

時空間周波数の異なる画像列からの高解像度動画の合成

Generation of High Resolution Video Sequence from Two Video Sequences with Different Spatio-temporal Frequencies

渡邊 清高[†]
Kiyotaka Watanabe

岩井 儀雄[†]
Yoshio Iwai

長原 一[†]
Hajime Nagahara

谷内田 正彦[†]
Masahiko Yachida

概要 本研究では高解像度・高フレームレート動画を取得するための新しいアプローチとして、全く同一の視野を持った2種類の動画(高解像度・低フレームレート動画, 低解像度・高フレームレート動画)を同時に撮影できる複合センサカメラを考える。そして、複合センサカメラから得られる2種類の動画から、画像処理によって高解像度・高フレームレート動画を生成する時空間アップサンプリングアルゴリズムを開発する。シミュレーション実験ではMPEGテストシーケンスを用いて画像生成を行い、さらに生成画像のピーク信号対雑音比を測定することによって提案手法の有効性を確かめる。

1. はじめに

デジタルビデオカメラの撮像素子として広く利用されているCCD(電荷結合素子)は、作製技術の発達に伴って高解像度化が進んできた。しかし、カメラの画像掃き出し速度には制限があるため、一般にカメラの解像度が高くなるとフレームレートが低下する。すなわち、高解像度と高フレームレートは両立しない。もちろん、デジタルシネマ等の特殊用途に対しては高解像度・高フレームレートのカメラが存在するが、非常に高価である。そこで、一般的なカメラから得られる低解像度動画あるいは低フレームレート動画を用いて、ソフトウェア処理によって高解像度動画あるいは高フレームレート動画を生成する手法の研究が昔から行われている。

画像処理によって高解像度の画像を得る単純な方法として補間法(内挿法)がある[1]。補間法は処理が簡単であるため計算コストは小さいが、ぼけた画像が生成されるため画質は良くない。信号処理の立場からのアプローチによって高解像度画像を得る手法に超解像があり、古くから研究されている。超解像では一般に、同一のシーンから複数の低解像度画像が得られるという仮定を設ける。そして、これらの低解像度画像の間にサブピクセルレベルの微小なずれが存在すれば、超解像による画像復元が可能となる。超解像の処理は、まず低解像度画像と高解像度画像の関係を表す画像劣化過程をモデル化し、その後モデルに従って複数の低解像度画像から高解像度画像を推定するという流れとなる。劣化モデルは静止画像に対するものと動画に対するものに大別される。超解像は複雑な処理を必要とするため、オフライン処理が前提となっている。このことから、超解像は主に衛星写真の高解像度化や軍事などの特殊な分野で利用されている。静止画像に対する超解像の従来研究は[2]にまとめられている。また、動画に対する超解像の研究として[3, 4]などがある。

一方、動画のフレーム数を向上させる処理はフレームレート変換を行うことに相当する。フレームレート変換はテレビ信号規格の相互変換やスローモーション映像生成などの目的で昔から研究されている。また、ハードウェアによるフレームレート変換器も実用化されている[5]。フレームレートを向上させるための単純な手法として、フレーム反復、時間方向フィルタリングによる線形補間があるが、フレーム反復により生成された動画は動きが滑らかでない。また、線形補間によって生成された動画は、シーン中に含まれる動物体がぼけてしまう。これらの問題を解決するために動き補償付きフレームレート変換が提案されており、従来研究として[6, 7, 8]などがある。

これら従来のアプローチのほとんどは、空間または時間周波数の一方を向上させるにとどまっていた。本研究では、高解像度・高フレームレートの動画を取得するために、2種類のカメラ(高解像度・低フレームレート, 低解像度・高フレームレート)により全く同一の視野を同時に撮影できる複合センサカメラを考える。この複合センサカメラから得られる2種類の動画から、画像処理によって高解像度・高フレームレート動画を生成するアルゴリズムを提案する。このような複合センサカメラは解像度やフレームレートが制限された既存のカメラを利用することができるため、低いコストで作成できる。また、動画を蓄積する場合、その符号量を小さくできるという利点もある。従来の超解像やフレームレート変換では空間または時間の一方のみをアップサンプリングするのに対して、本手法は時空間サンプリング周波数の異なる同一視点の2種類の動画から、高解像度・高フレームレート動画を生成する時空間アップサンプリングを特徴とする。提案するアルゴリズムは、高解像度動画配信システムの将来の実用化を想定し、ネットワーク配信を前提としたものとする。

本論文の構成は次の通りである。2章では本研究で想定する高解像度画像生成問題の設定を述べる。次に、3章では高解像度画像生成の提案手法を説明する。最初に提案手法のアイデアを述べ、その後アイデアを実現する方法を示す。4章では画像生成の実験方法について述べ、生成画像とピーク信号対雑音比(PSNR)の測定結果を示す。最後に、5章で結論と今後の課題を述べる。

2. 問題設定

本研究で想定する複合センサカメラは、図1に示すように高解像度・低フレームレートの動画を撮影できる高解像度CCDと、低解像度・高フレームレートの動画を撮影できる低解像度CCDの2つのCCDセンサをもつ。カメラに入射した光はプリズムによって分割され、

[†]大阪大学大学院基礎工学研究科システム創成専攻
Department of Systems Innovation, Graduate School of
Engineering Science, Osaka University.

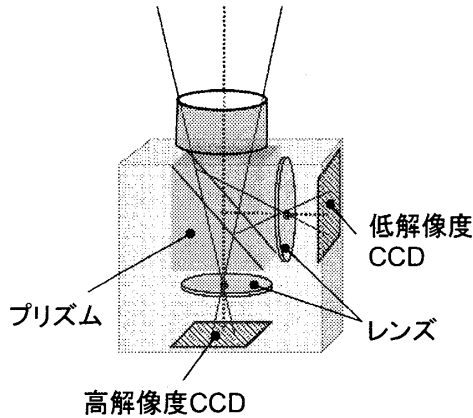


図 1: 複合センサカメラ

各 CCD に到達する。複合センサカメラについて以下の2つの仮定を設ける。

1. 複合センサカメラの2つの CCD から得られる画像は、解像度は異なるが同一の視野をもつとする。
2. 複合センサカメラは2つの CCD の外部同期を取ることができる。つまり、外部からパルスを入力することにより、時間的なずれのない高解像度画像と低解像度画像の組が得られるとする。本論文ではこのようなフレームをキーフレームと呼ぶこととする。

本研究では、以上の仮定の下で、同一の視野をもつ既知の高解像度・低フレームレート動画と低解像度・高フレームレート動画から未知の高解像度・高フレームレートの動画を生成するアルゴリズムを提案する。

3. 高解像度動画の生成

本研究では複合センサカメラから得られた動画をネットワーク配信することを想定している。動画をネットワーク配信するときに行われる動画符号化には、何らかの周波数変換により画像を周波数領域へ移す処理を含む。したがって、画像生成処理を画像空間領域と周波数領域の両方で行うことが可能である。

提案手法では、生成する高解像度画像の高周波成分と低周波成分を異なる方法によって推定する。

- 生成画像の高周波成分は、画像空間上で動き補償を行った高解像度画像のスペクトルをそのまま利用する。シーン中の動きは低解像度画像を用いて推定する。ただし、補償できない部分については低解像度画像から補間する。
- 生成画像の低周波成分は、画像空間上で動き補償された高解像度画像のスペクトルに低解像度画像のスペクトルを合成することによって生成する。

なお、本研究では周波数変換として DCT(離散コサイン変換)を用いる。

3.1 画像生成アルゴリズム

高解像度画像生成処理の流れを図2に示す。ただし、ここでは2種類の動画の画像サイズ比を2:1としている。高解像度画像の生成は次の手順で行う。

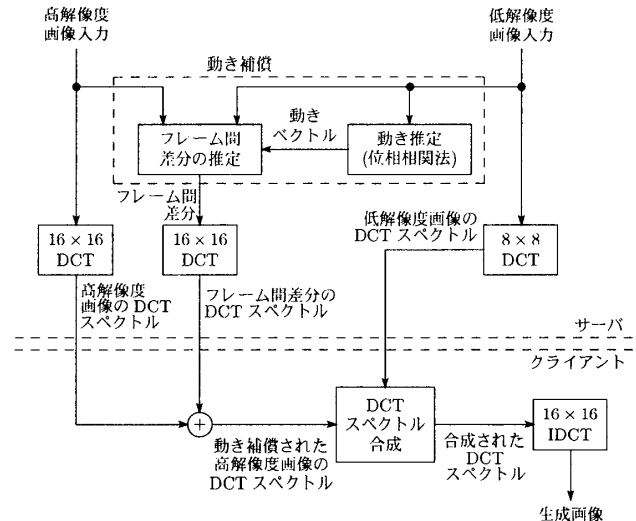


図 2: 高解像度画像生成処理の流れ

1. 位相相関法 [5, 9, 10] を用いて低解像度画像の各画素の動きベクトルを求める。
2. 1で求めた動きベクトルを用いて高解像度画像のフレーム間差分を推定する。ただし、動きベクトルが推定できなかった画素は、双一次補間法により低解像度画像から画素値を補間する。
3. 高解像度画像とフレーム間差分に対してサイズ16×16のブロックDCTを実行する。また、低解像度画像に対してサイズ8×8のブロックDCTを実行する。
4. 高解像度画像とフレーム間差分のそれぞれのDCTスペクトルをDCT領域で足し合わせることで、動き補償された高解像度画像のDCTスペクトルを求める。
5. 動き補償された高解像度画像のDCTスペクトルに対し、低解像度画像の対応する位置のDCTスペクトルを合成する。
6. 合成したスペクトルに対してIDCT(逆離散コサイン変換)を実行することにより、高解像度画像を生成する。

なお、動画のネットワーク配信を考える場合は、上の1から3までがサーバ側の処理、4から6までがクライアント側の処理となる。

3.2 フレーム間差分を用いた動き補償

提案手法ではフレーム間差分を用いて周波数領域で動き補償を行う。この原理について述べる。

今、 k 番目のフレームを I_k 、 k 番目と $k+1$ 番目のフレーム間差分を $S_{k,k+1}$ で表すと

$$I_{k+1} = I_k + S_{k,k+1} \quad (1)$$

の関係がある。すると、DCTの線形性より

$$C[I_{k+1}] = C[I_k] + C[S_{k,k+1}] \quad (2)$$

が成り立つ。ここで $C[\cdot]$ は DCT スペクトルを表す。すなわち、フレーム間差分の DCT スペクトルを前フレーム

の DCT スペクトルと足し合わせることによって、動き補償された画像の DCT スペクトルを得ることができる。

3.3 DCT スペクトル合成

一般的な画像信号は低周波成分を多くもつという性質がある。つまり、画像のもつ情報の大部分は低周波成分に集中しているということが言える。したがって、動き補償を行った高解像度画像の低周波成分に低解像度画像のスペクトルを合成することによって、生成画像を真の高解像度画像に近づけることができる。

今、動き補償された高解像度画像の DCT スペクトル (サイズ $N_h \times N_h$) を $C_h(u, v)$ とし、それに対応する位置の低解像度画像の DCT スペクトル (サイズ $N_\ell \times N_\ell$) を $C_\ell(u, v)$ とする。このとき、合成された DCT スペクトル $C(u, v)$ を

$$C(u, v) = \begin{cases} w_h(u, v)C_h(u, v) + \frac{N_h}{N_\ell}w_\ell(u, v)C_\ell(u, v) & \text{if } 0 \leq u, v < N_\ell \\ C_h(u, v) & \text{otherwise} \end{cases} \quad (3)$$

により求める (サイズ $N_h \times N_h$)。サイズの異なるスペクトルを合成するため、低解像度画像のスペクトル C_ℓ にエネルギー補正係数 N_h/N_ℓ をかける。また、 w_h, w_ℓ はスペクトル合成の重みである。本研究では $N_h = 16, N_\ell = 8$ とし、 w_h, w_ℓ として次式で表される関数を用いた。

$$w_h(u, v) = \begin{cases} 0 & \text{if } 0 \leq u, v < 4 \\ \frac{u-3}{5} & \text{if } u \geq 4 \text{ and } u \leq v \\ \frac{v-3}{5} & \text{otherwise} \end{cases} \quad (4)$$

$$w_\ell(u, v) = 1 - w_h(u, v) \quad (5)$$

4. 実験

現在、実際の複合センサカメラは開発中のため、シミュレーション実験により提案手法の評価を行った。なお、本実験では画質の評価基準として PSNR を用いた。

4.1 PSNR の定義

画像サイズが $M \times N$ 、フレーム数が K である 2 つの映像信号 $I_k(x, y)$ と $\hat{I}_k(x, y)$ の間の平均二乗誤差 σ_e^2 は次式で定義される。

$$\sigma_e^2 = \frac{1}{KMN} \sum_{k=1}^K \sum_{x=1}^M \sum_{y=1}^N [I_k(x, y) - \hat{I}_k(x, y)]^2 \quad (6)$$

PSNR は平均二乗誤差を用いて次式のように定義される。単位はデシベル (dB) である。

$$\text{PSNR} = 10 \log_{10} \frac{I_{\max}^2}{\sigma_e^2} \quad [\text{dB}] \quad (7)$$

ここで I_{\max} は信号強度の最大値で、一般的な 8 ビット信号の場合は $I_{\max} = 255$ である。

4.2 実験方法

本論文ではモノクロの MPEG テストシーケンス (30 [fps]) を用いたシミュレーション実験の結果を示す。サイズ $M \times N$ の画像からなるテストシーケンスの各フレー

表 1: 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 |
|---------------|-----------|-----------------|
| Flower garden | 352 × 240 | No. 1 ~ No. 115 |
| Football | 352 × 240 | No. 1 ~ No. 121 |
| Foreman | 352 × 288 | No. 1 ~ No. 295 |
| Hall monitor | 352 × 288 | No. 1 ~ No. 295 |

表 2: ピーク信号対雑音比の測定結果

| シーケンス名称 | PSNR [dB] | |
|---------------|-----------|-------------|
| | 提案手法 | LRI を 2 倍拡大 |
| Flower garden | 20.54 | 18.82 |
| Football | 24.42 | 23.38 |
| Foreman | 31.38 | 27.87 |
| Hall monitor | 34.58 | 25.19 |

ムに対して 50% の画像縮小を行い、サイズ $M/2 \times N/2$ の低解像度画像の系列を生成する。そして、元のテスト画像列から 6 枚毎に 1 枚フレームを取り出して 5 [fps] の高解像度動画像を作成する。また、縮小した画像を 30 [fps] の低解像度動画像とする。これら 2 種類の動画像を提案手法に適用し、サイズ $M \times N$ 、フレームレート 30 [fps] の高解像度動画像を生成する実験を行った。画像生成実験に使用した MPEG テストシーケンスを表 1 に示す。

4.3 実験結果

“Foreman” フレーム No. 82 の原画像、生成画像および拡大した低解像度画像を図 3 に示す。

“Foreman” のフレームごとの PSNR の変化を図 4 に示す。ただし、キーフレームは原画像と同一であるため、PSNR が計算できない。このため、図 4 において、キーフレームについては生成画像の PSNR (実線) を描画していない。つまり、実線の途切れている部分がキーフレームである。また、2 倍拡大した低解像度画像の PSNR を破線で示している。図 4 を見ると、200 フレームまでは提案手法の PSNR が低解像度画像を拡大したものに比べて平均 5 [dB] 程度上回っているが、200 フレーム以降において提案手法の PSNR が低解像度画像を拡大したものに比べて悪くなっている部分がある。これは、低解像度画像を用いた動き推定が失敗すると、動き補償された高解像度画像 (スペクトル合成を行う前の画像) の平均二乗誤差が大きくなる。このようなことが起こった結果、提案手法の PSNR が下がったと考えられる。より精度の高い動き推定の実現は今後の課題とする。また、各テストシーケンス全体の PSNR を表 2 に示す。表には、拡大した低解像度画像 (LRI) の PSNR も併記してある。表 2 に示しているように、全てのテストシーケンスについて、提案手法による生成画像の PSNR が 2 倍拡大した低解像度画像を上回った。

5. おわりに

本論文では、時空間周波数の異なる同一視点の 2 種類の動画像から、高解像度・高フレームレートの動画像を画像処理によって生成するアルゴリズムを提案した。そして、MPEG テストシーケンスを用いてシミュレーシ

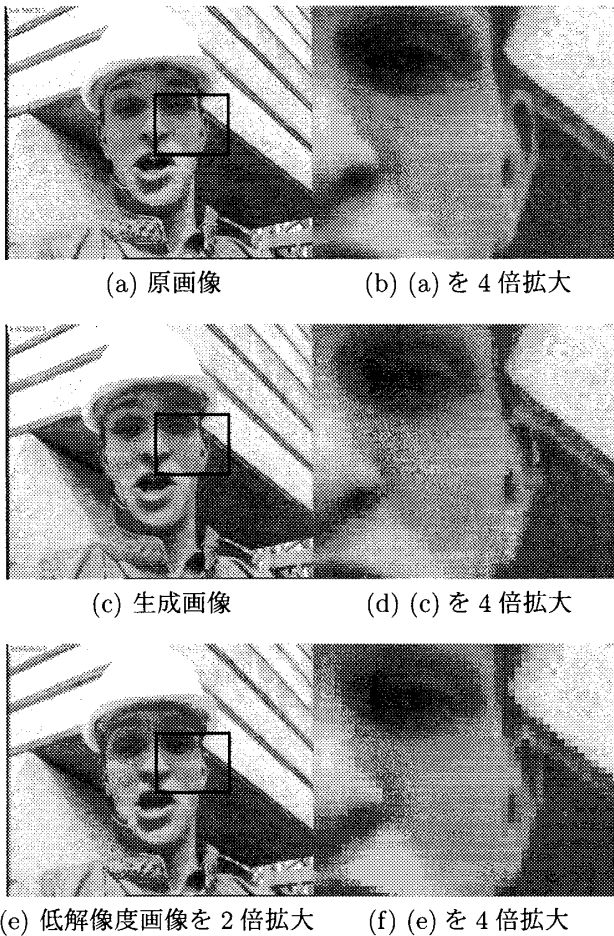


図 3: “Foreman” フレーム No. 82

ン実験を行った。PSNR を用いて生成画像を数値的に評価することにより、提案手法の有効性を確認した。

今後の課題として、画像生成処理の高速化および生成画像の高画質化が挙げられる。特に後者の問題については、動き推定の精度を向上させることによって解決を目指す。また、提案手法の時間的アップサンプリングの有効性を示すため、高解像度・低フレームレート動画に対して動き補償付きフレームレート変換を行うことによって高フレームレート動画を生成し、このようにして得られた動画と提案手法によって生成した動画を比較評価することを検討している。

謝辞

本研究の一部は独立行政法人情報通信研究機構「民間基盤技術研究促進制度」の援助を受けた。

参考文献

- [1] 高木幹雄, 下田陽久 監修, “画像解析ハンドブック”, 東京大学出版会, 1991.
- [2] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: A technical

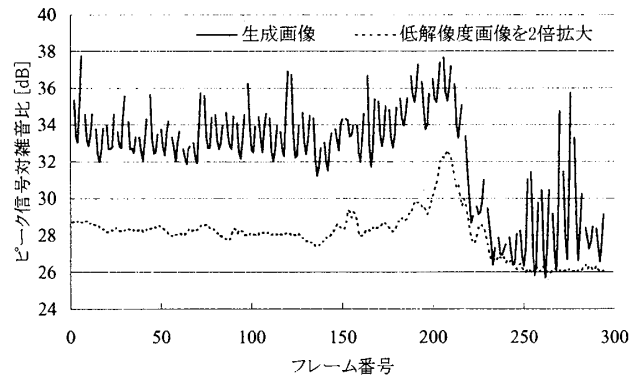


図 4: “Foreman” のピーク信号対雑音比

overview”, *IEEE Signal Processing Mag.*, Vol. 20, Issue 3, pp. 21–36, 2003.

- [3] H. Shekarforoush and R. Chellappa, “Data-driven multi-channel super-resolution with application to video sequences”, *J. Opt. Soc. Am. A*, Vol. 16, No. 3, pp. 481–492, 1999.
- [4] B. C. Tom and A. K. Katsaggelos, “Resolution enhancement of monochrome and color video using motion compensation”, *IEEE Trans. Image Processing*, Vol. 10, No. 2, pp. 278–287, 2001.
- [5] J. Watkinson, “The Engineer’s Guide to Motion Compensation”, Snell & Wilcox, 1994.
- [6] Y. K. Chen, A. Vetro, H. Sun, and S. Y. Kung, “Frame-rate up-conversion using transmitted true motion vectors”, In *Proc. IEEE 2nd Workshop on Multimedia Signal Processing*, Los Angeles, CA, pp. 622–627, 1998.
- [7] B. T. Choi, S. H. Lee, and S. J. Ko, “New frame rate up-conversion using bi-directional motion estimation”, *IEEE Trans. Consumer Electron.*, Vol. 46, No. 3, pp. 603–609, 2000.
- [8] S. H. Lee, O. Kwon, and R. H. Park, “Weighted-adaptive motion-compensated frame rate up-conversion”, *IEEE Trans. Consumer Electron.*, Vol. 49, No. 3, pp. 485–492, 2003.
- [9] G. A. Thomas, “Television motion measurement for DATV and other applications”, *BBC Research Department Report*, No. 1987/11, 1987.
- [10] M. Biswas and T. Nguyen, “A novel motion estimation algorithm using phase plane correlation for frame rate conversion”, In *Proc. 36th Annual Asilomar Conf. Signals, Syst. and Computers*, Vol. 1, pp. 492–496, 2002.