Digital Image Stabilization Based on Log-Polar Transform

Fei Yuan, Hong Zhang*, Ruiming Jia
Image Processing Center, BeiHang University, Beijing 100083, China
E-mail: flyuan@sa.buaa.edu.cn

Abstract

In this paper we present a novel log-polar transform based digital image stabilization algorithm to estimate large global transformations among consecutive frames. In our proposed algorithm, multi-resolution log-polar transform techniques in spatial domain are introduced as an initial motion estimation module to estimate arbitrary rotations, arbitrary translations and moderate scale changes. Then, a gradient-based nonlinear least squares optimization module is used to achieve sub-pixel accuracy. The experimental results show that the proposed algorithm can stabilize 15 frames per second in 320×240 image sequences with large interframe motions on a 3.0-GHZ Pentium 4 machine.

1. Introduction

Digital image stabilization (DIS) has become an integral component of some imaging devices, such as binoculars and commercial portable video recorders, which offers low cost and more flexibility than mechanical systems[1]. In addition, DIS can also be used as a front-end system in a variety of image analysis applications or simply as a visualization tool. The applications include the tele-operation of robotic vehicles, tracking moving objects from moving platforms and so on. The purpose of DIS is to remove the unwanted or unintended motion of the camera, so that the remaining contents of the sequence can be more easily viewed or further processed.

Our work is motivated by the problem of stabilizing image sequences captured by a CCD mounted on an airborne platform which undergoes large rotation and translation. The motions between the image sequences are very large. For example, the interframe rotation is as large as 72 degree. To handle the large displacements, we exploit a hierarchical approach for computational efficiency. First, we introduce multi-resolution log-polar transform techniques in spatial domain that serve as a preprocessing module to recover arbitrary rotations, arbitrary translations and moderate scale changes. Then, a gradient-based nonlinear least squares optimization technique is used to achieve sub-pixel accuracy. We conducted experiments to verify the performance of our proposed algorithm. The simulation results show that the estimation errors of scaling, rotational and translational motion parameters are no more than 0.01, 0.05 degree and 0.05 pixels respectively.

The rest of this paper is organized as follows. Section 2 describes our proposed algorithm based on multi-resolution log-polar transform. Section 3 presents some experimental results and Section 4 gives the conclusion of the paper.

2. Proposed Algorithm

Fig. 1 shows the outline of the proposed algorithm. The algorithm is composed of a motion estimation unit (ME) and a motion correction unit (MC). The ME unit consists of two stages. In the first stage, naming initial motion estimation, we introduce multi-resolution log-polar transform techniques in spatial domain that serve as a preprocessing module to recover arbitrary rotations, arbitrary translations and moderate scale changes. Then, a gradient-based nonlinear least squares optimization technique is used to achieve sub-pixel accuracy. The MC unit, composed of a motion compensation module and an image composition module, computes the global transformation and warps the current frame to generate the stabilized sequence. As the ME technique is the kernel of digital image stabilization, the remaining part of this section explains the proposed two-stage ME algorithm in detail.

---

1 Supported by national key laboratory (51476010105HK0101).
* Email: dmrzhang@buaa.edu.cn
2.1. Log-polar Transform

The log-polar transform (LPT) is a well-known space-variant image scheme used in computer vision systems[2]. Let \( I(x, y) \) be an image indexed by a Cartesian coordinate system, \( I^*(r, \theta) \) be the transformed image in the log-polar coordinate system. Then the log-polar transform with origin \((x_0, y_0)\) is:

\[
I^*(r, \theta) = L\{I(x, y); (x_0, y_0)\}
\]

(1)

Where, \( r = \log(\sqrt{(x-x_0)^2+(y-y_0)^2}) \) and \( \theta = \tan^{-1}\left(\frac{y-y_0}{x-x_0}\right) \).

LPT is useful for recovering rotation and scale because image rotation and scaling in the spatial domain results in translational shifts in the log-polar coordinate space. Fig. 2 shows an example. Fig. 2(a) is the original image, and Fig. 2(b) is a scaled, rotated and cropped version of Fig. 2(a). Their log-polar transforms are shown in Figs. 2(c) and 2(d). It is clearly seen that they differ by a phase shift. The best offset \( \Delta\theta \) and \( \Delta(\log r) \) which translates to scale and rotation in Cartesian space, can be easily found by cross-correlation in the log-polar space.

Fig. 1 Outline of the proposed algorithm

Fig. 2 Examples of Log-polar transform.

2.2. Initial Motion Estimation Based on Fast Multi-resolution Log-Polar Transform

From section 2.1, we can see the superior characteristic of log-polar transform in accounting for large rotation angles and scale translations. However, there is one difficulty that must to be solved: finding the corresponding origins of the two images from which the radial distance is measured. To solve this problem, Zokai S. etc[3] designed an algorithm to simultaneously find the best scale, rotation, and translation parameters for image registration. The algorithm first crops a circular template from the reference image and computes it’s log-polar transform, then for every position in the target image, a circular region is selected and compared against the circular template in log-polar space by normalized correlation. The maximum correlation position is used to determine the translation, rotation and scale parameters. Even though a coarse-to-fine multi-resolution framework is used to accelerate the process, Zokai’s algorithm is very time-consuming. Approximately 20 seconds is used to register a pair of 640×480 images on a 3.06-GHZ Pentium 4 machine in their implementation.

We introduce a rotation-invariant-features based fast search scheme to accelerate the search process.
while avoiding exhaustive search. The rotation-invariant-features based fast search algorithm positions the sliding window on areas of very high information content and cuts down the correlation sites by several dozens times. Our fast search scheme consists of two steps: feature extraction and circular block matching[6].

To deduce matching complexity and guarantee sufficient matching precision, two rotation-invariant features are extracted from input feature image. First, we smooth the input image by convolving it with a $5 \times 5$ Gaussian kernel operator. Then the smoothed image is convolved again with a $5 \times 5$ Laplacian kernel operator to generate the gradient image $G$. Two features are extracted by (2) from the circular block area (CBA) in gradient image $G$:

$$
\begin{cases}
  g_1 = \sum_{(x,y) \in \text{CBA}} (G(x,y) > 0) : 1 : 0 \\
  g_2 = \frac{1}{M} \sum_{(x,y) \in \text{CBA}} G(x,y)
\end{cases}
$$

(2)

Where, $M$ is the pixels number that the circular block area contains. $g_1$ means the numbers of strong edge pixels, and $g_2$ stands for the average gradient. Because Gauss operator[4] and Laplacian operator[5] are rotation invariant, $g_1$ and $g_2$ are rotation invariant.

We crop a circular template from the center of the reference image and compute its rotation-invariant-feature vector $\vec{g}_r = (g_1, g_2)^T$. Then perform circular block matching on the target image by two-dimensional motion search. The $l_1$-norm of the difference vector between the feature vectors is used as the matching criterion. The translation $(dx, dy)$ between the corresponding origins can be estimated by (3):

$$
(dx, dy) = \arg \min \{ \| \vec{g}_r - \vec{g}_t (\Delta x, \Delta y) \| \}
$$

(3)

Where, $\vec{g}_t = (\Delta x, \Delta y)^T$ stands for the feature vector of search block with offset $(\Delta x, \Delta y)$ from the center of target image.

To accelerate the searching process, we use the fast motion search program Lidong X.[6] proposed. If the difference between the feature $g_1$ of the circular template and that of a candidate block in search window is bigger than a threshold, the neighboring candidate blocks of this candidate block can be removed from the search window, because they are unlikely to be the best matching block.

After the translation $(dx, dy)$ between the corresponding origins is estimated, the log-polar correlation can be performed in a very small research window (e.g., $3 \times 3$). We implemented the proposed Log-Polar Transform algorithm in a three-level multi-resolution framework, and stabilized 15 frames per second in $320 \times 240$ image sequences on a 3.0-GHZ Pentium 4 machine.

### 2.3. Optimum Motion Estimation

Sub-pixel accuracy in digital image stabilization is very important for many subsequent applications. To achieve this goal, we introduce an efficient gradient-based nonlinear optimization algorithm in [7], called “Inverse Compositional Algorithm” (ICA), in which the time-consuming Hessian matrix of the cost function is not calculated in each iteration, but once in a pre-computation phase.

In our DIS system, we model the interframe motion by similarity model, and the parameterized set of allowed warps is:

$$
W(x; p) = \begin{bmatrix}
(1 + p_1) x & -p_2 y & +p_3 \\
-p_2 x & +(1 + p_1) y & +p_4
\end{bmatrix}
$$

(4)

Where, $x = (x, y)^T$ is a column vector containing the pixel coordinates, $P = (p_1, p_2, p_3, p_4)^T$ is the global motion parameter. The sum of squared differences (SSD) measure is used as the cost function, which is:

$$
E = \sum_x [R(W(x; \Delta p)) - I(W(x; p))]^2
$$

(5)

Where, $I(W(x; p))$ is the target image warped by an initial value $p$, $R(W(x; \Delta p))$ is the reference image warped by a incremental $\Delta p$, which can be estimated by Gauss-Newton algorithm:

$$
\Delta p = H^{-1} \sum_x [\nabla_R \frac{\partial W}{\partial p}]^T \left[ I(W(x; p) - R(x)) \right]
$$

(6)

Where, $H$ is the Hessian matrix of the reference image $R$:

$$
H = \sum_x [\nabla_R \frac{\partial W}{\partial p}]^T [\nabla_R \frac{\partial W}{\partial p}]
$$

(7)

Then updates the warp:

$$
W(x; p) \leftarrow W(x; p) \circ W(x; \Delta p)^{-1}
$$

(8)
(5) and (8) are iterated until the estimates of the parameters \( p \) converge. Because the initial value \( p \) is very close to the global minimum, accurate convergence can be reached fast, generally no more than four iterations.

3. Experimental Results

We evaluate the performance of the proposed algorithm on many synthetically generated test sequences. Fig. 3 shows two typical examples. Table 1 and Table 2 denote the quantitative analysis of the experimental results. It can be seen that the ME accuracy of the proposed algorithm is very high, with the estimation errors for scale, rotation and translation no more than 0.01, 0.05 degree and 0.05 pixels respectively.

![Scene One](image1)

![Scene Two](image2)

### Table 1 Quantitative analysis of Scene One

<table>
<thead>
<tr>
<th>Scene One</th>
<th>Ground truth</th>
<th>Estimated value</th>
<th>Estimated error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S ) (scale)</td>
<td>1.15</td>
<td>1.1483</td>
<td>0.0017</td>
</tr>
<tr>
<td>( \theta ) (degree)</td>
<td>120</td>
<td>120.0100</td>
<td>0.0100</td>
</tr>
<tr>
<td>( dx ) (pixel)</td>
<td>20</td>
<td>20.0132</td>
<td>0.0132</td>
</tr>
<tr>
<td>( dy ) (pixel)</td>
<td>-20</td>
<td>-19.9888</td>
<td>0.0112</td>
</tr>
</tbody>
</table>

### Table 2 Quantitative analysis of Scene Two

<table>
<thead>
<tr>
<th>Scene Two</th>
<th>Ground truth</th>
<th>Estimated value</th>
<th>Estimated error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S ) (scale)</td>
<td>0.85</td>
<td>0.8420</td>
<td>0.0080</td>
</tr>
<tr>
<td>( \theta ) (degree)</td>
<td>240</td>
<td>239.9640</td>
<td>0.0360</td>
</tr>
<tr>
<td>( dx ) (pixel)</td>
<td>-20</td>
<td>-19.9763</td>
<td>0.0237</td>
</tr>
<tr>
<td>( dy ) (pixel)</td>
<td>-20</td>
<td>-20.0258</td>
<td>0.0258</td>
</tr>
</tbody>
</table>

4. Conclusion and Future Work

The paper has presented a hierarchical DIS algorithm based on multi-resolution log-polar transform to recover large displacements among consecutive frames. The arbitrary rotations, arbitrary translations and moderate scale changes are first estimated by multi-resolution log-polar transform techniques in spatial domain. Then, a gradient-based nonlinear least squares optimization technique is used to achieve sub-pixel accuracy. The experimental results show that the proposed algorithm can fast and accurately stabilize image sequences with large interframe motions.

Future work will be done to recover large scale changes. Currently, even though the rotation-invariant-features based fast search scheme can accelerate the search process, it doesn’t work well under large scale changes.

References