length of decision steps to $1/e^{dS_\pi}$ of binary decision in the image plane. As shown in Table 1, the reduction rates for the scenes indicated in Figs. 1, 9 and 10 amount to $1/3 \sim 1/15$; the comparison between $dS_G$ and $dS_\pi$ implies that the complexity of $\hat{\psi}_\omega$ based pattern selection is reduced to $1/3 \sim 1/12$ of decision making based on unstructured brightness distribution $f_\omega$. The cue to the human’s selection is cooperatively generated via simulated saccade (9) on parallel distributed system (8).

7 Concluding Remarks

A stochastic scanning scheme was introduced for capturing saliency patterns in naturally complex scenes. Via multi-fractal articulation, the saliency patterns are localized within the perspective of the scene. Introduced $\hat{\psi}_\omega$-indexing significantly reduces the complexity of pattern selection.

References

Saliency-Based Boundary Object Detection in Naturally Complex Scenes

Figure 10: Boundary Object Detection (street view at night)