

# 超解像のための領域分割による動き検出法

田中 聡久 東京農工大学大学院共生科学技術研究院准教授

## 1 はじめに

マルチフレームの超解像は、フレーム間でサブピクセルのシフトを伴う複数の低解像度画像から、高解像度のフレームを推定する技術である。この超解像は、主に次の3つの要素技術に分けることができる: 1) 画像の位置あわせ, 2) 高解像度グリッドにおける低解像度画像の補間, 3) 画像復元 (ぼけと雑音の除去)。超解像技術については近年活発に研究が行われているが [1]-[6], [8]-[15], [17], マルチフレーム超解像に関しては、多くの研究がフレーム間の動きにあるモデルを仮定するものであった。それらの動きモデルは、多くが画像全体の動き (平行移動 [4] やアフィン変換 [5] など) を仮定している。しかしながら、動画はグローバルな動きよりも、局所的な動きの発生が頻繁に起こるものであるため、これを正確に推定することがマルチフレーム超解像実現の鍵となる。局所的な動きを推定する方法としては、オプティカルフローを用いたり [6], [12], ブロックマッチングを用いたり [13] するものが提案されている。しかしながら、前者は画素単位で動きを推定するために、同じオブジェクトに属す画素でありながら、まったく違う方向に動いていると推定する場合が発生する。後者は、オブジェクトの形状にかかわらず、画像をブロック分割することで動きを推定するため、特にオブジェクト境界において雑音が発生してしまう。

本研究報告では、これらの問題点を克服するために、画像を、任意形状をもつ複数の領域に分割し、それぞれの領域で動き推定を行うことで、画像の位置あわせを行う方法を提案する。さらに、この方法を超解像処理に適用することで、高解像度化をより高精度に行うことができることを示す。

## 2 問題設定

本稿で取り扱う問題は、次のように要約できる。同一シーンにおける連続する  $N$  フレームの低解像度画像を  $Y^k$  とし、 $M^2$  個の画素で構成されるとする。推定すべき高解像度をもつ原画像を  $X$  であらわし、 $L^2$  個の画素で構成されるとする。ここで、 $L > M$  である。 $N$  個のフレームが観測される過程は、次のように表現できる:

$$Y^k = DHF^k X + V^k,$$

ここで、 $F, H, D$  はそれぞれ、動き作用素、ぼけ作用素、ダウンサンプリング作用素である。また、 $V^k$  は加法性の雑音である。さらに、すべてのフレームは、同一のぼけとダウンサンプリング作用素をもち、それらは既知であると仮定する。また本稿では、ガウス性の雑音を仮定する。

## 3 領域分割を用いた動きの推定法

本節では領域マッチングに基づく局所動きの推定方法について述べる。この方法によって、局所的な動きがある画像系列においても、動きをうまく検出できることを示す。なお、式(1)のモデルにおいて推定すべき作用素は  $F^k$  である。

主要なステップを要約すると次のようになる。まず、図1と図2に示すように、参照フレームを任意形状の領域に分割する。領域分割には、rain-falling シミュレーションに基づく watershed 変換を用いる [18]。この watershed 変換は画像分割手法のひとつで、画像が領域 (オブジェクト) の境界で高い勾配をもち、領域内部で低い勾配をもつという現象を利用したものである。Watershed 変換による領域分割では、まず、画素値変化の小さい「弱い」エッジを除去し、ある閾値より小さい画素値をもつすべての画素をグルーピングすることで、「湖」と呼ばれる集合を形成する。このことで雑音の抑制や分割しすぎの防止を図っている。この

領域分割は、計算コスト削減のために、参照フレームのみに適用する。

このようにして得られた領域は、3 ステップ探索法 [20] や、全探索法 [21] を用いることで領域マッチングを行う。本稿では、最良の動きベクトルを得るために、全探索法の改良版を用いる。改良全探索法は次のように表現できる。各領域に対して、現在フレームと参照フレームの同一の位置で差分を計算する。この差分値が閾値より大きい場合は全探索を行い、そうでない場合は、この領域に対する動きベクトルを(0,0)であるとする。この閾値を用いることで、全探索アルゴリズムに要する計算時間を削減することが出来る。

マッチングに関して、参照フレームにおける対象領域にもっとも類似している領域への動き  $(d_x^*, d_y^*)_{R_i}$  は、二乗差分和 (SSD) の最小値

$$(d_x^*, d_y^*)_{R_i} = \operatorname{argmin}_{d_{\min} \leq d_x, d_y \leq d_{\max}} (SSD(d_x, d_y)_{R_i}), \quad (2)$$

で与えられる。ここで、 $d_{\min}$  と  $d_{\max}$  は、それぞれ最小・最大の移動量をあらわす。また、SSD は次のように定義される。

$$SSD(d_x, d_y)_{R_i} = \sum_{x,y \in R_i} |f_r(x,y) - f_c(x+d_x, y+d_y)|^2, \quad (3)$$

である。ここで、 $f_r$  は参照フレームを、 $f_c$  は現在フレームを表している。また、 $R_i$  はマッチング処理がおこなわれる領域である。

超解像への応用においては、動きベクトルの推定精度は、高解像度画像の座標精度と同等でなければならない。参照フレームも現在フレームも低解像度画像であるため、このまま動き推定を行うと、動きベクトルの精度は低いままである。したがって、低解像度フレーム  $Y^k$  を補間した画像フレーム  $\tilde{Y}^k$  を生成し、この画像列に対して前述の動き推定を行う必要がある。

#### 4 超解像への適用

前節では参照フレームからの動き  $F^k$  を求めた。これを超解像処理に適用する方法をここでは説明する。本稿では超解像処理を、以下に示す最適化問題を解くことで実現する。

$$\begin{aligned} \hat{X} &= \operatorname{argmin}_X \sum_{k=1}^N \psi(DHF^k X - Y^k) \\ &= \operatorname{argmin}_X \sum_{k=1}^N \psi(D(HF^k X - D^T Y^k)) \\ &= \operatorname{argmin}_X \sum_{k=1}^N \psi(D(HF^k X - \tilde{Y}^k)) \end{aligned} \quad (4)$$

ここで、 $\psi(\cdot)$  は画像の距離を決める関数で、 $\tilde{Y}^k$  は  $Y^k$  を補間したもの

$$\tilde{Y}^k = I(D^T Y^k) \quad (5)$$

である。 $I(\cdot)$  は補間作用素である。簡単のため、式(4)の最小化問題は、

$$\hat{X} = \operatorname{argmin}_X \sum_{k=1}^N \psi(D(HF^k X - \tilde{Y}^k)) \quad (6)$$

としてもよい。しかしながらここで、式(4)の代わりに、 $HF^k = F^k H$  を仮定することで、次の2つに分割した最適化問題を考える。

$$\hat{X} = \underset{X}{\operatorname{argmin}} \psi(HX - \hat{Z})$$

ここで,

$$\hat{Z} = \underset{Z}{\operatorname{argmin}} \sum_{k=1}^N \phi(F^k Z - \tilde{Y}^k)$$

であり,  $\phi(\cdot)$  は復元のためのコスト関数である.

文献[17]で得られた結果によると, 超解像処理によって生ずる雑音はガウス性とラプラス性のものが混合したものとなっている. したがって, L1 ノルムや L2 ノルムのような単独の評価基準を用いるよりも, これらを混合したノルムで評価関数を構成することが望ましいであろう. そこで, 超解像の補間問題は, 次のように表現できる.

$$\hat{Z} = \underset{Z}{\operatorname{argmin}} \left\{ \mu_1 \sum_{k=1}^N \|F^k Z - \tilde{Y}^k\|_1 + \mu_2 \sum_{k=1}^N \|F^k Z - \tilde{Y}^k\|_2^2 \right\},$$

ここで  $\mu_1$  と  $\mu_2$  は L1 ノルムと L2 ノルムの重要度を調整する重み係数である. この係数は Generalized

Likelihood Ratio Test (GLRT) [7] と呼ばれる数値を用いて決定する.  $\mu_1$  と  $\mu_2$  は  $\lambda_1^k$  と  $\lambda_2^k$  を  $k$  にわたって平均

した値となる. ここで,  $\lambda_1^k$  は

$$\lambda_1^k = \begin{cases} 0, & \text{GLRT}^k(Z) \geq 0.762 \\ 1 & \text{GLRT}^k(Z) < 0.762 \end{cases} \quad (7)$$

によって定義され,  $\lambda_2^k = 1 - \lambda_1^k$  である. また  $\text{GLRT}^k(Z)$  は  $k$  番目のフレームの位置あわせ誤差の GLRT 値である. この最適化問題の解は, 中間値と平均値を用いることで, 次のように与えられる.

$$\hat{Z}(i) = \mu_1 \operatorname{median}_k(F^{kT} \tilde{Y}^k(i)) + \mu_2 \operatorname{mean}_k(F^{kT} \tilde{Y}^k(i)) \quad (8)$$

ここで,  $\hat{Z}(i)$  は  $i$  番目の画素値をあらわしている.  $\tilde{Y}^k(i)$  についても同様である. このようにして低解像度

画像を統合することで, 当初の問題を, 劣化を受けた画像  $\hat{Z}$  から, 高解像度画像  $\hat{X}$  を推定する問題に変換することができる. ここでは, 正則化ぼけ除去アルゴリズムにより, この問題を解く. 解は次のように与えられる.

$$\hat{X} = \underset{X}{\operatorname{argmin}} \{ \mu_1 \|HX - Z\|_1 + \mu_2 \|HX - Z\|_2^2 + \|CX\|_2^2 \}, \quad (9)$$

ここで,  $C$  はラプラシアンであり, ハイパス作用素である. 式(9)は最急降下法で解くことができ, その繰り返しアルゴリズムは

$$X^{(n+1)} = X^{(n)} - \beta(\mu_1 H^T \operatorname{sign}(HX^{(n)} - \hat{Z}) + 2\mu_2 H^T (HX^{(n)} - \hat{Z})) + C^T CX^{(n)}$$

で与えられる.

#### 4 シミュレーション実験と結果

提案した動き検出法とそれを用いた超解像手法の有効性を確認するために, 人工的に低解像度画像を生成し, それらに提案手法を適用する. まず, 式(1)のモデルに基づき, 低解像度画像を生成した. ボケ作用素は  $7 \times 7$  のガウス作用素とし, 間引き率 2 でダウンサンプルした. さらに, 30dB のガウス性雑音を付加した. ここでは, 2 種類のテスト用ビデオシーケンスを用いた. ひとつは Table Tennis ( $240 \times 352$ , JPEG フォーマット) であり, もうひとつは Football ( $240 \times 352$ , QCI フォーマット) である.

動き検出の方法に関しては, 提案手法を, アフィン動きモデルを仮定した(画像全体が同一の動きをもつ)

場合、ブロックマッチング法、オプティカルフロー法と比較した。ブロックマッチング法では8×8のブロックを用い、オプティカルフロー法では文献[19]に示すものを利用した。

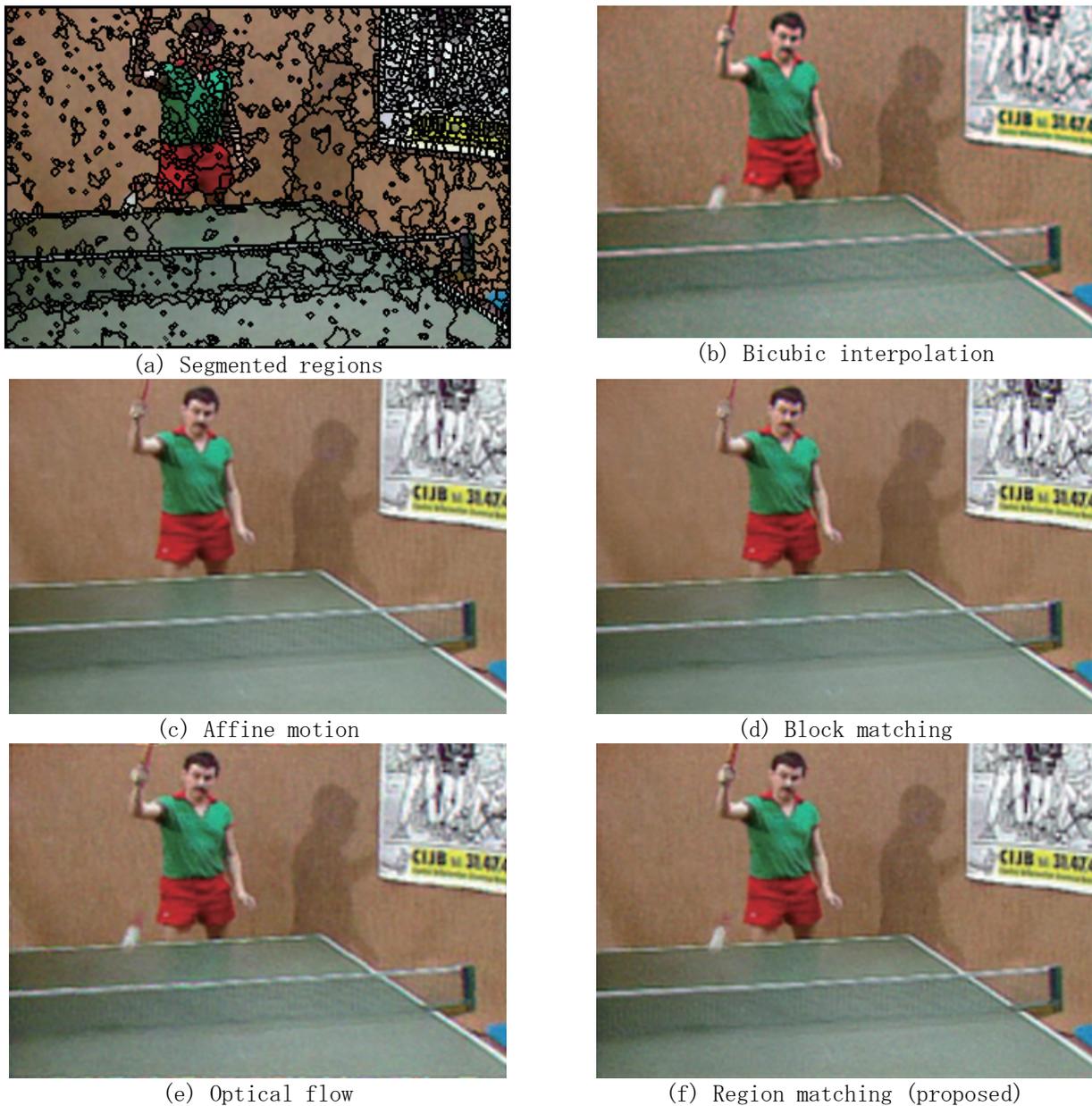
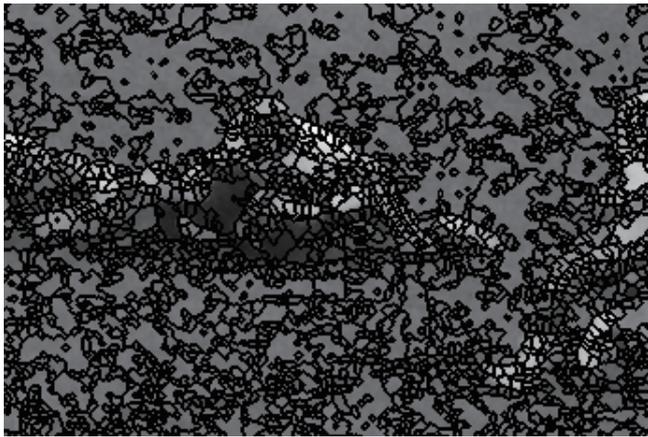


Fig. 1 Table Tennis

図1に、Table Tennisの画像列について、領域分割の結果とBicubic法、アフィン動きの場合、ブロックマッチング法、オプティカルフロー法、および提案手法による超解像処理の結果を示す。図1(c)からわかることは、画像全体の動きを仮定すると、局所的な動きに適応できないことである。図中のボールが消失することが、このことを顕著に表している。図1(d)から、ブロックマッチングは適切ではないことがわかる。画像中の物体は、一般に矩形のブロックを単位として動くわけではないので、この処理結果は当然の帰結であるといえる。オプティカルフロー法では、局所的な動きに対応できている部分もあるが、図1(e)に示すように、細かいノイズに敏感であるため、細部でエラーが起きている。図2に、Footballの結果を示す。Table Tennisの場合と同様の傾向を示していることがわかる。

次に、客観評価をするために、図1と2に示した画像の、原画像との類似度をPSNR (Peak signal-to-noise ratio) によって比較する。PSNRの測定結果を表1に示す。この表から、提案手法は客観評価においても、最も大きい値を示していることがわかる。



(a) Segmented regions



(b) Bicubic interpolation



(c) Affine motion



(d) Block matching



(e) Optical flow



(f) Region matching (proposed)

Fig. 2 Football

Table 1 PSNR measure for high-resolution images

Methods	Football sequence	Table Tennis sequence
Bicubic interpolation	27.1920	20.1066
Affine motion	24.9596	20.3655
Block matching	25.0817	20.5182
Optical flow	27.3101	19.9768
Region matching (proposed)	<b>28.0010</b>	<b>20.5343</b>

## 5 結論

本稿では、領域分割を用いた超解像の概念を提案し、領域分割と動きの推定法を提案した。また、この手法を動画像の超解像処理に適用をすることで、従来の動き推定法を用いる場合より、より高精度に高解像度画像を推定できることがわかった。

### 【参考文献】

- [1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21 - 36, May 2003.
- [2] Z. Jiang, T.T. Wong and H. Bao, "Practical super-resolution from dynamic video sequences," in *Proc. of IEEE Computer Vision and Pattern Recognition (CVPR 2003)*, Madison, Wisconsin, USA, June 2003.
- [3] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Dynamic demosaicing and color super-resolution of video sequences," *Proceedings of the SPIE conference on image reconstruction from incomplete data III*, Vol. 5562, October 2004.
- [4] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur," *IEEE Trans. Image Processing*, Vol. 10, no. 8, pp. 1187-1193, August 2001.
- [5] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Trans. on Image Processing*, vol. 13, no. 10, pp. 1327-1344, October 2004.
- [6] W. Zhao, H. Sawhney, "Is super-resolution with optical flow feasible?," *Proc. ECCV2002*, vol. 1, pp. 599-613, January 2002.
- [7] S. Farsiu, D. Robinson, M. Elad, P. Milanfar, "Robust shift-and-add approach to super-resolution," in *Proc. of the 2003 SPIE Conf. on Applications of Digital Signal and Image Processing*, San Diego, California, August 2003.
- [8] J. J. Clark, M. R. Palmer, and P. D. Lawrence, "A transformation method for the reconstruction of functions from non-uniformly spaced samples," *IEEE Trans. Acoust. Speech, Signal Processing*, vol.40, no.4, pp. 1151-1165, October 1985.
- [9] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 38, pp. 1013-1027, June 1990.
- [10] M. C. Hong, M. G. Kang, and A. Katsaggelos, "An iterative weighted regularized algorithm for improving the resolution of video sequences," *Proc. Int. Conf. Image Processing*, vol. 2, pp. 474-477, October 1997.
- [11] H. Stark and P. Oskoui, "High resolution image reconstruction from lower-resolution image sequences and space varying image restoration," *J. Opt. Soc. Am. A*, vol. 6, pp. 1715- 726, March 1989.
- [12] S. Baker and T. Kanade, "Super resolution optical flow," *Tech. Rep. CMU-RI-TR-99-36*, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, October 1999.
- [13] D. Barreto, L.D. Alvarez, and J. Abad, "Motion estimation techniques in super-resolution image reconstruction a performance evaluation," in *Virtual Observatory: Plate Content Digitalization, Archive Mining and Image Sequence Processing*, edited by Heron Press, vol. I, 254-268, Sofia (Bulgary), April 2005.
- [14] S. Farsiu, M. Elad, and P. Milanfar, "Video-to-video dynamic super-resolution for grayscale and color sequences," *EURASIP Journal of Applied Signal Processing, Special Issue on Superresolution Imaging*, Vol. 2006, Article ID 61859, Pages 1-15.
- [15] O. A. Omer and T. Tanaka, "Region-based sub-pixel motion estimation from noisy, blurred, and down-sampled sequences," *Lecture Notes in Computer Science*, ISSN 0302-9743, Vol. 4261/2006, pp. 229-236.

- [16] O. A. Omer and T. Tanaka, "Super-resolution based on region-matching motion estimation," in Proceedings of SPIE conference on Visual Communications and Image Processing (VCIP) 2007, vol. 6508, pp. 10-1-10-11, San Jose (CA), February 2007.
- [17] O. A. Omer and T. Tanaka, "Joint blur identification and high-resolution image estimation based on weighted mixed-norm with outlier rejection," in Proc. of 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008), pp. 1305-1308, Las Vegas, NV, USA, March 2008.
- [18] P. De Smet and D. De Vleschauwer, "Performance and scalability of highly optimized rainfalling watershed algorithm," Proc. Int. Conf. on Imaging Science, Systems and Technology, CISST98, pp. 266-273, Las Vegas, NV, USA, July 1998.
- [19] T. Gautama, and M.A. Van Hulle, "A phase-based approach to the estimation of the optical flow field using spatial filtering," IEEE Trans on Neural Networks, vol. 13, pp. 1127 - 1136, issue 5, Sept. 2002.
- [20] T. Koga, K. Linuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion compensated interframe coding for video conferencing," in Proc. Nat. Telecomm. Conf., New Orleans, LA, pp. G5.3.1-5.3.5., Nov. 29-Dec.3, 1981.
- [21] V. Bhaskaran and K. Konstantinides, "Image and video compression standards: algorithms and architectures," Kluwer Academic Publishers, 1997.

〈 発 表 資 料 〉

題 名	掲載誌・学会名等	発表年月
Extraction of High-Resolution Frame from Low-Resolution Video Using Region-Based Motion Estimation		(投稿中)