

Dynamic Markov random fields for stochastic modeling of visual attention

2008年11月27日

木村昭悟 (1) Derek Pang (1,2) 竹内龍人 (1)
大和淳司 (1) 柏野邦夫 (1)

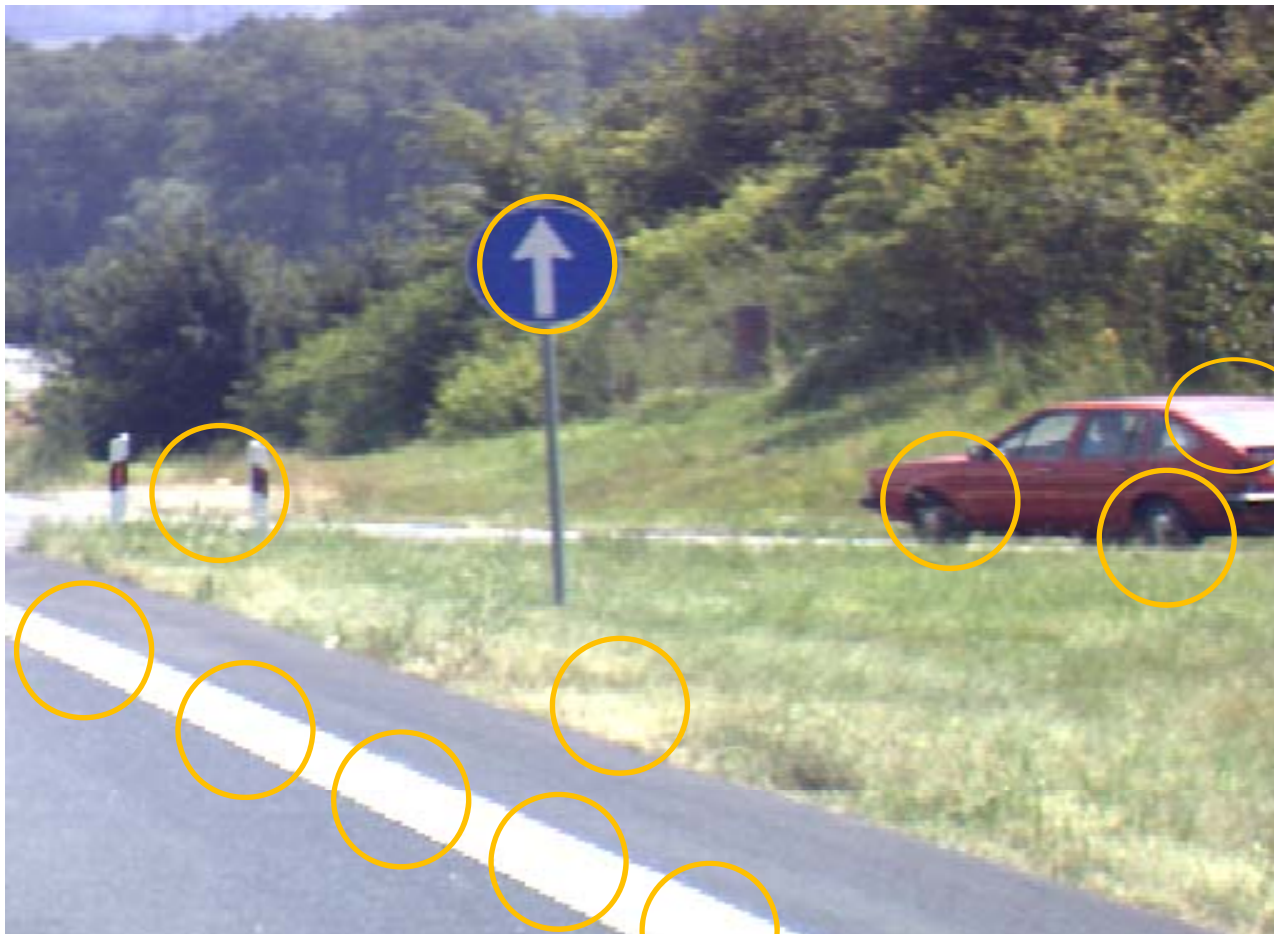
(1) 日本電信電話(株)NTTコミュニケーション科学基礎研究所
メディア情報研究部 メディア認識研究グループ



(2) Simon Fraser University
School of Engineering Science



Where would you focus?



人間は映像中から重要と思われる情報を瞬時に判断できる。

特徴統合理論

[Treisman & Gelade 1980]

- いくつかの基本的な特徴量(輝度・色など)を抽出し処理することで、各々 feature map を生成。
- Feature map を統合することで、saliency map(SM) を生成。
- Saliency map内で最も輝度値が大きくなる箇所に最初に(視覚的)注意が向けられる。



入力画像



Saliency map (extracted by [Itti et al. 1998])

特徴統合理論

[Treisman & Gelade 1980]

- いくつかの基本的な特徴量(輝度・色など)を抽出し処理することで、各々 feature map を生成。
- Feature map を統合することで、saliency map(SM) を生成。
- Saliency map内で最も輝度値が大きくなる箇所に最初に(視覚的)注意が向けられる。



入力画像

Saliency-based visual attention model

- 計算モデルの提案、心理物理学的検証 [Koch & Ullman 1984]
- 実装可能な計算モデルの提案 [Itti et al. 1998]
- その他関連研究 [Frintrop & Rome 2004] [Itti & Baldi 2005] [Leung et al. 2007]
- これらモデルの応用例:
robotics [Nagai & Rohlving 2007]
active vision [Takeuchi et al. 1997]
物体認識 [Frintrop et al. 2004]

Saliency map (extracted by [Itti et al. 1998])

従来研究の問題点

- 与えられた入力画像について決定論的にSMが計算される。
- SM内で最も輝度値が大きい領域に最初に注意が向く。
 - 同じ映像が与えられると、誰がいつその映像を見ても同じ場所に注意が向くことを主張
 - 現実の人間の行動とは異なる



入力画像



Saliency map (extracted by [Itti et al. 1998])

本研究の動機

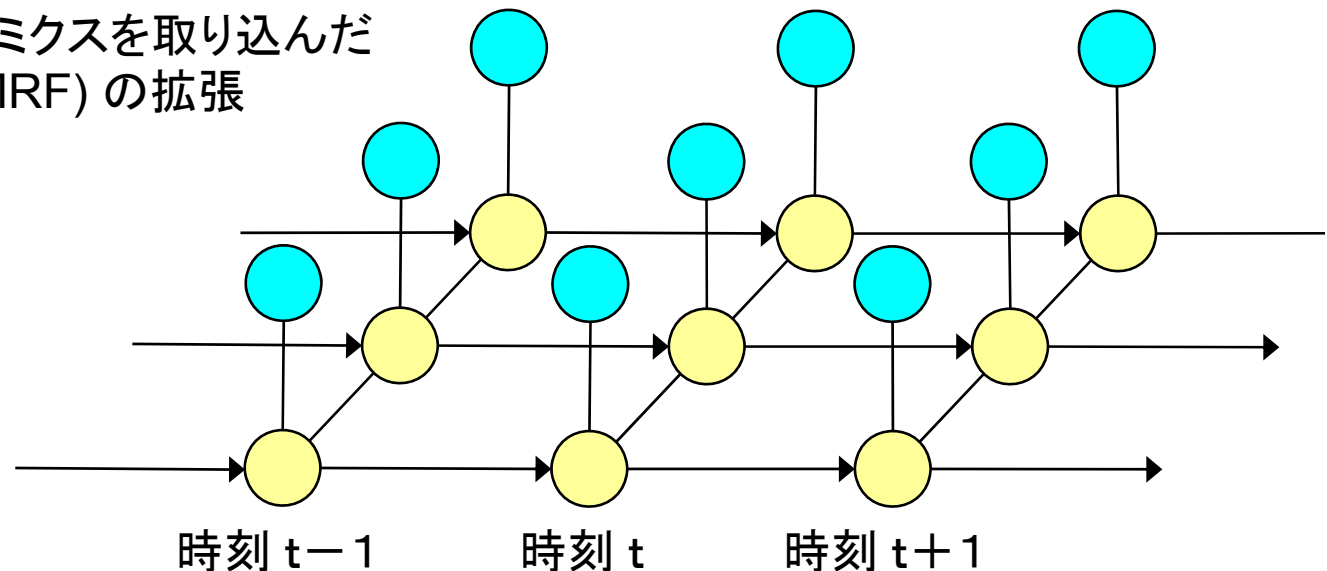
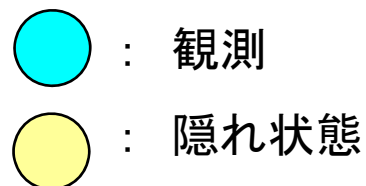
- 人間の視覚的注意の機構を確率的な挙動としてモデル化
→ 与えられた映像のみから
人間が注目しやすい領域をより正確に特定。
- 動的ベイジアンネットワークを用いた確率モデル
[Pang et al. 2008 @ PRMU June]
 - State space model と HMM を組み合わせた
ベイジアンネットワークによりモデル化
 - 人間が注目しやすい映像中の領域を自動的に推定
 - **空間的な関係性を考慮していない**
 - ← saliencyが高い箇所周辺のsaliencyが高いはず

提案モデルのポイント

- 動的マルコフ確率場によるsaliencyのモデル化
 - saliencyの時空間的な関係を統一的に記述
 - ナイーブ平均場近似により、[Pang et al. 2008] とほぼ同様のコストでsaliencyを推定

動的マルコフ確率場 (dynamic MRF)

時間方向のダイナミクスを取り込んだ
マルコフ確率場 (MRF) の拡張

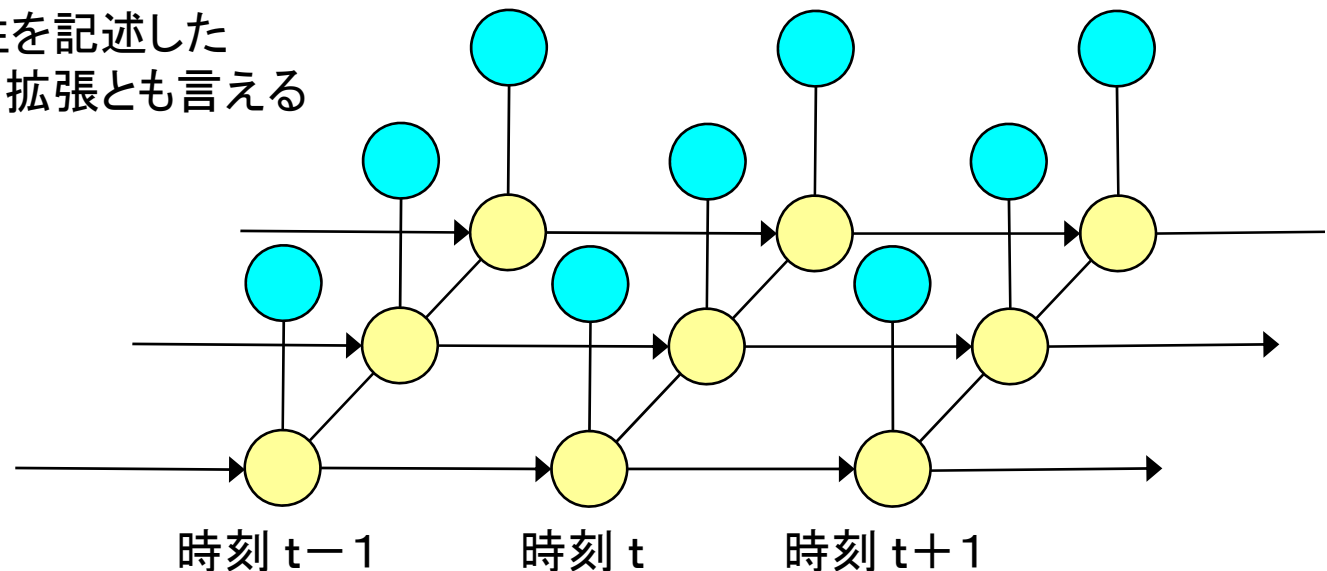
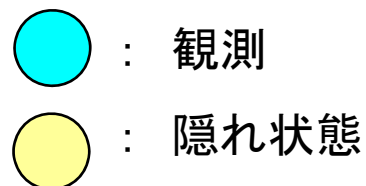


提案モデルのポイント

- 動的マルコフ確率場によるsaliencyのモデル化
 - saliencyの時空間的な関係を統一的に記述
 - ナイーブ平均場近似により、[Pang et al. 2008] とほぼ同様のコストでsaliencyを推定

動的マルコフ確率場 (dynamic MRF)

空間方向の関係性を記述した
状態空間モデルの拡張とも言える



提案モデルの概略

Top-down

Eye movement patterns (EMP)

- 視線移動の戦略を制御する人間の内部状態をモデル化（動かしたい or 動かしたくない）
- 映像入力とは独立に決定される

Eye-focusing density map

- Bottom-up/Top-down情報を統合することで、視線が向く確率の高い領域を推定

Saliency map (SM)

- 映像入力によって人間が受ける視覚刺激の強さを表現

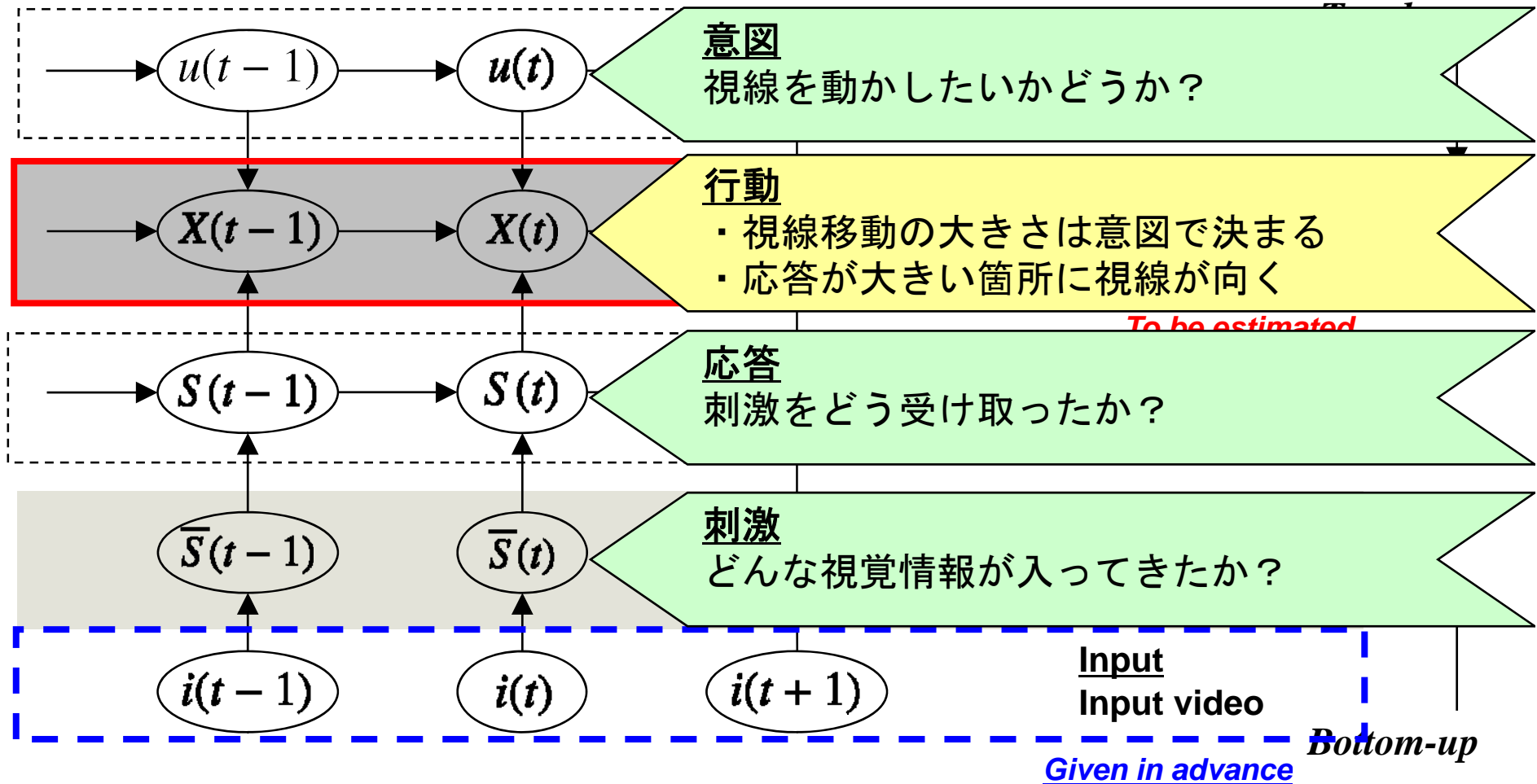
Stochastic saliency map (SSM)

- 信号検出理論 [Eckstein 2000] に基づき、刺激に対する応答をガウス分布でモデル化

Dynamic MRFの導入

Bottom-up

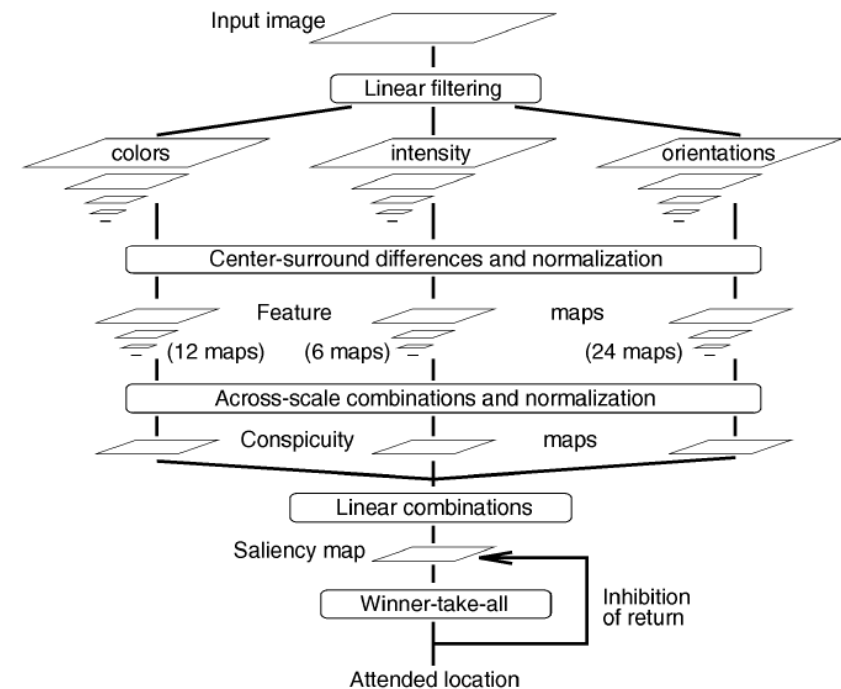
提案モデル



Saliency map の抽出

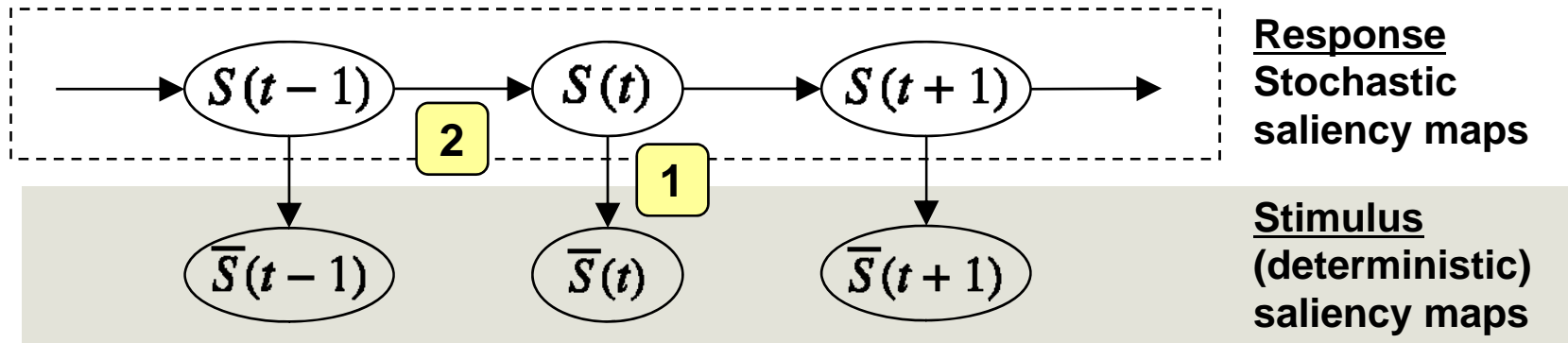
- Itti-Koch model [Itti 1998] を利用
 - 特徴統合理論に基づき、
映像の各フレームから独立にSMを生成
 - 基本特徴量の空間的なコントラストを
多重解像度処理によって抽出し統合

- 抽出に用いた基本特徴量
 - 輝度
 - 補色 (赤/緑、青/黄)
 - 方向 ($0, \pi/4, \pi/2, 3\pi/4$)
 - 運動 (水平、垂直)



Stochastic saliency map の推定

- (従来は) SMを観測とする pixel-wise state-space model



モデル

$$(1) p(\bar{s}(t, \mathbf{x}) | s(t, \mathbf{x})) = \mathcal{G}(\bar{s}(t, \mathbf{x}); s(t, \mathbf{x}), \sigma_{s2})$$

SSMがガウス分布を介し、SMとして観測される。

$$(2) p(s(t, \mathbf{x}) | s(t-1, \mathbf{x})) = \mathcal{G}(s(t, \mathbf{x}); s(t-1, \mathbf{x}), \sigma_{s1}).$$

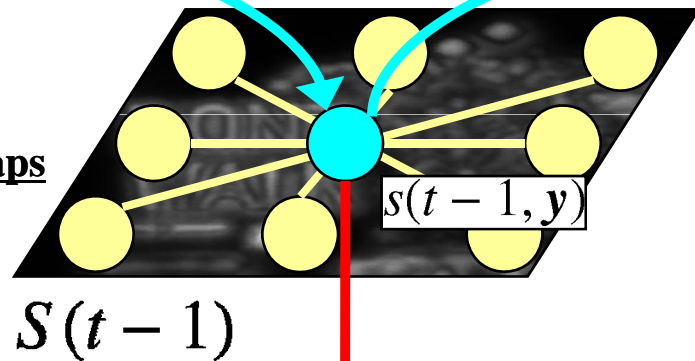
SSMの時間方向での連続性を仮定。

- 空間的依存性が考慮されていない！

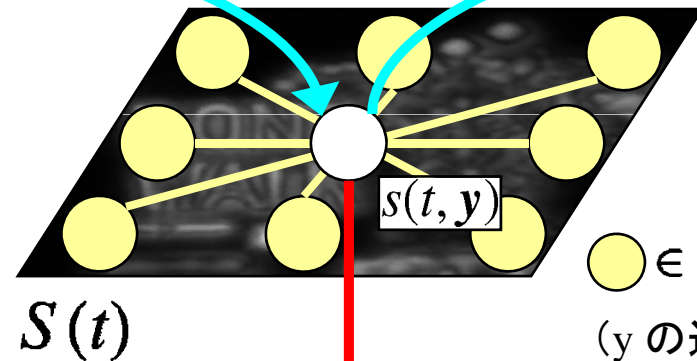
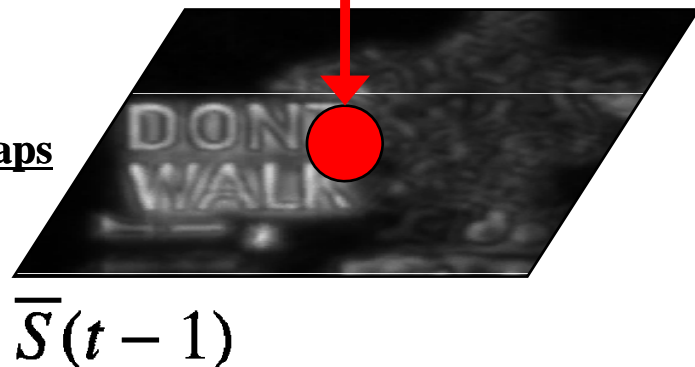
動的マルコフ確率場の導入

- SSMの時間方向での連続性を仮定。●
- SSMがガウス分布を介し、SMとして観測される。●
- SSMの空間的な連続性も同様に仮定。●

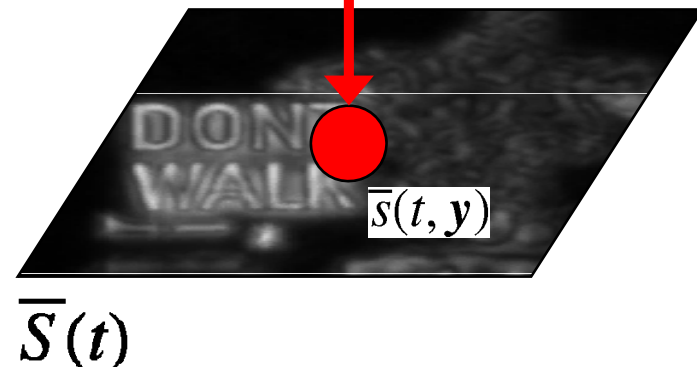
Stochastic saliency maps



Saliency maps



● $\in D(\mathbf{y})$
(\mathbf{y} の近傍)



SSM推定の定式化

- Gaussian dynamic Markov random field

モデル

$$p(S(t), \bar{S}(t) | S(t-1)) \propto \exp\{-\Phi(S(t), \bar{S}(t) | S(t-1))\},$$

$$\Phi(S(t), \bar{S}(t) | S(t-1)) = \sum_{y \in I} \left\{ \underbrace{f_1(s(t, y) | s(t-1, y))}_{\text{時間的連続性}} + \underbrace{f_2(\bar{s}(t, y) | s(t, y))}_{\text{SMを観測}} + \frac{1}{2} \sum_{\bar{y} \in D(y)} \underbrace{f_3(s(t, y) | s(t, \bar{y}))}_{\text{空間的連続性}} \right\},$$

$$f_i(s_1 | s_2) \propto -\log \mathcal{G}(s_1; s_2, \sigma_{si}), \quad (i = 1, 2, 3)$$

- Kalman filterと(ナイーブ)平均場近似を利用して推定

[Estimation step]

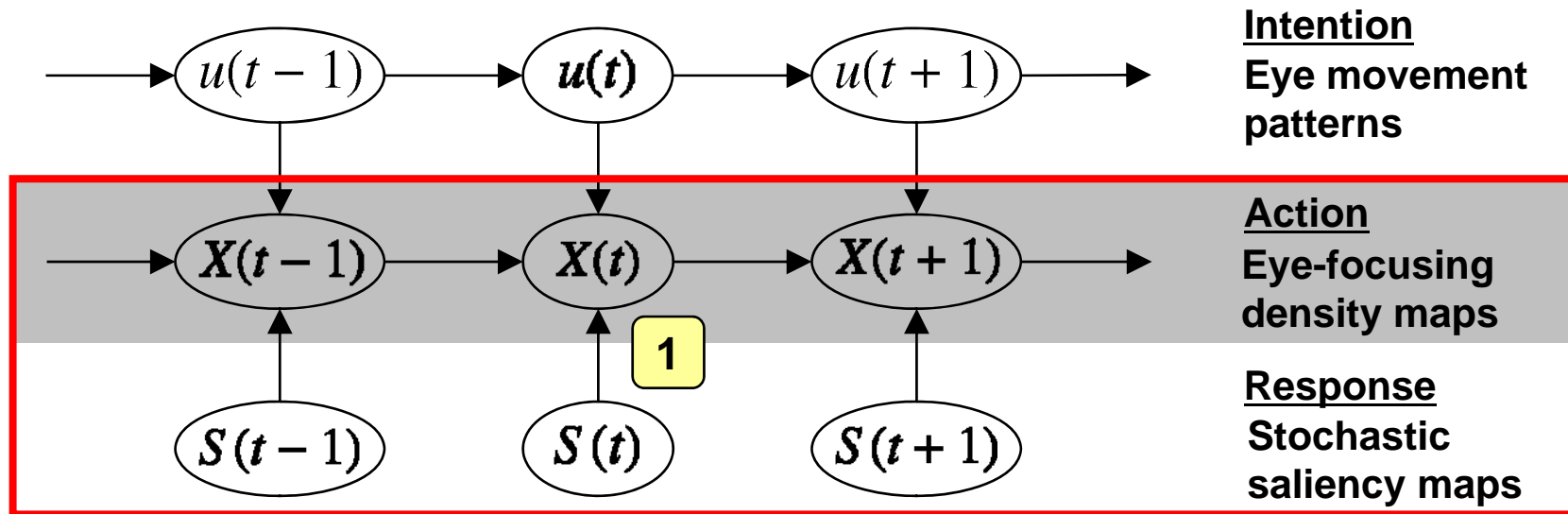
$$p(s(t, y) | \bar{S}(1:t-1)) = \mathcal{G}(s(t, y); \underbrace{\bar{s}(t, y | t-1)}_{\text{平均場近似が必要}}, \underbrace{\sigma_s(t, y | t-1)}_{\text{平均場近似が必要}}),$$

[Update step]

$$p(s(t, y) | \bar{S}(1:t)) = \mathcal{G}(s(t, y); \underbrace{\bar{s}(t, y | t)}_{\text{Kalman filterと同様}}, \underbrace{\sigma_s(t, y | t)}_{\text{Kalman filterと同様}}),$$

Eye-focusing density map の推定 (1)

- Eye movement patterns (EMP) を隠れ状態とするHMM



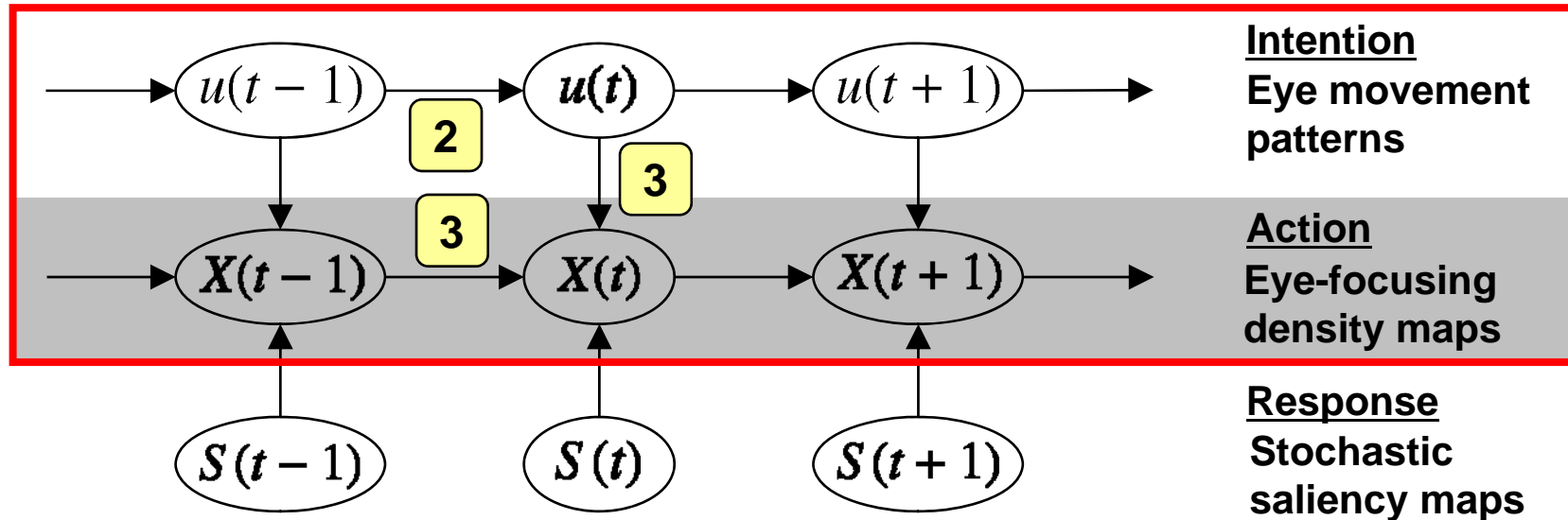
モデル

$$(1) \quad p(x(t)|p(S(t))) = \int_{-\infty}^{\infty} p(s(t, x(t))) \left\{ \prod_{\bar{x} \neq x(t)} P(s(t, \bar{x}) \leq s(t, x(t))) \right\} ds(t, x(t))$$

映像中の位置 $x(t)$ において実際に観測された応答 (=SSMの実現値)が、それ以外の位置での応答よりも大きくなるときに、位置 $x(t)$ に視線が向く。

Eye-focusing density mapの推定 (2)

- Eye movement patterns (EMP) を隠れ状態とするHMM



モデル

$$(2) \quad p(u(t)|u(t-1)) = \prod_{i=0}^1 \prod_{j=0}^1 \{\phi_{(i,j)}\}^{u(t)_i u(t-1)_j},$$

$$(3) \quad p(x(t)|x(t-1), u(t)) = \prod_{i=0}^1 \mathcal{L}(x(t); x(t-1), \gamma_{xi}, \sigma_{xi})^{u(t)_i}$$

入力と独立に遷移するEMPによって視線移動の大きさを制御

Eye-focusing density mapの推定 (3)

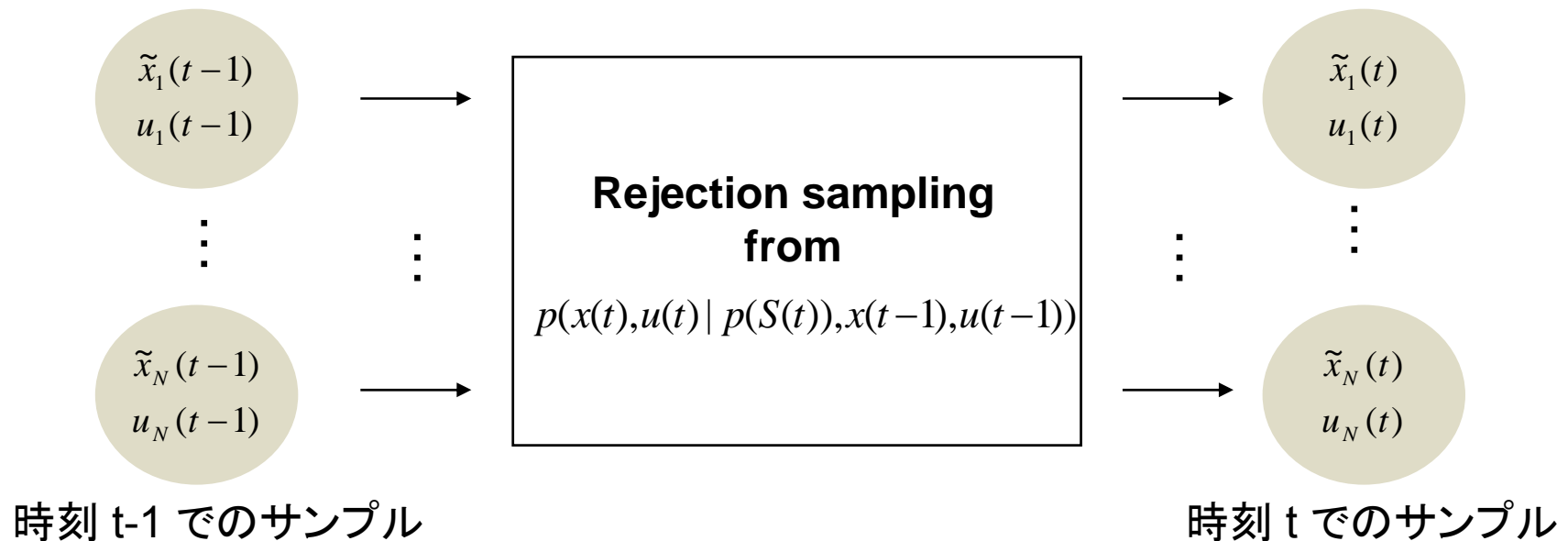
- 視線推定位置とEMPの組をサンプリングにより生成

$$p(\mathbf{x}(t), u(t) | p(S(t)), \mathbf{x}(t-1), u(t-1))$$

$$\stackrel{\text{def.}}{=} \frac{1}{Z} \underbrace{p(\mathbf{x}(t) | p(S(t)))}_{\text{Bottom-up}} \cdot \underbrace{p(u(t) | u(t-1)) \cdot p(\mathbf{x}(t) | \mathbf{x}(t-1), u(t))}_{\text{Top-down}},$$

Bottom-up

Top-down



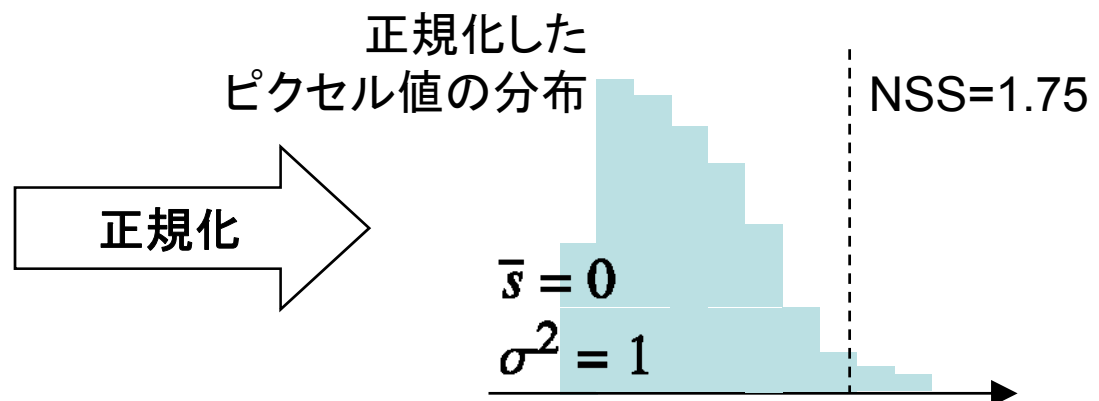
実験条件

- 6人の被験者に映像を提示、その視線位置を視線測定機器を用いて測定
 - 映像を液晶ディスプレイ上に表示
 - 視線測定機器： 角膜反射を利用、30fps [Ohno 2002]
 - 被験者の頭部は顎台によって固定
- 入力映像
 - 風景・動物等を含む自然映像8本
 - 640x480ピクセル、15fps、1本当たり30～90秒
- 映像視聴に際し、被験者への教示はなし。
- 計算機環境
 - Intel Core2 Duo E6850 3.0GHz, 3.0GB memory
 - Microsoft Visual C++ .NET, no optimization

評価尺度

- Normalized scanpath saliency (NSS)
 - ランダムな視線移動に対する有意差を測定する尺度
- 1. 出力画像のピクセル値を、平均=0、分散=1となるように正規化
- 2. 各フレームについて、被験者の視線位置での出力画像のピクセル値を抽出。
- 3. 上記ピクセル値のフレーム平均を取り、NSSを算出。

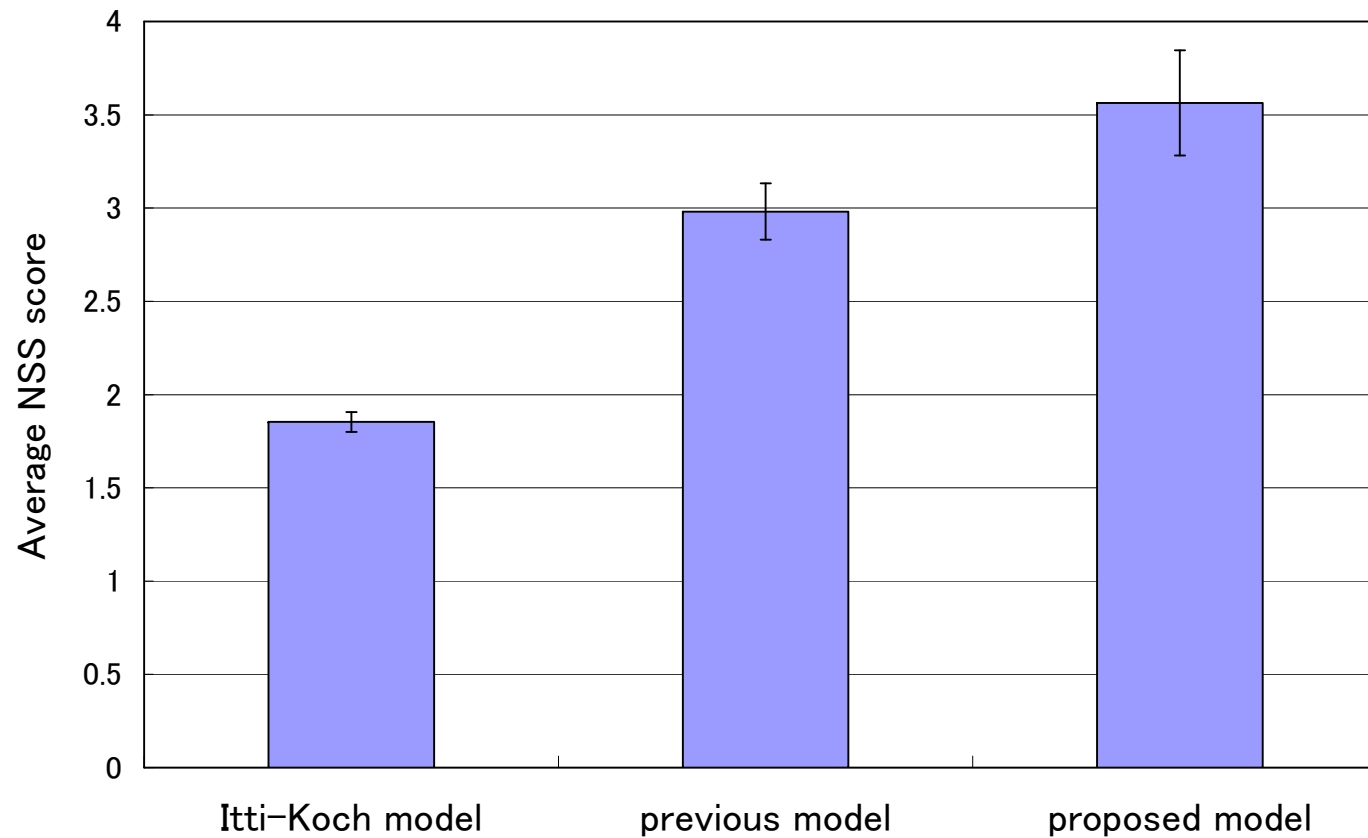
出力画像



実験結果 (1 / 3)

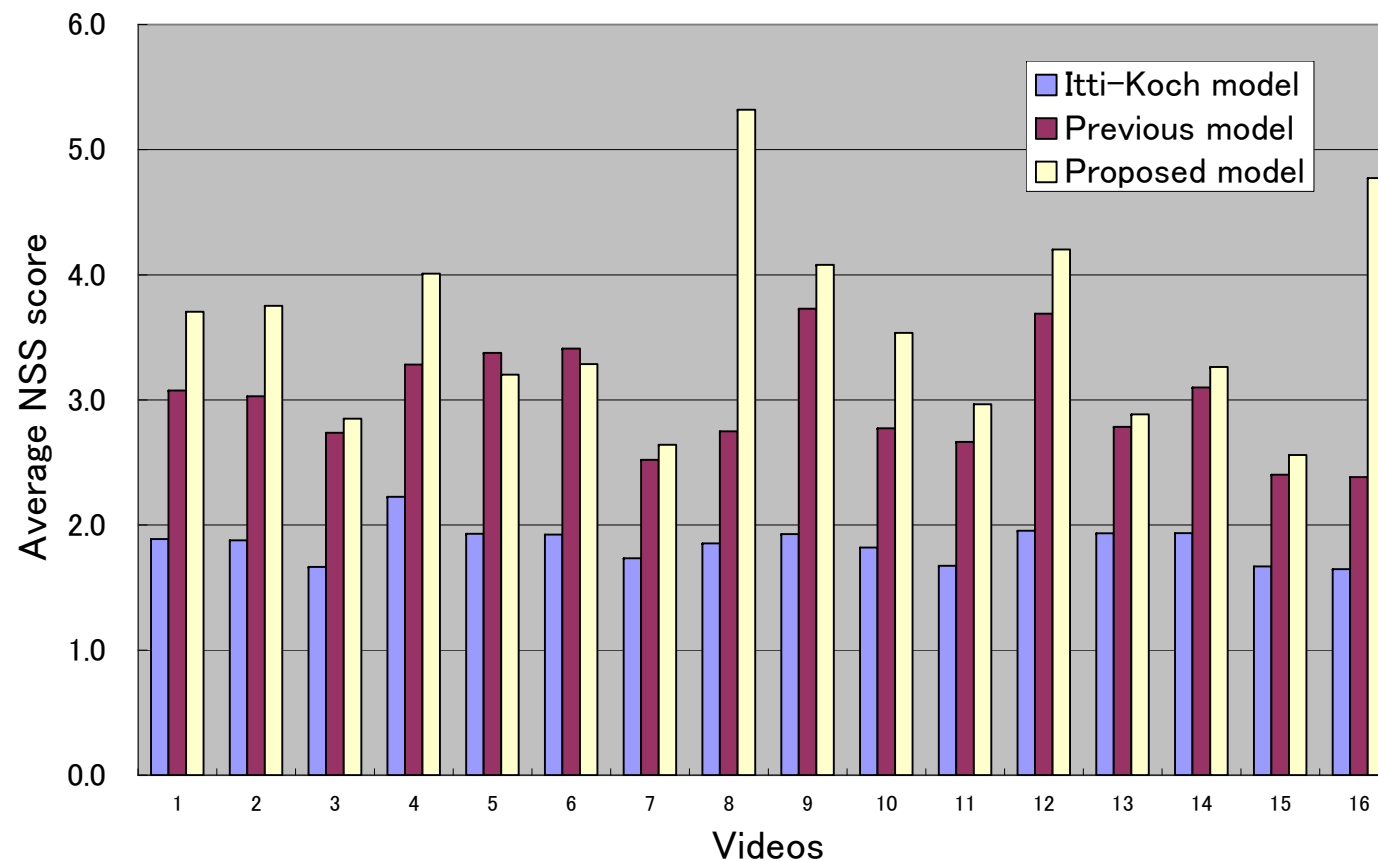
- 平均NSSの比較

- 提案法 with MRF >> Itti-Koch model (約2倍)
- 提案法 with MRF > 提案法 without MRF (約1.2倍)



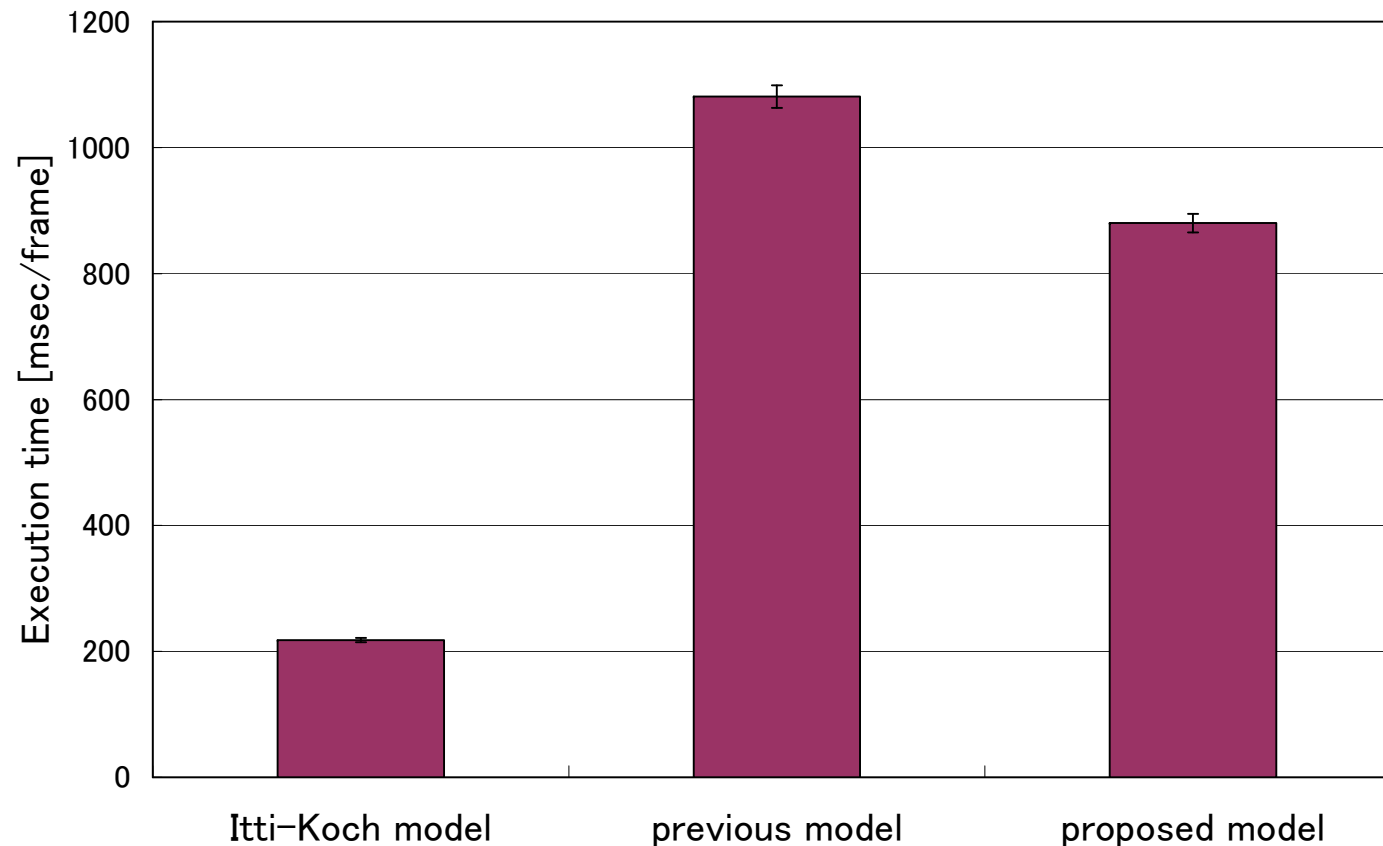
実験結果 (2/3)

- 各映像での平均NSSの比較
 - 多くのビデオで提案法の評価値が最も高い。



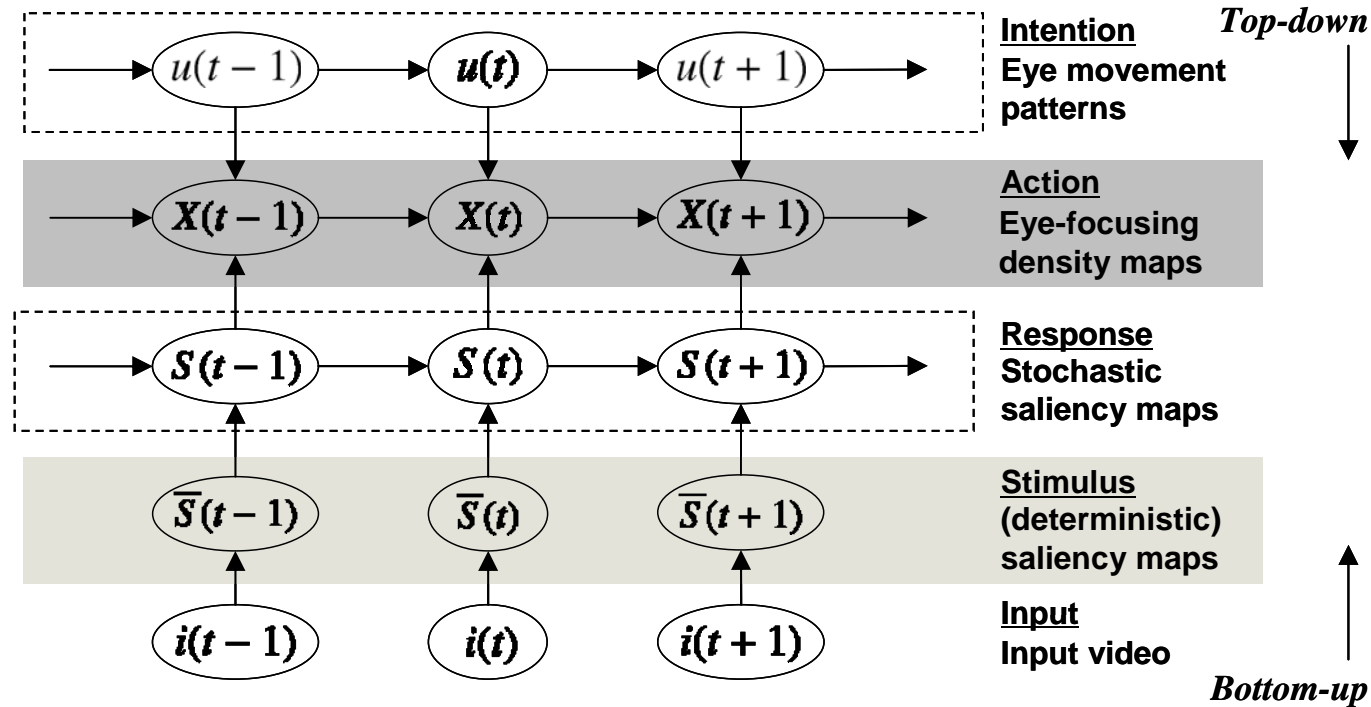
実験結果 (3/3)

- 平均計算時間の比較
 - 提案手法 with MRF < 提案手法 without MRF ... ? !



むすび

- 既提案の視覚的注意の確率モデルを拡張、映像顕著度の空間的な関係性を考慮
- 既提案技術に対しての優位性を確認。
 - 実際の人間の視線位置との一致性
 - 処理速度
- 今後の課題
 - トップダウン情報の拡張、ボトムアップ情報との関連性
 - GP-GPUなどを用いた高速実装



Thank you. Questions/Comments

E-mail: akisato <AT> eye brl ntt co jp