

複数フレーム型と学習型の組み合わせによる低解像度顔画像の超解像

○ 吉田智成†, 高橋友和‡, 出口大輔*, 井手一郎†, 村瀬洋†

○ Tomonari YOSHIDA†, Tomokazu TAKAHASHI‡, Daisuke DEGUCHI*,

Ichiro IDE† and Hiroshi MURASE†

†: 名古屋大学 大学院情報科学研究科, yoshidat@murase.m.is.nagoya-u.ac.jp,
{ide,murase}@is.nagoya-u.ac.jp

‡: 岐阜聖徳学園大学 経済情報学部, ttakahashi@gifu.shotoku.ac.jp

*: 名古屋大学 情報連携統括本部, ddeguchi@nagoya-u.jp

監視カメラ映像中の顔画像認識において、顔領域が低解像度であることは認識精度の低下を招く。この問題に対して我々は、顔向きの変化を利用した複数フレーム型超解像手法を提案してきた。しかし、複数フレーム型超解像には、画像が低解像度になるにつれ、高解像度化に必要な高周波成分を復元することが困難になるという問題がある。そこで本発表では、複数フレーム型超解像と学習型超解像を組み合わせた手法を提案する。これにより、複数フレーム型超解像では復元しきれなかった高周波成分の復元を図る。学習型超解像は、大量の顔画像から高解像度化に必要な顔の知識を学習し、それを用いて低解像度画像から高解像度画像を生成する技術である。実際に撮影した動画像を用いた実験の結果から提案手法の有効性を確認した。

<キーワード>複数フレーム型超解像, 学習型超解像, 非剛体変形, 顔画像認識

1. はじめに

近年、防犯対策を目的とした監視カメラ映像中の顔画像認識に対する需要が高まっている。監視カメラ映像中の顔領域は低解像度であることが多く、これは顔認識精度の低下を招く。この問題を解決するために、映像中の複数枚の低解像度画像から1枚の高解像度画像を生成する超解像技術[1-5]の利用が考えられる(図1)。この際、入力画像間で顔向きが変化する場合には、ほとんど変化しない場合に比べて、高解像度化に必要な情報をより多く取得できると考えられる。そこで我々は、顔向き変化を利用した複数フレーム型超解像手法を提案してきた[1]。複数フレーム型超解像手法[1-5]は複数の入力フレームから高解像度画像の各画素値を補完する手法であるが、解像度が低くなるにつれ、高解像度化に必要な高周波成分を復元することが困難になるという問題がある。

これに対して本発表では、複数フレーム型と学習型を組み合わせた顔画像に対する超解像手法を提案する。学習型超解像[6-8]は、大量の顔画像から高解



図1 超解像技術を利用した顔画像認識

像度化に必要な顔の知識を学習し、それを用いて低解像度画像から高解像度画像を生成する技術であり、入力画像の解像度にかかわらず高周波成分を復元できるという特徴を持つ。これにより、複数フレーム型超解像では復元しきれなかった高周波成分の復元を図る。

以降、2節で複数フレーム型と学習型の組み合わせによる低解像度顔画像の超解像手法について述べる。そして、3節で実験の方法を述べ、実験結果を考察する。最後に、4節でまとめと今後の課題を述べる。

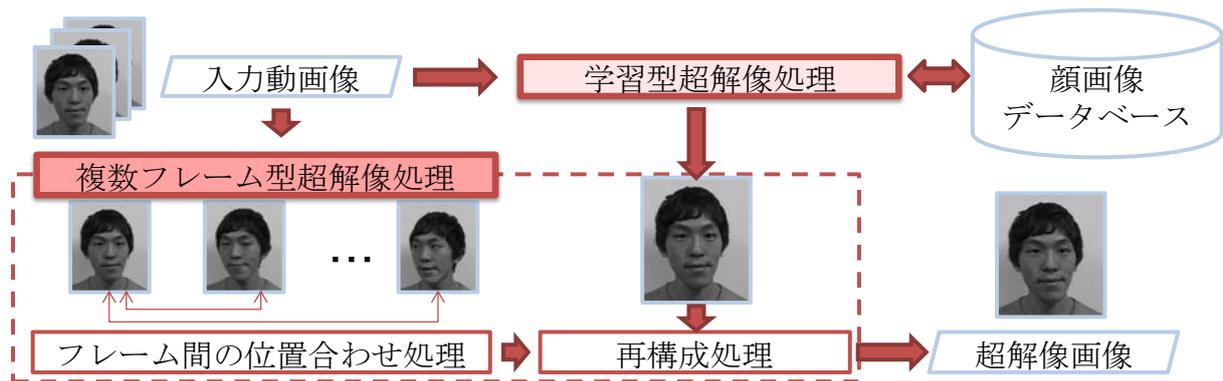


図2 提案手法における処理の流れ

2. 提案手法

2.1 概要

提案手法における処理の流れを図2に示す。提案手法は複数フレーム型超解像処理を元にして学習型超解像処理を組み合わせた手法である。まず入力に対して学習型超解像処理を行い高解像度画像を生成する。続いて、複数フレーム型超解像処理において、フレーム間の位置合わせ処理を行う。最後に、学習型超解像処理と位置合わせ処理の結果を用いた再構成処理を行い、超解像画像を生成する。以降、各処理について詳述する。

2.2 学習型超解像処理

学習型超解像処理は大量の顔画像から高解像度化に必要な顔の知識を学習し、それを用いて低解像度画像から高解像度画像を生成する技術である[6]。具体的には、まず事前処理として、大量の低解像度画像パッチと高解像度画像パッチの組からデータベースを作成する。入力画像をパッチに分割し、各パッチをそれと類似したデータベース中の複数の低解像度画像パッチの線形和で表現する。そして対応する高解像度画像パッチでそれらを置き換え、得られたパッチを統合することにより最終的な超解像画像を得る。これにより、入力画像の解像度にかかわらず高周波成分を復元することができる。学習型超解像処理の結果例を図3に示す。眼や口などの高周波成分が復元できていることが分かる。学習型超解像処理で得られた超解像画像を複数フレーム型超解像処理の中の再構成処理で利用することにより、最終的な超解像画像における高周波成分の復元を実現する。

2.3 複数フレーム型超解像処理

複数フレーム型超解像処理とは、複数の低解像度



(a) 入力画像 (b) 超解像画像 (c) 正解画像

図3 学習型超解像処理の結果例

画像をサブピクセル精度で位置合わせすることにより画素値を補完した後、ぼけを除去して高周波成分を復元することで高解像度画像を再構成する技術のことである。以降、フレーム間の位置合わせ処理と再構成処理について述べる。本手法では、再構成処理においてぼけを除去して高周波成分を復元するために学習型超解像処理の結果を用いる。

2.3.1. 位置合わせ処理

複数フレーム型超解像処理において、画素値を補完するためにフレーム間の位置合わせを高精度に行うことは重要である。特に顔画像は顔向きや表情の変化により2次元的に非剛体な変形をするため、単純な幾何変形では位置合わせが困難である。そのため本手法では非剛体変形を用いて基準フレームとその他のフレームの位置合わせをサブピクセル精度で行う[1]。非剛体変形は、画像上に配置した制御点を移動させることで画像の柔軟な変形を行う手法[9]であり、これにより顔向きや表情の変化に頑健な位置合わせを実現する。

2.3.2. 再構成処理

位置合わせ処理と学習型超解像処理の結果を用いて低解像度画像群から高解像度画像を推定する。再構成処理にはMAP (Maximum A Posteriori) 法[2]を

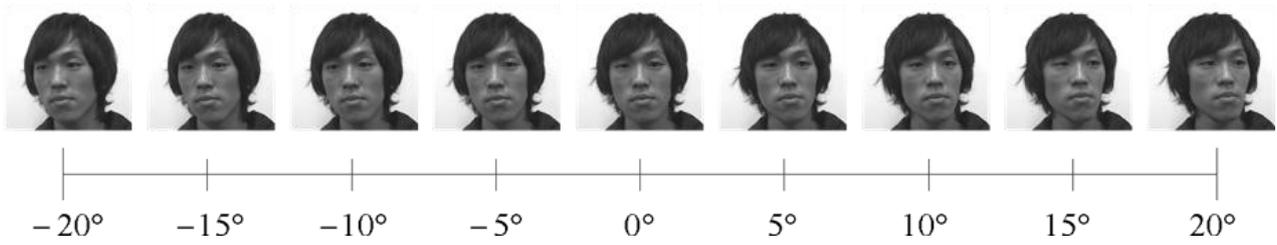


図4 撮影した動画像中のフレームの例

用いる。MAP法とは、入力画像を条件としたときの事後確率を最大にする高解像度画像を求める手法である。実際の処理では、事後確率から導かれる評価関数を最小化することにより高解像度画像を求める。本手法で用いる評価関数は以下の式で表される。

$$E = \sum_i \|SHW_i X - Y_i\|^2 + \alpha \|X - X_{\text{LSR}}\|^2 \quad (1)$$

ここで、 Y_i は*i*番目のフレームの低解像度画像であり、 X は求めたい高解像度画像である。 W_i は移動行列であり、フレーム間の位置合わせ処理により求める。 H は光学ぼけを表現する行列であり、ここではぼけを表すPSF (Point Spread Function) を位置によらず1つのGauss分布と仮定する。 S はダウンサンプリングを行う行列である。 X_{LSR} は位置合わせ処理の基準フレームに対する学習型超解像処理の結果画像であり、 α は第2項による影響を制御するパラメータである。本手法では式(1)の最小化に共役勾配法を用いる。 E が最小となるまで繰り返し高解像度画像 X を更新し、最終的な超解像画像を得る。

式(1)の第1項は一般的に誤差項と呼ばれるもので、高解像度画像の推定誤差を表す。また、第2項は高解像度画像の事前確率を基にした拘束項である。誤差項のみで高解像度画像を推定する場合は解が不安定になるが、誤差項と拘束項を両方用いることで推定誤差を小さくすると同時に、安定した解を求めることができる。さらに、本手法は拘束項で学習型超解像画像を顔の事前知識として利用することにより、顔画像中の高周波成分の復元を図る。

3. 実験

提案手法の有効性を調査するために、顔の向きや表情の変化を含む実際の映像を用いた評価実験を行った。実験では、低解像度の動画像から高解像度画像を生成し、生成した顔画像の認識率を算出した。

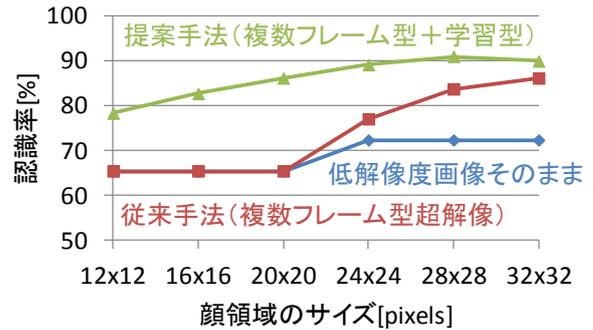


図5 解像度毎の認識率

3.1 実験方法

実験で利用する動画像は、三脚にデジタルカメラを固定し、手で焦点を合わせながら撮影した。顔向きを水平角度 $[-20^\circ, 20^\circ]$ の範囲で変化させながら29名分 \times 30本の動画像を得た。各々約30フレーム、顔領域の一边の解像度は約64pixelsであった。撮影した動画像中のフレームの例を図4に示す。

入力動画像として、撮影した高解像度の動画像の解像度をいくつかの異なる解像度に変換したものを使用した。具体的には、顔領域の一边の解像度がそれぞれ約12, 16, 20, 24, 28, 32 pixelsとなるようにダウンサンプリングし、平均0分散1のGauss雑音を加えたものを使用した。実験に用いた低解像度の動画像のフレームの例を図6, 7(a)に示す。

比較のために、以下に示す3つの手法により顔認識用の超解像画像をそれぞれ生成した。なお、超解像の倍率は入力動画像の解像度にかかわらず顔領域の一边が約64 pixelsとなるように設定した。

- 比較手法1: 低解像度画像そのまま
- 比較手法2: 複数フレーム型超解像[1]
- 提案手法: 複数フレーム型超解像+学習型超解像



(a) 低解像度画像（入力） (b) 比較手法による超解像画像 (c) 提案手法による超解像画像 (d) 元の高解像度画像

図6 人物毎の超解像画像例（顔領域の一边の解像度：20 pixels→64 pixels）

認識手法には、固有空間法[10]を利用した。ここで、固有空間の学習データと認識用の顔データベースとして、動画を撮影したものと異なるデジタルスチルカメラで撮影した正解データ 29 枚とその他の顔データ 300 枚の顔画像を用いた。なお、学習型超解像用の学習データには顔認識用の顔データベース中の人物は含まれていない。

3.2 実験結果

解像度毎の認識率を図5に示す。図5から、全ての解像度に対して、提案手法の認識率が比較手法よりも高いことが分かる。提案手法と複数フレーム型超解像の認識率の差が最も大きかった解像度は 20 pixels であり、そのときの認識率は複数フレーム型超解像が 65.5%であったのに対し、提案手法は 86.2%



(a) 低解像度画像（入力） (b) 比較手法による超解像画像 (c) 提案手法による超解像画像 (d) 元の高解像度画像

図 7 解像度毎の超解像画像例（上段：12x12 pixels，中段：24x24 pixels，下段：32x32 pixels）

であった。また、実験で得られた人物毎の超解像画像の例を図 6 に、解像度毎の超解像画像の例を図 7 に示す。これらから、眼、鼻、口などの高周波成分が正しく復元できていることが分かる。

3.3 考察

図 5 の解像度毎の認識率を見ると、入力画像の解像度が 20 pixels 以下のときに、複数フレーム型超解像ではあまり大きな向上が見られないのに対して、提案手法では 12% 以上認識率が向上していることが分かる。これは入力が低解像度になるにつれ、複数フレーム型超解像において、入力動画中の各フレームから高精度な認識に必要な情報を十分に得ることが困難となったためであると考えられる。これに対して提案手法では、複数フレーム型超解像と学習型超解像を組み合わせることにより、高精度な認識に必要な情報が十分に得られたと考えられる。

また、図 7 の解像度毎の超解像画像の例を見ると、

解像度が低くなるほど、正解画像とは異なる画像になっていることが分かる。これは、解像度が低くなるほど入力動画中に含まれる情報が少なくなるため、学習型超解像処理において誤った高周波成分が復元されたことが原因だと考えられる。この問題の解決法の 1 つとして、複数枚の低解像度画像を入力として学習型超解像処理を行うことが考えられる。学習型超解像処理による高解像度画像の推定を複数枚から行うことで、入力が非常に低解像度な場合でも正しい高周波成分の復元が実現できると考えられる。

4. むすび

本発表では、複数フレーム型と学習型を組み合わせた顔画像に対する超解像手法を提案した。従来の複数フレーム型超解像処理に加えて、学習型超解像処理を利用することにより、複数フレーム型超解像

処理では復元しきれなかった高周波成分の復元を実現した。

実動画像を用いた実験の結果から、複数フレーム型超解像手法[1]で生成した画像と比較して、認識率が最大で20.7%向上することを確認した。このことから、提案手法の有効性が確認できた。今後の課題として、学習型超解像処理の改良や処理の高速化などが挙げられる。

謝辞

日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に深く感謝する。本研究で使用する顔画像データを提供頂いたオムロン株式会社の関係者に感謝する。本研究の一部は、文部省科学技術戦略推進費、科学技術研究費補助金による。また、本研究では画像処理にMISTライブラリ(<http://mist.murase.mis.nagoya-u.ac.jp/>)を使用した。

参考文献

- [1] T. Yoshida, T. Takahashi, D. Deguchi, I. Ide and H. Murase, "Robust face super-resolution using free-form deformations for low-quality surveillance video," Proc. 2012 IEEE Int. Conf. on Multimedia and Expo, pp.368-373, July 2012.
- [2] R. R. Schulz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," IEEE Trans. on Image Processing, vol.5, issue 6, pp.996-1011, 1996.
- [3] F. W. Wheeler, X. Liu and P. H. Tu, "Multi-frame super-resolution for face recognition," Proc. of First IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems, pp.1-6, Sep. 2007.
- [4] P. Mortazavian, J. Kittler and W. Christmas, "3D-assisted facial texture super-resolution," 2009 British Machine Vision Conference, Sep. 2009.
- [5] 吉村真, 清水雅夫, 田中正行, 奥富正敏, "非剛体変形を利用した揺らぎ除去と高解像度画像生成," 画像の認識・理解シンポジウム (MIRU) 2007 論文集, pp.235-240, July 2007.
- [6] H. Huang, H. He, X. Fan and J. Zhang, "Super-resolution of human face image using canonical correlation analysis," Pattern Recognition, vol.43, issue 7, pp.2532-2543, 2010.
- [7] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H.

Hayes III and R. M. Mersereau, "Eigenface-based super-resolution for face recognition," Proc. 2002 IEEE Int. Conf. on Image Processing, vol.2, pp.845-848, Sep. 2002.

[8] A. Chakrabarti, A. N. Rajagopalan and R. Chellappa, "Super-resolution of face images using kernel PCA-based prior," IEEE Trans. on Multimedia, vol.9, issue 4, pp.888-892, 2007.

[9] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," IEEE Trans. Medical Imaging, vol.18, no.8, pp.712-721, 1999.

[10] M. Turk and A. Pentland, "Eigenfaces for recognition," Cognitive Neuroscience, vol.3, issue 1, pp.71-86, 1991.