The task of detection and tracking of a moving object is addressed. An algorithm has been developed which performs this task for monitoring and surveillance purposes. Prediction is also implemented in the algorithm to resolve the events of occlusion or masking, and also to increase the normal tracking performance. Real-time implementation generates deformation in the target appearance, and then a shape database is also used to improve losing target situation. A prototypical system has been developed that makes use of a moving camera located on a robotized system. A case study is presented about animal tracking in infrared live video.

Introduction

Real-time object tracking from image videos revealed in an open environment is still a challenging task. Current approaches are based on successive frame differences [1, 2], on trajectory tracking using weak perspective and optical flow [3], or on region approaches, defining active contours of the target objects, and neural networks to perform the movement analysis [4]. Also approaches using adaptive threshold techniques have been developed to detect the points that are moving in a coherent way through different frames [5], or to perform a motion detection, and a successive region segmentation [6].

Segmentation techniques have been also used to cluster pixels into regions corresponding to single objects on the basis of gray level and proximity, then regions are merged according to local motion estimation [7].

However, a changing background, related to inspecting the scene with a moving camera, can introduce great difficulties. In this paper we present an algorithm for object tracking in an image sequence acquired with a moving camera, when the background is variable. Preliminary results are also shown to the monitoring of animal movements at night in an open environment (i.e. natural reserves or parks) using near and far infrared (IR) vision.

Problem statement

Monitoring and tracking moving objects in a free and open environment can be subdivided into different sub-tasks:
- Target selection
- Target characterization
- Target tracking

The approach followed in this research lies on the hypothesis of working with infrared (IR) images, which make the system more robust and more invariant to light changes in the scene. The processed sequence is composed of gray levels images (i.e. frames or thermographs) of high temperature target (with respect to the major part of the scene).

Target selection, due to the characteristics of IR images which are clearly contrasted, can be obtained interactively by manual intervention of a human user.

Target characterization can be obtained through a simple and approximate segmentation together with a characterization of the target on the basis of an a priori knowledge base, to determine the object shape.

Target tracking procedure with a moving camera has to take into account the camera movements. These are typically in the opposite
direction with respect to the target motion (i.e. aiming to keep the target centered), causing jagged motion.

Algorithms for motion detection and tracking have to consider the presence or absence of all these aspects in order to have a significant performance.

**Approach and Techniques**

The initial selection of the target is made manually by the user, which selects a point internal to the target (called *selection point*). Following this selection an automatic segmentation is performed to obtain the contour of the target object (Fig. 1). The segmentation is based on an edge detection algorithm performing a gradient descent along all the directions starting from the selection point (red cross in Fig. 1).

Contextually with the target initial selection, the user performs also a target class selection, this means that, among predefined classes, the user chooses what kind of target has been selected (i.e. human, small animal, large animal, bird, car...). This information together with the target shape information is stored and will be used during the automatic target research phase.

The segmented target from the first image of the sequence is then tracked automatically in each following frame.

The features that are used for the automatic tracking are the following:
- Use of local maxima
- Movement prediction on the basis of previous steps’ movements
- Prior knowledge about specific targets

For each frame the algorithm performs the steps to correctly identify the target and to follow it.

First the algorithm selects a candidate point ($C_A$) to be the centroid of the target in the actual frame. This selection is made using local maximum criteria, that is considering the contour segmented in the previous frame the internal point being the most similar to the previous centroid is selected ($C_1$).

As a second step the algorithm takes into account the previous movements of the target (i.e. identified with the centroid). The past steps trajectory of the centroid points is stored and then adapted to the actual step, in this way a new candidate point is obtained ($C_B$). If $C_B$ is not coincident with $C_1$ then a new point ($C_2$) is calculated between them following Eq. 1.

$$C_2 = \alpha C_1 + \beta C_B$$  \hspace{1cm} (1)

Where $\alpha$ and $\beta$ represent the weight assigned to one point compared to the other, and $\alpha + \beta = 1$.

To be sure that $C_2$ belongs to a valid object a local maximum search is performed in a circular fixed neighborhood of it. The search finds the point ($C_N$) that has the gray level closest to the previous centroid. Then $C_N$ is the selected centroid, starting from this point again the edge detection segmentation, based on gradient descent, is performed and a new contour for the target is outlined.

It is possible that the outlined contour does not correspond to the previous target, but to a wrong one. To recognize this event statistical parameters of the target are computed at each frame on the region enclosed by the contour: area measure, perimeter measure, average brightness of the area, standard deviation measure, skewness, kurtosis, entropy.

Once the new contour is outlined these parameters are evaluated and compared to the ones of the previous frames. If the actual parameters exceed $p$ times the standard deviation of the parameters computed for the last $n$ frames, then an alert is sent and
The search for the correct target starts ($p$ and $n$ are prefixed parameters). This event usually corresponds either to an occlusion (or partial occlusion or masking) of the target in the scene, or to a quick movement in an unexpected direction. The second possibility represents a very hard task to solve, but for the first event the algorithm performs an automatic research of the target trying to make a forecast about its motion.

The phase of automatic target search is performed in two stages:
1. Research of the centroid of the real target
2. Parameters check-up and confirmation

The research of the centroid is performed following the hypothesis that the target has been occluded but it is still moving in the same direction it was previously. Thus an estimation of the research direction along the interpolated trajectory of the last $n$ frames, not counting the actual supposed to be wrong, is obtained. Following this direction the first point with a brightness close enough to the last valid centroid is selected. Then considering this point the contour is outlined and parameters computed.

The parameters check-up consists of two steps: a direct comparison of the new values with the old ones, and a shape comparison with the shapes data-base.

The first step takes also into account that the target could be still partially masked, thus having different values (i.e. lower area value). To take into account this possibility the threshold $p$ for the acceptance of consistent parameters is increased.

In the second step the shape of the candidate target is compared with the shapes recorded in the data-base. The shapes to compare with are retrieved in the data-base using the information stored at the time of the initial target selection.

In the data-base, for each target class, the possible variations of the initial shape are recorded. If the actual shape is among these shapes, or within a tolerance threshold, then it is valid (Fig. 2).

The check-up fails if both two steps give a “wrong target” result.

If the check-up fails (i.e. the target is not the correct one), the search starts again from stage 1. If after $j$ frames the correct target has not yet been grabbed, the algorithm gives the control back to the user. The value of $j$ can be computed in the following way: considering the distance between the last valid centroid and the edge of the image along the research direction, and dividing it by the average velocity of the target previously measured (Eq. 2).

$$j = \frac{D(C_N;E_r)}{Avg}$$  \hspace{1cm} (2)

Where: $D(x,y)$ is the Euclidean distance between points $x$ and $y$; $E_r$ is the point crossing the edge of the frame along the research direction $r$; and $Avg$ is as stated Eq. 3

$$Avg = \left(\frac{1}{n} \sum_{i=1}^{n} D(C_N^{i-1},C_N^{i})\right)$$  \hspace{1cm} (3)

That is the average step length of the centroid in the last $n$ frames ($C_N^i$ is the centroid at the step $i$).

Finally in Fig. 3 a block diagram of the implemented algorithm is shown.
The algorithm modeled has been applied to a real case study for the tracking of animal movements in an open environment during the night, for the fauna monitoring in natural parks.

The video images were acquired with a thermo-camera in the IR range (wavelength range: 8-12µm). The camera was mounted on a moving structure allowing 360° pan and 90° tilt movements. The camera allowed different optical lenses, the tests were performed with the 12° and 24° optics; the geometrical resolution was 320x240 pixel.

The acquired images were quite noisy and were not pre-processed, the hardware for the pan and tilt control of the camera was only a prototype, but still the performance of the algorithm was considered good and very promising to further improvements from the point of view of the mathematical model and from the mechanical one.

References