社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

フレーム選択型超解像処理を用いた低解像度文字認識手法の提案

大倉　直† 出口　大輔† 高橋　友和†† 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 1-38
E-mail: †{okura,ddeguchi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ††ttakahashi@gifu.shotoku.ac.jp

あらまし　ディジタルカメラやカメラ付き携帯電話といった携帯型ディジタル撮影機器の普及に伴い，撮影した画像中の文字を認識する技術の需要が高まっている．しかし，テキストの広範囲を一度に撮影し，その中の文字を認識する場合には，撮影文字画像の低解像度化と手ぶれによる画像の劣化という問題のため，文字の認識が困難となる．本稿では，単一の低解像度文字画像では認識が困難な場合でも，動画像中の複数フレームの情報を統合することで文字画像を高精度に認識する手法を提案する．具体的には，動画像中の劣化や変形が少ないフレームを選択して超解像処理に用いることで，手ぶれなどの影響を抑えた高解像度な入力文字画像を再構成する．さらに，これを複数枚用いて複数フレーム入力型部分空間法で認識を行う．実際に撮影した動画像を入力とした実験により，提案本手法の有効性を確認した．

キーワード　文字認識，超解像，低解像度文字，部分空間法

A Proposal of a Low-resolution Character Recognition Method by Super-resolution based on Frame Selection

Ataru OKURA†, Daisuke DEGUCHI†, Tomokazu TAKAHASHI††,
Ichiro IDE†, and Hiroshi MURASE†

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

Abstract　Portable cameras such as compact digital cameras and cell-phone cameras are widely used. This trend increases the demand for character recognition techniques using them as input devices. It is difficult to recognize characters accurately when we capture an image of a large area of a document and recognize characters in it, because both the resolution and the quality of each character image become low. In this report, we propose a method for recognizing very low-resolution characters by integrating information from multiple low-resolution character images in a video captured by a hand-held camera. In particular, a super-resolution based on frame selection reconstructs high-resolution character images from multiple low-resolution character images. The low-resolution frames to be used for super-resolution are selected so that the distorted frames should be excluded. Then, an input character from multiple super-resoluted character images is recognized by a multi-frame-based subspace method. The effectiveness of the proposed method was confirmed from results of experiment on actual captured character images.

Key words　Character recognition, Super-resolution, Low-resolution character, Subspace method

1. はじめに

近年，ディジタルカメラやカメラ付き携帯電話といった携帯型ディジタル撮影機器が低価格で入手できるようになり，これらの機器を日常的に携帯することが一般的になりつつある．これに伴い，これらの機器で撮影された画像からの文字認識技術

が注目を集めている．もし，これらの機器で撮影した文字画像の自動認識が可能になれば，日常生活の中で容易に利用可能なマンマシンインタフェースのための有用な技術になると考えられる．このようなカメラ入力型文字認識は，カメラ付き携帯電話機の普及に伴い，その需要が高まっている [1] [2].

しかし，カメラで撮影された文字画像を入力とする場合，主に次の 2 つの問題がある．

- 被写体との撮影距離に依存する文字画像の低解像度化
- 撮影時の手ぶれの影響による文字画像の劣化

前者の問題については，被写体に対して接写すれば，十分認識できる解像度の文字画像を得ることは可能である．また，テキスト全体などの広い範囲を高解像度で撮影する手法に，被写体を接写した画像や動画像のモザイキングが提案されている [3]．しかし，このような手法を用いて広い範囲を認識しようとした場合，被写体全体を覆うように局所的な撮影を繰り返す必要があり，面倒な作業になる．これに対し，カメラを用いて図 1 のようにテキスト全体を一度に撮影して認識できるシステムが実現できれば，より有用なユーザインタフェースとなり得る．このためには，低解像度の文字画像を高精度に認識する技術が必要である．

後者の問題について，カメラを手で持って撮影すると，手ぶれの影響などにより，文字画像の劣化のため，認識が困難なことが多い．したがって，手ぶれの影響を抑えることも，撮影文字を認識する際に必要である．

一方，近年のディジタルカメラやカメラ付き携帯電話機は，動画像を撮影可能なものが大半を占めている．文字を動画像で撮影した場合，動画像中の文字画像は，撮影時のシャッタスピードやサンプリングレートなどの内部要因と，手ぶれや照明変化などの外部要因から影響を受ける．このため同じ文字を撮影しても，各画像間でわずかに画素値が変化する．このような動画像から得られる複数フレームの画像情報を効果的に用いることができれば，単一の画像だけからでは認識が困難な低解像度文字に対しても，高精度な認識が可能になると考えられる．

そこで本研究は，単一の低解像度文字画像では認識することが難しい場合でも，動画像中の複数フレームの情報を用いて文字画像を高解像度化することにより，認識精度を向上させることを目的とする．

上記の 2 つの問題を以下の 2 つのアプローチによってそれぞれ解決する．

- 超解像処理による文字画像の高解像度化
- 超解像処理に利用するフレームの選択

まず，複数フレームに対して超解像処理により，高解像度の文字画像を作成する．具体的には，フレーム間の位置ずれをサブピクセル単位で算出し，それらの位置ずれを基に複数フレームを統合する．これにより，単一画像では認識に十分な情報が得られない場合でも，フレーム間で情報を補完することにより，認識に十分な解像度の文字画像を得る．

しかし，携帯カメラで撮影された画像は手ぶれの影響を受ける．そのため，動画像中のフレーム列にはぶれや回転などで強く劣化した超解像処理に悪影響を及ぼすフレームが含まれる．

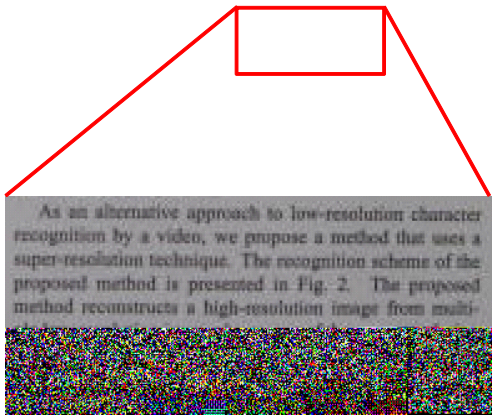

図 1　テキストの広範囲を撮影した例．各文字は低解像度になってしまう．

そこで，本研究で提案するフレーム選択型超解像処理では，この劣化を抑制するために，動画像中のフレーム列からそのような特異なフレームの除去を行う．

また，認識時には，複数枚の超解像文字画像を入力に使用し，それらの特徴空間内での分布の情報を用いることで，誤認識を誘発しやすい例外的な入力文字画像の影響を抑制する．

以降，2. では，提案手法であるフレーム選択型超解像処理を用いた低解像度文字認識手法について詳しく説明する．3. では，認識実験とその考察について述べ，最後に，4. でまとめる．

2. フレーム選択型超解像処理を用いた低解像度文字認識

2.1 概　　要

単一の低解像度文字画像では認識が困難な場合も，動画像中の複数フレームを統合することで文字画像を高精度に認識する手法を提案する．提案手法の処理は，図 2 に示す流れに従って行われる．処理はフレーム選択型超解像処理と文字認識処理の 2 つに分かれる．まず，フレーム選択型超解像処理では，動画像中のフレーム列から，超解像処理に悪影響を及ぼす回転やぶれなどの影響を強く受けた特異なフレームを除去し，残りのフレームを用いて高解像度化を行う．次に，文字認識処理では，高解像度化された超解像文字画像を用いて，柳詰らによって提案された複数フレーム入力型部分空間法 [4] により文字認識を行う．以下，それぞれの処理の詳細を述べる．

2.2 フレーム選択型超解像処理

位置ずれを含む複数枚の低解像度画像から高解像度画像を復元する処理として超解像処理がある．この超解像処理 [5] は，複数枚の低解像度画像間のサブピクセル単位での位置ずれの情報を利用し，個々の低解像度画像が持つ画素情報を統合することで，画像を高解像度化する．一般にこの超解像処理には次の 2 つの処理が含まれる．1 つ目は，複数枚の低解像度画像を用い

図 2　提案手法の処理の流れ

てサブピクセル単位での位置合わせをして統合し，高解像度画像を作成する．2 つ目は，作成された高解像度画像から光学ぼけを繰り返し計算で除去する．本研究では，1 つ目の高解像度画像を作成する処理に注目し，これを広義の超解像処理とする．

2.2.1　フレーム間の位置合わせ

ここでは，複数フレーム間の統合のために，フレーム間の位置合わせを行う．提案手法では，フレーム間の位置合わせに位相限定相関法 [6] を用いる．位相限定相関法は，画像の位相スペクトルのみを用いて 2 画像間の位置合わせを高精度に行う手法である．この手法は，照明変動や遮蔽の影響を受けにくい位相スペクトルを用いるため，ロバスト性が高いことが知られている．本研究では，複数フレーム中から基準画像 1 枚を任意に決定し，それに対して残りのフレームを位相限定相関法によって位置合わせする．

2.2.2　画像間類似度に基づくフレーム選択

携帯カメラで撮影された動画像中の各フレームには，撮影者の手ぶれにより，回転やぶれ，照明変化などの変動が含まれていると考えられる．複数フレームから超解像処理によって高解像度画像に統合する際に，これらの変動の影響を強く受けた特異なフレームが含まれていると，作成される高解像度画像の画質が劣化してしまう．提案手法では，そのような特異なフレームの影響を抑えることを目的として，超解像処理に用いるフレームを選択する．

まず，複数フレームから任意の基準画像を 1 枚選び，位相限定相関法によって求められた位置ずれ量を用いて，フレーム間の平均画像を作成する．動画像中の連続する I 枚のフレームを

$$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_I\} \tag{1}$$

とおく．サブピクセル単位でのずれを表現できるように，各画像の大きさを $N' \times N'$ に拡大する．以降，簡便のために $\boldsymbol{a}_1$ を基準画像として説明する．基準画像 $\boldsymbol{a}_1$ の原点（画像の左上）を基準とした注目画像 $\boldsymbol{a}_i$ のサブピクセル単位での位置ずれ量を $(\Delta x_i, \Delta y_i)$ とした場合，基準画像と注目画像の位置関係は図 3 のようになる．なお，注目画像中の基準画像との重複部分の画像領域を $\boldsymbol{a}'_i$ と定義する．平均画像 $\boldsymbol{\mu}$ は，この $\{\boldsymbol{a}'_1, \boldsymbol{a}'_2, \ldots, \boldsymbol{a}'_I\}$ から計算される．$\boldsymbol{\mu}$ の各画素値は次式によって計算される．

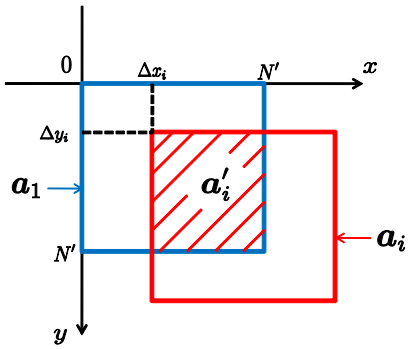

図 3　基準画像 $\boldsymbol{a}_1$ と注目画像 $\boldsymbol{a}_i$ の位置関係

$$\mu(x,y) = \frac{\sum_{i=1}^{I} a_i(x,y)}{\sum_{i=1}^{I} c_i(x,y)} \tag{2}$$

このとき，関数 a_i は，次式で定義される．

$$a_i(x,y) = \begin{cases} a'_i(x,y) & \Delta x \leqq x \leqq \Delta x + N', \\ & \Delta y \leqq y \leqq \Delta y + N' \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

これは，(x,y) が $\boldsymbol{a}'_i$ の範囲内であれば，$\boldsymbol{a}'_i$ の対応する画素値を得る関数である．また，関数 c_i は次式で定義される．

$$c_i(x,y) = \begin{cases} 1 & \Delta x \leqq x \leqq \Delta x + N', \\ & \Delta y \leqq y \leqq \Delta y + N' \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

これは，(x,y) が $\boldsymbol{a}'_i$ の範囲内であれば，1 を与える関数である．したがって，$\mu(x,y)$ は，各画素の位置に対して，重複した画像の画素値の総和を，重複した画像の枚数で割った値である．

次に，こうして求められた平均画像 $\boldsymbol{\mu}$ と重複部分 $\boldsymbol{a}'_i$ との画像間類似度を求め，それを用いて超解像処理に用いるフレームを選択する．ここで，図 4 に示すように，$\boldsymbol{a}'_i$ と対応する $\boldsymbol{\mu}$ 中の画像領域を $\boldsymbol{\mu}_i$ と定義する．このとき，$\boldsymbol{\mu}_i$ の大きさは $L_i \times M_i$ とする．L_i と M_i は以下の式で計算される．

$$L_i = N' - |\Delta x_i| \tag{5}$$

$$M_i = N' - |\Delta y_i| \tag{6}$$

画像間類似度 w_i を次式によって定義する．

$$w_i = \exp\left(-\frac{d_i}{v}\right) \tag{7}$$

$$d_i = \frac{||\boldsymbol{\mu}_i - \boldsymbol{a}'_i||^2}{L_i M_i} \tag{8}$$

ここで，v は実験的に決定する．以下では，$v = 0.0001$ とする．こうして w_i を平均画像と各フレームの類似度として扱い，平均画像との類似度が小さなフレームを特異なフレームとして除外することにより，高解像度化に用いるフレームを選択する．

表 1　撮影に使用したディジタルカメラの仕様

撮影機器	解像度	フレームレート
Canon PowerShot G9	360 × 240 pixels	30 fps

表 2　解像度の表記方法

表記	撮影距離	文字 'A' の平均解像度
size S	40 cm	6.0 × 6.0 pixels
size M	35 cm	6.5 × 6.5 pixels
size L	30 cm	7.0 × 7.0 pixels

被写体とした．文字の切り出しを行った結果，一つのフレームから全 62 文字を切り出すことができなかったフレームは実験対象から除外した．本実験では，カメラと被写体との距離を変化させることによって文字画像の解像度を調節した．実験で用いた文字の解像度は表 2 に示すように，size S，M，L と表記する．なお，解像度は，表 2 の撮影距離で撮影した低解像度文字画像の 'A' の縦幅の平均値を基準とした．また，次の 2 通りの条件で本実験の撮影を行った．

blur S:　手でカメラを持ち，なるべく静止して撮影する．

blur L:　手でカメラを持ち，揺らしながら撮影する．

これらの条件の下，部分空間の学習用画像と認識用の入力画像を撮影した．

3.2.2　学　　習

認識実験にあたり，部分空間の学習を行う必要がある．提案手法では，フレーム選択をしない単純な超解像処理により作成された超解像文字画像を用い，従来手法では，低解像度文字画像をそのまま用いた．また，両手法間での条件を統一するため，同じ動画像系列からそれぞれ学習した．このとき，提案手法では，まず，size L で撮影された動画像中の 30 枚の低解像度画像から 1 枚の超解像画像を作成した．次に，そこから切り出した超解像文字画像を全 62 カテゴリにつきそれぞれ 100 枚ずつ用意し，それらを用いて部分空間を学習した．なお，超解像処理では，4 × 4 倍の解像度に高解像度化した．一方，従来手法では，size L で撮影された低解像度画像からそのまま切り出し，全 62 カテゴリにつきそれぞれ 100 枚ずつを用いて，部分空間を学習した．どちらの手法でも，学習に用いる文字画像は 32 × 32 pixels に正規化し，予備実験の結果より，各カテゴリの部分空間を形成する固有ベクトルの数は，固有値が大きな順に 5 個とした．

3.2.3　認　　識

従来手法と認識率を比較することで，提案手法の有効性を評価した．両手法間の公平な比較を行うために，1 回の認識に使用する低解像度画像は同一のものを使用した．具体的には，どちらも動画像中の連続する 50 フレームを用いて実験した．このとき，提案手法では，その 50 フレーム中の 1 フレームを基準画像としてフレーム選択処理によって 30 枚を選択し，1 枚の超解像文字画像を作成した．これを，50 フレームそれぞれを基準画像として 50 回フレーム選択型超解像処理を行うことで，50 枚の超解像画像を作成し，複数フレーム入力型部分空間法で認識を行った．一方，従来手法では，50 フレームからそれぞれ切り出された文字画像 50 枚を複数フレーム入力型部分空間法で認識した．

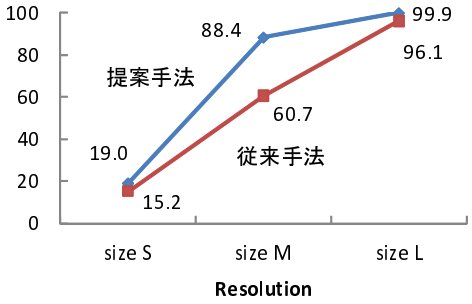

図 5　blur S での認識結果

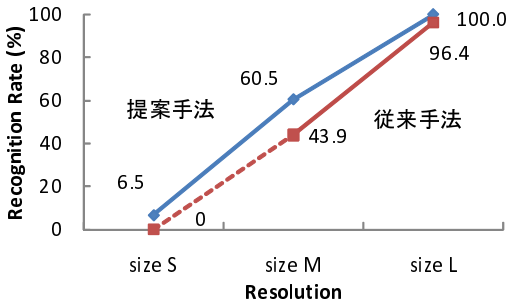

図 6　blur L での認識結果

本実験では，上記の条件で，3 つの解像度それぞれにつき撮影条件を変えて認識率を求めることで，認識性能の比較を行った．認識回数は，全 62 カテゴリそれぞれ 50 回，計 3,100 回とした．

3.3　実 験 結 果

3.3.1　blur S での結果

blur S で，撮影された文字画像を入力とした場合の認識率を図 5 に示す．どの解像度においても提案手法の方が従来手法よりも高い認識率が得られたことが分かる．特に，解像度 size M では 25% 以上の認識率の向上が見られた．

3.3.2　blur L での結果

blur L で，撮影された文字画像を入力とした場合の認識率を図 6 に示す．どの解像度においても提案手法の方が従来手法よりも高い認識率を得られたことがわかる．解像度 size M では 15% 以上の認識率の向上が見られた．また，解像度 size S の低解像度画像からは，画質の劣化と低解像度化の影響により，文字画像を切り出すことが不可能であったため，従来手法による認識率は 0 とみなした．

3.4　考　　察

3.4.1　提案手法と従来手法の比較

blur S，blur L どちらにおいても，提案手法は従来手法よりも良い結果が得られた．前者の解像度別認識結果を見ると，size S，size L ではあまり差が見られなかった．

size S に関して，この解像度は文字を画像として表現できる限界であると考えられる．例として，size S で切り出された文字画像 'B' とそれを提案手法で高解像度化した文字画像 'B' を図 7 に示す．これを見ると，撮影された低解像度文字画像は，目視でも文字カテゴリを判断することは難しい．このように低

(a) 低解像度文字画像　　(b) 超解像文字画像

図 7　size S の文字画像例（カテゴリ：‘B’）

(a) 低解像度文字画像　　(b) 超解像文字画像

図 8　size M の文字画像例（カテゴリ：‘B’）

(a) 低解像度文字画像　　(b) 超解像文字画像

図 9　size L の文字画像例（カテゴリ：‘B’）

解像度画像中の文字の画像情報が消失してしまった場合，これらに対して超解像処理を行っても，画像間で画素情報を補完しあうことができないため，認識率があまり改善されなかったと考えられる．一方，size L では，従来手法と提案手法の両手法で文字を認識するにあたり十分な解像度が得られたため，認識率に差が見られなかったと考えられる．

一方，size M では blur S，blur L ともに，提案手法と従来手法との認識率に大きな差が生じた．このとき，切り出された文字画像 ‘B’ とそれを提案手法で高解像度化した文字画像 ‘B’ を図 8 に示す．これは，複数フレーム入力型部分空間法での認識の際に，提案手法では，例外的な入力画像の影響を抑制することができたためと考えられる．入力文字画像の低解像度化とともに，複数フレーム中の特異な入力画像の割合は増加していく．そのようなフレームに対して複数フレーム入力型部分空間法を適用した場合，入力画像の分布は特異な入力画像の影響を強く受け，正しく認識されるはずの入力文字画像の効果を低下させてしまう．このような理由から，従来手法では size M から認識性能が大きく低下したと考えられる．これらのことから，提案手法は従来手法よりも低解像度な文字を安定して認識できることを確認した．

3. 4. 2　撮影条件の違い

blur S と blur L の実験結果を比較すると，後者では，文字画像の解像度が低下した場合の認識率の低下が著しいことがわかる．これは，テキストなどの被写体を広範囲に撮影した際に，撮影画像の品質が手ぶれの影響を大きく受けるようになったためと考えられる．blur L で撮影されたときの size M の結果を見ると，提案手法により超解像文字画像を作成して認識した方が高い認識率を得られたことがわかる．このことから，大きな手ぶれが発生する条件においても提案手法の有効性を確認した．

4. むすび

本稿では，フレーム選択型超解像処理を用いた低解像度文字認識の手法を提案した．撮影時のぶれや回転，照明変化などを要因として，劣化や変形が加わった特異なフレームが，超解像画像の劣化や変形の原因となるため，このような影響を抑制することを目的として超解像処理に用いるフレームの選択を行った．また，フレーム選択型超解像処理によって得られた複数枚の超解像文字画像を入力として，複数フレーム入力型部分空間法を用いて文字認識した．ここでは，入力文字画像の分布を推定し，その分布からの距離に応じて入力画像間に重みを付けることで，分布から外れた誤認識を誘発しやすい入力文字画像の影響を抑えた．

実際に撮影した動画像を入力として，提案手法による低解像度文字の認識実験を行った．実験の結果，提案手法による認識率は従来手法を上回り，最大で 25% 以上の認識率の向上を確認した．

今後の課題としては，超解像処理のフレーム選択において位置ずれを考慮した平均画像を推定する際，繰り返し処理によってその推定精度を向上させることが挙げられる．これにより，大きな手ぶれが生じたり，照明が変化したりするような環境においても，ロバストな超解像処理が可能となり，より実用的な文字認識システムの実現が期待できる．

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，科学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.ac.jp/`）を使用した．

文　献

[1] D. Doermann, J. Liang, and H. Li, “Progress in camera-based document image analysis,” Proc. 5th Int. Conf. on Document Analysis and Recognition, pp.606–616, Edinburgh, Scotland, August 2003.

[2] 黄瀬浩一, 大町真一郎, 内田誠一, 岩村雅一, “カメラを用いた文字認識・文書画像解析の現状と課題,” 電子情報通信学会技術研究報告, PRMU2004-246, February 2005.

[3] 池谷彰彦, 中島昇, 佐藤智和, 池田聖, 神原誠之, 横矢直和, 山田敬嗣, “カメラパラメータ推定による紙面を対象とした超解像ビデオモザイキング,” 画像の認識・理解シンポジウム（MIRU2004）講演論文集, pp.505–510, July 2004.

[4] S. Yanadume, Y. Mekada, I. Ide, and H. Murase, “Recognition of very low-resolution characters from motion images captured by a portable digital camera,” Proc. 2004 Pacific-Rim Conf. on Multimedia, Lecture Notes in Computer Science, vol.3332, pp.489–496, Springer-Verlag, December 2004.

[5] C. P. Sung, K. P. Min, and M. G. Kang, “Super-resolution image reconstruction: A technical overview,” IEEE Signal Process. Mag., vol.20, no.3, pp.21–36, May 2003.

[6] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, “High-accuracy subpixel image registration based on phase-only correlation,” IEICE Trans. Fundamentals, vol.E86-A, no.8, pp.1925–1934, August 2003.