

# Normalized Total Gradient: A New Measure for Multispectral Image Registration

Shu-Jie Chen, Hui-Liang Shen<sup>ID</sup>, Chunguang Li<sup>ID</sup>, *Senior Member, IEEE*, and John H. Xin

**Abstract**—Image registration is a fundamental issue in multispectral image processing, and is challenged by two main characteristics of multispectral images. First, the regional intensities can be essentially different between band images. Second, the local contrasts of two difference band images are inconsistent or even reversed. Conventional measures can align images with different regional intensity levels, but may fail in the circumstance of severe local intensity variation. In this paper, a new measure called normalized total gradient is proposed for multispectral image registration. The measure is based on the key assumption (observation) that the gradient of the difference between two aligned band images is sparser than that between two misaligned ones. A registration framework, which incorporates image pyramid and global/local optimization, is further introduced for affine transform. Experimental results validate that the proposed method is not only effective for multispectral image registration, but also applicable to general unimodal/multimodal image registration tasks. It performs better than or comparable to the existing methods, both quantitatively and qualitatively.

**Index Terms**—Multispectral image, multimodal image, image registration, band image, similarity measure, image gradient, normalized total gradient, affine transform, image pyramid, optimization.

## I. INTRODUCTION

MULTISPECTRAL color imaging has attracted intensive interest in recent years as it can acquire more spectral information than traditional RGB cameras. A multispectral imaging system can be set up by using a camera and an optical device that splits the visible spectrum reflected from the imaged object. Typical light-splitting devices include electronically controlled tunable filters [1], [2] or mechanically controlled filter wheels [3]. Compare with tunable filters, the filter wheel has the flexibility in filter selection and can achieve high spectral transmittance in the whole visible spectrum range. Fig. 1 shows a multispectral imaging system

that consists of a monochrome digital camera and a filter wheel installed with a series of optical filters. The filter wheel based multispectral imaging system usually suffers from image misalignment and out-of-focus blur. The misalignment between band images is caused by the non-coplanar placement of filters [3], [4]. The inter-band misalignment causes chromatic aberration (Fig. 14), and precludes the direct use of multispectral image. Although calibration technique is available for multispectral image alignment, it requires a particular calibration board, and in addition, the system must be recalibrated when the imaging distance changes. Hence, it is more convenient and favorable to explore a calibration-free multispectral image registration method for practical applications. On the other hand, the out-of-focus blur is caused by the different effective focal lengths at individual bands, which originates from the fact that the refractive indices of the lens are wavelength dependent. To deal with this problem, an autofocus method [5] has been employed to compute the focus positions of lens using a focusing device (i.e., step motor). Alternatively, when the band images are well aligned, the blur can also be computationally removed by multispectral image deblurring [6]. In this work, we will show in Section VI that multispectral images can be restored based on the proposed registration framework and the deblurring method introduced in [6]. In this way, neither the calibration board nor the focusing device is further needed in such multispectral imaging systems.

Image registration aims to find correspondences between two images by maximizing specific similarity measure (or equivalently, minimizing distance measure). Extensive surveys of registration techniques can be found in [7] and [8]. Registration techniques can be classified into two categories according to the descriptors that formulate the similarity measures [8], i.e., the feature-based one and intensity-based one. Feature-based methods usually involve feature detection, feature matching, transform model estimation, and resampling. Generally these methods are more efficient but less accurate when compared to the intensity-based methods. As one of the most successful feature-based methods, scale invariant feature transform (SIFT) has been widely used in a variety of matching tasks in computer vision [9]. Intensity-based registration methods have been applied in medical imaging [7]. Similarity measures, including sum-of-squared-differences (SSD), cross correlation (CC), correlation ratio (CR), and mutual information (MI), have been widely employed in either unimodal or multimodal image registration [10]–[13].

Multispectral image challenges the conventional intensity-based registration methods at two aspects. First, the intensities

Manuscript received October 20, 2015; revised June 2, 2017, August 9, 2017, and September 28, 2017; accepted November 17, 2017. Date of publication November 22, 2017; date of current version December 27, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61371160, in part by the Zhejiang Provincial Key Research and Development Project under Grant 2017C01044, and in part by the Fundamental Research Funds for the Central Universities under Grant 2017XZZX009-01. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xudong Jiang. (Corresponding author: Hui-Liang Shen.)

S.-J. Chen, H.-L. Shen, and C. Li are with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: incredible77@zju.edu.cn; shenhl@zju.edu.cn; cgli@zju.edu.cn).

J. H. Xin is with the Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong (e-mail: john.xin@polyu.edu.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2017.2776753

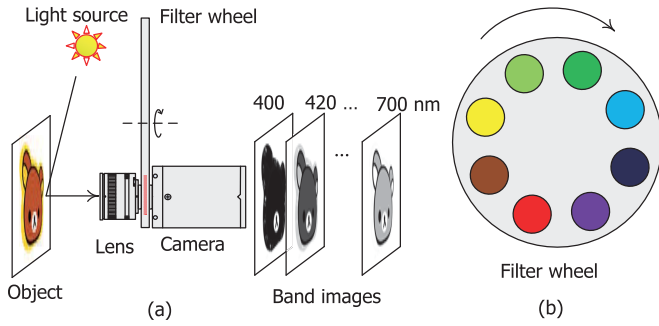


Fig. 1. The multispectral imaging system (a) and filter wheel (b) used in this work. Note that in our system totally 16 filters are installed on the wheel, but only 8 filters are shown for illustration.

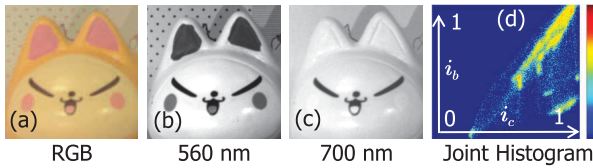


Fig. 2. Local contrast inconsistency of an aligned multispectral image. (a) Multispectral image displayed in RGB. (b) Band image at wavelength 560 nm. (c) Band image at wavelength 700 nm. (d) Joint histogram of intensities of image (b) and (c). Warmer color denotes larger number of counts.

of same local region in different band images can be essentially different. Second, the local intensity may vary in some band images while keeps constant in others, leading to inconsistent or even reverse contrast. As an example, Fig. 2 illustrates the phenomenon of contrast inconsistency. Because the spectral responses of yellow face and pink regions on the face are diverse at 560 nm but similar at 700 nm, the pink regions are visible in the band image at 560 nm, but vanish in the band image at 700 nm. According to [14], image registration methods cannot deal with obvious structure inconsistency in multispectral or multimodal images when without using an effective similarity measure.

This paper proposes a new intensity-based method to resolve the mentioned challenges for multispectral image registration. The normalized total gradient (NTG), which is applied on the difference between two band images (referred as *difference image* hereafter), is introduced to measure the alignment degree of two images. The employment of NTG is based on the key assumption that the gradient of difference image between two aligned images is sparser than that between two misaligned images. This assumption is validated by the statistical investigation on a large number of multispectral images of real-scenes.

Based on the NTG measure, an image registration framework that incorporates image pyramid and global/local optimization is introduced for general affine transform, as shown in Fig. 3(a). During the optimization, an upside-down image pyramid is first constructed. Global optimization algorithm (e.g., differential evolution [15]) is used to find a reliable starting point  $\mathbf{p}_0$  at the bottom layer of the pyramid. To update the parameters at the  $k$ th layer of the pyramid, local optimization method (e.g., Newton's method) is implemented on the  $k$ th image pair whose floating image is first warped

using the parameters  $\mathbf{p}_{k-1}$  transferred from the lower layer. The parameter transfer process abridges the parameter gap between the current image and its down sampled image. Local optimization stage continues until it reaches the top layer of the pyramid. Fig. 3(b) shows the computation of the NTG value, which will be discussed in detail in Section III.

The success of NTG measure is due to two reasons. First, the gradient operation relieves the influence of the local intensity that varies slowly. It is straightforward that when the local intensity keeps constant or varies slowly, its gradient scarcely changes. Second, the global sparseness measure used in NTG bears a rapidly varying local intensity such as contrast reversal. Since the rapid intensity variation is local, it contributes rather few unpleasant edges to the gradient of difference image. As a result, the sparseness of the gradient of difference image is dominated by the counteraction of those common structures. In other words, the sparseness is mainly determined by the degree of alignment.

We note that gradient sparseness of difference image is a common attribute of unimodal or multimodal images. Though the NTG measure is originally developed from multispectral images, it can be successfully applied to general image registration tasks. For illustration the registration of medical image pairs, flash/no-flash image pair, and multi-sensor image pair are given in Section VI.

The main contributions of this work can be summarized as follows:

- NTG is introduced as a registration measure based on the assumption that the gradient of difference image is most sparsely distributed when two images are well aligned.
- A framework, which is comprised of image pyramid and global/local optimization, is proposed for multispectral image registration.
- In addition to multispectral images, the registration framework is also applicable to unimodal and multimodal images.

The remainder of this paper is organized as follows. Section II reviews the related work on multispectral/multimodal image registration. The NTG measure is introduced in Section III and the image registration framework is described in Section IV. Section V shows the experimental results and Section VI illustrates the applications of the proposed image registration method. Finally, Section VII concludes the paper.

## II. RELATED WORK

Intensity-based image registration can be categorized into unimodal and multimodal image registrations, judged by the criterion whether corresponding pixels have similar intensity values. In this sense, multispectral image registration belongs to the multimodal registration category. In the following only the most relevant multimodal image registration methods are reviewed.

### A. Mutual Information (MI)

MI was introduced for medical image registration in [12] and extensively investigated in [13]. It is defined as

$$I(A, B) = H(A) + H(B) - H(A, B), \quad (1)$$

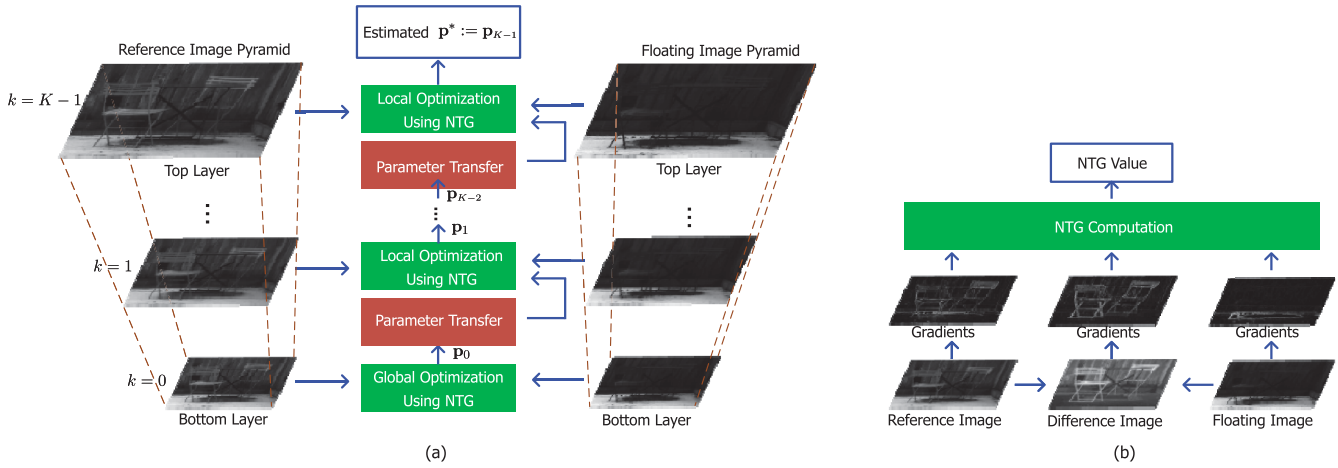


Fig. 3. Multispectral image registration using NTG. (a) Registration framework. (b) NTG computation.

where  $H(\cdot)$  denotes the entropy of image, and  $H(A, B)$  denotes the entropy of the joint histogram of image  $A$  and  $B$ . When two images are correctly registered, corresponding regions should overlap and the joint histogram shows certain clusters for the intensities of those regions. This results in a least joint entropy  $H(A, B)$  [13]. Thus maximizing mutual information can be viewed as registering images  $A$  and  $B$  such that, in the overlap, the information ( $H(A)$  and  $H(B)$ ) provided by the images is large and the local regions are in good match. To further increase immunity against overlap, normalized mutual information has been explored [16].

As a variant of MI, regionalized mutual information was introduced to cope with the local intensity variation [17]. By assuming that local intensity varies slowly across the image, global mutual information is constructed from local regional distribution. Alternatively, the impact of slowly varying local intensity can also be relieved by computing MI in regions of large spatial variations (e.g., corners and edges) [18], [19]. In [18] MI is augmented with the Harris operator and in [19] MI is modified with the assistance of saliency measure. Such methods perform well on images with slow intensity variation at the cost of high computation complexity.

### B. Correlation Ratio (CR)

While MI describes the clustering property of joint histogram using entropy, CR characterizes such property using variance [11]. CR stems from the law of total variance

$$\text{var}(A) = \mathbb{E}_B(\text{var}(A|B)) + \text{var}_B(\mathbb{E}(A|B)). \quad (2)$$

When the joint histogram is well clustered, for every single intensity that represents some local region in  $B$ , the variance of intensity in the same local region of  $A$  should be small. This leads to a small value of  $\text{var}(A|B)$ . For all intensity values of  $B$  in the region of overlap, the mean variance of  $A$  denoted by  $\mathbb{E}_B(\text{var}(A|B))$  is hence the smallest when registered. Accordingly, the proportion of the second term of the right hand side in (2) arrives its maximum, which indicates

the correlation ratio

$$CR = \frac{\text{var}_B(\mathbb{E}(A|B))}{\text{var}(A)} \quad (3)$$

is maximized.

Recently, CR was thrown light upon again in the view of pattern matching. The work [20] starts from the discussion of matching by tone mapping (MTM) and ends up with a matching metric in form of CR.

Both MI and CR describe the regional correspondence of two different modal images. When the intensity of two images can be well approximated by functional mapping and the variation of regional intensity follows some form of distribution, both MI and CR can achieve good registration. In CR-based registration, variance is used to best measure the variation of regional intensity that follows the normal distribution. In MI-based registration, entropy is employed to measure the sparseness of the joint histogram and bears a wider range of distributions. However, functional correspondence does not exist in multispectral images. As an example, Fig. 2(d) shows the intensity correspondences of two aligned band image patches (560 nm and 700 nm). Though the joint histogram exhibits clusters, no functional mapping is observed.

### C. Local Normalized Cross Correlation

While MI and CR are global intensity-based registration methods, local normalized cross correlation (NCC) [10] has been introduced to measure the correspondence between two modal images locally. Based on the assumption that the directional derivative energy maps are locally linearly correlated, NCC can be used to measure the similarity between two multimodal patches. Since not all patches satisfy this assumption, a build-in Hessian-based outlier rejection mechanism is employed in [10] when constructing a global cost function.

A variant of NCC called robust selective normalized cross correlation (RSNCC) [14], was recently introduced to solve the dense matching problem of both multispectral and natural images. To deal with structure inconsistency and notable gradient variation, a robust function is designed to balance

the local normalized cross correlation between intensities and gradients. The robust selective function aims to reduce the risk of registration inaccuracy, which can be caused in regions violating the linearity assumption.

#### D. Residual Complexity (RC)

Our proposed NTG is closely related to [21] in the sense of measuring the complexity of difference image. In [21] the difference of two unimodal images with local intensity variation, referred as the residual image, was used. It was shown that the complexity of RC reduces to its minimum when two images are aligned. RC is computed by measuring the sparseness of discrete cosine transform (DCT) coefficients of difference image. However, when two images are of different modality, the sparseness of DCT coefficients is no longer guaranteed. The reason is that in multimodal cases the intensities of two corresponding regions do not counteract with each other in difference image when registered. Compared to RC, the proposed NTG measures the sparseness of the gradient of difference image. By taking the gradient, the residual intensity counteracts with its neighbors so that multimodality induces less impact on sparseness measurement.

#### E. Phase Correlation (PC)

PC is a well known FFT-based unimodal image registration technique [22] and has been extended to cover multimodal image by taking gradients/edges of multimodal images as inputs [23]. Assuming the gradients of two multimodal images are global phase-correlated, phase correlation technique can find the rigid transform between two multimodal images quickly. In the case of multispectral image registration, the spatial contrast differences may violate the global phase-correlated assumption. Besides, PC is not applicable to nonrigid deformation (e.g., affine transform), which is actually a common circumstance in multispectral image registration.

### III. NORMALIZED TOTAL GRADIENT (NTG)

Statistics of image features are of significant use in the image processing and computer vision fields. The sparseness of gradients of natural images has been prevalently applied in single image deblurring [24], denoising [25], and inpainting [26]. Statistics of multispectral images have also been exploited in efficient image representation [27]. In this work, the feature of gradient sparseness, which is computed on difference image, is explored for multispectral image registration.

#### A. Sparseness

To validate the sparseness assumption of the gradient of difference image, a multispectral image database with 77 real-world scenes [27] is employed. Some images are shown in Fig. 4(a). Each multispectral image consists of 31 band images that cover the visible spectrum from 420 nm to 720 nm at interval 10 nm. Image intensities are normalized to the range [0, 1]. All multispectral images are originally aligned. The misaligned multispectral images are generated by imposing displacements on all bands except for the reference band.

Without loss of generality, we use the band image at 560 nm as the reference image as it is usually less affected by noise. The *difference image* is then computed as the difference between the reference image and another band image.

Fig. 4 shows the distributions of gradient of difference image when the imposed displacements (with  $\Delta x = \Delta y$ ) vary from 0 to 8 pixels. The distribution is more heavy-tailed when the displacement becomes smaller. There will be more zero gradients and less large gradients in difference image when two images are gradually aligned. The distribution of gradients is the sparsest in case of perfect alignment. As a consequence, the absolute sum of the gradients of the aligned difference image is smaller than that of the misaligned difference image. Let  $f_R(\mathbf{x})$  denotes the reference image with spatial coordinates  $\mathbf{x} = (x, y)^T$ ,  $f(\mathbf{x})$  and  $\tilde{f}(\mathbf{x})$  denotes the aligned and misaligned floating images, this observation can be formulated as

$$\sum_{\mathbf{x} \in \Omega} |\nabla_l(f(\mathbf{x}) - f_R(\mathbf{x}))| \leq \sum_{\mathbf{x} \in \Omega} |\nabla_l(\tilde{f}(\mathbf{x}) - f_R(\mathbf{x}))|, \quad (4)$$

where operator  $\nabla_l$ , with  $l \in \{x, y\}$ , denotes the derivative along the direction  $l$ , and  $\Omega$  denotes the effective domain of image. Using the  $L_1$ -norm, (4) can be written as

$$\|\nabla_l(\mathbf{f} - \mathbf{f}_R)\|_1 \leq \|\nabla_l(\tilde{\mathbf{f}} - \mathbf{f}_R)\|_1, \quad (5)$$

where  $\mathbf{f}$ ,  $\mathbf{f}_R$ ,  $\tilde{\mathbf{f}}$  respectively denotes the vectorized form of image  $f(\mathbf{x})$ ,  $f_R(\mathbf{x})$ ,  $\tilde{f}(\mathbf{x})$ , and

$$\|\nabla_l(\mathbf{f} - \mathbf{f}_R)\|_1 = \sum_{\mathbf{x} \in \Omega} |\nabla_l(f(\mathbf{x}) - f_R(\mathbf{x}))|.$$

By summing over all directional derivatives we have

$$\sum_l \|\nabla_l(\mathbf{f} - \mathbf{f}_R)\|_1 \leq \sum_l \|\nabla_l(\tilde{\mathbf{f}} - \mathbf{f}_R)\|_1. \quad (6)$$

Naturally the  $L_1$ -norm plays a role of global sparseness measure on the gradients of difference image. In the literatures the  $L_1$ -norm of the gradient of the image along all directions ( $x$  and  $y$  in this work) is referred as the anisotropic total gradient of the image [28]. Since it is deduced from the distribution of the gradients of difference image in this work, we refer it the total gradient (TG) of difference image. According to (6) the TG of the aligned different image is smaller than that of the misaligned one, hence minimizing the TG of different image is equivalent to image pair alignment.

The relationship between image alignment and gradient sparseness can be intuitively explained. It is expected that, when two images are aligned, the gradients of these images overlap with each other. Consequently, the number of nonzero gradients of difference image (measured by  $L_1$ -norm in TG computation) reduces.

#### B. Normalization

A measure is computed on the pixels in the region of overlap between two images. It would be of limited use if it is sensitive to overlap variation. The work [16] presents an extensive discussion on the overlap, and introduces the normalized mutual information (NMI) to reduce the effect of varying overlap on entropy measure.

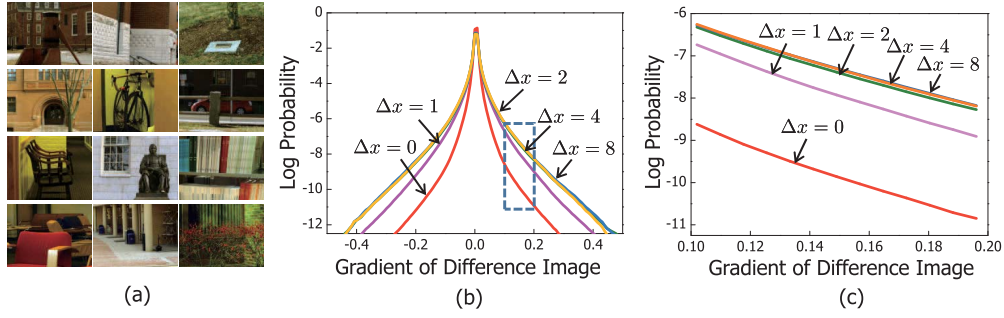


Fig. 4. (a) Multispectral images of real-world scenes (displayed in RGB). (b) Distributions of gradients of difference image in cases of various displacements. (c) Close-up view of the highlighted area in (b). Note that we set the displacements  $\Delta x = \Delta y$  in the simulation, for clarity only  $\Delta x$  is shown here.

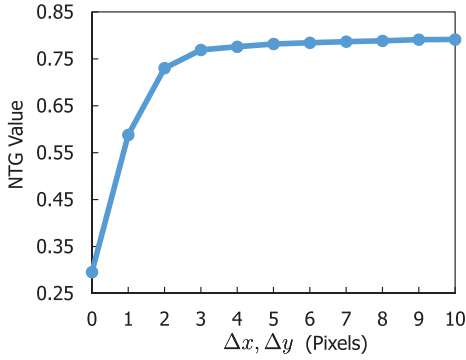


Fig. 5. Average NTG values on 77 multispectral images with respect to spatial displacements  $\Delta x$  and  $\Delta y$ .

The TG of difference image is also sensitive to the overlap because the summation of absolute gradients is defined on the region of overlap  $\Omega$  when two images are misaligned. The value of total gradient reduces when the region of overlap become smaller, and eventually falls to zero when there is no overlap between two images. This case should be definitely avoided during the registration process. Hence, normalization with respect to overlap is necessary. If we treat the  $L_1$ -norm as a kind of energy measure, the total energy in two images can be computed as  $\sum_l (\|\nabla_l \mathbf{f}\|_1 + \|\nabla_l \mathbf{f}_R\|_1)$  and the normalized total gradient (NTG) of difference image is defined as

$$NTG(f, f_R) = \frac{\sum_l \|\nabla_l (\mathbf{f} - \mathbf{f}_R)\|_1}{\sum_l (\|\nabla_l \mathbf{f}\|_1 + \|\nabla_l \mathbf{f}_R\|_1)}. \quad (7)$$

In (7), the numerator is the total gradient of difference image and the denominator is the total energy that normalizes the measure. It can be easily verified that  $0 \leq NTG \leq 1$ . Minimizing NTG is equivalent to minimizing the relative total gradient of difference image under the normalization of total energy.

To verify the effectiveness of NTG, we compute the NTG values on 77 multispectral images [27] with respect to different displacements (with  $\Delta x = \Delta y$ ). Fig. 5 shows the trend of average NTG values with respect to different displacements. It is observed that the NTG value decreases slowly when the displacement varies from 10 to 4 pixels. However, it changes abruptly when the displacement is less than 4 pixels, and reaches minimum in the case of perfect alignment ( $\Delta x =$

$\Delta y = 0$ ). This trend is consistent to the distributions of gradients in Fig. 4. The gradients and corresponding NTG values, which are computed from two band images with various displacements, are illustrated in Fig. 6. As shown, a smaller displacement results in a sparser gradient image and accordingly a smaller NTG value.

Based on the above analysis, NTG is an excellent indicator for perfect image alignment, but cannot easily distinguish large image displacements. Consequently, the objective function defined by NTG (see (10)) is not convex with respect to transform parameters. To cope with this characteristic, we use image pyramid and global optimization in the registration framework. By sub-sampling, the image pyramid relieves the impact of large displacements and makes the global region of convergence be easily traced. The global optimization (such as differential evolution [15]) further reduces the risk of being stuck into local minimums. The success of this strategy will be validated by experiments.

#### IV. MULTISPECTRAL IMAGE REGISTRATION

The problem of multispectral image registration can be cast as the minimization of NTG measure with respect to either parametric or non-parametric transform. In the following we focus on image registration with global affine transform. Affine transform has been frequently used as the pre-processing of medical image registration [29]. It has been verified that affine transform suffices in aligning band images [30].

The aim of image registration is to find a transform with parameter  $\mathbf{p}$  such that the transformed floating image  $g(\mathbf{x}, \mathbf{p}) = f(u(\mathbf{x}, \mathbf{p}), v(\mathbf{x}, \mathbf{p}))$  matches the reference image  $f_R(\mathbf{x})$ . The affine transform can be formulated as

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{pmatrix} \mathbf{p}, \quad (8)$$

where  $\mathbf{p} = (p_1, p_2, \dots, p_6)^T$ . Then aligning the floating and reference images is equivalent to solve the parameter  $\mathbf{p}^*$ ,

$$\mathbf{p}^* = \underset{\mathbf{p}}{\operatorname{argmin}} J(\mathbf{p}), \quad (9)$$

where the objective function  $J(\mathbf{p})$  is defined as

$$\begin{aligned} J(\mathbf{p}) &:= NTG(g, f_R) \\ &= \frac{\sum_l \sum_{\mathbf{x} \in \Omega(\mathbf{p})} |\nabla_l (g(\mathbf{x}, \mathbf{p}) - f_R(\mathbf{x}))|}{\sum_l \sum_{\mathbf{x} \in \Omega(\mathbf{p})} (|\nabla_l g(\mathbf{x}, \mathbf{p})| + |\nabla_l f_R(\mathbf{x})|)} \end{aligned} \quad (10)$$

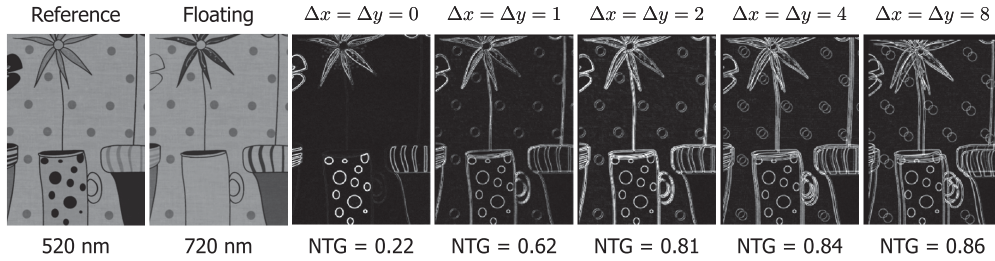


Fig. 6. Gradients of difference image and corresponding NTG values in cases of various displacements.

and  $\Omega(\mathbf{p})$  denotes the region of overlap. In this case, the region of overlap is defined as a set of points who have valid values after parametric transform, i.e.  $\Omega(\mathbf{p}) := \{(\mathbf{x}, \mathbf{y}) | (u(\mathbf{x}, \mathbf{p}), v(\mathbf{x}, \mathbf{p})) \in \Omega\}$  and  $\Omega$  is the domain of image.

#### A. Framework

The image registration framework (Fig. 3) consists of global optimization and local optimization. In the framework, images are first subsampled into  $K$  layers to form an upside down image pyramid whose lower layer is the subsampled version of the upper one. Global optimization is performed on the bottom layer to obtain a good initial estimation for the successive optimization. The initial point is transferred to the upper layer and refined by local optimization until reaching the top layer. This hybrid optimization strategy balances the computation efficiency and accuracy of image registration. In this work we use the differential evolution (DE) [15] and the well-known Newton's method to serve as the global and local optimizers, respectively. Differential evolution (DE) is a genuinely useful global optimization algorithm and has earned a reputation as a very effective and reliable global optimizer [15]. Other global and local optimization algorithms can also be incorporated into our framework.

#### B. Global Optimization

In global optimization, we use the DeMat library [15] to search the initial point at the bottom layer. The parameters of the DE algorithm are set according to practical requirement. For our captured multispectral images, we found that a population of 30 individuals and 200 generations are usually adequate for deriving a good initial point. Other parameters are set by default [15]. Since no severe deformation is observed and large displacements have been reduced by down-sampling, the searching space of  $\mathbf{p} = (p_1, \dots, p_6)^T$  is constrained from  $[0.95, -0.05, -10, -0.05, 0.95, -10]$  to  $[1.05, 0.05, 10, 0.05, 1.05, 10]$ . Larger displacements are adopted in the experiment to extensively evaluate the registration algorithm.

The choice of number of pyramid layers,  $K$ , is relevant to the sampling factor and image size. In this work the sampling factor is simply set as  $\alpha = 2$ . A large number of pyramid layers can accelerate the global optimization process, but image structure will be lost in the bottom layer. This will consequently result in unreliable initial point for the local

optimization process. According to our experience, a four-layered image pyramid (i.e.,  $K = 4$ ) is appropriate for our acquired multispectral images with  $1392 \times 1040$  pixels. It well balances the computational efficiency and accuracy.

#### C. Local Optimization

In local optimization, the parameters at each layer is updated as

$$\mathbf{p}^{t+1} = \mathbf{p}^t - (\mathbf{H}_{\mathbf{p}}^J)^{-1} J_{\mathbf{p}}|_{\mathbf{p}=\mathbf{p}^t}, \quad (11)$$

where  $\mathbf{H}_{\mathbf{p}}^J$  and  $J_{\mathbf{p}}$  are, respectively, the Hessian matrix and gradient of the objective function  $J(\mathbf{p})$  with respect to parameter  $\mathbf{p}$ . The superscript  $t$  denotes the iteration index. Since the global optimization has provided a reliable initial point, we found that 6 iterations are usually sufficient for Newton's method to refine the parameter. Readers are suggested to refer Appendix A for more detailed implementation of Newton's method.

#### D. Parameter Transfer

There is a gap between the initial parameter of images at  $k$ th layer and the estimated parameter of images at  $(k-1)$ th layer due to the down sampling operation. When affine transform is used, we simply scale the translation parts of the parameter with the sampling factor  $\alpha$  to abridge this gap. More specifically, if the translation parameters estimated at the  $(k-1)$ th layer are  $p_3$  and  $p_6$ , their initial values are set as  $\alpha p_3$  and  $\alpha p_6$  at the  $k$ th layer.

### V. EXPERIMENTS

Experiments were conducted on our multispectral images and the CAVE database [31]. Our multispectral images (see Fig. 7) were captured using the imaging system illustrated in Fig. 1. The acquired band images are sharp and not affected by out-of-focus blur with a mechanical focusing device [5]. The CAVE database consists of 31-band multispectral images (400 – 700 nm, at 10 nm intervals) of 31 static scenes (totally 961 band images) and is publicly available.<sup>1</sup> Fig. 8 illustrates eight sample images.

The MI [12], RC [21], CR [11], RSNCC [14], and proposed NTG measures are incorporated into the same image registration framework (global optimization only) for fair comparison.

<sup>1</sup><http://www1.cs.columbia.edu/CAVE/databases/multispectral/>.

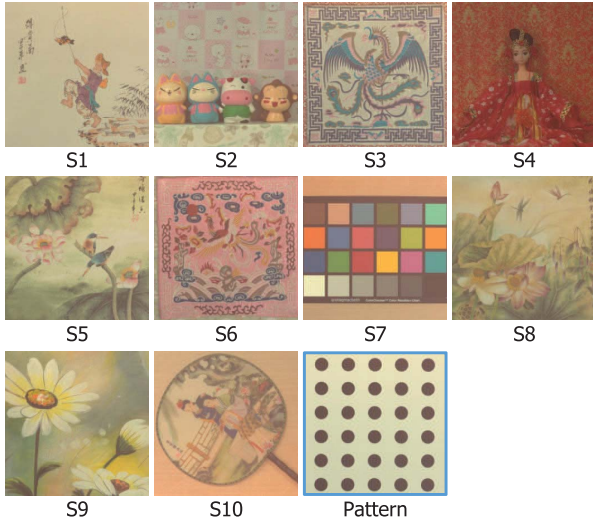


Fig. 7. Ten sample images (S1, S2, ..., S10) and a pattern image (displayed in RGB) used in experiments. The pattern image is used to compute the ground truth transform.



Fig. 8. Sample images (displayed in RGB) in the CAVE database.

As PC [22] does not have a specific measure form, its image registration algorithm is individually implemented. We compute image edges according to [23] and implement PC using the function `imregcorr` in Matlab R2016a. Our exploration indicates that the original RSNCC [14] suffers the overlap problem, i.e., its value reduces when overlap region shrinks. Hence we revise RSNCC by normalizing it with the size of overlap region to increase its robustness.

In the experiments, different measures are evaluated using both synthetic and real data. In the synthetic data experiment, we use similarity transform as the simulated deformation so that PC [22] can be quantitatively evaluated. In the real data experiment, we use affine transform to approximate the deformation in our captured multispectral images and investigate the registration accuracy. Similarity transform is assumed when applying PC [22]. The NTG-based image registration method is additionally evaluated in a practical circumstance that the multispectral imaging system suffers out-of-focus blur.

#### A. Synthetic Data

The synthetic data experiment is first conducted on the S2 multispectral images in Fig. 7 using simulation. The band image at 400 nm is used as the reference image, and the band images at 480 nm, 560 nm, 620 nm, and 700 nm are used

as the floating ones. As observed in the top left of Fig. 9, there are obvious local contrast inconsistency in the band images. To further simulate local intensity variation, we add dark clouds on the floating images (see bottom left of Fig. 9). The dark cloud is generated by Gaussian model with standard deviation  $\sigma$ , as described in Appendix B. The synthetic transforms imposed on floating images include spatial translation and rotation/scaling. To evaluate different measures on severe image deformation, the translations  $\Delta x$  and  $\Delta y$  are set in the range of  $[-20, 20]$  pixels. The rotation angle  $\theta \in [-10^\circ, 10^\circ]$  and the scaling factor  $s \in [0.5, 1.5]$ .

Fig. 9 shows the cost maps of different measures in the cases of translation and rotation/scaling transforms.<sup>2</sup> The measure values are first normalized to range  $[0, 1]$  and then color encoded for better comparison. For MI and CR, the quantities  $1/\text{MI}$  and  $1/\text{CR}$  are used as their corresponding measures to keep the consistency of all measures (reaching minimum when aligned). It is observed that MI [12], RC [21], and CR [11] are more or less influenced by contrast inconsistency and slow intensity variation. This observation is expected, as these measures are computed on image intensities that are inconsistent among band images. In comparison, the revised RSNCC and proposed NTG measures produce the correct transforms. Their robustness to local intensity variation is mainly due to the incorporation of gradient operations.

We further evaluate the robustness of different measures by increasing the intensity and contrast inconsistency between band images. When the standard deviation  $\sigma$  increases the dark cloud becomes larger, and consequently the contrast inconsistency and local intensity variation become severer. Fig. 10 shows the registration error curves of different measures with respect to  $\sigma$ . The registration error is computed as the mean radial distance between the pixel positions of the reference image and those of the clouded floating images after registration. It is observed that both RSNCC [14] and NTG achieve the highest registration accuracy and perform constantly stable in case of various dark cloud levels.

The measures are also extensively evaluated on the CAVE database. For all scenes, the band images at 560 nm are used as the reference images. The floating images are generated from the 961 band images (31 scenes  $\times$  31 bands) by imposing similarity transforms with known specific parameters. The reference and floating images are then aligned using PC [22] and different measures. The evaluation is conducted on medium displacement ( $s = 1.1$ ,  $\theta = 10^\circ$ ,  $\Delta x = -10$ ,  $\Delta y = 10$ ) and large displacement ( $s = 1.25$ ,  $\theta = 30^\circ$ ,  $\Delta x = -20$ ,  $\Delta y = 20$ ).

Fig. 11 shows the registration errors of the 961 band images in the cases of medium and large displacements. For better visual comparison, the registration errors of each measure are plotted in ascending order. The registration beyond 3 pixel error is regarded as failure in the experiments. In case of medium displacement, the revised RSNCC [14], PC [22], and NTG run successfully on more than 89% images. NTG produces the lowest registration error (0.17 pixels), followed by RSNCC (0.29 pixels) and PC (0.47 pixels). In case of large

<sup>2</sup>The PC algorithm does not have a measure form, and thus the cost map cannot be plotted.

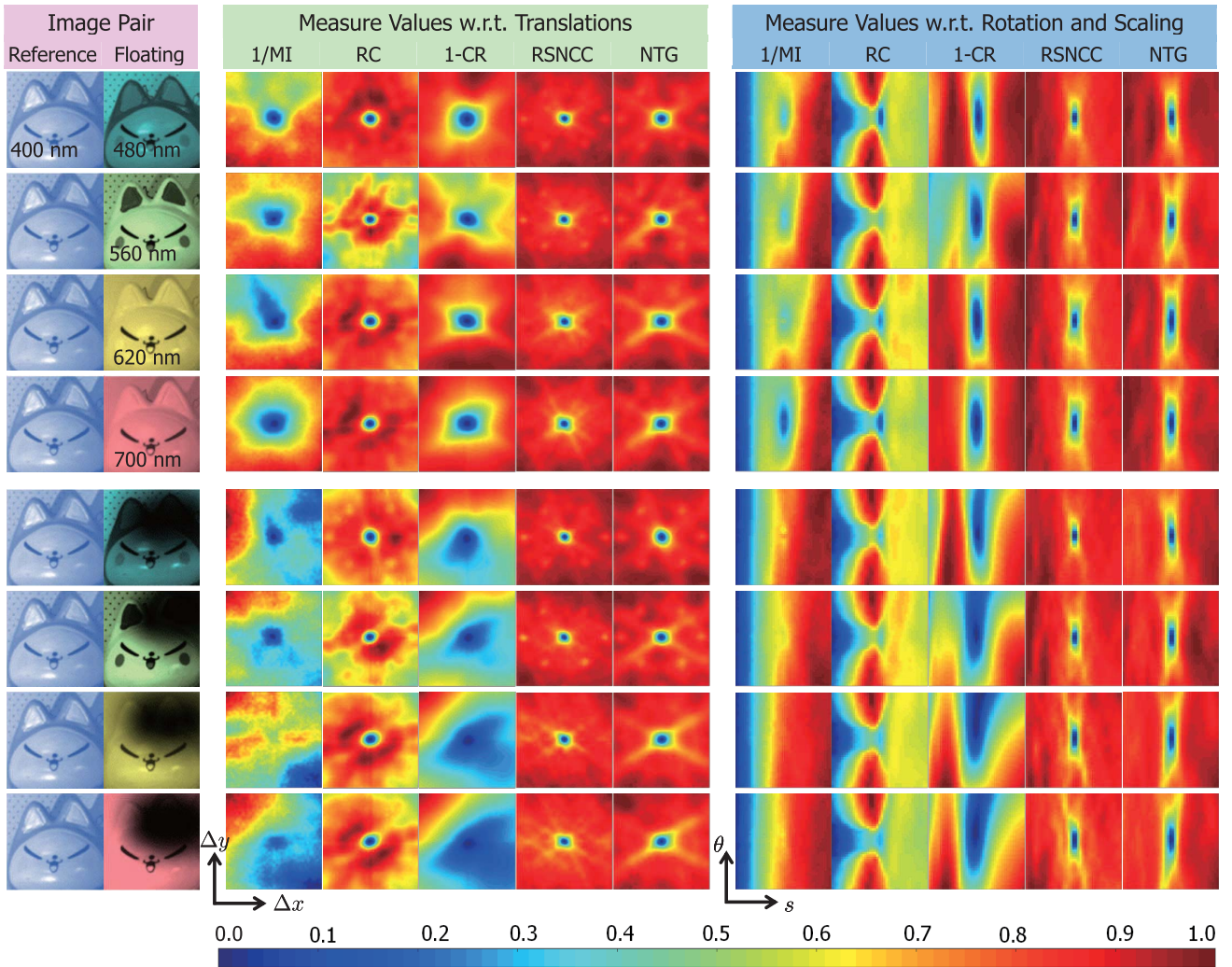


Fig. 9. Cost maps of the MI [12], RC [21], CR [11], RSNCC [14], and proposed NTG measures. In translation transforms, the measures are computed with respect to translations  $\Delta x$  and  $\Delta y$ . In rotation/scaling transforms, the measures are computed with respect to rotation angle  $\theta$  and scaling factor  $s$ . The measure values are color encoded according to the color bar in the normalized range [0, 1]. The ground truth translation is (0, 0) pixels and the ground truth rotation/scaling is ( $0^\circ$ , 1.0). These ground truth transforms locate at the centers of the cost maps. See text for details.

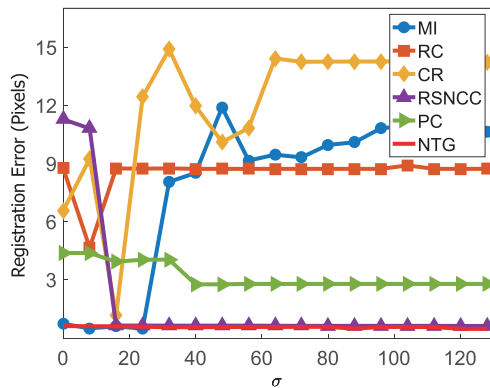


Fig. 10. Registration error curves of the MI [12], RC [21], CR [11], RSNCC [14], PC [22], and proposed NTG measures with respect to the standard deviation  $\sigma$  of dark cloud. The reference and floating images are the band images of S2 at 400 nm and 560 nm, respectively.

displacement, the registration error of NTG is also the lowest. The above evaluation on the CAVE database verifies that the NTG measure is in general superior to the competitors.

### B. Real Data

We capture the multispectral images of 10 real scenes and a pattern board (see Fig. 7) using our imaging system. The imaging condition is kept fixed and the ground truth transform parameters are obtained from the pattern image by calibration. The transform parameters are also individually estimated from the sample images ( $S_1, S_2, \dots, S_{10}$ ). Then the ground truth parameters and estimated parameters are applied on the centers of the circles in the pattern band images. The mean radial displacement between corresponding remapped circle centers is regarded as the registration error. To conduct a more accurate registration, in the experiment we estimate the parameters of affine transform (or similarity transform in case of PC [22]). Table I lists the median registration error (in pixels) of the 10 samples using different registration measures. It is observed that NTG yields subpixel accuracy and performs better than the competitors.

As a detailed example, Fig. 12 shows the registration errors of the sample  $S_1$  and  $S_2$  along different bands. The registration errors by the proposed NTG measure are generally smaller

TABLE I  
MEDIAN REGISTRATION ERRORS USING DIFFERENT MEASURES (IN PIXELS). THE SMALLEST ERRORS ARE IN BOLD

Band No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
MI [12]	1.15	0.68	0.66	0.61	0.61	0.51	0.58	0.56	-	0.68	0.58	0.66	0.61	0.85	1.47	1.27
RC [21]	2.12	1.34	1.24	1.26	1.17	0.75	0.52	0.58	-	0.99	0.81	0.94	1.23	0.81	1.05	1.02
CR [11]	3.50	1.81	1.95	1.47	1.22	0.82	0.62	0.48	-	0.79	0.60	0.98	0.82	1.20	2.12	2.98
RSNCC [14]	<b>0.38</b>	0.90	0.92	1.28	1.49	1.65	1.89	1.89	-	1.74	1.20	0.85	1.30	0.90	1.56	<b>0.43</b>
PC [22]	0.63	0.48	1.01	0.51	0.44	<b>0.30</b>	0.33	0.27	-	1.03	0.85	0.52	0.41	0.42	1.19	1.00
Proposed NTG	0.80	<b>0.37</b>	<b>0.43</b>	<b>0.32</b>	<b>0.26</b>	0.35	<b>0.23</b>	<b>0.24</b>	-	<b>0.36</b>	<b>0.32</b>	<b>0.33</b>	<b>0.34</b>	<b>0.27</b>	<b>0.95</b>	0.45

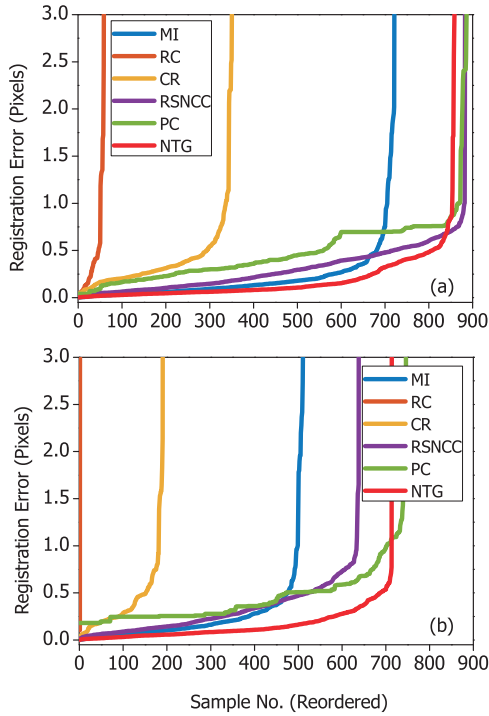


Fig. 11. Registration error curves of the MI [12], RC [21], CR [11], RSNCC [14], PC [22], and proposed NTG measures on the CAVE database. (a) Medium displacement. (b) Large displacement. The sample numbers of each measure are individually reordered with respect to the registration errors produced by that measure.

than others. The close-up views of image registration results of the sample S1 and S2 are further shown in Fig. 13, where the floating image is at band 420 nm and the reference image is at band 560 nm. The image pairs before and after registration are displayed in composite RGB. It is observed that the proposed NTG measure always produce high registration accuracy. For the competitors, small misalignment exists in sample S1 while large misalignment occurs in sample S2.

### C. Robustness to Blur

It is known that the filter wheel based multispectral imaging system suffers from out-of-focus blur, which is caused by the varying effective focal lengths in different bands [5]. For each band image, the out-of-focus blur can be modeled using a Gaussian kernel [6]. To explore the robustness of the

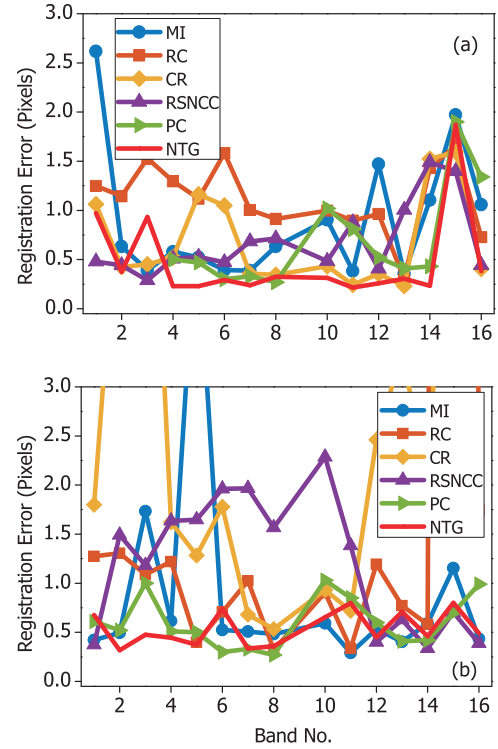


Fig. 12. Registration errors (in pixels) produced by different measures on (a) Sample S1 and (b) Sample S2.

proposed registration framework with respect to image blur, we manually blur the captured multispectral images in Fig. 7 using Gaussian kernels, whose standard deviations  $\sigma$  (blur levels) range from 0 to 4.75 pixels. Without loss of generality, we keep the band image at 560 nm sharp and use it as the reference image. We then estimate the transform parameters from the blurred images, and compute registration errors using the procedure mentioned above.

The distribution of mean registration error, which is computed from all band images of the ten sample images using the NTG-based registration method, is plotted in Fig. 15. It is observed that the registration errors are around 0.4 pixels, and become only slightly larger when increasing the blur levels. This excellent attribute is due to the image pyramid employed in the proposed registration framework. In the lower layers, the image blur is reduced by down sampling and thus the transform parameters can still be reliably estimated.

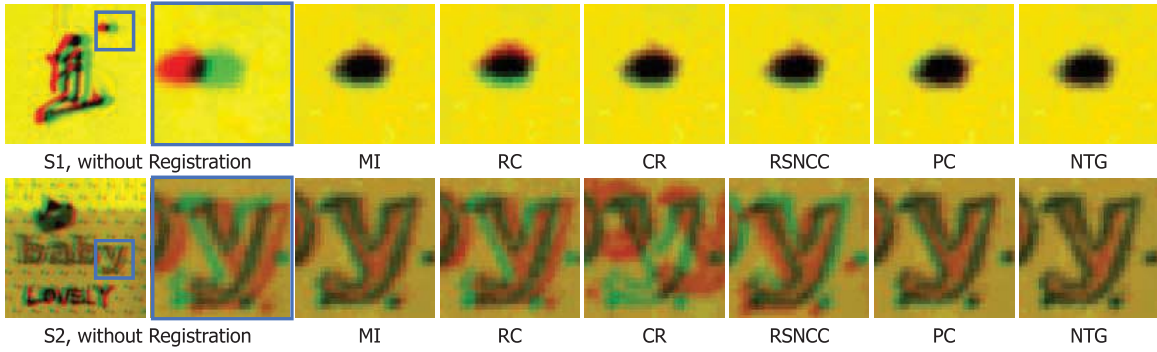


Fig. 13. Close-up views of multispectral image registration results of samples S1 and S2 using MI [12], RC [21], CR [11], RSNCC [14], PC [22], and the proposed NTG. Blending results of the floating image (band 420 nm) and the reference image (band 560 nm) are displayed in RGB.

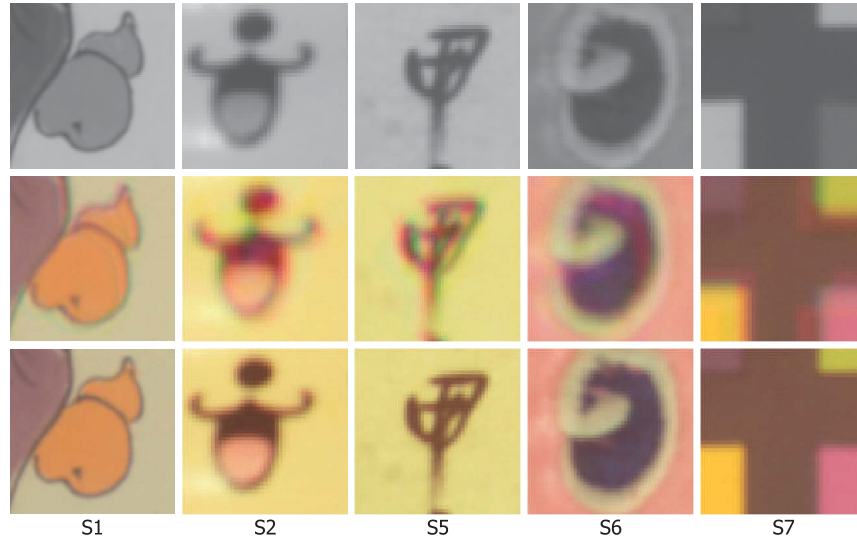


Fig. 14. Close-up views of multispectral registration results using the NTG-based framework. First row: reference images (band No. 9). Second row: original multispectral images. Third row: multispectral images after registration. Multispectral images are displayed in RGB for visualization.

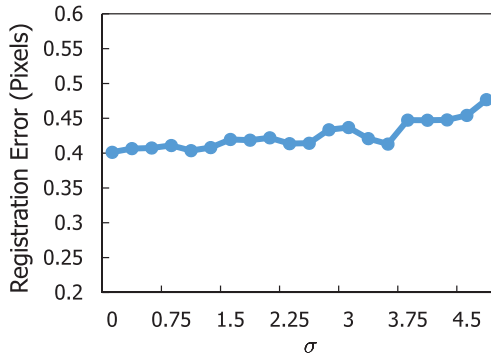


Fig. 15. Registration errors with respect to blur levels. The errors are computed from all band images of the ten samples in Fig. 7 using the NTG-based registration method.

By transferring the parameters from the lower layers to the upper layers, image registration can keep its robustness to the out-of-focus blur. The effectiveness of similar strategy has also been validated in recent works such as patch matching [32] and image deblurring [33].

## VI. APPLICATIONS

The multispectral images captured by filter wheel based imaging systems may suffer from both misalignment and

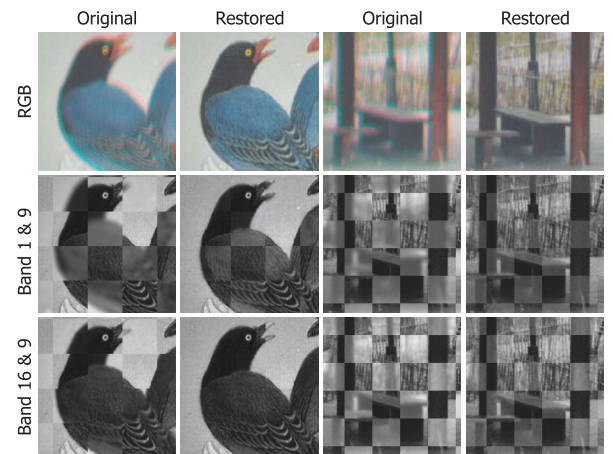


Fig. 16. Multispectral image restoration (registration and deblurring) results. Multispectral images are displayed in RGB. Blending results of the reference image (No. 9) and two blurriest band images (No. 1 and No. 16) are displayed in composited chessboards.

out-of-focus blur. We discuss the application of the proposed method in two circumstances. In the first circumstance (Section VI-A) the multispectral images are well

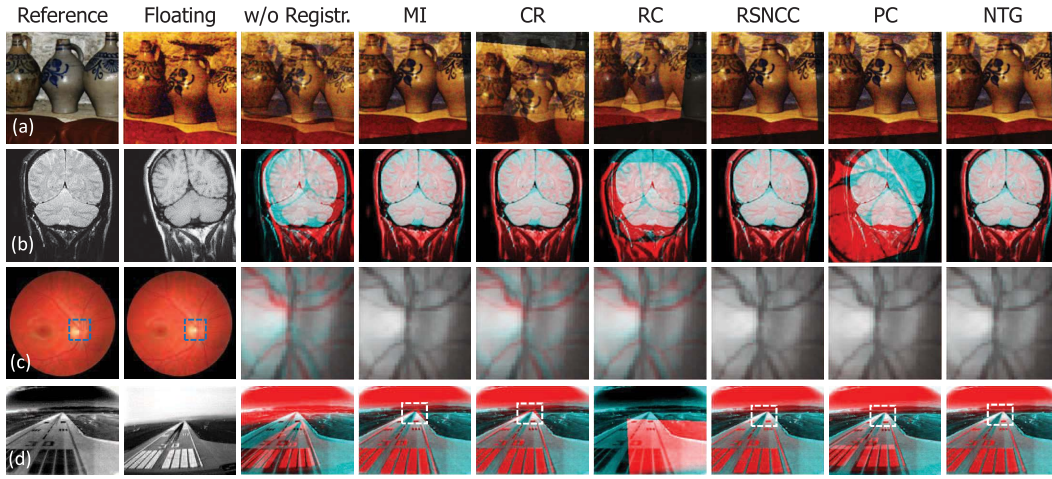


Fig. 17. Other registration tasks using MI [12], RC [21], CR [11], RSNCC [14], PC [22], and the proposed NTG. (a) Flash/no-flash image pair. (b) MR brain image pair. (c) Retinal image pair. For better comparison only the close-up views of registration results in blue boxes are displayed. (d) The multi-sensor (infrared and visible) image pair. The white boxes highlight the details for comparison.

focused by using a focusing device [5] and image registration is applied on sharp band images. In the second circumstance (Section VI-B) no focusing device is employed and thus image restoration must deal with both misalignment and out-of-blur. Section VI-C further shows that, in addition to multispectral images, the proposed method is applicable to multimodal and other image registration tasks.

#### A. Multispectral Image Registration

Fig. 14 shows the registration results produced by the proposed framework. The chromatic aberration caused by image misalignment is obvious in the original acquired multispectral images. After registration, the band images are well aligned and accordingly the chromatic aberration in the RGB images is eliminated.

#### B. Multispectral Image Restoration

In multispectral restoration, our aim is to remove both misalignment and out-of-focus blur. As discussed in Section V-C, the proposed framework is robust to image blur, hence image registration is directly applied on the original captured multispectral images without any preprocessing. After image registration, sharp band images can be further obtained using the recent multispectral image deblurring algorithm [6]. The multispectral image restoration results on real multispectral images are shown in Fig. 16. For illustration, the two blurriest images (band No. 1 and No. 16) and the reference image (band No. 9) are displayed in a form of composite chessboard. It is observed that the quality of the multispectral image (displayed in RGB) are greatly improved after image restoration.

#### C. Other Registration Tasks

Though designed for multispectral images, the proposed framework is also applicable to other multimodal/unimodal registration tasks. The registration results of flash/no-flash image pair, medical image pair, and multi-sensor image

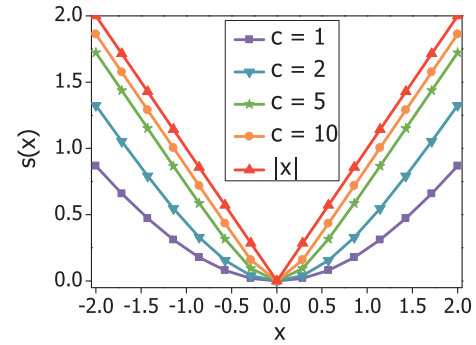


Fig. 18. Function  $s(x)$  with respect to different values of  $c$ . The value of  $x$  is in the range  $[-2, 2]$ . The absolute function  $|x|$  can be well approximated by  $s(x)$  when  $c = 10$ .

pair are shown in Fig. 17. The flash/no-flash image pair in Fig. 17(a) are available in [34]. The floating image is generated by imposing a  $5^\circ$  rotation on the no-flash image. The MR brain image pair in Fig. 17(b) are provided by the Center for Morphometric Analysis at Massachusetts General Hospital.<sup>3</sup> The floating image is the T1 image rotated by  $-5^\circ$ . The retinal image pair in Fig. 17(c) is available in [35], and the multi-sensor image pair in Fig. 17(d) is available in [10].

Fig. 17 shows that the MI [12], RSNCC [14], and NTG based registration methods achieve plausible alignment results on all image pairs. The RSNCC [14] and proposed NTG measures produce better alignment results on the multi-sensor image pair (Fig. 17(d)) when compared with the MI measure [12]. PC [22] achieves good registration results on the flash/no-flash and retinal image pairs, but fails on multimodal image pairs. CR [11] does not produce satisfactory results on the flash/no-flash and retinal image pairs. RC [21], as a unimodal registration measure, only succeeds on retinal image pair but the result is still inferior to RSNCC [14], PC [22], and NTG. In summary, the proposed NTG measure performs better

<sup>3</sup>Available at <http://www.cma.mgh.harvard.edu/ibsr/>

than the conventional MI [12], RC [21], CR [11] measures and PC [22] method, and performs comparatively to the most recent RSNCC [14] on these image pairs.

## VII. CONCLUSIONS

This paper proposes a new measure, namely normalized total gradient (NTG), for multispectral image registration. The employment of NTG is based on the observation that the gradient of difference image is the sparsest when two images are perfectly aligned. A registration framework, which consists of image pyramid and global/local optimization, is introduced for affine transform. Experimental results validate that the proposed NTG measure is well suited to multispectral image registration and outperforms its competitors. As an extension, the NTG can also be applied in other multimodal/unimodal registration tasks.

There are two limitations in this work. First, the NTG measure may fail in some special cases like periodic pattern images. Fortunately, it is seldom that a real image contains purely periodic patterns, and hence the proposed NTG measure works in most cases. Second, the proposed method is only implemented for affine transform, which is only applicable to static and rigid objects. In the future we plan to incorporate the NTG measure into nonrigid registration frameworks to further extend its application.

## APPENDIX A

### DETAILS OF NEWTON'S METHOD

To minimize the objective function  $J(\mathbf{p})$  in (10) we first denote  $m(\mathbf{p})$  and  $n(\mathbf{p})$  as the numerator and denominator of  $J(\mathbf{p})$ , respectively, as follows,

$$\begin{aligned} m(\mathbf{p}) &= \sum_{\mathbf{x} \in \Omega(\mathbf{p})} (|g_x(\mathbf{x}, \mathbf{p}) - f_{R,x}(\mathbf{x})| + |g_y(\mathbf{x}, \mathbf{p}) - f_{R,y}(\mathbf{x})|), \\ n(\mathbf{p}) &= \sum_{\mathbf{u} \in \Omega(\mathbf{p})} (|g_x(\mathbf{x}, \mathbf{p})| + |g_y(\mathbf{x}, \mathbf{p})| + |f_{R,x}(\mathbf{x})| + |f_{R,y}(\mathbf{x})|), \end{aligned}$$

where the subscripts  $x$  and  $y$  denote the partial derivatives along the  $x$  and  $y$  directions. With these notations, the objective function (10) is represented as  $J(\mathbf{p}) = \frac{m(\mathbf{p})}{n(\mathbf{p})}$ . By taking the derivative with respect to  $\mathbf{p}$  we have

$$J_{\mathbf{p}} = n^{-2}(nm_{\mathbf{p}} - mn_{\mathbf{p}}), \quad (12)$$

where  $m_{\mathbf{p}}$  and  $n_{\mathbf{p}}$  denote the gradient of  $m$  and  $n$  with respect to  $\mathbf{p}$ . Note that in (12) the parameter  $\mathbf{p}$  in parenthesis is omitted for notation simplification. The  $(i, j)$ th entry in Hessian matrix  $\mathbf{H}_{\mathbf{p}}^J$  is

$$\begin{aligned} \mathbf{H}_{\mathbf{p}}^J(i, j) &= J_{p_i, p_j} = n^{-2}(nm_{p_i, p_j} - mn_{p_i, p_j} + 2mn^{-1}n_{p_i}n_{p_j} \\ &\quad - m_{p_i}n_{p_j} - n_{p_i}m_{p_j}). \end{aligned}$$

Hence the Hessian matrix  $\mathbf{H}_{\mathbf{p}}^J$  becomes

$$\begin{aligned} \mathbf{H}_{\mathbf{p}}^J &= n^{-2}(n\mathbf{H}_{\mathbf{p}}^m - m\mathbf{H}_{\mathbf{p}}^n + 2mn^{-1}n_{\mathbf{p}}n_{\mathbf{p}}^T \\ &\quad - m_{\mathbf{p}}n_{\mathbf{p}}^T - n_{\mathbf{p}}m_{\mathbf{p}}^T), \quad (13) \end{aligned}$$

where  $\mathbf{H}_{\mathbf{p}}^m$  and  $\mathbf{H}_{\mathbf{p}}^n$  denote the Hessian matrices of  $m$  and  $n$  with respect to  $\mathbf{p}$ .

To this end, in order to derive the gradient  $J_{\mathbf{p}}$  and the Hessian matrix  $\mathbf{H}_{\mathbf{p}}^J$  we resort to computing  $m_{\mathbf{p}}$ ,  $n_{\mathbf{p}}$ ,  $\mathbf{H}_{\mathbf{p}}^m$ , and  $\mathbf{H}_{\mathbf{p}}^n$ , which will be elaborated below.

### A. Computing $m_{\mathbf{p}}$ and $n_{\mathbf{p}}$

Let  $d(\mathbf{x}, \mathbf{p})$  denote the difference image, i.e.,

$$d(\mathbf{x}, \mathbf{p}) = g(\mathbf{x}, \mathbf{p}) - f_R(\mathbf{x}),$$

then it is easy to derive that

$$\begin{aligned} d_x &= g_x - f_{R,x}, \quad d_y = g_y - f_{R,y}, \\ d_{x,\mathbf{p}} &= g_{x,\mathbf{p}}, \quad d_{y,\mathbf{p}} = g_{y,\mathbf{p}}. \end{aligned} \quad (14)$$

Let  $\rho_1(\cdot)$  denote the first order derivative of the absolute function  $|\cdot|$  (see Appendix A-C). Since the overlap changes slightly in the two consecutive updates of  $\mathbf{p}$ , i.e.  $\Omega_{\mathbf{p}}(\mathbf{p}') \approx \mathbf{0}$ , by taking the derivatives of  $m$  and  $n$  with respect to  $\mathbf{p}$ , we have

$$m_{\mathbf{p}} \approx \sum_{\mathbf{x} \in \Omega(\mathbf{p})} (\rho_1(d_x)g_{x,\mathbf{p}} + \rho_1(d_y)g_{y,\mathbf{p}}) \quad (15)$$

and

$$n_{\mathbf{p}} \approx \sum_{\mathbf{x} \in \Omega(\mathbf{p})} (\rho_1(g_x)g_{x,\mathbf{p}} + \rho_1(g_y)g_{y,\mathbf{p}}). \quad (16)$$

We narrow our focus down to the computation of  $g_{x,\mathbf{p}}$  and  $g_{y,\mathbf{p}}$  since all the other terms are known given the current transform parameter  $\mathbf{p}'$ . By implementing  $g(\mathbf{x}, \mathbf{p}) = f(u(\mathbf{x}, \mathbf{p}), v(\mathbf{x}, \mathbf{p}))$ , we have

$$\begin{pmatrix} g_x \\ g_y \end{pmatrix} = \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} \begin{pmatrix} f_u \\ f_v \end{pmatrix}.$$

Consequently it can be derived that

$$\begin{aligned} \begin{pmatrix} g_{x,p_i} \\ g_{y,p_i} \end{pmatrix} &= \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} \begin{pmatrix} f_{uu} & f_{uv} \\ f_{vu} & f_{vv} \end{pmatrix} \begin{pmatrix} u_{p_i} \\ v_{p_i} \end{pmatrix} \\ &\quad + \begin{pmatrix} u_{x,p_i} & v_{x,p_i} \\ u_{y,p_i} & v_{y,p_i} \end{pmatrix} \begin{pmatrix} f_u \\ f_v \end{pmatrix}, \quad (17) \end{aligned}$$

where  $p_i \in \{p_1, p_2, \dots, p_6\}$ . In (17), the gradient of the objective function is finally decomposed to the transformed first order and second order gradients of the floating image, and the gradients of the transform model.

According to the affine transform model (8), we have

$$\begin{aligned} \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} &= \begin{pmatrix} p_1 & p_4 \\ p_2 & p_5 \end{pmatrix}, \\ u_{\mathbf{p}} &= (x, y, 1, 0, 0, 0)^T, \quad v_{\mathbf{p}} = (0, 0, 0, x, y, 1)^T, \\ u_{x,\mathbf{p}} &= (1, 0, 0, 0, 0, 0)^T, \quad u_{y,\mathbf{p}} = (0, 1, 0, 0, 0, 0)^T, \\ v_{x,\mathbf{p}} &= (0, 0, 0, 1, 0, 0)^T, \quad v_{y,\mathbf{p}} = (0, 0, 0, 0, 1, 0)^T. \end{aligned}$$

To this end, the gradient of the objective function can be computed by inserting (15), (16), and (17) into (12).

### B. Computing $\mathbf{H}_p^m$ and $\mathbf{H}_p^n$

Let  $\rho_2(\cdot)$  be the second order derivative of the absolute function  $|\cdot|$ . By using (14) and (15), the  $(i, j)$ th entry in  $\mathbf{H}_p^m$  can be computed as

$$\mathbf{H}_p^m(i, j) = m_{p_i, p_j} \approx \sum_{\mathbf{x} \in \Omega(\mathbf{p})} (\rho_1(d_x)g_{x, p_i, p_j} + \rho_2(d_x)g_{x, p_i}g_{x, p_j} + \rho_1(d_y)g_{y, p_i, p_j} + \rho_2(d_y)g_{y, p_i}g_{y, p_j}).$$

Hence the Hessian matrix  $\mathbf{H}_p^m$  becomes

$$\mathbf{H}_p^m \approx \sum_{\mathbf{x} \in \Omega(\mathbf{p})} (\rho_1(d_x)\mathbf{H}_p^{g_x} + \rho_2(d_x)g_{x, \mathbf{p}}g_{x, \mathbf{p}}^T \rho_1(d_y)\mathbf{H}_p^{g_y} + \rho_2(d_y)g_{y, \mathbf{p}}g_{y, \mathbf{p}}^T), \quad (18)$$

where  $\mathbf{H}_p^{g_x}$  and  $\mathbf{H}_p^{g_y}$  are the Hessian matrices of  $g_x$  and  $g_y$ .

Similarly, we have

$$\mathbf{H}_p^n(i, j) = n_{p_i, p_j} \approx \sum_{\mathbf{x} \in \Omega(\mathbf{p})} (\rho_1(g_x)g_{x, p_i, p_j} + \rho_2(g_x)g_{x, p_i}g_{x, p_j} + \rho_1(g_y)g_{y, p_i, p_j} + \rho_2(g_y)g_{y, p_i}g_{y, p_j}),$$

and thus

$$\mathbf{H}_p^n \approx \sum_{\mathbf{x} \in \Omega(\mathbf{p})} (\rho_1(g_x)\mathbf{H}_p^{g_x} + \rho_2(g_x)g_{x, \mathbf{p}}g_{x, \mathbf{p}}^T \rho_1(g_y)\mathbf{H}_p^{g_y} + \rho_2(g_y)g_{y, \mathbf{p}}g_{y, \mathbf{p}}^T). \quad (19)$$

According to the affine model given in (17), we can derive that

$$\begin{pmatrix} g_x & p_i & p_j \\ g_y & p_i & p_j \end{pmatrix} = \begin{pmatrix} u_{x, p_j} & v_{x, p_j} \\ u_{y, p_j} & v_{y, p_j} \end{pmatrix} \begin{pmatrix} f_{uu} & f_{uv} \\ f_{vu} & f_{vv} \end{pmatrix} \begin{pmatrix} u_{p_i} \\ v_{p_i} \end{pmatrix} + \begin{pmatrix} u_x & v_x \\ u_y & v_y \end{pmatrix} \begin{pmatrix} f_{uu, p_j} & f_{uv, p_j} \\ f_{vu, p_j} & f_{vv, p_j} \end{pmatrix} \begin{pmatrix} u_{p_i} \\ v_{p_i} \end{pmatrix} + \begin{pmatrix} u_{x, p_i} & v_{x, p_i} \\ u_{y, p_i} & v_{y, p_i} \end{pmatrix} \begin{pmatrix} f_u & p_j \\ f_v & p_j \end{pmatrix}, \quad (20)$$

where

$$\begin{pmatrix} f_u & p_j \\ f_v & p_j \end{pmatrix} = \begin{pmatrix} f_{uu} & f_{uv} \\ f_{vu} & f_{vv} \end{pmatrix} \begin{pmatrix} u_{p_j} \\ v_{p_j} \end{pmatrix},$$

$$\begin{pmatrix} f_{uu} & p_j \\ f_{vu} & p_j \end{pmatrix} = \begin{pmatrix} f_{uuu} & f_{uuv} \\ f_{vu} & f_{vv} \end{pmatrix} \begin{pmatrix} u_{p_j} \\ v_{p_j} \end{pmatrix},$$

and

$$\begin{pmatrix} f_{uv} & p_j \\ f_{vv} & p_j \end{pmatrix} = \begin{pmatrix} f_{uvu} & f_{uvv} \\ f_{vvu} & f_{vvv} \end{pmatrix} \begin{pmatrix} u_{p_j} \\ v_{p_j} \end{pmatrix}.$$

It is clear that the Hessian matrix of the objective function can finally decomposed to the transformed gradients of the floating image up to third orders. To this end, the Hessian matrix can be computed by inserting (18), (19), and (20) into (13).

### C. Form of $\rho_1$ and $\rho_2$

Since the absolute function  $|x|$  is not differentiable at  $x = 0$ , we use the function

$$s(x) = x + \frac{2}{c} \log\left(\frac{1 + e^{-cx}}{2}\right) \quad (21)$$

to approximate it. The function  $s(x)$  with respect to different values of  $c$  is illustrated. The value of  $x$ , which is the gradient

of difference image, is in the range  $[-2, 2]$ . As illustrated,  $s(x)$  gradually approximates  $|x|$  as  $c$  increases. Theoretically larger  $c$  values will result in a better approximation. Our investigation indicates that the value  $c = 10$  suffices for NTG computation. The first and second derivatives of  $s(x)$  are, respectively,

$$\rho_1(x) = \frac{1 - e^{-cx}}{1 + e^{-cx}} \quad (22)$$

and

$$\rho_2(x) = \frac{2c}{e^{-cx} + e^{cx} + 2}. \quad (23)$$

## APPENDIX B DARK CLOUD GENERATION

When added by dark cloud, the intensity of floating image  $f(x, y)$  at spatial coordinate  $(x, y)$  is changed to

$$f'(x, y) = (1 - \gamma(x, y))f(x, y) + \gamma(x, y)f_{\min}.$$

Here,  $f_{\min}$  is the minimal intensity of image  $f(x, y)$ , and  $\gamma(x, y)$  is the Gaussian image defined as

$$\gamma(x, y) = \exp\left(-\frac{(x - x_c)^2 + (y - y_c)^2}{2\sigma^2}\right),$$

where  $(x_c, y_c)$  denotes the center of the Gaussian image, and  $\sigma$  denotes the standard deviation.

## REFERENCES

- [1] J. Y. Hardeberg, F. Schmitt, and H. Brettel, "Multispectral color image capture using a liquid crystal tunable filter," *Opt. Eng.*, vol. 41, no. 10, pp. 2532–2548, 2002.
- [2] J. Katrašnik, F. Pernuš, and B. Likar, "Radiometric calibration and noise estimation of acousto-optic tunable filter hyperspectral imaging systems," *Appl. Opt.*, vol. 52, no. 15, pp. 3526–3537, 2013.
- [3] J. Brauers and T. Aach, "Geometric calibration of lens and filter distortions for multispectral filter-wheel cameras," *IEEE Trans. Image Process.*, vol. 20, no. 2, pp. 496–505, Feb. 2011.
- [4] J. Brauers, N. Schulte, and T. Aach, "Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2368–2380, Dec. 2008.
- [5] H.-L. Shen, Z.-H. Zheng, W. Wang, X. Du, S.-J. Shao, and J. H. Xin, "Autofocus for multispectral camera using focus symmetry," *Appl. Opt.*, vol. 51, no. 14, pp. 2616–2623, 2012.
- [6] S.-J. Chen and H.-L. Shen, "Multispectral image out-of-focus deblurring using interchannel correlation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4433–4445, Nov. 2015.
- [7] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, no. 1, pp. 1–36, Mar. 1998.
- [8] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, pp. 977–1000, Oct. 2003.
- [9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [10] M. Irani and P. Anandan, "Robust multi-sensor image alignment," in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 959–966.
- [11] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *Proc. 1st Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 1496, Cambridge, MA, USA, 1998, pp. 1115–1124.
- [12] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.
- [13] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
- [14] X. Shen, L. Xu, Q. Zhang, and J. Jia, "Multi-modal and multi-spectral registration for natural images," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 309–324.

- [15] K. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*. Heidelberg, Germany: Springer, 2006.
- [16] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, 1999.
- [17] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, "Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change," *IEEE Trans. Med. Imag.*, vol. 25, no. 5, pp. 626–639, May 2006.
- [18] J. Woo, M. Stone, and J. L. Prince, "Multimodal registration via mutual information incorporating geometric and spatial context," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 757–769, Feb. 2015.
- [19] H. Luan, F. Qi, Z. Xue, L. Chen, and D. Shen, "Multimodality image registration by maximization of quantitative–qualitative measure of mutual information," *Pattern Recognit.*, vol. 41, no. 1, pp. 285–298, 2008.
- [20] Y. Hel-Or, H. Hel-Or, and E. David, "Matching by tone mapping: Photometric invariant template matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 317–330, Feb. 2014.
- [21] A. Myronenko and X. Song, "Intensity-based image registration by minimizing residual complexity," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1882–1891, Nov. 2010.
- [22] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.
- [23] P. Nandy, B. N. Post, J. L. Smith, and M. L. Decker, "Edge-based correlation image registration algorithm for the multispectral thermal imager (MTI)," *Proc. SPIE*, vol. 5425, pp. 197–207, Aug. 2004.
- [24] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 787–794, 2006.
- [25] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [26] A. Levin, A. Zomet, and Y. Weiss, "Learning how to inpaint from global image statistics," in *Proc. Int. Conf. Comput. Vis.*, 2003, pp. 305–312.
- [27] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 193–200.
- [28] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.
- [29] 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. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.
- [30] J. Klein and T. Aach, "Multispectral filter wheel cameras: Modeling aberrations for filters in front of lens," *Proc. SPIE*, vol. 8299, p. 82990R, Jan. 2012.
- [31] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
- [32] O. Lotan and M. Irani, "Needle-match: Reliable patch matching under high uncertainty," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 439–448.
- [33] T. Michaeli and M. Irani, "Blind deblurring using internal patch recurrence," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 783–798.
- [34] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 664–672, Aug. 2004.
- [35] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "Retinal image registration through simultaneous camera pose and eye shape estimation," in *Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc.*, Aug. 2016, pp. 3247–3251.



**Shu-Jie Chen** received the B.Eng. degree in information engineering from Zhejiang University in 2012, where he is currently pursuing the Ph.D. degree with the College of Information Science and Electronic Engineering. His research interests are multispectral imaging and image processing.



**Hui-Liang Shen** received the B.Eng. and Ph.D. degrees in electronic engineering from Zhejiang University, China, in 1996 and 2002, respectively. He was a Research Associate and a Research Fellow with The Hong Kong Polytechnic University from 2001 to 2005. He is currently a Professor with the College of Information Science and Electronic Engineering, Zhejiang University. His research interests include multispectral color imaging, image processing, and 3D computer vision.



**Chunguang Li** (M'14–SM'14) received the M.S. degree in pattern recognition and intelligent systems and the Ph.D. degree in circuits and systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2002 and 2004, respectively.

He is currently a Professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. His current research interests include statistical signal processing, machine learning, wireless sensor network, and image processing.



**John H. Xin** received the Ph.D. degree from the University of Leeds, U.K., in 1989. He is a Chair Professor with the Institute of Textiles and Clothing, The Hong Kong Polytechnic University. His main research interests include color difference evaluation, color communication, multispectral color imaging, and coloration technology. He is a fellow and the Chartered Colorist of the Society of Dyers and Colourists, U.K.