Perceptual Equivalence of Scale and Chromatic Aspects of Environmental Saliency Arising in Naturally Complex Scenes

Kohji Kamejima

November 7, 2015

Abstract: A new framework is presented for saliency based analysis of naturally complex scenes. In this framework, the fluctuation of local scale shift and the diversity of subtle chromatic scattering are extracted as the environmental saliency: viewer specific visualization of sign patterns to be associated with landmark objects. Being guided by the probability distribution of the scale shift, the scene image is partitioned into a fractal attractor spanning the roadway area and associated distribution of boundary objects. The chromatic diversity of the scene image is evaluated within a probabilistic color space as the support of a saliency index. The saliency index is applied to the estimate of the invariant measure to capture the sign patterns as fractal attractors within the boundary distribution. Thus, the scale-chromatic aspects of the environmental saliency are articulated into a system of fractal attractors of perceptually equivalence to jointly identify ground-object structure. Through experimental studies, it has been demonstrated that the complexity of decision steps in the detection of landmark objects can be significantly reduced by the saliency based articulation. Adding to it, the perceptual equivalence between the scale- and chromatic-aspects of the environmental saliency is testable via the multi-fractal articulation; to this end, it is sufficient to match the fractal attractors designed in 2D chromatic aspect with 3D landmarks through the projection into 2.5D scale aspect. By this perceptual equivalence, a self-reflective mechanism is induced in the two-aspect representation of the environmental saliency.

Keywords: environmental saliency, naturally complex scene, scale-chromatic complexity, multi-fractal articulation.

Table of Contents

1 Introductory Remarks 2
2 Statement of Problem 2
3 Randomness-Based Approach 3
   3.1 Scale Randomness for Ground-Object Separation 3
   3.2 Chromatic Complexity for Palette Generation 5
4 Saliency Pattern Filtering in Chromatic Aspect 6
5 Multi-Fractal Coding 7
6 Perceptual Equivalence 9
7 Experiment 10
8 Discussions 13
9 Concluding Remarks 15
Appendix 16
A Laplacian-Gaussian Analysis 16

2Professor emeritus, Osaka Institute of Technology, kamejima@k.zaq.jp
1 Introductory Remarks

By invoking recent development of various space systems, we can extend the scope of decision making for geometric planning [1] as well as autonomous vehicle guidance [2] beyond physical perspective in complex roadway scene. For instance, the interactive mapping system developed for field automation [3] can be applied to cooperative maneuvering within the satellite-roadway-vehicle network [4] as shown in Fig. 1; a multitude of on-vehicle perceptual segments are connected with the roadway segment in which a cut of the satellite image is downloaded as a common basis of local terrains; the scenes surrounding the vehicles are localized on the satellite image by the global positioning system (GPS). The scheme has been applied to the implementation of anticipative road following where the roadway model identified in the scene image is transferred to the satellite image to extend the roadway pattern to possible destinations prior to physical access [5].

Through the anticipative road following process, the vehicles simulate a transition of local terrains along a ‘future’ trajectory. Within the consistency of GPS localization, the future trajectory provides effective information to activate the geometric planning process and/or to program the autonomous vehicle guidance to cope with various complex situations. However, due to the essential uncertainty in the distribution of temporal and/or moving objects, it is not practical to utilize the satellite image as the reference of the vehicle control for the subsequent maneuvering process. As a result, the scope of on-vehicle vision systems is still restricted within a physical-geometric perspective.

To expand the decision space through the satellite-vehicle-roadway network, in this paper, we introduce a new framework to identify vehicle specific scenes spanning over the tempo-spatio discrepancy. The first problem to be tackled is the preset of the on-vehicle vision system based on the structural representation of not-yet-visible scenes along the future trajectory. Invoking the knowledge established in the ecological optics [6], we assume that the generic structure of the scenes is substantiated via a two-phase process: (i) recognition of an object free space confined by unstructured boundary and (ii) articulation of the boundary distribution into landmark objects.

2 Statement of Problem

To activate the two-phase perception process, the vision system should be grounded on the surroundings through various aspects of the observed image. Despite the diversity of appearances, natural scenes are repleted by environment specific landmarks sufficient to activate individual decision making towards the viewer specific goal. To control the attention within such a complex scene, it has been pointed out that perception processes apply ‘feature integration’ schemes [7] to ‘visual saliency’ randomly distributed in noisy images [8]. Through the investigation of the early vision, it has been revealed that the multi-scale Gaussian filter is implemented by inherent parallel computation mechanism [9], [10] to extract 2.5 dimensional cue to object detection [11], [12]. Such a ‘fast’ scale information has been applied to early structural analysis including ground-object separation of complex scene imagery [13]. Simultaneously, the same observed image is accepted by a more sophisticated system where spectral diversity of incoming light is factorized within a system of primary colors [14]. Despite the loss of the depth information, saliency patterns extracted as ‘matted’ images [15] can be associated with 3D objects and localized in the perspective of natural scenes.

In the conventional integration scheme, observed images are assumed to be structured in terms of ‘sensible saliency’ including key-points of well-organized objects and chunks of attractive colors; such
Perceptual Equivalence of Scale and Chromatic Aspects of Environmental Saliency Arising in Naturally Complex Scenes

Environmental Saliency

Figure 2: Schematics of preestablished restriction.

sensible saliency can be integrated via top down process for the selection of visual attention in complex scenes [16]. In many natural scenes, the sensible features should support a multitude of viewer specific decision makings. However, the maintenance of the consistency is still an open problem in the feature representation spanning unpredicted discrepancy between viewer specific photographing conditions. Adding to it, practical scene images exhibit superfluous saliency patterns for individual decision makers. In such a situation, the contextual analysis of a priori fixed 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 a working hypothesis, assume that possible features to be focused arise from the real world of the two aspects: ‘physical-geometric environment’ and ‘co-evolution environment’, as illustrated in Fig. 2; landmark objects in the physical-geometric environment are suffering from the iteration of ‘deformation and/or degeneration processes’ to exhibit as-is images (as indicated by down arrow); the co-evolution environment is resulted from the collaboration of intentional and/or contingent participants. The two aspect model is endowed with the following self-organization mechanism; to maintain a multitude of mutual coexistence, the participants substantiate warning ‘sign’ and/or informative ‘design’ as landmark objects within the physical-geometric environment (left arrows); simultaneously, the participants have competitively developed sophisticated sensing devices including eyes, in particular, to accept the sign/design conveyed by the degenerated version of the landmarks (up arrows).

Based on the working hypothesis, a multitude of perception processes within the framework of ecological optics are required to restrict the focus on an invariant structure called ‘environmental saliency’ underlying scene images; landmark objects to be focused, from the other point of view, should present the entire participants with ‘readable image’ of the environmental saliency. Noticing that the fluctuation of local scale shift and the diversity of subtle chromatic scattering arising in naturally complex scenes can jointly be matched with the environmental saliency spanning significant tempo-spatio discrepancy [4], the problem to be considered is stated as follows: how to extend the scope of the ‘bottomup’ image processing to randomness-based observation of the environmental saliency?

3 Randomness-Based Approach

Within the schematics of reestablished restriction indicated in Fig. 2, the environmental saliency is matched with a set of generic rules to identify the ground-object structure in vehicle specific views. By invoking empirical knowledge on neural sensitivity to fractal patterns [17], [18], combined with emotional preference to fractal skylines [19] and finite programmability of a class of self-similarity imaging processes [20], in what follows, the range of environmental saliency is supposed to be articulated into a system of fractal attractors spanning the expansion of a horizontal plane and an aggregation of boundary objects.

As a theoretical basis for viewer independent scene analysis, in this section, some preliminary results are summarized for the detection and organization of randomness distributed in the scene images.

3.1 Scale Randomness for Ground-Object Separation

Let a segment of the roadway be anticipatively identified in the satellite image [5] and suppose that the scene image is localized at the segment. Such a segment can be projected into the image plane \( \Omega = \{ \omega = (x, y) \mid x \in X, y \in Y \} \) as indicated in Fig. 3 (a) where expansion and depth of the open space
are indexed along horizontal and vertical axes of the image plane, respectively; the projection of the roadway segment is visualized as a vector $\vec{\Omega}^d$ with the projected origin $d_0 = (x_0, y_0)$ at the base of the image plane towards an a priori estimate of the vanishing point at $d_\infty = (x_\infty, y_\infty)$. In natural scenes, the perception of a connected open space is evoked by the convergence of the scale information towards the vanishing point. This generic rule is formalized by the following perceptual linearity

$$\bar{\sigma}_d = \max \left( \kappa \sigma_0 \left| \frac{d_\infty - d}{d_\infty - d_0} \right|, \sigma_0 \right),$$

with respect to the depth parameter $d$ along the vector $\vec{\Omega}^d$. The nominal representation (1) can be extended in unstructured scene images where the perceptual linearity is diffused in underlying noisy patterns with spectrum spanning the interval $\Sigma_0^\infty = [\sigma_0, \kappa \sigma_0]$. By assigning the lower limit of noise spectrum to the scale at the vanishing point $\sigma_0$, thus, we have a robust representation of the perceptual linearity for the identification of the generic structure in complex scene imagery. In various scene images, we can fix the range of relative scale shift by $\kappa = 2$.

Consider the brightness of the incoming light $f_\omega$ observed at the pixel $\omega \in \Omega$. Noting that many imaging devices including human eye detect the brightness distribution through a system of the Gaussian point spread functions, we can estimate the upper bound of the scale information $\hat{\sigma}_\omega$, $\omega \in \Omega$, by

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

where $\Delta f_\omega$ denotes the Laplacian of $f_\omega$ (see Appendix A). This implies that the consistency of the generic model (1) can be evaluated by the following probability density function

$$g_d(\omega|\sigma_0) = \frac{1}{\sqrt{2\pi \sigma_d^2}} \exp \left[ -\frac{\left| \hat{\sigma}_\omega - \sigma_d \right|^2}{2\sigma_d^2} \right],$$

where the fluctuation of $\hat{\sigma}_\omega$ is simulated by Gaussian random variable taking its value in the interval $[0, 2\sigma_d]$ with probability 0.7. In natural scene, the perceptual linearity suffers from breakdown due to occlusion, e.g., by something perpendicular to the open space. Let $\omega_i = (x, y)$, $y \in Y$ be a ground pixel and define the vertical chain of pixels $\langle \omega_i \rangle$ with size $\| \langle \omega_i \rangle \|$ by

$$\langle \omega_i \rangle = \{ \omega_1, \ldots, \omega_k, \ldots \}, \quad k = 1, 2, \ldots, \| \langle \omega_i \rangle \| - 1,$$
where $\omega^\uparrow = (x, (\cdot)), (\cdot) > y$ stands for the upward extension. Then the probability for the pixel $\omega = (x, y)$ to be a breakdown is indexed by

$$
g_b(\omega|\sigma_0) = \begin{cases} \max_{(\omega_i) \in B_{\omega_0}^{\sigma_0}} \prod_{\omega \in \omega_i} G_b(\omega^\uparrow|\delta_{\omega_i}, \sigma_0); & \text{for } B_{\omega_0}^{\sigma_0} \neq \emptyset, \\
0; & \text{otherwise,} \end{cases} \tag{4}
$$

where $B_{\omega}^{\sigma_0} = \{ (\omega_i) \mid (\omega_i) \cap \{ \omega \} \neq \emptyset, \| (\omega_i) \| \geq \sigma_0 \}$ and

$$
G_b(\omega^\uparrow|\delta_{\omega_i}, \sigma_0) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_{\omega_i}^2}} \exp \left[ -\frac{|\delta_{\omega_i} - \delta_{\omega}^\downarrow|^2}{2\sigma_{\omega_i}^2} \right]; & \text{for } \delta_{\omega_i} \in \Omega_{\omega_0}^\infty \\
0; & \text{otherwise}. \end{cases}
$$

Hence, the support of the perceptual linearity is extended to the distribution $g_b(\omega|\sigma_0)$ with the restriction $g_b(\omega|\sigma_0)$ to visualize the ground-object structure as shown in Fig. 3(b). In this figure, a connected open space within the roadway area is represented by fractal attractor $\Xi_d$ satisfying the following self-similarity:

$$
\Xi_d = \bigcup_{\mu_i \in \nu_d} \mu_i(\Xi_d), \tag{5a}
$$

where $\nu_d$ is a set of contraction mappings of the following form:

$$
\mu_i(\omega) = \frac{1}{2} [\omega + \omega_i^f]. \tag{5b}
$$

The range of the attractor $\Xi_d$ is successively generated under the control of the fixed point $\omega_i^f$; the ranges of mappings $\mu_i \in \nu_d$ are distinguished by specific colors. By identifying the $g_d(\omega|\sigma_0)$-distribution with an invariant measure associated with the self-similarity (5), the consistency between the fractal code and the generic model (1) is verified through the generation of a closed graph on the space; simultaneously, the expansion of breakdown pixels is evaluated in terms of the distribution $g_b(\omega|\sigma_0)$ and visualized to focus on not-yet-identified objects. As shown in this figure, we have a stochastic description $\{ g_d(\omega|\sigma_0), g_b(\omega|\sigma_0) \}$ of the ground-object structure arising in the naturally complex scenes. Thus, the random fluctuation of the scale variation $\delta_{\omega}$ is available as an access path to the scale aspect of environmental saliency.

### 3.2 Chromatic Complexity for Palette Generation

The perception of chromatic diversity is essentially mental processes; for most human observers only three primaries are required to match a test light [21]. To precisely manipulate the information conveyed by chromatic diversity, let the coloring of the reflected light be identified within the nonnegative subset of 3D Euclid space $R^3_\omega$:

$$
f^{\text{RGB}}_\omega = [R_\omega, G_\omega, B_\omega]^T, \quad 0 \leq R_\omega, G_\omega, B_\omega \leq 1,
$$

where $R_\omega$, $G_\omega$, $B_\omega$ are the intensities of three primaries at the pixel $\omega$ in the image plane $\Omega$. Define the chromatic information associated with the representation $f^{\text{RGB}}_\omega$ by

$$
\phi_\omega = \phi(f^{\text{RGB}}_\omega) = \frac{f^{\text{RGB}}_\omega}{|f^{\text{RGB}}_\omega|}.
$$

Noticing that the totality of $\phi_\omega$ is identified with the surface of the positive unit sphere $\partial^1_\omega$, by definition, let the diversity of chromatic information around $\phi_\omega$ be evaluated in terms of the following 2D Gaussian probability density function

$$
g_\alpha(\phi|\phi_\omega) = \frac{1}{2\pi\alpha} \exp \left[ -\frac{|\phi - \phi_\omega|^2}{2\alpha} \right], \quad \phi \in \partial^1_\omega, \tag{6}
$$

with sensitivity parameter $\alpha > 0$. Following experimental studies using various roadway scene images, the sensitivity factor $\alpha$ should be adjusted to $1/10 - 1/100$ [5]. For sufficiently small deviation $|\phi - \phi_\omega|$, the measure $g_\alpha(\phi|\phi_\omega)$ approximates the Gaussian distribution on the tangential space at $\phi_\omega$. 

---

*Perceptual Equivalence of Scale and Chromatic Aspects of Environmental Saliency Arising in Naturally Complex Scenes*
In the color space endowed with the locally Gaussian probability (6), we can evaluate the diversity of the chromatic information in terms of a collection of color samples, called a palette \( s = \{ f_i^{RGB}, i = 1, 2, \ldots, ||s|| \} \) with the following diversity index

\[
R_s = \frac{1}{2\pi\alpha} \exp \left[ -\frac{\sigma^2_{\phi\phi}}{2\alpha} \right],
\]

\[
\sigma^2_{\phi\phi} = \frac{1}{||s||(||s||-1)} \sum_{1 \leq i, j \leq ||s||, i \neq j} |\phi(f_i^{RGB}) - \phi(f_j^{RGB})|^2.
\]

For a sufficiently rich palette \( s \), by definition, the index \( R_s \) provides a statistical estimate for the ranges of chromatic diversity to be assigned to each palette color. Define

\[
G_\alpha(\omega|s) = \max_{f_i^{RGB} \in s} \left[ \phi(f_i^{RGB}) | \phi(f_j^{RGB}) \right].
\]

Then, we can introduce the following matching rule

\[
G_\alpha(\omega|s) > R_s \Rightarrow f_i^{RGB} \sim R_s \sim s,
\]

with respect to the scene specific palette \( s \). The index (7) can be applied to reduce the chromatic complexity of the scene image as shown in Fig. 4 where the chromatic diversities of scene image and extracted palette \( s \) are displayed in upper- and lower-subwindows, respectively. In this case, 995 samples of unique chromatic information were collected on the fractal attractor spanning the entire image plane \( \Omega \) as displayed in upper-subwindow; the samples were refined to select 90 representatives as the scene specific palette \( s \) with respect to the equivalence relation (7b); the matching rule (7) is applied to the following re-coloring process:

\[
f_\omega^{RGB} \rightarrow f_i^*,
\]

where \( f_i^* \) is the nearest representative satisfying

\[
g_\alpha \left[ \phi(f_\omega^{RGB}) | \phi(f_i^*) \right] = G_\alpha [\omega|s],
\]

in the refined palette \( s \). In this scene, 1/10 representatives are sufficient to regenerate an equivalent impression. Thus, we can invoke the chromatic information \( \phi_\omega \) as an access path to the chromatic aspect of the environmental saliency.

### 4 Saliency Pattern Filtering in Chromatic Aspect

Being guided by the 2.5D representation of the ground-object structure, the 2D distribution of the boundary objects can be classified with respect to a scene specific palette as shown in Fig. 4. In contrast with the perceptual linearity, a precise representation of a geometric rule, it is well known that the color matching is an essentially mental process. Based on the schematics of Fig. 2, the perception of chromatic diversity should be supervenient on the anatomic structure developed within the co-evolution environment [22].

Recent advancements in neuronal physiology reveal that the early stage of color perception is substantiated by a stochastic mechanism [14] where three or four types of photopigments are captured by networked neurons. By identifying the neuronal computation steps with an aggregation of stochastic
processes on the locally Gaussian color space, let the complexity for presetting saliency colors be indexed by

$$\psi^H_\omega = e^{-\mathcal{H}(\phi_\omega)} \int_\Omega e^{-\mathcal{H}(\phi_\omega)} d\omega,$$

where the saliency index $\psi^H_\omega$ is attributed to a probability in viewer specific observation: $p^c_\omega(\omega) = (c_\omega/f_{RGB(\omega)})^2, c \in \{R,G,B\}$. Noticing that the ‘negative entropy’ is available as a global evaluation of predictability, we can apply the saliency index to ‘tricolor matting’ of a scene as illustrated in Fig. 5 where a weighted distribution of the brightness $\psi^H_\omega f_{RGB}$ is visualized. The combination of Figs. 4 and 5 implies that we can select saliency patterns supporting an equivalent impression of the objects via $\psi^H$-filtering of $g_\alpha$-image.

5 Multi-Fractal Coding

As demonstrated in Figs. 3 and 4, we can extract distributed information $(\hat{\sigma}_\omega, \phi_\omega)$ as an access path to the environmental saliency. Noticing that the scene specific features can be efficiently extracted via the fractal sampling as demonstrated in Fig. 4, consider fractal coding of saliency pattern enhanced through $\psi^H$-channel. The problem is to design a set of contraction mappings $\nu_c$ for generating a fractal attractor $\Xi_c$ satisfying the self-similarity (5) within a saliency pattern. Since the imaging process of the attractor $\Xi_c$ is controlled by the fixed point $\omega_\mu$, as indicated in (5), we can design the mapping set $\nu_c = \{\mu_i\}$ generating a saliency pattern spanning an object image; by using such an explicit formulation, the expansion of fractal attractor can be visualized on a connected surface as well as the roadway area illustrated in Fig. 3(b).

It seems to be an easy task for human vision endowed with not-yet-explicated inherent information processing mechanism to articulate unstructured distribution of the boundary objects (Fig. 3) into landmarks as enhanced in Fig. 5. To maintain the close cooperation with such a sophisticated process, suppose that a representative $f^* \in \mathcal{S}$ is selected in the enhanced image $\psi^H f_{RGB}$. Define a $\mathcal{S}$-pattern by the distribution of equivalent pixels satisfying

$$\mathcal{D} = \{\omega \in \Omega \mid f_\omega \overset{R_\omega}{\sim} f^* \in \mathcal{S}\}.$$  

In many natural scenes, various objects are wrapped by similar saliency colors. To separate a landmark object within the distribution $\mathcal{D}$, we introduce the following articulation algorithm consisting of two successive steps; expansion and unification.

**Successive Expansion:** Let $\partial^D \mathcal{D}$ be the Laplacian-Gaussian boundary associated with the attractor $\Xi_c$ (Appendix A). Noticing the attractor should be expanded nondeterministically by the fixed points,
we have the following allocation scheme:

\[
\Omega_{t+1}^* = \Omega_t^* \cup d\Omega_t^*,
\]
\[
d\Omega_t^* = \left\{ \partial \omega^* \mid \forall \partial \omega : \frac{\partial}{\partial \omega} \left( \omega^*, \Omega_t^* \right) \geq \frac{\partial}{\partial \omega} \left( \omega, \Omega_t^* \right) \right\},
\]
\[
\Omega_0^* = \left\{ \partial \omega^*_0 \mid \forall \partial \omega_0 : G_\alpha \left( \partial \omega_0 \{ f^* \} \right) \geq G_\alpha \left( \partial \omega_0 \{ f^* \} \right) \right\},
\]
with respect to the following ‘semi-distance’ introduced between a point \( \omega \in \Omega \) and a subset \( \Lambda \subset \Omega \):

\[
\frac{\partial}{\partial \omega} \left( \omega, \Lambda \right) = \min_{\lambda \in \Lambda} |\omega - \lambda|.
\]

The expansion process (11) mutually separates the fixed points in the fixed set \( \partial^5 \Omega \) and halts at \( T \) when the increment \( d\Omega_t^* \) satisfies the following sub-scale condition:

\[
\max_{\omega \in d\Omega_t^*} \frac{\partial}{\partial \omega} \left( \omega, \Omega_t^* \right) < \sigma_0.
\]

**Successive Unification:** Many physical entities can be surrounded by convex contours in the scene images. This implies that the final set \( \Omega_f^* \) should be articulated according to a convexity criterion. To this end, we invoke the following nondeterministic algorithm:

\[
\Omega_{t+1}^f = \Omega_t^f \cup d\Omega_t^f,
\]
\[
d\Omega_t^f = \left\{ \partial \omega^f \mid \forall \partial \omega : \frac{\partial}{\partial \omega} \left( \omega^f, \Omega_t^f \right) \leq \frac{\partial}{\partial \omega} \left( \omega, \Omega_t^f \right) \right\},
\]

The successive process (13) expands the initial set \( \Omega_0^f = \Omega_0^f \) towards a convex set within the final expansion \( \Omega_f^* \). To maintain an on-going articulation \( \Omega_f^* \) within an object image, the process (13) is interrupted by the following breakdown criterion:

\[
\min_{\xi \in d\Xi_t} G_\alpha \left( \xi | s \right) < \min_{\xi \in \Xi_t} G_\alpha \left( \xi | s \right),
\]

where \( \Xi_t \) and \( d\Xi_t \) denote fractal attractors generated by the mapping sets \( \nu_t \), and \( d\nu_t \) associated with on-going and testing fixed points, i.e., \( \Omega_t^f \) and

\[
d\Omega_t^f \cup \left\{ \partial \omega^f \mid \forall \partial \omega : \frac{\partial}{\partial \omega} \left( \omega^f, \Omega_t^f \right) \leq \frac{\partial}{\partial \omega} \left( \omega^f, d\Omega_t^f \right) \right\},
\]

respectively. The unification process is finally halted by the singularity condition: \( \| \Omega_f^* - \Omega_f^f \| < 2 \).

An example of the allocation results by the algorithm (11)–(14) is indicated in Fig. 6. In this case, first, the sample \( f^* \) nearest to \( \mathbb{R} \)-primary is selected to confine \( \Omega \) as displayed in (a); the boundary \( \partial^5 \Omega \) is displayed by white lines in the main window with \( \Omega_f^* \) (blue dots on \( \partial^5 \Omega \)) and local maxima of associated Gaussian field (red dots distributed in \( \Omega \)); the chromatic diversities of \( \Omega \) and associated \( g_\alpha \)-sampling by (7) are visualized in upper- and lower subwindows, respectively. Next, \( \Omega_f^* \) is successively unified as illustrated in (b) where the chromatic diversities on \( \Xi_t \) and \( d\Xi_t \) are compared in upper- and lower subwindows, respectively. Finally, the convexity of \( \Omega_f^* \) at the first breakdown articulation by (14) was verified via the sampling test on associated attractor as shown in the main window of (c) with subwindows exhibiting associated chromatic diversities. As shown in (a) of this figure, it is not easy to articulate the pattern \( \Omega \) into landmark objects via conventional local maxima separation by the boundary \( \partial^5 \Omega \). This yields an over estimate as shown (b) where the final set \( \Omega_f^* \) may generate a fractal attractor spanning three objects wrapped by the same saliency color. Through the unification algorithm (13) we have an articulation as displayed in (c); the final set \( \Omega_f^* \) is partitioned into a convex subset \( \Omega_f^* \) on a surface rectangle of the object and ‘the rest’ at the first interruption. The resulted fixed point \( \omega_f^* \in \Omega_f^* \) is used to design a contraction mapping \( \mu_s \in \nu_s \); the distribution \( g_s \left[ \phi_s \mid \phi \left( f^* \right) \right] \) is restricted to the invariant measure with respect to \( \nu_s \); under the breakdown criterion (14), the mapping set \( \nu_s \) generates a fractal attractor \( \Xi_s \) satisfying the constraint (5) within the surface image. By iterating this unification-interruption steps, thus, we have an articulation of the boundary distribution into a system of fractal attractors.
Perceptual Equivalence of Scale and Chromatic Aspects of Environmental Saliency Arising in Naturally Complex Scenes

(a) Fixed point allocation $\Omega^*_T$ via successive expansion.

(b) Interruption of successive unification process at $G_\alpha$-breakdown.

(c) Articulation of fixed point allocation: $\Omega^*_t \rightarrow \Omega^f_t$.

Figure 6: Fractal articulation of saliency distribution $\mathcal{D}$.

Table 1: Two aspect structure of environmental saliency.

<table>
<thead>
<tr>
<th>aspect</th>
<th>scale aspect</th>
<th>chromatic aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>visibles</td>
<td>$f_\omega \sim f^{vis}_{\omega}$</td>
<td></td>
</tr>
<tr>
<td>control parameter</td>
<td>$\hat{\sigma}_\omega$</td>
<td>$\phi_\omega$</td>
</tr>
<tr>
<td>receptor</td>
<td>rhodopsin</td>
<td>photopsin</td>
</tr>
<tr>
<td>access path</td>
<td></td>
<td>$\hat{\sigma}_\omega$</td>
</tr>
<tr>
<td>structure</td>
<td>ground-object</td>
<td>illuminant array</td>
</tr>
<tr>
<td>generic rule</td>
<td>${ \sigma_d, \sigma_0 }$</td>
<td>${ R_\alpha, \mathcal{g} }$</td>
</tr>
<tr>
<td>measure</td>
<td>${ g_d, g_b }$</td>
<td>${ g_\alpha, \psi^{RC} }$</td>
</tr>
<tr>
<td>fractal model</td>
<td>$\nu_d \leftarrow \Omega^f \rightarrow { \nu_c }$</td>
<td></td>
</tr>
</tbody>
</table>

6 Perceptual Equivalence

As shown in Figs. 3, 4 and 6, the ground-object structure can be identified through bottom up organization of the environmental saliency; the fluctuation of the scale information $\hat{\sigma}_\omega$ is matched with the generic model $\hat{\sigma}_d$ to yield a probabilistic evaluation of a connected open space $g_d (\omega | \sigma_0)$; the diversity of chromatic information $\phi_\omega$ is efficiently sampled to generate the as-is palette $\mathcal{g}$ with nondeterministically precise measure $G_\alpha (\omega | \mathcal{g})$. The two aspect structure of the environmental saliency is summarized in Table 1. Under the control of $\hat{\Omega}^d$ (trichromatic primary), the environmental saliency is grounded on the
visible \( f_\omega \) \( f_{\omega^{RGB}} \) through the access path \( \tilde{\omega} \) \( \phi_\omega \); this process is substantiated by the fast (slow) channel accepting rhodopsin (photopsin). In response to the stimuli \( f_\omega \) \( f_{\omega^{RGB}} \), the scale- (chromatic-) aspects yields a stochastic representation of the ground-object structure (illuminant array) to be matched with the generic rule \( \{ \sigma_d, \sigma_0 \} \) \( \{ \mathbb{R}_a, \mathbb{S} \} \) through the measure \( g_d \) \( g_a \) (\( \psi^{\omega^{RGB}} \)). On resulted set of fixed points \( \Omega^f \), we can design fractal models to articulate the scene image into a connected open space in the roadway area \( \nu_d \) and a system of saliency objects \( \{ \nu_c \} \).

As indicated in Table 1, the environmental saliency induces a multi-representation of a common basis of fractal modeling \( \Omega^f \) supported by two observations of the brightness distribution \( f_\omega \) \( f_{\omega^{RGB}} \). Under the schematic of Fig. 2, perception via the scale- and chromatic-aspects should generate preestablishingly consistent descriptions of the environmental saliency. By invoking the generic structure based on the ecological optics, (i) \( (g_d, g_b) \)-based recognition of a object free space should be confirmed via \( g_b \)-based analysis of the scale aspect model; simultaneously, (ii) \( (g_a, \psi^{RGB}) \)-based boundary articulation should be associated with the landmark objects relative to the 2.5D perspective of the open space through \( (g_d, g_b) \)-based depth analysis.

Suppose that the expansion of the roadway area and boundary objects are represented by self-similarity processes visualized through the invariant measures \( g_d (\omega|\sigma_0) \) and \( g_b (\omega|\sigma_0) \), respectively. Within the framework of such a fractal modeling, it has been pointed out that the probability distributions to capture the roadway and boundary objects can be generated via the following system of diffusion equations [23]:

\[
\begin{align*}
\frac{\partial \varphi_d (t; \omega|\nu_d)}{\partial t} &= \frac{1}{2} \varphi_d (\omega|\nu_d) + \rho_d [g_d (\omega|\sigma_0) - \varphi_d (t; \omega|\nu_d)], \\
\varphi_d (0; \omega|\nu_d) &= 0, \\
\frac{\partial \varphi_b (t; \omega|\nu_b)}{\partial t} &= \frac{1}{2} \varphi_b (\omega|\nu_b) + \rho_b [g_b (\omega|\sigma_0) - \varphi_b (t; \omega|\nu_b)], \\
\varphi_b (0; \omega|\nu_b) &= 0,
\end{align*}
\]

(15a)

where \( \rho_d \) is adjusted to the Kolmogorov’s complexity [24] of the generic roadway model \( \nu_d \) [5]; following the empirical knowledge on the emotional perception [19], the complexity factor \( \rho_b \) of the mapping set \( \nu_b \) should be adjusted to the fractal dimension of skylines, i.e., \( \rho_b = 1.3 \). Let \( \varphi_d (\varphi_b) \) be a version of the steady state solution to (15)

\[
\varphi_d (\omega|\nu_d) \to C_\nu \varphi_d (\omega|\nu_d), \quad (t \to \infty)
\]

with normalization constant \( C_\nu C_b \). With the complexity factor \( \rho_d, \rho_b \), \( \varphi_d (\varphi_b) \) yields the probability distribution to capture an aggregation of Brownian motion processes confined by the mapping sets \( \nu_d (\nu_b) \). Based on statistical moments of the conditional distribution \( \varphi_d (\omega|\nu_d) \), we have a design of the fractal model \( \nu_d \) spanning over a connected object free space. Let the boundary distribution be restricted to a saliency pattern \( \mathbb{D} \) by \( \psi^{\omega^{RGB}} \)-filtering and suppose that a fractal attractor \( \nu_c \) is designed for generating an attractor \( \Xi_c \subset \mathbb{D} \). The consistency of the chromatic aspect model \( \nu_c \) can be verified through a finite computational test on the local maxima \( \Theta \) of the capturing probabilities \( \varphi_d (\varphi_b) \). By invoking a necessary condition for the existence of the self-similarity structure [23], we can test the inter-aspect consistency of the boundary object model; for the set \( \nu_c \) generating a fractal attractor within the support of the capturing probability there exists a subset \( \Theta_c \subset \Theta \) satisfying the following finite invariance condition:

\[
\Theta_c = \left\{ \theta \in \tilde{\Theta} \mid \exists \mu_i \in \nu_c : \mu_i^{-1}(\theta) \in \Theta_c \right\}.
\]

(16)

Due to the finite testability of the self-similarity, the object models within \( \psi^{\omega^{RGB}} \)-image can be efficiently matched with the ground-object structure. Noticing that the 2.5 dimensionality of the capturing probability, and resulted set \( \tilde{\Theta} \) as well, maintains the perspective of the scene to be structured, we can introduce a self-reflective mechanism in the multi-aspect representation of the environmental saliency.

Thus, the environmental saliency to be resulted in the schematics of Fig. 2 is identified through a bottomup integration process of perceptually equivalent channels; the two aspects indicated in Table 1 are mutually associated through computationally independent analysis of scale- and chromatic-randomness distributed in scene images.

7 Experiment

The perceptual equivalence of the multi-fractal coding process was verified through experimental studies using various scene images. Noting the two phase organization consisting of ground-object partitioning and landmark articulation, first, the equivalence in the ground driven process is considered; the problem
Perceptual Equivalence of Scale and Chromatic Aspects of Environmental Saliency Arising in Naturally Complex Scenes

Figure 7: Fractal sampling of a distant view of the scene. \((\nu_d^1, \Xi_d^1)\)

(a) \(g_a\)-matting \((\nu_d^2 \leftarrow s_d \leftarrow \Xi_d^1)\).

(b) Invariant feature \(\Theta_d\) on \(g_a\)-matting image \((\nu_d^2, \Xi_d^1)\).

Figure 8: Equivalence of \(\bar{\sigma}_d - \bar{s}\)-rule in ground driven perception.

is to verify the consistency of a chromatic aspect model induced on the open space via the scale aspect analysis. Next, the object driven process is verified; the problem is to localize the multi fractal articulation of the \(\psi_{\omega}^{\omega} f_{\omega}^{\omega}\)-distribution through the finite invariance analysis on the scale aspect information \((\varphi_d, \varphi_b)\).

Ground Driven Perception: Figures 7 and 8 show a part of experimental results for the fractal coding in the distant view of the scene displayed in Fig. 3. In this scene image, it has been demonstrated that a fractal model \(\nu_d^1\) to control the imaging process (5) within the roadway area has been designed and verified through the finite invariance test (16) on the the distribution \(\{g_d(\omega|\sigma_0), g_b(\omega|\sigma_0)\}\) [25]; by sampling a palette \(s_d\) on the associated attractor \(\Xi_d^1\) as shown in Fig. 7, a \(g_a\)-matted image was generated as shown in Fig. 8(a); by applying \(\{g_d, g_b\}\)-based fractal modeling to this restriction, again, we have another version of the mapping set \(\nu_d^2\) which generates a finite closed graph connecting a discrete set \(\Theta_d\) called an invariant feature through the self-similarity mechanism (5) as shown in (b) where \(\Theta_d\) and links are marked by large dots and meshed lines, respectively. The existence of the invariant feature \(\Theta_d\) demonstrates that the palette \(s_d\) in the chromatic aspect provides an essentially equivalent information for fractal modeling within the scale aspect; both \(\nu_d^1\) and \(\nu_d^2\) generate nonempty invariant features with
Perceptual Equivalence of Scale and Chromatic Aspects of Environmental Saliency Arising in Naturally Complex Scenes

Figure 9: Saliency distribution by $\psi^{\mathcal{H}}$-filtering in a distant view.

(a) Invariant feature $\Theta_c$ generated by $\nu_c$ on $g_b$-information.

(b) Invariant feature $\Theta_c$ generated by $\nu_c$ on $g_d$-information.

Figure 10: Finite invariance test of fractal articulation $(\Omega^f, \nu_c)$

Object Driven Perception: The perceptual equivalence has been demonstrated through experimental analysis using a close-up of the scene (Fig. 3). In this case, the $G_{\alpha,\omega}|s\rangle$-distribution is generated in the chromatic aspect to activate the successive expansion-unification process illustrated in Fig. 6. The experimental results of such an object driven process are summarized in Figs. 10 and 11; a mapping set respect to the common measure $g_d(\omega|\sigma_0)$.

By applying the $\psi^{\mathcal{H}}$-filter directly to the scene image, we have the distribution of the saliency patterns as illustrated in Fig. 9. In this case, however, the patterns was too small to yield the fixed points passing through the halting condition (12).

The combination of Figs. 7, 8 with 9 implies that the perceptual equivalence on the environmental saliency induces a generic perception strategy: the fast decision makings should be supervenient on the ground-object structure based on the scale information; even for too distant landmarks, $\psi^{\mathcal{H}}$-filtering should be exploited to preset the fractal articulation process prior to physical approach. Thus, the first equivalence of the scale- and chromatic-aspect models was verified on the open space in the scene image shown in Fig. 3.
nu was designed with respect to the first cluster of the model parameter Omega^f in the chromatic aspect. By applying the set nu to the scale aspect representation, the consistency of nu with the depth information was verified as shown in Fig. 10; the designed fractal code was demonstrated to yield a sufficient invariant links within the two versions of the local maxima with respect to the breakdown distribution phi_b(omega|nu_b) and the open space evaluation phi_d(omega|nu_d) as exhibited in (a) and (b), respectively: (a) implies that there exists a perpendicular plane supporting the fractal attractor to be generated by the model nu; simultaneously, (b) indicates that the plane has ground pixels within the perspective of the roadway area. This implies that the chromatic aspect coding nu is essentially consistent with the breakdown criterion of the perceptual linearity underlying the scene image; thus, essentially 2D model via chromatic aspect analysis can be localized within the context of 2.5D perspective induced on the scale aspect of the environmental saliency.

As a result, we can visualize an object at the saliency pattern within the context of the ground-object structure as illustrated in Fig. 11 where the ranges of mappings gamma^c_i in nu are jointly marked by small dots with different colors; the range of the attractor is confined by a subwindow avoiding the distribution of the rest Omega^t - Omega^f computed as shown in Fig. 6. Despite of iconic discrepancy, thus, the scope of object image analysis can be parametrically confined and separated from the roadway area (Fig. 3) by nu. The chromatic diversity of the fractal model is displayed in the subwindows; despite considerable diversities of fractal sampling (upper-subwindow) and associated palette (lower-subwindow), the connectedness of the psi^H_omega-pattern is supported by the existence of Theta as shown in Fig. 10.

Figures 12 and 13 show the results of other experiments. In these experiments, saliency patterns are extracted as indicated in (a); associated fractal models are verified on the scale space information and visualized in the ground-object structure as illustrated in (b) and (c), respectively. The experimental results demonstrate that the environmental saliency provides effective cue for analyzing even ill-conditioned scenes where landmark objects are observed as non-dominant patterns; the post is rather ‘low-keyed’ object in a night view as shown in Fig. 12; a warning color of vehicles sometimes yields smaller image than attractive distractions as displayed in Fig. 13. Thus, the second equivalence was demonstrated via the 2.5D localization of the chromatic aspect models in the scene images successfully.

8 Discussions

In the current implementation, the fractal articulation scheme is activated in response to human’s selection of a palette color f^* on the psi^H_omega-images. The significance of psi^H-filtering was verified by using various types of scene images as summarized in Table 2. In this table, the reduction of the unpredictability is evaluated in terms of the relative complexity index dS_{G} = S_{G} - S_{G}; S_{G} and S_{G} designate the Shannon’s entropy with respect to the uniform distribution and the gray level distribution f_{omega}, respectively; S_{H} stands for the following entropy

\[ S_{H} = - \int_{Omega} \psi^{H}_{omega} \log (\psi^{H}_{omega}) \, domega, \]

\[
\begin{array}{|c|c|c|}
\hline
\text{scene} & dS_{G} & dS_{H} \\
\hline
\text{shopping street} & 0.131549 & < 0.897254 \\
\text{post office} & 0.200029 & < 1.794611 \\
\text{street view at night} & 0.148731 & < 1.278006 \\
\text{industrial park} & 0.103450 & < 1.620070 \\
\hline
\end{array}
\]
Perceptual Equivalence of Scale and Chromatic Aspects of Environmental Saliency Arising in Naturally Complex Scenes

Figure 12: \(g_d - \psi^{3c}\) equivalence induced by scale-chromatic randomness.

(a) Saliency distribution via \(\psi^{3c}\)-filtering.

(b) Invariant feature \(\Theta_c\) generated by \(\nu_c\) on \(g_d\)-information.

(c) Contextual visualization of \(\nu_c\)-model (street view at night).

with respect to the saliency index. As shown in Table 2, the \(\psi^{3c}\)-filter concentrates the information distributed in the image plane into a set of saliency patterns of average size \(dS_G = 2.4\) to 6; this implies that the essential length of decision steps on the \(\psi^{3c}f_{RGB}\)-image is reduced to 0.4–1/6 of random search in the image plane; the computational complexity is no greater than 1/2–1/5 the decision steps on the gray scale image indicated by \(dS_G\). As demonstrated in this table, the length of the decision steps in relative simple scenes as indicated in Figs. 7 and 12 is reduced to 0.5–0.6 in the naturally complex scenes as indicated in Figs. 5 and 13. This may imply the existence of ‘naturally designed scenes’: random allocation of landmark objects exhibiting easy-to-read representation of sign and/or design as the environmental saliency. From the view point of the acceptance by the human’s perception process, the saliency patterns should be efficiently scanned all over the image. As summarized in Table 2, we can introduce the saliency index to concentrate the machine and/or neuronal computational resource to the possible landmark images.

In this paper, the future trajectory is utilized to preset the on-vehicle vision system. Noticing that the fractal coding process is supported by the field dynamics (15), we can exploit the ‘initial model’ as the prediction of the object images in the subsequent maneuvering process; by adjusting the fixed point along the optical flow associated with the view point shift [13], the fractal models can be parametrically updated without re-clustering via the expansion-unification steps. This implies that the deviation of the vanishing point due to GPS instability is successively reduced through the adaptation of the \(\bar{\sigma}_d\)-model to the \(g_b\)-distribution.

As indicated in Table 1, the saliency distribution is extracted and articulated within a cut of scene images. Such a two dimensionality implies that small signs of landmark objects can be magnified via
Perceptual Equivalence of Scale and Chromatic Aspects of Environmental Saliency Arising in Naturally Complex Scenes

Figure 13: $g_d - \psi^{\mathcal{K}}$ equivalence induced by scale-chromatic randomness

re-sampling the image; by applying the clustering scheme (11) combined with (13) to the ‘close-up’ of each object observed in a distant view (Fig. 9), we have the multi-fractal articulation of the saliency distribution prior to physical approach to the scene; being supported by the perceptual equivalence, the object models are ‘collaged’ within the perspective of the ground-object structure as displayed in Fig. 11.

In contrast to the dynamic adaptation in the scale-aspect process, however, the consistency of the saliency patterns is restricted within a specific cut of image; i.e., the expansion of the saliency pattern is explicitly determined by the palette sampled in the specific scene image. Noticing that a compact pallet extracted from naturally complex scenes can be adapted to the satellite image spanning tempo-spatio discrepancy [5], we can utilize the palette from the bird’s eye view as a prediction of the perspectives of the scenes. In the experimental results displayed in Figs. 6, 12 and 13, 0.2% of chromatic information is sufficient for the recognition of a landmark object as well.

The over-the-horizon adaptation of the palette to the chromatic aspect of the environmental saliency is the next problem. Adding to it, the development of the cooperative, and compensative as well, scanning mechanism is left to future investigations on the inherently understandable vision systems.

9 Concluding Remarks

A new framework was introduced for saliency based analysis of naturally complex scenes. As the basis of structural description, in this framework, the scale shift and the chromatic diversity are extracted via randomness based image feature analysis and organized into the environmental saliency. By associating the scale-chromatic complexity of the environmental saliency with the invariant measure, the randomness
distribution is articulated into a system of fractal attractors. Through experimental studies, it has been demonstrated that the multi-fractal articulation induces the perceptual equivalence between the scale-and chromatic-aspects of the environmental saliency. By this perceptual equivalence, a self-reflective mechanism is implemented in the two-aspect representation of environmental saliency.

References


A Laplacian-Gaussian Analysis

By using the following approximation

\[ g_\tau (\omega) - \delta \sim \frac{\tau}{2} \Delta g_\tau (\omega), \quad \omega \in \Omega, \]  

\[(A1)\]
where \( g_\tau(\omega) = e^{-|\omega|^2/2\tau} \), we have
\[
\sigma^2 = \tau \leq \frac{2g_\tau(\omega)}{|\Delta g_\tau(\omega)|}.
\] (A2)

By identifying the brightness distribution \( f_\omega \) with a smoothed image of an aggregation of delta distributions, we can extend the upper boundary (A2) as stated in (2).

Furthermore, by applying the zero-cross detection [9] to the smoothed point image \( g_\tau \), we have the following representation of the zero-cross boundary
\[
\partial g = \left\{ \omega \in \Omega \mid g_\tau(\omega) = \bar{p} \right\}, \quad \bar{p} = \frac{1}{\nu} \max \nu g_\tau(\omega),
\]
without explicit dependence of the variance \( \tau \). Noticing the association \( \phi_\nu(\omega|\nu) \sim g_\nu(\omega) \) induced by (A1), we can extend the zero-cross detection to the smooth distribution \( \phi_\nu(\omega|\nu) \) to yield the following Laplacian-Gaussian boundary:
\[
\partial g_D = \left\{ \omega \in \Omega \mid \phi_\nu(\omega|\nu) = \bar{p} \right\},
\] (A3)
where the level \( \bar{p} \) is determined prior to exact specification of the mapping set \( \nu \) [23].