Robust Multi-Sensor Image Registration by Enhancing Statistical Correlation

Kyoung Soo Kim, Jae Hak Lee, and Jong Beom Ra, Senior Member, IEEE

Abstract—This paper deals with robust registration of the images acquired by different sensors, namely, the electro-optic (EO) and infrared (IR) ones. In this paper, we propose the two preprocessing schemes to improve the performance of normalized mutual information (NMI) based registration. Both schemes try to enhance the statistical correlation between a pair of EO and IR images for accurate and fast registration. The first scheme, extraction of statistically correlated regions (ESCR), extracts the regions in an image that are highly correlated to their corresponding regions in the other image. This extraction procedure is performed for each image, and the commonly extracted regions are used for calculating NMI. The second scheme, enhancement of statistical correlation by filtering (ESCF), adaptively filters out the pair of images to enhance the statistical correlation between them. The proposed schemes are applied to NMI-based registration and the results are prospective for various pairs of EO/IR sensor images in terms of registration accuracy, robustness, and speed.

Index Terms—Image registration, normalized mutual information, statistical correlation, electro-optic and infrared images.

I. INTRODUCTION

MULTI-SENSOR IMAGES have been actively used in many image-based applications, such as surveillance, automatic target recognition (ATR) and tracking, and medical image analysis, since they provide complementary information. In those applications, image registration is a very fundamental and important part. Image registration may be regarded as a process to establish the spatial correspondence between the two images of the same scene at different view points, at different times or by different-modality sensors. Multi-sensor image registration is considered a very difficult task, because the relationship between intensities of the images acquired by different sensors is usually complicated and not well-known. However, it is one of challenging research topics due to its importance in many multi-sensor image based applications.

There are two broad categories of image registration for multi-sensor images. One of them is the feature-based image registration which extracts and uses the features having common information in the two images. The feature-based algorithms utilize the features such as feature points [1], contour feature [2], edge [3], oriented edge vector fields [4], and gradient [5]. In those algorithms, selection accuracy of corresponding features directly affects the performance of image registration. The other is the intensity-based image registration which utilizes pixel intensities in the overlapped region of two images. However, due to the complicated relationship between the intensities of corresponding pixels [4], a simple registration technique based on area correlation cannot be directly applied. Therefore, the algorithms using the statistical information have been proposed. Among them, the algorithms based on mutual information (MI) [6]-[10] or normalized mutual information (NMI) [11]-[13] are the representative ones.

In many applications, EO/IR image registration has been widely used as an important part of information fusion, since the information of EO/IR images has a complementary nature. Namely, IR images provide temperature information, while EO images provide the reflection and radiation information of visible rays. However, due to the different characteristics of EO/IR images, image registration for EO/IR images is known to be quite difficult. Since the temperature of an object depends on the environment (such as weather, local time, and atmospheric temperature), the intensity in IR image is very sensitive to the environment. And, in IR images, some salient edges are often missing, since the temperature is not abruptly changed at object edges. Therefore, some features in an EO image cannot be found in the corresponding IR image, and vice versa. If corresponding pairs of features between the two images are not extracted exactly, the feature-based image registration cannot provide a good performance. Due to these characteristics, feature-based image registration algorithms may be inadequate to register a pair of EO/IR images.

Meanwhile, the relationship between the intensities of pixels in a pair of EO/IR images is usually complex. Contrast reversal may occur in some local regions of the two images, and multiple intensities in one sensor image may correspond to a single intensity in the other image [4]. Due to these characteristics, the intensities between a pair of EO/IR images cannot be globally correlated. Moreover, they are often not even statistically correlated [4]. Therefore, intensity-based algorithms cannot be easily applied to EO/IR registration. Even the NMI-based image registration algorithm, which has been known to be accurate and robust for multimodality image registration [11]-[13], may result in a poor performance, because a pair of EO/IR images does not have the global statistical correlation. Therefore, this paper proposes the two preprocessing schemes, extraction of statistically correlated

K. S. Kim, J. H. Lee, and J. B. Ra are with Korea Advanced Institute of Science & Technology 373-1, Guseong-dong, Yuseong-gu, Daejeon, South Korea (phone: +82-42-869-5434; fax: +82-42-869-8034; e-mail: {kskim, jhlee}@issserver.kaist.ac.kr, jbra@ee.kaist.ac.kr).

Manuscript received May 8, 2005. This research was supported by the Agency for Defense Development, Korea, through the Image Information Research Center at Korea Advanced Institute of Science & Technology.
regions (ESCR) and enhancement of statistical correlation by filtering (ESCF), to improve the performance of NMI-based image registration. The organization of this paper is as follows. Section II reviews an NMI based image registration algorithm. In section III, the proposed algorithm is described. And section IV provides the experiment results. Finally, the paper is concluded in section V.

II. NMI-BASED IMAGE REGISTRATION

We first review an NMI-based image registration algorithm which is the basis of the proposed preprocessing schemes. The NMI-based algorithm uses the statistical information of pixel intensities. Let us consider a pair of images acquired at different positions. We regard the one as a fixed image \( F \) and the other as the moving image \( M \) to be aligned. Then, the registration process is to find appropriate parameter values \( p \) of a predefined transformation model, which establishes the spatial correspondence between the two images. Namely, the process can be described as

\[
\text{p_{final}} = \arg \max_p S(F, M; p),
\]

where \( S \) denotes a similarity measure of the two images, \( F \) and \( M \) transformed with \( p \). The NMI-based algorithm uses NMI as the similarity measure.

Shannon’s marginal entropy \( H \) is defined as

\[
H(X) = -\sum_x p_x(x) \log p_x(x),
\]

(2)

where \( p_x(x) \) denotes the probability mass function (pmf) of discrete random variable \( X \). And the joint entropy is defined as

\[
H(X, Y) = -\sum_{x,y} p_{xy}(x,y) \log p_{xy}(x,y),
\]

(3)

where \( p_{xy}(x,y) \) denotes the joint pmf of two discrete random variables \( X \) and \( Y \). Then, MI is represented as

\[
I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y).
\]

(4)

If we assume that random variables \( (X, Y) \) are the intensity pair of \( F \) and \( M \), MI can be interpreted as the statistical correlation between the two images. Hence, as \( F \) and \( M \) become aligned, the statistical correlation, or MI, increases [6]. Therefore, it is possible to use MI as a similarity measure. However, since MI is sensitive to the size of overlapped region, it may not be a good similarity measure in some cases. This problem has been solved by introducing an overlap invariant entropy measure NMI [11], i.e.,

\[
I_s(X; Y) = \frac{H(X) + H(Y)}{H(X, Y)}.
\]

(5)

To obtain MI or NMI, we can use a joint histogram. The joint histogram is obtained from intensities of pixel pairs in a fixed image \( F \) and the corresponding \( M \). Since pixel grids of \( F \) and \( M \) may not be aligned, we need to perform interpolation in \( M \). And a joint histogram estimated by using improper interpolation may introduce a local maximum problem in NMI. To solve this problem, partial volume interpolation is generally used in many image registration algorithms, because it changes the joint histogram gradually during the registration [6], [7].

A proper selection of transformation model is important in the registration. If the scene can be approximated by a planar surface or the baseline between two sensors is small relative to their distance from the scene, a single 2D parametric transformation model can be used [4]. In this paper, we adopt the perspective transformation model to express the spatial correspondence between two images, namely,

\[
\begin{bmatrix}
  x''_m \\
  y''_m
\end{bmatrix} =
\begin{bmatrix}
  1 & x'_f & y'_f & 0 \\
  0 & 0 & 1 & x' f y' f - x' f^2 - y' f^2
\end{bmatrix}
\begin{bmatrix}
  x'_m \\
  y'_m
\end{bmatrix}.
\]

(6)

where \((x'_f, y'_f)\) and \((x''_m, y''_m)\) represent \( F \) and \( M \) transformed with \( p = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8)^T \), respectively. Here, parameters \( p_1, p_2, p_3 \), and \( p_4 \) represent the translation, scaling, and shearing along the \( x \)-axis, respectively. And \( p_5, p_6, p_7 \) represent the translation, shearing, and scaling along the \( y \)-axis, respectively. \( p_7 \) and \( p_8 \) express the perspective information. In other words, the former corresponds to the \( y \)-axis rotation, and the latter to the \( x \)-axis rotation.

The final goal of image registration is to find the parameter values maximizing a similarity measure. To improve the speed in finding those parameter values, many optimization techniques have been proposed. In MI or NMI-based image registration, the multi-dimensional downhill simplex method is popularly used due to its high speed and robustness [8], [14].

III. PROPOSED ALGORITHM

As mentioned above, NMI based algorithms are preferred in multi-modality image registration, due to their accuracy and robustness. Those algorithms are based on the assumption that the statistical correlation between the two images is global. Therefore, a direct use of NMI is not efficient for the EO/IR image registration, because EO and IR images do not often satisfy this assumption [4]. To solve this problem, we adopt the two preprocessing schemes so that we may extract and use only partial regions having high statistical correlation for NMI based registration.

A. Overall Structure of the Proposed Algorithm

The overall structure of the proposed registration algorithm is shown in Fig. 1. The algorithm consists of two stages. In the first stage, ESCR and ESCF are performed for each image, \( F \) and \( M \). ESCR efficiently extracts the regions in one image that are highly statistically correlated to their corresponding regions in the other image. And ESCF adaptively performs median filtering for each image, to enhance the statistical
correlation between \( F \) and \( M \). Then, in the second stage, the two output images from the first stage are registered based on NMI measure.

\[
\text{IR image (fixed image } F) \quad \text{EO image (moving image } M) \quad \Downarrow \quad \text{ESCR} \quad \Downarrow \quad \text{ESCF} \quad \Downarrow \quad \text{NMI-based image registration}
\]

Figure 1. Overall structure of the proposed image registration algorithm.

B. ESCR

The conditional entropy is defined as

\[
H(X | Y) = \sum_y p_Y(y) H(X | Y = y),
\]

where \( H(X | Y = y) \) denotes the entropy or uncertainty of \( X \) when \( Y \) is known. Note in Eq. (7) that the conditional entropy is the average of \( H(X | Y = y) \) for all \( y \). Hence, if \( H(X | Y = y) \) is low, the uncertainty in \( X \) given knowledge of \( Y \) is low. Similarly, if the statistical correlation between two images \( F \) and \( M \) is high, for known pixel values in \( F \), the uncertainty for the corresponding pixel values in \( M \) is low. In a pair of EO/IR images, features in one image may not exist in the other. Also, a single pixel value in one image may correspond to multiple pixel values in the other [4]. Therefore, the uncertainty for the pixel values in \( M \) for known pixel values in \( F \) is often high. Therefore, even for a well-registered EO/IR image pair, the conditional entropy can be high, or the statistical correlation can be low.

Now let us consider the relationship between the conditional entropy and NMI. From Eq. (4), MI can be rewritten as

\[
I(X;Y) = \frac{H(X) + H(Y)}{2} - \frac{H(X | Y) + H(Y | X)}{2}.
\]

By using Eqs. (5) and (8), NMI can be described as

\[
I_p(X;Y) = \frac{H(X) + H(Y)}{H(X) + H(Y)} - \frac{H(X | Y) + H(Y | X)}{H(X) + H(Y)}.
\]

The NMI newly described in Eq. (9) consists of marginal entropies \( H(X) \) and \( H(Y) \), and conditional entropies \( H(X | Y) \) and \( H(Y | X) \). In the image registration procedure, the marginal entropies are hardly changed compared to the conditional entropies. Therefore, we can notice from Eq. (9) that the conditional entropy is inversely proportional to NMI.

In using NMI as a similarity measure in the registration procedure, we assume the two conditions: NMI (conditional entropy) has the maximum (minimum) value when a pair of images is perfectly aligned, and it decreases (increases) smoothly as the two images become misaligned. However, due to the regions having low statistical correlation, the conditional entropy may not have the minimum value even if the two images are perfectly aligned, and also it may not smoothly change as the images become misaligned. Therefore, for a robust NMI-based multi-sensor image registration, it is important to effectively remove the regions having the high conditional entropy.

Fig. 2 shows a couple of pairs of real EO/IR images. In the figure, the pairs of elliptical regions between EO/IR images represent the ones having low statistical correlation or high conditional entropy, because an intensity value in one region corresponds to multiple intensity values in the corresponding region in the other image. To alleviate this undesirable phenomenon, we may select and remove the pairs of regions having low statistical correlation. However, it is practically difficult to find those pairs before image registration. Hence, we remove the regions having low intensity variation in each image so that the corresponding regions, which may have multiple intensity values in the other image, can be automatically removed in estimating the joint histogram. However, this straightforward removing process can also eliminate the pairs of rectangular regions having high statistical correlation, since those regions have low intensity variation. Fortunately, pairs of large rectangular regions do not affect the performance of image registration, because the joint histogram is hardly changed by these pairs. On the contrary, pairs of small rectangular regions should not be removed.

By considering the observation above, we propose the ESCR scheme as shown in Fig. 3. First, each image is smoothed by using a Gaussian smoothing filter to reduce the noise, and salient edges are then extracted by using Canny edge operator [15]. Second, the scheme produces the distance map by calculating the distance from each pixel to the nearest edge. The regions having large values in the distance map can be considered regions of large area with low intensity variation. Hence, ESCR defines and removes only the regions whose pixels have larger distances than predetermined \( d^p \).
Here, the value of $d^h$ is determined depending on the degree of initial misregistration and the size of overlapped region between the two images. Fig. 4 demonstrates the result by applying ESCR to a pair of EO/IR images. It can be shown in the figure that ESCR can successively select and remove the regions having low intensity variation only of large area, while keeping the regions having high statistical correlation. Note that small areas are not removed even if they have low intensity variation (for example, car window, etc.)

![Figure 4](image)

Figure 4. Application results of ESCR to (a) an EO image and (b) the corresponding IR image. (c), (d) Edge maps. (e), (f) ESCR masks. (g), (h) Regions extracted by ESCR with $dh$ of 10.

C. ESCF

EO and IR images usually have small intensity variation even in a homogeneous region, due to several reasons such as non-uniform illumination, temperature variation, and sensor noise. Therefore, the 2D joint histogram of a pair of EO/IR images is undesirably dispersed. This dispersion causes the high conditional entropy which may reduce the robustness of registration. And it also undesirably increases the frequency of one-to-many intensity correspondences between two images, which may increase the number of the evaluations in the optimization procedure. For example, if an object $A$ has intensity $a$ in both $F$ and $M$, and an object $B$ has intensity $a$ in $F$ but has intensity $b$ in $M$, $(a, a)$ and $(a, b)$ in the joint histogram have high densities, or high probabilities, when the two images are exactly aligned. However, if object $A$ in $F$ and object $B$ in $M$ are overlapped in the middle of image registration processing, $(a, b)$ in the joint histogram has high density even though two images are misaligned. This can provide a potentially poor registration result and an unnecessary increase of the number of evaluations in the optimization procedure [13].

In order to alleviate the dispersion of the joint histogram, we propose the ESCF scheme as shown in Fig. 5. In the first step, ESCF finds the salient edges. Then in the second, it adaptively performs 2D median filtering except the edges in the images. The median filter effectively removes the unwanted noise while preserving even small edges which are not detected in the first step. Thereby, we can flatten homogeneous regions while preserving discontinuous regions, and consequently reduce the histogram dispersion. Fig. 6 demonstrates the result of ESCF. By comparing the joint histograms given in Fig. 6(c) and (f), we can notice that ESCF can reduce the histogram dispersion as desired.
Figure 5. The proposed ESCF procedure.

Figure 6. Joint histogram before and after applying ESCF. (a) An EO image, (c) the corresponding IR image, and (e) their joint histogram. (b) The EO and (d) IR images after ESCF, and (f) their joint histogram.

IV. EXPERIMENT RESULTS AND ANALYSIS

For experiment, we use the four test pairs of EO/IR images as shown in Fig. 7. The size of all the images is 320 x 240. IR images are acquired by using either a mid-wave infrared (MWIR: about 3 ~ 5 μm) or long-wave infrared (LWIR: about 7 ~ 14 μm) camera.

To examine the spatial correspondence between a pair of EO/IR images, the perspective transformation model in Eq. (6) is adopted. We use the partial volume interpolation in estimating the joint histogram, because it changes the histogram gradually [6], [7], [9], [10], [16], [17]. And the multi-dimensional downhill simplex method is adopted in the optimization procedure to maximize NMI. A proper selection of \( F \) and \( M \) from EO/IR images is important to image registration performance. Since, in estimating the joint histogram, interpolation is performed in \( M \), the image of higher resolution is considered adequate to \( M \). In the experiment, IR images are assigned to \( F \) and EO images to \( M \).

We experimentally analyze the effects of the two proposed preprocessing schemes, ESCR and ESCF, to the NMI-based registration, by applying ESCR only or applying both ESCR and ESCF. Fig. 8 shows the image registration result by applying ESCR to image_pair_1. We note that the intermediate result of applying ESCR is already shown in Fig. 4. To examine the degree of registration, we display the alternating patches of an IR image and the registered EO image as shown in Fig. 8(a) and (b). The figures show that, as expected, the ESCR-applied registration algorithm provides a more accurate result than the conventional algorithm. Fig. 9 shows NMI curves for image_pair_1 depending on several geometric transformations, namely translation, scaling, and rotation. In the conventional algorithm, NMI curves do not have the maximum value when the two images are exactly aligned, and do not decrease smoothly as the images become misaligned.

On the other hand, by using partial regions selected by ESCR, the NMI curves become smooth and have a maximum value when the two images are exactly aligned.

Figure 7. Image pairs of EO/MWIR images or EO/LWIR images.
Figure 8. Comparison between the conventional image registration using the whole region and the registration using partial regions selected by ESCR for image_pair_1. (a) Registration result by the conventional registration algorithm using whole region. (b) Registration result using partial regions selected by ESCR.

Meanwhile, Figs. 10 and 11 illustrate the image registration results for image_pair_2 and 3 of different characteristics, respectively. Since image_pair_2 has low statistical correlation, the proposed algorithm improves the registration performance compared to the conventional algorithm (compare Fig. 10(c) with (e)). On the other hand, in the experiment for image_pair_3 with high statistical correlation, the conventional algorithm provides an accurate registration result similar to the one from the proposed one (compare Fig. 11(c) with (e)). However, the proposed algorithm is faster, because partial regions selected by ESCR, instead of the whole image, are used for NMI calculation and the statistical correlation of image_pair_3 is enhanced by ESCF. Table 1 also demonstrates that ESCF can make the image registration faster. Here, we consider the number of evaluations a measure of the processing speed. Note that the proposed algorithm also performs well for a pair of EO and LWIR images (see Fig. 12.).
which require accuracy and/or speed. Medical image and multi- or hyperspectral image registration useful for many multi-modality image registration, such as algorithm without preprocessing. The algorithm would be preprocessing schemes can provide more accurate and robust, proposed registration algorithm based on the two correspondences are reduced by ESCF. In summary, the number of evaluations in the optimization process can be performed only for the regions extracted by ESCR. Also, application of ESCR makes the computation time of each evaluation reduced, because the joint histogram estimation can make NMI-based registration more accurate and robust. The experimental examination as well as theoretical analysis, we find out that ESCR can efficiently extract the regions in an image that are highly correlated to their corresponding regions in the other image and ESCF can enhance the statistical correlation between corresponding regions. By applying the two preprocessing schemes, the NMI curve is varying smoothly and has the maximum value when a pair of images is exactly aligned, contrary to the conventional registration without preprocessing. Thereby, the two proposed schemes make NMI-based registration more accurate and robust. The application of ESCR makes the computation time of each evaluation reduced, because the joint histogram estimation can be performed only for the regions extracted by ESCR. Also, the number of evaluations in the optimization process can be reduced, because the cases of one-to-many intensity correspondences are reduced by ESCF. In summary, the proposed registration algorithm based on the two preprocessing schemes can provide more accurate and robust, and faster image alignment results than the conventional algorithm without preprocessing. The algorithm would be useful for many multi-modality image registration, such as medical image and multi- or hyper-spectral image registration which require accuracy and/or speed.

V. CONCLUSIONS

In this paper, we propose the two preprocessing schemes, ESCR and ESCF, to improve the performance of NMI-based image registration. Both schemes try to enhance the statistical correlation between a pair of EO/IR images. From the experimental examination as well as theoretical analysis, we find out that ESCR can efficiently extract the regions in an image that are highly correlated to their corresponding regions in the other image and ESCF can enhance the statistical correlation between corresponding regions. By applying the two preprocessing schemes, the NMI curve is varying smoothly and has the maximum value when a pair of images is exactly aligned, contrary to the conventional registration without preprocessing. Thereby, the two proposed schemes make NMI-based registration more accurate and robust. The application of ESCR makes the computation time of each evaluation reduced, because the joint histogram estimation can be performed only for the regions extracted by ESCR. Also, the number of evaluations in the optimization process can be reduced, because the cases of one-to-many intensity correspondences are reduced by ESCF. In summary, the proposed registration algorithm based on the two preprocessing schemes can provide more accurate and robust, and faster image alignment results than the conventional algorithm without preprocessing. The algorithm would be useful for many multi-modality image registration, such as medical image and multi- or hyper-spectral image registration which require accuracy and/or speed.

Table 1. Comparison of the number of evaluations

<table>
<thead>
<tr>
<th>Image pairs</th>
<th>Without ESCF</th>
<th>With ESCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>284</td>
<td>194</td>
</tr>
<tr>
<td>2</td>
<td>179</td>
<td>161</td>
</tr>
<tr>
<td>3</td>
<td>109</td>
<td>94</td>
</tr>
</tbody>
</table>

REFERENCES


