Simultaneous Identification and Tracking of Moving Targets

Ahmed Shalaby
CVIP Lab
University of Louisville
KY, USA
ahmed.shalaby@louisville.edu

Asem Ali
CVIP Lab
University of Louisville
KY, USA
amali003@louisville.edu

Aly Farag
CVIP Lab
University of Louisville
KY, USA
aly.farag@louisville.edu

Abstract

This paper describes a framework for simultaneous identification and tracking of moving targets in random media. Video and IR thermal sensors are used to obtain the target signature. Classical Kalman filtering methods are implemented on targets with unknown trajectories. Computer vision methodologies are proposed to design a smart interceptor which identifies the targets based on shape and thermal signatures. The paper also describes a platform for basic studies in tracking of targets using vision-guided robotics. The system enables multiple object tracking and recognition.

1. INTRODUCTION

Automatic target tracking (ATR) is a real-world problem that takes various forms on continuous basis; examples include tracking ships in the ocean, commercial airplanes, unlawful intrusions into national airspaces, and satellite systems, to name a few. Tracking moving targets is quite difficult in adverse situations, and the field has progressed immensely since World War II. Tracking and identification of targets takes another dimension with camouflage, deformation and speed, which result in lack of discriminatory features for positive identification. For example, the well-being of satellite systems may be affected by debris in space. Likewise, the threat of stray rockets, surprise attacks and cyber warfare is not something unthinkable in the world we live in.

Computer vision methodologies are used in various real world applications for identification of targets before decision making. For example, military strikes in a dynamic theatre or battle field are based on the loop: intelligence, target identification and tracking, and action [11]. Another less dramatic example is in biomedical applications where certain blood cells are tracked (based on injection of Nano material) and identified for radiation therapy [6]. In the field of Radars, tracking systems are complex and dynamic with both radar and computing resources being shared by a large number of tasks with widely varying requirements. Managing these tasks so as to obtain the maximum benefit from the available resources is a difficult challenge. Environmental factors such as noise, heating constraints of the radar and the speed, distance and maneuverability of tracked targets dynamically affect the mapping between the level of service and resource requirements [4]. In this paper, we study intelligent interceptors of high speed moving targets. The approach hinges on conceptual smart sensor suite attached to the interceptor for target identification, and a robust mechanism for tracking the identified target. Operational scenarios follow the process of positive identification of the target.

Real world scenarios that require simultaneous identification and tracking are abundant. For example, astronomers are fond of studies of comets, asteroids and other objects/phenomenon that may hit the earth and affect the human life [7]. All these examples illustrate that viewing of high speed targets using smart sensors on the interceptors is already in practice and may be incorporated in various applications. Figure 1 illustrates generic systems for simultaneous identification and tracking.

In the following sections, we examine the various components of the above system and propose a vision-based system for simultaneous object identification and tracking.

2. OBJECT DETECTION

Tracking depends on the type of sensors used and the data at hand. In video tracking, for example, a common ap-
Table 1. Object detection Classes.

<table>
<thead>
<tr>
<th>Class</th>
<th>Related work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point detectors</td>
<td>Harris Detector (1988)</td>
</tr>
<tr>
<td></td>
<td>Affine Invariant Point Detector (02)</td>
</tr>
<tr>
<td></td>
<td>Scale Invariant Feature Transform (04)</td>
</tr>
<tr>
<td>Segmentation</td>
<td>Active Contours (1995)</td>
</tr>
<tr>
<td></td>
<td>Mean Shift (1999)</td>
</tr>
<tr>
<td></td>
<td>Graph Cut (2000)</td>
</tr>
<tr>
<td>Supervised</td>
<td>Support Vector Machines (1998)</td>
</tr>
<tr>
<td>Classifiers</td>
<td>Neural Network (1998)</td>
</tr>
<tr>
<td></td>
<td>Adaboost (2003)</td>
</tr>
</tbody>
</table>

proach for object detection is to use information in a single frame. However, some object detection methods make use of the temporal information computed from a sequence of frames in order to reduce the number of false detections. This temporal information is usually in the form of frame differencing, which highlights changing regions in consecutive frames. Given the object regions in the image, it is then the tracker’s task to perform object correspondence from one frame to the next to generate the tracks. Table I shows several common object detection methods in the computer vision literature [12]. Object detection is the starting point of tracking where many tools have been developed for this work.

2.1. Object Segmentation

Maximum a-posteriori (MAP) based segmentation framework of multimodal images is one of the important tools that have been developed in this area. In this frame, a joint Markov Gibbs Random Field (MGRF) model is used to describe the image. The main contribution of the work is more accurate model identification. For a known number of classes in the given image, the empirical distributions of this image signals are precisely approximated by a Linear Combination of Gaussian (LCG) distributions with positive and negative components. Gibbs potential, which is used to identify the spatial interaction between the neighboring pixels, is analytically estimated. Finally, an energy function using the previous models is formulated and is globally minimized using graph cuts. Many experiments have been conducted to validate the framework. This work is used as a preprocess step in many applications (e.g., medical imaging and scene understanding) [3].

2.2. Segmentation using Active Contours

Shape-based segmentation is a technique that uses both image and shape prior information. Our approach for the shape-based segmentation problem is based on a specially designed level set function format. This format permits us to better control the process of object registration which is an important part in the shape-based segmentation framework. The method depends on a set of training shapes used to build a parametric shape model. The color is taken into consideration besides the shape prior information. The shape model is fitted to the image volume by registration through an energy minimization problem. The approach overcomes the conventional methods problems like point correspondences and weighing coefficients tuning of the partial differential equations (PDE’s). Also it is suitable for multi-dimensional data and computationally efficient. The approach is validated against different image modalities that include different objects [8].

3. OBJECT TRACKING

The aim of an object tracker is to generate the trajectory of an object over time by locating its position in every frame of the video. Object tracker may also provide the complete region in the image that is occupied by the object at every time instant. The tasks of detecting the object and establishing correspondence between the object instances across frames can either be performed separately or jointly. In the first case, possible object regions in every frame are obtained by means of an object detection algorithm, and then the tracker correspond objects across frames. In the latter case, the object region and correspondence is jointly estimated by iteratively updating object location and region information obtained from previous frames. In both tracking approaches, the objects are represented using the shape and/or appearance models.

The model selected to represent object shape limits the type of motion or deformation it can undergo. For example, if an object is represented as a point, then only a translational model can be used. In the case where a geometric shape representation like an ellipse is used for the object, parametric motion models like affine or projective transformations are appropriate. These representations can approximate the motion of rigid objects in the scene.

For a nonrigid object, silhouette or contour is the most descriptive representation and both parametric and nonparametric models can be used to specify their motion [12]. As we will see in section 5, it will present some experimental results for target tracking based on classical Kalman filter.

4. OBJECT IDENTIFICATION

This section deals with object modeling as viewed by optical and IR sensors. Section 4.1 and 4.2 summarize the optical imaging process, whereas Section 4.3 summarizes the process of thermal imaging using light wave IR sensors.

4.1. Camera Model

As a simulation for image formation, we use the pinhole camera model. The image formation process in the ideal pinhole camera system constitutes the basics for the conventional perspective camera model, which is used in com-
computer vision. In this model there are three main components, the camera center, which is the perspective projection center, and the retinal plane (image plane), which is located at a distance equal to the focal length from the camera center. The image of a 3D scene point is the intersection of a line connecting this 3D scene point and the camera center (As shown in the Figure 2.)

The calibration process is to find the model parameters. These parameters mainly consists of two sets: the extrinsic parameter \([R|T]\) (rotation and translation that convert the 3D scene point from an arbitrarily world coordinates system to camera coordinates system.) and the intrinsic parameters \(K\) (focal length and viewing angles, pixel size).

4.2. Simulation of different operational scenarios of ballistic missile

In this section, we study and simulate two different scenarios of simultaneous identification and tracking of objects, Intercept Scenario and Rendezvous Scenario.

4.2.1 Intercept Scenario

For the first scenario, as shown in Figure 3, the sensor will be at a certain distance from the object and it will remain within this distance for a certain period of time. Then sensor intercepts the object. To detect the object according to this scenario, we need a special sensor (lens with large focal length and with high resolution image). We can simulate the trajectories of the sensor and the object using the model of two-body motion with first body guided to second body. Figure 4 illustrates simulated trajectories for the object and the sensor and an example of simulated images for a sphere object.

According to the simulation of this scenario (as shown in the Figure 4) the sensor will be at distance 1500 meter from the object after 32.4 seconds. After that, the sensor will remain within this distance for about 0.6 seconds. Then sensor intercepts the object at the 33 second. Therefore, for this scenario the sensor should have the following specs:
- Range to detect a target is 1500m.
- Frame rate greater than 2 Fps.
- Field of View greater than 0.02 X 0.1 degree.

4.2.2 Rendezvous Scenario

For the second scenario, as shown in Figure 5, the sensor will be at a certain distance from each object and it will remain within this distance for a certain period of time. Then sensor move away from the object. To detect the object according to this scenario, we need a special sensor (lens with large focal length and with high resolution image). The sensor should have a large field of view to detect all the target objects. Figure 6 illustrates simulated trajectories for the objects and the sensor and an example of simulated images for five objects (e.g., sphere, cone and cylinder). According to the simulation of this scenario (as shown in the Figure 6), the sensor should have the following specs:
- Range to detect a target is 13000 m.
- Frame rate greater than 0.5 Fps.
- Field of View greater than 0.02 x 0.04 degree.

Which it requires that using cameras with very large focal lengths as shown in Table 2 for the different objects.

One can notice that these focal lengths need huge lenses which are not practical. So, we need to focus on IR sensor as we will discuss in the next section.
4.3. Thermal Image formation process

A blackbody is essentially an opaque object in the spectrum under consideration, which means it has a high emissivity; ideal black bodies absorb all radiation incident on them and therefore have an emissivity $\varepsilon = 1$. The blackbody radiation’s spectral distribution depends only on the absolute temperature of the blackbody [5] and is given by Planck’s blackbody radiation law in Equation 1. Wien’s law in Equation 2 is used to determine the location of the maximum amplitude of Planck’s law for a given temperature/wavelength pair. These equations are illustrated in Figure 7 for temperatures suitable for humans.

$$P(\lambda, T) = \frac{8\pi hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1} \text{[Watts / m}^2]$$ (1)

$$\lambda_m T = 2.898 \times 10^{-3} \text{[mK]}$$ (2)

The thermal image formation process is illustrated in Figure 8. To acquire a thermal image, an object that emits thermal IR radiation is placed in front of an IR lens. The IR lens focusses the radiation onto the focal plane array (FPA) IR sensitive detectors. The FPA utilizes the photoelectric effect to generate electrical signal. These signals are multiplexed and converted to a digital electrical signal, which is processed further to handle internal noise sources before being output to a video processor. The video processor converts the data into a 2D video frame, which is then captured via a frame-grabber in the computer. These images are returned as gray-levels, where the intensity of the gray level is proportional to the temperature on the object’s surface [1].

The newest generation of quantum thermal IR FPAs is capable of detecting two or four spectral bands and can render a multicolor image, i.e. one per band. Both MCT photodiodes and quantum well infrared photodetectors (QWIPs) offer the multi-color capability to image in the MWIR and LWIR bands, respectively. Gunapala et al. even recently reported the recent development of a dualband, one megapixel MWIR/LWIR FPA [9]. In addition to larger FPA sizes, the general goal of these third-generation systems is to provide enhanced capabilities such as higher frame rates, better thermal resolution, multi-color functionality and on-chip functions. The current definition of third generation is defined to maintain the technological advantage held by the US and allied forces and applies to cooled and un-cooled FPAs [10].

In this work, we examined the possibility of extracting object signatures from IR sensors. We report some preliminary results on small spherical objects with different colors, surface thickness and contents.

Table 2. The focal lengths of the required cameras.

<table>
<thead>
<tr>
<th>Object</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focal length(m)</td>
<td>550</td>
<td>3000</td>
<td>200</td>
<td>65</td>
<td>50</td>
</tr>
</tbody>
</table>

In order to create what could be an IR signature for objects, we imaged the objects using our LWIR camera at various temperatures. The histograms were obtained at each temperature and linear combinations of Gaussian (LCG) kernel fitting for the average histogram were performed [2].

Figure 9 illustrates that although the histogram of visible images are almost identical in cold and hot cases, the average histogram of thermal images are different. That’s called the IR signature of the object. This is very important factor for positive identification of the desired target to be tracked.
Figure 9. Generating a signature from LWIR images of test objects. A linear combination of Gaussian Kernel fitting of the average histogram shows distinctions between objects.

5. EXPERIMENTAL RESULTS

As shown in Figure 9, the modes of the LCG and their locations show distinctions with objects, and therefore, they may be used as an aspect of the object signature. These signatures should be used in conjunction with other features for object identification. Tracking, which starts from actual or estimated object locations in the field of view of the sensors, should be affected by the identification process. In fact, the decision to further track or collide the object will depend on the identification step.

On the other hand, in order to test the interplay between object characteristics, trajectory and the tracking mechanism we established a vision-guided robotic system formed of a robot arm on which the control is vision-guided. The system is shown in Figure 10. Objects were mounted in the ceiling by a fine wire enabling free motion with controlled speed and pattern. The objects are to have different shapes, appearance and temperatures, and to move on random trajectories.

The system is quite flexible in terms of range of motion and targets. Speed is the major limitations due to the work space constraints and the on board robotics interface. Yet, the system provides a controlled mechanism to test and validate the Kalman filtering approach to object tracking based on system identification of the objects from LWIR images. Figure 11-15 show some off-line experiments that illustrate the overall process of identification and tracking using vision-guided robotics.
6. CONCLUSION AND EXTENSIONS

In this paper we examined video sensors and thermal imagery for identification, and Kalman filter for tracking. The vision-guided robotics platform described in the experimental part is an attempt to study, at a down-to-earth scale of size, motion and weight, some of the issue involved in simultaneous identification and tracking of targets. The platform allowed study of thermal and video sensors and extraction of "signatures" of targets at different speeds and temperatures. The robot arm enabled validation of the tracking mechanism since the exact route of the targets may be known a priori, and the robot arm will enable recording the behavior of the tracking mechanism.

Considerable algorithmic development is needed to make the robotic platform viable for simulation and validation of various scenarios. Likewise, various high resolution IR cameras may be incorporated to extract the heat pattern for the objects, in order to assist in signature construction and testing.

The current setup needs to be fitted with the latest wireless transmission technology, in order to make the robot motion easier to record and validate the performance of the tracking algorithm. Of course, the characteristics of the atmosphere and the types of uncertainties in the targets should be examined in order to reduce false positive detection. Therefore, better problem definition would be essential to make use of the proposed robotics platform.

On the sensors side, a combination of thermal, video, hyperspectral and GPS/positioning sensors may be used in designing a sensor suite for the interceptors, which will provide various capabilities to identify the targets of interest.

On the intelligence side, a number of approaches can be discovered for identification and tracking. Computer vision methodologies and machine learning techniques may be used to decipher the characteristics of the objects/targets in the theater. Once potential targets are detected, immediate identification would be needed in order to decide a plan of action by the interceptor.

References