An Improved Method for Spatial Measurement in Surveillance Video

Guang-jun LIAO¹, Min YANG¹ and Wei CHEN²

¹Faculty of Forensic Science and Technology, Guangdong Police College, Guangzhou, China
²School of Automation, Guangdong University of Technology

Keywords: Video detection technology, Perspective transformation, Close-range photogrammetry, Spatial Measurement.

Abstract. Video detection technology plays an increasingly significant role in forensics and criminology. Measurement of real distance from single image frames is a key technology for many applications. There are several methods for the measurement, including proportion measurement, investigative experiment, perspective and projection transformation. Although these methods are applicable to the special scenarios, there are still shortcomings of their own. Therefore, we propose an approach called two perspective transformations to achieve the real size of objects in 3-D world measurement. Firstly, we use close-range photogrammetry method to obtain 3-D structure of the multiple feature points at the crime scene. Then the 3-D structure will be corresponded to 2D feature points in the measured single image frames, and the collinear equations are used to establish the perspective transformation between 2D features in the measured image frame and 3-D structure in 3-D real world. Finally, we use the perspective transformation matrix to reconstruct the 3-D structure of any feature point in the measured image frame. The method we proposed is successfully applied in a corpse-dismantling case for measuring a box’s height.

Introduction

Video detection technology plays a more and more important role in forensics and criminology. With the increasing possibility of revealing the modus operandi, surveillance video has been used to record crime on spots. For instance, FBI identified two suspects with the help of surveillance videos just two days after Boston Marathon bombings (April 15, 2013). London bombings event happened on 7 July 2005, in which surveillance also helped the police investigate the suicide attacks. On March 1, 2014, in Kunming railway station, the terrorists attacked and killed the innocent civilians, which as a result caused a total of 31 civilians’ deaths and 141 injured. The surveillance video played a key role during the whole process when the suspects were tracked.

In fact, there are many automated processes related to surveillance videos, including facial recognition¹, license plate recognition², spatial measurement³, biometric data mining⁴,⁵ and etc. Furthermore, measurement of real distance from images is a crucial technology for many applications, such as crime detection. Major methods for the measurement are listed as followed:

Proportion measurement: Finding a precise scale object (using Photoshop or other image process software) at the crime scene from the captured video, and estimating the dimension of the suspicious object by the comparing to the object.

Investigative experiment: Setting up an additional calibration object that they can be easily identified, and capturing images for the object, then compare with the evidential frame to estimate the size of suspicious object.

Perspective projection transformation: Using perspective transformation to obtain the 3-D structure of video image feature points, on the premise of at least 6 known feature points with 3-D structure. So, this method is used in the situation when people or objects appear in a structural environment, such as corridor, door frame, etc. Otherwise, we need to use expensive 3-D laser scanner to obtain accurate 3-D information.
All the methods mentioned above are applicable to the special scenes, but there are still shortcomings of their own. For the first method, the ideal reference object may not always be either existing or accurate, leading to deviation and low precision. Moreover, calibration object or placement position in investigative experiment may also be proper for estimating the dimension, which results in a confusing effect. For the last method, 6 points in structural environments are needed, whereas they are not always satisfied. Otherwise, this method is limited by expensive 3-D structure acquisition equipment.

Unlike the above existing method, we propose an approach to obtain the spatial measurement of suspicious objects from single frames which only depends on visual information without any reference objects. Firstly, we use close-range photogrammetry method on crime spots to recover 3-D structure of the multiple feature points. Then the 3-D structure of the feature points will be remapped to the evidential video frames, and the camera imaging is used to establish the perspective transformation between 2D image frame and the 3-D real world. Finally, we use the perspective transformation matrix to calculate the 3-D information of any feature point in the evidential image frames. The method we proposed has been successfully applied in the measurement of the suspicious box, which provided important clues for the investigation of the case.

**Method**

In this study, based on close-range photogrammetry and spacial mapping, an improved method is presented to measure 3-D structure from a single evidential image in surveillance video. The close-range photogrammetry is similar to the theodolite measuring system, which usually uses the triangle intersection method, and records two or more images from suitable camera viewpoints for the same crime spot. After the images being pre-processed, feature points matched, perspective transformation calculated, the 3-D structure of the feature points in the scene can be worked out. The images acquisition equipment can be classified as digital camera of professional measurement and non-measurement digital camera. Taking generality into account, the proposed method chooses the non-measurement digital camera. Firstly, calibrate the non-measurement camera and record images. Secondly, establish the perspective transformation relation between the image points coordinate and the space points coordinate through the camera imaging model. Thirdly, assign a desired scale and XYZ coordinate system. Finally, produce the desired outcome: scaled XYZ feature point coordinates.

Close-range photogrammetry is an active measurement technology to obtain the 3-D information of feature points. However, in the video investigation, it is often necessary to obtain the 3-D information of object, which is generally not repeatable. So, the combination of the perspective transformation principle with collinear equation to calculate the 3-D information of the feature points in the video image is proposed.

![Figure 1. Coordinate system of the world, camera and image.](image)

The image points coordinate \( v_p^e = \begin{bmatrix} x_p^e \\ y_p^e \\ 1 \end{bmatrix} \) can be connected to the space points coordinate \( v = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \) by the collinear equations (1):

\[
\begin{pmatrix}
0 & y_v & -z_v \\
-x_v & 0 & -z_v \\
-1 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
x_p^e \\
y_p^e \\
1
\end{pmatrix} = \begin{pmatrix}
x \\
y \\
z
\end{pmatrix}
\]
\[ x - x_0 = f \cdot \left( a_{11}(X_A - X_S) + a_{22}(Y_A - Y_S) + a_{33}(Z_A - Z_S) \right) \]
\[ y - y_0 = f \cdot \left( a_{11}(X_A - X_S) + a_{22}(Y_A - Y_S) + a_{33}(Z_A - Z_S) \right) \]

\( a_q \) are elements of the rotation matrix \( R \), which is the function for the elements \( \omega, k, \varphi \) of exterior orientation.

\[
R = \begin{bmatrix}
     a_{11} & a_{12} & a_{13} \\
     a_{21} & a_{22} & a_{23} \\
     a_{31} & a_{32} & a_{33}
\end{bmatrix}
= \begin{bmatrix}
     \cos \varphi \cos k - \sin \varphi \sin w \sin k & \sin \varphi \cos w & -\cos \varphi \sin k - \sin \varphi \sin w \cos k \\
     -\sin \varphi \cos k - \cos \varphi \sin w \sin k & \cos \varphi \cos w & \sin \varphi \sin k - \cos \varphi \sin w \cos k \\
     \cos \varphi \sin k & \sin w & \cos w \cos k
\end{bmatrix}
\] (2)

\( x_0, y_0, f \) are inner orientation elements. \( X_S, Y_S, Z_S \) and \( \omega, k, \varphi \) are exterior orientation elements. The number of internal and external orientation elements usually is 9. We choose the Direct Linear Transformation (DLT) algorithm to connect the formula (1) and (2), which is widely used in 3-D reconstruction of planar images.

\[
x + \Delta x + \frac{L_1 X + L_2 Y + L_3 Z + L_4}{L_5 X + L_6 Y + L_7 Z + 1} = 0
\]
\[
y + \Delta y + \frac{L_1 Y + L_2 Z + L_4}{L_5 Y + L_6 Z + L_7 + 1} = 0
\]

\( L_i (i = 1, 2, \ldots, 11) \) are unknown. \( x, y \) are image points coordinates. \( X, Y, Z \) are space points coordinates in the scene. \( \Delta x, \Delta y \) are objective lens distortion and negative differential deformation respectively, belonging to the systematic error. There are 11 unknown parameters when \( \Delta x, \Delta y \) without consideration, so we need at least 6 pairs of corresponding points to establish the connection between image points coordinate and space points coordinate, or we can use some fitting method such as the least-square method to control the error while more than 6 pairs are selected.

Establishing the perspective transformation relation between space points coordinate and image points coordinate of the feature point by means of close-range photogrammetry measurement, we need 6 pairs of corresponding points at least. Mapping feature points to video scene, then we can establish the perspective transformation relation between video images and real world. By means of collinear equations and the least-square method, we get coefficients of the perspective transformation matrix. Finally, we can work out the 3-D information of any feature points in measured video images through perspective transformation matrix.

**Case Report**

In a suspected murder case, the height of the box appearing in the surveillance video is an important clue to estimate whether the box is enough to contain an adult female or not.
Figure 2. A suspect with a box appear the surveillance video.

In Fig.2, the suspect was walking, there isn’t suitable reference object for proportion measurement. In order to verify investigative experiment, we choose an investigator, whose height is same as the suspect, to perform the experiment. However, as the posture of suspect is changeable, the experimental error is very large. Besides, the video screen is an irregular space, which is difficult to obtain 6 feature points’ 3-D information for softwares such as ViSystem or Conigtech to calculate the height of box.

Figure 3. Investigative experiment.

The fundamental principle of stereo vision is using triangulation measurement to calculate the error between pixel position and the images, which are the pictures of object shot from two or more points of view, before we can obtain the 3D depth information of the object. We propose achieving three-dimensional calibration of the feature points of the scene by the three-dimensional reconstruction on stereo vision, and then realizing three-dimensional measurement of objects in video image according to the perspective relationship. The steps are as follow:

1) We should obtain two or more images of the object from different views;
2) Calibrate the camera so that we can establish the mapping relationship between 3D and 2D;
3) The 3D information of the object can be obtained through stereo matching after extracting the feature points;
4) Map the obtained three-dimensional data to the video image to be measured (figure 1). It can be achieved through the spatial measurement, and the height of the box is 15.5cm.

The box was unable to contain an adult female whole body. We supplied an important clue, which may be a corpse-dismantling case, to polygraph technicians. After hard working, the suspect admit the criminal fact that he killed a woman and dismembered the corpse. To identify the crime scene,
investigator found the box in the video (Fig.2), whose actual height was 16.2 cm. Thus, the error is less than 5%.

Figure 4. The actual height is 16.2cm.

Summary

We propose a method based on two perspective transformation. Firstly, with close-range photogrammetry measurement, we can establish a perspective relation between image points coordinate and space points coordinate, and we can work out the 3-D information of feature points. Secondly, information of feature points and collinear equations are used to calculate the coefficients of perspective matrix between space points coordinate and video image points coordinate. Finally, we can work out 3-D information of any point in video image through the perspective matrix. Our method breaks through some limitations, and it doesn’t need a proportional reference which is necessary for proportion measurement or investigative experiment.

Acknowledgements

In this paper, the research was sponsored by the National Natural Science Foundation of China (No.61471134), the Innovative Projects of Department of Education of Guangdong Province (Project No. 2015KTSCX096), the Innovative Projects of Department of Public Security of Guangdong Province (Project No. TKY2015002).

Reference