Superresolution Reconstruction of Video Sequence Using a Coarse-to-Fine Registration and Optimal Interpolation Strategy

Zunshang Zhu1,*, Xiaolin Liu2, Yun Yuan1, Xianwei Zhu1 and Qifeng Yu1

1 College of Aerospace Science and Engineering & Hunan Key Laboratory of Videometrics and Vision Navigation, National University of Defense Technology, Changsha, China
2 College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha, China
* Corresponding author E-mail: zzs2623@nudt.edu.cn

Received 29 Jun 2012; Accepted 27 Mar 2013

DOI: 10.5772/56451

© 2013 Zhu et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract In this paper, a novel video superresolution (SR) reconstruction approach using a coarse-to-fine registration and optimal interpolation strategy is proposed. Firstly, initial correspondence between two consecutive frames is identified by scale invariant feature transform (SIFT) features under the homography transformation model. Then, a least square matching (LSM) method is employed to achieve a high-precision estimation of homography. Finally, optimal interpolation using the iterative back projection-based deblurring method was designed to obtain a high-resolution grid interpolation and to improve the visual quality of the final SR results. The experiment results demonstrate that the proposed method provides an available SR approach in engineering applications.

Keywords Superresolution Reconstruction, Image Registration, Homography, Optimal Interpolation

1. Introduction

In recent years, rapid improvements in video and image processing technologies have taken place. High resolution (HR) images commonly contain more details of the scene, which are significant for analysis and post-processing. However, in practical applications, caused by the limitation of optical elements’ volume and the bandwidth of the communication channel in the video surveillance system, image resolution cannot meet our requirements. Designing SR techniques to obtain an HR image from observed multiple low-resolution (LR) images have been studied in various ways. SR generates the HR image by combining multiple LR images with sub-pixel shifts from each other, since each LR image can provide similar but different information of the scenes.

The major advantage of the SR approach is that flexible, cheaper and existing LR imaging systems can still be utilized, especially for some systems where equipment is
already fixed. Some video surveillance systems worked in the ‘push boom’ work style and provided many overlapped images for the target region. Therefore, using the SR technique to improve the resolution of the targets can be considered.

Most SR algorithms are restricted to planar scenes with the assumption of the homography or affine transformation model. In other words, image registration with sub-pixel accuracy is a prerequisite to SR algorithms; otherwise, inaccurate registration may degrade the reconstructed image quality severely. High precision registration has directly caused several difficulties when applying these methods in real video applications. The first difficulty is the location precision of corresponding points under the different views; the second difficulty is the lack of an accurate model to describe the transformation between the LR images under the conditions of a moving camera and scene variations; the third difficulty is restoring the HR image from multiple registered frames with certain extent errors.

In this paper, we aim to excavate the validate information from multiple video frames with overlapped regions to the maximum extent, and propose a solution to deal with the non-strict transformation between video frames, such as terrain undulation in satellite video and depth variation in a normal video. We mainly focus on the two special issues on video sequence SR: on the one hand, we aim to diminish the errors caused by homography estimation, on the other hand, we use an optimal interpolation method to reduce the disturbance of registration errors and noises.

The paper is organized as follows. Section 2 gives a review of the registration, interpolation and restoration algorithms used in the literature on SR. Section 3 presents the proposed method based on coarse-to-fine registration and optimal interpolation. The experiment results and discussions are given in Section 4. Section 5 concludes the paper.

2. Related work

Surveys of SR methods can be found in much of the literature such as [1,2]. Most SR methods exploit the high-frequency information of the LR-aliased images to reconstruct the HR image. These methods can be classified into two categories: frequency domain approaches and space domain approaches [3]. Tsai and Huang [4] proposed a frequency domain video SR approach, which exploits the shift property of the Fourier transform and developed a set of system equations in the frequency domain that relates the HR video frames to the observed LR video frames. The performance of frequency-domain video SR is usually limited by the global translation model. The space domain approaches can treat the complicated motion model and deal with the process of corresponding interpolation, filtering and resampling; their observation model involves global and partial motions, optical blur, infra-frame motion blur, spatial variable point spread function, etc. Major algorithms of the spatial domain SR include the non-uniform interpolation method [5,6], the probabilistic method [7], and the MAP/POCS combination method [8], etc. Compared with the frequency-domain approaches, the space domain approaches can yield better results because of their high inclusiveness and flexibility.

Precise image registration is a critical procedure of SR. Registration algorithms mainly fall into two groups: region-based and feature-based registration. Region-based registration methods attempt to iteratively estimate the movement by minimizing an error function based on the intensity difference in the overlapping area. Region-based registration methods have the advantage of being able to deal with complex motion with higher accuracy, but the computational cost for the entire image is too high for real-time implementation. Feature-based methods are based on correspondences between points, lines or other geometry primitives. A typical approach is to match extracted Harris corners utilizing a normalized cross-correlation of the local intensity values. Recently there has been great progress made in the employment of invariant features [9,10] for image registration, for instance, Lowe’s scale invariant feature transform (SIFT) features[11], which are geometrically invariant under similarity transforms and invariant under affine changes in intensity. These features have been proven to be robust and reliable. Feature-based techniques are fast, but fail to provide accurate results when the illumination or viewpoints significantly change. In remote-sensing image registration, the least square matching (LSM) [12] method based on affine transformation is frequently adopted, which conducts calculations using information within the image window and produces satisfactory matching results in the case of approximate planar scenes. The precision can reach 1/10-pixel and even higher, but this iterative algorithm needs a proper initial estimation to guarantee the convergence.

Another major issue regarding SR algorithms is their excessive dependence on an accurate registration. If the displacements between images are inaccurately estimated, the SR algorithm may lead to image degradation and not improvement. Some recent SR algorithms have been devised to improve the robustness to cope with such errors. Wang and Qi [13] proposed a robust SR algorithm based on the Kalman filter. The registration uncertainties are included in the filter equations, based on the error estimation of the dynamical model due to the registration error. In [14], Lee and Kang defined a constrained least squares problem as a signal
dependent regularization functional, which incorporated the registration errors. Bose and Ahuja [15] used moving least squares to reconstruct the HR image, and proposed a technique to adaptively select the scale and order parameters for estimating the image intensity at each point of the HR grid. Costa [16] made a statistical analysis of the mean square algorithm, which is applied to the image sequence with global translation.

Rew [17] used the least square fitting method with polynomial function to enhance the illumination uniformity. However, the least square fitting method is known to be sensitive to outliers. To deal with this, RANdom SAmple Consensus (RANSAC) was introduced by Fishler and Bolles [18] as the most popular robust estimation technique to estimate transformations in datasets containing outliers. This strategy has been widely used in the curve and surface fitting process, since it is reliable and easy to implement. It operates in a hypothesize-and-verify framework, and given a set of coarse correspondence, RANSAC randomly samples a minimal subset to hypothesize a geometric model. This model is then verified against the remaining correspondences. This process is iterated until a certain criterion is met. The original RANSAC algorithm is proposed for the robust fitting of models with a good tolerance to outliers. We extended this method to solve the interpolation problem with the assumption of an intensity distribution model.

3. The proposed approach

In this paper, we aim to deal with the distance of the scene from the camera which is much greater than the motion of the camera between views, and the parallax effects caused by the three-dimensional character of a scene which is negligibly small, such as in satellite imagery. We adopted the homography model to describe the transformation between LR frames, and employed an optimal interpolation strategy to tackle the limitations of the unified transformation model. Firstly, initial correspondence between two consecutive frames is obtained by using scale invariant feature transform (SIFT) features under the homography transformation model. Then, a least square matching (LSM) method is employed to achieve a high-precision estimation of homography. Finally, optimal interpolation with the iterative back projection-based deblurring method was designed to obtain a high-resolution grid interpolation and to improve the visual quality of the final SR results.

The proposed pipeline contains several processing steps from registration of the LR images to reconstruct an HR image. Some pre-processing is also performed to enhance the quality of the video sequence. The basic steps of our algorithm are depicted in Fig.1 and described in more detail in the remaining parts of this section.

![Figure 1. The procedure of SR method](image)

3.1 Registration

SIFT-based image registration is invariant to translation, rotation, scaling and illumination changes in images. However, the detection of the extrema is in the discrete space. The extrema may not be the real extrema in the continuous space and this is especially so for the boundaries of objects, where the location of the keypoints will move slightly under different viewpoints. Inspired by the region-based image registration techniques, we exploited the local region information around the SIFT points using an iterative method to obtain higher accuracy match point locations. In real data experiments, we observed that the accuracy of the SIFT registration method is not sufficient for SR, therefore, a LSM-based refinement stage is performed to achieve the required accuracy. Our coarse-to-fine registration procedure is shown in Fig.2.

![Figure 2. The procedure of registration](image)

A. Initial registration and homography estimation

The SIFT algorithm is a popular candidate for extracting the interest points which are invariant to translation, rotation, scaling and illumination changes. It firstly constructs a Gaussian scale-space pyramid from the input image while also calculates the gradients and difference of Gaussian (DoG) images at these scales. Interest points are detected at the local extrema within the DoG scale space. Once multiple keypoints have been detected at different scales, the image gradients in the local region around each feature point are encoded using orientation histograms and represented in the form of a rotationally invariant feature descriptor. The initial correspondences can be obtained by the SIFT detector and matching approach.
We assume that the transformation between image sequence can be approximated to the planar homography. This is appropriate to a freely moving camera viewing a very distant scene. In all cases, it is assumed that the images are obtained by a perspective pin-hole camera.

A point is represented by a homogeneous coordinate, such as \((x,y,1)\). Conversely, the point \((x_1, x_2, x_3)\) in the homogeneous coordinate corresponds to the inhomogeneous point \((x_1/x_3, x_2/x_3)\).

Under a planar homography, points are mapped as:

\[
\begin{bmatrix}
    x' \\
    y' \\
    1
\end{bmatrix} = 
\begin{bmatrix}
    h_{11} & h_{12} & h_{13} \\
    h_{21} & h_{22} & h_{23} \\
    h_{31} & h_{32} & h_{33}
\end{bmatrix} 
\begin{bmatrix}
    x \\
    y \\
    1
\end{bmatrix}
\]  
(1)

or

\[
x' = Hx
\]  
(2)

and the \(H\) is called the homography matrix. The planar homography has eight degrees of freedom. Each point correspondence generates two linear equations for the elements of \(H\), hence four correspondences are enough to solve the homography directly. If more than four points are available, a least squares solution can be found by linear methods. The method of RANSAC was used to obtain a unique estimate of \(H\) from noisy data. Full details are given in Hartley and Zisserman [19].

B. Refine homography with least square matching

The least squares matching (LSM) method attempts to match windows of pixels by establishing a correspondence between them and by minimizing the difference of their intensity values. The precision can reach 1/10-pixel and even higher. The LSM method is appropriate for video sequences which have some overlap regions and enough texture information, especially for those SIFT keypoints.

Because the distance of the scene from the camera is relatively large, the transformation between the local region around the SIFT keypoints can be approximated to an affine transformation. The affine transformation is defined as:

\[
\begin{bmatrix}
    x' \\
    y' \\
    1
\end{bmatrix} = 
\begin{bmatrix}
    a_0 & a_1 & a_2 \\
    b_0 & b_1 & b_2 \\
    1
\end{bmatrix} 
\begin{bmatrix}
    x \\
    y \\
    1
\end{bmatrix}
\]  
(3)

where \((x,y)\), \((x',y')\) are the coordinates in image pairs and \(a_0, a_1, a_2, b_0, b_1, b_2\) are affine transformation parameters. For satellite video sequences, the time interval and location variation between two frames is small, and the illumination changes slowly and continuously. Considering the noise and geometry distortion, the intensity variation between image pairs can be approximately described as:

\[
g_i(x,y) + n_i(x,y) = g_i(x',y') + n_i(x',y')
\]  
(4)

where \(n_1, n_2\) are the random noise and \(g_1, g_2\) are the intensity values in image pairs.

The LSM method estimated the movement by iteratively modifying the affine transformation and radiometric shift parameters to minimize the intensity difference in the correspondence area. The error equation for each pixel can be described as:

\[
\Delta g = g_i(a_0x + a_1y + a_2, b_0x + b_1y + b_2) - g_i
\]  
(5)

After linearization, the equation became:

\[
v = c_0da_0 + c_1da_1 + c_2da_2 + c_3db_0 + c_4db_1 + c_5db_2 - \Delta g
\]  
(6)

where \(da_0, \ldots, db_2\) were the parameters' deviation and the initial values can be \([a_0, a_1, a_2, b_0, b_1, b_2] = [0, 1, 0, 0, 1]\), \(\Delta g = g_i(x,y) - g_i(x,y)\) was the difference of intensity.

The parameters of the error equation are given as:

\[
[c_0, c_1, c_2, c_3, c_4, c_5] = [x \frac{\partial g_2}{\partial x}, y \frac{\partial g_2}{\partial x}, x \frac{\partial g_2}{\partial y}, y \frac{\partial g_2}{\partial y}, \frac{\partial g_2}{\partial y}]
\]  
(7)

According to the theory of the least square matching, we optimized the objective function:

\[
\sum vv = \text{min}
\]  
(8)

For all the pixels within the match window, Eq.(6) can be rewritten into a matrix form .

\[
CX = L
\]  
(9)

where

\[
C = \begin{bmatrix}
    x \frac{\partial g_{21}}{\partial x} & y \frac{\partial g_{21}}{\partial x} & \frac{\partial g_{21}}{\partial y} & \frac{\partial g_{21}}{\partial y} \\
    x \frac{\partial g_{22}}{\partial x} & y \frac{\partial g_{22}}{\partial x} & \frac{\partial g_{22}}{\partial y} & \frac{\partial g_{22}}{\partial y} \\
    \vdots & \vdots & \vdots & \vdots \\
    x \frac{\partial g_{2n}}{\partial x} & y \frac{\partial g_{2n}}{\partial x} & \frac{\partial g_{2n}}{\partial y} & \frac{\partial g_{2n}}{\partial y}
\end{bmatrix},
\]

\[
X = [da_0, da_1, da_2, db_0, db_1, db_2]^T,
\]

\[
L = [g_{11} - g_{21}, g_{12} - g_{22}, \ldots, g_{1n} - g_{2n}]^T.
\]
If \( n > 6 \), Eq. (9) was an over-determined equation and the solution of \( X \) can be computed by conventional linear least squares:

\[
X = (C^T C)^{-1} C^T L
\]  

(10)

3.2 RANSAC interpolation

Once the relative motion information is estimated with sub-pixel accuracy, it requires interpolation when the fractional unit of motion is not equal to the HR grid. Simple methods, such as zero-order hold, bilinear or the least square fitting interpolation, have been used in the literature based on the assumption that the relative shifts were known precisely, and the registration errors can be ignored. However, in real applications, these methods tended to smoothen and blur image details, as well as performed inefficiently in the presence of noise.

Considering we use planar homographies to represent the geometric transformations between any LR images to the reference LR image, the re-projection errors of each pixel in the overlap region are different. These points with larger re-projection errors are regarded as outliers. In general, the bilinear interpolation and the least square fitting methods are known to be sensitive to outliers, so it was essential for us to select the points with higher precision for an accurate and robust interpolation.

Fischler proposed the RANSAC method for robust fitting of models. It is robust in the sense of good tolerance to outliers in the coarse data and capable of interpreting and smoothing data containing a significant percentage of gross errors. The algorithm has been widely applied to parameter estimation problems in computer vision, such as feature extraction and matching. In order to reduce or eliminate the effect of registration errors, now we introduce this idea to our optimal interpolation.

In our application, we can obtain 20 or more frames with overlapped regions. For each location of the HR grid, we can obtain more than one correspondence location in other frames. For the sake of accuracy, we find the neighbour integral pixels with the minimum Euclidean distance in each frame. Since the pixel location is relevant to the intensity value in the two dimensional image, the deviation of the location can also be regarded as the intensity shift.

We attempted to adopt a second order polynomial model to fit the intensity distribution in the HR grid. Some researchers suggest that using a higher order polynomial can fit the rapid changes more accurately. From the point of view of an engineering application, the magnification factor (MF) of SR technique is less than 2.0. Zhouchen Lin et al. [20] give explicit bounds of the magnification factor and investigate a sufficient number of LR images. They prove that the practical limit is 1.6, if the noise removal and registration is not good enough. For a smaller MF, second order approximation is sufficient for our application. The second order polynomial model can be described as:

\[
f(x, y) = c_0 + c_1x + c_2y + c_3x^2 + c_4xy + c_5y^2
\]  

(11)

where \( x, y \) are the pixel coordinates and \( 1, x, y, x^2, xy, y^2 \) are the basis functions. \( f(x, y) \) is the intensity value. Given \( m \) points, and \( m > 6 \), the equation can be written into a matrix form:

\[
MC = B
\]  

(12)

where

\[
M = \begin{bmatrix}
1 & x_1 & y_1 & x_1^2 & x_1y_1 & y_1^2 \\
1 & x_2 & y_2 & x_2^2 & x_2y_2 & y_2^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_m & y_m & x_m^2 & x_my_m & y_m^2
\end{bmatrix},
\]

\[
C = [c_0 \ c_1 \ c_2 \ c_3 \ c_4 \ c_5]^T, \ B = [f_0 \ f_1 \ f_2 \ f_3 \ f_4 \ f_5]^T
\]

Thus, the solution of \( C \) can be computed by conventional linear least squares.

\[
C = (M^T M)^{-1} M^T B
\]  

(13)

We have made the assumption that the intensity distribution of the local grid is submitted to a second order polynomial model, then the RANSAC method was employed to optimize the parameters of the model.

The RANSAC interpolation algorithm is summarized in algorithm 1.

**Require:** The set of correspondence pixels for current grid point.

**Ensure:** The optimal interpolation for current grid point.

1: for \( i = 1 \) to \( M \) (\( M \) is the required sampling times) do
2: Select a sample of pixels in correspondence pixels set;
3: Estimate the hypothesis parameters based on the sample;
4: Calculate the error of each correspondence point in all frames with regard to the hypothesis;
5: Determine the number of inlier pixels consistent with the hypothesis;
6: end for
7: Select the model hypothesis with largest inlier pixels support, then re-estimate model parameters using all the inliers.

**Algorithm 1** The RANSAC multi-frame interpolation

For the RANSAC multiple frame interpolation, the required number of samples \( M \) is related to the hypothesis model and noise level. In the optimal interpolation procedure, we have taken four neighbour points from the reference image, which can be treated as
control points. Then two or more points were randomly sampled from other images based on the registration results. This solution can reduce the sample times as well as lead to robust fitting. Robustness and flexibility in modeling noise characteristics and a random sample consensus about the solution are the major advantages of the optimal interpolation approach.

3.3 Iterative blind de-convolution

Now we will deal with the restoration problem with the aim to remove the blurring and noise of the HR image obtained by the non-uniform interpolation approach.

HR images from optimal interpolation on non-uniformly sampled data still have noise and blurring, so de-convolution restoration of the image is adopted as the final step to estimate the SR images and eliminate the blurring effect for better image quality. The image degeneration can be described by the process of the point-spread function (PSF).

Image restoration is an inverse process to estimate the primitive image from the degenerated image. It is literally a regular de-convolution process if the PSF is determined. Since the PSF of video sequence remains unknown, we restore the image by the iterative back projection (IBP) method [21,22] using empirical parameters. This process is repeated iteratively to minimize the energy of the error. The IBP method can be expressed as:

\[ f^{n+1}(i,j) = f^n(i,j) + \frac{1}{K} \sum_{(m_1, m_2) \in R^{(i,j)}} (g_k(m_1, m_2) - g^n_k(m_1, m_2)) \ast h_{BP}(m_1, m_2, i, j) \]  

(14)

where \( g_k^{(i,j)} \) are simulated LR images from the approximation off after \( n \) iteration, \( y^{(i,j)} \) denotes the local window \( m_1, m_2 \in g_k \mid m_1, m_2 \) is influenced by pixel, \( h_{BP}(m_1, m_2, i, j) \) is a back projection kernel that determines the contribution of the error \( g_k(m_1, m_2) - g^n_k(m_1, m_2) \) to \( f^n(i, j) \).

It needs to be pointed out that the choice of \( h_{BP} \) plays an important role in the reconstruction. Our algorithm was mainly focused on down sample degeneration, so we take a Gaussian blurring kernel as the PSF.

4. Experimental results

In order to evaluate the performance of our approach for SR, a series of simulations and real data experiments were carried out to verify them.

4.1 Simulation experiments

We simulated the sub-pixel displacement shown in Fig.3, the sub-pixel displacement is simulated by down sampling the high-resolution grid from the odd-even field. Then, the registration results obtained by the proposed method are compared to those obtained by the SIFT matching method, and illustrated in Fig.4. By using the proposed method, the motion estimation errors and re-projection errors have been reduced significantly, and the registration errors are mainly within 1/10-pixel.

![Figure 3. Simulation for sub-pixel displacement](image)

![Figure 4. Simulation for registration.](image)
Figure 5. The RANSAC interpolation of the HR grid. (a) One unit of interpolation, the four green circles are pixel from the reference image, and the red stars are from the other frame by registration, (b) compare the RANSAC interpolation with the least square fitting method, (c) the least square fitting interpolation, (d) the proposed method.

Simulation experiments were also employed to verify the proposed method. The simulation results show a comparison between the proposed algorithms and those of the three existing algorithms on four LR Lena images. The peak signal to noise ratio (PSNR) metrics is used as an indicator of performance. The visual quality of the HR images and PSNR metrics were demonstrated in Fig.6 and Table 1.

Figure 6. Simulation results showing a comparison between the proposed method and the existing SR method on a Lena image. (a) Four down sampled LR images of Lena, (b) original HR Lena, (c) bilinear, (d) bicubic, (e) POCS with Vandewalle registration, (f) our method.
Table 1. PSNR of the different methods

| Technique                  | PSNR |
|----------------------------|------|
| The bilinear interpolation | 22.33|
| The bi-cubic interpolation | 22.81|
| The Vandewalle POCS method | 25.91|
| The proposed method        | 28.17|

The PSNR was defined by the mean squared error (MSE) to the original reference image. The PSNR can be obtained from equation (16)-(18):

$$\text{MSE} = \frac{1}{r \times c} \sum_{i=1}^{r} \sum_{j=1}^{c} [H(i,j) - F(i,j)]^2$$  \hspace{1cm} (16)

$$\text{RMSE} = \sqrt{\text{MSE}}$$  \hspace{1cm} (17)

$$\text{PSNR} = 20 \log_{10} \left( \frac{255}{\text{RMSE}} \right)$$  \hspace{1cm} (18)

4.2 Real data experiments

The proposed method has been applied to the real scene image sequence and satellite video (provided by LAPAN TUBSAT), and the HR image is generated from multiple LR frames. Fig. 6 provided a comparison between the SR image and the primitive low resolution image which was amplified to the size of the reconstructed image. From Fig. 6(a) and (b) we can find that the depths of the objects are different, so the transformation between two frames cannot be regarded as homography. However, our algorithm can achieve a good result for almost the whole image. This verified that the proposed method can reduce the shortcomings of the aforementioned homography model, as well as select the most effective information from multiple images to generate a HR image. Then we applied the algorithm to a real satellite video, and the results are shown in Fig.6(c) and (d). Fig.7(a), 7(b), 7(c), 7(d) give a comparison of the details in the primitive and reconstructed images, the HR image contains more distinguishable details such as recognizable objects.

Finally, the evaluation quality of the reconstructed image and the interpolation-enlarged image is assessed by entropy, average gradient and edge gradient. The average gradient parameter reflects the minute contrast and texture changes of the image. The information entropy reflects the richness of the image information, and the edge gradient is based on Wang Xing’s method [23], which is based on the smoothing effect of blur on edges. Table 2 shows that the evaluation results of the reconstructed image are markedly improved.

| Test image | Entropy | Average gradient | Edge gradient |
|------------|---------|------------------|---------------|
|            | bilinear | bilinear | bilinear | bilinear |
| #1         | 2.10     | 12.39     | 14.23     | 35.46     |
| #2         | 2.18     | 12.01     | 16.93     | 35.65     | 42.85 |
| #3         | 2.17     | 19.22     | 20.48     | 46.59     | 50.23 |
| #4         | 2.11     | 11.59     | 15.64     | 34.22     | 40.35 |

Table 2. Image Quality of the different methods

4.3 Computational complexity analysis

The procedure of SIFT initial match, RANSAC interpolation and IBP restoration tend to be computationally costly. The computation of the SIFT extractor, initial match and IBP restoration have a good property of computational consistency. This property motivates us to transfer these algorithms to a largely parallel platform such as graphic process units (GPU). We adopted the Marten Bjorkman’s CUDA SIFT, and implemented the IBP algorithm on GPU to improve the computation efficiency.
The LSM refinement and RANSAC interpolation were an iterative-based method with a data dependent conditional branch. The branch divergence would sharply decrease the efficiency, so we implemented these stages on CPU.

The overall execution time and the distribution of each stage are shown in Table 3.

Hardware platform: Intel Core i5-2430M CPU with 2 GB L2 Cache, NVIDIA NVS 4200M.

Experiment condition: Input nine LR images, image size: 256×256, RANSAC iteration threshold: 10.

| Stage                          | Value  |
|-------------------------------|--------|
| CUDA SIFT initial match (sec) | 0.248  |
| LSM refinement (sec)          | 2.252  |
| RANSAC interpolation (sec)    | 7.907  |
| IBP (sec)                     | 0.023  |
| Total time (sec)              | 10.428 |
| Frame rate (frames/sec)       | 0.096  |

Table 3. Execution time distribution

5. Conclusion

This paper proposed a new superresolution method for video sequences. The coarse-to-fine registration gives precise motion estimation for the continuous images, and the random sample consensus interpolation have good tolerance to the noise caused by sampling as well as the registration errors, especially for continuous images with terrain undulation. Experiments on the LAPAN TUBSAT satellite videos demonstrate the promising ability to obtain HR images of the ground scene. It also gives a clear boundary and more details for analysis of the ground surface.

However, this paper was designed to deal with scenes that are far from the camera and with a negligibly small three dimensional character to the scene. The proposed method may often suffer from a sort of degradation because it cannot inherently handle the perspective distortion caused by depth variation. These problems deserve further investigation.

6. Acknowledgment

The authors would like to extend gratitude to Marten Bjorkman aka Celebrandil for their provision of the CUDA SIFT extractor and LAPAN TUBSAT for their provision of satellite videos. Relevant work of the paper is funded by the National Basic Research Program of China under grant no. 2013CB733100 and the Open Research Foundation of Science and Technology on Aerospace Flight Dynamics Laboratory under grant no. 2012afd1041.

7. References

[1] Park S C, Park M K, Kang M G. Super-resolution image reconstruction: a technical overview[J]. Signal Processing Magazine, IEEE, 2003, 20(3): 21-36.
[2] Van Ooijenkerk J D. Image super-resolution survey[J]. Image and Vision Computing, 2006, 24(10): 1039-1052.
[3] Vandewalle P, Sui S, Vetterli M. A frequency domain approach to registration of aliased images with application to super-resolution[J]. EURASIP Journal on Advances in Signal Processing, 2006.
[4] Tsai R Y, Huang T S. Multiframe image restoration and registration[J]. Advances in computer vision and Image Processing, 1984, 1(2): 317-339.
[5] Li L C, Yu Q F, Yuan Y, et al. Super-resolution reconstruction and higher-degree function deformation model based matching for Chang’E-1 lunar images[J]. Science in China Series E: Technological Sciences, 2009, 52(12): 3468-3476.
[6] Cheng M H, Chen H Y, Leou J J. Video super-resolution reconstruction using a mobile search strategy and adaptive patch size[J]. Signal Processing, 2011, 91(5): 1284-1297.
[7] Schultz R R, Stevenson R L. A Bayesian approach to image expansion for improved definition[J]. Image Processing, IEEE Transactions on, 1994, 3(3): 233-242.
[8] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections[J]. JOSA A, 1989, 6(11): 1715-1726.
[9] Lindeberg T. Feature detection with automatic scale selection[J]. International journal of computer vision, 1998, 30(2): 79-116.
[10] Matas J, Chum O, Urban M, et al. Robust wide baseline stereo from maximally stable extremal regions[C]. British machine vision conference. 2002, 1: 384-393.
[11] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.
[12] Ackermann F. Digital image correlation: performance and potential application in photogrammetry[J]. The Photogrammetric Record, 1984, 11(64): 429-439.
[13] Wang Z, Qi F. Super-resolution video restoration with model uncertainties[C]. Image Processing, 2002. Proceedings. 2002 International Conference on. IEEE, 2002, 2: II-853-II-856 vol. 2.
[14] Lee E S, Kang M G. Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration[J]. Image Processing, IEEE Transactions on, 2003, 12(7): 826-837.
[15] Bose N K, Ahuja N A. Superresolution and noise filtering using moving least squares[J]. Image Processing, IEEE Transactions on, 2006, 15(8): 2239-2248.
[16] Costa G H, Bermudez J C M. Statistical analysis of the LMS algorithm applied to super-resolution image reconstruction[J]. Signal Processing, IEEE Transactions on, 2007, 55(5): 2084-2095.
[17] Rew K, Nam T, Joo H N, et al. Enhancement of Illumination Irregularity for the 2D Blot Detection Under Low Contrast[J]. Journal Korean Society of Precision Engineering, 2007, 24(4): 29.

[18] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381-395.

[19] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision Second Edition. Cambridge University Press, 2004.

[20] Lin Z, Shum H Y. Fundamental limits of reconstruction-based superresolution algorithms under local translation[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2004, 26(1): 83-97.

[21] Hsieh C C, Huang Y P, CHEN Y, et al. Video Super-Resolution by Motion Compensated Iterative Back-Projection Approach[J]. Journal of Information Science and Engineering, 2011, 27: 1107-1122.

[22] Irani M, Peleg S. Motion analysis for image enhancement: Resolution, occlusion, and transparency[J]. Journal of Visual Communication and Image Representation, 1993, 4(4): 1993-12.

[23] Wang X, Tian B, Liang C, et al. Blind image quality assessment for measuring image blur[C]. Image and Signal Processing, 2008. CIISP’08. Congress on. IEEE, 2008, 1: 467-470.