

疎表現理論に基づく知能的視覚情報処理システムに関する研究 — 動画像疎表現の統計的モデリングとその応用 —

齊藤 隆弘*

Study on Intelligent Visual Information Processing Systems Based on Sparse-Coding Theory: Statistical Modeling of Moving-Image Sparse Coding and Its Applications

Takahiro SAITO*

1. はじめに

標記の研究を、工学研究所 研究プロジェクト A として、計 6 名の研究組織にて、2016 年 4 月より 2019 年 3 月までの 3 年間に亘って実施した。その研究内容は多岐に亘るが、本報告では、その内で最も基礎的且つ基盤的な研究として位置づけた“動画像疎表現の統計的モデリングとその応用”に焦点を絞り、得られた主要な研究成果を、研究の背景と動機とともに要約して示す。詳細については、研究代表者らが公表した研究論文等を参考されたい。

2. 研究の背景と動機

近年、f-MRI 脳機能イメージングや経頭蓋磁気刺激法等の非侵襲的に脳活動を観測する技術の進歩により、脳神経科学が急速に進展し、とくに視覚系の神経回路網の構成とその動作メカニズムが明らかにされつつある。20 世紀の脳神経科学では、フィルタリング理論や標準正規化理論をその理論的基盤とした D. Marr の視覚計算モデルが最も有力な作業仮説とされていたが、1990 年代半ばに B. A. Olshausen と D. J. Field は、『感覚信号の脳内表現は疎表現である』とする作業仮説を提唱した^[1]。現在では、この疎表現仮説が、最も有力な作業仮説と評価されている。

疎表現とは、『信号がごく少数の記述子の線形結合で表現される』ことを意味する。また、脳神経科学における疎表現仮説とは、『構造的な特徴をもち、脳にとって有意な信号には必ず疎な表現が存在し、一方出鱈目で脳にとって無意味な信号には疎な表現は存在しない』とする作業仮説である。

さらに、その後、疎表現の概念を導入することで、多層神経回路網を局所最適解に陥らせることなく効率的に学習させることが可能な“深層学習法”が開発され、これを用いた教師学習によって“静止画像から物体を認識する精度”が格段に向上することが明らかされた。しかし、動画像シーケンスを対象とした深層学習は、現時点では、発展途上にある。

一方、研究代表者らは、人の視覚を超えた人工視覚の実現を目指

し、2007 年頃から疎表現理論に基づく各種の画像復元技術を構成するとともに、これにより画像復元性能が飛躍的に向上することを明らかにしてきた。本研究プロジェクトでは、研究代表者らのこれまでの研究成果をさらに発展させる基礎的且つ基盤的な研究課題として“動画像疎表現の統計的モデリング”を取り上げ、その統計的手法を理論的且つ構造的に整備し、さらに動画像復元の高度化に応用展開するための基本的枠組みを構築した。

3. 3 次元 DFT 領域における動画像疎表現の統計的モデリング

研究代表者らは、2016 年に準備的な研究として、動画像復元に適した 3 次元平均値分離型短時間 DFT を構成し、これを動画像雑音除去へと適用し、世界最高水準の動画像雑音除去手法と対等以上の優れた雑音除去性能を示すことを実験的に明らかにした^[2]。この研究により 3 次元 DFT が動画像復元の強力なツールであることが実証されたことを受け、3 次元 DFT 係数の統計的モデリングに研究の焦点を絞り、複素 DFT 係数の 2 次元確率分布を球対称分布として統計的にモデル化する手法の開発に着手した。

具体的には、2 成分混合型球対称ガウス (2-C S²GM: 2-Component Spherically-Symmetric Gaussian Mixture) 分布モデルを採用した。2-C S²GM モデルは、汎用的な混合型確率分布モデルの一つであるが、その中に原信号の“疎性の尺度”と“観測劣化モデル”を陽な形式で含ませることができる。2 次元 2-C S²GM モデルでは、疎性を有する 2 次元原信号 $\bar{\mathbf{x}}$ の事前確率分布 $p_x(\bar{\mathbf{x}})$ を、有意性の低い信号に対応した比較的小さな分散 $\sigma_{s,[0]}^2$ の 2 次元球対称ガウス分布 $g_2(\bar{\mathbf{x}}; \bar{\mathbf{0}}, \sigma_{s,[0]}^2)$ と、有意性の高い信号に対応した比較的大きな分散 $\sigma_{s,[1]}^2$ の 2 次元球対称ガウス分布 $g_2(\bar{\mathbf{x}}; \bar{\mathbf{0}}, \sigma_{s,[1]}^2)$ を $P_0 : P_1$ の混合比で確率的に混合した確率密度関数として、次式にて定義する。

$$p_x(\bar{\mathbf{x}}) := P_0 \times g_2(\bar{\mathbf{x}}; \bar{\mathbf{0}}, \sigma_{s,[0]}^2) + P_1 \times g_2(\bar{\mathbf{x}}; \bar{\mathbf{0}}, \sigma_{s,[1]}^2); \\ 0 < P_0 < 1, 0 < P_1 < 1, P_0 + P_1 = 1, \sigma_{s,[0]}^2 < \sigma_{s,[1]}^2 \quad (1)$$

上式では、事前確率分布が単一の球対称ガウス分布に縮退する“特異な場合”を排除するように、モデルパラメータ $P_1, \sigma_{s,[0]}, \sigma_{s,[1]}$ の許容範囲を設定している。

原信号 $\bar{\mathbf{x}}$ に雑音 $\bar{\mathbf{w}}$ が加わった劣化信号 $\bar{\mathbf{y}} = \bar{\mathbf{x}} + \bar{\mathbf{w}}$ が観測される条

*研究代表者 教授 電気電子情報工学科

Research Representative, Professor, Dept. of Electric, Electronic and Information Engineering

件付確率 $p_w(\bar{\mathbf{y}}|\bar{\mathbf{x}})$ が、平均 $\bar{\mathbf{x}}$ 、分散 $\sigma_w^2 (>0)$ の 2 次元球対称ガウス分布 $g_2(\bar{\mathbf{y}}|\bar{\mathbf{x}}, \sigma_w^2)$ に従うものと仮定する。このとき、劣化信号 $\bar{\mathbf{y}}$ が観測される確率 $p_y(\bar{\mathbf{y}})$ は、次式で与えられる。

$$p_y(\bar{\mathbf{y}}) = P_0 \times g_2(\bar{\mathbf{y}}; \bar{\mathbf{0}}, \sigma_0^2) + P_1 \times g_2(\bar{\mathbf{y}}; \bar{\mathbf{0}}, \sigma_1^2); \quad (2)$$

$$\sigma_0^2 := \sigma_w^2 + \sigma_{s,[0]}^2 > 0, \quad \sigma_1^2 := \sigma_w^2 + \sigma_{s,[1]}^2 > \sigma_0^2$$

この確率分布を簡潔に $\mathcal{GM}_2(\bar{\mathbf{0}}; \sigma_0^2, \sigma_1^2; P_0, P_1)$ と表記する。この 2-C S²GM モデルは、三つの独立なモデルパラメータ $P_1, \sigma_0^2, \sigma_1^2$ をもつ。

モデル化対象の各動画シーケンスの複素数の 3 次元交流 DFT 係数のサンプル系列から求めた“その大きさの 1 次, 2 次, 4 次モーメントのサンプル推定値”を、各々 $a_R, \sigma_R^2, \mu_{R,4}$ とする。このとき、同一の統計量を与える 2 次元 2-C S²GM 分布モデルの三つのモデルパラメータ $P_1, \sigma_0^2, \sigma_1^2$ を求めることを通し、複素数の 3 次元交流 DFT 係数の統計的モデリングが行われる。2 次元 2-C S²GM 分布モデルのモデルパラメータの統計的推定法の詳細については、研究代表者らの文献[3]を参照されたい。

統計的モデリングによって一意に求められた二次元 2-C S²GM 分布モデル $\mathcal{GM}_2(\bar{\mathbf{0}}; \sigma_0^2, \sigma_1^2; P_0, P_1)$ の確率密度関数の“形状”を評価するため、二次元信号 $\bar{\mathbf{y}}$ の大きさの 2 次モーメントが次元数 2 と一致するように 2 次元 2-C S²GM 分布モデルを規格化する。この規格化 2 次元 2-C S²GM 分布モデルは、次式にて与えられる。

$$\mathcal{GM}_2(\bar{\mathbf{0}}; S, T; P_0, P_1); \quad S = \frac{\sigma_0^2}{P_0 \sigma_0^2 + P_1 \sigma_1^2}, \quad T = \frac{\sigma_1^2}{P_0 \sigma_0^2 + P_1 \sigma_1^2} \quad (3)$$

$$P_0 + P_1 = 1 \quad \& \quad P_0 + P_1 = 1, \quad 0 < S < 1, \quad T > 1, \quad 0 < P_0 < 1, \quad 0 < P_1 < 1$$

なお、独立なモデルパラメータは、 P_0, S の二つのみである。

図 1 には、種々の周波数インデックス (l, m, n) について、“各動画シーケンスの複素数の 3 次元交流 DFT 係数に当て嵌められた規格化 2 次元 2-C S²GM 分布”の二つの独立なモデルパラメータ P_0, S を 2 次元座標上にプロットした。これにより、複素数の 3 次元交流 DFT 係数の疎性を視覚的に把握できる^[4]。図 1 は、3 次元 DFT の解析サブブロックのサイズを $8 \times 8 \times 8$ とし、三種類の ITE 標準モノクローム動画シーケンスを対象とした結果である。図 1 に示したように、各々の周波数インデックス (l, m, n) の 3 次元交流 DFT 係数に対して求められたモデルパラメータは、その多くが $(P_0, S) = (1, 0)$ 付近に偏在しており、このことは非常に疎性が強く、且つ非常に裾野の長い確率分布としてモデル化されていることを意味している。

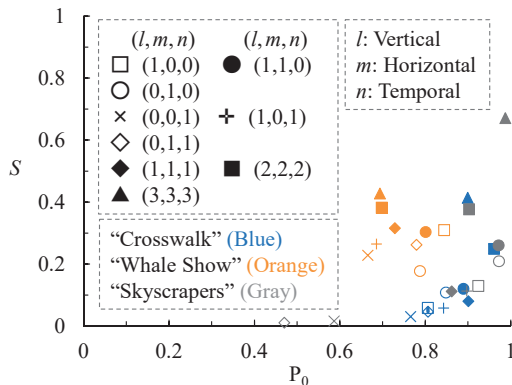


図 1 周波数インデックス (l, m, n) の 3 次元交流 DFT 係数に当て嵌められた規格化 2 次元 2-C S²GM 分布のモデルパラメータ P_0, S 。

4. ベイズ LS 推定関数の構成と動画復元への応用

雑音 $\bar{\mathbf{w}}$ の分散 σ_w^2 が既知であると仮定し、2 次元 2-C S²GM モデルの下で観測劣化信号 $\bar{\mathbf{y}}$ から原信号 $\bar{\mathbf{x}}$ を復元する問題をベイズ統計学の枠組みで統計的最適化問題として定式化し、復元誤差の二乗平均を最小化するという意味での最適性を有したベイズ LS (Least Squares) 推定を解析的に閉じた関数として導出する^[5]。とくに、3 次元 DFT を用いた動画復元への応用では、式(2)の二次元 2-C S²GM 分布モデルでモデルパラメータ $\sigma_{s,[0]}^2$ を強制的に 0 に固定した“簡易化モデル”に対して構成した次式のベイズ LS 推定関数が

$$\hat{\bar{\mathbf{x}}}(\bar{\mathbf{y}}) = \int_{\mathbb{R}^2} \bar{\mathbf{x}} p_{\bar{\mathbf{x}}|\bar{\mathbf{y}}}(\bar{\mathbf{x}}|\bar{\mathbf{y}}) d\bar{\mathbf{x}} = c_1 \cdot \frac{\bar{\mathbf{y}}}{1 + c_2 \exp(-c_3 \|\bar{\mathbf{y}}\|^2)}, \quad (4)$$

$$c_1 = \frac{\sigma_{s,[1]}^2}{\sigma_w^2 + \sigma_{s,[1]}^2}, \quad c_2 = \frac{P_0}{P_1} \cdot \frac{\sigma_w^2 + \sigma_{s,[1]}^2}{\sigma_w^2}, \quad c_3 = \frac{1}{2} \cdot \frac{\sigma_{s,[1]}^2}{\sigma_w^2 (\sigma_w^2 + \sigma_{s,[1]}^2)}$$

望ましい結果をもたらす。このベイズ LS 推定関数の出力ベクトル $\hat{\bar{\mathbf{x}}}(\bar{\mathbf{y}})$ は、入力 $\bar{\mathbf{y}}$ と同一方向である。よって、式(4)は、入力 $\bar{\mathbf{y}}$ の大きさのみに作用する位相保存型 Shrinkage の意義を有する。なお、観測劣化動画に混入している雑音 $\bar{\mathbf{w}}$ の分散 σ_w^2 は、高周波数の 3 次元交流 DFT 係数から高精度に推定できる。

詳細は省略するが、雑音に汚された観測劣化動画の 3 次元平均値分離型短時間 DFT において、各々の周波数インデックスの DFT 係数への二次元 2-C S²GM 分布の簡易化モデルの当て嵌めを通して式(4)のベイズ LS 推定関数を構成し、これを零出力 Shrinkage や線形 Shrinkage と合わせて階層化し、DFT 係数に位相保存型 Shrinkage として適応的に適用することで、雑音除去性能が適応的且つ統計的に最適化される^[5]。また、全ての処理パラメータは、処理対象の観測劣化動画のみから自動的に設定される。こうして構成された 3 次元平均値分離型短時間 DFT 動画雑音除去法は、2018 年時点で世界最高の雑音除去性能を示した^[5]。

5. むすび

本研究で開発した“動画疎表現の統計的モデリング技法”は、動画復元の高度化に欠かせない基盤的な技術であり、今後、各種の動画復元課題へと応用されることを期待したい。

6. 参考文献

- [1] B. A. Olshausen and D. J. Field, Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?, Vision Res., 37 (23), 3311-3325 (1997).
- [2] 小松隆, 張鍵, 齊藤隆弘, 平均値分離型三次元 ST-DFT を用いた動画復元法, 電子情報通信学会論文誌, J100-D (9), 842-845 (2017).
- [3] 齊藤隆弘, 小松隆, ランダム信号の DFT 係数のための多次元二成分混合型球対称ガウス分布モデルとそのパラメータ推定, 電子情報通信学会論文誌, vol.J101-D (9), 印刷中 (2018).
- [4] 齊藤隆弘, 小松隆, [招待講演] 動画の 3-D DFT 係数の統計的モデリングとその動画復元への応用, 電子情報通信学会・画像工学研究会, IE2018-6 (岐阜, 2018.5).
- [5] T. Komatsu and T. Saito, Statistical modeling for 3-D DFT Coefficients of Moving-Image Sequences and Its Application to Denoising, Proc. IEEE ICIP 2018, in print (Athens, 2018.10).