OBJECT TRACKING AND POSITIONING ON VIDEO IMAGES

Chi-Farn Chen, Min-Hsin Chen
Center for Space and Remote Sensing Research, National Central University, Chung Li, TAIWAN
cfchen@csrsr.ncu.edu.tw
55

Commission PS ICWG V/III

KEY WORDS: Video, Target, Tracking, Object, Matching

ABSTRACT:

This study presents an application of digital video camera for object tracking and positioning. The main purpose of the study is to automatically trace and position a motionless object on the video images that are recorded on a moving vehicle. Because of the change of the focal length of camera and the movement of the vehicle, the appearance of a motionless object on the video images firstly will gradually change its shape and completely lose its trace after a period of time. The variation of the shape apparently will bring about the complexity for developing the automated algorithm to trace and position object on the video image. In this study, we develop a shape-based tracking technique to implement the tracking task. The technique uses the shape matrix algorithm (SMA) that has scale and rotation invariant characteristics to calculate the similarity of variant shapes between adjacent video frames. After the object is traced in every video frame, the photogrammetric collinearity condition equations are used to transform the object from the image coordinates to the ground coordinates. An experiment is performed to trace a motionless ship in the open sea. The result shows that the proposed method can successfully trace and position the ship even the ship had become entirely out of shape on the video images.

1. INTRODUCTION

Recently the digital video (DV) camera has become a popular monitoring tool because it is smaller, lighter and easier using than traditional one. By combining DV with a moving vehicle (e.g., helicopter or airplane), it can record a sequence of images conveniently and effectively. This study presents a delicate algorithm that can automatically trace and position a motionless object on the video images, which are recorded on a moving helicopter. The proposed method may be divided into three sequent steps: (1) the acquisition of the position and angular orientation of the camera; (2) the estimation of the image coordinates of the object; (3) the construction of the ground coordinates of the object. The first step is designed to find out the position and angular parameters of the video camera. In order to meet the strict and complex recording environment on the helicopter, a hardware system integrates tilt-meter, GPS, and digital compass with DV is developed in this study. The main function of the system is to record the video images and the camera orientation synchronously. The second step is to estimate the image coordinates of the object on the video images. The main goal of the step is to implement the task of object tracking on the video images. Since the object in our study changes the shape all the way in the recording period, this study uses color-based segmentation (Pei, 1999) and shape-based matching algorithm (Flusser, 1992; Flusser, 1995) to trace and locate the object. The final step is to calculate the ground coordinates of the object on the image sequences. Linking up the camera's orientations (from steop1) with object image coordinates (from step 2), the collinearity condition equations can be used to calculate the ground coordinates of the object.

The organization of this paper is as follows. Section 2 introduces the information of the hardware system. Section 3 describes the proposed object segmentation process. In section 4, the shape-based object matching and tracking techniques are shown in detail. The positioning method will be shown in section 5. Finally, the experiment results and conclusions are addressed in section 6 and section7.

2. THE HARDWARE SYSTEM

In order to record the video images and orientation information of the camera synchronously, wherefore this study integrates GPS, digital compass and the tilt meter with DV through an encoded-decoded hardware device. The type of digital video camera used in recording data is SONY PC115. 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 +/- 3°, and the GPS accuracy is about 10~15 meters. The integrated system basically transfer the orientation data from digital format, acquires from the devices mentioned above, into analog voice. Therefore the video data and orientation parameters can be recorded through both video and audio channels synchronously. Eventually, after the recording procedure is accomplished, by making use of the decoded device and Microsoft DirectX component, the camera's orientation and the corresponding video signals can be reconstructed from analog data to digital images.

3. OBJECT SEGMENTATION

In this section, the detail of object segmentation methods will be presented. Dissimilar to the tracking object on the unchanging background, it is difficult to use simple differentiating method to segment the object on the varied background. Therefore, we use two-stage segmentation procedure to track the object on a sequence of images. The first stage uses the color features to segment possible regions of the target object from the images. The second stage employs Active Contour Model to describe the contour shape of the target object.

3.1 The Color Feature Segmentation

In the early perception stages of human beings, similar colors are always grouped together for further analysis. Based on this assumption, desired object can be identified by extracting the characteristic colors as the features. In this study, the desired object is manually selected by means of marking a rough polygon in the very first image. Then an unsupervised classification method is used to extract major color classes of the object and background. Accordingly, these color classes can be used to segment the object or background on the next images. The followings are the brief descriptions of the method.

For a given pixel **P** is one of pixels in the search window on next image. **{O$_c$}** is the class centers of the object and **{B$_c$}** is the class centers of background. **T** is the threshold of spectrum distance.

If (P.Minimum Distance to {B$_c$}<T)
P belongs to background class.
If (P.Minimum Distance to {O$_c$}<T)
P belongs to object class.
Else
P is a part of background class.

After the color-based segmentation is finished, it is possible to obtain a binary image with object separated from background. Normally, the desired object segment should be a solid region; therefore, the erosion, dilation and eight-neighbor object labeling algorithms are employed to eliminate the error pixels. Moreover, it is necessary to describe the object as a close contour, an energy-minimizing edge segmentation algorithm, which is applied to represent the shape of the object, will show in next sub-section.

3.2 Shape Feature Extraction

Active Contour Model (ACM) (Kass, 1987) is one kind of parametric curves presentations, which is defined within a curve domain. The curves of ACM can move under the influence of internal forces cause by the initial curve and external forces supplied by image data. ACM is widely used in image processing applications, such as edge detection, segmentation and particularly to locate object boundaries. ACM transfers the boundary detection problem in image domain to energy-minimizing problem in curve domain. The traditional energy functions in ACM are defined as follows:

$$E_{Snake} = \sum_{i=1}^{N} (E_{int}(s_i) + E_{ext}(s_i)) \quad (1)$$

$$E_{\text{int}}(S_i) = \alpha \left| S_i - S_{i-1} \right|^2 + \beta \left| S_{i-1} - 2S_i + S_{i+1} \right|^2 \quad (2)$$

$$E_{ext}(S_i) = -\left| \nabla f(S_i) \right| \quad (3)$$

Where E_{Snake} = the snake energy of the contour.

S_i = the ith position of contour.

N = number of snaxel.

E_{int} = the internal energy at snaxel S_i .

E_{ext} = the external energy at snaxel S_i .

α, β = the weighting functions are defined to control the relative importance of the elastic and bending terms.

Traditional ACM has two problems, initialization and convergence in concave regions. The initialization problem means that the initial contour has to close to the object, because the potential force of traditional ACM is generally small. Due to the ACM has no extra pressure force on concave region, the contour is often across the boundary concave. In order to solve the problems, an extra external force, which is named Gradient Vector Flow (GVF) (Xu and Prince, 1998) can be used to improve the result of ACM. GVF is a diffusive filed, which is computed by gradient vector of a gray-level map derived from the image. When the point in the field is near to object's boundary, the GVF field will move toward the boundary. Moreover, GVF field will change smoothly in the homogenous region of the image. Therefore, it indicates that GVF not only can provide a larger buffer in initial contour but also can converge to the concave region.

By using GVF Active Contour Model, the shape of possible objects, which are segmented by color features on next image, can be extracted automatically. Fig.1 shows the flowchart of object segmentation.

4. SHAPE-BASED OBJECT MATCHING AND TRACKING

It must be noted that the color segmentation may generate a number of possible objects on the subsequent images. Moreover, the movement of the video camera definitely will keep changing the shape of the object. Apparently both various possibilities and varied shapes of the objects will bring about the complexity for tracking the desired object on the subsequent images. This study uses the shape derived from ACM along with Shape Matrix Algorithm (**SMA**) to trace the desired object between adjacent images.

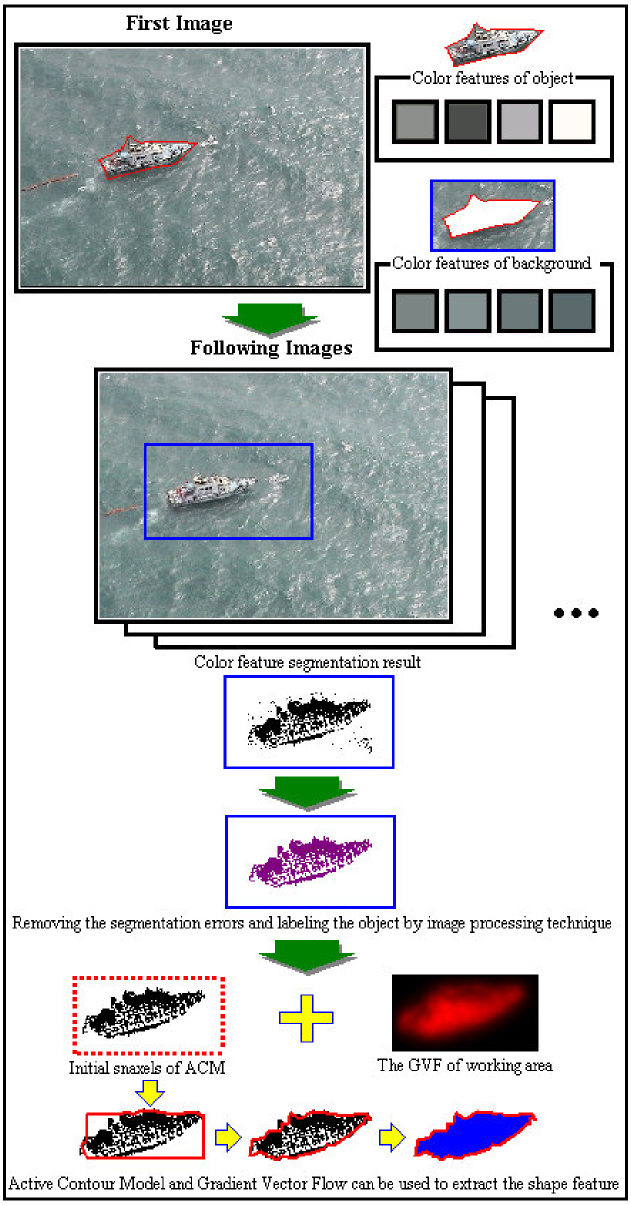

Fig.1 Flow chart of object segmentation.

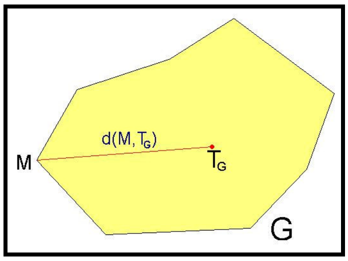

Fig.2 Object G in the image domain.

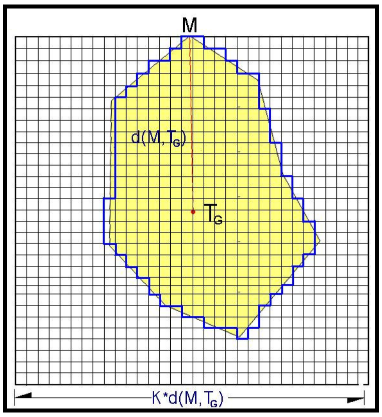

Fig.3 Shape of object G can be reconstructed from SMA.

4.1 Shape-based Matching

SMA is a kind of invariant shape description, which defines the gravity center of object as the origin and the longest axis of object as main axis. The shape of object on the image can be reconstructed aas a shape matrix from SMA (Fig.2, Fig.3). By comparing the shape matrix B^{OT} with B^{OS}, the degree of similarity between two objects OT and OS can be calculated. Because the dimensions of the matrixes have to be equal, hence the dimension of the matrix **n** and the similarity p (B^{OT}, B^{OS}) are defined by following formulas:

$$n = K * \max[\max_{A \in OT} d(A, T_{OT}), \max_{A \in OS} d(A, T_{OS})] \quad (4)$$

$$p(B^{OT}, B^{OS}) = 1 - \frac{1}{n \times n} \sum_{j=1}^{n} \sum_{i=1}^{n} \left| B_{ij}^{OT} - B_{ij}^{OS} \right| \quad (5)$$

Where d(A,T) = the distance from A to T.
K = the similarity accuracy factor.

The factor K influences the accuracy of similarity; the higher K can describe the shape more exhaustively. Considering saving the compute time, we set K=3 in this study. By using SMA, the similarity between two objects can be calculated, accordingly, the object region on the next image can be found too. SMA not only finds the most similar object on the next image, but also figures out the difference of scaling and rotation between the two shapes.

4.2 Feature Point Tracking

The shape region can be extracted automatically by preceding steps, consequently, the next step is to track the feature point of the object. The area-based matching algorithm can match point feature precisely, nevertheless, the success of area-based matching algorithms highly depends on the invariance of the target window and search window. By means of the coefficients of SMA, it is posssible to eliminate the shape change of the object. Eventually, the center position of the best matching block which with highest probability is the tracking result.

The size of target block is 7 by 7 in this research, and the target block is given by the result of TDGO (Lue, 1988). TDGO is an interest operator, which can find the most obvious feature point in the image. The boundary of search window is decided by the object's boundary rectangle on the next image. This study selects the Mean Square Error (MSE) to be the objective function of area-based matching. MSE measures the magnitude of error as a result of two blocks' comparison, lower error means higher probability of two blocks are similar. The details of MSE are shown as follows.

Let B_{pq} is the candidate block with position (p,q) in next image and T_{rs} is the target block with position (r,s) in current image. The Mean Square Error of two blocks is given as:

$$MSE(B_{pq}, T_{rs}) = \frac{1}{N \times M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(B_{pq}^{ij} - T_{rs}^{ij} \right)^2 \quad (6)$$

Where: M,N are the dimensions of the block.
i,j represent the position of pixels in blocks.

5. OBJECT POSITIONING

As mentioned before, the orientations of the camera can be obtained from the hardware system, furthermore, the shape-based matching and feature point tracking can be used to compute the image coordinates of the object. Consequently, for each image that has orientation data, will give rise to set three equations. These will be derived from the following relationships between the recording position and the object's location in local geodetic system (E, N, H).

$$\vec{T} = \vec{P}^{\,i} + S^{i} \cdot \vec{U}^{\,i} \quad (7)$$

where

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

$\vec{P}^{\,i}$ = The ***i***th frame position vector $[P_E^i \ P_N^i \ P_H^i]$

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 $[U_E^i \ U_N^i \ U_H^i]$.

It can be shown as $\vec{U}^{\,i} = R_{FB}^{i} \cdot \overrightarrow{Obs}^{\,i} \,/\, |\overrightarrow{Obs}^{\,i}|$,

where R_{FB}^{i} =

$$\begin{bmatrix} c(\phi^i)c(\kappa^i) & s(\omega^i)s(\phi^i)c(\kappa^i)+c(\omega^i)s(\kappa^i) & -c(\omega^i)s(\phi^i)c(\kappa^i)+s(\omega^i)s(\kappa^i) \\ -c(\phi^i)s(\kappa^i) & -s(\omega^i)s(\phi^i)s(\kappa^i)+c(\omega^i)c(\kappa^i) & c(\omega^i)s(\phi^i)s(\kappa^i)+s(\omega^i)c(\kappa^i) \\ s(\phi^i) & -s(\omega^i)c(\phi^i) & c(\omega^i)c(\phi^i) \end{bmatrix}$$

R_{FB}^{i} is a rotation matrix and use the shorthand **c** for cos and **s** for sin.

If there are **N** images that have orientation data, it will exist 3***N** equations and **N**+3 unknowns. The unknowns include the object's three dimension coordinates and **N** scaling factors. The least squares adjustment is suitable to solve the following linear equations.

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

6. EXPERIMENT RESULTS AND CONCLUSIONS

The experiment of this study was designed to trace and position a motionless ship near the coast. The video images were recorded at a frame rate of 10 fps and with image size of about 320 pixels by 240 pixels. The ground survey was also used to obtain the coordinates of the ship that will be used as the accuracy check. A part of the tracking results are shown in Fig.4 (a), Fig. 4(b) and Fig. 4 (c). The red contour on the image represents the result of tracking, and the green cross is the result of feature point tracking. The results indicate that even though the shape of the object changes entirely, the proposed algorithm not only can trace the contour of ship, but also can trace the feature point on the ship. The ground coordinates of feature point on ship are also calculated by the collinearity equations. A position comparison is made between the feature point on the ship from the video images and the ship itself from the ground survey. The comparison indicates that a position discrepancy of about 35 meters can be found. The position discrepancy basically can be attributed to the limited precision of the orientation recording devices. However, if the position of ship includes the range of the whole ship body (the ship has the length of at lest 100 meter), the position discrepancy can be considered acceptable.

Figure 4(a). 1ST image

Figure 4(b). 147th image

Figure 4(c). 258th image

7. CONCLUSIONS

In this study, we have presented the color-based and shape-based algorithms for object tracking and positioning on the video images recorded from a moving helicopter. The methods use color features as the basis to segment the desired object on the images and then employ shape features to track the object and position the feature point of the object on every image frames. The feature point of the object is then transformed to the ground coordinates with the aid of the orientation parameters of the camera. The position result from the video images is compared with that of the ground survey. The quantitative assessment indicates that the proposed method can position the object on the video image within the acceptable level of the accuracy.

8. REFERANCES

Flusser, J., 1992. "Invariant Shape Description and Measure of Object Similarity"., In : *4th International Conference on Image Processing and its Applications*. (IEE Conference Publication 354). IEE, pp. 139-142.

Flusser J., 1995, " Object matching by means of matching likelihood coefficients ", *Pattern Recognition Letters*, 16,pp.893-900.

Kass, M. A. and D. Terzopoulos, 1987. "Snakes: Active contour models", in Proc. *1st International Conference Computer Vision*, London, pp. 259-268.

Lue, Y., 1988. "Interest Operator and Fast Implementation, International". *Archives of Photogrammetry and Remote Sensing , International Archives of Photogrammetry and Remote Sensing*, 27(□)□491-500.

Pei, S. C, 1999. "Extracting Color Feature and Dynamic Matching for Image Data-Base Retrieval". *IEEE Transactions on circuits and systems for video technology*, vol. 9, NO 3.

Xu, C. and J.L. Prince, 1998. "Snakes, shapes, and gradient vector flow" *IEEE Trans Image Proc*. Vol 7,pp 359-369.

Xu, C. and J. L. Prince, 1998. *"Generalized Gradient Vector Flow External Forces for Active Contours", Signal Processing— An International Journal*, vol. 71, no. 2, pp. 131-139.