

顔の非剛体変形に頑健な複数フレーム超解像の検討

吉田 智成[†] 高橋 友和^{††} 出口 大輔[†] 井手 一郎[†] 村瀬 洋[†]

[†] 名古屋大学 大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町
^{††} 岐阜聖徳学園大学 経済情報学部 〒500-8288 岐阜県岐阜市中鶉1丁目38番地
 E-mail: †yoshidat@murase.m.is.nagoya-u.ac.jp,
 †{ddeguchi,ide,murase}@is.nagoya-u.ac.jp, ††ttakahashi@gifu.shotoku.ac.jp

あらまし 監視カメラ映像から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、動画像を用いた超解像技術を利用することで認識精度が向上できると考えられる。しかし、従来の超解像は、平面物体を撮影した動画像を対象としている場合が多く、顔画像のように向きや表情の変化によって2次元的に非剛体変形する動画像を扱うことが困難である。そこで本報告では、局所領域毎に位置合わせをすることで、顔向きや表情の変化に頑健な超解像を行う手法を提案する。具体的には、フレーム間の位置合わせに非剛体レジストレーションを用いることで顔画像の非剛体変形に柔軟に対応する。実際に撮影した動画像を用いた実験の結果、超解像の性能向上が見られ、提案手法の有効性が確認できた。

キーワード 複数フレーム超解像, 非剛体変形, 顔画像認識

A study of robust multi-frame super resolution for non-rigid deformation of face

Tomonari YOSHIDA[†], Tomokazu TAKAHASHI^{††}, Daisuke DEGUCHI[†], Ichiro IDE[†], and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University
 Furo-cho, Chigusa-ku, Nagoya, Aichi, 464-8601 Japan
^{††} Faculty of Economics and Information, Gifu Shotoku Gakuen University
 Nakauzura 1-38, Gifu, Gifu, 500-8288 Japan
 E-mail: †yoshidat@murase.m.is.nagoya-u.ac.jp,
 †{ddeguchi,ide,murase}@is.nagoya-u.ac.jp, ††ttakahashi@gifu.shotoku.ac.jp

Abstract In a face recognition system with surveillance video cameras, the decrease in resolution of face images degrades the recognition accuracy. To overcome this problem, multi-frame super-resolution techniques could be used to improve the accuracy. However, most super-resolution techniques assume that a planar object is captured in input images. Therefore, it is difficult to apply them to face images that include non-rigid deformations caused by changes of face poses and expressions. In this report, we propose a multi-frame super-resolution method that can deal with changes of face poses and expressions. To achieve this, alignment of each local region between video frames is performed by using a free-form deformation method. Thus, the proposed method can easily deal with the non-rigid deformation of face images. Experimental results demonstrate that the proposed method improved the performance of super resolution for actual videos. From this, we confirmed the effectiveness of the proposed method.

Key words multi-frame super-resolution, free-form deformations, face recognition

1. まえがき

近年、防犯対策として、銀行や地下鉄、街頭などの多くの場所に監視カメラが設置されており、監視カメラ映像から人物を特定する技術に対する需要が高まっている。監視カメラ映像からの高精度な人物特定技術により、例えば街中の監視カメラ映像を用いた指名手配犯の搜索、犯罪現場の監視カメラ映像を用いた犯人の特定が可能となると考えられる。実際に、ロンドンで起きた爆破テロ事件の犯人検挙にも、監視カメラ映像が大きく貢献したとされている [1]。

監視カメラ映像から人物を特定するには、顔画像認識が有効であると考えられる。これは映像から得られる情報の中で、顔が個人の特徴を最もよく表していると考えられるからである。しかし、監視カメラの設置位置や台数には制限があり、1台のカメラは広範囲を撮影するため、図1に示すように、実際に監視カメラから得られる映像中の人物の顔画像領域は低解像度であることが多い。一般に、顔領域が低解像度になるほど顔画像の認識性能は低下する。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 [2] を利用することが考えられる。超解像技術とは、複数枚の低解像度画像をサブピクセル精度で位置合わせすることで画素値を補完した後、ぼけやノイズを除去することで高解像度画像を生成する技術である。しかしながら、これまで提案されてきた超解像技術の多くは、平面物体を撮影した画像を対象としている場合が多く、顔画像のように向きや表情の変化によって2次元的に非剛体変形する画像を扱うことが困難である。

2次元的な非剛体変形を考慮した超解像手法としては、特徴点ベースで位置合わせを行うもの [3, 4] と、領域ベースで位置合わせを行うもの [5] がある。Frederickらは顔画像間の位置合わせにAAM (Active Appearance Model) を利用することで、高解像度な顔画像を生成する手法を提案している [3]。また、Pouriaらは顔の3次元モデルを用いた超解像処理を行うことで、後の認識処理に有効な画像を生成する手法を提案している [4]。しかしながら、これらの手法では画像が低解像度になるほど、顔の特徴点を正確に抽出することが難しくなり、超解像処理に必要な画像間の位置合わせ精度が低下すると考えられる。一方で、田中らは領域ベースの位置合わせを用いた超解像手法を提案している [5]。この手法では、B-Spline関数を利用した非剛体変形を用いて、空間的に滑らかで不規則な変形(揺らぎ)を高精度に推定し、超解像を行う。また、遠方を撮影し

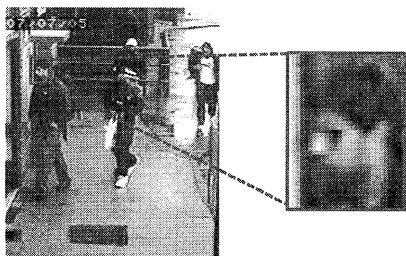


図1 監視カメラから得られる画像の例

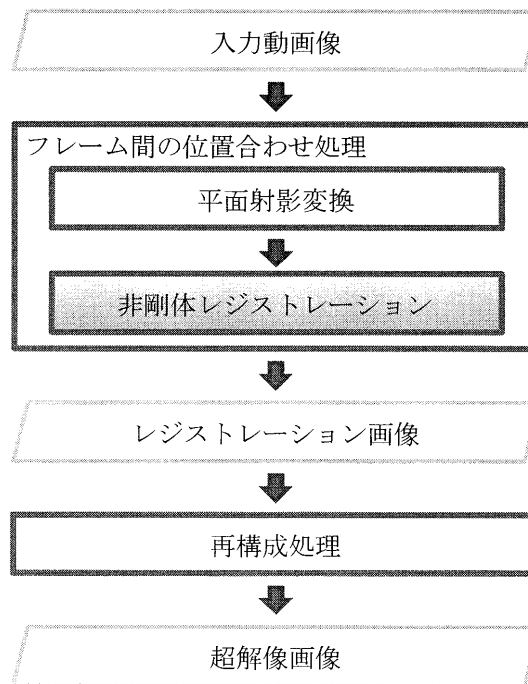


図2 提案手法による超解像処理の流れ

た時系列画像や水面を通して撮影した時系列画像を対象とした実験において、大気や水面の揺らぎを高精度に除去することに成功している [5]。本研究で対象とする顔画像においても、向きや表情の変化により非剛体な変化が生じる。よって本稿では、このような変化を非剛体変形モデルで表現することにより、顔画像に対する超解像処理を高精度に行う手法を提案する。

以降、2節では提案手法について述べる。3節では提案手法の有効性を確認するための評価実験について述べ、4節で考察を加える。最後に5節でまとめる。

2. 提案手法：非剛体変形を用いた超解像

提案手法における処理の流れを図2に示す。提案手法は従来の超解像処理の枠組みを拡張したものである。以降、従来の超解像処理について述べ、その後、従来の超解像処理と提案手法の差異を述べる。超解像処理は大きく分けてフレーム間の位置合わせ処理と再構成処理の2つの処理から構成される。まず、フレーム間の位置合わせ処理では、入力動画像中のあるフレームを基準として、基準フレームと各フレームとの位置合わせをサブピクセル精度で行う。その後、全てのフレームの位置合わせの結果として得られる画像の平均画像(以降、レジストレーション画像と呼ぶ)を作成する。レジストレーション画像は、複数の低解像度画像から高解像度画像の各画素値を補完した画像である。続いて、再構成処理では、レジストレーション画像からぼけやノイズを除去しながら最終的な高解像度画像の推定を行う。以上が従来の超解像処理である。

以下、従来の超解像処理と提案手法の差異について述べる。従来の位置合わせ処理では、平面物体を撮影した画像を対象としているため、平面射影変換を用いて位置合わせを行う。これに対して、提案手法では、まず前処理として平面射影変換

による位置合わせを行い、その後、非剛体レジストレーション (FFD; Free-Form Deformations) により局所領域毎に詳細な位置合わせを行う。非剛体レジストレーションの前処理として平面射影変換により位置合わせを行う理由は、画像間に2次元的な回転や平行移動がある場合に非剛体レジストレーションのみでは対応が困難となるためである。以降、各処理について詳述する。

2.1 平面射影変換の推定

まず、前処理としての位置合わせを行うために基準フレームとそれ以外の各フレームの間の平面射影変換行列を推定する。具体的には、位置合わせの基準となる画像と、位置合わせを行いたい画像 (参照画像) を入力とし、参照画像を平面射影変換した画像と基準画像との差が小さくなるように、射影変換のパラメータを反復更新する。本手法では、平面射影変換行列を求める手法として ICIA (Inverse Compositional Image Alignment) [6] を用いる。

2.2 顔の非剛体変形に頑健な位置合わせ

顔向きや表情の変化により画像内に非剛体な変化が存在する場合には、平面射影変換のみによる詳細な位置合わせは困難となる。この位置合わせ精度の低下は、後の処理に悪影響を及ぼす。そこで、非剛体レジストレーションを用いた詳細な位置合わせを行う。非剛体レジストレーションは、医療画像などの3次元画像間の位置合わせに利用されており [7]、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い、画像間の位置合わせを行う手法である。

図3に非剛体レジストレーションの処理の流れを示す。非剛体レジストレーションは以下の手順で行われる。

Step 1 入力画像 I と基準画像 T を用意し、入力画像上に格子状に制御点を $L \times M$ 個配置する。 l 行 m 列目の制御点の座標を $p_{l,m}$ ($l = 1, \dots, L, m = 1, \dots, M$) とする。

Step 2 各 $p_{l,m}$ に関して、以下の式を計算して制御点の位置を更新する。 $D_{l,m}$ は制御点 $p_{l,m}$ の近傍にある画素の位置の集合である。

$$p_{l,m} = p_{l,m} + \mu \frac{\nabla c}{\|\nabla c\|} \quad (1)$$

$$\nabla c = \frac{\partial}{\partial p_{l,m}} \sum_{D_{l,m}} |I'(x,y) - T(x,y)| \quad (2)$$

ここで、 $I'(x,y)$ 、 $T(x,y)$ はそれぞれ I' 、 T の (x,y) における画素値である。また、 I' は制御点を移動して入力画像 I を変形した後の画像であり、以下の式により求める。

$$I'(x,y) = I(w(x,y)) \quad (3)$$

$$w(x,y) = \sum_{i=0}^3 \sum_{j=0}^3 B_i(u') B_j(v') p_{u+i,v+j} \quad (4)$$

$$B_0(t) = (1-t)^3/6 \quad (5)$$

$$B_1(t) = (3t^3 - 6t^2 + 4)/6 \quad (6)$$

$$B_2(t) = (-3t^3 + 3t^2 + 3t + 1)/6 \quad (7)$$

$$B_3(t) = t^3/6 \quad (8)$$

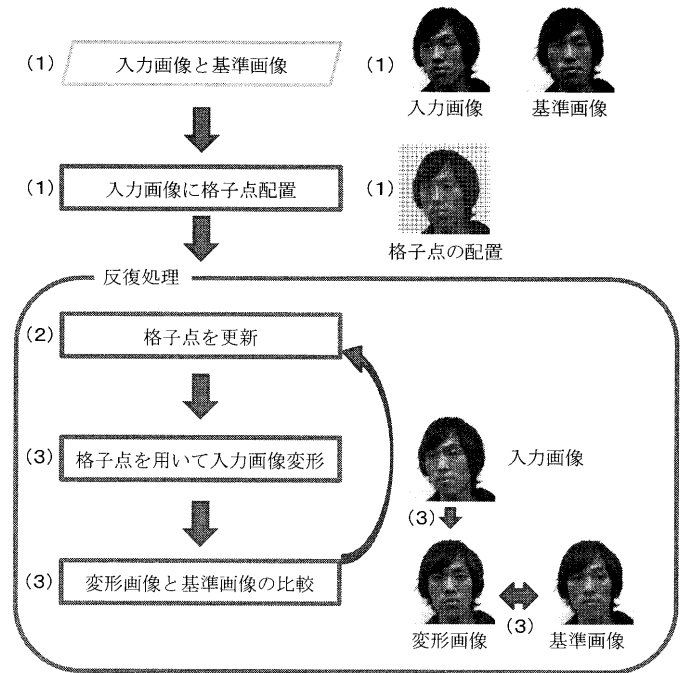


図3 非剛体レジストレーションの流れ

ここで、 $u = \lfloor x/L \rfloor - 1, v = \lfloor y/M \rfloor - 1, u' = x/L - \lfloor x/L \rfloor, v' = y/M - \lfloor y/M \rfloor$ である。

Step 3 変形画像 I' と基準画像 T の距離 d を計算する。

$$d = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y |I'(x,y) - T(x,y)| \quad (9)$$

Step 4 d がしきい値以下になるか、反復回数の上限に達するまで Step 2~Step 3 を繰り返す。

以上の処理を、あるフレームを基準画像として、それ以外の全てのフレームとの間で行う。位置合わせ後の全てのフレームの平均画像をレジストレーション画像とする。

2.3 再構成処理

レジストレーション画像から光学ぼけを除去する。再構成処理には MAP (Maximum A Posteriori) 法 [2] を用いる。MAP 法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り、事後確率から導かれる評価関数を最小化することで高解像度画像を求める。

MAP 法の評価関数は以下の式で表される。

$$J = \sum_{x=1}^{X'} \sum_{y=1}^{Y'} [b(x,y)^T H - R(x,y)]^2 + \lambda D(H) \quad (10)$$

ここで、 H はレジストレーション画像 R に対応する高解像度画像 ($X' \times Y'$ pixels) であり、 $b(x,y)$ は位置 (x,y) における光学ぼけを表す PSF (Point Spread Function) カーネルである。ここでは、PSF を位置に依らず 1 つのガウス分布で仮定する。また、 λ は第 2 項による影響を制御するパラメータである。式 (10) の第 1 項は一般的に誤差項と呼ばれるものであり、高解像度画像の推定誤差を表す項である。また、第 2 項は事前確率を基にした拘束項であり、 $D(H)$ は、BTV (Bilateral Total Variation) 法 [8] の正則化項である。この項はエッジの滑らか

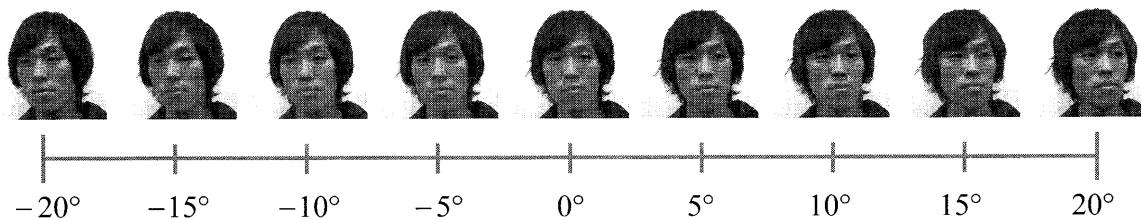


図4 撮影した動画のフレームの例

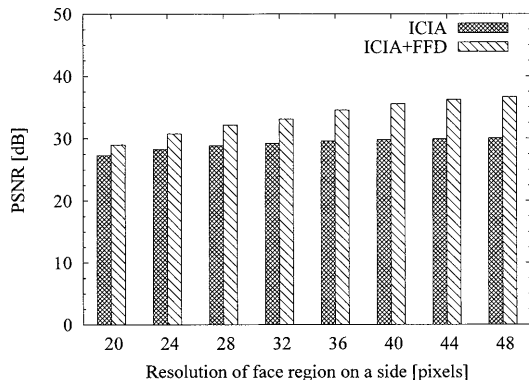


図5 実験結果：解像度毎のPSNR

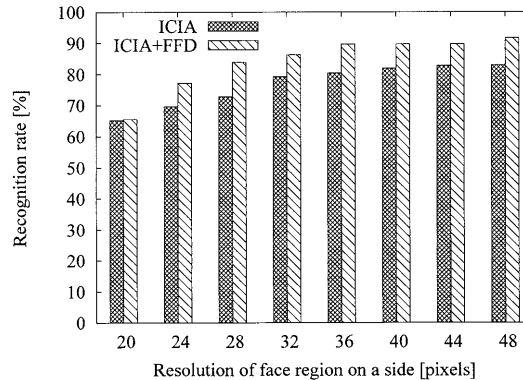


図6 実験結果：解像度毎の認識率

さの拘束を表している。 J を最小化することで光学ぼけを除去する。

この最小化には、共役勾配法 [9] を用いる。この場合、 J の勾配は以下の式で計算される。

$$\frac{\partial J}{\partial \mathbf{H}} = 2 \sum_{x=1}^{X'} \sum_{y=1}^{Y'} \mathbf{b}(x, y) \left[\mathbf{b}(x, y)^T \mathbf{H} - \mathbf{R}(x, y) \right] + \lambda \frac{\partial D(\mathbf{H})}{\partial \mathbf{H}} \quad (11)$$

J が最小となるまで繰り返し高解像度画像 \mathbf{H} を更新する。以上の処理により、超解像画像 \mathbf{H} を得る。

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による2次元的な非剛体変形を含む映像を用いた評価実験を行った。実験では、低解像度の動画から高解像度画像を生成し、生成した画像のPSNR (Peak Signal-to-Noise Ratio) と認識率を算出した。

3.1 実験方法

実験で利用する動画は、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本の動画を得た。各々約30フレーム、顔領域の一辺の解像度は約64 pixelsであった。撮影した動画中のフレームの例を図4に示す。

入力動画として、撮影した高解像度の動画をいくつかの異なる解像度に加工したものを使用した。具体的には、顔領域の一辺の解像度が約20, 24, 28, 32, 36, 40, 44, 48 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。高解像度の動画から生成した低解像度の動画のフレームの例を図8に示す。

以下に示す2つの位置合わせ手法により、各入力動画中の顔領域の一辺の解像度が約64 pixelsとなるように超解像を行い、高解像度画像を生成した。

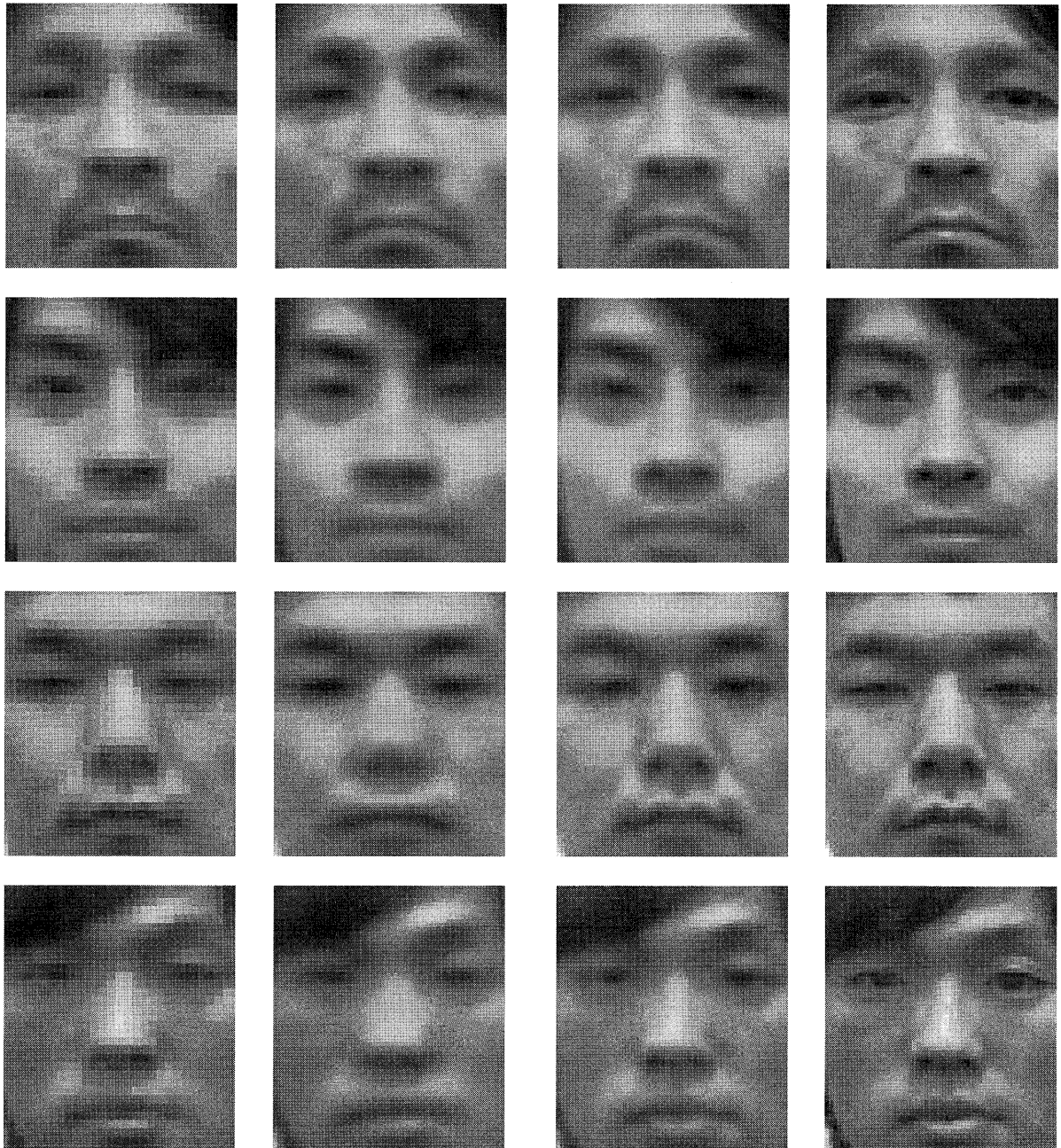
- 比較手法：ICIAのみ (従来の超解像)
- 提案手法：ICIA+FFD

なお、本実験では超解像処理の基準フレームは顔向きが正面のフレームを手で選択して使用した。生成した画像の評価基準として、元の画像からの劣化度合を表す画質指標として広く用いられているPSNRと、顔画像認識による認識率を用いた。PSNRの真値には、低解像度に加工する前の高解像度画像を利用し、認識手法には、顔画像認識において広く用いられている固有空間法 [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には、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枚と既存の顔データベース^(注1)中の300枚を用いた。

3.2 実験結果

解像度毎のPSNRを図5に、認識率を図6にそれぞれ示す。また、実験で得られた超解像画像の例を図7に示す。図5から、全ての解像度に対して、提案手法のPSNRが比較手法よりも高いことが分かる。提案手法と比較手法のPSNRの差が最も大きかった解像度は44 pixelsであり、そのときのPSNRは比較手法が30.01dBであったのに対し、提案手法は36.64dBであった。また、図6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。これらのことから、提案手法の有効性が確認できた。

(注1)：(財) ソフトピアジャパン (<http://www.softopia.or.jp/>)



(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 超解像処理結果例 (顔領域の一边の解像度: 32pixels → 64pixels)

4. 考 察

4.1 提案手法による PSNR と認識率の向上

比較手法に対して提案手法では, PSNR が平均 4.37dB, 認識率が 7.3% 向上し, 比較手法よりも提案手法の方が PSNR と認識率ともに高い結果となった. これは, 提案手法が非剛体変形を用いた結果, 顔向きや表情の変化に頑健な画像間の位置合わせが実現できたことが要因だと考えられる. 実際に, 図 7 (b) および (c) から, 比較手法よりも提案手法の結果の方がぼけが少なく, 高周波成分をより再現できていることがわかる.

4.2 解像度毎の認識率の差異

図 6 から, 顔領域の一边の解像度が 20 pixels の場合には,

比較手法に対する提案手法の認識率の大きな向上は見られなかった. このときの入力の低解像度画像を図 8 に示す. 認識率の向上があまり見られなかった原因として, 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が十分に得られなかったことが考えられる. 解像度が低下すると, 顔の向きや表情の変化が発生していても観測される各画素値の変化は小さくなる. そのため, サブピクセル精度の位置合わせを行ったとしても高解像度画像の画素値を十分に補完することができなかったと考えられる. この問題の解決法の一つとして, 入力以外の顔画像の情報を利用することが考えられる. これに関して, 大量の顔画像からなるデータベースから顔画像が共通して持つ特徴を学習し, それを再構成処理に利用する学習型超解像手法が提案さ

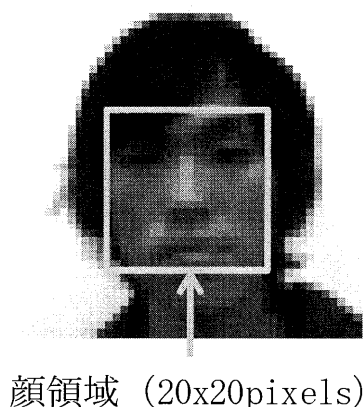


図8 顔領域の一辺の解像度が20 pixelsの低解像度画像の例

れている [11] [12] [13]. これらの手法のように, 入力動画像以外の顔画像から得られる情報を利用することで超解像の性能が向上できると考えられる.

4.3 位置合わせ処理における問題

位置合わせ処理を行う際, 基準画像と参照画像の間で瞬きなどの急激な表情変化が発生した場合, 正確な位置合わせが行われない場合があった. これは, 急激な表情変化によって, 一方の画像では存在しない画素値がもう一方の画像では存在するような状況が発生したことが原因であると考えられる. このような状況では, 非剛体レジストレーションにおける制御点の変位が正しく計算されないため, 結果として位置合わせ精度が低下することが想定される. この問題の解決法の一つとして, 顔特有の知識を利用することが考えられる. 顔のパーツの形状や動きなどの顔特有の情報に基づいて非剛体レジストレーションに制約を加えることで, 急激な表情変化にも対応可能な位置合わせ処理が実現できると考えられる.

5. むすび

本報告では, 顔向きや表情の変化に頑健な複数フレーム超解像手法を提案した. 提案手法は, フレーム間の位置合わせに非剛体レジストレーションを用いることで顔画像の非剛体変形に柔軟に対応するものである.

実験では, 顔画像の非剛体変形を含む低解像度の動画像を用いて提案手法の有効性を調査した. 平面物体を対象とした従来の超解像処理で生成した画像の PSNR が 30.01dB であったのに対し, 提案手法では 36.64dB であった. また, 従来の超解像処理で生成した画像の認識率が 82.9% であったのに対し, 提案手法では 91.6% であった. これらのことから, 提案手法の有効性が確認できた.

今後の課題として, 超解像の更なる性能向上を目的とした, 学習型超解像手法の利用や, パーツの形状や動きなど顔特有の情報の利用等が挙げられる.

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. 本研究の一部は, 文部科学省科学技術振興調整費, 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.m.is.nagoya-u.](http://mist.murase.m.is.nagoya-u.ac.jp/)) を使用した.

文 献

- [1] 橋本隆男: “ロンドンでのテロ対策”, Japan Railway Engineers' Association, **49**, 2, pp. 31459–31461 (2006).
- [2] R. R. Schulz and R. L. Stevenson: “Extraction of high-resolution frames from video sequences”, *IEEE Transactions on Image Processing*, **5**, 6, pp. 996–1011 (1996).
- [3] F. W. Wheeler, X. Liu and P. H. Tu: “Multi-frame super-resolution for face recognition”, *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pp. 1–6 (2007).
- [4] P. Mortazavian, J. Kittler and W. Christmas: “3D-assisted facial texture super-resolution”, *British Machine Vision Conference* (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 (MIRU2007) 論文集*, pp. 235–240 (2007).
- [6] S. Baker and I. Matthews: “Lucas-Kanade 20 years on: A unifying framework”, *International Journal of Computer Vision*, **56**, 3, pp. 221–255 (2004).
- [7] 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 Transactions on Medical Imaging*, **18**, 8, pp. 712–721 (1999).
- [8] S. Farsiu, M. D. Robinson, M. Elad and P. Milanfar: “Fast and robust multiframe super resolution”, *IEEE Transactions on Image Processing*, **13**, 10, pp. 1327–1344 (2004).
- [9] 坂和正敏: “数理計画法の基礎”, 森北出版 (1999).
- [10] M. Turk and A. Pentland: “Eigenfaces for recognition”, *Journal of Cognitive Neuroscience*, **3**, pp. 71–86 (1991).
- [11] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes III and R. M. Mersereau: “Eigenface-based super-resolution for face recognition”, *Proceedings of IEEE International Conference on Image Processing*, Vol. 2, pp. 845–848 (2002).
- [12] A. Chakrabarti, A. N. Rajagopalan and R. Chellappa: “Super-resolution of face images using kernel PCA-based prior”, *IEEE Transactions on Multimedia*, **9**, 4, pp. 888–892 (2007).
- [13] H. Huang, H. Hea, X. Fan and J. Zhang: “Super-resolution of human face image using canonical correlation analysis”, *Journal of Pattern Recognition*, **43**, pp. 2532–2543 (2010).