

映像閲覧行動の時空間ずれ構造モデルを用いた注視点予測 (author's version)

米谷 竜*

川嶋 宏彰†

松山 隆司‡§

概要

本資料は電子通信情報学会技術報告 映像閲覧行動の時空間ずれ構造モデルを用いた注視点予測 (信学技報, vol. 113, no. 196, IEICE-PRMU2013-41, pp. 57-62, 2013 年 9 月) の著者バージョンです。最終稿については <http://www.ieice.org/ken/paper/20130902MBfp/> をご覧ください。

人間が一般の映像を閲覧する状況では、映像中において人間の視覚的注意を引きつける顕著領域と、実際の注視点との間に時空間的なずれが生まれる。このずれは、顕著領域の変動 (顕著性変動) に対する反応遅れや動き予測に起因するものであり、変動の種類によってその程度が異なる。本研究では、顕著領域・注視点間のずれが持つ構造と顕著性変動の関係性に基づいて、映像から注視点を予測する手法を提案する。提案手法は、注視点に対して時空間ずれをともなって観測される顕著性変動パターンを、単純な変動パターンの組み合わせによって表現する。これにより、「注視点の近傍でどのような顕著性変動パターンがどのような時空間ずれとともに観測されやすいか」をデータドリブンに学習することによる注視点予測が実現できる。

1 はじめに

人間の視線運動は、その時々興味や集中といった心的状態のみならず、外界の動的な事象を反映して振る舞うものとして、視覚心理をはじめ様々な分野において研究されている [1]。我々は現在、人間がスポーツ映像やサーベイランス映像といった一般映像を閲覧する際の視線運動をとりあげ、コンテンツ中において視覚的注意が向けられうる箇所 (注意点) と実際に視線が向けられる箇所 (注視点) との間に存在する時空間的なずれについて研究している [2]。このようなずれの持つ構造をモデル化し、統計的に学習することにより、新たに得られた映像においてどこが注視されやすいかを予測することが可能であり、視線運動に基づく特定のタスクに対する熟練度の推定 [3] や ADHD といった障害の診断 [4]、視線運動を考慮したコンテンツ設計 [5] といった応

用につながることを期待できる。

視線運動と視覚的注意の時空間ずれ自体は古くより研究されており [6]、たとえば動く対象を目で追う際に反応遅延が発生すること [7] や、逆に対象の動き軌跡を予測して先に視線運動が開始されること [1] が知られている。これらは人工的な視覚刺激を用いた実験から得られた知見であるが、時空間ずれは一般映像を閲覧する際にも確認できる。図 1 は、映像においてバスが画面左から右に通過する際の注視点 (10 名, 白点) を示したものである。このとき、第 2, 3 フレームにおけるいくつかの注視点が、その前フレームにおいてバスの存在した箇所に依然として存在することが分かる。これは動く対象に対する視線運動の遅延によるものであり、実際の注視点と注意が向けられた顕著領域 (ここではバス) との間に存在する時空間ずれが現れたものとして捉えることができる。このように、注視点と注意点の間の時空間ずれは、注意対象である顕著領域が持つ変動 (顕著性変動) や視線運動の特性を反映するものとなる。

本研究では、上述の時空間ずれが持つ構造を解析

* yonetani@vision.kuee.kyoto-u.ac.jp

† kawashima@i.kyoto-u.ac.jp

‡ tm@i.kyoto-u.ac.jp

§ 京都大学 大学院情報学研究科 〒606-8501 京都市左京区吉田本町



図 1 映像閲覧における時空間ずれの例．白点は各フレームにおける 10 名の被験者の注視点を表す．

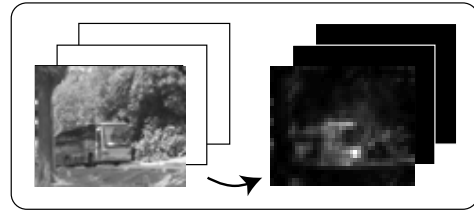
するためのモデル(時空間ずれ構造モデル)および、ずれ構造の統計的学習に基づいて注視点予測を行う手法を提案する．先行研究 [2] と同様、本研究では時空間ずれ構造の解析にあたって、顕著性マップ [8] などを用いて映像中における顕著領域を抽出し、視覚的注意が向けられうる候補領域として扱う．そのうえで本稿では、顕著性変動と時空間ずれの関係を表現・解析するための新たなモデルを導入する．提案モデルは、ある点を中心とした時空間パッチ内における顕著性変動パターンを、単純かつ局所化されたパターン (localized primitive patterns; LPP) を用いて記述する．LPP は与えられた点に対する顕著領域の相対位置と変動パターンの種類を同時に表現するものであり、注視点が与えられた場合、その近傍においてどのような変動パターンを持つ顕著領域がどの程度の時空間ずれを引き起こすかを解析できるようになる．また、大量の映像から LPP のコードブックを学習することで、ずれ構造の効率的な表現が可能になる．

提案手法の概要を図 2 に示す．本研究では、ある点の時空間近傍における顕著性変動パターンをベクトル表現しておく (図 2(A-1))．これにより、大量の変動パターンサンプルから構成されるデータ行列の因子分解に基づいて LPP のコードブックを学習可能になる (図 2(A-2))．

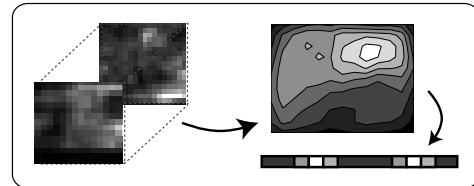
コードブックの学習により、時空間ずれを LPP の組み合わせにより表現できるようになる (図 2(B-1))．提案手法ではこの組み合わせの情報を特徴として用いることで、教師有り学習の枠組みに基づいて注視点予測を行う．具体的には、学習データとして与えられた映像と注視点系列に対して、注視点近

(A) Learning LPPs for gap structure models

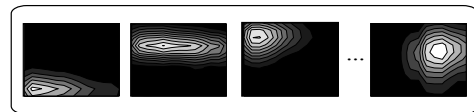
(A-0) Calculating saliency maps from videos



(A-1) Extracting saliency dynamics

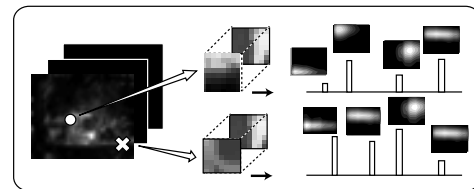


(A-2) Learning LPPs from data



(B) Gaze point prediction

(B-1) Feature extraction with LPPs



(B-2) Learning a discriminant function

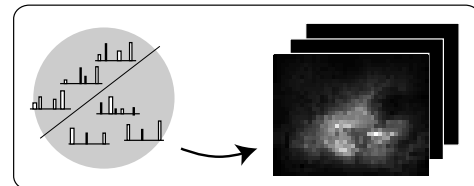


図 2 時空間ずれ構造の学習に基づく注視点予測

傍の LPP を正例、映像からランダム抽出された点の近傍における LPP を負例とし、識別関数を学習する (図 2(B-2))．これにより、映像のみが新たに与えられたとき、各ピクセルにおける注視のされやすさが計算でき、各フレームにおいてどこが注視されやすいかを分布として予測することが可能になる．

2 時空間ずれ構造モデル

人間がディスプレイを通してスポーツ映像やサーベイランス映像といった一般映像を閲覧する状況を考える．ディスプレイ下にはイトラッカーが設置され，視線運動がディスプレイ上の注視点系列として精度よく計測できることを前提とする．本研究や [2, 9] で取り上げられている注視点予測とは，与えられた映像中のどこが注視されやすいかを分布として推定する問題であり，手法の性能は，イトラッカーにより得られた注視点と予測結果との整合性を normalized scanpath saliency (NSS) [10] などで測ることにより評価できる*1．

我々の目的は，視覚的注意が向けられうる顕著領域と実際の注視点との間に存在する時空間ずれが持つ構造をモデル化し，その統計的学習に基づいて注視点予測を行うことである．時空間ずれ構造のモデル化にあたっては，以下の2点が問題となる．

- (A) 時空間ずれの構造をどのように表現するか
- (B) 顕著性変動や視線運動の種類が時空間ずれに及ぼす影響をどのように表現するか

先行研究 [2] は (A) を中心的に扱うものであり，ずれ構造を表現する関数を従来の知見に基づいてトップダウンに仮定することで，関数当てはめの枠組みでモデルの学習を行っている．これに対して本稿で提案するモデルは，注視点近傍における顕著性変動パターンから，顕著領域の典型的なずれを学習するボトムアップ型アプローチをとる（詳細は 2.1, 2.2 節）．[2] のアプローチは学習の簡便さおよび事前知識を導入することの容易さ，本稿のアプローチはモデルの表現能力においてそれぞれ有利である．さらに本研究では，前節に述べた LPP により顕著領域

*1 NSS は本来，ある画像特徴の持つ視覚的顕著性と注視点との相関性を測る指標として導入されたものであり [10]，注視点は人間の顕在的注意 (overt attention) として扱われている．また，現在では視覚的顕著性マップの良さを測る指標として多くの研究において採用されている．詳細は <https://sites.google.com/site/saliencyevaluation/evaluation-measures> などを参考にされたい．

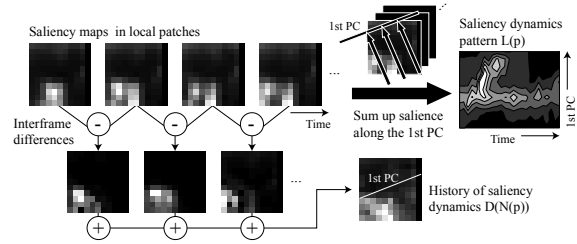


図3 顕著性変動パターンの2次元表現

のずれと変動パターンを同時に表現することにより，(B) のうち顕著性変動がずれに及ぼす影響を扱うことが可能である（視線運動の種類が及ぼす影響については，稿を改めて論じる）．

2.1 顕著性変動パターンの抽出・表現

映像中のある一点を $p = (x, y, t)$ と表現する． $(x, y) \in \mathbb{R}^2$ はフレーム中のある位置， $t \in \mathbb{N}$ はフレーム ID を表す．また， $p = (x, y, t)$ を中心とした時空間パッチを $\mathcal{N}(p) = \{(u, v, \tau) \mid \|x - u\| \leq \delta_x, \|y - v\| \leq \delta_y, \|t - \tau\| \leq \delta_t\}$ とおく ($\delta_x, \delta_y, \delta_t$ はそれぞれ近傍のサイズを示す)．このとき，点 p 近傍における顕著性変動は， $\mathcal{N}(p)$ 中の顕著性マップ系列により与えられることになる．もし p が注視点であるならば， $\mathcal{N}(p)$ 中の顕著性変動パターンは注視点近傍で顕著領域がどのような相対位置にあり，どのような変動をしているかを表現しているという点から，顕著性変動を考慮した時空間ずれを示すものとなる．

このようなずれの持つ構造を可視化するとともに後段の処理を簡易化するため，本研究では $\mathcal{N}(p)$ 中の変動パターンを時間，空間の軸からなる2次元ユークリッド空間に埋め込んで表現する（図3）．まず， $\mathcal{N}(p)$ 中の顕著性マップ系列のフレーム間差分の絶対値を計算し，それを時間方向に積分することで，顕著性変動の履歴 $D(\mathcal{N}(p))$ を得る．そして， $D(\mathcal{N}(p))$ を多数サンプルで近似し，その第一主成分を計算することで，変動パターンを最も良く表現する軸を得る．最後に，得られた軸と直交する方向に沿って， $\mathcal{N}(p)$ 中の各フレームにおける顕著度を積分することで，顕著性変動パターンの2次元表現 $L(p)$ を得る．

図4に得られた変動パターン $L(p)$ と対応する映像・視線データの一部を示す．ここでは，図中各ブロック2列目における注視点近傍の変動パターンを，同4列目にいくつかのスケール（時空間パッチの大きさ）で示した．同一点の近傍における顕著性変動であっても，スケールによって現れる変動パターンが異なる．また，複数の顕著領域が観測されることもある．

2.2 LPP コードブックの学習

本研究では，図4に示したような顕著性変動パターン $L(p)$ を，図5に示されるような単純かつ局所化されたパターン（LPP）に分解して表現する．さらに $L(p)$ を効率よく表現するにあたって，ここではLPPのコードブックを導入する．まず， $L(p)$ の1次元ベクトル表現を $l(p) \in \mathbb{R}_+^K$ とおく．そのうえで， I 個のLPPからなるコードブックを $\mathcal{M} = \{m_i \in \mathbb{R}_+^K \mid i = 1, \dots, I\}$ と表す． m_i はLPPを1次元ベクトル表現したものであり， $l(p)$ と同じ空間に定義されるものとする．コードブック \mathcal{M} を用いることで， $l(p)$ は m_i の活性度 a_i から構成されるベクトル $a(p) = (a_1, \dots, a_I)^T \in \mathbb{R}_+^I$ として表現できるようになる． p が注視点である場合， $a(p)$ は注視点近傍における時空間ずれ構造を表しており，学習データ中のある注視点 p に対して活性度 a_i が高い場合，注視点の周りに m_i により表される時空間ずれ・変動パターンを持つ顕著領域が存在することを示している．

コードブック \mathcal{M} の学習において，本研究では非負値行列行列因子分解（non-negative matrix factorization; NMF）[11]を利用する．NMFは顔解析[11]のみならず，文章データ・音の解析においてもしばしば利用されており[12]，与えられた非負行列をスパース構造化された非負因子とその活性化係数を表現する非負因子に分解する手法である．いま， N 個のサンプル $\mathcal{L} = (l(p_1), \dots, l(p_N)) \in \mathbb{R}_+^{K \times N}$ を考える．これに対して，NMFにより以下のような2個の非負因子が得られる．

$$\mathcal{L} = MA \quad (1)$$

このとき， $M \in \mathbb{R}_+^{K \times I}$ はLPPの系列 $M =$

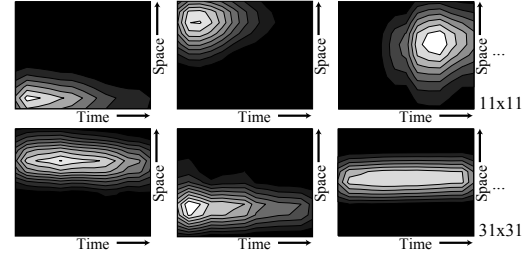


図5 コードブックとして得られたLPPの一部

(m_1, \dots, m_I) （すなわちコードブック \mathcal{M} ）に対応し， $A \in \mathbb{R}_+^{I \times N}$ は活性度ベクトルの系列 $A = (a(p_1), \dots, a(p_N))$ となる．

図5に得られたコードブックの一部を示す． M および A の推定には multiplicative update rules [13] を利用した．

3 時空間ずれ構造モデルを用いた注視点予測

これまでの手続きにより，ある点 p の近傍における時空間ずれ構造がLPPの活性度 $a(p)$ により表現できる．これを用いて，本研究では教師有り学習の枠組みに基づく注視点予測を行う．具体的には，注視点近傍におけるLPPの活性度を正例，映像からランダムに抽出した点近傍での活性度を負例とし，注視点近傍でどのようなLPPが強く現れやすいかを識別関数として学習する．これにより，新たに得られた映像中のどこが注視されやすいかという情報が，分布として得られるようになる．

図4に示した通り，注視点近傍の顕著領域数やその変動パターンは，近傍のスケールによって異なる．これらは全て視覚的注意を引きつける可能性がある一方で，実際にどの領域に注意が向けられるかは未知である．そこで本研究では， H スケールからなる複数の近傍 $\mathcal{N}_h(p)$ ($h = 1, \dots, H$) におけるずれ構造を同時に学習するアプローチをとる．具体的にはまず， $\mathcal{N}_h(p)$ における顕著性変動パターン $L(p)$ をそれぞれ同一のパッチサイズに縮小したのち，スケールごとに異なるLPPのコードブックを学習する．そして，各スケールについて活性度ベクトル $a_h(p) \in \mathbb{R}_+^{I_h}$ を計算する (I_h はスケール h における

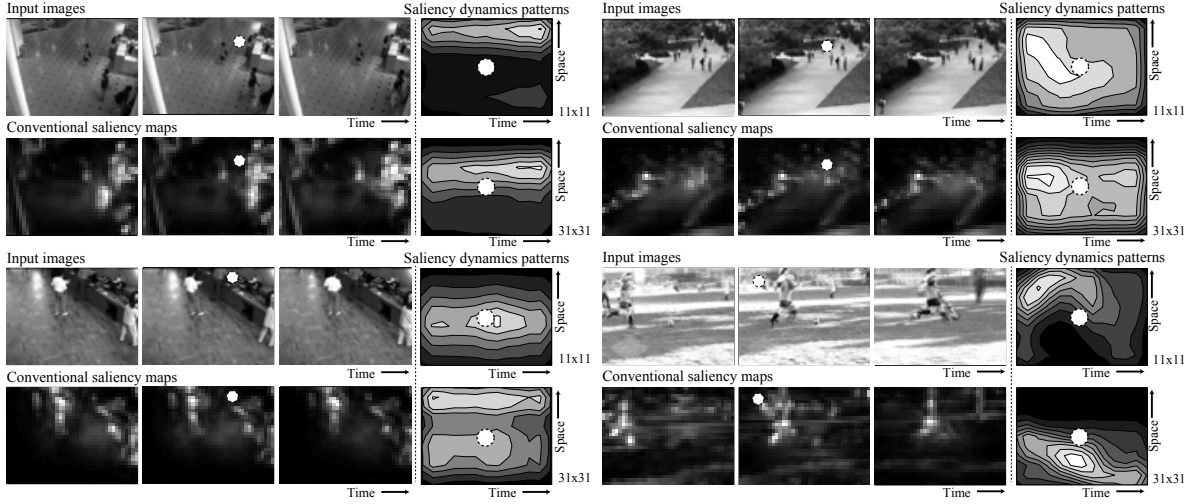


図4 顕著性変動パターン例. 各ブロックについて2列目の白点は1名の被験者の注視点を示す. 4列目は注視点(中心の白点)の近傍における顕著性変動パターンを示す. 上段では 11×11 ピクセル $\times 0.4$ 秒, 下段: 31×31 ピクセル $\times 0.4$ 秒の時空間パッチを用いた.

コードブックサイズ). 最後に, これらの活性度ベクトルを結合し, $\mathbf{a}(p) = (\mathbf{a}_1(p)^T, \dots, \mathbf{a}_H(p)^T)^T \in \mathbb{R}_+^{I'}$ を得る ($I' = \sum_h I_h$).

点 p における注視のされやすさは以下のように評価される.

$$F(p) = \beta^T \bar{\mathbf{a}}(p) \quad (2)$$

ただし, $\bar{\mathbf{a}}(p)$ は学習データにおいて $\mathbf{a}(p)$ を平均0, 標準偏差1に正規化した特徴ベクトルである. また, $\beta \in \mathbb{R}^{I'}$ はLPPの寄与率を表すパラメタである. 本研究では, 正例の点 p (注視点) について $g(p) = 1$, 負例の点 p (ランダム抽出された点) について $g(p) = -1$ となるような識別関数 $g(p) = \text{sgn}(F(p) + \beta_0)$ として β を学習する (ただし $\beta_0 \in \mathbb{R}$ はバイアス項). これにより, 新たに得られた映像の各ピクセルにおける注視のされやすさ $F(p)$ が計算でき, 各フレームにおいてどこが注視されやすいかが分布として得られる.

4 評価実験

提案手法の有効性を検証するため, 本稿では AS-CMN database [14] (24種類の映像, 10人の被験者) を用いた評価実験を行った.

4.1 評価方法

映像に対する汎化性を検証するため, 本研究では C 個の映像のうち $C-1$ 個を用いてモデルの学習を行い, 残り1個の映像を用いて学習されたモデルを評価する形で leave-one-out 交差検証を行った. モデルのパラメタは学習データにおいて受信者操作特性曲線の曲線下面積 (area under the curve; AUC) が最も高くなるものとして推定した. そのうえで, 学習結果に基づいて得られた注視点予測の分布を NSS によって評価した. モデルの最終的な評価は, 各映像をテストデータとした際の NSS を全映像について平均することで得た.

4.2 モデルの実装・学習

入力として与える顕著性マップの計算においては, Ittiらのモデル [8] を利用した. そのうえで, 提案手法では以下のパラメタを設定・学習する必要がある.

- 各スケールにおける時空間近傍 $\mathcal{N}_1(p) \dots \mathcal{N}_H(p)$ およびスケール数 H .
- LPPのコードブック $\mathcal{M}_1, \dots, \mathcal{M}_H$ およびそのサイズ I_1, \dots, I_H ,
- 識別関数のパラメタ β

$\mathcal{M}_1, \dots, \mathcal{M}_H$ については, I_1, \dots, I_H それぞれについてあらかじめとりうる値を設定したのち, 正例・負例をあわせた全サンプルを用いてそれぞれの場合を学習した. そして, 学習された β を用いて学習データを評価することで得られる AUC に基づいて最適なものを選択した. また, β の学習にあたってはフィッシャーの線形識別を利用した. H は経験的に $H = 2$ と設定するとともに, \mathcal{N}_1 および \mathcal{N}_2 における空間方向のサイズ (δ_x, δ_y) は $(5, 5)$ ピクセル (11×11 ピクセルのパッチ) および $(15, 15)$ ピクセル (31×31 ピクセルのパッチ) に設定した (フレームサイズは 80×60 ピクセル). また, 時間方向のサイズ δ_t は $\mathcal{N}_1, \mathcal{N}_2$ とともに 0.4 秒に設定した.

4.3 比較手法

本実験では (1) 時空間ずれ構造を利用することの効果, (2) LPP を用いたモデル化の効果の検証を目的として, 2 種類の比較手法 SM, GSM1 を導入した. さらにベースライン (手法 BS) として, [8] の顕著性マップを利用した. 手法 SM は (1) を検証するものであり, 文献 [14] にて提案されている手法を修正して用いた. 文献 [14] では与えられたデータに対して NSS が最も高くなるよう顕著性マップを平滑化するが, その際の平滑化パラメタは学習によって得られるものではない. そこで手法 SM では, 提案手法と同様, 学習データにおける AUC が最も高くなるように平滑化パラメタを定めた. そのうえで, 平滑化された顕著性マップの顕著度を特徴として, 提案手法と同様に識別関数を学習することによって, 最終的な出力を得た. 一方, 手法 GSM1 は (2) に関するものであり, [2] のモデル化法を踏襲し, \mathcal{N}_h 中において [8] を用いて得られる顕著度が最も高くなる点を注意点として検出するとともに, その点を持つ時空間ずれを 2 次関数を用いてモデル化した. また, 注視点予測にあたっては, 検出された注意点における顕著度を時空間ずれの構造と同時に学習した (詳細は文献 [2] を参照されたい).

4.4 結果と考察

表 1 に手法ごとの NSS スコアを示す. 手法 BS と比較して手法 SM のスコアが高くなっていることから, 注視点と注意点の間に時空間ずれが存在

表 1 手法ごとの NSS スコア. 値が高いほどモデルが注視点を精度良く予測していることを示す.

BS [8]	SM	GSM1 [2]	GSM2 (proposed method)
0.63	0.81	0.94	0.97

することが分かる. さらに, 手法 SM および手法 GSM1 と比較して提案手法 GSM2 のスコアが高くなっていることから, 上に述べた時空間ずれには特定の構造が存在し, それを学習することで注視点予測の性能を向上が可能であること, また LPP を用いた構造のモデル化が有効であることが示された.

図 6 に手法ごとの出力例を示す. 1, 2, 3 行目では, 被験者の多くは画面外に走り去る車や人間を追跡しており (白点線で囲まれた点群), その際時空間ずれが発生している. このような場合, 手法 BS, SM と比較して, 時空間ずれ構造を用いる GSM1, GSM2 では注視がされやすいと予測される箇所が注視点近傍を捉えていることが確認できる. 一方, 4 行目のように視線運動の遅延があまりに大きい場合 (黒点線で囲まれた点群), 提案手法を用いても正しく注視点予測ができていない. これについては, 本実験で経験的に決定している時空間近傍のサイズを調整することで改善される可能性がある. さらに 5 行目のように, 被験者が対象を凝視 (fixation) しており, もともとの入力 (BS) が注視点をよく捉えている場合, 提案手法を用いることで逆に誤った箇所が注視されやすいと予測され, スコアが低下するという結果も確認された. これに関しては文献 [2] においても議論されている通り, 視線運動の種類に応じて時空間ずれの程度が異なることを考慮する必要があると考えられる. 今後は, 顕著性変動パターンから視線運動の種類を推定し, 注視点推定に用いる枠組みを検討する.

最後に, 注視点予測において正例と負例を分離する時空間ずれを図 7 に示す. 学習されたパラメタ $\beta = (\beta_1, \dots, \beta_I)$ の各要素 β_i は, 対応する LPP m_i の寄与率と考えることができるため, ここでは全 m_i について β_i による重み付け和を計算するこ

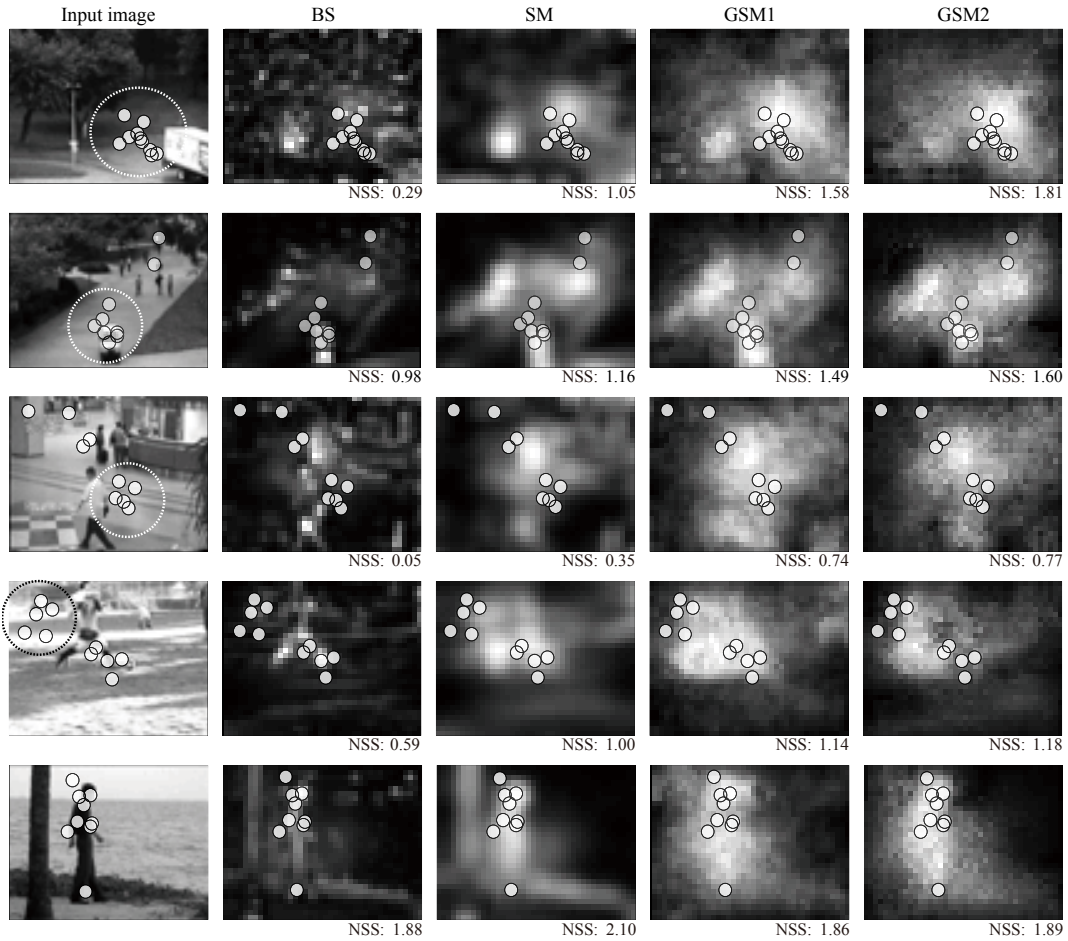


図6 手法ごとの出力例．白点は10名の被験者の注視点を示す．

とで、予測に寄与するLPPを可視化した． 11×11 ピクセルのパッチにおける時空間ずれにおいては、学習データに依らず視覚的注意の向きうる顕著領域に対して注視が時間的に遅延する結果が得られている．その一方で、 31×31 ピクセルのパッチでは、詳細は学習データによって異なるものの、注視が顕著領域に対して時間的に先行する点が全体に共通する特性として挙げられる．さらに、 11×11 ピクセルのパッチと異なり、注視点と顕著領域の間には空間的なずれも存在する．すなわち実験において被験者は、ある程度顕著領域の動きを予測して注視を移動させることはできるものの、遅延なく領域を追跡することは難しく、時空間的なずれをとまって注視を行っていたことが示唆される．また、これらの時空間ずれは全てが必ずしもGSM1のモデル(2次

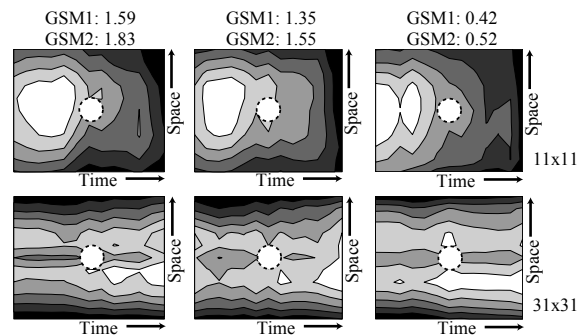


図7 注視点予測に寄与するLPPの可視化結果および手法GSM1, GSM2のNSSスコア

関数)によって精度良く表現できるものではない点から、LPPを用いたボトムアップなモデル化アプローチの有効性が確認できる．

5 むすび

映像閲覧行動における注視点と注意点の時空間ずれが持つ構造を統計的に学習することにより、新たに得られた映像に対して注視点予測を行う手法を提案した。提案手法では、注視点の近傍において視覚的注意をひきつける顕著領域の変動パターンを、単純な変動パターンの組み合わせによって表現する。これにより、「注視点の近傍でどのような顕著性変動パターンがどのような時空間ずれとともに観測されやすいか」をデータドリブンに学習し、注視点予測に用いることが可能になった。

謝辞 本研究の一部は、科学研究費補助金 特別研究員奨励費 24・5573 の補助を受けて行った。

参考文献

- [1] E. Kowler, “Eye movements: the past 25 years,” *Vision Research*, vol.51, no.13, pp.1457–1483, 2011.
- [2] R. Yonetani, H. Kawashima, and T. Matsuyama, “Learning Spatiotemporal Gaps between Where We Look and What We Focus on,” *IPSJ Trans. on CVA*, vol.5, pp.1–5, 2013.
- [3] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen, “Gaze-based interaction for semi-automatic photo cropping,” *CHI*, pp.771–780, 2006.
- [4] P.H. Tseng, I.G. Cameron, G. Pari, J.N. Reynolds, D.P. Munoz, and L. Itti, “High-throughput classification of clinical populations from natural viewing eye movements,” *J. Neurol.*, vol.260, no.1, pp.275–284, 2013.
- [5] J. Simonin, S. Kieffer, and N. Carbonell, “Effects of display layout on gaze activity during visual search,” *INTERACT*, vol.3585, pp.1054–1057, 2005.
- [6] J. Hoffman, “Visual Attention and Eye Movements,” *Attention*, ed. by H. Pashler, vol.31, chapter 3, pp.119–153, Psychology Press, 1998.
- [7] C. Rashbass, “The relationship between saccadic and smooth tracking eye movements,” *The Journal of Physiology*, vol.159, pp.326–338, 1961.
- [8] L. Itti, C. Koch, and E. Niebur, “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis,” *TPAMI*, vol.20, no.11, pp.1254–1259, 1998.
- [9] T. Judd, K. Ehinger, F. Durand, and A. Torralba, “Learning to Predict Where Humans Look,” *ICCV*, pp.2106–2113, 2009.
- [10] D. Parkhurst, K. Law, and E. Niebur, “Modeling the role of salience in the allocation of overt visual attention,” *Vision Research*, vol.42, no.1, pp.107–123, 2002.
- [11] D. Lee and H. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol.401, no.6755, pp.788–791, 1999.
- [12] 澤田 宏, “非負値行列因子分解 nmf の基礎とデータ/信号解析への応用,” *電子情報通信学会誌*, vol.95, no.9, pp.829–833, 2012.
- [13] D. Lee and H. Seung, “Algorithms for non-negative matrix factorization,” *NIPS*, pp.556–562, 2001.
- [14] N. Riche, M. Mancas, and D. Culibrk, “Dynamic Saliency Models and Human Attention: A Comparative Study on Videos,” *ACCV*, pp.586–598, 2012.