IMAGE REGISTRATION UNDER LOCAL ILLUMINATION V ARIATIONS USING ROBUST BISQUARE $M$-ESTIMATION

Mohamed M. Fouad, Richard M. Dansereau and Anthony D. Whitehead

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

ABSTRACT

In this paper, we present a registration approach for images having arbitrarily-shaped locally variant illuminations. These variations tend to degrade the performance of geometric registration precision (GRP) and impact subsequent processing. Traditional registration approaches typically use a least-squares estimator that is sensitive to outliers. Instead, we propose using a robust bisquare $M$-estimator, as it differently penalizes the small and large residuals. The proposed approach shows clear improvements over competing approaches in terms of GRP and illumination correction, using simulated and real image pairs.

Index Terms— Image registration, bisquare $M$-estimation, illumination correction.

1. INTRODUCTION

Geometric image registration (GIR) of a set of images is a common pre-processing step in many applications, such as super-resolution, image stitching, and remote-sensing applications. For these and similar applications, sub-pixel accuracy in registration is necessary for satisfactory post-processing results.

The GIR process is impacted by the presence of locally variant illumination changes. In [1, 2], a global illumination model (GIM) is presented to relate the intensity levels of an image pair, where pre-defined neighborhoods with an imposed smoothness constraint are proposed in [1]. In [3], an affine illumination model (AIM) is given with triangular or quadratical region support. In [4], we present an approach considering arbitrarily-shaped local illumination variations (ASLIV) with two illumination levels (i.e., shadowed and non-shadowed). Note that the approaches in [1–4] use a least-squares estimator (LSE), which is very sensitive to outliers since its influence function (i.e., first derivative) is proportional to the residuals’ magnitude. In [5], we proposed replacing the LSE by a robust estimator, the Huber $M$-estimator, that has a greater resistance to outliers. The Huber function differently penalizes the residuals. However, its influence function assigns a constant weight to the high residuals, thus still impacting the GRP.

In this paper, we present a registration approach that can cope with multiple shading levels using an $M$-estimator with a robust objective function, such as the bisquare function [6]. The advantage of this function is that its influence function assigns zero weight to the high residuals, thus improving the GRP. The proposed approach is cast in an iterative coarse-to-fine scheme. A comparison with competing approaches is also presented.

This paper is organized as follows. In Section 2, we propose a registration approach using bisquare $M$-estimation. In Section 3, we develop the experiments using real and simulated data sets. Finally, conclusions are given in Section 4.

2. PROPOSED APPROACH WITH BISQUARE $M$-ESTIMATION

In this section, we propose a registration model to improve the GRP and obtain illumination correction, while using bisquare $M$-estimation.

2.1. Model

Consider that we have two $N \times M$ input images, $I_1$ and $I_2$, captured for the same scene at two different times. Both images may have distinct illumination regions with arbitrary shapes. We can formulate our extended registration model that relates the intensity levels of $I_1$ and $I_2$ as the matrices,

$$I_2(x_2) \approx B(x_1) I_1(x_1) + C(x_1),$$

where $x = (x, y)$ is a pixel location, and $\{I_1, I_2, B, C\} \in \mathbb{R}^{N \times M}$. As presented in [4, 5], the illumination variations, $B$ and $C$, are modeled as matrices

$$B = \begin{bmatrix} b_{11} & \ldots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{N1} & \ldots & b_{NM} \end{bmatrix}, \quad C = \begin{bmatrix} c_{11} & \ldots & c_{1M} \\ \vdots & \ddots & \vdots \\ c_{N1} & \ldots & c_{NM} \end{bmatrix}.$$ (2)

Many motion models could be used to model the motion in (1). However, we chose the 6-parameter affine motion model as shown in [1–4] (i.e., $x_2 = A(x_1)$, see Eqs.(3) and (4).
in [4]). In addition, in [4], the illumination changes were limited to two shading levels. In this paper, the approach is generalized such that each image $I_p$ can contain $V_p$ distinct levels of illumination.

To demonstrate the idea of the proposed model, we will start with the case that a *priori* known information of distinct shading levels is available. Fig. 1(a) shows an example of a masked image $I_1$ segmented into three distinct illumination levels: $L_1^1$, $L_1^2$, and $L_1^3$ (i.e., $V_1=3$). Similarly, Fig. 1(b) depicts another masked image $I_2$ segmented, as well, into a different set of distinct illumination levels: $L_2^4$, $L_2^5$, and $L_2^6$ (i.e., $V_2=3$). With a rough geometric registration for $I_1$ and $I_2$, an absolute image difference (AID), $I_{1,2}$, is created having a set of overlapping regions, $R_j$, such that $I_{1,2} = \bigcup_{j=1}^J R_j$ (see Fig. 1(c)), where $J$ refers to the number of resulting overlapping regions. Similarly, in a real imaging scenario, we propose a simple method of segmenting the AID of the roughly aligned images using, for instance, the $k$-means algorithm [7]. Thus, the illumination regions, $R_j$, are obtained: first estimated and then iteratively refined.

Note that the sum of the number of distinct illumination levels in both images (i.e., $V_1 + V_2$) does not express the number of illumination regions, $J$, as the latter depends on the former as well as the intersections among distinct levels. We’d like to stress that each resulting illumination region, $R_j$, is assumed to have its own constant brightness $b_j$ and contrast $c_j$. A special case was discussed in [4], where $J = 2$.

For pixel domain mathematical manipulations, each $R_j$ can be represented by a binary mask, $Q_j(x)$, such that

$$Q_j(x) = \begin{cases} 1, & \forall x \in R_j \\ 0, & \text{otherwise} \end{cases}, \quad \text{where } Q_j \in \mathbb{R}^{N \times M}. \quad (3)$$

Thus, we can constrain $B$ and $C$ in (1) to $J$ regions as

$$B = \sum_{j=1}^J b_j Q_j, \quad C = \sum_{j=1}^J c_j Q_j. \quad (4)$$

The following section presents an approach to estimate the unknown vector $\Phi = [a_1, \ldots, a_6, b_1, \ldots, b_J, c_1, \ldots, c_J]^T$, which contains the 6 parameters of brightness, $J$ for shading and $J$ for contrast, respectively.

### 2.2. Iterative Scheme Using Bisquare $M$-estimator

To estimate the unknown vector $\Phi$, we present an iterative framework using bisquare $M$-estimator

$$\min_{\Phi} \left\{ \mathcal{L} = \sum_{x} \Psi(E(\Phi, x); \alpha) \right\}, \quad (5)$$

where

$$\Psi(t; \alpha) = \begin{cases} s \left( 1 - \left( \frac{t}{\alpha} \right)^2 \right)^3, & |t| \leq \alpha \\ s, & |t| > \alpha \end{cases},$$

$$\Phi=x$$

where $\Psi(\cdot)$ denotes the bisquare function [6], $\alpha$ is a positive tuning threshold, and $s = \frac{1}{\alpha^2}$. A residual located at position $x$ can be given by $E(\Phi, x)$, such as

$$E(\Phi; x) = I_2(A(x)) - B(x) I_1(x) - C(x). \quad (7)$$

Following [8]'s approach, (5) can be rewritten as

$$\min_{\Phi} \left\{ \mathcal{L} = \sum_{x} \left[ s \left( 1 - \left( \frac{E(\Phi, x)}{\alpha} \right)^2 \right)^3 \cdot F + s \tilde{F} \right] \right\}, \quad (8)$$

where $(\cdot)$ refers to element-by-element matrix product and $E$, $F \in \mathbb{R}^{N \times M}$. The elements of the low-residual selective matrix, $F$, can be obtained by

$$f_{xy} = \begin{cases} 1, & E(\Phi; x, y) > \alpha \\ 0, & |E(\Phi; x, y)| \leq \alpha \end{cases}. \quad (9)$$

Whereas the high-residual selective matrix, $\tilde{F}$, is the binary complement of $F$. We can estimate $\Phi$ using the Gaussian-Newton method [9] to solve the non-linear minimization problem in (8). Note that $\hat{\Phi}$ is updated by $\Delta$ at each iteration, $i$, such that

$$\hat{\Phi}_i = \hat{\Phi}_{i-1} + \Delta_i. \quad (10)$$

Replacing $E(\cdot)$ in (8) by its 1st order Taylor series expansion, then $\mathcal{L}$ can be rewritten as

$$\mathcal{L} = \sum_{x} \left[ s \left( 1 - \frac{1 - \frac{1}{\alpha^2} E(\Phi_{i-1})}{\alpha^2} \right)^3 \cdot F + s \tilde{F} \right].$$

As with setting the gradient of $\mathcal{L}$ w.r.t. $\Delta$ to zero, four imaginary roots are rejected. Then, set the real root to zero to obtain

$$- \sum_{x} \left[ (E(\Phi_{i-1}) \nabla_{\Phi} E(\Phi_{i-1})) \cdot F \right]$$

$$= \Delta_i^T \sum_{x} \left[ (\nabla_{\Phi} E(\Phi_{i-1}) \nabla_{\Phi} E^T(\Phi_{i-1})) \cdot F \right]. \quad (11)$$

Fig. 1. With *a priori* known information, (a,b) two different sets of distinct illumination levels of two input images (i.e., $V_1=V_2=3$), yielding (c) a masked AID with seven arbitrarily-shaped regions, each has its own constant illumination (i.e., $J=7$).
Thus, we can write (12) in matrix notation as

\[-Y P = (Y H) \hat{\Delta}, \]

where \(H = [H_{11}, H_{12}, \ldots, H_{NM}], Y = [Y_{11}, Y_{12}, \ldots, Y_{nm}], P = [P_{11}, P_{12}, \ldots, P_{NM}], \) and \(Y_{nm} = f_{nm} H_{nm}, \)

\[H_{nm} = [n I_x, n I_y, m I_x, m I_y, -I_1 Q_1, \ldots, -I_1 Q_J, -Q_1, \ldots, -Q_J]^\top, \quad (14)\]

and \(P_{nm} = f_{nm} I_2(n, m), \quad 1 \leq n, m \leq N, M. \quad (15)\)

Eq. (13) can be used to perform one iteration for finding a solution of \(\hat{\Delta}\) to update \(\Phi\). As well, our approach uses a coarse-to-fine [10] framework to cope with large motions with \(r\) resolution levels. The convergence is achieved at the last resolution level, when either the cost function in (8) is updated by less than a predefined threshold, \(\varepsilon\), or a maximum number of iterations, \(i\), has been reached. The unknown parameters are iteratively accumulated yielding final estimates. The proposed approach is referred to as BM-ASLIV, where the superscript accounts for the \(6+2J\) parameters in \(\Phi\).

3. EXPERIMENTAL RESULTS

3.1. Data Set Description & Performance Evaluation

The image data sets used in the experiments include two categories: real and simulated. The first category involves eight real 600 × 600 LANDSAT satellite images [11] with unknown ground truth motions. The second category includes fifty 500 × 500 simulated pairs acquired from 3000 × 3000 pixel IKONOS satellite images for the Pentagon and its surroundings as shown in [4]. Each simulated pair is subjected to varying \(V_1\) and \(V_2\) levels of illumination, respectively, in different areas in both images according to the desired number of illumination regions \(J\); set to 3 and 4. Note that all the simulated pairs are directly presented to competing models without pixel quantization error.

For the real and simulated data sets, we compute the correlation between the overlapping area of the two registered images using the PSNR and normalized cross-correlation (NCC) [12]. As well, the average of absolute error (AAE) of the estimated affine parameters and their ground truth values is computed for the simulated data sets. Those AAEs are determined to show the mis-estimation in geometric registration using competing models as

\[\text{AAE}(a_t) = \frac{1}{D} \sum_{d=1}^{D} |a_t^d - a_t^d|, \quad t \in \{1, \ldots, 6\}, \quad (16)\]

where \(D\) is the number of pairs per data set (=50) and \(a_t^d\) denotes the parameter, \(a_t\), for pair \(d\). Note that the lower the value of the AAEs, the better the performance of the approach.

3.2. Implementation Setup & Discussion

Our implementation runs on a 2 GHz Pentium IV Core 2 Duo, with 2 GB of RAM. We compare the proposed approach, BM-ASLIV, to the GIM [1, 2], the approach in [4] (referred to as LS-ASLIV) and the approach in [5] (i.e., referred to as HM-ASLIV). The unknown vector, \(\Phi^0\), is initialized by an identity vector, such that the two input images are aligned and no illumination variations exist (i.e., \(\Phi^0\) is set to \([1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0]\) and to \([1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0]\) with ASLIV, ASLIV-based and ASLIV, ASLIV-based approaches, respectively). In addition, the number of resolution levels, \(r\), is set to 5, the predefined threshold, \(\varepsilon\), is set to 0.1 and the max number of iterations, \(i\), is set to 10. As well, the tuning threshold of the bisquare function, \(\alpha\), is set to 4.685 \(\sigma\) [6] when using the BM-ASLIV approach, where \(\sigma = \text{MAR}/0.6745\) and MAR is the median absolute residual at each iteration.

Table 1 shows that the proposed approach, BM-ASLIV, provides more precise motion estimates, in the presence of local illumination variations, over competing approaches for the “\(J=3\)” and “\(J=4\)” data sets. In addition, Tables 2 and 3 show that the BM-ASLIV outperforms competing approaches in terms of PSNR with a slight increase in computational time.

The gamma-corrected AIDs of the registered pairs, of the pair in Fig. 2-(a,b), are manifested in Fig. 2-(e,f,g,h) using GIM, LS-ASLIV, ASLIV and BM-ASLIV approaches, respectively (\(\gamma = 1.85\) for better visualization). Similarly, the normalized AIDs of the aligned pairs, for the pair in Fig. 2-(c,d), are shown in Fig. 2-(i,j,k,l) using GIM, LS-ASLIV, ASLIV, and BM-ASLIV approaches, respectively. Since the less the brightness, the more accurate the GRP, Fig. 2 shows the effectiveness of the BM-ASLIV approach versus competing approaches.

4. CONCLUSIONS

In this paper, we present an image registration model that can cope with images having arbitrarily-shaped local illumination variations, ASLIV. The proposed model is cast in an iterative coarse-to-fine approach using a robust bisquare M-estimator. The proposed approach is converged with a trivial initialization of the unknown parameters. Both real and simulated data sets are used in the experiments. The results show that the proposed approach achieves clear improvements in terms of geometric registration precision by an average increase of 90.13%, 23.45%, and 12.16% compared to the approaches in [1, 2], [4], and [5], respectively, with a reasonable increase in computational time.
Table 1. The AAEs ($\times 10^{-4}$) of the estimated affine parameters of the “$J=3$” and “$J=4$” data sets, using (i) GIM, (ii) LS-ASLIV, (iii) HM-ASLIV and (iv) BM-ASLIV approaches.

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>$\gamma$</th>
<th>$J=3$ data set</th>
<th>$\gamma$</th>
<th>$J=4$ data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$15.3858$</td>
<td>$1.7205$</td>
<td>$1.0253$</td>
<td>$1.1013$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$12.0018$</td>
<td>$1.0344$</td>
<td>$0.9789$</td>
<td>$1.1615$</td>
</tr>
<tr>
<td>$a_3$</td>
<td>$14.6081$</td>
<td>$1.1661$</td>
<td>$1.0859$</td>
<td>$1.1564$</td>
</tr>
<tr>
<td>$a_4$</td>
<td>$19.1138$</td>
<td>$2.0500$</td>
<td>$1.8699$</td>
<td>$2.1703$</td>
</tr>
<tr>
<td>$a_5$</td>
<td>$9.1907$</td>
<td>$2.0750$</td>
<td>$1.7235$</td>
<td>$2.1321$</td>
</tr>
<tr>
<td>$a_6$</td>
<td>$26.8872$</td>
<td>$2.1611$</td>
<td>$2.0933$</td>
<td>$2.1382$</td>
</tr>
</tbody>
</table>

Table 2. PSNR values for a random subset of the “$J=3$” data set using the (i) GIM, (ii) LS-ASLIV, (iii) HM-ASLIV and (iv) BM-ASLIV approaches.

<table>
<thead>
<tr>
<th>Pair</th>
<th>PSNR (dB)</th>
<th>Computational Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#10</td>
<td>22.53</td>
<td>29.02</td>
</tr>
<tr>
<td>#17</td>
<td>20.01</td>
<td>28.33</td>
</tr>
<tr>
<td>#28</td>
<td>21.76</td>
<td>29.42</td>
</tr>
<tr>
<td>#34</td>
<td>21.91</td>
<td>27.01</td>
</tr>
<tr>
<td>#38</td>
<td>23.47</td>
<td>28.63</td>
</tr>
<tr>
<td>#42</td>
<td>22.95</td>
<td>29.25</td>
</tr>
<tr>
<td>#48</td>
<td>21.94</td>
<td>29.69</td>
</tr>
</tbody>
</table>

5. REFERENCES


Table 3. NCC [12] values for the real pairs using (i) GIM, (ii) LS-ASLIV, (iii) HM-ASLIV and (iv) BM-ASLIV approaches.

<table>
<thead>
<tr>
<th>Pair</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.8819</td>
<td>0.9226</td>
<td>0.9245</td>
<td>0.9758</td>
</tr>
<tr>
<td>#2</td>
<td>0.9642</td>
<td>0.9789</td>
<td>0.9802</td>
<td>0.9919</td>
</tr>
<tr>
<td>#3</td>
<td>0.9799</td>
<td>0.9853</td>
<td>0.9870</td>
<td>0.9861</td>
</tr>
<tr>
<td>#4</td>
<td>0.9550</td>
<td>0.9644</td>
<td>0.9659</td>
<td>0.9729</td>
</tr>
</tbody>
</table>

Fig. 2. (a,b) A real pair (#B). (c,d) A “$J=3$” simulated pair (#34). Using competing approaches, (e,f,g,h) gamma-corrected AIDs of the pair #B, $\gamma = 1.85$, and (i,j,k,l) normalized AIDs of the pair #34.


