Representation and Extraction of Multi-Scale Image Feature Associated with Maneuvering Affordance

Kohji Kamejima
Faculty of Engineering, Osaka Institute of Technology
5-16-1 Omiya, Asahi, Osaka 535-8585 JAPAN
Phone: +81(Japan)-6-6954-4330 Fax: +81(Japan)-6-6952-6197
e-mail: kamejima@dim.oit.ac.jp

Abstract: A detection scheme is presented for maneuvering affordance in noisy imagery. Under the assumption that image features to be observed are generated as fractal attractors induced by mental maneuvering process, environment features are represented in directional scale images. For eliminating background noise in this sampling process, pixelwise filter is designed based on probabilistic complexity analysis. The detectability of affordance patterns has been verified through experimental studies.

Keywords: Maneuvering Affordance; Directional Scale Image; Environment Features; Noise Level Estimation; Fractal Pattern Detection

1 Introductory Remarks

Visual perception organizes randomly distributed image features into cues to subsequent decisions. One of the essential capabilities of ‘real world intelligence’ is to activate feasible operations based on the belief on real world. Such phenomenal supervenience to reality is expected to be implemented in artificial real world intelligence intended for field automation [8], social safety [2] and interactive education [1], e.g.

The real world intelligence is essentially intentional within the framework of a priori knowledge; It applies partial description of the environment to encountered scene in generating action plan. The real world intelligence, simultaneously, is nondeterministic within the environment. The decision process is activated prior to the comprehension of the situations. Such intentional – nondeterministic scheme provides logical basis for the existence of the real world intelligence.

The intentional – nondeterministic scheme, on the other hand, introduces serious contradiction: self-referential structure in decision process. To attain admissible goals within the intentional – nondeterministic scheme, real world intelligence is required to maintain decision space in consideration of possible versions of the world. For this purpose, environment should be described in terms of symbols associated with distributed objects. To a priori endorse planning steps to be resulted, the symbolic description should convey enough information for restoring scene image. The capability of the computation on a posteriori definite environment description, the logical consequence of the supervenience to real world, requires chicken-and-egg process in implementing artificial vision systems.

Physically, the object distribution in natural scene establishes complete description of environment. This description should be essentially reconfigurable so that all observers can afford to make their own consistent decisions. As an as-is part of real world, thus, natural environment itself maintains consistency despite of continuous and simultaneous maneuvering of observers. Following ecological optics [3], in fact, perception process is supported by the texture distribution in natural scene. The description of environment is directly observed as the distribution of optical array filling up the scene. The information conveyed by such ecological optics transfers the implications of environment as the affordance: the message to be captured for decision makers.

Following the consideration of Gibson, the process of ‘seeing’ is nonintentional in contrast with intentional decision process. By looking around in the cockpit, for instances, the human pilot recognizes the situation of satisfactory flight as well as a bird maintaining stable flight in complex landscape. Usually, the affordance is considered to be sufficient for decision maker to recognize the problem to be solved in encountered situations. To provide such informative affordance, the nature is assume to compute entire affordance in the ‘light-speed’ via the transformation of optical array.

As one of basic processes in environment understanding, the binding between intentional locomotion and optic array provides innate description of ‘maneuverable area’. For example, the texture on a roadway yields a
Maneuvering area is specified in terms of random distribution of image features and observed with many "distracting" objects. Despite the unexpectability of image features, the distribution is regulated by general geometric rule, i.e., perspective projection. Due to the generality, the feature distribution exhibits complex appearance. To concentrate stable maneuvering, perception channels are required to evaluate the expansion of image features of encountered roadway in "background noise" [6].

distribution of similar patterns as shown in Fig. 1. Although the position of each pattern on the roadway in image plane is dependent on the viewpoint, the proportion of patterns is well regulated through perspective projection.

This proportionality, thus, is available as a robust representations for general classes of the roadway scene. This abstract representation implicitly specifies the expansion of maneuverable area. Despite abstract representation, such implicit specification can be identified with generic dynamics governing the visual complexity. Since this visual complexity is not explicitly dependent on a specific texture pattern, we can discriminate the roadway pattern from a not-yet-identified texture by testing the proportionality in the observed image.

2 Evoked Mental Dynamics Model

The existence of the affordance provides logical basis for information processing model of visual perception. Following Neisser [9], for instance, perception is a dynamic process that updates internal description of surroundings, called scheme, through continual matching with observation.

Dynamic perception system intentionally observe the scene by projecting the images of environment. Visual perception converts the totality of optical array into brightness distribution. The observation is completed through matching the brightness distribution with the projection. Intentional access to environment information, simultaneously, induces mental motion on image plane. By defining the view with respect to maneuvering observer as shown in Fig. 2, X-Y coordinate system on image plane is associated with Width-Depth information of the scene. Such intentional coordinate system provides 3D interpretation of not-yet-structured image feature distribution.

Furthermore, the knowledge of perspective projection combined with mental motion within the schema evokes the following explorative access to the scene image:

(i) Recall the roadway pattern in current schema
(ii) Reduce the pattern to generate in situ template for scale-independent image features
(iii) Match the top of reduced pattern to recognize the extension of maneuverable area.
(iv) Shift reduced pattern to left and right sides to reconfirm the width of steps.

If reduced patterns cover target roadway without any overlapping, the imaging process generate complex pattern with minimal length of program [5]. Such minimal program was proved to generate fractal attractor:

Proposition 1 (Open Set Condition) [4] Assume that
the pattern $\Lambda$ satisfies the open set condition with respect to a set of contraction mappings $\nu = \{\mu_i : \Omega \rightarrow \Omega\}$, i.e.,

$$\Lambda_\nu = \bigcup_{\nu} \mu_i(\Lambda_i) \subset \Lambda,$$

with

$$\mu_i(\Lambda_i) \cap \mu_j(\Lambda_j) = \emptyset,$$

for any $\mu_i, \mu_j \in \nu$.

Then there exists a fractal attractor $\Xi \subset \Lambda$ satisfying the following condition:

$$\Xi = \bigcup_{\nu} \mu_i(\Xi).$$

Thus, underlying proportionality in scene imagery induces a dynamical system for 'computing' a self-similar pattern based on compact a priori model.

The existence of the self-similarity implies the visibility of scene image in terms of fractal attractor. In roadway scene, target attractor $\Xi$ is embedded in noisy background as shown in Fig. 1. For detecting target pattern, we can exploit the capturing probability $\varphi$ of not-yet-identified attractor. Within the framework of entropy maximization, the capturing probability $\varphi$ is given as the solution to the following elliptic equation:

$$\frac{1}{2} \Delta \varphi(\omega|\nu) + \rho(\chi_\Xi - \varphi(\omega|\nu)) = 0, \quad (1)$$

where $\rho = \log \|
u\|; \|\nu\|$ is the size of mapping set $\nu = \{\mu_i\}$, and $\chi_\Xi$ denotes the invariant measure associated with the self-similarity process. Despite the generativity of unexpected attractor image, hence, the computation process can be specified in terms of complexity factor $\rho$. By a priori evaluating imaging complexity of not-yet-identified self-similarity process, we have

**Proposition 2 (Pointwise Filter) [7]** Assume the background noise $\chi_\Omega$ is uniformly distributed in the image plane $\Omega$ and suppose that observed measure $\chi_\Lambda$ is represented by

$$\chi_\Lambda = \chi_\Xi + \chi_\Omega. \quad (2)$$

Then the noise level is estimated by

$$\tilde{\gamma} = C_{\Omega}^\varphi \tilde{\rho}_\nu, \quad (3a)$$

where

$$C_{\Omega}^\varphi = \int_{\Omega} \varphi(\omega|\nu) d\omega, \quad (3b)$$

$$\log \tilde{\rho}_\nu = 1 - \frac{1}{2} (1 - e^H - R e^{-D}) - \tilde{H}. \quad (3c)$$

The combination of the random feature distribution with intentional coordinate system implies that the maneuvering affordance can be extracted from scene image without a priori geometric model. As a universal dynamics, self-similarity processes can afford to generate arbitrary visible information in the scene. The elimination of background noise, thus, results in the detection of a region supporting not-yet-identified fractal attractors.

### 3 Directional Scale Imaging

Consider the transformation of the brightness distribution into the measure $\chi_\Lambda$. Let $f(x,y), (x,y) \in \Omega$ be the gray level distribution in image plane $\Omega$. For square summable distributions, generally, we have the following representation in associated frequency space:

$$F(\omega_x, \omega_y) = \int_{-\infty}^{\infty} \chi_\Lambda(\omega_x, y)e^{-\omega_y y} dy,$$

$$\chi_\Lambda(\omega_x, y) = \int_{-\infty}^{\infty} f(x,y)e^{-\omega_x x} dx,$$

where $\omega_x$ and $\omega_y$ denote spatial frequencies associated with $x$ and $y$ coordinates, respectively. It is well known that the complete information conveyed by the representation $F(\omega_x, \omega_y)$ satisfies the following scaling rule:

**Proposition 3 (Similarity)** Suppose that

$$x = k(y)x_0,$$

where $k(y) > 0$, for any $y$. Then

$$\chi_{k(y)x}(\omega_x, y) = k(y)^{-1}\chi(k(y)^{-1}\omega_x, y). \quad (4)$$

Define scale parameter $(\lambda_x, \lambda_y)$ by

$$\lambda_x = \frac{1}{\omega_x},$$

$$\lambda_y = \frac{1}{\omega_y}.\quad (5)$$

Then, in the image plane fixed to the maneuverer (Fig. 2), complete information conveyed by the feature distribution has the following scale-image representation:

**Proposition 4 (Directional Scale Image)** Let $f \in L^2$ be the gray level distribution in the image plane $\Omega$. Then, we have the following representation

$$F(\lambda_x, \lambda_y) = \int_{-\infty}^{\infty} \chi_\Lambda(\lambda_x, y) \exp \left[ -\frac{\tilde{y}}{\lambda_y} \right] dy,$$

with respect to the horizontal and vertical scales, $\lambda_x$ and $\lambda_y$, where $\chi(\lambda_x, y)$ denotes the directional scale image...
with respect to the horizontal scale $\lambda_x$ and vertical position $y$, given by

$$\chi_L(\lambda_x, y) = \int_{-\infty}^{\infty} f(x, y) \exp \left( -\frac{jx}{\lambda_x} \right) dx.$$ 

Let the distributions with respect to $(x, y)$ and $(\lambda_x, y)$ are imaged on common plane $\Omega$. By definition, then, the aggregation of the directional scale image

$$\mathcal{S} = \left\{ \chi(\lambda_x, y) \mid (\lambda_x, y) \in \Omega \right\},$$

represent the totality of the information

$$\mathcal{I} = \left\{ f(x, y) \mid (x, y) \in \Omega \right\}.$$ 

Based on induced mental dynamics (i) - (iv), we can extract a sequence of line images following the proportionality. For this purpose, the scale of the image features are normalized by the width of image plane. By this normalization, the self-similarity of image features is represented in terms of the linear proportionality in scale parameter $\lambda_x$. Thus, we have the following representation of maneuvering affordance in terms of a directional scale image:

**Proposition 5 (Similarity)** Suppose that

$$x = k(y)x_0,$$ 

where $k(y) > 0$, for any $y$. Then

$$\chi_{k(y)x}(\lambda_x, y) = k(y)^{-1} \chi(k(y)\lambda_x, y).$$ 

Hence, the detection of maneuvering affordance of self-similarity results in the analysis of the similarity (6) in a directional scale image.

4 Experiments

Detection scheme of maneuvering affordance has been verified through experimental studies. A part of results is shown in Figs. 3 - 7. In these results, random dots on a flat plane were observed as shown in Fig. 3 then digitized in a buffer of 640 x 480 resolution.

$$f = 0.299 \times [R] + 0.587 \times [G] + 0.114 \times [B].$$ 

By the perspective projection, we can expect that the texture is governed by the following scaling rule:

$$x \rightarrow k(y)x, \quad y < 480,$$ 

$$k(y) = \begin{cases} 1 - \left( \frac{y}{y_0} \right); & \text{for } y < y_0, \\ 0; & \text{otherwise.} \end{cases}$$

By the perspective projection, we can expect that the texture is governed by the following scaling rule:

$$x \rightarrow k(y)x, \quad y < 480,$$ 

$$k(y) = \begin{cases} 1 - \left( \frac{y}{y_0} \right); & \text{for } y < y_0, \\ 0; & \text{otherwise.} \end{cases}$$

In this figure, power spectra of line images are arranged in scanning order.
By invoking the maximal entropy criterion, the level of the background noise was estimated as $\gamma$. The output filter separates dominant points in directional Fourier images from background noise. By this separation, background noise with higher spatial frequency has been eliminated. Despite such reduction, the depth information was not deformed.

As indicated in this figure, the flat zone is characterized by the power spectrum exhibiting frequency shift. In addition, the depth of the flat zone is shown to be determined in terms of the intensity of power spectra clearly.

For specifying the depth of flat zone, the power spectra was converted into binary images. In figure 5, the detection level were adjusted to the level $\gamma$ given by (3). By evaluating the mapping size $\|\nu\| = 3$, the filtering level $\gamma$ was evaluated as follows:

$$\gamma = 14.3,$$

for this version of observation. In this binary image, structured area and noisy background are discriminated in terms of the filtering level $\gamma$.

As demonstrated in figure 5, directional Fourier images detect the environment feature specifying the flat zone. The experimental results demonstrate that the dominant areas of the directional Fourier images are available as robust perception of a flat zone from the current position. Thus, we can available binary image of the directional Fourier images as a version of affordance for inducing stable maneuvering.

Figure 6 shows a directional scale image generated by the random dot image. In this figure, power spectra of line images with respect to scale parameter $\lambda_x$ are arranged in scanning order. As indicated in this figure, the flat zone is characterized by the power spectrum exhibiting scale shift. In addition, the depth of the flat zone is shown to be determined in terms of the intensity of power spectra clearly.

For specifying the depth of flat zone, the power spectra was converted into binary images. In figure 7, on the other hand, the detection level were adjusted to the level $\gamma$ given by:

$$\gamma = 81.0.$$

In this binary image, structured area and noisy background are discriminated in terms of the filtering level $\gamma$.

As demonstrated in figure 7, directional scale images detect the environment feature specifying the flat zone. In contrast with the directional Fourier image 4, furthermore, the proportionality is represented by a triangle with random texture. This implies that the feature distribution can be coded in terms of mapping set generating fractal collage of the maneuvering feature. The experimental results demonstrate that the dominant areas of the directional scale images are available as robust perception of a flat zone from the current position. Thus, we can available binary image of the directional scale images as a version of affordance for inducing stable maneuvering.

5 Concluding Remarks

Directional scale transform was applied to extract maneuvering affordance in noisy imagery. By identifying
By invoking the maximal entropy criterion, the level of the background noise was estimated as $\tilde{\gamma}$. The output filter separates dominant points in directional scale images from background noise. By this separation, background noise with large scale parameter has been eliminated. Despite such reduction, the depth information was not deformed.

the brightness distribution of observed patterns with the capturing probability of unknown fractal attractor, pointwise filters are introduced for eliminating background noise. The detectability of affordance patterns has been verified through experimental studies.

References


