GEOMETRIC REGISTRATION OF IMAGES WITH ARBITRARILY-SHAPED LOCAL INTENSITY VARIATIONS FROM SHADOWS

M. M. Fouad, R. M. Dansereau and A. D. Whitehead

Department of Systems and Computer Engineering, Carleton University
Ottawa, Ontario, Canada K1S 5B6

ABSTRACT

In this paper, we focus on the sub-pixel geometric registration of images with arbitrarily-shaped local intensity variations, particularly due to shadows. Intensity variations tend to degrade the performance of geometric registration, thereby degrading subsequent processing. To handle intensity variations, we propose a model with illumination correction that can handle arbitrarily-shaped regions of local intensity variations. The approach is set in an iterative coarse-to-fine framework with steps to estimate the geometric registration with illumination correction and steps to refine the arbitrarily-shaped local intensity regions. The results show that this model outperforms linear scalar model by a factor of 6.8 in sub-pixel registration accuracy.

Index Terms— Image registration, local intensity correction, photometric correction, panoramic image mosaics.

1. INTRODUCTION

Geometric image registration of a set of images is a common pre-processing step in many applications, such as generating panoramic image mosaics, performing super-resolution enhancement, or for change detection for environmental monitoring. For many of these applications, sub-pixel accuracy in registration is necessary for satisfactory post-processing results. Unfortunately, for some image sets the local and global intensity variations between the images tend to degrade or even spoil the resulting geometric registration.

Although a number of researchers have incorporated global [1] and semi-local photometric/illumination correction into the geometric image registration problem [2, 3], the resulting registration from these models tends to fail when dramatic and disjoint shifts in local intensity exist. Many of the existing models assume slowly varying intensity changes, which isn’t true for all imaging scenarios. Dramatic intensity shifts occur in many image sets, such as shadows in outdoor images taken at widely varying times of the day (see Fig. 1).

Another limitation of the existing image registration techniques is that they operate either on a global illumination model or on a local illumination model that doesn’t easily allow arbitrarily-shaped regions. For instance in [2], semi-local intensity changes are modeled within pre-defined neighborhoods with an imposed smoothness constraint between neighborhoods. In [3], a semi-local illumination model is given with triangular or quadrilateral region support. While these approaches to modeling intensity variations are valuable for slowly changing intensities differences, dramatic intensity shifts over arbitrarily-shaped regions are poorly represented.

In this paper, we focus on the geometric registration of images with disjoint local intensity shifts, particularly due to large shadow differences, such as from time of day capture. One important assumption made to simplify our modeling is that a single point light source is assumed, which is reasonable for daytime outdoor imaging with the sun as our natural light source. We are interested in correcting the shadows in the target image and performing sub-pixel accurate geometric registration with global and local intensity differences estimated and compensated in the source image being registered. Also, the regions with local intensity shifts can be of arbitrary shape and are extracted through an iterative framework to allow the geometric registration to converge to sub-pixel accuracy.

This paper is organized as follows. In Sec. 2, we present our proposed model that deals with arbitrarily-shaped local intensity changes jointly with geometric registration in a coarse-to-fine framework, with segmenting the intensity varying regions to constrain the solution for the single point light source assumption. In Sec. 3, we present experimental results using natural images and satellite image pairs with simulated shadows that allow us to compare the results to ground truth results. Finally, conclusions are discussed in Sec. 4.

2. GEOMETRIC REGISTRATION WITH ARBITRARILY-SHAPED LOCAL INTENSITY-BASED MODEL

Consider that we have two input images, $I_1$ and $I_2$, captured for the same scene. Following the models in [2, 3], we can formulate our generalized geometric registration and arbitrarily-shaped local intensity model that relates $I_1$ to $I_2$ as

$$I_2(x_2, y_2) \simeq b I_1(x_1, y_1) + c$$

(1)
Fig. 1. (a,b) Two natural canyon images captured at different times of the day, and (c,d) source and target satellite images with simulated shadows, respectively.

where the intensity variations are modeled with

\[
\begin{bmatrix}
    b_{11} & \ldots & b_{1M} \\
    \vdots & \ddots & \vdots \\
    b_{N1} & \ldots & b_{NM}
\end{bmatrix}, \quad \begin{bmatrix}
    c_{11} & \ldots & c_{1M} \\
    \vdots & \ddots & \vdots \\
    c_{N1} & \ldots & c_{NM}
\end{bmatrix},
\]

(2)

While the motion in (1) could be modeled in many ways, such as is discussed in [3], we use the 6-parameter affine model

\[
\begin{align*}
x_2 &= x_1 + a_1 x_1 + a_2 y_1 + a_5 \\
y_2 &= y_1 + a_3 x_1 + a_4 y_1 + a_6
\end{align*}
\]

(3)

where \([a_1, a_2, \ldots, a_6]^T\) is the motion vector field.

The primary difference between our model and those presented in [2, 3]\(^1\) is that \(b\) and \(c\) are matrices as opposed to scalars. Having \(b\) and \(c\) as matrices allows for arbitrary intensity changes at a pixel-by-pixel level between \(I_1\) and \(I_2\). Unfortunately, having \(b\) and \(c\) as matrices makes solving (1) an ill-posed problem. However, if we make the assumption that the scene is illuminated with a single point source light, such as the sun, and that specular reflection and diffusion in the scene creates reasonably constant ambient light in the shadows, then we can constrain the solution of (1). This illumination scenario is not uncommon for daytime outdoor imaging, which can include aerial and satellite imaging. By assuming these illumination conditions, we can argue that \(I_1\) and \(I_2\) will contain areas of direct illumination from the point source light and areas of partial illumination primarily by the ambient light alone (i.e., shadows). The model could be extended to multiple point source lights, which would create additional regions of illumination, but this paper focuses on the case of a single point source light.

With the single point source light assumption, image \(I_i\), \(i \in \{1, 2\}\), can be segmented into regions of direct illumination \(R_i^1\) and regions of ambient illumination \(R_i^a\). This segmentation normally corresponds to non-shadow and shadow regions, respectively. With this segmentation for the two images, rough geometric registration of \(I_1\) and \(I_2\) will produce overlapping regions, where some regions would have a similar illumination

\[
R_{1\oplus 2}^1 = (R_1^1 \cap R_2^1) \cup (R_1^1 \cap R_2^a) \quad (5)
\]

and other regions would have a dissimilar illumination

\[
R_{1\oplus 2}^d = (R_1^1 \cap R_2^a) \cup (R_1^a \cap R_2^1), \quad (6)
\]

Using the regions defined by \(R_{1\oplus 2}^1\) and \(R_{1\oplus 2}^d\), we have arbitrarily-shaped regions that define the illumination differences between \(I_1\) and \(I_2\) in terms of two states. We can thus constrain \(b\) and \(c\) in (1) to these two states as follows

\[
b(x, y) = \begin{cases} 
    b_s, & \forall (x, y) \in R_{1\oplus 2}^s \\
    b_d, & \forall (x, y) \in R_{1\oplus 2}^d
\end{cases} \quad (7)
\]

\[
c(x, y) = \begin{cases} 
    c_s, & \forall (x, y) \in R_{1\oplus 2}^s \\
    c_d, & \forall (x, y) \in R_{1\oplus 2}^d
\end{cases} \quad (8)
\]

where \(b_s\) and \(b_d\), and \(c_s\) and \(c_d\) are scalars for the similar and dissimilar illumination states for \(b\) and \(c\), respectively. Note that, for pixel domain mathematical manipulations, \(R_{1\oplus 2}^1\) and \(R_{1\oplus 2}^d\) bring out zero-one masks, \(Q_s\) and \(Q_d\), respectively such that

\[
Q_s(x, y) = \begin{cases} 
    1, & \forall (x, y) \in R_{1\oplus 2}^s \\
    0, & \text{otherwise}
\end{cases}
\]

(9)

\[
Q_d(x, y) = \begin{cases} 
    1, & \forall (x, y) \in R_{1\oplus 2}^d \\
    0, & \text{otherwise}
\end{cases}
\]

(10)

where \(1 \leq x \leq N\), and \(1 \leq y \leq M\). Thus, we can rewrite (7) and (8) as

\[
b(x, y) = b_s Q_s(x, y) + b_d Q_d(x, y),
\]

(11)

\[
c(x, y) = c_s Q_s(x, y) + c_d Q_d(x, y). \quad (12)
\]

Thus, \(\Phi = [a_1, a_2, a_3, a_4, a_5, a_6, b_s, b_d, c_s, c_d]^T\) which is the unknown vector can be optimized by minimizing the sum of the squared differences between \(I_1\) and \(I_2\), yielding a quadratic cost function

\[
\tilde{\Phi} = \arg \min_{\Phi} \mathcal{L}(\Phi; x, y) \quad (13)
\]

where \(\mathcal{L}(\Phi; x, y) = \sum_{x,y} e(\Phi; x, y)^2\), we can write

\[
e(\Phi; x, y) = I_x (a_1 x + a_2 y + a_5) + I_y (a_3 x + a_4 y + a_6)
\]

\[
+ I_2 - b_s Q_s I_1 - b_d Q_d I_1 - c_s Q_s - c_d Q_d \quad (14)
\]

\(^{1}\)To simplify the presentation in this paper, our model also excludes some other components, such as lens distortion correction in [3].
after substituting (3), (4), (11), and (12) into (1) and then obtaining the 1st-order truncated Taylor series expansion of $I_2$, where $I_x = \partial I_2(x, y) / \partial x$ and $I_y = \partial I_2(x, y) / \partial y$.

Then, we can estimate $\hat{\Phi}$ using the Gaussian-Newton method [4] to solve the non-linear least-squares problem in (13). Note that, $\hat{\Phi}$ is updated in each iteration $i$ in each level of resolution;

$$\hat{\Phi}_{i} = \hat{\Phi}_{i-1} + \Delta_i$$

In (14), $e(\hat{\Phi}; x, y)$ can be approximated by its 1st-order truncated Taylor series expansion yielding,

$$\mathcal{L} = \sum_{x,y \in I} \left[ e(\hat{\Phi}_{i-1}; x, y) + \Delta_i^T \frac{\partial}{\partial \hat{\Phi}} e(\hat{\Phi}_{i-1}; x, y) \right]^2$$

Setting the gradient of $\mathcal{L}$ w.r.t. $\Delta$ to zero, we obtain

$$-\sum_{x,y \in I} \left[ e(\hat{\Phi}_{i-1}; x, y) \frac{\partial}{\partial \hat{\Phi}} e(\hat{\Phi}_{i-1}; x, y) \right] = -\sum_{x,y \in I} \left[ \frac{\partial}{\partial \hat{\Phi}} e(\hat{\Phi}_{i-1}; x, y) \right] \frac{\partial}{\partial \hat{\Phi}} e(\hat{\Phi}_{i-1}; x, y)$$

We can rewrite (17) in matrix notation, such as

$$-H \Delta^T \Phi = (H H^T) \Delta$$

such that $H = [H_{1,1}, H_{1,2}, \ldots, H_{N,M}]$, where

$$H_{n,m} = [nI_x, mI_x, nI_y, mI_y, I_x, I_y, I_1Q_s, I_2Q_d, Q_s, Q_d]^T$$

and

$$K = [I_1(1, 1), I_1(1, 2), \ldots, I_1(N, M)].$$

Equations (18) to (20) can be used to perform one iteration for finding a solution of $\Delta$ to update $\hat{\Phi}$ in (15). However, $Q_s$ and $Q_d$ in (19) must also be determined. So, an iterative framework is proposed by first solving for the unknown parameters in $\Phi$ and then determining the similar and dissimilar illumination regions for $R_{\text{I}2}^{\text{IEC}}$ and $R_{\text{I}2}^{\text{IBD}}$, respectively. These two steps can then be iterated to refine the parameters $\Phi$ and refine the arbitrarily-shaped regions $R_{\text{I}1}^{\text{IEC}}$ and $R_{\text{I}1}^{\text{IBD}}$.

Assuming rough geometric registration for $I_1$ and $I_2$, the simplest method is to find the difference between the geometrically registered $I_1$ and $I_2$ and then use a clustering algorithm, such as the k-means algorithm [5] to split the resulting difference into two states/regions, for $R_{\text{I}2}^{\text{IEC}}$ and for $R_{\text{I}2}^{\text{IBD}}$, respectively. With noise or misalignment of the images, the difference between the roughly geometrically registered $I_1$ and $I_2$ may not allow for good clustering. An alternative to improve robustness could be to form neighborhood averages with the registered $I_1$ and $I_2$ and use the difference of these averages as input to the clustering algorithm. Other alternatives exist, but we will use the aforementioned approach for simplicity in this paper.

Our approach uses the model in (1) in cooperation with a coarse-to-fine scheme [6] to cope with large motions. A Gaussian pyramid is first built for both source and target images, and the full registration is estimated at the coarsest level. This estimate is used to warp the source image in the next finer level of the pyramid. The process of computing a new estimate is performed at each level of the pyramid. The affine and illumination parameters at each level of the pyramid are accumulated yielding a single final transformation.

3. EXPERIMENTS AND RESULTS

Two natural $120 \times 300$ images of a canyon, captured at different times of the day, are shown in Fig. 1(a,b). Note that Fig. 1(b) has slightly different orientation, global exposure, and includes different shadows compared to Fig. 1(a). In addition, our experiments include 51 simulated image pairs that are $500 \times 500$ pixels acquired from $3000 \times 3000$ pixel IKONOS satellite images. One example is shown in Fig. 1(c,d) showing the Pentagon and its surrounding area. In these 51 image pairs, the source images are subjected to simulated shadows (e.g., Fig. 1(e)). The target images are deformed using a priori known affine transformation parameters and then globally and locally exposed to intensity variations in different regions producing shadows (e.g., Fig. 1(d)). These simulated 51 image pairs are used since they allow ground truth verification of the proposed approach, which would be difficult with natural images such as with Fig. 1(a,b).

Our implementation runs on a 2 GHz Pentium IV Core 2 Duo, with 2 GB of RAM. In our experimental runs, we initialized $\Phi_0$ to $[1, 0, 0, 1, 0, 0, 1, 0, 0, 0]$ and stopped the iterations when the estimate updates become smaller than a predefined threshold (set to 0.1) or a maximum number of iterations (set to 10) is reached. Furthermore, the coarse-to-fine framework is exploited with 5 resolution levels. All 52 image pairs are tested with our approach as well as an approach where b and c in (1) are scalars, such as in [2, 3], which we will call the linear scalar model for illumination.

In Table 1 are results on the average of the absolute difference error between the estimated affine parameters computed using both our approach and the linear scalar model, compared to the a priori known transformation parameters of the 51 pairs of simulated images. Our proposed method provides an average 6.8 factor improvement on the 6 parameters compared to the linear scalar model. These results show that our model provides more precise motion estimates in the presence of arbitrarily-shaped local intensity variations of this type.

In addition, we also took the estimated parameters and generated image mosaics. These image mosaics are generated by applying the affine motion transform and illumination compensation to the target image, and then creating an overlaid average of the source and target images. For the test images in Fig. 1, image mosaics are created using both LSM and the proposed model and shown in Fig. 2.
The image mosaics following the proposed model in Fig. 2(b,d) are not only visually better, but as Table 2 shows, also have image quality measures that outperform those for Fig. 2(a,c) that use the linear scalar model. These image quality measures (SSIM [7] and PSNR) are used to compare the source image and target image after full geometric registration and illumination correction with the models. Table 2 also shows results for a random set of our data set.

### Table 1. The average of the absolute difference error between the estimated affine parameters computed by both models, and those of the a priori known parameters for the 51 pairs.

<table>
<thead>
<tr>
<th>Param.</th>
<th>Linear scalar model</th>
<th>Proposed model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$10.1514 \times 10^{-4}$</td>
<td>$1.6573 \times 10^{-4}$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$9.5571 \times 10^{-4}$</td>
<td>$1.4512 \times 10^{-4}$</td>
</tr>
<tr>
<td>$a_3$</td>
<td>$9.8017 \times 10^{-4}$</td>
<td>$1.5436 \times 10^{-4}$</td>
</tr>
<tr>
<td>$a_4$</td>
<td>$16.9127 \times 10^{-4}$</td>
<td>$2.2740 \times 10^{-4}$</td>
</tr>
<tr>
<td>$a_5$</td>
<td>$89.5918 \times 10^{-2}$</td>
<td>$7.6021 \times 10^{-2}$</td>
</tr>
<tr>
<td>$a_6$</td>
<td>$21.8505 \times 10^{-2}$</td>
<td>$7.8053 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

### Table 2. SSIM and PSNR for source and registered target images using (i) linear scalar model, and (ii) proposed model.

<table>
<thead>
<tr>
<th>SSIM</th>
<th>PSNR (dB)</th>
<th>Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>(ii)</td>
<td>(i)</td>
</tr>
<tr>
<td>Fig. 1-(a,b)</td>
<td>0.211 0.612</td>
<td>15.87 20.73</td>
</tr>
<tr>
<td>Fig. 1-(c,d)</td>
<td>0.739 0.863</td>
<td>17.74 26.12</td>
</tr>
<tr>
<td>Pair 3</td>
<td>0.680 0.910</td>
<td>19.32 26.15</td>
</tr>
<tr>
<td>Pair 8</td>
<td>0.704 0.902</td>
<td>20.23 26.35</td>
</tr>
<tr>
<td>Pair 12</td>
<td>0.718 0.895</td>
<td>21.73 27.06</td>
</tr>
<tr>
<td>Pair 29</td>
<td>0.749 0.898</td>
<td>23.23 28.01</td>
</tr>
<tr>
<td>Pair 38</td>
<td>0.721 0.905</td>
<td>21.59 27.68</td>
</tr>
<tr>
<td>Pair 41</td>
<td>0.713 0.899</td>
<td>20.08 27.86</td>
</tr>
<tr>
<td>Pair 47</td>
<td>0.813 0.905</td>
<td>22.79 29.67</td>
</tr>
</tbody>
</table>

registration accuracy compared to the linear scalar model.

## 5. REFERENCES


