Architectural Building Detection and Tracking under Rural Environment in Video Sequences Taken by Unmanned Aircraft System (UAS)

Qiang He¹, Chee-Hung Henry Chu², Aldo Camargo³
¹Department of Mathematics, Computer and Information Sciences
Mississippi Valley State University, Itta Bena, MS 38941
²The Center for Advanced Computer Studies
University of Louisiana at Lafayette, Lafayette, LA 70504-4330
³Electrical Engineering Department, School of Engineering and Mines
University of North Dakota, Grand Forks, ND 58202-7165

Abstract - An Unmanned Aircraft System (UAS) is an aircraft or ground station that can be either remote controlled manually or is capable of flying autonomously under the guidance of pre-programmed GPS waypoint flight plans or more complex onboard intelligent systems. The UAS aircrafts have recently found extensive applications in military reconnaissance and surveillance, homeland security, precision agriculture, fire monitoring and analysis, and other different kinds of aids needed in disasters. Through surveillance videos captured by a UAS digital imaging payload over the interest areas, the corresponding UAS missions can be conducted. In this paper, we present an effective method to detect and extract architectural buildings under rural environment from UAS video sequences. The SIFT points are chosen as image features. The planar homography is adopted as the motion model between different image frames. The proposed algorithm is tested on real UAS video data.

Keywords: Unmanned Aircraft System (UAS), Object Detection and Tracking, Planar Homography

1 Introduction

An Unmanned Aircraft System (UAS) [1] is an aircraft or ground station that can be either remote controlled manually or is capable of flying autonomously under the guidance of pre-programmed GPS waypoint flight plans or more complex onboard intelligent systems. The UAS aircrafts have recently found extensive applications in military reconnaissance and surveillance, homeland security, precision agriculture, fire monitoring and analysis, and other different kinds of aids during disasters.

Here three major applications from UAS are listed and discussed as follows.

1. Military reconnaissance and surveillance:
The different kinds of UASs, such as Predator, Hunter, and Global Hawk, have been developed and successfully applied in many military operations. They provide supports on intelligent surveillance, target acquisition, tactical/strategic reconnaissance, and strike capability. The appearance of UASs is transforming the way of traditional war-fighting.

2. Homeland security:
The UAS could be deployed along the boundaries and coastal lines to protect the territory out of attacks from terrorists. The UAS installed with color and thermal IR digital cameras can provide airborne surveillance and positioning data about those suspicious targets as fast as in almost real time. The acquired data will be transmitted into the ground station and the subjects can be identified or apprehended.

3. Fire monitoring and analysis:
It is reported that The 2011 Texas wildfires has destroyed more than 1500 homes, burned over 34,000-acre land, and killed 2 people. The wildfire results in serious loss of property, human life and agriculture. The UAS mounted with thermal infrared imaging camera can assist the wildfire management, such as Forest Fire Damage Assessment, Forest Fire Mapping, Forest Fire Communications, to reduce the loss of life, property, and forest natural resources.

Under the appearance of natural disasters such as a wildfire, a typhoon, and an earthquake, the communication systems are always damaged at first. The UAS becomes a very helpful tool to assist people and government to conduct the salvage. An UAS can be sent out to disaster-affected areas to collect data for further action and set up the communication between the people in disaster areas and the outside rescue people and the rescue commander center. The residential architectural buildings are very important targets under the UAS surveillance (refer to Fig. 1). The information about the residential architectural buildings inside the disaster areas
provides us a direction to rescue people and reduce the property loss. Therefore, how to detect and extract architectural buildings from UAS surveillance videos becomes a very useful task. Here we are, in particular, interested in the detection and tracking of architectural buildings under rural areas or suburbs where the transportation cannot get access to and the communication is not easy to get recovery quickly soon.

This paper proceeds as follows. Section 2 describes the basic characteristics of man-made architectural buildings, which could characterize them for detection and tracking. Section 3 presents the technique to detect and track buildings from UAS videos in nearly real time. Section 4 summarizes the algorithm. The experimental results on real UAS data are presented in Section 5. We draw a conclusion for this paper in Section 6.

![Fig. 1. Architectural buildings locate in rural areas seen from an UAS video (from internet).](image)

## 2 Characteristics of Man-made Architectural Buildings

As we already know that the Hurricane 2005 and the Texas wildfire 2011 has resulted in serious influence and huge loss of human-being life, residential property, agricultural yield harvest, etc. The information about the residential architectural buildings inside the disaster areas provides us a direction to rescue people and protect the property. Therefore, how to detect and extract man-made buildings from UAS surveillance videos becomes a very meaningful task. We are particularly interested in the detection and tracking of architectural buildings under rural areas or suburbs where the transportation cannot get access to and the communication is not easy to get recovery.

There are many very special characteristics for man-made buildings distinguished from the natural environment. We summarize some important ones as follows.

1. Planar surfaces abound in man-made architectural buildings such as brick walls, window glasses, roofs, etc.
2. Parallel lines are widely found in man-made structures. So vanishing points and vanishing lines can be produced and used as a clue to locate the architectural buildings.
3. Special angels exist in man-made structures. For example, the right angles are always formed around the corners from different sides of brick walls.
4. Special geometric shapes such as rectangles, circles, ellipses, distribute in the building surface.

All of this special information from architectural buildings provides us many useful cues to implement the related object detection and tracking.

## 3 Detection and Tracking of Architectural Buildings

The architectural buildings stand out of natural scene with their man-made characteristics such as special colors, shapes, structures, etc. With the appearance of natural disasters such as earthquakes, hurricanes, tornados, the architectural buildings are always the targets which a disaster attacks and people rescue on the other side. In this section, we will discuss the detection and tracking of architectural buildings from UAS video image sequences. The detection and tracking of architectural buildings refers to extracting architectural buildings as foreground objects and tracking them in image sequence through robust invariant features under geometric transformation. There are three major tasks to detect and track architectural buildings: (1) feature detection and correspondence [3,5,7,8]; (2) motion estimation between different frames; and (3) object segmentation. In our research, the SIFT (Scale Invariant Feature Transform) corners [5] and planar surfaces [7,8] are selected as image features to characterize man-made architectural buildings. The planar homography [4,7] is chosen as the motion model since it captures the motion between planar surfaces in different views. The intelligent scissor method [6] is applied to extract buildings from the Delaunay triangulation [9] on SIFT feature points.

### 3.1 Distinctive Features for Architectural Buildings

For classical feature detectors such as Harris corner detectors, the localization and scale are assumed constant through affine transformation of the local image structures. So they are, in fact, very sensitive to changes in image scale. On the other side, the SIFT (Scale Invariant Feature Transform) feature descriptor is invariant to scaling, rotation, affine transformation, and partially invariant to illumination changes. It can robustly identify objects in the situation of clutter and partial occlusion. The SIFT feature has been widely applied in object recognition, robot navigation, image mosaicking, video tracking, etc. We adopt the SIFT feature here to identify and track local features in architectural buildings from UAS video sequences. Planar surfaces abound in man-made environments, in particular, in architectural buildings, manufactured objects, indoor environments, etc. The planar surface is a very important characteristic to encapsulate 2-dimensional information for objects. In computer vision field, the planes have been successfully used on camera calibration [10]. The plane-based calibration is easy to be implemented and satisfied solution has been reached. Here planar surfaces are applied to capture texture or high-level properties of architectural buildings in comparison...
with that the SIFT feature captures the local characteristics of architectural buildings.

3.1.1 SIFT corner

The SIFT feature is detected through Scale Invariant Feature Transform (SIFT). It is invariant to scale and rotation and can be used as a reliable feature to perform matching and/or tracking between different views of an object or scene. The Scale Invariant Feature Transform (SIFT) consists of following four steps.

1. **Scale-space extrema detection**: The scale-space extrema are the maximum or minimum values of the difference-of-Gaussian images at multiple scales. They are potential distinctive features and detected by comparing a pixel with its 26 neighboring pixels in the current and adjacent scales.

2. **Key point localization**: The key points are selected through two-order Taylor expansion. Given the difference-of-Gaussian $D(x,y,\sigma)$, its Taylor expansion is

\[
D(x) = D(0) + \frac{\partial^2 D}{\partial x^2} x + \frac{1}{2} x^T \frac{\partial^3 D}{\partial x^3} x
\]

In extrema position, we have $\frac{\partial D(x)}{\partial x} = 0$ and solve $x$ as

\[
\hat{x} = \left(\frac{\partial^2 D}{\partial x^2}\right)^{-1} \frac{\partial D}{\partial x} \tag{1}
\]

3. **Orientation assignment**: One or more orientations of each key point are selected with the peaks of smoothed histogram of local gradients.

4. **Key point descriptor**: The gradient distribution around a key point is used as descriptor. There are eight gradient directions for an image pixel. Here a 4 pixel by 4 pixel window around each key point is chosen to capture its characteristics. Therefore, the key points are described as a 16x8=128 dimension vector.

3.1.2 Planar surface

The planar surface is an important characteristic capturing 2-d information for objects. In addition, they abound in man-made environments, in particular, in architectural buildings, manufactured objects, indoor environments, etc. There are two major problems for planar surface detection. One is to compute the planar homography between different images based on entity correspondences. The other is to extract the whole plane from background through the planar homography. Our method uses SIFT feature as correspondences to compute homography between planar surfaces for all potential planes in different views. Then a Delaunay triangulation is conducted on these planar inliers. In order to extract the whole planes, we apply intelligent scissor method [6] to segment the original images. The segmented regions with maximum fitting to Delaunay triangulation areas are regarded as the buildings.

3.2 Planar Homography and Properties

From the knowledge given in [4], there is a planar homography among planar surfaces taken from different viewpoints. That is,

\[
x' = Hx
\]

where $x$ and $x'$ are the homogeneous coordinates for the projections of three-dimensional point $X$ on a world plane in the first image and the second one. In general, homography $H$ is a 3 by 3 non-singular matrix and hence its rank is 3. So there exists a one-to-one point correspondence between the left and right images. If the rank of $H$ is 2, then all the points of the plane in the first image will be projected onto a line in the second image. If the rank of $H$ is 1, then all the points of the plane in the first image will be projected onto a point in the second image.

In practice, we find that the following two problems have to be addressed for computation of planar homography. First, there is a homography from the first image to the second one under pure rotation of camera. This homography is stronger than any inter-image planar homography since all image points satisfy this homography. We should ignore it in order to detect those inter-image planar homographies. Second, if the rank of a homography is less than 3, it is degenerate. We should check this kind of homographies and ignore them. Here both the pure rotation and degeneracy cases of $H$ are discussed as follows.

3.2.1 Homography from pure rotation

If the only motion of camera is a pure rotation about its center, then there is a planar homography between the left image and the right image.

Define the project matrices for the left image and the right image respectively as

\[
P_1 = [K| -\tilde{C}] \text{ and } P_2 = [KR| -\tilde{C}]
\]

where $K$ is the camera calibration matrix associated with its internal parameters. $K$ is nonsingular. $\tilde{C}$ is the coordinates of the camera center in the world coordinate system. $R$ is the rotation to the world coordinate system.

Then given a world point $X$ and its two projections $x$ and $x'$ on the left and the right images, we have

\[
x = P_1 X = [K| -\tilde{C}]X \text{ and } x' = P_2 X = [KR| -\tilde{C}]X \tag{5}
\]

and further

\[
x' = P_2 X = [KR| -\tilde{C}]X = KR^{-1}[K| -\tilde{C}]X = KR^{-1}x = Hx \tag{6}
\]

where $H = KR^{-1}$ defines the planar homography between the whole left image and the whole right image.
In order to detect inter-image homographies on planar surfaces, we should ignore this major homography. We can segment the original images first and only consider those potential homographies within the different segmented components.

3.2.2 Homography degeneracy

As we know from the above, homography $H$, in general, is nonsingular and hence its rank is 3. Then we have a one-to-one point matching between the left and right images. Now, we consider the degeneracy case of $H$, that is, the rank of $H$ is less than 3. If the rank of $H$ is 2, all the points of the plane in the first image will be projected onto a line in the second image. If the rank of $H$ is 1, all the points of the plane in the first image will be projected onto a point in the second image. Here, the degenerate cases of $H$ are not of interest to us. As a result, we will check and then ignore them. No matter what the rank of $H$ is 1 or 2, the three row vectors of $H$ are collinear. This is a guide to detect the degeneracy of $H$.

3.3 Object Segmentation

In order to encapsulate the planar surfaces, we compute the Delaunay triangulation [9] on inliers for each detected homography. The Delaunay triangulation of a point set is a collection of edges, where each edge has a circle containing the edge's endpoints but not containing any other points. However, in general, the inliers-based triangulation does not enclose the whole planar surfaces. We further use some segmentation method to extract the whole planes. The Delaunay triangulation areas using the Delaunay triangulation with vertices of inliers of the planar homography.

4 Algorithm

In summary, the major stages of the algorithm to extract and track architectural buildings are listed as follows:

1. Detect SIFT points in video frames.
2. Estimate the planar homography between frames with most SIFT point fits using RANSAC [2].
3. Save the homography and make Delaunay triangulation with vertices of inliers of the planar homography.
4. If the maximum number of homography is reached or the ratio between inliers and total points is less than a threshold, then STOP.
5. Remove inliers and consider those points, which are not satisfied with the previous homography (outliers of it), GO TO stage 2.

6. Segment Delaunay triangulation areas using intelligent scissor. Those segmented components will be regarded as the final whole planes.

The fitting error of homography is defined as the sum of forward-projection fitting errors and backward-projection fitting errors. If the fitting error of homography is too small, one plane may be fitted by several homographies. If the fitting error of homography is too large, two or more planes may be fitted by one homography. In order to detect all planes, we set the fitting errors as small values.

5 Experimental Results

The proposed algorithm to detect and track architectural buildings in UAS video sequences was tested on two sets of real video data captured by an experimental small UAS operated by Lockheed Martin Corporation flying a custom-built electro-optical (EO) and uncooled thermal infrared (IR) imager. The time series of images are extracted from videos with resolution 320 x 240. Our algorithm was developed using MATLAB and was performed in nearly real time. We implemented our algorithm on a Dell Optiplex 760 workstation with an Intel Core 2 Duo CPU running at 3.00 GHz and 1.97 Hz and a 2.96 GB RAM. If we ported the algorithm into the C programming language, the algorithm would execute much more quickly.

Test data from small UAS aircraft are highly susceptible to vibrations and sensor pointing movements. As a result, the related video data are some kind of blurred and the interesting targets are not easy to be identified and recognized sometimes. The SIFT feature shows a robustness here. The experimental results for the first data set are given in Figures 2, 3, and 4. The experimental results for the second data set are provided in Figures 5, 6, and 7.

In the first test (refer to Figures 2, 3, and 4), the two image frames from original UAS videos are displayed in Figure 2. The 173 SIFT features and 172 ones are detected in the first and second images respectively. Among these features 123 matches are found through nearest-neighbor method, which is illustrated in Figure 3. The red cross sign “x” represents the identified SIFT features while the blue line connects the corresponding SIFT matches from two different image frames. The planar homographies are computed based on these matches using RANSAC method. The inliers associated with strong homography is connected through Delaunay triangulation to capture the rough area of architectural buildings. And then the intelligent scissor is applied to draw the contours of Delaunay triangulation area accurately and extract the architectural buildings out of the image. The segmented man-made buildings is displayed in Figure 4. Here two neighboring buildings are regarded as one.
Fig. 2. (a) and (b) are two image frames from original UAS videos. The size is reduced by 50% for visualization purpose.

Fig. 3. SIFT feature extraction and correspondence. Here there are totally 123 matches with nearest-neighbor method.

Fig. 4. Extraction of architectural buildings. The intelligent scissor is applied to draw the contours of Delaunay triangulation area and then segment the architectural buildings out of the image. There are totally 32 SIFT features for Delaunay triangulation.

In the second test (refer to Figures 5, 6, and 7), the two image frames from original UAS videos are displayed in Figure 5. There are 456 SIFT features and 434 ones detected in the first and second images respectively. Among these features, totally 267 matches are found through nearest-neighbor method. The correspondence is illustrated in Figure 6. The planar homographies are computed based on these matches using RANSAC method. The inliers associated with strong homography is connected through Delaunay triangulation to locate the rough area of architectural buildings. And then the intelligent scissor is applied to draw the contours of Delaunay triangulation area accurately. The segmented area is treated as the potential architectural buildings. The segmented man-made building is displayed in Figure 7.

Fig. 5. (a) and (b) are two image frames from original UAS videos. The size is reduced by 50% for visualization purpose.

Fig. 6. SIFT feature extraction and correspondence. Here there are totally 267 matches with nearest-neighbor method.

Fig. 7. Extraction of architectural buildings. The intelligent scissor is applied to draw the contours of Delaunay triangulation area and then segment the architectural buildings out of the image. There are totally 75 SIFT features for Delaunay triangulation.

6 Conclusions

In this paper, we present an effective method to detect and track architectural buildings from UAS videos taken under rural environment. The SIFT points are chosen as image features. The planar homography is adopted as the motion model between image frames. The intelligent scissor is used for segmentation; however, it is not an automatic segmentation method. The algorithm is tested on real UAS video data. In the future work, we will explore other methods to locate architectural buildings.
7 References


