Panorama Recovery from Noisy UAV Surveillance Video

Yi Wang\(^1\), Richard R. Schultz\(^2\) and Ronald A. Fevig\(^3\)

\(^1\)2 Department of Electrical Engineering, University of North Dakota
\(^3\) Department of Space Studies, University of North Dakota

\{yi.wang@und.edu, richardschultz@mail.und.edu, rfevig@aero.und.edu\}

Abstract

This paper proposes an efficient and robust algorithm to recover a panorama from poorly-obtained UAV video frames contaminated with significant noise. In this algorithm, the eigen-space based neighborhood region will be introduced with our novel feature-based random M least-squares (RMLS) registration technique. Meanwhile, the corresponding similarity regions will be assigned weights according to the similarity between these neighboring regions. Next, Bayesian multi-frame sampling will be implemented utilizing the homography estimated by the frame registration. Finally, the sub-region in each frame which is applicable to the multi-frame sampling will be stitched utilizing multi-resolution blending.

Keywords: Mosaic, Multi-scale Neighborhood Region, Random M Least-squares, Unmmaned Aerial Vehicle

1. Introduction

UAV digital payload imaging have found more and more utility in the modern surveillance system. It is not terribly difficult for a small UAV payload to obtain a stream of video covering a region of interest. However, due to both of the limitation of payload cameras and the environment where the video is taken, it is usually unlikely to obtain a single frame covering the whole region of interest. Even more problematic, due to the high ISO speed, the video recorded by the UAV is usually contaminated by noise, which decreases the utility of the video. Basically, if we evaluate each single frame captured by UAV payloads, they can be useless because of the noise and resolution. Thus, the recovery of the video frames and stitching them together to obtain the panorama is essentially important for our further analysis. Fortunately, the video we record contains many frames, and more information can also be studied by multi-frame technology.

In this paper, we will first recall our feature-based (SIFT) registration method called random M least squares [2] [3]. Then, our unique multi-resolution neighborhood structure (MRSN) will be proposed to represent a more comprehensive and precise neighborhood structure for each pixel. The neighborhood structure is then vectorized to a noisy patch, and a Bayesian method to estimate the optimized state patch will be proposed. In Bayesian maximum a posteriori (MAP) estimation, multi-frame registration will be selected as the candidate for the state, which we named the Multi-Frame Bayesian Prior Candidates (MBPC). After that, a non-Gaussian similarity weight function is proposed to calculate a more reasonable estimation of weight (as the conditional probability) for all of the similar regions of each specific pixel. Once the joint regions of the frames have been removed of noise, they are stitched together by multi-resolution blending [4] to generate the panorama.

This paper is organized as follows. Section 2 will describe the RMLS method while Section 3 will cover the multi-resolution neighborhood region and similarity weighting. Section 4 contains our noise-reduction mosaicking algorithm and the experimental results.

To demonstrate the effectiveness of our algorithm, we provide three panoramas and their enlarged sub-regions: noisy panorama, the recovered panorama by non-local means de-noising [5] (the most efficient de-noising method in recent years), and the recovered panorama by our proposed method. From the results, the dominance of our method will be easily evaluated.

2. Frame Registration

Scale-Invariant Feature Transformation (SIFT) [1] approach is used to detect local feature points, which is invariant to rotation, noise, illumination, and scaling. Once the feature points are found, they are merged into a k-d [6] tree. The four nearest neighbors are found with a computational complexity of \(O(N\log N)\). Next, a standard RANSAC routine is chosen to reject the feature points outside the overlapping region. The homography transformation model is adopted as in equation (1), and then rewritten as equation (2) for least-square solution. The algorithm of feature-based registration is summarized in Table [1].

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

Equation (1) can be rewritten as
Table 1. Random M-Least Square Algorithm for Image Registration

1. Using all of the match pairs in the match set, generate $A_{total}$ according to the matrix on the lefthand side of equation (2) and $B_{total}$ according to the vector on the righthand side of equation (2), and assign the accepted error according to the desired precision, $\epsilon$.

2. Randomly choose $M$ (approximately one quarter of the total number of matched pairs, based on our testing results) points from the match set, to generate the matrix $A_M$ and vector $B_M$, calculate the transformation matrix using the least squares fit and the total tolerance:
   
   $h = \text{LeastSquareSolve}(A_M, B_M)$

3. Use the parameters derived in Step 2 to fit all of the elements in the match set and find the total fitted number less than the accepted error:

   $Total_{num} = \text{Count}(\|A_{total}h - B_{total}\|^2 < \epsilon^2)$

4. Repeat the above steps a suitable number of times, updating the transformation matrix when the following condition is satisfied:

   $\|tol_{new}\| < tol$ and $Total_{num_{new}} > Total_{num}$

   
   $\left[ \begin{array}{cccc} x & y & 1 & 0 \\ 0 & 0 & x & y \\ \vdots & \vdots & \vdots & \vdots \end{array} \right] \left[ \begin{array}{c} h_{11} \\ h_{12} \\ h_{13} \end{array} \right] = \left[ \begin{array}{c} \frac{x}{f} \\ \frac{y}{f} \\ \frac{x'}{f} \\ \frac{y'}{f} \end{array} \right]$

From this representation, we see that if we have more than four matching pairs, this problem is overdetermined, and we use the algorithm described in Table 1 to derive the required transformation matrix from coarse to fine.

3 Multi-Scale Neighborhood Region and Non-Gaussian Similarity Function

3.1 Multi-Scale Neighborhood Region

Most local-fusion denoising methods suffer from loss of details. Recently, a non-local means method was proposed by A. Buades [5]. This method focuses on weighing the non-local similarity region according to a Gaussian procedure and generating favorable results compared with the local-fusion method. However, due to the structure of their neighborhood region, their assumption of independence between these regions and their insufficient sampling numbers, their method generates obvious artifacts (Figure 2-a,b) in our experiments and still has room for improvement.

The Multi-Scale Neighborhood Region (MSNR) is constructed to find similar regions and weigh the similarities.

$$MSNR(x, y) = \left( \frac{G(x, y, \sigma)}{D_{2x}(x, y)} \frac{L(x, y, \sigma)}{D_{1y}(x, y)} \right)$$

$$G(x, y, \sigma) = g(x, y, \sigma) * I(x, y)$$

$$L(x, y, \sigma) = G(x, y, k\sigma) - G(x, y, \sigma)$$

Here, $G(x, y, \sigma)$ and $L(x, y, \sigma)$ represent the Gaussian and Laplacian pyramids, respectively, while $D_{1y}(x, y)$, $D_{2x}(x, y)$, $D_{1y}(x, y)$ and $D_{2y}(x, y)$ represent the first-order and second-order derivatives.

3.2 Optimal Bayesian Solution

Basically, the denoising procedure attempts to recover the ground truth image from the observations. We use $u(x)$, $z(x)$ and $v(x)$ to denote the ground-truth image, noisy image, and noise, respectively. Their relationship is modeled as:

$$z(x) = f(u(x), v(x))$$

Actually, we will never know the value of $u(x)$, since our observations are from an ultra-noisy video. Here, the estimated ground-truth value is denoted by $\hat{u}(x)$, and the expected value of error function is given as:

$$E[error(u(x), \hat{u}(x))] = \sum error(u(x), \hat{u}(x))p(u(x)|z(x));$$

The denoising process is the same as minimizing equation (7). According to the Bayesian posteriori model, the (7) can be modified as

$$E[error(u(x), \hat{u}(x))] = \sum \frac{error(u(x), \hat{u}(x))p(z(x)|u(x))p(u(x))}{p(z(x)|u(x))p(u(x))}$$

Now it is clear that the results of our denoising is decided by three factors: the error function $error(u(x), \hat{u}(x))$, the similarity function $p(z(x)|u(x))$, and the prior probability $p(u(x))$. Stated as above, the value of $u(x)$ is unknown, thus, to choose a reasonable candidate for $u(x)$ is critically important.

3.3 Multi-Frame Bayesian Prior Candidates

In section 3.2, the optimized estimation $u_{opt}(x)$ can be estimated by minimizing the

$$u_{opt}(x) = \text{ArgMin} \sum error(u(x), \hat{u}(x))p(z(x)|u(x))p(u(x))$$

Ideally, the error function can be defined as

$$error(u(x), \hat{u}(x)) = \begin{cases} 1 & \text{if } u(x) = \hat{u}(x) \\ 0 & \text{otherwise} \end{cases}$$
However, no estimation can be perfect in the real-world, so we adopt a hybrid error function to replace it. This error function discards the imprecise estimation while decreasing quadratically the patch’s difference.

\[ \text{error}(u(x), \hat{u}(x)) = \begin{cases} ||u(x) - \hat{u}(x)||^2, & \text{if} ||u(x) - \hat{u}(x)|| < \gamma \\ 0, & \text{otherwise} \end{cases} \]  

(11)

Assuming this piecewise error function, the optimal estimation will be adopted as:

\[ u_{\text{opt}}(x) = \text{ArgMin} \sum_{u(x)} \frac{||u(x) - \hat{u}(x)||^2 p(z(x)|u(x)) p(u(x))}{p(z(x)|u(x)) p(u(x))} \]

(12)

Although \( p(z(x)|u(x)) \) and \( p(u(x)) \) can not be obtained from the UAV video itself, they can be estimated by fusing the corresponding pixels of the registered frames, which we call the multi-frame Bayesian prior candidates. Here, according to our experiments, we replace \( u(x) \) by

\[ u(x) \rightarrow \sum_{\Omega(x)} z(x) \]

(13)

\( \Omega(x) \) represents the adjacent domain of \( u(x) \), which is exactly the same as the patch of the corresponding pixels in the registered adjacent frames. \( N(x) \) represents the semi-local neighbor fields of \( u(x) \) and its corresponding counterparts. Utilizing sufficient information, we can find a suitable candidate for \( u(x) \) to optimize our estimation. Then (12) can be recast as:

\[ u_{\text{opt}}(x) = \text{ArgMin} \sum_{\Omega(x)} \sum_{z(x)} N(x) \sum_{\Omega(x)} u(x_{i,j}) \]

(14)

\[ p(z(x)|\Omega(x) \sum_{\Omega(x)} \sum_{N(x)} z(x)) \]

(15)

Up to now, the multi-frame neighborhood candidates for the estimation has been derived and the optimal estimation can be calculated by fusing the information of the multiple corresponding neighborhood region.

3.4 Hybrid Similarity Function

To solve (14), the implementation of the conditional probability \( p(z(x)|\Omega(x) \sum_{\Omega(x)} \sum_{N(x)} z(x) \) is critical. Originally, Baudes [5] assumes the patch \( z(x) \) and \( \sum_{\Omega(x)} \sum_{N(x)} z(x) \) are Gaussian independent, as he adopts the Euclidean distance to estimate the similarities. However, according to our experiment, we find it is more reasonable to assume each single elements in patch \( z(x) \) and \( \sum_{\Omega(x)} \sum_{N(x)} z(x) \) are Gaussian independent to each other. Then according to the relationship between the Gaussian and Chi-square distribution, we can assume that \( p(z(x)) \) and \( \sum_{\Omega(x)} \sum_{N(x)} z(x) \) satisfy the Chi-square distribution with a degree of freedom \( k \in \Omega(x) \sum_{\Omega(x)} \sum_{N(x)} z(x) \).

Now the similarity function can be defined by:

\[ p(z(x)|\Omega(x) \sum_{\Omega(x)} \sum_{N(x)} z(x)) = \exp \left( \frac{-1}{2k} ||z(x) - \Omega(x) \sum_{\Omega(x)} \sum_{N(x)} z(x)||^2 \right) \]

(16)

\[ \sum_{\Omega(x)} \sum_{N(x)} \sum_{\Omega(x)} \sum_{N(x)} z(x) \]

4 Overview of Panorama Recovery Algorithm and Results

In this section, the algorithm for our panorama recovery will be stated first, and then the experimental results will be presented.

we will demonstrate the effectiveness of our algorithm by using it on the UAV video captured by our custom-built UAV payloads. RMLS registration helps to find more high-weight similar regions to eliminate the artifacts. A Bayesian conditional probability and prior probability are proposed to solve the problem. The multi-frame neighborhood region is utilized to model the initial guess of ground truth (prior probability) to improve the results significantly. A Chi-square distribution with a a large degree of freedom is utilized to evaluate the similarities properly.

Actual UAV images are shown in Figure 1 to demonstrate our algorithms effectiveness. Figures 1 (a)-(i) shows the noisy input frames. Figure 2 (a) results from mosaicking the noisy UAV data. Figure 2 (b) shows an enlarged sub-region of the noisy panorama. Figure 3(a) results from mosaicking the de-noised UAV data by non-local means. Figure 3(b) is an enlarged sub-region of Figure 3(a). Figure 4(a) results from mosaicking the de-noised UAV data utilizing our methods. Figure 4 (b) show an enlarged sub-region of Figure 4(a). While comparing Figures 3 and 4, we can see significant improvement.

Figure 1. Input Noisy Video Data
methods since it is based on a novel feature-based registration algorithm, a robust multi-scale neighborhood region (MSNR), and a reasonable approximation of the Bayesian conditional and prior probability. And this algorithm is especially useful for refining the data from a large set of poor observations.

References


