Multi-Fractal Articulation of Environmental Saliency Arising 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 new framework is presented for saliency based object detection in naturally complex scenes. In this framework, the random distribution of the scale and chromatic aspects of image complexity are exploited for the multi-fractal articulation of scene image into the ground-object structure and object images, respectively. Due to the intrinsic coherence of the scale and chromatic aspects, detected object images are well organized within the ground-object structure to indicate various types of maneuvering context arising in the scenes.

Keywords: Environmental Saliency; Scale-Chromatic Complexity; Multi-Fractal Articulation

Table of Contents

1 Introductory Remarks  2
2 Locally Gaussian Palette  3
3 Chromatic Complexity Generator  4
4 Multi-Fractal Articulation  7
5 Perceptual Equivalence  9
6 Concluding Remarks  10

1 Introductory Remarks

Despite infinite diversity of appearance, natural scenes exhibit environment specific landmarks to be identified within individual intention of viewers. To control the focus to such a landmark object, perception processes should gather randomly distributed image features and apply ‘feature integration’ schemes to generate ‘visual saliency’ associated with the complex scene [5]. Due to the redundancy of natural scene relative to decision makings by the viewers, computational feature integration processes easily fall into combinatorial explosion.

At the practical implementation of the basic capability of early vision, thus, it is pertinent to transform the diversity of the image features into the space of intermediate representation called ‘environmental saliency’: a distributed representation of universal rules nondeterministically governing physical-geometric structure of the natural scenes. For instance, the scale-chromatic randomness has been matched with the universal rules to organize the environmental saliency as a transferable description of the ground-object structure. Such a randomness-based representation provides the theoretical basis of an anticipative road following system illustrated in Fig. 1; the environmental saliency extracted in a scene image is transferred to a cut of satellite image to extend a visible segment of the roadway pattern beyond the physical-geometric perspective.

Let \((\hat{\omega}_t, \hat{\theta}_t)\) be the estimate of the current position and direction in the satellite image with local roadway segment \(\hat{v}_t\). Suppose that the local segment is extended to generate a chain \(\bar{\mathcal{V}}\) to a possible destination in the bird’s eye view. By this symbolic segmentation, geometric complexity of the local terrain is reduced to a graph on which we can constrain the diversity of future scenes prior to physical access. Adding to it, we can associate generic structure of ‘naturally designed scenes’ in the multi-viewpoint images spanning the temporal-spatio discrepancy including the distribution of moving objects.

In this paper, we introduce a new framework for identifying vehicle specific scenes based on the transferred generic structure. We assume that the randomness distributed in the bird’s eye view can be downloaded to the on-vehicle vision systems as a stationary description of the scene structure. Following empirical knowledge of ecological optics [3] combined with recent advancements in machine perception [7] and emotional perception [4], we can expect that the generic structure of maneuverable scenes is described in terms of a set of fractal codes specifying the expansion of horizontal plane and the aggregation of boundary objects. Noting this, the randomness of the scale information is extracted as a version of latent images in the gray level distribution to estimate the fractal code spanning a connected open space [9]; downloaded segment \(\hat{v}_t\) is mapped into the scene image (Fig. 2) to design a fractal code bounded by breakdown pixels (Fig. 3). To cooperate with human’s inherent perception, the breakdown pattern
should be articulated into a system of fractal codes to be associated with the landmark objects. As the cue to the multi-fractal coding, in what follows, another version of latent images, random distribution of chromatic complexity, is extracted as the observables of not-yet-identified landmark objects.

2 Locally Gaussian Palette

As the results of the evolution in the really existing world, human’s vision system is equipped with a sophisticated information processing mechanism for understanding the scenes thronged with friendly or undesirable neighbors [12]. Through the iteration of selection pressure, simultaneously, the neighbors have developed various types of micro structures for the luminescence of emotional expressions. As intentional participants of the co-evolution process without the ‘intelligent designer’, human’s vision seems to be able to control the attention within the surroundings [11], on the premise that imminent decision making should be evoked by early perception of unstructured ambient light [3] and universal preference to a class of fractal patterns [4]. The existence of such pre-computational schemes implies that the attention control process can be anticipatively adapted to the co-evolutionarily designed scene through the multi-viewpoint observation.

Following the anatomy of early vision system, the neuronal system is activated by two aspects of incoming light: brightness and spectrum [2]; the system is sensitive to brightness distribution and subtle to spectrum shift. This implies that even weak light is sufficient to generate the following scale information \( \hat{\sigma}_\omega \) in the early scene analysis:

\[
\hat{\sigma}_\omega \sim \sqrt{\frac{2f_\omega}{|\Delta f_\omega|}},
\]

where \( f_\omega \) denotes the brightness of the incoming light at the pixel \( \omega \) in a image plane \( \Omega \). By matching the scale information \( \hat{\sigma}_\omega \) with a generic perspective model, we have a probability distribution evaluating
the expansion of a connected open space as shown in Fig. 3.

Let R, G, B be three primaries, and suppose that the light is identified with a linear combination of the primaries by humans. This implies that the information conveyed by the spectral distribution be described as follows: \( f_{\text{RGB}} = \begin{bmatrix} f_R^\omega \\ f_G^\omega \\ f_B^\omega \end{bmatrix} \), where \( f_\omega^\) designates subjective weight of the primary \( \omega \). Define \( \phi_\omega = f_{\text{RGB}} / |f_{\text{RGB}}| \). By identifying the totality of the chromatic information \( \phi_\omega \) with the positive part of a unit spherical surface, we can induce the following measure [8]:

\[
g_\alpha (\phi_i | \phi_j) = \frac{1}{2 \pi \alpha} \exp \left[ -\frac{|\phi_i - \phi_j|^2}{2 \alpha} \right]. \tag{2}
\]

For sufficiently small \( |\phi_i - \phi_j| \), the measure \( g_\alpha (\phi_i | \phi_j) \) approximates the Gaussian distribution on local tangential space at \( \phi_j \). Noticing the following constraint

\[
|\phi_\omega|^2 = \sum_{\text{RGB}} \left( \frac{f_\omega^\) \text{RGB}}{|f_{\text{RGB}}|} \right)^2 = 1,
\]

we have the following index for evaluating a kind of coloring saliency:

\[
\psi_{\text{RGB}} = \exp \left[ -\mathcal{H}_{\text{RGB}}^\right], \tag{3}
\]

\[
\mathcal{H}_{\text{RGB}}^\right = -2 \sum_{\text{RGB}} \left( \frac{f_\omega^\) \text{RGB}}{|f_{\text{RGB}}|} \right)^2 \log \left( \frac{f_\omega^\) \text{RGB}}{|f_{\text{RGB}}|} \right).
\]

In this indexing, the substantial process in the retina system generates the ‘square root of capturing probability’ \( \sim \phi_\omega \) to yield the random distribution of Shannon’s entropy \( \mathcal{H}_{\text{RGB}} \); the complexity measure is transformed to a saliency probability \( \psi_{\text{RGB}} \) via vitals specific ‘neg-entropy’ generation. By using the saliency indexing (3), we can detect landmark objects in well-structured scene as illustrated in Fig. 4 where the chromatic complexity (=simplicity, in this case) of the ‘block world’ is represented in a conventional color space (b) for ‘matting’ the saliency patterns in noisy background (c). The saliency indexing can be applied to naturally complex scene to visualize the distribution \( f_{\text{RGB}}^\text{RGB}, \psi_{\text{RGB}} \) as shown in Fig 5 where the complexity of the scene indicated in Fig. 2 is reduced via the \( \psi_{\text{RGB}} \)-filtering to visualize partial distribution of landmark objects. As demonstrated in Figs. 4 and 2, the saliency probability is sensitive to the objects painted by preassigned primaries. To correct such pigmentation-level bias, we need the adaptation of the primary system to the global complexity arising in the entire observed scene.

### 3 Chromatic Complexity Generator

Figure 4 implies the existence of a chromatic complexity generator; the diversity of the chromatic information is expanded towards the set of the primaries. Noticing this, consider an inverse problem; select a set of fixed points to regenerate the chromatic diversity as a fractal attractor in the color space. For such a global analysis of chromatic diversity, let the chromatic complexity index \( \phi_\omega \) be identified with the following planar representation of the color space:

\[
\Gamma \ni \gamma = \epsilon_\text{RGB} \phi_\omega, \quad \epsilon_{\text{RGB}} = \begin{bmatrix} \epsilon_R \\ \epsilon_G \\ \epsilon_B \end{bmatrix}, \quad \epsilon^(\) = \begin{bmatrix} \cos \theta (\) \\ \sin \theta (\) \end{bmatrix}^T, \tag{4}
\]

with \( \) a priori \( \theta_R = \pi/2, \theta_G(B) = \theta_R (\) = \) \( \pi/3 \). The diversity of the incoming light \( f_{\text{RGB}} \) is represented in the color space \( \Gamma \) through the linear transform (4). The representation in \( \Gamma \) is restored through the following procedure:

\[
\tilde{\phi}_\omega = \tilde{\phi}_\gamma + \overline{\phi}_\omega 1_{\text{RGB}}, \quad 1_{\text{RGB}} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T, \tag{5}
\]

where \( \tilde{\phi}_\gamma = \frac{2}{3} \left( \epsilon_{\text{RGB}} \right)^\gamma \). In (5), \( \overline{\phi}_\omega \) designates a nominal brightness level given as the solution to the following

\[
3\overline{\phi}_\gamma^2 + 2\overline{\phi}_\gamma 1_{\text{RGB}} \cdot \overline{\phi}_\omega + \left| \overline{\phi}_\gamma \right|^2 = 1.
\]

Thus, the representation in \( \Gamma \) can enhance the spectrum shift independent of the variation of the brightness level in observed imagery.
Let a set of samples $\mathbf{s}$ be collected in a scene image to generate the following field on $\Gamma$:

$$\frac{\partial \varphi}{\partial t} = \frac{1}{2} \Delta \varphi + \rho [\chi - \varphi],$$

(6)

where $\chi$ denotes the aggregation of Dirac’s delta measure distributed on the set $\{ \gamma(\phi) \mid \phi \in \mathbf{s} \}$ and $\rho$ is adjusted in terms of the size of the primary set. By identifying the distribution $\chi$ with an observation of the invariant measure associated with not-yet-identified fractal attractor in the color space $\Gamma$, we can specify a set of control parameter $\hat{\Pi} = \{ \hat{\pi} \}$ for regenerating the diversity of the chromatic information in the color space $\Gamma$. To this end, first, a possible fixed point is located as an initial set of the control point, $\hat{\Gamma}_0$, on the Laplacian-Gaussian boundary $\partial \varphi$. The initial set is expanded via the following successive scheme:

$$\hat{\Gamma}_{t+1} = \hat{\Gamma}_t \cup d\hat{\Gamma}_t,$$

(7)
Figure 5: $f_{\text{RGB}}^{\text{RGB}}$-image (naturally complex scene)

where

$$d\tilde{\Gamma}_f = \left\{ \tilde{\gamma}_f \mid \forall \tilde{\gamma} : \tilde{\eta} \left( \tilde{\gamma}, \tilde{\Gamma}_f \right) \geq \tilde{\eta} \left( \tilde{\gamma}, \tilde{\Gamma}_f \right) \right\},$$

and $\tilde{\eta} (\gamma, A) = \min_{\lambda \in A} |\gamma - \lambda|$. Next, the set of vertices $\tilde{\Gamma} = \left\{ \tilde{\gamma}_k \in \tilde{\Gamma}_f \right\}$ satisfying the following conditions are selected to minimize the mapping set for regenerating the distribution $\chi_s$:

$$\forall m, k : \theta_{mk} - \theta_{nk} < \pi,$$

$$(\tilde{\gamma}_k)^{(i)} - \hat{\gamma}_k = |(\tilde{\gamma}_k)^{(i)} - \hat{\gamma}_k| \exp^{i(\theta_{i,k} + \theta_k)},$$

$$\hat{\gamma}_k = |\hat{\gamma}_k| \exp^{i\theta_k}.$$

Finally, the distribution of $\tilde{\Gamma}$ is expanded along the following repulsive force:

$$d\hat{\gamma}_k = \sum_{\tilde{\gamma}_j \in \tilde{\Gamma}} (\hat{\gamma}_k - \hat{\gamma}_j) g_\alpha (\phi_k | \phi_j),$$

within the possible coloring circle $|\hat{\gamma}_k| \leq 1$. Following the repulsive force, the vertices $\{ \hat{\gamma}_k \}$ are separated each other to yield a set of as-is primaries $\hat{\Pi}$. The scheme (7) combined with (8) yields a set of fixed points to be associated with a set of contraction mapping for regenerating the distribution $\chi_s$ in the color space $\Gamma$. By adding the dynamics (9), we have a meta-process: a generator of the fractal dynamics controlled by the as-is primaries $\hat{\Pi}$.

The implication of the chromatic complexity generator is demonstrated in Fig. 6 where the distribution of the samples $s$ is indicated in (a); in (b) the associated field $\varphi_{\rho}(\gamma | s)$ is displayed with the distribution of possible fixed points $\tilde{\Gamma}_f$; in this figure, the vertices $\{ \hat{\gamma}_k \}$ are selected and separated to yield the as-is primaries $\hat{\Pi}$.

The effectiveness of the as-is primaries is illustrated in Fig. 7 where the landmark objects are detected based on the as-is saliency index given by

$$\hat{\psi}_\omega = \exp \left[ -\hat{\mathcal{F}}_\omega \right],$$

$$\hat{\mathcal{F}}_\omega = -\sum_{\hat{\pi}_i \in \hat{\Pi}} p (\gamma_\omega | \hat{\pi}_i) \log p (\gamma_\omega | \hat{\pi}_i),$$

with the following distributed representation of the primary selection complexity:

$$p (\gamma_\omega | \hat{\pi}_i) = \frac{g_\alpha (\gamma_\omega | \hat{\pi}_i)}{\sum_{\hat{\pi}_i \in \hat{\Pi}} g_\alpha (\gamma_\omega | \hat{\pi}_i)}.$$
Multi-Fractal Articulation of Environmental Saliency Arising in Naturally Complex Scenes

4 Multi-Fractal Articulation

To detect a set of landmark objects in the $f_{\omega}^{\hat{\phi}_\omega}$-image, let the saliency pattern

$$\mathcal{O} = \left\{ \omega \in \Omega \mid f_\omega \sim \hat{\pi}_1 \in \tilde{\Pi} \right\}$$

with respect to an as-is primary $\hat{\pi}_1 \in \tilde{\Pi}$ be articulated into fractal attractors generated by contraction mappings $\mu_i : \Omega \mapsto \Omega$ of the following form:

$$\mu_i(\omega) = \frac{1}{2} [\omega + \omega_{\mu_i}^{\mu_i}], \quad i = 1, 2, \ldots, ||\nu||,$$

where $\omega_{\mu_i}^{\mu_i}$ is the fixed point of the mapping $\mu_i$. Noticing that the fixed points should be located the Laplacian-Gaussian boundary $\partial^\theta \mathcal{O}$, we can apply the following allocation scheme (7) to the saliency
pattern, as well:

\[
\begin{align*}
\Omega_{t+1}^{f} &= \Omega_t^{f} \cup d\Omega_t^{f}, \\
\partial\Omega_t^{f} &= \left\{ \partial \omega^* \in \partial^\circ \mathcal{O} \, | \, \eta \left( \partial \omega^*, \Omega_t^{f} \right) \geq \eta \left( \partial \omega, \Omega_t^{f} \right), \partial \omega \in \mathcal{O}_t \right\}, \\
\eta \left( \partial \omega^*, \Omega_t^{f} \right) &= \eta \left( \partial \omega, \Omega_t^{f} \right).
\end{align*}
\]  

(13)

Consider the articulation of the unstructured objects \( \mathcal{O} \) detected as shown in Fig. 8 in terms of separated fractal attractors. As the representation of the generator of one of the attractors, let \( \nu_t \) be the mapping set designed by the on-going allocation \( \Omega_t^{f} \). The connectedness of the fractal coding can be evaluated in terms of the following \( g_\alpha \)-measure:

\[
g_\alpha \left( f_t | s \right) = \min_{f \in \mathcal{H}} g_\alpha \left( \phi \left( f^{RGB} \right) | s \right),
\]

(14)

where \( \mathcal{H} \) denotes the attractor generated by \( \nu_t \). Define \( d\nu_t^{*} \) as the mapping set designed based on the set \( d\Omega_t^{f} \cup \Omega_t^{*} \) where \( \Omega_t^{*} \) designate the origin of the minimal span satisfying the following condition

\[
\Omega_t^{*} = \left\{ \omega^* \in \Omega_t^{f} \, | \, \eta \left( \omega^*, d\Omega_t^{f} \right) \leq \eta \left( \omega^*, d\Omega_t^{f} \right) \right\},
\]

(15)

for arbitrary \( \omega^f \in \Omega_t^{f} \). By using the fractal attractor \( d\mathcal{C}_t^{*} \) associated with the fractal code \( d\nu_t^{*} \), we have the halting condition for the expansion process (13) as follows:

\[
g_\alpha \left( d\nu_t^{*} | s \right) < g_\alpha \left( f_t | s \right),
\]

(16)

where \( d\nu_t^{*} \) and \( f_t \) denote the fractal sampling associated with \( d\nu_t^{*} \) and \( \nu_t \), respectively. By applying the halting mechanism (16) to the fixed point estimates (Fig. 8), the unstructured scene image is articulated into a system of ‘saliency pattern’ wrapping landmark objects as shown in Fig. 9.
5 Perceptual Equivalence

As shown in figures 2, 3 and 5, the scale and chromatic aspects of the environmental saliency are intrinsically consistent to yield a system of the fractal models: one is the ground model for separating a connected open space from boundary objects; the other is saliency pattern model for concentrating computational resources to the detection of landmark objects. In the multi-fractal articulation process, the connectedness criterion (14) combined with the halt condition (16) is effective to separate a landmark object in complex background including distractive objects painted by similar colors as shown in Fig. 10. Furthermore, the fractal code is consistent with the ground-object structure identified through scale complexity analysis; the fractal model based on the chromatic complexity analysis can yield an invariant features on the scale evaluation (1) as demonstrated in figure 11. This implies that the environmental saliency can be exploited as the basis for the contextual visualization of naturally complex scenes. For instance, figure 12 displays a version of the maneuvering context: the left side of boundary objects are articulated to confine an area of a post office; a post should be separated from a sign board to follow the
open space. The multi-fractal articulation algorithm has been applied to various natural scenes to detect landmark objects to notice for pedestrians and vehicles going through the roadways [10]. These results imply the computability of an emotional expression required for an intelligent vehicle [1] including the values and/or special taste, the expression of space and movement, and feeling of fear.

6 Concluding Remarks

A multi-fractal coding was applied for saliency based object detection in naturally complex scenes. By identifying the chromatic diversity with a fractal attractor in a color space, an as-is primary system is selected to discriminate the image of landmark objects in noisy background. Supported by the consistency with ground-object structure based on the scale complexity, the multi-fractal model of the object images can be exploited to indicate various types of maneuvering context arising in the scenes.

References