REGISTRATION OF HYPERSONTAL AND TRICHROMATIC IMAGES VIA CROS CROSS CUMULATIVE RESIDUAL ENTROPY MAXIMISATION

Mahmudul Hasan¹, Mark Pickering¹, Antonio Robles-Kelly¹,²,³, Jun Zhou¹,²,³, Xiuping Jia¹

¹ School of Engineering and Information Technology, University College, The University of New South Wales, Australian Defence Force Academy, Canberra, ACT, Australia.
²Canberra Research Laboratory, NICTA, Canberra, ACT, Australia.
³College of Engineering and Computer Science, The Australian National University, Canberra, Australia.

ABSTRACT

In this paper we address the problem of image fusion between imagery acquired by trichromatic sensors and hyperspectral imagers. We do this by presenting a method aimed at registering a high-resolution trichromatic image with lower resolution hyperspectral data. The method presented here maps the hyperspectral image into the grayscale image so as to employ the cross cumulative residual entropy for purposes of multimodal registration. We illustrate the utility of our approach by presenting registration results on a set of surveillance image pairs consisting of a set of high-oblique colour and hyperspectral images.

Index Terms— CCRE, Mutual information, Image registration, Hyperspectral imagery.

1. INTRODUCTION

The use of hyperspectral and thermal cameras along with trichromatic cameras is becoming more and more common in surveillance applications. To interpret information effectively in an automated multi-sensor surveillance system, a robust multi-sensor image fusion system is essential. Image registration is a vital pre-processing step in this fusion process. Popular image registration algorithms for monochromatic imagery provide good performance for single sensor applications but their use in multi-sensor applications is limited due to their reliance on intensity-based similarity measures. These intensity-based similarity measures include normalized cross-correlation [1], sum-of-the-squared difference (SSD) [2], and optical-flow estimation [3], amongst others. Due to brightness regularity assumptions of the image pixels, these algorithms are not suitable for multi-modal image registration.

Multi-modal image registration is a more challenging task and several interesting methods have been proposed in the literature to address this problem, especially for medical image processing. Mutual information (MI) based on Shannon’s entropy is a popular method for multimodal image registration which was first introduced by Viola and Wells [4]. The cross cumulative residual entropy (CCRE) was introduced by Wang and Vemuri [5] and proved to be more robust than mutual information in medical imaging applications. Application of both MI and CCRE similarity measures for image registration in remote sensing with multi-spectral and multi-modal image data by Pickering et al [6, 7] also showed promising results. The key strengths of the CCRE over other popular mutual information methods are that CCRE has significantly larger noise immunity and a larger convergence range over the field of parameterized transformations.

In this paper, we evaluated image registration methods which use CCRE and MI as similarity measures for the registration of trichromatic and hyperspectral surveillance images.

2. IMAGE REGISTRATION

Since automatic image registration algorithms usually require two grayscale images, several techniques have been proposed to convert hyperspectral and trichromatic images to grayscale images. Amongst them, the simplest method is to take an average across all the bands or channels. Another option is to use the CIE 1931 transfer table [8] so as to estimate the image-band contributions to form a trichromatic image [9]. In either case, upon obtaining the grayscale images from the hyperspectral and trichromatic images, registration can be performed using different methods.

Note that the image registration performance can be affected by both the registration algorithm and the quality of grayscale images to be registered. In this paper, we focus on the performance of the registration algorithm. In the following section we discussed two multi-modal information-theoretic similarity measures – mutual information and cross-cumulative residual entropy. When the images are perfectly registered these similarity measures will have a maximum value. If the images are misaligned the similarity measure value will be reduced. So, the image registration problem is now converted to a non linear optimization problem using the negative of the similarity measures. As the gradient and Hessian matrix of these two
similarity measures can be computed analytically using Parzen-window estimation [10, 11], Newton’s optimization method can be applied for solving this nonlinear optimization. The relationship between the two sets of image coordinates \((x', y')\) and \((x, y)\) is often modelled using an affine transform, i.e.

\[
\begin{bmatrix}
  x' \\
  y' \\
  1
\end{bmatrix} = \begin{bmatrix}
  m_1 & m_2 & m_3 \\
  m_4 & m_5 & m_6 \\
  0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
  x \\
  y \\
  1
\end{bmatrix} \tag{1}
\]

The optimal set of affine transformation parameters \(m\) is found when the negative of the similarity measure is minimised.

2.1. Mutual Information

In medical imaging, Viola and Wells [4] applied mutual information (MI) to register magnetic resonance (MR) images with a 3D model of a human skull. For multi-modal medical applications, several techniques have been proposed that use mutual information (MI) as the similarity measure [11, 12]. The algorithms can be used in a wide variety of imaging conditions. MI is defined in terms of Shannon entropy and joint probability distribution. Let \(u\) and \(v\) be the \(i\)th pixel value of target image \(I(x_i, y_i)\) and the reference image \(R(x_i, y_i)\) respectively and let \(L\) be the maximum pixel value of the both images. The joint and marginal probability distribution functions are then defined by:

\[
P(u, v) = p(I = u, R = v) \tag{2}
\]

\[
P_j(u) = \sum_{v=1}^{L} P(u, v) = \sum_{v=1}^{L} p(I = u, R = v) \tag{3}
\]

\[
P_R(v) = \sum_{u=1}^{L} P(u, v) = \sum_{u=1}^{L} p(I = u, R = v) \tag{4}
\]

where \(p(\ )\) is the probability mass function (pmf). According to Thévenaz and Unser [10], the mutual information between \(I\) and \(R\) is then defined as

\[
S = \sum_{u=1}^{L} \sum_{v=1}^{L} P(u, v) \log \left( \frac{P(u, v)}{P(u)P_R(v)} \right) \tag{5}
\]

To make the measure differentiable with respect to the motion parameters it is necessary to estimate the probability distribution function of an image using the Parzen-window method. Thévenaz and Unser [10] chose a cubic B-spline function denoted by \(\beta^3(\ )\) to estimate the joint probability distribution function of two images using the Parzen-window method and this term is given by,

\[
\hat{P}(u, v) = \frac{1}{N} \sum_{i=1}^{N} \beta^3(u - I(x'_i, y'_i))\beta^3(v - R(x_i, y_i)) \tag{6}
\]

As shown by Thévenaz and Unser [11] and Pickering et. al [6], the gradient and Hessian matrices can be calculated analytically using equations (5) and (6) and Newton’s optimization method can then be used to obtain a set of optimal transformation parameters.

2.2. Cross-Cumulative Residual Entropy

Wang and Vemuri [5] proposed an image registration method using an information-theoretic similarity measure known as cross cumulative residual entropy (CCRE). The CCRE is defined using cumulative residual distributions rather than probability density functions. Cumulative residual distributions are more regular and robust than probability densities in presence of noise. The CCRE similarity measure for two images \(I(x', y')\) and \(R(x, y)\) is defined using cumulative residual distributions by the following equation:

\[
C = \sum_{u=1}^{L} \sum_{v=1}^{L} G(u, v) \log \left( \frac{G(u, v)}{G_j(u)P_R(v)} \right) \tag{7}
\]

where \(L\) is the maximum pixel value of the images. The joint and marginal cumulative residual distributions are given as [6],

\[
G(u, v) = p(I > u, R = v) \tag{8}
\]

\[
G_j(u) = \sum_{v=1}^{L} G(u, v) = \sum_{v=1}^{L} p(I > u, R = v) \tag{9}
\]

In the algorithm proposed by Wang and Vemuri [5], the Parzen-window estimate of the joint cumulative residual distributions is given by

\[
\hat{G}(u, v) = \frac{1}{N} \sum_{i=1}^{N} \phi(u - I(x'_i, y'_i))\beta^3(v - R(x_i, y_i)) \tag{10}
\]

Here function \(\beta^3(\ )\) is a cubic B-spline function and \(\phi(\ )\) is the cumulative residual function [6, 7]. As shown by Wang and Vemuri [5] and Pickering et. al [6], the elements of the gradient vector \(\nabla C\) and Hessian matrix \(\nabla^2 C\) can be calculated analytically for use in Newton’s optimization method.

3. EXPERIMENTS

Seven sets of hyperspectral and tri-chromatic images were collected for our experiments. The hyperspectral imagery was acquired using an imaging system comprising of an Acousto-Optic Filter fitted to a FireWire camera. The images were captured at 21 bands ranging from 430nm to 630nm wave lengths. The trichromatic imagery was acquired using a Nikon D80 digital SLR.
The hyperspectral and trichromatic cameras were positioned on a high-oblique point overlooking a car park and its surrounding area so as to capture images over the day. Due to the differences in the camera positions, the focal lengths and the acquisition times between the hyperspectral camera and the Nikon DSLR, the images obtained from the two cameras showed a fair amount of variation. Some image samples are shown in Figure 1.

Prior to registration, the images were converted to grayscale images. For each image pair the images were first manually registered and then a set of 50 artificial affine transformations were applied on the tri-chromatic image. The ranges of these artificial transforms were: up to 120 pixels translation, between $-25^\circ$ to $25^\circ$ degree rotations and from 85% to 115% scaling. To speed up the registration, the DSLR images were first scaled down to 0.125 of the original. After registering at this scale, images were registered in 0.25 and 0.5 scales for good precision.

The performance of the CCRE and MI based image registration were evaluated with the above mentioned datasets. Figure 2 shows a pseudo-colour sample result of the CCRE based registration method. It can be observed that although the images were captured at quite different scales and from a different viewpoint, the registration is successful.

Figure 1. Sample images from the dataset. The left-hand column shows sample pseudo-colour imagery for the hyperspectral camera. The right-hand column shows samples from the Nikon DSLR.

Figure 2. (a) Pseudo-colour image of a sample hyperspectral image of size 1280x960 pixels (b) corresponding tri-chromatic SLR image resized to 1437x1139 pixels and (c) registration result.

Figure 3 shows the average registration error at each iteration for both methods. Both methods successfully registered the image, however, the CCRE based registration method converges faster than the MI based registration method. Note that the image scale is increased from 12.5% to 25% after 200 iterations, which results in a sharp drop in registration error for both the methods.

Figure 4 shows a plot of the final maximum error versus initial maximum error and Figure 5 shows the registration success rate for the two registration methods. Note that, for displacements of up to 70 pixels, the CCRE based registration method effectively has a 100% success rate and for larger initial displacements the CCRE method outperforms the MI based method. It should also be noted that a single-sensor registration algorithm using SSD as the similarity measure failed to register the images in all cases.
4. CONCLUSION

In this paper we addressed the problem of finding a suitable multi-sensor registration algorithm to automatically register images captured using trichromatic and hyperspectral cameras. We evaluated the performance of registration methods using two information theoretic similarity measures. The CCRE based image registration method showed better performance by showing faster convergence, better success rate and better registration accuracy when compared to a MI based image registration method.

This accurate registration procedure will allow more sophisticated automatic analysis of surveillance imagery captured using these sensors.

5. REFERENCES


