Digital Image Stabilization Based on Phase Correlation

Ruiming Jia, Hong Zhang, Lei Wang  
Image Processing Centre  
Beihang University  
Beijing, China  
jamin_han@163.com

Abstract—In this paper, we present a method for digital image stabilization based on Fourier-Mellin transform and phase correlation. We acquire the rotating angle and scaling factor firstly by phase correlation between Fourier-Mellin transform images of the reference and observed images. After rotating and scaling the observed image, implement the phase correlation again to compute the spacial translation. Because of spectral periodicity, the correlation between Fourier-Mellin transform images will be weaken and maybe lead to wrong result when the rotating angle is close to 90°. So we add a course search before phase correlation to avoid this situation. And we use a smooth window to reduce the noise in phase correlation. Furthermore, a coordinate mapping chart is recommended to avoid the unnecessary computation cost during creating log-polar image. The experiment results show the proposed method can acquire accurate motion parameters between two adjacent frames.

Keywords—Image stabilization, Fourier-Mellin Transform, Phase correlation.

I. INTRODUCTION

The purpose of digital image stabilization is to reduce or remove the unintended negative influence of the camera movement on images captured. So the stabilized contents of the sequence can be appropriate for the further processing. The stabilization technique can be utilized for the binoculars and commercial portable video recorders by offering low cost and more flexibility than mechanical system [1], or as a front-end system for image analysis applications. The applications include the tele-operation of robotic vehicles, tracking moving objects from moving platforms [2] and so on.

This paper presents an improved method of image stabilization based on Fourier-Mellin transform and phase correlation which computes transformation parameters between a reference image and its rotated, scaled and translated image. The image stabilization method based on Fourier-Mellin Transform (FMT) and phase correlation is a global method, which use phase shifts between reference and observed images to determine the image motion parameters [3][4]. This paper proposed several ways to improve the algorithm efficiency and reduce the computation at the same time.

This paper is organized as follows. First, in Section 2, we introduce the principle of FMT and phase correlation, and then give the basic method. Section 3 describes our optimization algorithm in detail, including the ways to improve the detecting accuracy, reduce the noise and the computation cost. Section 4 shows the experiment results of the artificial and real image pairs. Section 5 gives the conclusion of this paper.

II. PROPOSED METHOD

Consider the observed image \( s(x,y) \) is a rotated, scaled and translated (RST) replica from the reference image \( f(x,y) \). Suppose the rotation angle is \( \phi \), the scaling coefficient is \( \alpha \), and the spatial shift is \( (x_0, y_0) \). Now we introduce the Fourier-Mellin Transform firstly.

A. Fourier-Mellin Transform

Fourier-Mellin Transform (FMT) is a technique which turns the rotation and scaling in spatial domain into phase shifts[6].

In the spatial domain \( s(x,y) \) can be described by

\[
\begin{align*}
  s(x,y) &= f(\alpha x \cos \phi + y \sin \phi + x_0, \\
  &\quad \alpha y - x \sin \phi + y \cos \phi + y_0)
\end{align*}
\]

(1)

After Fourier transform, in the frequency domain the relation in (1) is described by

\[
\begin{align*}
  S(u,v) &= \alpha^{2u} e^{-2\pi i u x_0} \cdot F(\alpha^v (u \cos \phi + v \sin \phi)), \\
  \alpha^v &= (-u \sin \phi + v \cos \phi)
\end{align*}
\]

(2)

Consider the magnitude of \( F(u,v) \) and \( S(u,v) \), we have

\[
|F(u,v)| = |\alpha| \cdot |F(\alpha^v (u \cos \phi + v \sin \phi))|
\]

(3)

The spectral magnitude is translation invariant.

From (2) we can see that a rotation of the image in the spatial domain rotates the same angle in spectral magnitude, and a scaling by \( \alpha \) scales the spectral magnitude by \( |\alpha| \) in the spectra domain. The rotation and scaling will change into translation after mapping the spectral magnitude into the logarithmic-polar coordinate \((c, \theta)\).

Consider the Log-Polar Transform (LPT).

\[
\begin{align*}
  u &= e^c \cos \theta \\
  v &= e^c \sin \theta
\end{align*}
\]

(4)

From (3) and (4), the LPT of spectral magnitude is shown as (5). The log-polar mapping of the spectral corresponds to a physical realization of the FMT [7][8]. In this log-polar domain presentation, both rotation and scaling are turned into translations in parameter domain.

Supported by the National Natural Science Foundation of China (Grant No. 60872079) and National Key Laboratory on Optical Features of Environment and Target Foundation (No. 9140C6105020805).
\[ |F| = |F(e^{i\phi} f(x, y))| = |F(e^{i\phi} f(x, y))| \]

\[ F(e^{i\phi} f(x, y)) = e^{-i\phi} F(e^{i\phi} f(x, y)) \]

B. Symmetric Phase-Only Matched Filtering

The SPOMF [4] is an improved method based on phase correlation to calculate the translation parameters between two images.

Consider \( s(x, y) \) differs from reference image \( f(x, y) \) only by a displacement \((x_0, y_0)\), and then

\[ s(x, y) = f(x + x_0, y + y_0) \]

In the frequency domain, the equation becomes

\[ S(u, v) = e^{-i\phi} \frac{F(u, v)}{F'(u, v)} \]

The cross-power spectrum is defined as

\[ Q(u, v) = \frac{S(u, v)}{S(u, v)} \frac{F(u, v)}{F'(u, v)} \]

The phase of the cross-power spectrum is equivalent to the phase difference between two images [5] [6].

The inverse Fourier transforms of \( Q(u, v) \) is a Dirac \( \delta \)-function centered at \((x_0, y_0)\) with a sharp peak.

C. Computing Motion Parameters by FMT and SPOMF

As Fig.1 shows, the process of computing motion parameters is divided into two steps.

1. Corrected image is acquired by rotating and scaling the observed image, and SPOMF is implemented again between the corrected image and reference image to figure out the translation.

   Step 1: Firstly we compute the FMT of observed and reference images, and then SPOMF is applied to the two FMT images for determining the rotation angle and scaling factor.

   Step 2: Corrected image is acquired by rotating and scaling the observed image, and SPOMF is implemented again between the corrected image and reference image to figure out the translation.

III. IMPROVED ALGORITHM

The improved algorithm shows as Fig. 2. Firstly a Gauss Window is applied to the images along each coordinates for reducing the noise due to sampling and truncation. Because of spectral periodicity, the correlation between Fourier-Mellin transform images will be weaken and maybe lead to wrong result when the rotating angle is close to 90°. So we add a coarse search before phase correlation to avoid this situation. After SPOMF operation, a Gauss smooth filter is applied to weaken the noise. Otherwise, a sub-pixel method is adopted to improve the detecting precision.

![Image](image.png)

Figure 2. The flow chart of proposed method

A. Gauss Window Applying

There are some defects in the procedure of FMT and SPOMF. For example, several direct and inverse Fourier transforms are implemented in the algorithm, and the sampling and truncation bring some kind of noise in the result of SPOMF. So a Gauss window is applied to reduce the negative effect. For comparing the effect, we define the SNR as flow,

\[ \text{SNR} = \frac{C_{\text{SNR}} - M}{\sigma} \]

\( C_{\text{SNR}} \) presents the maximum of the pulse, the \( M \) is the mean value and the \( \sigma \) is standard deviation.

Comparing the results in Fig.3, it shows that the result in (b) has a higher SNR than (a).

2. Sub-pixel Accuracy Obtaining

The sub-pixel estimation [7] supposes that the real matching position is an average of the neighborhood around impulse. Firstly, the phase correlation result of SPOMF needs to be smoothed, as the Fig.4 shows.
Here we choose 5x5 Gauss smooth filter to process the output image of SPOMF. From Fig.4 (b), it is clear that the noise peak is weakened, so we can segment the impulse neighbor by a proper threshold.

Considering the correlation result image is \( q(x, y) \), and there are \( N \) points having been segmented. The sub-pixel coordinates can be calculated as follow.

\[
\begin{align*}
\mathbf{\xi} &= \frac{\sum_{i=0}^{N} \sum_{j=0}^{N} (x+i) \cdot q(x+i, y+j)}{\sum_{i=0}^{N} \sum_{j=0}^{N} (x+i)} \\
\mathbf{\eta} &= \frac{\sum_{i=0}^{N} \sum_{j=0}^{N} (y+j) \cdot q(x+i, y+j)}{\sum_{i=0}^{N} \sum_{j=0}^{N} (y+j)}
\end{align*}
\]

(10)

Here, \( (\mathbf{\xi}, \mathbf{\eta}) \) is the sub-pixel coordinates of the segmented points.

![Figure 3. (a) The SPOMF result between the original images, SNR=13.4. (b) The SPOMF result after adding a Gauss Window, SNR=19.2](image)

![Figure 4. (a) The correlation result of SPOMF. (b) The smoothed result with 5x5 Gauss smooth filter](image)

**B. Coarse Search before SPOMF**

For the periodicity of the Fourier spectral magnitude, only half of the magnitude field is used in the FMT and the image of FMT has a height of 0-180\(^\circ\) degree range. So when the rotating angle is close to 90\(^\circ\), as the Fig.5 shows, the pulse in SPOMF result appears ambiguous, because the correlation between the two images is lower.

The low SNR will lead to wrong result. Here a coarse search method is applied to give an approximate shift value between the two FMT images. As the Fig.6 shows, when rotation angle is \( \lambda \), the FMT image has a \( \lambda \) shift of cycle translation.

![Figure 5. The SPOMF results with different rotating angles, SNR\(_{0\degree}=20.1, SNR\(_{20\degree}=12.4, SNR\(_{75\degree}=11.3, SNR\(_{85\degree}=9.7 \).](image)

![Figure 6. The cycle translation of the FMT image with \( \lambda \) shift.](image)

The coarse search adopts the SAD (Sum of Absolute Difference) to detect the approximation of \( \lambda \). Suppose the FMT of reference \( r(x, y) \) and FMT of observed image \( f(x, y) \) with resolution of \( M \times N \). The replica \( g(x, y) \) of cycle translation of \( f(x, y) \) is shown as follow,

\[
\begin{align*}
g(x, y) &= \begin{cases} 
 f(x, y+\lambda) & y < M - \lambda \\
 f(x, M-y) & y \geq M - \lambda 
\end{cases}
\end{align*}
\]

(11)

The SAD is defined as

\[
SAD = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |g(i, j) - r(i, j)|
\]

(12)

Set \( \lambda = 1, 2, \ldots, M \) to calculate all of the SAD, and the minimum of the SAD presents the approximation translation between two of FMT images. Fig.7 gives an example.

![Figure 7. The SAD between the FMT images, the observed image’s rotating angle is 85\(^\circ\) and the coarse search result is 81\(^\circ\)](image)
As Fig.7 shows, the coarse search figures out the approximate value 81°. Then offset the FMT image 81 pixel by cycle translation and acquire the corrected image \( g(x,y) \). The translation between \( g(x,y) \) and \( r(x,y) \) is only around 4 pixel, and precise result will be acquired by phase correlation.

Suppose the coarse search give the approximation translation \( \lambda_y \). Correct the observed FMT image by cycle translation and apply SPOMF to acquire the precise translation \( \hat{\lambda}_y \). So the last translation of FMT is

\[
\hat{\lambda} = \lambda_0 + \hat{\lambda}_y \quad (13)
\]

C. Update the Reference Image Properly

In process of image sequence stabilization, the reference image needs to be replaced regularly to reduce the accumulative error. Consider images 1-5 in a sequence, as the Fig.7 shows.

![Figure 8.](image)

The updating way in Fig. 8(b) will have lower accumulative error than (a), as the table I. shows.

<table>
<thead>
<tr>
<th>Processing 15 frames totally</th>
<th>A step per 1 frame</th>
<th>A step per 3 frames</th>
<th>A step per 5 frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulative Error</td>
<td>8.6°</td>
<td>3.5°</td>
<td>2.7°</td>
</tr>
</tbody>
</table>

D. Reducing Computation Cost

1) Using pre-calculated map for LPT

If giving the image resolution and the detecting precision, the relationship between Cartesian and log-polar coordinate has been determined. During LPT, it’s not necessary to calculate log-polar coordinates every time. A transform map between Cartesian and log-polar should be created before processing to avoid the repeat computation.

2) Down-sampling image if possible

There is a relationship between the image resolution and the precision of the detecting parameters. As Fig.9 (a) shows, there is a relationship between the image resolution and log-polar coordinates every time. A transform map between Cartesian and log-polar should be created before processing to avoid the repeat computation.

![Figure 9.](image)

D. Reducing Computation Cost

1) Using pre-calculated map for LPT

If giving the image resolution and the detecting precision, the relationship between Cartesian and log-polar coordinate has been determined. During LPT, it’s not necessary to calculate log-polar coordinates every time. A transform map between Cartesian and log-polar should be created before processing to avoid the repeat computation.

2) Down-sampling image if possible

There is a relationship between the image resolution and the precision of the detecting parameters. As Fig.9 (a) shows, the minimum detecting angle by one pixel is \( \theta \) in an image with \( M \times M \) resolution.

Fig. 9 (b) shows that the minimum detecting angle changes when the image resolution varies. It is clear that the theoretic precision has been determined if the image resolution is defined. For example, the minimum detecting angle \( \theta \) is around 0.8° when image resolution equals 100.

IV. EXPERIMENTAL RESULTS

A. Using Artificial Image Pairs

Firstly we use the synthetical image to evaluate the algorithm. The input image is obtained by rotating 15°, scaling 1.25, translating (-10, 15) and adding Gauss white noise with SNR=10dB. The experiment result is shown in Fig.10.

![Figure 10.](image)

Figure 10. The motion estimate result of the artificial image pairs: (a) Reference image; (b) the input image; (c) the phase correlation result of FMT of (a) and (b), angle =15°; (d) the phase correlation result for detecting spatial translation, shift=(-9, 15).

According to the experiment result, the scaling has an error around 0.0007, which leads to the translation error of one pixel on x axis.

B. Using Real Image Pairs

Experiment result of real image pairs is shown in Fig.11. We can find that the noise in phase correlation result is stronger than the artificial experiment result, because the real observed image has extra textures.
Figure 11. The experiment result of real image pairs: (a) the reference image; (b) the observed image; (c) the phase correlation of FMT of (a) and (b); (d) the phase correlation result for detecting spatial translation; (e) the corrected image; (f) the residual image between (a) and (e).

From experimental results, the method is proved effective when images in sequence have a motion of rotating, scaling and translating.

REFERENCES