Target Positioning Using Object Tracking Technique for Video Images

Mr. Min-Hsin Chen*
Prof Chi-Farn Chen*
Mr. Chao-Yao Cheng*
Mr. Chun Lu Lin*

*Center for Space and Remote Sensing Research
National Central University Chung Li,
Tel:(886)-3-4227151-7659
e-mail: u3260424@cc.ncu.edu.tw
China Taipei

Abstract:

Digital video camera recently has become a popular monitoring tool because of its low cost and relatively high resolution. Comparing with the photogrammetric photos, digital video camera is difficult to locate the target position in the video image. In this study, we integrate the digital video camera, the digital compass and GPS (Global Positioning System) during the recording process. The frames of video have high correlation, so they can be used to track the object and image coordinate of the object's center in the video images. Using the frame number, GPS and orientation data, we can find the interior orientation of the frame. Then using least squares adjustment, we can calculate the position of the object quickly. An experiment is performed to locate the ship's positioning in the open sea; the result shows that the proposed method can locate the target's position successfully without the ground control points.

1. Introduction

Sometimes we take the helicopter and use the digital video camera to monitor in the sky. Though the digital video camera can get a large number of video images and can be used conveniently, we can't know the position of the object we are interested in. By equipping with GPS, tilt meter and digital compass, we can know the orientation and location of shooting. We can exchange the digital data to voice, so it can be recorded on the voice channel synchronously by using the encoder. After capturing and decoding from the digital tape, user can choose the object on the video images. Because the frames of video have high correlation, they can be used to track the object and get the image coordinate of the object's center in the video images. Using the image coordinate, the focal length (interior orientation), the digital camera's location and orientation (exterior orientation), we can calculate the object's coordinate by ways of least squares adjustment

2. Hardware of the System

The digital video camera of this positioning system we used is the SONY TRV20, equipped with GPS, digital compass and the tilt meter. SONY TRV20 has 10X optical zoom lens (f=4.2mm~42mm) but in this study we can't know the focal length during zooming. Hence we shot in fixed focal length (f=4.2mm). The angle of the tilt meter output which points up or down, left or right is related to the horizon. The accuracy of the tilt meter is +/- 1° with in a range of +/- 20° from the horizon. The accuracy of the digital compass is about +/- 5° and accuracy of GPS is about 15 meter. The entire data update rate is 1 Hz. When the data was input into the encoder, the encoder exchanged the data into voice and output to the microphone. Besides, the data was recorded on the voice channel at the same time. Therefore, the position, the orientation data and the video images are recorded on the tape synchronously.

After shooting, we used the capture card to compress the data on the tape to a video file. Using the program in C++ and the decoder, by playing the video file, we got the data of shooting per second and the corresponding frame number.

3.Object Tracking

When the video file is playing, the user can pause the movie and mark the target on the video image. The frames of video with high correlation can use image matching to track the object. Due to the most length of motion vector of the video images is less than 12 pixels [1], we can use the rectangle that the user pens in to be the target window and according to the target window size extent 12 pixels to be the search window. We move the target window to match the search window on the next frame and use the matched rectangle to replace the target window until the object disappears on the image or the result of matching fails.

The difference between the photo and video images matching is that the video images is sequence image; as for the sequence images, the shape change of the object which is caused by shooting angle change and without leading to match failure is OK. However, the sequence matching result on the former frame will effect the latter one, so we decided to choose "Normalized Cross Correlation (NCC)" to be the matching method. The NCC matching method is the method that chooses the rectangle, which has the highest correlation coefficient between the target window and the search window to be the matching result. It has the best accuracy.

Sometimes the shock of camera will cause vague images, and then the highest correlation might not be the correct matching result. One of the principles to recognize object in naked eyes is the object's color. Thus, if the highest correlation coefficient is smaller than 0.8, expect for calculating the correlation coefficient of the matching rectangle in 3 bands(R, G, B), also calculate the mean of the target window and the corresponding rectangle on the search window in 3 bands. If the result of matching is correct, the sum of difference between two means should be minimum .

$$\sum_{Band=1}^{3} (\text{Mean of Target Window [Band]} - \text{Corresponding Rectangle on the Search Window [Band]})^2$$

=Minimum So find all the rectangles that the difference of correlation coefficient between the highest one is less than 0.05 in the search window, the rectangle, which has the minimum sum of mean difference will be the target window in next frame. If the highest correlation coefficient is less than 0.7 then the result of matching is fail. Let the center of the frame be the origin of the image plane (S axis and L axis), then can use the coordinate of matching rectangle's center multiply the real pixel size ($1.5875*10^{-5}$ meter/pixel) and the focal length to build the observing vector $\overrightarrow{Obs}$ (Fig 1). $\overrightarrow{Obs}^i = \begin{bmatrix} x^i & f & y^i \end{bmatrix}$, where

$x^i = S^i$ *pixel size, f=focal length, $y^i = L^i$ *pixel size for ***i***th frame.

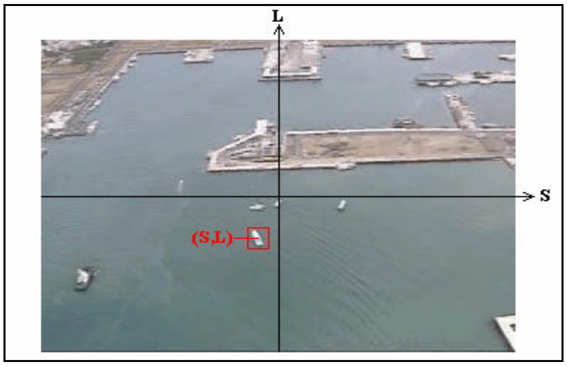


Figure1. (S, L) is the image coordinate of the matching rectangle's center, origin of the image coordinate is the center of the frame.

4.Positioning

For each frame that has position and orientation data, will give rise to set of three equations. These will be derived from the following relationship between the shooting position and the target's location in local geodetic system (E, N, H). (Fig 2):

$$\vec{T} = \vec{P}^{\,i} + S^i \cdot \vec{U}^{\,i}$$

where

$\vec{T}$ = Target position vector $[T_E \quad T_N \quad T_H]$

$\vec{P}^{\,i}$ = The *i*th frame position vector $[P^i_E \quad P^i_N \quad P^i_H]$

S^i = Scaling factor of *i*th frame

$\vec{U}^{\,i}$ = The unit vector pointing from the camera to the target of *i*th frame $[U^i_E \quad U^i_N \quad U^i_H]$

Can be shown as $\vec{U}^{\,i} = R^i_{FB} \cdot \overrightarrow{Obs}^{\,i} \;/\; |\overrightarrow{Obs}^{\,i}|$, where

$$R^i_{FB} = \begin{bmatrix} c(f^i)c(k^i) & s(w^i)s(f^i)c(k^i)+c(w^i)s(k^i) & -c(w^i)s(f^i)c(k^i)+s(w^i)s(k^i) \\ -c(f^i)s(k^i) & -s(w^i)s(f^i)s(k^i)+c(w^i)c(k^i) & c(w^i)s(f^i)s(k^i)+s(w^i)c(k^i) \\ s(f^i) & -s(w^i)c(f^i) & c(w^i)c(f^i) \end{bmatrix}$$

R^i_{FB} is a rotation matrix and use the shorthand c for cos and s for sin.

If there are [n] frames that have position and orientation data, we have 3n equations to solve n+3 unknowns that contains the targets three dimension coordinates and N unknown scaling factors. Can use least squares adjustment to calculate the unknowns.

$$\begin{bmatrix} 1 & 0 & 0 & -U_E^1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & -U_N^1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -U_H^1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & -U_E^2 & \cdots & 0 \\ 0 & 1 & 0 & 0 & -U_N^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 & 0 & \cdots & -U_H^n \end{bmatrix}_{[3n][n+3]} \begin{bmatrix} T_E \\ T_N \\ T_H \\ S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix}_{[n+3][1]} = \begin{bmatrix} P_E^1 \\ P_N^1 \\ P_H^1 \\ P_E^2 \\ P_N^2 \\ \vdots \\ P_H^n \end{bmatrix}_{[3n][1]}$$

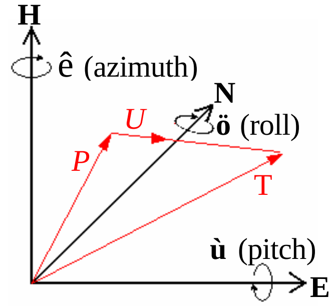
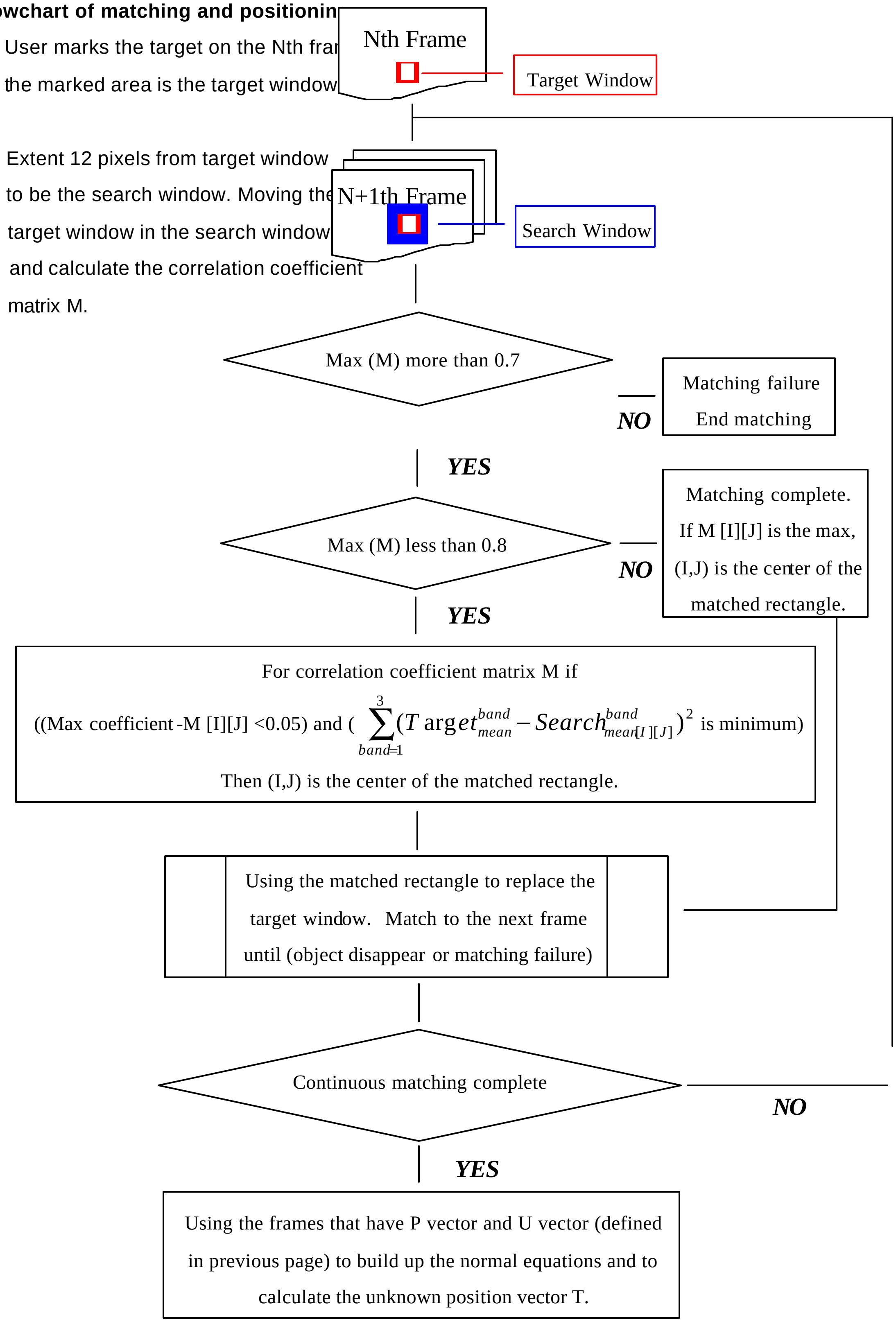


Figure2. The geometry of target observation and the angle definition

5.Flowchart of matching and positionin

- User marks the target on the Nth fra
  the marked area is the target window
- Extent 12 pixels from target window
  to be the search window. Moving the
  target window in the search window
  and calculate the correlation coefficient
  matrix M.



6.Results and Conclusion

Figure3. Mark 3 targets on the frame

We choose 3 targets on the frame for test (Fig3). The red arrow on Fig4 means the camera. The black circle means the estimative target position and the black rectangle means the real position of the target.

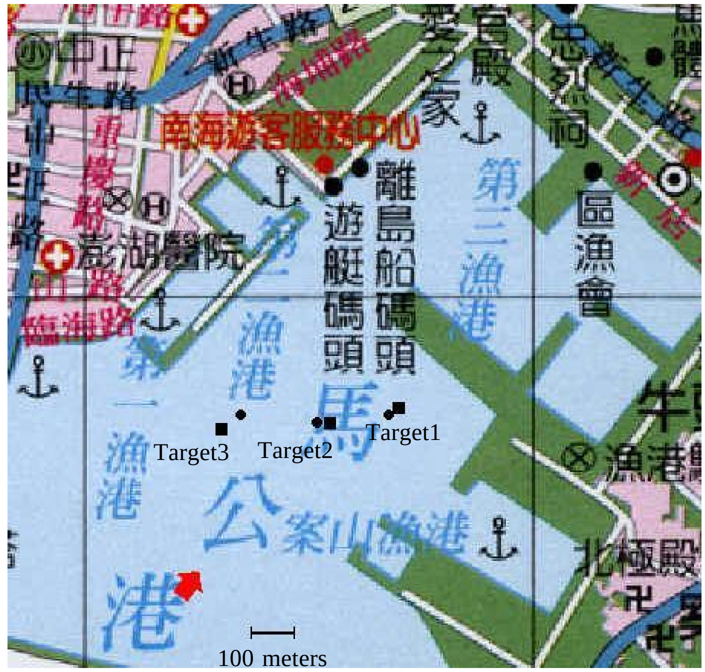


Figure4. Digital map of test area

The position error of target1 is about 23.65 meters and target2 is 26.16 meters but 49.8 meters in target3. Error consists of GPS error, tile meter error a nd matching error. If the target is not obvious enough to the frame, it will lead matching failure. The proposed method canlocate the target's position roughly and without any ground control point

Reference:

[1] A. Sharifinejad and H. Mehrpour , "A Fast full search block matching algorithm using three window search based on the statistical analysis of the motion vectors", IEEE International Conference on 2002 Page(s): 104 -108 vol.1

[2] Lurng-Kuo Liu; Feig, E.,"A block-based gradient descent search algorithm for block motion estimation in video coding", Circuits and Systems for Video Technology, IEEE Transactions on, Issue: 4, Aug. 1996 Page(s): 419 –422 vol.6