

GPS-Based Multi-viewpoint Integration for Anticipative Scene Analysis

Kohji Kamejima

Osaka Institute of Technology, Faculty of Information Science and Technology,
1-79-1 Kitayama, Hirakata 573-0196 Japan,
kamejima@is.oit.ac.jp
<http://www.is.oit.ac.jp/~kamejima>

Abstract. A multi-viewpoint integration scheme is introduced to recognize scene features *prior to* physical access. In this schematics, chromatic complexity of vehicle's- and bird's-eye-views of roadway scenes are matched to extend GPS tracks towards possible destinations. Saliency patterns arising in destination images are anticipatively extracted to control the focus of inherent and machine vision to what to be analyzed.

Keywords: Multi-viewpoint Image, Scene Analysis, Chromatic Complexity, GPS Signal Processing, Image Saliency.

1 Introductory Remarks

By networking on-board vision system with global positioning and earth observation systems, we can localize landmark objects in distant scenes along roadway patterns to be followed. This implies that a multitude of intelligent vehicles can be exploited for over-the-horizon cooperation of the maneuvering processes; the scene images provided by probe vehicles are re-used by future visitors as anticipative visualization of landmarks to be recognized. Such augmented perception can be utilized as the basis of spontaneous coordination of participant vehicles through the localization [8] and regulation [9] of maneuvering processes. To activate the localization and/or regulation processes, the future visitors should retrieve the probed images along feasible trajectories and anticipatively control the focus of recognition.

Within the framework of the the satellite-roadway-vehicle networking, we can implement an access path to 'future' scenes as illustrated in Fig. 1; a scene image is matched with a local section of satellite image including GPS track; detected roadway pattern is extended towards possible destinations to observe landmark distribution in the probed scenes *prior to* physical access. It has been pointed out that image features can be organized to extract saliency distribution including various types of objects-to-be-observed [3,7]. However, due to significant discrepancy of photographing conditions, it is not easy to predict what will be observed in the future scenes through the analysis of the probed images; based on such ill conditioned observation, in many practical situations, the on-vehicle vision are

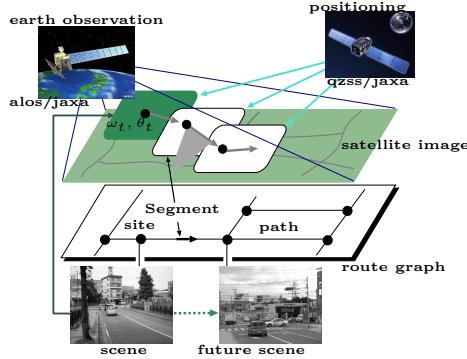


Fig. 1. Schematics of Anticipative Scene Analysis

required to yield a man readable visualization to maintain the consistency of the over-the-horizon maneuvering processes. Noticing that human's inherent vision is endowed with not-yet-explicated focusing mechanism in chromatic diversity, in this paper, we consider saliency based integration of multi-viewpoint images connected by GPS tracks.

2 GPS Tracking on Satellite Image

Suppose that bird's-eye-views and perspectives of roadway scenes are jointly captured through the earth observation systems and on-vehicle cameras, respectively, to generate multi-viewpoint images on an image plane Ω endowed with longitude-latitude or camera specific coordinate system. Let GPS tracks be identified with stochastic processes to be segmented in terms of 2D vectors $\mathbf{v}_t = (\omega_t, \theta_t)$, $t = 0, 1, 2, \dots, k$, with origin ω_t and direction θ_t . By non-deterministically associating a sequence of \mathbf{v}_t within GPS residual [4], we have a noisy observation of not-yet-identified roadway pattern on a cut of satellite image.

To correct the GPS residual, the scene image captured at ω_t is matched with a local section of the satellite image. In this association process, first, an open space in a roadway area is recognized in the scene image and scanned to collect a set of color samples $S = \{f_i^{\text{RGB}}, i = 1, 2, \dots\}$ with size $\|S\|$. Let the pixel value f^{RGB} be mapped into the positive part of a unit sphere; through the mapping $\phi(f^{\text{RGB}}) = f^{\text{RGB}} / |f^{\text{RGB}}|$, we can evaluate the local complexity of the sample S in terms of 2D Gaussian measure

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

Define local similarity by $R_S = e^{-\sigma_{\phi\phi}^2/2\alpha}/2\pi\alpha$ based on the following statistics

$$\sigma_{\phi\phi}^2 = \frac{1}{\|S\|(\|S\| - 1)} \times \sum_{f_i^{\text{RGB}}, f_j^{\text{RGB}} \in S, f_i^{\text{RGB}} \neq f_j^{\text{RGB}}} |\phi(f_i^{\text{RGB}}) - \phi(f_j^{\text{RGB}})|^2.$$

By selecting the representatives $f_{i^*}^{\text{RGB}} \in S$ with respect to the equivalence criterion R_S , we have a palette $\mathfrak{s} = \{f_{i^*}^{\text{RGB}}\}$ to nondeterministically re-draw the essential part of the roadway area. Let the consistency of a pixel f_{ω}^{RGB} , $\omega \in \Omega$ with \mathfrak{s} be evaluated in the locally Gaussian space as follows:

$$\max_{f_j^{\text{RGB}} \in \mathfrak{s}} g_{\alpha} [\phi(f_{\omega}^{\text{RGB}}) | \phi(f_j^{\text{RGB}})] > R_S \Rightarrow f_{\omega}^{\text{RGB}} \stackrel{R_S}{\sim} \mathfrak{s}. \quad (2)$$

To associate the bird's eye view with the perspective of the vehicle, a local section including ω_t is scanned to generate another palette $\check{\mathfrak{s}} = \{\check{f}_i^{\text{RGB}}\}$. By applying the association rule (2), we have an estimate of the roadway palette:

$$\hat{\mathfrak{s}} = \left\{ \check{f}_i^{\text{RGB}} \stackrel{R_S}{\sim} \check{\mathfrak{s}} \mid \check{f}_i^{\text{RGB}} = f_i^{\text{RGB}} + (\bar{\check{\mathfrak{s}}} - \bar{\mathfrak{s}}), \quad f_i^{\text{RGB}} \in \mathfrak{s} \right\}, \quad (3)$$

where $\bar{\mathfrak{s}}$ and $\bar{\check{\mathfrak{s}}}$ designate mean colors of \mathfrak{s} and $\check{\mathfrak{s}}$, respectively. In (3), the spectral shift due to the disparity of photographing conditions are compensated via the mean value adaptation; samples from the exterior of the not-yet-identified roadway area are eliminated as non-equivalent pixels. By evaluating the consistency within the bird's eye view in terms of $(\hat{\mathfrak{s}}, g_{\alpha})$ -information, thus, we can apply Hough voting to roadway pattern detection; $\mathbf{v}_t(\omega_t, \theta_t)$ is adapted to $\hat{\mathbf{v}}_t(\hat{\omega}_t, \hat{\theta}_t)$ spanning significant chromatic discrepancy between scene- and satellite images.

An example of experimental results of roadway pattern detection is shown in Fig. 2; a GPS track from the 'gateway' to the 'branch' is marked on the satellite image in terms of \mathbf{v}_t -sequence in (b); a perspective image at the branch is captured by the camera to generate the palette \mathfrak{s} on a sampling area indicated in (a); the palette \mathfrak{s} is adapted to $\check{\mathfrak{s}}$ extracted from the satellite image as illustrated in (c); resulted estimate $\hat{\mathfrak{s}}$ is applied to detect a segment of the roadway pattern at the branch area. As indicated in (d), the segment was allocated along really existing roadway patterns; furthermore, by iterating the segmentation process, a 'robotic probe' was deployed to simulate the $\hat{\theta}_t$ -sequence of a GPS track towards the 'cross' as indicated in (e); thus, the GPS tracks was extended to connect possible destinations. By invoking experimental results using various complex roadway scenes, we can simulate feasible trajectories of 1.3–1.9 km length towards future scenes [5]; via the adaptation-and-elimination process (3), 1/2,000–1/2,500 of color samples are used in the Hough voting process; in the adaptation process shown in (c), the complexity of roadway area is finally represented by 29 feature colors.

3 GPS-Based Graph Generation

Suppose that possible destinations $N = \{n_i^d, i = 1, 2, \dots\}$ are localized in the satellite image. By identifying the simulated GPS track with a route graph

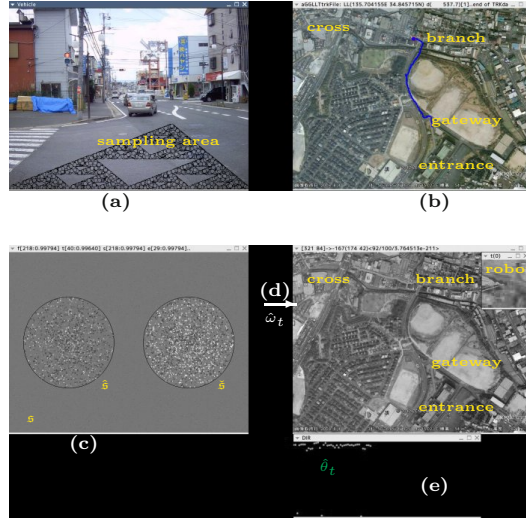


Fig. 2. GPS Track Extension via Multi-viewpoint Association

spanning the distribution of N , we can extract a image of ‘future’ scenes which has been captured by probe vehicles. To implement efficient search scheme, we introduce a dynamic matching algorithm in a stochastic sense.

Conditional Brownian Motion Model: Let a landmark $n \in N$ be localized at $n_{\perp} \in \Omega$ and consider the probability distribution to capture 2D Brownian motion with origin landmark n at time σ , i.e.,

$$g_{\sigma}(\omega|n) = \frac{1}{2\pi\sigma} \exp \left[-\frac{|\omega - n_{\perp}|^2}{2\sigma} \right]. \quad (4)$$

Noticing the following approximation

$$g_{\sigma}(\omega|n) - \delta_{n_{\perp}} = \int_0^{\sigma} \frac{1}{2} \Delta g_s(\omega|n) ds \sim \frac{\sigma}{2} \Delta g_{\sigma}(\omega|n),$$

and extending the representation (4) to the distribution of the localization image $\chi_N = \frac{1}{\|N\|} \sum_{n \in N} \delta_{n_{\perp}}$, we have the following probability distribution to finally capture at least one Brownian motion shifted from N :

$$\frac{\sigma}{2} \Delta \varphi_{\sigma}(\omega|N) + [\chi_N - \varphi_{\sigma}(\omega|N)] = 0. \quad (5)$$

By identifying the information arising from the landmarks $N - n$ with independent noisy background, we have the following decomposition:

$$\varphi_{\sigma}(\omega|N) = g_{\sigma}(\omega|n) + \chi_{\Omega}, \quad (6)$$

where χ_{Ω} denotes the uniform distribution standing for the background noise.

Dynamic Detection: Let the origin be randomly re-selected in each unit time interval. In such a situation, the probability to capture a Brownian motion iteratively conditioned until time $t > 0$ is given by

$$\tilde{g}_t^n(\omega|N) = \left(\frac{1}{\|N\|} \right)^t g_t(\omega|n) = e^{-\rho t} g_t(\omega|n), \quad \rho = \log \|N\|. \quad (7)$$

Thus, we can apply the dynamic likelihood test [10] to the identification of the segment $\hat{\mathbf{v}}_t$ with the conditional Brownian motion process. To this end, first, the variation is evaluated as follows:

$$d\varphi_t = -\nabla\varphi_\sigma(\hat{\omega}_t|N)^T \hat{\mathbf{v}}_t dt. \quad (8)$$

Noticing $\varphi_\sigma \sim g_\sigma$ within the area $|\omega - \hat{n}_\perp| < \sqrt{\sigma_N}$ and

$$\nabla g_\sigma(\omega|n) = -\frac{1}{\sigma}(\omega - \hat{n}_\perp) g_\sigma(\omega|n),$$

next, we have the following stochastic differential:

$$d\tilde{\varphi}_t = \frac{d\varphi_t}{\varphi_t} = \tilde{\phi}_t dt + \sqrt{\sigma_w} dw_t,$$

where $\tilde{\phi}_t = (\omega - \hat{n}_\perp)^T \hat{\mathbf{v}}_t$ and the fluctuation due to the background distribution χ_Ω is simulated in terms of Wiener process w_t with $\sqrt{\sigma_w}$. By modifying the Wiener measure in terms of the iteration factor $e^{-\rho t}$ induced in (7) and noticing that the likelihood ratio satisfies the stochastic differential equation

$$\frac{d\Lambda_t}{\Lambda_t} = \frac{\tilde{\phi}_t}{\sigma_w} d\tilde{\varphi}_t - \rho dt, \quad (9)$$

finally, the nearest localization \hat{n}_\perp can be detected in a rectangle $d\hat{\Omega}_t$ with diagonal vertexes $\hat{\omega}_t$ and $\tilde{\omega}_t$ given by

$$\tilde{\omega}_t = \hat{\omega}_t + 2\sqrt{-2\sigma \log \left[\frac{\varphi_\sigma}{\varphi_{\max}} \right] \frac{\nabla\varphi_\sigma}{|\nabla\varphi_\sigma|}}, \quad (10)$$

where $\varphi_{\max} = \max_{\omega \in \Omega} \varphi_\sigma(\omega|N)$. The estimation process (10) is activated at each segment $\hat{\mathbf{v}}_t$ consistent with one of $g_\sigma(\omega|n)$, $n \in N$; this implies that the landmark search process is controlled by the dynamical system (9) to confine within a restricted region $d\hat{\Omega}_t$.

The performance of the detection scheme is illustrated in Fig. 3 where a series of GPS track data (\mathbf{v}_t) is matched with landmark set; in this case, the distribution of 18 landmarks are allocated in a cut of satellite image of 640×480 resolution to yield the field information φ_σ in subwindow (a) where the GPS track consisting of 336 $\hat{\mathbf{v}}_t$ -vectors is matched with φ_σ . The transition of Λ_t is displayed in (b); in this subwindow, the level of background noise χ_Ω and $d\hat{\Omega}_t$ -generation time are indicated by thin white line and large white circles, respectively; the scene image of the latest landmark is displayed in the subwindow (c). Throughout the GPS track analysis, a route graph spanning 5 landmarks was generated as an access path to the future scenes along roadway pattern.

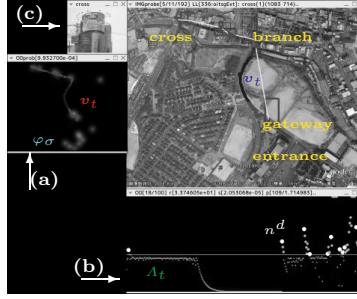


Fig. 3. GPS-based Route Graph Generation

4 Saliency Indexing for Scene Feature Detection

In many practical roadway scenes, various types of objects to be recognized are randomly distributed in complex background. This implies that the probed images should be articulated into a not-yet-identified set of landmark objects to adapt the on-vehicle vision to the future scenes. To this end, the probe vehicles are required to recognize the diversity of naturally complex scenes in terms of transferable information.

It should be noted that the perception of chromatic diversity is essentially mental processes supported by stochastic neuronal dynamics [2]. To visualize the performance of the mental process, let the chromatic diversity of naturally complex scenes be represented in a linear color space

$$\Gamma \ni \gamma = e^{\text{RGB}} \phi, \quad e^{\text{RGB}} = [e^{\text{R}} \quad e^{\text{G}} \quad e^{\text{B}}],$$

where $e^{(\cdot)} = [\cos \theta_{(\cdot)} \quad \sin \theta_{(\cdot)}]^T$ with $\theta_{\text{R}} = \pi/2, \theta_{\text{G(B)}} = \theta_{\text{R}} + (-)2\pi/3$. In this color space, the chromatic diversity is visualized as a random aggregation of δ -measure $\chi_s = \sum_{\phi \in s} \delta_{\gamma(\phi)}$ as well as the N -allocation. By identifying the

distribution χ_s with a degenerate version of a not-yet-identified fractal attractor, we can design a set of fixed points of an iterated function system [1]; through the adaptation of the fractal dynamics to a non-degraded version of the distribution χ_s , a global features, called *as-is* primaries $\hat{H} = \{\hat{\pi}_i\}$ [6], are recognized.

According to the fractal visualization of the color perception process, the chromatic diversity spanning the roadway scenes should be accepted via parametric restoration of a fractal attractor; hence, the complexity of ‘neuronal computation’ can be evaluated in terms of the *as-is* primary. To this end, first, the probability of the neuronal photopigment selection is evaluated by $p(\phi|\hat{\pi}_i) = (\phi^T \hat{\pi}_i)^2$. Next, the *a posteriori* probability of the neuronal selection under the observation ϕ is evaluated by

$$p(\hat{\pi}_i|\phi) = p(\phi|\hat{\pi}_i) \frac{p(\hat{\pi}_i)}{\sum_{\hat{\pi}_i \in \hat{\Pi}} p(\phi|\hat{\pi}_i) p(\hat{\pi}_i)}, \quad p(\hat{\pi}_i) = \frac{1}{\|\hat{\Pi}\|}.$$

As the result, finally, we can index the complexity of the neuronal process in terms of the Shannon's entropy:

$$\mathcal{S}_\phi = - \sum_{\hat{\pi}_i \in \hat{\Pi}} p(\hat{\pi}_i|\phi) \log p(\hat{\pi}_i|\phi), \quad (11)$$

The complexity index (11) is applied to the extraction of saliency distribution in roadway scenes as demonstrated in Fig 4; the fluctuation of γ -representation arising in the scene image Fig. 2(a) is displayed in (c); resulted distribution χ_s is identified with a fractal attractor to specify a scene specific deviation of the primary as indicated in (b). By using the evaluation (11), the following pixel-wise filter is designed and applied to the distribution $\phi(f^{\text{RGB}})$ at $\omega \in \Omega$:

$$\psi_\omega = 1 - \exp \left[-\frac{1}{2} \left(\frac{\log \|\hat{\Pi}\| - \mathcal{S}_{\phi_\omega}}{\log \|\hat{\Pi}\| - \mathcal{S}_m} \right)^2 \right], \quad (12)$$

where \mathcal{S}_m designates the mean value of $\mathcal{S}_{\phi_\omega}$ on entire Ω . In this filtering, the computational cost for neuronal color preference is evaluated in terms of \mathcal{S}_ϕ to emphasize the 'easy-to-select' pixels. As demonstrated in the main window, Fig. 4(a), the ψ_ω -filtering is effective to extract random distribution of landmarks to be noted by on-vehicle vision and inherent perception as well. The result of RGB-based recognition is displayed in Fig. 5 where the ψ -distribution is evaluated in terms of conventional RGB primary; the comparison with the RGB-based recognition demonstrates that the *as-is* primary based saliency distribution provides more sensitive filter to 'low-keyed' landmarks, in particular. Thus, we can exploit the *as-is* primary to control the focus of the on-vehicle vision.

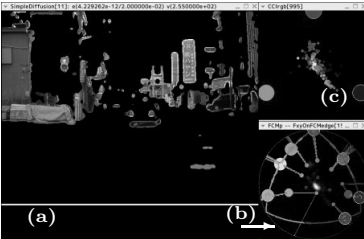


Fig. 4. *As-is* Primary Based Saliency Pattern



Fig. 5. RGB-based Recognition

5 Transferability of *As-is* Primary

To support the over-the-horizon cooperation, the *as-is* primary should be transferred to future visitors spanning significant discrepancy of photographing conditions and object allocation. The robustness of the *as-is* primary is demonstrated through experimental studies. Figure 6 indicates a part of recognition results where the *as-is* primary extracted in the branch scene, Fig 2(a), is applied to various perspectives of the cross scene; these results show that the *as-is* primary can be transferred to distant scenes spanning considerable discrepancy in photographing conditions and physical degeneration processes. Figure 7 illustrates another result where the *as-is* primary extracted in previously probed image of cross scene; these results demonstrate that the *as-is* primary provided by probe vehicles can be re-used by future visitors to the scenes.



Fig. 6. Transferability of *As-is* Primary: branch→cross

The effectiveness and transferability of the *as-is* primary is evaluated in terms of the complexity arising in the saliency distribution (\cdot) given by

$$\mathcal{S}_{(\cdot)} = -\frac{1}{C_{(\cdot)}} \int_{\Omega} (\cdot)_{\omega} \log(\cdot)_{\omega} d\omega + \log C_{(\cdot)}, \quad (13)$$

where $C_{(\cdot)}$ denotes the normalization constant. The experimental results are summarized in Table 1 where the reduction of the computational complexity is evaluated in terms of the difference between the Shannon's entropy with respect to the uniform distribution \mathcal{S}_{\emptyset} , i.e., $d\mathcal{S}_{(\cdot)} = \mathcal{S}_{\emptyset} - \mathcal{S}_{(\cdot)}$. In this table, saliency distribution and the gray level distribution are invoked to yield \mathcal{S}_{ψ} and \mathcal{S}_G via the evaluation (13), respectively; the distribution ψ is computed by using the *as-is* primary detected in each scene image; ψ^b and ψ^p are associated with perspective images transferred from branch and cross-p scenes, respectively; ψ^{Δ} designates the saliency distribution based on RGB primary. As shown in this table, the ψ_{ω} -filter well concentrate the information distributed in the image plane into a set

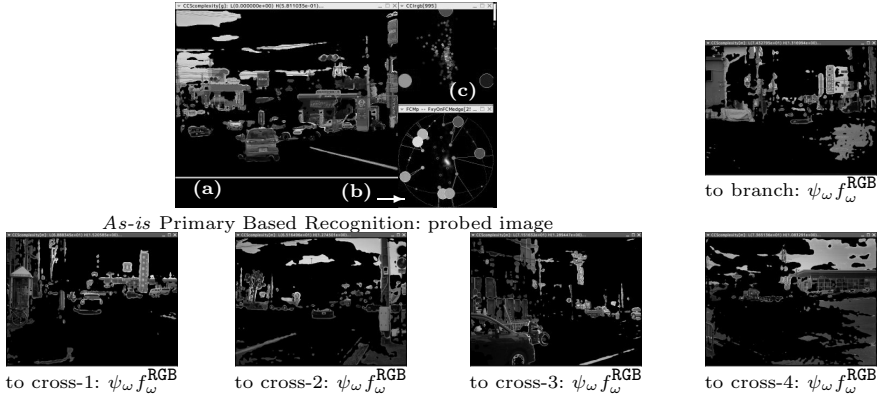


Fig. 7. Transferability of *As-is* Primary: probe→visitor

of saliency patterns to reduce the essential length of focus control programs to $e^{-1.3} - e^{-0.42}$ of random search in image plane; by this concentration, on-vehicle vision can reduce the delay to activate object recognition processes no less than 30% – 72% of focus control based on gray level distribution. The comparison of dS_{ψ} - and dS_{ψ^b} -results implies that future visitors can concentrate the focusing area to 6–68% of saliency patterns in destination images via the prediction by using observed scene image; this implies that the forward transfer of the *as-is* primary along the simulated GPS track yields a efficient focus control to scene-transversal object images; while the reference to probed image can be exploited as an effective backup in the sampling-based palette generation. In comparison with the RGB based complexity reduction, the saliency indexing based on native- and/or transferred *as-is* primary was proved to maintain sensitivity to considerably ‘low-keyed’ object images.

Table 1. Relative Complexity Reduction

scene	dS_{ψ}	dS_{ψ^b}	dS_{ψ^p}	$dS_{\psi^{\triangle}}$	$\ \hat{I}\ $	dS_G
branch	1.342628	–	0.310515	1.435720	7	0.080535
cross-1	0.583304	0.640835	0.514303	2.276305	8	0.138044
cross-2	0.829417	0.900230	0.474707	1.828937	4	0.162454
cross-3	0.421347	0.947673	0.454722	1.576738	4	0.182304
cross-4	0.434097	0.542789	0.290063	0.577541	5	0.131293
probed	0.587654	1.315543	–	2.084636	7	0.115865

As demonstrated in Fig. 4, we can exploit the ψ_{ω} -filter for the extraction of the saliency distribution including various types of objects to be observed; in contrast with the conventional method, resulted $\psi_{\omega} f_{\omega}^{\text{RGB}}$ -pattern is controllable in terms of the *as-is* primary. The transferability of the *as-is* primary makes it

possible to anticipatively control the focus on the saliency patterns *prior to* physical access as demonstrated in Figs. 6 and 7. Despite that the ψ_ω distribution is not sufficient for unique identification of natural objects, we can significantly reduce the complexity of decision steps by machine vision exhibiting man readable visualization.

6 Concluding Remarks

Complexity-based image analysis and GPS-based route graph generation jointly provide an access path to destination images through satellite-roadway-vehicle network. By simulating neuronal computation process of chromatic diversity in terms of a fractal dynamics, the *as-is* primary can be detected to extract saliency patterns in noisy background. Experimental results demonstrated that we can recognize and transfer the *as-is* primary to control the focus of machine vision to various types of man readable saliency patterns.

References

1. Barnsley, M.F., Demko, S.: Iterated function systems and the global construction of fractals. The Proceedings of the Royal Society of London A-399, 243–275 (1985)
2. Ebner, M.: Color Constancy. John Wiley and Sons, West Sussex (2007)
3. Itti, L., Kock, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-20(11), 1254–1259 (1998)
4. Kamejima, K.: GPS/GNSS residual analysis via competitive-growth modeling of ionosphere dynamics. In: Signal Processing XIII: Proceedings of the 13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey, pp. WedAmOR4–2(4 pages). EURASIP (2005)
5. Kamejima, K.: Generation and adaptation of transferable roadway model for anticipative road following on satellite-roadway-vehicle network. SICE Journal of Control, Measurement, and System Integration 4(2), 97–104 (2011)
6. Kamejima, K.: Multi-fractal articulation of environmental saliency for contextual visualization of naturally complex scenes. International Journal of Innovative Computing, Information and Control 8(3B), 2233–2248 (2012)
7. Kwak, S., Ko, B., Byun, H.: Automatic salient-object extraction using the contrast map and salient points. In: Proceedings Pacific-Rim Conference on Multimedia, pp. II:138–II:145 (2004)
8. Tan, H.-S., Huang, J.: DGPS-based vehicle-to-vehicle cooperative collision warning: Engineering feasibility viewpoints. IEEE Transactions on Intelligent Transportation Systems ITS-7(4), 415–428 (2006)
9. Wang, J., Schroedl, S., Mezger, K., Ortloff, R., Joos, A., Passegger, T.: Lane keeping based on location technology. IEEE Transactions on Intelligent Transportation Systems ITS-6(3), 351–356 (2005)
10. Wong, E.: Stochastic Processes in Information and Dynamical Systems. McGraw-Hill, New York (1971)