

## 三次元平均値分離型短時間 DFT による動画像の復元性能 Performance of moving image restoration using 3D mean separation type ST-DFT

小松 隆<sup>†</sup>      齊藤 隆弘<sup>†</sup>  
Takashi Komatsu   Takahiro Saito

### 1. はじめに

動画像復元において、疎表現効率の観点から三次元 DFT は三次元 DCT よりも優位であることを示した[1]. さらに、動画像復元への応用と復元性能の改善のため、DFT 係数を混合型球対称ガウス分布により統計的モデリングする手法提案した[2].

本稿では、実際の劣化動画像に対して三次元平均値分離型短時間 DFT 復元法を適用する具体的な処理手順について述べると共に、シミュレーションにより、既存の動画像復元手法 BM4D との性能比較を行った。

### 2. Short Time DFT

画像からブロックを切り出し二次元 DFT すると、ブロック境界における不連続の影響が多数の DFT 係数に及ぶ。短時間 DFT(Short Time DFT, 以下では ST-DFT) は、このブロック境界における不連続の影響を軽減するため、切り出されたブロック内の信号に対してブロック境界でほぼ 0 となる重み窓を掛けてから DFT する方式である。ST-DFT は 1980 年に J.S.Lim により提案され、静止画像の復元に応用された[3]. 静止画像に対しては二次元 ST-DFT を用いても二次元 DCT との復元性能に優位差はほとんど見られない。

動画像に対しては、動画像シーケンスから三次元ブロックを切り出し、三次元 ST-DFT を適用する。画面全体が平行移動している場合、切り出されたブロック内の  $n$  フレームと  $n+1$  フレームのデータの二次周波数の大きさは概ね等しく、位相のみが異なる。このため三次元 ST-DFT は、動画像をより少ない変換係数で表現(疎表現)できる。一方、動画像に対する三次元 DCT は、三次元ブロックを週対称に折り返す操作の影響で、フレーム間の動きが複数の周波数成分に影響を与える。動画像を表現するために三次元 DCT は三次元 ST-DFT よりも多くの変換係数を必要とする。よって、動画像復元には三次元 DCT よりも三次元 ST-DFT を用いたほうが疎表現の観点から明らかに有効な変換であるといえる。

### 3. 平均値分離 ST-DFT 動画像復元法の手順

図 1 に平均値分離 ST-DFT 動画像復元法の処理手順を示す。

①で雑音付加動画像に対して別途付加雑音の標準偏差  $\sigma_w$  の推定を行なう。具体的には、雑音付加動画像から  $16 \times 16 \times 16$  の三次元ブロックを冗長に切り出し、切り出された各々のブロックに対して 3 次元 DFT を適用する。得られた DFT 係数のうち時空間周波数の高い複素変換係数(実際にはインデックスが[8,8,7]の変換係数)の実部と虚部の値の絶対値を、大きさの順に並べた時のメディアン値を 0.6745 で割ることで  $\sigma_w$  の値をロバスト推定している。

②では雑音付加動画像から  $16 \times 16 \times 16$  の三次元ブロックを冗長に切り出す。実際には、水平・垂直・フレームの各

方向とも 2 画素飛ばしで三次元ブロックの切り出しを行っている。

③では切り出された三次元ブロックごとにブロック内の画素値の平均を求め、ブロック内の各画素値から引く操作を行う。

④では平均値が分離された三次元ブロックに対して、三角重み窓を水平・垂直・フレーム方向に各々適用した三次元配列に対して三次元 DFT を施す。

⑤では三次元 DFT 係数に位相保存型のベイズ Shrinkage を施す。①で求めた付加雑音の標準偏差の推定値  $\hat{\sigma}_w$  と、DFT 係数のインデックスごとの 2 次モーメント  $M_{R,2}$  と 4 次モーメント  $M_{R,4}$  から ST-DFT の変換係数ごとに最適な MMSE ベイズ推定関数が得られる。この推定関数により Shrinkage を施す。

⑥では⑤の結果に対して 3 次元逆 DFT を施す。

⑦では冗長なブロック分割により個々の画素に対して複数の復元値が得られるので、窓関数による重みを考慮して荷重平均値を求める。

⑧では冗長ブロックごとに分離された平均値を補償するための加算を行う。

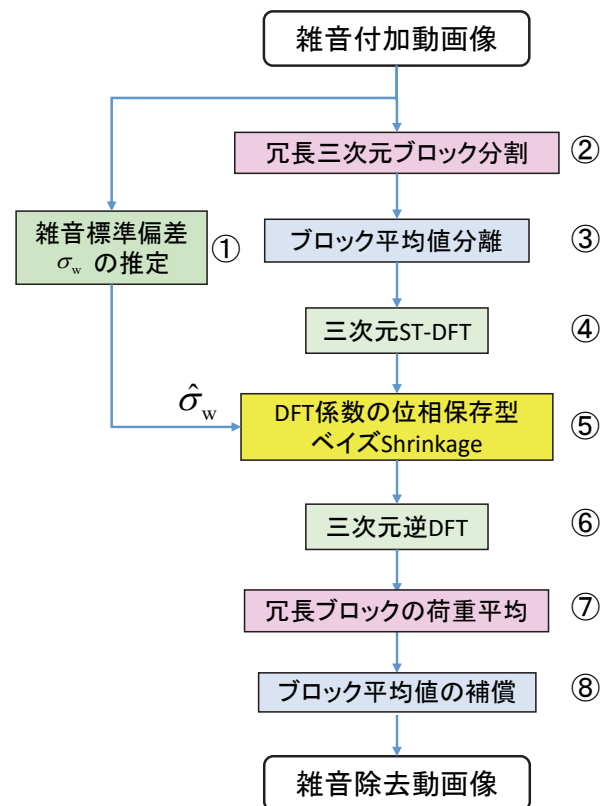


図 1 提案法のブロック図

#### 4. 三次元平均値分離型 ST-DFT

画像は大きな値の直流成分を持っている。ST-DFT では DFT の前にブロック境界で 0 となる重み窓を掛けるが、直流に重み窓を掛けると、窓関数の形状に由来する交流成分を発生してしまう。この影響を避けるためには、重み窓を掛ける前に、三次元ブロックごとに平均値を差し引いておく必要がある。復元の際には、三次元ブロックごとに差し引かれた平均値分を加算する必要がある。提案法ではこの処理を③と⑧の部分で行っている。

#### 5. 窓関数

提案法では、ST-DFT の窓関数に三角重み窓関数を用いる。タップサイズ 16 の一次元の三角重み窓関数を式(1)に示す。

$$\frac{1}{8} \frac{2}{8} \frac{3}{8} \frac{4}{8} \frac{5}{8} \frac{6}{8} \frac{7}{8} \frac{8}{8} \frac{7}{8} \frac{6}{8} \frac{5}{8} \frac{4}{8} \frac{3}{8} \frac{2}{8} \frac{1}{8} \quad (1)$$

動画像から切り出され、平均値が分離された  $16 \times 16 \times 16$  のブロックに対して、式(1)の重み窓を垂直・水平・フレーム方向に掛ける。この窓関数はブロックの切り出しを 2 のべき乗飛ばしで行っても完全再構成性を有している。提案法では各方向とも 2 画素飛ばしでブロックの切り出しを行っている。2 画素とばしでブロックの切り出しを行う場合、演算量は画素とばしを行なわない場合と比較し、演算量を 1/8 に削減できる。この時、復元性能の極めて僅か劣化に留まる[5]。

#### 6. ST-DFT 係数の位相保存型 Shrinkage

動画像から切り出された三次元ブロックに対して、平均値分離の後、三次元 ST-DFT すると、変換係数は複素数となる(ただしブロックサイズが偶数の場合、直流変換係数と 7 つの交流変換係数は実数となる)。変換係数に含まれる雑音由来の成分を除去するために、複素数の振幅成分に対してのみ Shrinkage 処理を行う。この処理を位相保存型 Shrinkage と呼ぶ。

#### 7. DFT 係数のモデルパラメータの推定と MMSE ベイズ推定関数

##### 7.1 モデル化

ST-DFT の複素変換係数を式(2)に示すように、二次元混合型球対称ガウス分布とモデル化する。

$$p(\vec{y}) = P_0 \times g_n(\vec{y}; \vec{0}, \sigma_w^2) + P_1 \times g_n(\vec{y}; \vec{0}, \sigma_w^2 + \sigma_s^2) \quad (2)$$

$P_0$  は付加雑音  $\sigma_w^2$  のみが観測される確率を、 $P_1$  は雑音の加わった信号  $\sigma_w^2 + \sigma_s^2$  が観測される確率を表す。

モデルパラメータ  $P_0, P_1, \sigma_w^2, \sigma_s^2$  の値が判ると、式(3)にて観測信号  $\vec{y}$  から原信号  $\vec{x}$  を推定する MMSE ベイズ推定関数が求まる[4]。

$$\vec{\hat{x}}(\vec{y}) = c_1 \cdot \frac{\vec{y}}{1 + c_2 \cdot \exp(-c_3 \cdot \|\vec{y}\|^2)} \quad (3)$$

$$; c_1 = \frac{\sigma_s^2}{\sigma_w^2 + \sigma_s^2}, c_2 = \frac{P_0}{P_1} \left( \frac{\sigma_w^2 + \sigma_s^2}{\sigma_w^2} \right), c_3 = \frac{1}{2} \frac{\sigma_s^2}{\sigma_w^2 (\sigma_w^2 + \sigma_s^2)}$$

##### 7.2 モデルパラメータ推定

$P_0 + P_1 = 1$  の関係と、新たなパラメータ  $\rho$  とにより、 $\sigma_w^2, \sigma_s^2$  の関係を式(3)とすると

$$\sigma_s^2 = \rho \cdot \sigma_w^2, \quad \rho > 0 \quad (4)$$

$\{P_0, P_1, \sigma_w^2, \sigma_s^2\}$  の 4 つのパラメータ値の推定する問題は、 $\{P_1, \sigma_w^2, \rho\}$  の 3 つのパラメータの推定問題となる。ある DFT 係数  $\vec{y}$  の一次、二次、四次モーメントを  $M_{R,1}, M_{R,2}, M_{R,4}$  とすると、各モーメントとパラメータ  $\{P_1, \sigma_w^2, \rho\}$  とには式(4)の連立方程式の関係が成り立つ。

$$\begin{cases} M_{R,2} = 2\lambda_{(n),2} \cdot (1 + \rho P_1) \cdot \sigma_w^2 \\ M_{R,4} = 4\lambda_{(n),4} \cdot (1 + 2\rho P_1 + \rho^2 P_1) \cdot \sigma_w^4 \\ M_{R,1} = \sqrt{2}\lambda_{(n),1} \cdot \{(1 - P_1) + P_1 \sqrt{1 + \rho}\} \cdot \sigma_w \end{cases} \quad (5)$$

$\sigma_w^2$  は信号  $\vec{y}$  中に含まれる雑音の電力を表す。画像信号は主に低い周波数成分に集中し、極めて高い周波数成分は概ね雑音に由来する。静止画像の場合高い周波数を有する変換係数の値から雑音の標準偏差を指定できる。動画像の場合、フレーム方向に高い周波数成分が表れることはほとんど無く、時空間周波数が共に高い変換係数のサンプル値から、混入している雑音の標準偏差  $\sigma_w$  を容易に推定できる[6]。  $\sigma_w$  の値が不明の場合、式(5)から  $\{P_1, \sigma_w^2, \rho\}$  を求めるためには、Stefensen の反復解法が必要となる。しかし、 $\sigma_w$  の値が何らかの手法によりあらかじめ求めることができる場合、式(4)の連立方程式の解  $\{P_1, \rho\}$  は  $\sigma_w^2$  と二次、四次モーメント  $M_{R,2}, M_{R,4}$  から、式(6)により反復演算を行うことなく求めることができる。この導出の際、一次モーメント  $M_{R,1}$  の値は不要となる。

$$\rho = \frac{\lambda_{(n),2} M_{R,4} - 4\lambda_{(n),4} M_{R,2} \sigma_w^2 + 4\lambda_{(n),4} \lambda_{(n),2} \sigma_w^2}{2\lambda_{(n),4} \sigma_w^2 (M_{R,2} - 2\lambda_{(n),2} \sigma_w^2)} \quad (6)$$

$$P_1 = \frac{\lambda_{(n),4} (\sigma_w^2 - 2\lambda_{(n),2} \sigma_w^2)^2}{\lambda_{(n),2} (\lambda_{(n),2} M_{R,4} - 4\lambda_{(n),4} M_{R,2} \sigma_w^2 + 4\lambda_{(n),4} \lambda_{(n),2} \sigma_w^2)}$$

$\sigma_w^2$  と  $\rho$  の値が判ると式(3)より  $\sigma_s^2$  が求まる。  $P_1$  の値が判ると、  $P_0 + P_1 = 1$  より、  $P_0$  の値も定まる。  $P_0, P_1, \sigma_w, \sigma_s$  の値が決定すると、式(2)にて観測信号  $\vec{y}$  から原信号  $\vec{x}$  を推定する MMSE ベイズ推定関数が求まる。

#### 8. ST-DFT 係数の位相保存型 Shrinkage の処理手順

本提案法では、ブロックサイズ  $16 \times 16 \times 16$  で切り出しを行い、平均値分離 3 次元 ST-DFT を行う。動画像画像シーケンスに全体で ST-DFT 係数のインデックスごとに二次、四次モーメントを求め、7. の手順で求めたベイズ推定関数とその係数の振幅成分のみに適用することで位相保存型 Shrinkage を行う。ただし、式(2)のベイズ推定関数を導出する際に式(7)、式(8)の 2 つの制約条件を仮定している。

$$M_{R,2} > \sigma_w^2 \quad (7)$$

$$M_{R,4} > 2M_{R,2}^2 \quad (8)$$

式(1)条件を満たさない場合、即ち ST-DFT 係数の二次モーメントが雑音の分散値未満の場合、ST-DFT 係数は雑音成分のみを含むと見なされ式(9)の NULL Shrinkage が適用される。

$$\tilde{\mathbf{x}}(\mathbf{y}) = \mathbf{0} \quad (9)$$

式(7)は満たされるが式(8)の条件が満たされない場合、式(2)において  $P_0=0, P_1=1$  と考え、 $c_2=0$  より、その ST-DFT 係数に対しては式(10)に示す単一の線形 Shrinkage を適用する。

$$\tilde{\mathbf{x}}(\mathbf{y}) = \frac{\sigma_s^2}{\sigma_w^2 + \sigma_s^2} \cdot \mathbf{y} \quad (10)$$

式(7),式(8)を共に満たす場合、式(3)のベイズ Shrinkage を適用する。以上の処理手順を図 2 に示す。

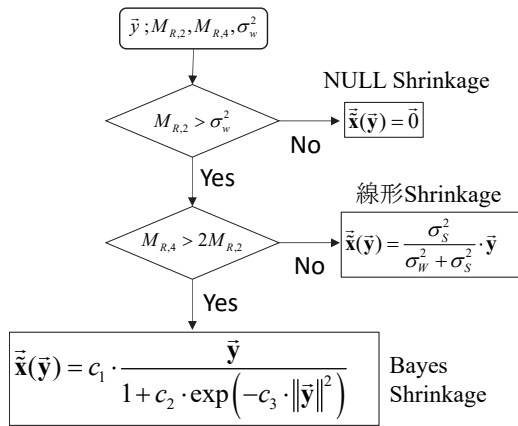


図 2 位相保存型 Shrinkage の処理手順

### 9. 冗長ブロックの荷重平均

位相保存型 Shrinkage が適用された ST-DFT 係数に三次元逆 DFT 変換を施すと、サイズ  $16 \times 16 \times 16$  の雑音除去された冗長ブロック動画像が得られる。各画素に対して複数のブロックから輝度値が得られるので、窓関数による重みを考慮して、これらの重み付き平均を求めることで、ブロック平均値の分離された復元動画像が得られる。

### 10. ブロック平均値の補償

③でブロックごとに分離した平均値に相当する値を⑦の復元動画像に加えることで最終的な復元動画像が得られる。

### 11. シミュレーション

ITE の標準動画像を用いて提案法の性能評価を行った。実験には表 1 の動画像を用いた。各動画像の輝度信号  $Y$  のみをモノクローム原動画像として用いた。

モノクローム原動画像に標準偏差約 13.6 の白色ガウス性雑音を加え雑音付加テスト動画像とした。雑音の加えられた動画像の PSNR は約 25.42dB であった。

#### 11.1 付加雑音の標準偏差の推定

雑音付加テスト動画像を用いて、付加雑音の標準偏差の推定を行った (図 1 の①)。表 2 に時空間周波数の共に高い 3 次元 DFT 係数から推定した付加雑音の標準偏差と、実

際に付加された雑音 (= 雑音付加画像 - 原画像) から求めた付加雑音の標準偏差の真値を示した。表 2 の結果より、推定法により概ね正しく付加雑音の標準偏差が推定されていることがわかる。

#### 11.2 平均 PSNR による復元性能の比較

雑音付加動画像に提案法を適用し、最終的に得られた復元動画像と原動画像との PSNR を求めた。現在最高水準の動画像雑音除去法とされている BM4D 法[7]による結果との

表 1 実験に使用した画像

| 動画像名                    | フレーム番号     |
|-------------------------|------------|
| Walk through the square | # 600~#695 |
| Whale Show              | # 0~#95    |
| Harbor                  | # 600~#695 |
| Bird Cage               | # 600~#695 |
| Crosswalk               | # 600~#695 |
| Intersection            | # 600~#695 |
| Flamongoes              | # 600~#695 |
| Driving                 | # 600~#695 |
| JapaneseRoom            | # 600~#695 |

表 2 付加雑音の標準偏差の推定精度

| テスト動画像名                 | 付加雑音の標準偏差 |       |
|-------------------------|-----------|-------|
|                         | 推定値       | 真値    |
| Walk through the square | 13.83     | 13.79 |
| Whale Show              | 13.85     | 13.80 |
| Harbor                  | 13.61     | 13.65 |
| Bird Cage               | 13.79     | 13.84 |
| Crosswalk               | 12.94     | 13.08 |
| Intersection            | 13.69     | 13.67 |
| Flamongoes              | 13.50     | 13.53 |
| Driving                 | 13.74     | 13.73 |
| JapaneseRoom            | 13.91     | 13.83 |

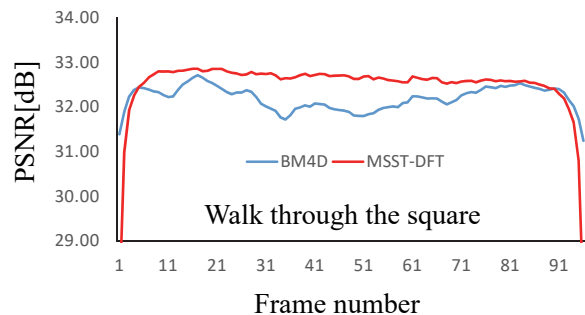


図 3 復元画像のフレームごとの PSNR

比較評価を行った。図 3 に雑音付加テスト動画像 Walk through the Square のフレームごとの PSNR を示した。赤が提案法、青が VB4D 法である。図 3 より提案法では、先頭

と最後の約 7 フレームの PSNR が大きく低下していることがわかる。提案法では三次元 ST-DFT を行う際に動画像の先頭と末尾のフレームを巡回シフトにより繋げて処理するため、この部分では不連続フレームの影響を被る。そこで、以下の評価では、全フレームの PSNR の平均値に加えて、先頭と最後の 7 フレームを除いたフレームのみから求めた PSNR の平均値も評価した。表 3 に BM4D と提案法 (MSST-DFT) による復元動画像の平均 PSNR を示した。両方式とも上段が全 96 フレームの、下段が中央 82 フレームの PSNR の平均値を表している。PSNR の高い方式の値を太字で示している。Crosswalk と Driving 以外の動画像に対しては、提案法により高い平均 PSNR が得られていることがわかる。

### 11.3 平均 $R_{MAE\_IFD}$ による復元性能の比較

これまで動画像の画像品質は、個々のフレームの静止画像としての PSNR の平均値、またはフレームごとの変動の様子のグラフにより評価されてきた。筆者らは、動画像のフレーム間での画素値の変動の忠実性を評価する客観的評価法を提案した[8]。この評価法では式(11)に示す  $R_{MAE\_IFD}$  を評価する。

$$R_{MAE\_IFD} = 20 \cdot \log_{10}(255 / MAE\_IFD) \quad (11)$$

$$MAE\_IFD(\mathbf{F}, \mathbf{G})_{n,n+1} = \frac{1}{M} \sum_{(i,j) \in \Omega} \left| (F_{i,j}^{(n)} - F_{i,j}^{(n+1)}) - (G_{i,j}^{(n)} - G_{i,j}^{(n+1)}) \right|$$

$F^{(n)}, F^{(n+1)}$  : 復元動画像  $\mathbf{F}$  のフレーム  
 $G^{(n)}, G^{(n+1)}$  : 原動画像  $\mathbf{G}$  のフレーム  
 $\Omega$  : 画像領域  
 $M$  :  $\Omega$  内の画素数

$R_{MAE\_IFD}$  は PSNR と同様に、大きな値ほど、フレーム間における画素値の変動が原動画像に忠実なものとなっていることを表す。表 4 に BM4D と提案法 (MSST-DFT) による復元動画像の  $R_{MAE\_IFD}$  を示した。両方式とも上段が全 95 フレームの、下段が中央 81 フレームの  $R_{MAE\_IFD}$  を表している。 $R_{MAE\_IFD}$  の高い方式の結果を太字で示している。

表 3 平均 PSNR による性能評価 (単位: dB)

| PSNR     | Walk Square  | Whale show   | Harbor       | Bird Cage    | Crosswalk    | Intersection | Flamingoes   | Driving      | Japaneseroom |
|----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| VBM4D    | 32.19        | 29.00        | 30.93        | 33.71        | <b>33.41</b> | 33.61        | 32.49        | <b>32.64</b> | 34.73        |
|          | 32.21        | 28.83        | 30.97        | 33.76        | <b>33.45</b> | 33.69        | 32.50        | <b>32.67</b> | 34.81        |
| MSST-DFT | <b>32.48</b> | <b>29.60</b> | <b>32.08</b> | <b>33.96</b> | 32.30        | <b>34.42</b> | <b>33.15</b> | 31.81        | <b>34.93</b> |
|          | <b>32.67</b> | <b>29.64</b> | <b>32.37</b> | <b>34.19</b> | 32.42        | <b>34.52</b> | <b>33.31</b> | 31.90        | <b>35.23</b> |

表 4 平均  $R_{MAE\_IFD}$  による性能評価 (単位: dB)

| $R_{MAE\_IFD}$ | Walk Square  | Whale show   | Harbor       | Bird Cage    | Crosswalk    | Intersection | Flamingoes   | Driving      | Japaneseroom |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| VBM4D          | <b>33.34</b> | 30.52        | 30.52        | 34.64        | <b>36.53</b> | 38.32        | 35.55        | <b>33.95</b> | 36.32        |
|                | 33.39        | 30.45        | 30.45        | 34.61        | <b>36.58</b> | 38.36        | 35.57        | <b>33.97</b> | 36.36        |
| MSST-DFT       | <b>33.34</b> | <b>30.75</b> | <b>30.75</b> | <b>35.33</b> | 35.79        | <b>39.85</b> | <b>36.29</b> | 33.39        | <b>36.80</b> |
|                | <b>33.82</b> | <b>30.86</b> | <b>30.86</b> | <b>35.50</b> | 35.93        | <b>39.92</b> | <b>36.40</b> | 33.47        | <b>37.06</b> |

## 12. むすび

筆者らは、三次元 DFT が三次元 DCT よりも優位であることを示してきた。また、DFT 係数の確率分布モデルとして、雑音を陽な形で含む 2 成分混合型球対称ガウス分布モデルを想定することで、最小二乗誤差の観点から最適なベイズ推定関数を求める手法を提案した。本稿では、三次元平均値分離型短時間 DFT 復元法の具体的な処理手順について述べた。また、実際の動画像を用いた雑音除去シミュレーションにより、BM4D との復元性能比較を行ない、多くの動画像に対して提案法により高い評価値が得られることが分かった。

### 参考文献

- [1] 小松隆, 張鍵, 齊藤隆弘 「平均値分離型三次元 ST-DFT を用いた動画像づく現法」 電子情報通信学会論文誌 D, Vol. J100-D, No. 9, pp. 842-845, 2017 年 9 月.
- [2] 齊藤隆弘, 小松隆 「劣化画像復元のための DFT 係数の確率分布モデル: 多次元混合型球対称ガウス分布モデルとそのパラメータ推定」 FIT2017, I-002, 2017 年 9 月.
- [3] J.S.Lim, "Image restoration by short space spectral subtraction", IEEE Trans. ASSP, Vol. ASSP-28, no. 2, pp. 191-197, April 1980.
- [4] 齊藤隆弘, 小松隆, 「劣化画像復元のための DFT 係数の確率分布モデル: 多次元混合型球対称ガウス分布モデルとそのパラメータ推定」, FIT2017, 4H-1, (2017)
- [5] 小松隆, 齊藤隆弘 「平均値分離型三次元 ST-DFT の演算量削減」 電子情報通信学総合大会, D-11-29, 2018 年 3 月.
- [6] 小松隆, 齊藤隆弘 「三次元平均値分離型 ST-DFT 変換係数の雑音分散推定」 映像情報メディア学会年次大会(発表予定), 2018 年 9 月.
- [7] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video Denoising, Deblocking and Enhancement Through Separable 4-D Nonlocal Spatiotemporal Transforms", IEEE Trans. Image Process., vol. 21, no. 9, pp. 3952-3966, September 2012.
- [8] 野村祐介, 小松隆, 齊藤隆弘 「雑音除去動画像の客観的画質評価法の一検討」, 電子情報通信学会総合大会, D-11-14, 2017 年 3 月.