**Dynamic Markov random fields for stochastic modeling of visual attention**

Akisato Kimura (1), Derek Pang (2), Tatsuto Takeuchi (1), Junji Yamato (1), Kunio Kashino (1)

(1) NTT Communication Science Laboratories, NTT Corporation (2) School of Engineering Science, Simon Fraser University

**Background:** Developing an accurate computational model of human visual system is a long-standing challenge.

→ Such a model enable any system to select just relevant information from a complex and cluttered visual scene.

**Problem:** Different people may attend to different regions at the same time!

→ However, most previous computational models only select a fixed attended location every time for the same input. (cf. [Itti 1998] [Frintrop 2005])

**Contributions:** A stochastic model of saliency-based human visual attention with a dynamic Bayesian network to automatically predict the likelihood of where humans typically focus on, only from an input video

1. Simulate and combine the visual saliency response and cognitive state
2. Introduce a dynamic Markov random field (dynamic MRF) model to describe a spatiotemporal property of saliency responses.

---

**Proposed stochastic model**

A stochastic model of saliency-based human visual attention with a dynamic Bayesian network

1. **A cognitive state** that governs the strategy of eye movements
2. **A density map** that indicates the probable human-attended regions
3. **Saliency responses** perceived through a kind of stochastic processes
4. **Saliency** idealized as the average strength of the visual stimulus

---

**Saliency maps (SM)**

Idealize the average strength of visual stimulus

- Intensity, color opponents (red/green, blue/yellow), orientations (0, n/4, n/2, 3n/4), motion energies (horizontal, vertical)

- **Stimulus**
  - Intensity
  - Color opponents
  - Orientation
  - Motion energy

- **Saliency (deterministic) maps**

**Stochastic saliency maps (SSM)**

Saliency response through a Gaussian random process

- **Response**
  - Stochastic saliency maps
  - Saliency maps

**Dynamic Gaussian Markov random field (MRF) model**

\[ p(x_t, y_t | x_{t-1}, y_{t-1}) = \exp(\text{log likelihood}(x_t, y_t | x_{t-1}, y_{t-1})) \cdot \text{prior}(x_{t-1}, y_{t-1}) \]

- **Temporal smoothness**
- **Spatial smoothness**

- The state of the stochastic saliency map can be predicted using Kalman Filter and a (naïve) mean field approx.

---

**Experiments**

Evaluating the accuracy by comparing to human eye motions

- **Input**
  - 8 video clips with natural scenes (640x480, 15fps, 30-90sec)

- **Device**
  - 1280x1024 LCD display, 6 subjects, no specific instructions

- **Metric**
  - Normalized scan-path saliency (NSS) [Peters et al. 2007]

- **Main issue**

**Average NSS scores**

- **NSS Score for each video**

**Average execution time**

- **MFT method**
- **proposed method**

---

**Contact:** Akisato Kimura (akisato@ieee.org, http://www.brl.ntt.co.jp/people/akisato/)