

DCTによる次元圧縮と事例選択を用いた ビデオ超解像アルゴリズムの高速化

渡邊清高^{†,††} 岩井儀雄^{††} 羽下哲司[†] 谷内田正彦^{††}

[†] 三菱電機株式会社先端技術総合研究所 ^{††} 大阪大学大学院基礎工学研究科

学習ベース超解像アルゴリズムは、事例を蓄積するために大量のメモリを消費する点と、データベース中の事例を最近傍探索するための計算負荷が大きい点が問題であった。本論文では、メモリ消費量と計算負荷の小さい、学習ベースのビデオ超解像アルゴリズムを新たに提案する。次元圧縮を行うために、特徴ベクトルの成分として DCT 係数を利用する。また、コンパクトなデータベースを構築するための事例選択手法を与える。MPEG テストシーケンスを用いたシミュレーション実験により、提案手法は生成画像の画質を保ちながら超解像処理の効率を改善できることが示された。

A Fast Algorithm of Video Super-Resolution Using Dimensionality Reduction by DCT and Example Selection

Kiyotaka Watanabe^{†,††} Yoshio Iwai^{††} Tetsuji Haga[†] Masahiko Yachida^{††}

[†] Advanced Technology R&D Center, Mitsubishi Electric Corporation

^{††} Graduate School of Engineering Science, Osaka University

In learning-based super-resolution algorithms, there are two major problems. One is that they require a large amount of memory to store examples, the other is high computational cost to find nearest neighbors in the database. In this paper, we propose a novel learning-based video super resolution algorithm with less memory requirements and computational cost. To this end, we adopt discrete cosine transform (DCT) coefficients for feature vector elements. Moreover, we design an example selection procedure to construct a compact database. We conducted evaluative experiments using MPEG test sequences to synthesize high resolution video. Experimental results show that our method can improve effectiveness of super-resolution algorithm, while preserving the quality of synthesized image.

1 はじめに

近年のマルチメディア技術の発達に伴って、従来よりも高品質な映像を扱える製品が民生用でも普及し始めている。例えば、デジタルビデオカメラでは約 200 万画素のフル HD (High Definition) カメラが販売されている。スチルカメラでは 1000 万画素を超える製品も市販されているが、このような高解像度で動きの滑らかな動画画像を撮像することはできない。これは、カメラの画像掃き出し速度に制限があり、解像度とフレームレートの両方を同時に向上させることが困難なためである。デジタルシネマ等の特別な用途では高解像度と高フレームレートの両立が実現されているが、このようなシステムは特殊で高価な機材を用いて実現されており、一般的な利用はコストの面で非現実的である。そこで、ハードウェアの限界を超える高解像度の映像を得るため、画像処理によって高解

像度化する手法が研究されている。

画像の高解像度化の簡便な手法として、最近傍法、双一次補間法、3 次スプライン法などの補間法(内挿法)[1]がある。これらの手法は処理が比較的単純で広く用いられているが、ぼけた画像が生成される。

より高度な手法として、一枚以上の低解像度画像を用いて高解像度画像を得る超解像(super-resolution)が研究されている。超解像のアルゴリズムは、再構成ベースの手法と学習ベースの手法に大別される。

再構成ベースの手法は信号処理のアプローチによって高解像度画像を生成する手法で古くから研究されている。しかし、サブピクセルのずれをもつ複数の低解像度画像が得られているという前提が必要のため、適用できる範囲は限定的である。

一方、学習ベースの手法は、低解像度画像と高解像度画像間の相関関係を直接利用して画像を高

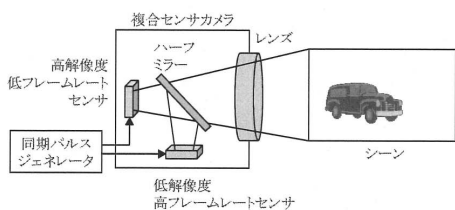


図 1: 複合センサカメラの概念図

解像度化するアプローチで、近年活発に研究が進められている。まず学習段階では、低解像度画像とそれに対応する高解像度画像を訓練用データとして組で与える。そして、これらの画像から小ブロックを切り出して特徴量を計算し、1組の事例として対応付けてデータベースに蓄える。高解像度化処理の段階では、低解像度画像に対する最近傍をデータベース中から探索し、それに対応する高解像度画像のデータを用いることにより高解像度画像を生成する。Bakerらは再構成ベース超解像の理論限界を示すと同時に、大量の事例を用いることにより文書画像や顔画像の高解像度化が行えることを示した[2]。また、Freemanらは学習による超解像の概念を一般の画像に適用した結果を示した[3]。その他、一般の静止画を対象とした手法として[4, 5, 6]などが、動画画像を対象とした手法として[7, 8]などがある。学習ベース超解像は、事例データベースを用いることによって一枚の画像から高解像度化が行える点が特徴である。

筆者らのグループでは、高品質な動画画像を得るアプローチとして、時空間分解能の異なる2種類の動画画像を同一視野で同時に撮像できる複合センサカメラ(図1)を提案している[9]。複合センサカメラは、高解像度低フレームレート動画画像と低解像度高フレームレート動画画像を同一視野で同時に撮像できるカメラである。2つのセンサに同期パルスを入力することにより、センサ間で同期したフレーム(キーフレーム)を取得できる。また、これら2種類の動画画像から高解像度高フレームレート動画画像を生成するアルゴリズム(図2)も提案している[10, 11, 12]。これらの手法では、高フレームレート動画画像を用いて推定した動きの情報とキーフレームの高解像度テクスチャを利用し、高解像度フレームを合成している。本論文では、図2に示す高解像度高フレームレート動画画像生成の問題に対し、学習ベース超解像の概念を適用した新しい画像生成アルゴリズムを提案する。

学習ベース超解像の従来手法の問題点として、事

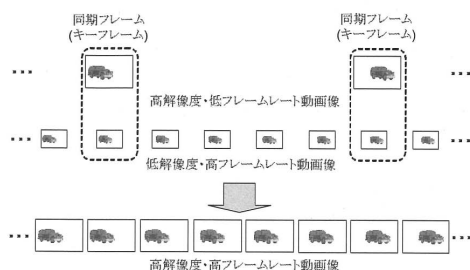


図 2: 2種類の動画画像からの高解像度動画画像生成

例を蓄積するために大量のメモリを消費するという点と、事例の最近傍探索のための計算負荷が大きいという点が挙げられる。これらの問題の解決策として、以下のアプローチが考えられる。

1. 最近傍探索の高速アルゴリズムを採用する。
2. 事例の特徴ベクトルの次元を削減する。
3. データベースに蓄積する事例数を削減する。

1については、Freemanら[3]が多次元データに対する最近傍探索の高速アルゴリズム[13]を用いている例がある。本研究では、近似的に最近傍探索を行うANN(Approximate Nearest Neighbor)探索アルゴリズム[14]を利用する。ANNは、微小な誤差を許容することにより最近傍の事例を高速に求めることができる。

2および3は、事例データベースのメモリ消費量および最近傍探索の計算量が、いずれもデータベース内の事例数と特徴ベクトルの次元のオーダーとなることから、有効な解決策である。2については、例えばBishopら[7]は主成分分析を用いて147次元の特徴ベクトルを20次元に圧縮することにより、事例探索のコスト削減を図っている。しかし主成分分析はその計算自体のコストが大きい。そのため、動画画像のように大量のデータを扱う場合には不適である。そこで提案手法では、画像の小ブロックに対してDCT(discrete cosine transform: 離散コサイン変換)を適用し、その低周波成分の係数のみの特徴量として用いることにより次元を圧縮する。DCTには高速演算アルゴリズムが存在するため、入力画像からの特徴量の計算、および次元圧縮処理の高速化を図ることができる。

3は、従来の学習ベース超解像の手法では考慮されていないアプローチである。従来研究では、訓練データから得られる多様な事例を十分に格納できるだけのメモリ容量を利用している。本論文では、蓄積できる事例数に制約があるという条件下

で、画像の高解像度化のために有効な事例を選択するアルゴリズムを提案する。

以降、提案するビデオ超解像アルゴリズムを2章で説明する。3章ではMPEGテストシーケンスを用いたシミュレーション実験の結果を示す。最後に、4章で本論文をまとめる。

2 高解像度動画画像生成アルゴリズム

提案手法では、静止画を対象とした学習ベース超解像の手法[4, 5]を動画画像へ拡張し、さらに特徴量としてDCT係数を利用する。図2に示すように、高解像度低フレームレート動画画像と低解像度高フレームレート動画画像が得られており、それらは同一視野であるとする。また、同期撮像されたフレーム(キーフレーム)がある一定時間間隔で得られるとする。提案手法では、キーフレームを訓練用画像として用いて事例データベースを構築し、このデータベースを用いて低解像度画像の高解像度化を行う。

なお、人間の視覚は色の空間的な変化に鈍感なことから、本章で述べる超解像処理は画像の輝度成分のみに対して行う。色差成分については超解像処理を行わず、低解像度画像からの双三次スプライン補間によって生成する。これにより、高解像度化処理の計算負荷の軽減を図る。

2.1 キーフレームを用いた学習

学習ベース超解像の手法では、多様な事例をデータベースに蓄積するためには大量のメモリ容量を確保する必要がある。もし十分なメモリ容量があれば、訓練用画像(キーフレーム)から得られる全ての事例をデータベースに格納できる。ここでは、メモリ容量に制約があるために全ての事例を格納できない場合について考える。本論文では、以下の2つのデータベース

- キーフレーム内からランダムに選択した事例からなるデータベース $\mathcal{D}_1$
- $\mathcal{D}_1$ 内の最近傍事例とのユークリッド距離が大きい事例からなるデータベース $\mathcal{D}_2$

を構築し、これらを統合することにより事例データベース $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ を得る方法を提案する。以後、 $\mathcal{D}$ に蓄積できる事例数の上限を n 、 $\mathcal{D}$ に蓄積されている事例数を $|\mathcal{D}|$ と表記することとする。

高解像度画像 I_H と低解像度画像 I_L の組(キーフレーム)が入力されたときの学習の手順は以下の通りである。

1. I_L を双三次スプライン補間により拡大し、 I_H と同じサイズの拡大画像 I_H^L を生成する。

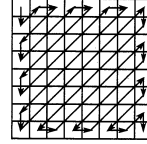


図 3: DCT 係数のジグザグスキャン

2. はじめに $\mathcal{D}_1 = \mathcal{D}_2 = \emptyset$ とする。また、 $\mathcal{D}_1, \mathcal{D}_2$ に蓄積させる事例数の上限 n_1, n_2 ($n = n_1 + n_2$) を適当に定める。
 3. (事例のランダムサンプリング) $|\mathcal{D}_1| = n_1$ になるまで以下の処理を繰り返す。
 - (a) I_H^L の中からサイズ $K \times K$ の小ブロック B^ℓ を切り出す。ただし、ブロックの切り出しはランダムかつ重複しないように行う。
 - (b) ブロックのコントラスト $c^\ell = E[|x^\ell - \mu^\ell|]$ を求める。ここで $\mu^\ell = E[x^\ell]$ はブロック内の画素 $x^\ell \in B^\ell$ の平均値である。
 - (c) c^ℓ が閾値 θ_c より小さければ、このブロックについては以降の処理を行わない。そうでなければ、このブロックをデータベースに追加するため、(d) に進む。
 - (d) B^ℓ に対して DCT を実行し、DCT 係数 $C[B^\ell]$ を求める。そして、 $C[B^\ell]$ の AC 成分のうち、低周波成分から順番に d 個の係数を取り出して d 次元ベクトル $\mathbf{v}^\ell$ を作る。係数を選択する順序は、JPEG 符号化で用いられているジグザグスキャン(図3参照)に従う。
 - (e) 高解像度画像 I_H についても同様に、 B^ℓ と同じ位置からサイズ $K \times K$ のブロック B^h を切り出し、DCT 係数 $C[B^h]$ を求める。
 - (f) $(\mathbf{v}^\ell, C[B^h])$ を1組の事例としてデータベース $\mathcal{D}_1$ に追加する。
- ステップ3で切り出された B^ℓ のうち、コントラスト c^ℓ が閾値 θ_c 以上で、事例として $\mathcal{D}_1$ に追加されたブロックの集合を $\mathcal{B}_1$ と表記する。
4. (最近傍探索による事例選択) 優先度付き待ち行列 $\mathcal{Q}$ を用意し、はじめは $\mathcal{Q} = \emptyset$ とする。 I_H^L に対して以下のステップをラスタスキャンの順番で実行する。
 - (a) I_H^L の中からサイズ $K \times K$ の小ブロック $B^\ell \notin \mathcal{B}_1$ を切り出す。
 - (b) B^ℓ に対してステップ3(b)からステップ3(e)までの処理を実行し、事例 $(\mathbf{v}^\ell, C[B^h])$ を求める。
 - (c) $\mathbf{v}^\ell$ の最近傍を $\mathcal{D}_1$ の中から探索する。見つ

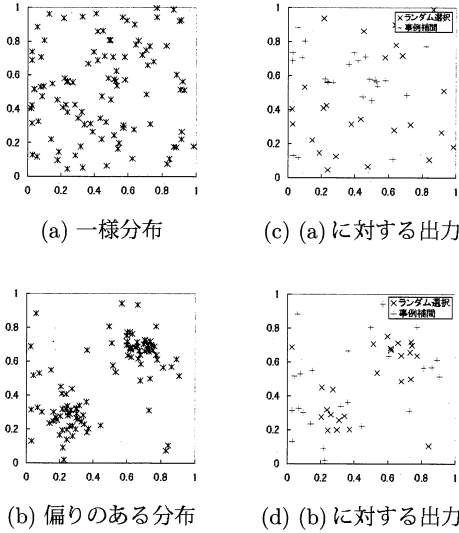


図 4: データ選択アルゴリズムの実行例

- かった最近傍の事例を v_{NN}^ℓ と表記する。そして、 v^ℓ と v_{NN}^ℓ の間のユークリッド距離 $\text{dist}(v^\ell, v_{NN}^\ell)$ を測る。
- (d) $\text{dist}(v^\ell, v_{NN}^\ell)$ を優先度として、 $(v^\ell, C[B^h])$ を Q に追加する。ここで、距離が大きいほど優先度が高い (距離が大きいものを優先的に Q に残す) とする。
- 上の手順が全てのブロックについて終了した時点で、 Q に格納されている優先度の高い n_2 個の事例によりデータベース $\mathcal{D}_2$ を構築する。
5. $\mathcal{D}_1$ と $\mathcal{D}_2$ の事例を統合し、事例データベース $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ を構築する。

図 4 は、ランダムサンプリングと最近傍探索による事例選択の概念を 2 次元データに対して適用した例である。図 4(a) は 100 点の一樣分布のデータ、(b) は偏った分布 (20 点の一樣分布と 40 点の正規分布 2 つ) をもつ 100 点のデータである。(c) と (d) は、それぞれ (a) および (b) から事例選択アルゴリズムによって 50 点を選択した例である。ここで、 $n_1 = n_2 = 25$ としている。最初のランダムサンプリングによって、元の分布において密な部分から多くの事例が選択される。その後の最近傍探索による事例選択では、分布が疎な部分の事例が優先的に選択される。すなわち、空間をできるだけ均質に埋めるようにデータを選択するよう機能する。

2.2 低解像度画像の高解像度化

キーフレーム以外のフレームについては、低解像度画像 I_L^ℓ を以下の手順で高解像度化する。

1. I_L^ℓ を双三次スプライン補間により拡大し、画像 $I_H^{\ell'}$ を生成する。
2. $I_H^{\ell'}$ からサイズ $K \times K$ の小ブロック B_q^ℓ を切り出し、以下の処理をブロック毎に実行する。ここで、ブロックは w 画素の幅で重複させながら切り出す。
 - (a) B_q^ℓ のコントラスト c_q^ℓ を求める。 $c_q^\ell < \theta_c$ ならばこの部分については以降の超解像処理を行わず、 $I_H^{\ell'}$ の画素値をそのまま生成画像に適用する。そうでなければ、この部分の超解像処理を行うため、(b) 以降の処理に進む。
 - (b) B_q^ℓ に対して DCT を実行し、DCT 係数ブロック $C[B_q^\ell]$ を求める。そして、 $C[B_q^\ell]$ の AC 成分のうち、ジグザグスキャン順序に従って低周波成分から d 個の係数を取り出し、 d 次元ベクトル v_q^ℓ を作る。
 - (c) (k -最近傍探索) v_q^ℓ をクエリとして事例データベース $\mathcal{D}$ に対する k -最近傍探索を実行し、 k 個の最近傍事例 $v_1^\ell, v_2^\ell, \dots, v_k^\ell$ を求める。
 - (d) (Local neighbor embedding) Gram 行列 G_q を以下のように定義する。

$$G_q = (v_q^\ell \mathbf{1}^T - L)^T (v_q^\ell \mathbf{1}^T - L) \quad (1)$$

ここで $\mathbf{1}$ は全ての要素が 1 の d 次元ベクトル、 L は k 個のベクトル $v_1^\ell, v_2^\ell, \dots, v_k^\ell$ を並べて作った $d \times k$ 行列である。連立方程式 $G_q w_q = \mathbf{1}$ を w_q について解き、 $\sum w_{qi} = 1$ ($i = 1, 2, \dots, k$) となるように正規化する。

- (e) (直流成分補正) 最近傍事例 v_i^ℓ に対応する高解像度画像の DCT 係数ブロックを $C[B_i^h]$ とする。 k 個の DCT 係数 $C[B_i^h]$ について、その直流成分 (DC 係数) をクエリ $C[B_q^\ell]$ の直流成分で置き換える。係数置換後の DCT 係数ブロックを $C[B_i^{h'}]$ とする。
- (f) $C[B_i^{h'}]$ の、重み w_q による線形結合

$$C[B_i^h] = \sum_{i=1}^k w_{qi} C[B_i^{h'}] \quad (2)$$

を求める。

- (g) $C[B_i^h]$ に対して IDCT を実行し、画像空間に変換する。

表 1: 実験に使用した MPEG テストシーケンス

シーケンス名称	画像サイズ	使用フレーム
Coast guard	352 × 288	No.0 ~ No.270
Football	352 × 240	No.1 ~ No.121
Foreman	352 × 288	No.0 ~ No.270
Hall monitor	352 × 288	No.0 ~ No.270

3. ブロックを重複させながら貼り合わせる．重複部分の画素値を平均化することにより，ブロックの継ぎ目を滑らかにする．

3 実験

提案手法の有効性を検証するため，MPEG テストシーケンスを用いたシミュレーション実験を行った．プログラムは Visual C++ 2005 により実装し，CPU: Intel Pentium4 3.0 [GHz]，RAM: 512 [MB]，OS: Windows XP の PC 上で実行させて実験を行った．なお，以下の評価実験において，各パラメータを $\theta_c = 8.0$ ， $w = 4$ ， $k = 2$ と設定した．また，DCT ブロックのサイズ K は 8 とした．

3.1 従来手法との比較

MPEG テストシーケンスを用いたシミュレーション実験により，従来手法との比較を行った．実験には表 1 に示すシーケンスを使用した．複合センサカメラから得られる動画をシミュレートする 2 種類の動画を，次のような手順に従って作成した．サイズ $M \times N$ ，フレームレート 30 [fps] のテストシーケンスの各フレームを縦横共に 4 分の 1 に縮小し，サイズ $M/4 \times N/4$ の低解像度高フレームレート動画を作成する．また，元のテストシーケンスから ρ フレーム毎にフレームを取り出し，高解像度低フレームレート動画を作成する．これら 2 種類の動画を提案手法に適用し，サイズ $M \times N$ ，フレームレート 30 [fps] の高解像度高フレームレート動画を作成する実験を行った．なお，本節で述べる実験では，最近傍探索による事例選択を行わずに，キーフレームから得られるすべての事例をデータベースに蓄積して動画の高解像度化を行った．(すなわち， n_1 を十分に大きく取り，かつ $n_2 = 0$ とした．) また，特徴ベクトルの次元 (DCT 係数の数) は $d = 20$ と設定した． d の値による画質の変化については次節で述べる．

テストシーケンス “Foreman” フレーム No. 45 の生成画像を図 5 に示す．比較のため，原画像を (a) に，双三次スプライン補間により拡大した低解像度画像を (b) に，DCT 合成法 [10] により生成した画像を (c) に示す．また，(a) から (d) を拡大し

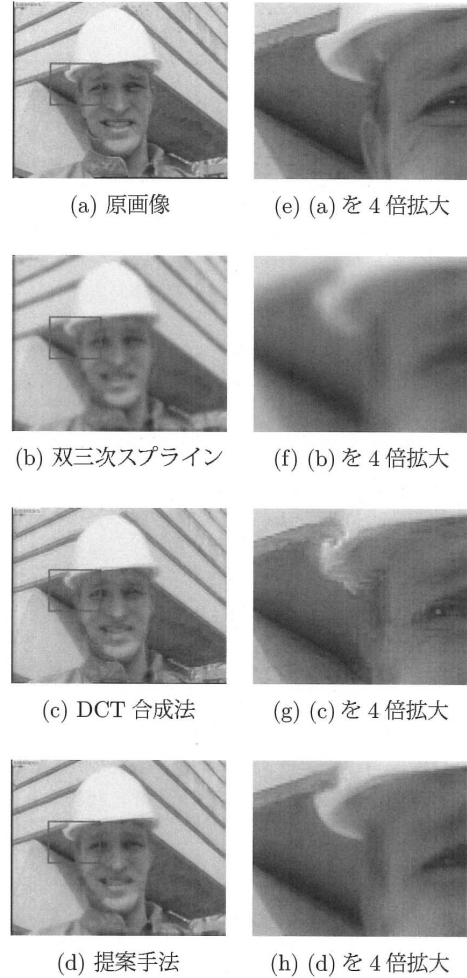


図 5: “Foreman” フレーム No. 45

たものを，それぞれ (e) から (h) に示す．提案手法による生成画像 (d)(h) を見ると，背景の斜め方向のエッジが鮮鋭に生成されていることが分かる．ただし，(h) に示す画像の中央の部分などでは鮮鋭なエッジが生成されず，双三次スプライン補間の画像 (b)(f) と同様にぼけている．これは，生成画像と時間的に対応する低解像度画像が，このテクスチャを復元するための十分なコントラストを持っていない (閾値 θ_c 以下であった) ため，この部分で超解像処理が行われなかったからである．提案手法による生成画像において，背景の壁面の細かいテクスチャが得られていないのも同様の理由による．

表 2: PSNR の測定結果

シーケンス 名称	最近傍法	双三次 スプライン	提案手法			DCT 法 [10]		
			$\rho = 7$	$\rho = 10$	$\rho = 30$	$\rho = 7$	$\rho = 10$	$\rho = 30$
Coast guard	22.01	22.05	24.72	24.49	23.99	24.59	24.14	22.53
Football	20.75	20.77	22.23	22.07	21.72	20.68	20.64	20.63
Foreman	26.13	26.26	29.93	29.76	29.21	27.41	27.01	26.46
Hall monitor	23.09	23.06	29.57	29.38	28.94	30.54	29.90	28.57

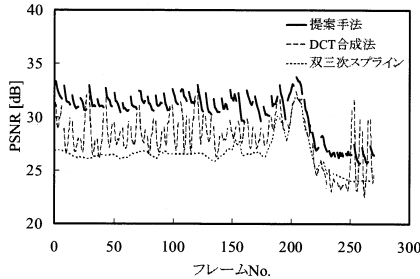


図 6: “Foreman” の PSNR の変動

各手法により高解像度化した画像と原画像の間の PSNR (Peak Signal to Noise Ratio) をシーケンス全体について測定した結果を表 2 に示す。提案手法および DCT 合成法 [10] については、入力 of 2 種類の動画のフレームレート比 ρ を変えながら画像を生成し、PSNR を測定した。テストシーケンス “Hall monitor” を除いて、提案手法は従来手法 (DCT 合成法) よりも高い PSNR を示した。本実験に使用したテストシーケンスは, “Hall monitor” は固定カメラにより撮影されたシーンであるが, その他はカメラ自体が動いたり, 被写体が動いているシーンである。DCT 合成法は動きの情報を生かすため, シーンに激しい動きが含まれているほど動き推定に失敗しやすくなり, 生成画像の画質が低下する。一方, 提案手法は動きの情報を生かすことなく高解像度化処理を行う。従って, 提案手法はシーン中に多くの動きが含まれていても高品質な映像を生成できるということが言える。なお, “Hall monitor” の PSNR についても, $\rho = 30$ の場合は提案手法が DCT 合成法を上回った。これは, DCT 合成法がキーフレームに時間的に近いフレームから順番にフレームを生成していくため, キーフレームから離れているほど生成画像の画質が低下するためである。

テストシーケンス “Foreman” を各手法に適用したときの PSNR の変動を図 6 に示す。提案手法,

表 3: 次元圧縮の影響

次元 d	Foreman		Hall monitor	
	PSNR [dB]	処理時間 [msec]	PSNR [dB]	処理時間 [msec]
3	27.89	47	26.41	48
5	29.40	56	28.95	56
10	29.88	97	29.54	89
15	29.92	114	29.56	107
20	29.93	123	29.57	111
25	29.93	129	29.57	116

および DCT 合成法については, $\rho = 7$ の入力動画画像を与えたときの PSNR であり, 線が途切れている部分がキーフレームに対応する。従来手法 (DCT 合成法) では, 生成画像とキーフレームが時間的に離れているほど動き推定の誤差が累積し, 画質が低下する。その結果, PSNR は弓なりの変動を示す。一方, 提案手法では動きの情報は生かさないため, 生成画像とキーフレームの時間的な離れ度合いに依らず PSNR は高い値を維持したままで, 変動が小さい。

3.2 次元圧縮の効果と影響

特徴ベクトルの次元 d の選び方による, 画質や処理時間への影響を調べる実験を行った。テストシーケンス “Foreman” および “Hall monitor” を利用し, 特徴ベクトルの次元 d を変化させて高解像度動画画像を生成し, PSNR および処理時間を測定した。本実験では $\rho = 7$ とし, その他の条件は前節の実験と同様である。また, 前節の実験と同様に事例選択を行わず, キーフレームから得られる全ての事例を蓄積して画像生成を行った。

表 3 に PSNR と処理時間の測定結果を示す。表 3 の処理時間は, 1 枚の高解像度画像を生成するために要した処理時間の平均である。いずれのシーケンスについても, d が 15 より小さくなると PSNR が低下していくが, $d = 15$ 以上であれば PSNR はほとんど変化しない。従って, 画像から切り出したブロックに含まれる情報のほとんどは, 15 個程度の DCT 係数で表現できるということになる。 d を 20 以上と設定しても最近傍探索の計算時間を要

表 4: データベースのサイズ配分の影響

サイズ		PSNR [dB]	
n_1	n_2	Foreman	Hall monitor
2000	8000	29.56	28.03
4000	6000	29.59	28.21
6000	4000	29.57	28.24
8000	2000	29.54	28.07
10000	0	29.41	27.69

するだけで、画質の向上は見込めないということが言える。

3.3 事例選択アルゴリズムの検証

テストシーケンス “Foreman” および “Hall monitor” を利用し、ランダムサンプリングによるデータベース D_1 と事例選択によるデータベース D_2 のサイズ配分に関する実験を行った。特徴ベクトルの次元を $d = 20$ 、データベース全体のサイズを $n = 10000$ に設定し、 n_1 と n_2 を変化させて画像を生成し、生成画像の PSNR を測定した。なお、その他のパラメータについては前節の実験と同様である。表 4 に PSNR の測定結果を示す。データベースに蓄積する事例をすべてランダムに選択する単純な方法 ($n_1 = 10000, n_2 = 0$) と比較すると、提案する二段階の事例選択アルゴリズムを用いた方が PSNR が高くなった。よって、蓄積できる事例数に制約がある場合には、データベースに蓄積する事例を単にランダムに選択するよりも提案手法を用いた方が、より画質の高い画像が生成できることが示された。また、“Foreman” および “Hall monitor” のいずれについても、 n_1 と n_2 を同程度とした場合の PSNR が最大となった。

次に、ベクトルの次元とデータベースのサイズによる画像生成処理の性能変化を検証した。実験はテストシーケンス “Foreman” を使用し、ベクトルの次元 d およびデータベースのサイズ n_1, n_2 を変化させながら画像を生成し、生成画像の PSNR と画像生成に要した時間、および使用メモリ量を測定した。なお、前の実験結果を踏まえて、 $n_1 = n_2$ と設定した。その他のパラメータについては前の実験と同様である。また、使用メモリ量の計算については、一つの実数を 8 バイトとして計算した。一組の事例データ ($v^{\ell}, C[B^k]$) は $d + K^2$ 個の実数からなることから、 n 個の事例をデータベースに蓄えるために必要な容量を $n(d + K^2) \times 8/1024$ [kB] として求めた。なお、ここでは事例そのものの蓄積に必要なメモリ量を計算しており、最近傍探索のデータ構造に必要なメモリ量などは考慮していない。

(a) $d = 5$, $n_1 = n_2 = 1000$ 

(c) (a) を 4 倍拡大

(b) $d = 10$, $n_1 = n_2 = 5000$ 

(d) (b) を 4 倍拡大

図 7: 次元圧縮と事例選択を用いた生成画像

表 5 に性能変化の実験結果を示す。表中の全事例蓄積の項目は、事例選択アルゴリズムを使用せずに画像を生成した場合の PSNR である。なお、全事例蓄積の場合は $n = 100000$ においてメモリ量を計算した。

表より、 d を小さくするほど、あるいは $n_1 = n_2$ を小さくするほど高速で、かつメモリ消費量を抑えながら画像を生成できるが、生成画像の PSNR が低下することが分かる。つまり、 d や n_1, n_2 の選び方により、画質を優先するか、計算時間およびメモリ消費量を優先するかの調整が可能となる。ただし、次元圧縮や事例選択を行って生成した画像 (図 7) を見ると、主観的な画質の劣化はそれほど大きくない。全ての事例を蓄積し次元を $d = 20$ として生成した画像 (図 5) を図 7 の画像と比較しても、画像のテクスチャが大きく歪むような劣化は見られない。すなわち、提案する次元圧縮と事例選択アルゴリズムを採用することにより、主観的な画質の低下を抑えつつ高速に高解像度動画像を生成できると言える。

4 おわりに

本論文では、学習ベース超解像のアプローチに基づく高解像度動画像生成アルゴリズムを提案した。従来の学習ベース超解像アルゴリズムにおける計算負荷とメモリ消費量の問題への対策として、DCT を用いた次元圧縮と事例選択手法を提案した。そして、複合センサカメラを想定したシミュレー

表 5: ベクトルの次元とデータベースサイズによる性能変化

データベース のサイズ $n_1 = n_2$	$d = 5$			$d = 10$			$d = 15$		
	PSNR [dB]	処理時間 [msec]	容量 [kB]	PSNR [dB]	処理時間 [msec]	容量 [kB]	PSNR [dB]	処理時間 [msec]	容量 [kB]
1000	28.45	39	1078	28.67	43	1156	28.70	46	1234
2000	28.75	42	2156	29.04	49	2312	29.08	52	2469
5000	29.12	46	5391	29.54	61	5781	29.58	66	6172
10000	29.35	49	10781	29.82	73	11563	29.85	82	12344
全事例蓄積	29.40	56	53906	29.88	97	57813	29.92	114	61719

ション実験により提案手法の有効性を検証した。実験により、動きの激しいシーンについては、動きの情報をういた従来手法よりも本手法の方が高品質な映像を生成できることを示した。また、提案手法は生成画像の画質を保ちながら超解像処理の効率を改善できることを示した。

本論文では、複合センサカメラにより撮像された2種類の動画像を入力動画像として想定したが、問題設定を変更することにより汎用のカメラに適用することも可能となる。最近のデジタルビデオカメラやスチルカメラ、携帯電話などでは、撮影モードを切り替えることにより静止画と動画像の両方を撮影することができる。静止画モードにより撮影した高解像度画像を用いて学習を行った後、動画像モードで撮影された低解像度フレームに対して高解像度化を行う、といった使い方も考えられる。今後、汎用のカメラにより得られた画像を提案手法に適用することを検討する。

参考文献

- [1] 高木幹雄, 下田陽久 (監修), “新編画像解析ハンドブック”, 東京大学出版会, 2004.
- [2] S. Baker and T. Kanade, “Limits on super-resolution and how to break them”, IEEE Trans. Pattern Anal. & Mach. Intell., Vol. 24, No. 9, pp. 1167–1183, 2002.
- [3] W. T. Freeman, E. C. Pasztor and O. T. Carmichael, “Learning low-level vision”, Int. J. Comput. Vision, Vol. 40, No. 1, pp. 25–47, 2000.
- [4] J. Sun, N.-N. Zheng, H. Tao and H.-Y. Shum, “Image hallucination with primal sketch priors”, In Proc. IEEE Conf. CVPR, Vol. 2, pp. 729–736, 2003.
- [5] H. Chang, D.-Y. Yeung and Y. Xiong, “Super-resolution through neighbor embedding”, In Proc. IEEE Conf. CVPR, Vol. 1, pp. 275–282, 2004.
- [6] C. V. Jiji and S. Chaudhuri, “Single-frame image super-resolution through contourlet learning”, EURASIP J. Applied Signal Process., Vol. 2006, pp. 1–11, 2006.
- [7] C. M. Bishop, A. Blake and B. Marthi, “Super-resolution enhancement of video”, In Proc. 9th Int. Conf. Artificial Intell. & Statistics, 2003.
- [8] D. Kong, M. Han, W. Xu, H. Tao and Y. Gong, “A conditional random field model for video super-resolution”, In Proc. ICPR, pp. 619–622, 2006.
- [9] 重本倫宏, 星川章, 長原一, 岩井儀雄, 谷内田正彦, 鈴木俊哉, “時間的・空間的分解能の異なる複合センサカメラシステム”, 情報学論:CVIM, Vol. 47, No. SIG5 (CVIM13), pp. 35–44, 2006.
- [10] K. Watanabe, Y. Iwai, H. Nagahara, M. Yachida and T. Suzuki, “Video synthesis with high spatio-temporal resolution using motion compensation and spectral fusion”, IEICE Trans. Inf. & Syst., Vol. E89-D, No. 7, pp. 2186–2196, 2006.
- [11] 渡邊清高, 岩井儀雄, 長原一, 谷内田正彦, 鈴木俊哉, “ウェーブレット領域での動き補償と画像統合による高解像度高フレームレート動画像の生成”, 情報学論:CVIM, Vol. 47, No. SIG10 (CVIM15), pp. 195–207, 2006.
- [12] 松延徹, 長原一, 岩井儀雄, 谷内田正彦, 鈴木俊哉, “モーフィングによる高解像度高フレームレート動画像の生成”, 信学論 D, Vol. J90-D, No. 4, pp. 1073–1084, 2007.
- [13] S. A. Nene and S. K. Nayer, “A simple algorithm for nearest neighbor search in high dimensions”, IEEE Trans. Pattern Anal. & Mach. Intell., Vol. 19, No. 9, pp. 989–1003, 1997.
- [14] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman and A. Y. Wu, “An optimal algorithm for approximate nearest neighbor searching in fixed dimensions”, J. ACM, Vol. 45, No. 6, pp. 891–923, 1998.