Accurate Registration using Adaptive Block Processing for Multi-spectral Images

Ik-Hyun Lee, Student Member, IEEE, and Tae-Sun Choi, Senior Member, IEEE

Abstract—Image registration is a challenging task, with applications in surveillance, motion estimation, and fusion systems. Due to the diversity of sensors, local distortions and large image size, satellite images are often difficult to accurately register. In literature, local descriptor based processing techniques such as scale-invariant feature transform (SIFT) have been applied to register satellite images, which provide robust features. However, these techniques suffer from a high computational cost, lack of features and low distribution quality, which affect the registration accuracy. In this paper, we develop an algorithm to register satellite images based on adaptive block processing to increase the number of features and improve the distribution quality. In addition, outlier removal using statistical masks are associated with classical random sample consensus (RANSAC); a subsequent comparative analysis demonstrates the accuracy of the proposed method. Typically, a classical SIFT prevents its wide application in recent remote sensing, though this is no longer case with the proposed adaptive block processing method.

Index Terms—Multi-spectral images, registration, adaptive block searching, outlier removal.

I. INTRODUCTION

Image registration is the spatial alignment of corresponding images of the same scene acquired from different times, views, and/or sensors. In general, registration methods can be divided into two categories: 1) area-based methods and 2) feature-based methods. Area-based methods use the pixel intensity of corresponding region, in which the similarity measure is key factor for registration accuracy. In contrast, feature-based methods use points, curves, lines, branches, and regions. The critical point of feature-based methods is to extract correct correspondent features between two or more images. Normally, feature-based methods consist of four steps:

i. Feature extraction – Common, but distinctive objects are considered as features, such as points, edges, curves, lines, branches, and regions.
ii. Feature matching – Using features from the previous step, the common features between the reference image and the sensed image are matched. Registration accuracy depends on the correctness of feature matching. To improve the matching quality, outlier removal techniques are normally integrated.
iii. Transformation estimation – Parameters of the mapping function are estimated using the matched features, and the sensed image is then transformed using the estimated transformation.
iv. Image re-sampling – The transformed image is finally re-sampled using various interpolation methods.

Unlike medical images, satellite image registration using area-based methods are difficult to register accurately due to the large image size with local distortion [1]. Due to the limitation of area-based methods, satellite image registration commonly uses feature-based methods. Satellite image registration is an especially challenging task due to the local distortion and high computational complexity. A high-resolution satellite image can be several hundred megapixels in size and occupy several spectral bands. And though high-resolution images provide detailed information, it is not efficient to process the entire image due to limited resources such as storage, memory. High-resolution satellite images also contain local distortions due to different sensors having different paths, angles, and terrain relief; i.e., the number of feature points and distribution quality affect the accuracy. In general, corresponding feature points should be distributed throughout the image. In literature [2]–[5], the effect of feature point number was analyzed in terms of registration error and accuracy, in which a larger number of matching points was found to provide better results. These findings indicate that feature points should be well-distributed due to local geometric distortion.

Recently, a scale-invariant feature transform (SIFT) [6] was applied to satellite images based on non-block processing [7]–[11]. Note, however, that SIFT-based methods suffer from high complexity [12] when extracting feature points and computing descriptors. It also has drawbacks such as lack of the feature points and its distribution quality; to date, there have been few attempts to overcome these problems. Of possible methods, Ke and Sukthankar [13] proposed PCA-SIFT to compute descriptors in a similar manner to SIFT. To reduce the size of the feature space, a principal component analysis (PCA) was applied by choosing important features. Reduction of the feature space size is an important role in lowering the time complexity; PCA is a method used to reduce the dimensions by selecting only important features. In [13], 41 × 41 patches for the horizontal and vertical gradient derivatives were extracted and the 3042-dimensional vector was reduced to 36 using PCA. Speeded up robust features (SURF) [14] is a further solution as it allows for a very fast computation of detectors using integral images [15] and a Hessian matrix. However, both PCA-SIFT and SURF have a low accuracy and their...
processing times are only slightly improved compared to SIFT [16], [17]. As such, these methods are not fully suitable for high resolution satellite image registration. Yet, even though SIFT has weaknesses due to high computational complexity and distribution quality of features, it remains one of the best methods for satellite image registration due to the robustness of distinctive features. And though there have been some attempts to overcome these problems, they are not fully applicable for satellite image registration due to the variation of feature points and distribution.

To overcome these limitations, we propose an adaptive block processing algorithm based on a geostatistical analysis such as semi-variance [18] and semi-variograms [19], [20]. The proposed method adaptively determines the appropriate block size through a geostatistical SIFT (GSIFT), and provides more features and well-distribution than could be previously obtained. With respect to the registration accuracy, we also proposed an outlier removal algorithm using sign and statistical mask. It is simple, but it provides efficient and accurate result than classical methods such as a random sample consensus (RANSAC).

In the remainder of this paper, Section II discusses the motivation, and the proposed adaptive block processing and efficient outlier removal methods are explained in Section III. Section IV then presents the experimental results and a comparative analysis. Finally, Section V concludes this study.

II. MOTIVATION

Block based methods [10], [21] have been applied to complex satellite images using SIFT. Non-block based methods having a small image size are typically inefficient in remote sensing, as the complexity of satellites images has increased due to increases in high resolution imaging and thus contains a greater amount of information. To overcome this problem, block based methods for large image size have proposed that the entire image \( X \times Y \) be divided into smaller image blocks \( M \times N \), such that

\[
IB(x, y) = \left\{ (\xi, \zeta) : |\xi - x| \leq \frac{M}{2} \land |\zeta - y| \leq \frac{N}{2} \right\}.
\]

where \((\xi, \zeta)\) denotes a pixel within the image block \( M \times N \). Figure 1 shows the graphical representation of an image block. In literature, the block size is assumed to be square (i.e., \( M \) equals \( N \)). In addition, the number of sub-images are heuristically determined, such as \( 3 \times 3 \) [21], \( 5 \times 5 \) [10] and \( 8 \times 8 \) [22]. The major concern of these block based methods is the need to reduce the processing time, and there is less or a concern about the number of features and their distribution quality. The processing time, feature points, and their distribution quality using SIFT are affected as follows:

- **Detection step** – SIFT uses the scale-space within number of octaves and number of levels per octave to find local extrema of the difference of Gaussian (DoG). Within the process, Gaussian convolutions and down-sampling are repeated. Due to this process, a smaller image block incurs less computational complexity for convolution and sampling. Although a smaller image block provides a low processing time, it does not always provide feature points within octaves and levels. With approach to coarser octaves and levels, the image becomes smaller and more blurred; the coarse scaled features can thus be discarded.

- **Descriptor step** – The SIFT descriptor for an \( F_{num} \times 128 \) dimensional vector, where \( F_{num} \) is determined by the number of extracted features. In the case of sample 1 using a \( 100 \times 100 \) image patch size with the first block of multi-spectral image \( U_{100}\{1\}\{1\} = U_{100}(1 : 100, 1 : 100) \) is 43, whereas the \( 200 \times 200 \) image patch size with first block of a panchromatic image \( U_{200}\{1\}\{1\} = U_{200}(1 : 200, 1 : 200) \) is 167. In other words, the descriptor size of \( U_{100}\{1\}\{1\} \) is \( 128 \times 43 \), while \( U_{200}\{1\}\{1\} \) is \( 128 \times 167 \). Generally, a larger image has more information and more features can be extracted; a smaller image patch allows less dimensionality due to the number of extracted features. We confirm here that a smaller image patch allows a smaller descriptor.

- **Matching step** – NNDR is used for descriptor matching. A smaller descriptor dimension requires less comparison processing time for the first and second descriptors.

Table I presents a comparison between non-block and block based processing according to the processing time and feature points. Initial matching points were obtained using the nearest-

<table>
<thead>
<tr>
<th>Block size ((M \times N))</th>
<th>(3792 \times 4000)</th>
<th>(100 \times 100)</th>
<th>(150 \times 150)</th>
<th>(200 \times 200)</th>
<th>(250 \times 250)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial feature points</td>
<td>64605</td>
<td>61644</td>
<td>65134</td>
<td>67346</td>
<td>68304</td>
</tr>
<tr>
<td>Initial matching points</td>
<td>4441</td>
<td>5984</td>
<td>6707</td>
<td>6959</td>
<td>6849</td>
</tr>
<tr>
<td>Processing time (s)</td>
<td>899.974</td>
<td>54.1824</td>
<td>67346</td>
<td>71.5425</td>
<td>75.0336</td>
</tr>
<tr>
<td>Descriptor dimension</td>
<td>64603 (\times 128)</td>
<td>40 (\times 128)</td>
<td>96 (\times 128)</td>
<td>177 (\times 128)</td>
<td>284 (\times 128)</td>
</tr>
</tbody>
</table>

Fig. 1. Block-wise satellite image having \(X \times Y\) size and \(M \times N\) block size.

Copyright (c) 2013 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.
neighborhood distance ratio (NNDR)\(^2\) [16]. The processing time includes the period from feature extraction to feature matching, and the descriptor dimension is the average of each block obtained by block processing. The size of the processed image block affects the computational time and the number of feature points. From the table, non-block based processing is seen to be the most time consuming due to the high descriptor dimension; the entire image processing (non-block based) time is 899, 974 s, whereas the block based processing time ranges from 54.1824 s to 75.0336 s. Block based processing enables a dramatically reduced processing time, as it can be observed that the processing time for the smallest block size is almost as fast as for a larger block size. Although the larger block size processing generally provides more feature points and matching points, this is not always the case. The dependence of the block size can be observed in Fig. 2. The figure illustrates the difficulty and importance of determining an appropriate block size for the number of feature points and its distribution. Figure 2(a) shows non-block based processing, and the others are block based processing with 200 \(\times\) 200 (Fig. 2(b)), 250 \(\times\) 250 (Fig. 2(c)) block sizes. Note that even though the block sizes are similar in both cases, the number of correct matching points and its distribution are different. Processing the 200 \(\times\) 200 and 250 \(\times\) 250 blocks produces 13 and 7 points, respectively; there are fewer non-block based correct matching points than block based matching points. In addition, the 200 \(\times\) 200 block provides well-distributed matching points. In summary, the smaller block size has a smaller number of features, low dimensional descriptors, and low complexity for descriptor comparison for matching; i.e., block size plays an important role in determining the overall accuracy and processing time. Generally, a smaller block size can reduce the processing time, but may not provide accurate results due to the lack of feature points and its distribution. In some cases, a smaller block size does not provide any feature points due to the small correspondence region between the reference image and the sensed image.

III. PROPOSED METHOD

The main drawbacks with conventional methods are lack of the number of feature points and the feature points may not be well distributed, thereby affecting the registration accuracy. To improve the registration accuracy, feature points are extracted using a variable block size with SIFT; the SIFT descriptor is used for feature matching via the block based nearest-neighborhood distance ratio (NNDR) [16]. Next, outliers are removed using random sample consensus (RANSAC) [23] and the proposed statistical masks. Then, local transformations using B-spline [24] and image re-sampling using a bicubic interpolation [25] are performed. Figure 3 shows the flowchart of the proposed method.

A. Feature Extraction and Matching using Adaptive Block Processing

Conventional methods have difficulty automatically determining an appropriate block size. We propose here the determination of an appropriate block size based on a geostatistical analysis, based on the semi-variance; the semi-variance can be used to interpret the spatial relationship between image data based on the specific distance (separation distance), and as such can be used to measure the similarity between data. The semi-variogram is the plot of the semi-variance against the distance; in a typical semi-variogram, in which a spatial analysis, based on the semi-variance; the semi-variance can be used to interpret the spatial relationship between image data based on the specific distance (separation distance), and as such can be used to measure the similarity between data. The semi-variogram is the plot of the semi-variance against the distance; in a typical semi-variogram, in which a spatial relationship exists in the data, the semi-variance will increase as the lag between points increases. Figure 4 presents the semi-variogram. The nugget, denoted as \(C_0\), is the variance at zero distance; the range, denoted as \(a\), is the distance at which the semi-variogram level reaches a constant, and thereby indicates the distance at which the data are no longer correlated. And the sill, denoted as \(C_0 + C_1\), is the constant semi-variance value beyond the range. Here, the semi-variance defined as

\[
\gamma(l) = \frac{1}{2} \text{Var} \left[ z(x + l) - z(x) \right].
\]

(2)

where \(\gamma(l)\) can also be calculated using half of the squared difference of the value, i.e.,

\[
\gamma(l) = \frac{1}{2} \text{E} \left[ z(x + l) - z(x) \right]^2.
\]

(3)

---

2A detailed explanation of NNDR is found in Section III-A
Fig. 5. Semi-variance analysis: Semi-variograms of a (a) homogeneous area \((l = 297)\), and (b) heterogeneous area \((l = 281)\).

The semi-variance at distance \(l\) is then calculated by summing the squared differences of the values of all data, i.e.,

\[
\gamma(l) = \frac{1}{2} \left[ \sum_{i=1}^{p} [z(x_i) - z(x_i - l)]^2 \right],
\]

where \(p\) is the number of pairs. Note that this value is calculated separately for each \(l\) along a line centered on the point of investigation with a horizontal \(h\) and vertical line \(v\) as:

\[
\gamma^{\text{hor}}(h) = \frac{1}{2} \left[ \sum_{i=1}^{p} [IB_M(x_i) - IB_M(x_i - h)]^2 \right],
\]

\[
\gamma^{\text{ver}}(v) = \frac{1}{2} \left[ \sum_{i=1}^{p} [IB_N(x_i) - IB_N(x_i - v)]^2 \right].
\]

Since the semi-variance measures the similarity of pixels distributed in a block, a higher value indicates that the pixels are less related, whereas a lower value indicates that the pixels are more related. The horizontal and vertical block sizes \(M'\) and \(N'\) are determined from \(h\) and \(v\), which maximize \(\gamma^{\text{hor}}(h)\) and \(\gamma^{\text{ver}}(v)\) as:

\[
M' = \arg \max_h \left( \gamma^{\text{hor}}(h) \right)
\]

\[
N' = \arg \max_v \left( \gamma^{\text{ver}}(v) \right).
\]

Finally, the adaptive image block \(AIB(x, y)\) is

\[
AIB(x, y) = \{ (\xi, \zeta) | |\xi - x| \leq \frac{M'}{2} \land |\zeta - y| \leq \frac{N'}{2} \}.
\]

Figure 5 shows that the maximum criterion can be determined using the proposed algorithm. Figure 5(a) is a mountainous area, whereas Fig. 5(b) is an urban area. The urban area is seen to be heterogeneous (more pixel variation) due to the many objects, whereas the mountainous area is homogeneous (less pixel variation). In the area having a higher semi-variance, the homogeneous area is around 15,000, whereas the heterogeneous area is 2,400. The maximum semi-variance of the heterogeneous area is notably higher than for the homogeneous area, indicating that the area with more pixel variation has a higher semi-variance. In the area having a higher semi-variance (Fig. 5(b)), 808 more distinctive features are obtained due to the more pixel variation, whereas there were 559 features in Fig. 5(a); i.e., if the pixel variation is high, more distinctive features can be extracted. Figure 6 shows the various block sizes determined by adaptive block processing, where \(M'\) and \(N'\) indicate the block size determined by semi-variogram analysis. The number of horizontal and vertical blocks is the index of each block; there are various block sizes along the horizontal and vertical lines. Matching points are obtained using the NNDR [16], which uses the threshold in the ratio between the first and the second nearest neighbor descriptors. The NNDR can be defined as

\[
NNDR = \frac{||\text{descr}_A - \text{descr}_B||}{||\text{descr}_A - \text{descr}_C||}.
\]

where \(\text{descr}_A - \text{descr}_B\) and \(\text{descr}_A - \text{descr}_C\) are the distances to the nearest and second nearest neighbors, respectively. \(\text{descr}_A\) is the base descriptor, and \(\text{descr}_B\) and \(\text{descr}_C\) are
Fig. 6. Block size variation for computing appropriate block size along (a) horizontal line, and (b) vertical line.

Fig. 7. Number of correct matching points for block sizes: (a) non-block, (b) 200 × 200 block, (c) 250 × 250 block, and (d) adaptive block processing.

its closest two neighbors. Figure 7(d) shows the feature points obtained using the proposed adaptive block size. From the figure, the proposed adaptive block based method provides more feature points and better distribution than either the non-block based or other block based methods.

B. Outlier Removal using Statistical Analysis

In literature [26], the random sample consensus (RANSAC)\(^2\) [23] has been applied to initial correspondence sets to estimate the homography for outlier removal. The number of samples is adaptively set as the proportion of outliers is determined from each consensus set. RANSAC works well under conditions of small different correspondences and less local distortion. Also, RANSAC [27]–[29] can handle a moderate number of outliers. Figure 8(a) presents a superimposed image in which RANSAC has been applied. Note, however, that even though RANSAC was applied to remove outliers, outliers still exist; to more accurately remove outliers, in this study we propose the use of sign and statistical masks, as in Algorithm 1. Having few (or no) feature points in an area makes it difficult to calculate statistics. To overcome this issue, we merged the feature points within three row blocks—from \(R\{i\}\{j\}\) to \(R\{i + 2\}\{j\}\). This merged block provided more accurate statistics.

1) **Sign Mask**: Due to the panchromatic and multi-spectral CCD position and angle, there are differences along the \(x\) axis matching points between the panchromatic \(R(x, y)\) and multi-spectral \(U(x, y)\) images. In the figure, the difference in the matching point axis between the panchromatic (red circle) and multi-spectral (green circle) images is negative (I marked line in Fig. 8(a)) due to the multi-spectral image being shifted to the right. When aligning matching points between the multi-spectral image \(U(257.2, 860.1)\) and panchro-

\[\text{Algorithm 1 Outlier removal using sign and statistical masks.}\]

1: \textbf{procedure} \textsc{Sign \& Statistical Masks} \\
2: \text{\(R(x, y)\)} \triangleright \text{Input image: panchromatic \(R(x, y)\) with size \(X \times Y\)} \\
3: \text{\(U(x, y)\)} \triangleright \text{Input image: multi-spectral \(U(x, y)\) with size \(X \times Y\)} \\
4: \((M \times N)\) \triangleright \text{Image block size \(M \times N\)} \\
5: \text{for} \(i = 1 : 3 : \left\lfloor \frac{N}{3} \right\rfloor\) \text{do} \triangleright \text{Number of image block along the row, increment is 3} \\
6: \text{for} \(j = 1 : 1 : \left\lfloor \frac{M}{1} \right\rfloor\) \text{do} \triangleright \text{Number of image block along the column} \\
7: \text{\(Axis1\)} ← \(R\{i\}\{j\}\) to \(R\{i + 2\}\{j\}\) \triangleright \text{Matching points \(x\) axis in \(R\{i\}\{j\}\) to \(R\{i + 2\}\{j\}\)} \\
8: \text{\(Axis2\)} ← \(U\{i\}\{j\}\) to \(U\{i + 2\}\{j\}\) \triangleright \text{Matching points \(x\) axis in \(U\{i\}\{j\}\) to \(U\{i + 2\}\{j\}\)} \\
9: \text{\(AxisDiff1\)} ← \(Axis1 - Axis2\) \triangleright \text{Difference of \(Axis1\) and \(Axis2\)} \\
10: \text{if} \text{\(AxisDiff1\)} < 0 \text{then} \\
11: \text{\(Axis1’\) ← \(Axis1\)} \triangleright \text{Outlier removed new features \(Axis1\)} \\
12: \text{\(Axis2’\) ← \(Axis2\)} \triangleright \text{Outlier removed new features \(Axis2\)} \\
13: \text{end if} \\
14: \text{\(AxisDiff2\)} ← \(Axis1' - Axis2'\) \triangleright \text{Difference of \(Axis1'\) and \(Axis2'\)} \\
15: \text{\(AxisMean\)} ← \text{mean}(\text{AxisDiff2}) \triangleright \text{Mean of \(AxisDiff2\)} \\
16: \text{\(AxisStd\)} ← \text{STD}(\text{AxisDiff2}) \triangleright \text{Standard deviation of \(AxisDiff2\)} \\
17: \text{\(RangeLower\)} ← \text{AxisMean} - \text{AxisStd} \triangleright \text{Lower boundary of range} \\
18: \text{\(RangeUpper\)} ← \text{AxisMean} + \text{AxisStd} \triangleright \text{Upper boundary of range} \\
19: \text{if} \text{\((RangeLower < Axis1' < RangeUpper)\)} \text{& \((RangeLower < Axis2' < RangeUpper)\)} \text{then} \\
20: \text{\(Axis1''\) ← \(Axis1'\)} \triangleright \text{Outlier removed final axis 1} \\
21: \text{\(Axis2''\) ← \(Axis2'\)} \triangleright \text{Outlier removed final axis 2} \\
22: \text{end if} \\
23: \text{end for} \\
24: \text{end for} \\
25: \text{end procedure}
we further propose a statistical approach to remove outliers. If the difference is lower (α marked line in Fig. 8(a)), the matching point is considered to be an outlier. From Fig. 8(b) we observe that the marked outliers are accurately eliminated using the proposed sign mask.

2) Statistical Mask: Figure 9(a) also shows that even though a sign mask was applied to remove outliers, outlier α and β marked lines exist, in which the axis difference is already negative. To overcome this complicated condition, we further propose a statistical approach to remove outliers. To define the range of matching point difference, the feature points that the outlier previously removed using a sign mask, are used to calculate the mean and standard deviation (STD). The lower boundary of the range is calculated using the mean subtracting STD, whereas the upper boundary is computed using the mean adding STD. If the difference is lower (α marked line in Fig. 9(a)) or higher (β marked line in Fig. 9(a)) than the proposed range, the matching points are considered as outliers. Figure 9(b) shows the matching points after applying the proposed algorithm. As the matching points have the same direction and length, the figure confirms that the matching points are accurate and correct, i.e., that the outliers are accurately eliminated by the statistical mask.

C. Transformation Model Estimation and Re-sampling

The B-Spline method proposed by Rueckert et al. [24] serves as an alternative to Demons [30]. This method requires a rough pre-alignment to bring the images together such that only local deformations are present, and these local deformations can then be modeled using B-Spline. This method is more robust to noise than Demons and is not as dependent on texture. In contrast to thin-plate splines [31], B-splines are locally controlled, which makes them computationally efficient even for a large number of feature points.

The spatial resolution of the reference image is 1 m and the sensed image is 4 m. The transformed sensed image is interpolated at 4 times magnification via image re-sampling. Among various methods, pixels created in a bicubic interpolation [25] use $4 \times 4$ neighborhood pixels. Pixel variation can decrease the number of different interpolations because it gives weight to both sides. Note that though there is possibility to further improve the output image quality, the computing time required is larger than methods such as the nearest neighbor interpolation and bilinear interpolation. Despite this drawback, however, it gives the highest accuracy and a smoother surface.

IV. RESULTS AND DISCUSSIONS

A. Parameters and System

The proposed algorithm was implemented using Matlab on a PC with a quad-core processor, 8 GB RAM, and a 64-bit Linux operating system. To evaluate the performance significance, state-of-the-art methods such as Hessian-affine (HA) [32], [33], SURF [34], SIFT [6], and ASIFT [35] are compared to the proposed GSIFT. The parameters affecting each method on the feature extraction performance and descriptor construction. Two of the original SIFT parameters were modified to obtain more feature points. Boundary points were not removed, and the local extrema threshold was 0.001. The Hessian threshold was also modified to 500; more details of the parameters can be found in [36], [37]. The NNDR is 1.4; if this value is increased, the number of matching points will increase, though more processing time is also required and the process becomes less accurate.

B. Data Sets

In order to evaluate the proposed adaptive block processing, three different data sets were used, as shown in Fig. 10. An urban area, road, buildings, sea, and mountain regions can be seen. The size of each sample panchromatic image $R(x, y)$ is $15,023 \times 16,000$ and requires around 500 MB of storage. The sample multi-spectral image $U(x, y)$ is $3,792 \times 4,000$ and consumes approximately 30 MB of memory. The satellite images consist of one panchromatic image (spectral region: $0.5 - 0.9 \mu m$) and 4 spectral bands: blue (band 1, $0.45 - 0.52 \mu m$), green (band 2, $0.52 - 0.60 \mu m$), red (band 3, $0.63 - 0.69 \mu m$), and near-infrared (band 4, $0.76 - 0.90 \mu m$). The spatial resolution of the panchromatic image is 1 m and the multi-spectral image is 4 m. Geometric distortion is included in all images.
C. Distribution Quality Comparison

Table II shows the number of initial feature points (IFPs) and correct matching points (CMPs) of set 2 using both conventional methods and the proposed GSIFT. From the table, ASIFT provides the largest number of feature and correct matching points, followed by GSIFT, SIFT, SURF, and HA. Roughly, the number of correct matching points are similar to the proposed GSIFT. Using the proposed method, the average number of CMPs increased by 27% relative to the original SIFT. Figure 11 highlights the distribution of correct matching points; for a clear visualization, we used 1200 sampled feature points. Figure 11 visually shows the distribution of feature points; however, it was difficult to show how the feature points were distributed along the whole image. To measure the feature distribution, we proposed an algorithm referred to as the non-feature points area ratio (NR). This algorithm searches areas that have no feature points, as in Algorithm 2. This small value represents a greater spread of feature points. Figure 12 compares NR values to determine the distribution quality of feature points. The proposed method provides the lowest values, which indicates well-distributed feature points.

Algorithm 2 Non-feature points area ratio (NR)

1: procedure NON-FEATURE POINTS AREA RATIO
2: \( R(x, y) \) \( \triangleright \) Input image with size \((X, Y)\)
3: \((M' \times N')\) \( \triangleright \) Initial non-feature points area detection block size
4: \( NFP_A \leftarrow 0 \) \( \triangleright \) Initialize non-feature points area (NFP) count
5: for \( i = 1 \rightarrow \lceil \frac{X}{M'} \rceil \) do \( \triangleright \) Block number of image row
6: for \( j = 1 \rightarrow \lceil \frac{Y}{N'} \rceil \) do \( \triangleright \) Block number of image column
7: if feature points do not exist in \( I\{i\}\{j\} \) then
8: \( NFP_A \leftarrow NFP_A + 1 \) \( \triangleright \) Update NFP count
9: else
10: \( NFP_A \leftarrow NFP_A \)
11: end if
12: end for
13: end for
14: \( NR = \frac{NFP_A}{M'N'} \) \( \triangleright \) Non-feature points area ratio
15: end procedure

D. Accuracy Comparison

The root mean square error (RMSE) is a metric used to describe the accuracy of a registration process by measuring the location error.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left\| P_i - \hat{P}_i \right\|^2}.
\]  

(9)

where \( n \) is the number of points, \( P_i \) are points in the reference images, and \( \hat{P}_i \) are estimated points in the transformation model. Circular errors with a 90% confidence (CE90) [38], [39] are another metric used to describe the accuracy. A CE90 value is the minimum diameter of the horizontal circle that can be centered on feature points. Its mathematical representation is

\[
Pr \left( \left\| P_i - \hat{P}_i \right\| \leq CE90 \right) = 90\%
\]  

(10)

where \( Pr \) is the probability. Apart from RMSE, CE90 is only affected by the majority (90%); thus, lower RMSE and CE90 values indicate a more precise estimation of correct values. A third metric for comparing the registration accuracy is the
Fig. 13. RA1 comparison of previous methods (HA, SURF, SIFT, ASIFT), and proposed method (GSIFT).

Fig. 14. Warped RGB images of data set 2 (first row) using: HA (first column), SURF (second column), SIFT (third column), and ASIFT (last column), with their 4 times magnified images (second row).

probability at the 1 pixel error of RMSE.

$$RA1(\%) = \frac{\text{Number of points less than RMSE 1 error}}{\text{Number of points (n)}}.$$  (11)

Table III compares the conventional and proposed methods in terms of the registration accuracy between RMSE and CE90. From the table, performance of the proposed method is best, followed by ASIFT, SIFT, and HA. The RA1 comparison (Fig. 13) shows that the proposed method provides more accurate registration results.

E. Visual Comparison

Figure 14 compares the accuracy for the conventional methods. The first row of each is in color (red: multi-spectral band 3 image; green: multi-spectral band 2 image; blue: multi-spectral band 1 image). The second row shows the 4 times magnified images of its red square area in the RGB images; if registered accurately, there should be less of each color. Figure 15 demonstrates the proposed adaptive block processing and outlier elimination using statistical masks. In Fig. 15(b), the warped image is not correctly registered using standard RANSAC. In contrast, Fig. 15(c) shows the correct warped RGB image using the proposed outlier elimination. The proposed method can be clearly distinguished visually, and provides better warped RGB images compared to the conventional methods (HA, SURF, SIFT, and ASIFT).

V. CONCLUSIONS

In this article, we introduced an adaptive block processing in an attempt to improve accuracy using a geostatistical analysis.

In addition, we also proposed an outlier removal algorithm based on statistical masks. Subsequently, the proposed GSIFT was able to obtain more accurate results than methods such as HA, SURF, SIFT, and ASIFT. From the experimental results, the RMSE and CE90 value of the proposed method was less than half compared of that obtained using conventional methods. Through the distribution quality analysis, the proposed method also extracts more distinctive features than the previous methods, indicating that the proposed method provides more accurate registration results. Importantly, the high computational cost and less distributed feature points of the original SIFT prevented its wide application in recent remote sensing; this issue has been overcome with the proposed adaptive block, which accurately and efficiently extracts well-distributed feature points.

APPENDIX A

SCALE-INVARIANT FEATURE TRANSFORM

The scale-invariant feature transform (SIFT) consists of a feature detector and descriptor. The detector can extract features such as blobs/regions in the scale-space, ensuring invariance to image scaling, translation, and 2D rotation. The descriptor is also partially invariant to illumination and 3D viewpoint. Figure 16 represents the flowchart of the SIFT process.

A. Detector

- Scale-space: The scale-space \(L(x, y, \sigma)\) is constructed by the convolution between an input image \(I(x, y)\) and a variable-scale Gaussian \(G(x, y, \sigma)\) [40] as

\[
L(x, y, \sigma) = G(x, y, \sigma) \ast I(x, y),
\]

\[
G(x, y, \sigma) = \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}.
\]  (12)

where \(\ast\) is the convolution operator. Figure 17 presents an example of constructing a scale-space.
TABLE III
QUALITY COMPARISON FOR CONVENTIONAL METHODS AND PROPOSED GSIFT.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) RMSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HA</td>
<td>1.6059</td>
<td>1.7283</td>
<td>2.1925</td>
<td>1.6316</td>
<td>1.6852</td>
<td>2.0139</td>
<td>1.4094</td>
<td>1.546</td>
<td>2.2249</td>
<td>2.7117</td>
<td>2.6052</td>
<td>1.6914</td>
</tr>
<tr>
<td>SURF</td>
<td>1.806</td>
<td>1.1806</td>
<td>1.9255</td>
<td>1.2004</td>
<td>1.7549</td>
<td>1.9674</td>
<td>0.7897</td>
<td>1.2415</td>
<td>1.9961</td>
<td>2.4417</td>
<td>0.6528</td>
<td>1.423</td>
</tr>
<tr>
<td>SIFT</td>
<td>1.0228</td>
<td>1.2962</td>
<td>1.5140</td>
<td>1.0954</td>
<td>1.5134</td>
<td>1.8139</td>
<td>0.7307</td>
<td>1.2331</td>
<td>1.6588</td>
<td>2.2252</td>
<td>0.6462</td>
<td>1.366</td>
</tr>
<tr>
<td>ASIFT</td>
<td>1.0023</td>
<td>0.9379</td>
<td>1.1080</td>
<td>0.7111</td>
<td>1.6625</td>
<td>1.5993</td>
<td>0.6258</td>
<td>1.0003</td>
<td>3.0146</td>
<td>3.8997</td>
<td>0.4803</td>
<td>2.701</td>
</tr>
<tr>
<td>GSIFT</td>
<td>0.6205</td>
<td>0.6157</td>
<td>0.7269</td>
<td>0.5225</td>
<td>0.6810</td>
<td>0.9201</td>
<td>0.3284</td>
<td>0.5997</td>
<td>0.8779</td>
<td>1.1427</td>
<td>0.3527</td>
<td>0.6594</td>
</tr>
</tbody>
</table>

b) CE90
| HA      | 1.1964 | 1.3176 | 1.5821 | 1.2490 | 1.2759 | 1.5911 | 1.0406 | 1.1736 | 1.5347 | 1.9446 | 1.3166 | 1.2843 |
| SURF    | 0.8581 | 0.8393 | 1.4127 | 0.8617 | 1.1999 | 1.2951 | 0.5121 | 0.6594 | 1.0228 | 1.2962 | 0.5121 | 0.6594 |
| SIFT    | 0.7347 | 0.9344 | 1.1085 | 0.7744 | 1.1403 | 1.3375 | 0.5203 | 0.8442 | 1.0228 | 1.2962 | 0.5121 | 0.6594 |
| ASIFT   | 0.5267 | 0.6908 | 0.829  | 0.5185 | 0.9027 | 1.1038 | 0.3928 | 0.6869 | 1.1507 | 1.5124 | 0.4702 | 1.0124 |
| GSIFT   | 0.3526 | 0.4412 | 0.5234 | 0.3839 | 0.4821 | 0.6618 | 0.2399 | 0.4066 | 0.5842 | 0.735  | 0.2444 | 0.4421 |

Fig. 17. Constructed scale-space.

Fig. 18. DoG example.

- Difference of Gaussian (DoG): Blob is detected as local extrema of the DoG scale-space as
  \[ D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \ast I(x, y), \]
  \[ = L(x, y, k\sigma) - L(x, y, \sigma). \]  
  An example of a DoG is shown in Fig. 18. To extract the local extrema in a DoG scale-space, \(3 \times 3 \times 3\) neighborhoods are compared to each point of a \(3 \times 3\) window. The points having local maxima or minima are considered keypoints.

- Localization: To ensure sufficient contrast and edge points, keypoints having an unstable extremum of less than 0.03 and a principal curvature ratio at an extremum of more than 10 are rejected.

B. Descriptor

- Assign orientation: Each candidate keypoint can be assigned a magnitude \(m(x, y)\) and orientation \(\theta(x, y)\), such that
  \[ m(x, y) = \{((L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2)^{\frac{1}{2}}, \]
  \[ \theta(x, y) = \tan^{-1} \left( \frac{(L(x, y) - L(x - 1, y))}{(L(x + 1, y) - L(x, y - 1))} \right). \]  
  - Histogram: For orientation invariance, the sampling grid for the histograms is rotated to the main orientation of each keypoint. The grid is a \(4 \times 4\) gradient window obtained by using \(4 \times 4\) sample cells of 8-bin orientation histograms, which produces 128-dimensional feature vectors. Figure 19 presents a graphical representation of the descriptor.
  - Normalization: To achieve invariance to illumination changes, the descriptor is normalized with respect to the unit length.
  - Gaussian weighting: This function is applied to give less importance to gradients farther from the descriptor center and to avoid sudden changes.

APPENDIX B
RANDOM SAMPLE CONSENSUS (RANSAC)

Before applying a random sample consensus (RANSAC) [23], putative correspondences are obtained by extracting feature points independently in both the reference and sensed images. In our case, we used the nearest-neighborhood distance ratio (NNDR) [16]. To eliminate the outliers, RANSAC is used to cope with a large proportion of outliers; the objective is to provide an initial point correspondence set. The RANSAC
algorithms are subsequently applied to the initial correspondence set to estimate the homography model $H$, which can be represented by

$$H = \begin{bmatrix} H_1 \\ H_2 \\ H_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}.$$  \hfill (16)

The coordinates of $u$ and $v$ represent the feature points in the sensed image, whereas $x$ and $y$ represent the coordinates of the feature points in the reference image. Using homography, each point pair is computed as

$$x = \frac{h_{11}u + h_{12}v + h_{13}}{h_{31}u + h_{32}v + h_{33}}, \quad y = \frac{h_{21}u + h_{22}v + h_{23}}{h_{31}u + h_{32}v + h_{33}}.$$  \hfill (17)

The process of RANSAC is as follows:

i. Randomly choose 4 putative correspondent point pairs between the reference and sensed images.

ii. Compute the homography model $H$ using the selected sample set.

iii. Calculate the distance for each putative correspondence.

iv. Classify the consensus set $S_c$ into inliers and outliers using the distance threshold.

v. If the size of the number of inliers is greater than some threshold, re-estimate the model using all the feature points in $S_c$, and terminate.

vi. After repeating this process a number of times, the largest consensus set $S_c$ is selected, and the model is re-estimated using all the feature points in the subset $S_c$.

ACKNOWLEDGMENT

This research was supported by the Space Core Technology Development Program through the National Research Foundation of Korea by the Ministry of Education, Science and Technology (No.2012M1A3A3A02033352).

REFERENCES


Ik-Hyun Lee received a B.S. degree in Control and Instrument Engineering from Korea University, Korea, in 2004, and an M.S. degree in School of Information and Mechatronics at Gwangju Institute of Science and Technology, Korea, in 2008. He is currently doing Ph.D. research study at Signal and Image Processing Lab. in School of Information and Mechatronics at Gwangju Institute of Science and Technology, Korea. His research interests include image registration, multi focus image fusion and 3D shape recovery.

Tae-Sun Choi received a B.S. degree in Electrical Engineering from the Seoul National University, Seoul, Korea, in 1976, an M.S. degree in Electrical Engineering from the Korea Advanced Institute of Science and Technology, Seoul, Korea, in 1979, and a Ph.D. degree in Electrical Engineering from the State University of New York at Stony Brook, in 1993. He is currently a professor in the School of Information and Mechatronics at Gwangju Institute of Science and Technology, Gwangju, Korea. His research interests include image processing, machine/robot vision, and visual communications.