Fusion of Multiple Mobile Cameras for Object Tracking

Jorge Eduardo Xalteno Altamirano[1], Leopoldo Altamirano Robles[1], Carlos Alberto Reyes[1], Luis Carlos Altamirano Robles[2]

[1] Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE)
[2] Benemerita Universidad Autónoma de Puebla (BUAP)

Abstract:. In this work, we propose a novel way to use sensor fusion based on DSmT approach, in order to track objects using mobile cameras. We use two kinds of camera movements, namely, rectangular and circular movement, and we show that camera movement is a very practical way to deal with partial occlusion. By using accumulative evidence, our approach can detect objects in scenes even if objects are partially occluded, tracking objects where static cameras can not. Experimental results show that our proposal is better than approaches based on static camera sensor fusion for object tracking.

Keywords: Dezert-Smarandache Theory, sensor fusion, mobile sensors.

1 Introduction

Surveillance task can be seen as the process of detecting objects of interest in a sequence of photos (video); these objects can be in motion or they can be statics; if objects are in motion, the task is to follow these objects in order to determine if the behavior or movement of the object is of interest for the observer. Objects tracked are usually people or vehicles; in the case of people, it could be interesting to detect if the person is making a felony, for instance; in the case of automobiles, it would be interesting to determine if the automobile is committing a traffic foul.

Usually, the surveillance of places is done with a sensor placed strategically, in such a way, that the scene of interest is covered properly by the sensor. However, in many scenes it is impossible to cover the whole field of interest with only one sensor. Therefore, it is necessary to use multiple sensors placed properly [7].

In applications that use several sensors, information from one sensor is used to combine and complement the information from other sensor [2]; There are several ways to combine information from multiple sensors as Dempster-Shafer or Dezert-Smarandache approaches [3]; in this work, we use the last one approach because it was proved that is better than other ones, for working with uncertain information [3].

There exists another problem, it's the one of object partial occlusion, it happens when an object is placed between the target and the camera for a while, there are approaches to

Special Issue in Advances in Pattern Recognition
Research in Computing Science 44, 2009, pp. 37-44

solve this problem that use a predictor to approximate the position of the object, however, if the object has a non-predictable movement, the approximation will not be accurate.

In this work we propose an approach based on the cameras movement and in sensor fusion to solve this problem, this approach helps to follow the object and to place it more accurate on the scene.

2 Sensor Position and Movement

The sensors are placed in such way that they cover the same area viewed from different places; in this way, they can have a different sight of the scene and contribute with their information to the sensor fusion. We used two sensors to do our experiments; these sensors were moved in two ways: rectangular and circular styles of movement. These kinds of movement were selected not only for experimental proposes, but also for analyzing different sources of evidence.

In Fig 1.a we show the first kind of movement used in this work (rectangular); in these case the sensors move from side to side horizontally; the movement describes a straight line and the sensors are moved with a regular speed.

The second kind of movement is in circular way; in this kind of movement the sensors move around a point placed in the center of the scene. Fig. 1.b shows a scheme of this movement.

It is important to mention that with mobile sensors, we can cover a wider Field Of View (FOV), in opposition to the approaches where the sensors are static. The movements are based on real life, when a person wants to see an object that is occluded by another object; the person usually moves its head to be able to see the target; in the case of sensors, the approach is the same: we move the sensor to have a better FOV but we also use the movement to deal with the problem of partial occlusion accumulating evidence during movement of sensors for object tracking.

3 Object Movement Detection

The object movement in a sequence of images can be seen as the change of coordinates of an object in the current image respect to the previous image. The movement detection is an important step in many autonomous surveillance systems. In this work, we used two different ways to detect object movement.

3.1 Images difference

The most obvious method to detect changes in a sequence of images is to compare two corresponding pixels to determine if they have the same gray value. In the simplest way, a binary difference $DP_{jk}(x,y)$ between two images $F(x,y,j)$ and $F(x,y,k)$ is obtained by:

$$DP_{jk}(x,y) \begin{cases} 1 & \text{if} | F(x,y,j)-F(x,y,k)>r \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where r is a threshold to be defined [6]. In the difference of images the pixels with value 1 are considered as the result of the object movement.

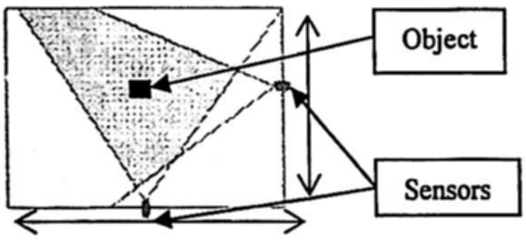


Fig. 1.a) Horizontal sensor movement

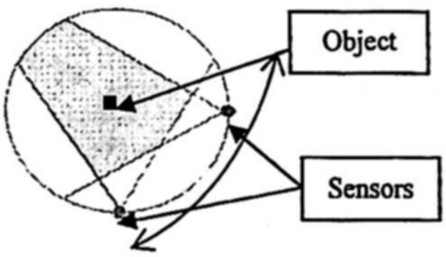


Fig. 1.b) Circular sensor movement

3.2 Size Filter

The result of the previous step is a binary image contaminated with noise; this noise is due to small changes between images; to eliminate this noise, we used a size filter. Simply, the pixels that do not belong to a minimum size group of connected pixels are eliminated.

4 Belief Creation

4.1 Object Characteristics

After the movement detection, we calculate characteristics of the object in the images. Namely: the vertical axis and the first and second statistical moments.

The vertical axis is calculated after the object is detected as [1]; vertical axis is obtained from the outer rectangle (the smallest rectangle that rounds the object), determining the line that crosses the rectangle from up to bottom exactly in the middle. An example of the vertical axis is shown in the (Fig. 2).

Fig 2. Outer rectangle and vertical axis of an object.

Other side, the statistical moments are calculated as follows:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \tag{2}$$

The first and second statistical moments are m_{00} and m_{01} respectively.

4.2 Believes or Masses

A belief is a data that tells us the certainty of an object to exist in the image [3]; it can be derived from any object characteristic. In this work, it has been done using the characteristics mentioned above.

For the vertical axis, the whole image is partitioned in cells with 20x20 pixels in size. Then believes are created according to the cells that the vertical axis crosses. It means that if the vertical axis passes through three cells, then there are three believes. The number of cells where the axis passes is the number of elements of the frame of discernment; this frame will be used in the fusion step. The values of the elements in the frame are the number of pixels that belong to the vertical axis in that cell.

For the case of moments, as we consider only two momentums, therefore there are only two elements in the frame of discernment. The values of these elements are the values of the moments.

Both cases, the vertical axis and the moments, are normalized so that the quantities of their elements are equal to 1.

5 Sensor Fusion

5.1 Dezert-Smarandache Theory (DSmT)

In 2002 Jean Dezert and Florentin Smarandache proposed a novel Information Fusion Model [3]; it is an evolution of the Dempster-Shafer model. The DSmT uses a frame of discernment Θ; this frame is a set of propositions $\Theta=\{\theta_1, \theta_2, \ldots, \theta_n\}$, that they are used to carry out the information fusion, DSmT has three main characteristics:

- The evidence combination refutes the third logical principle; this principle establishes that something "is or is not". The DSmT does not require this principle, therefore it can consider concepts related to the fuzzy logic.
- It is proposed a new combination rule.

- The elements from the frame of discernment are not necessarily exhaustive and exclusive.

The model DSm considers that the frame of discernment is dynamic. In this case some elements from Θ could be exclusive at a time, and they can not exist anymore at other time; this model is known as the free model ($\mathcal{M}^f(\Theta)$) of the Dezert-Smarandache Model.

5.1.1 Classical DSm Combination Rule

When the free model $\mathcal{M}^f(\Theta)$ is used, the combination rule used for a frame Θ with k elements is defined as in the equation (3):

$$m_{\mathcal{M}^f(\theta)}(C)=\sum_{\substack{x_1,x_2,\ldots,x_k\in D^\theta \\ x_1\cap x_2\cap\ldots\cap x_k=A}}\prod_{i=1}^{k}m_i(X_i) \tag{3}$$

6 Proposed Solution

The proposed solution has been implemented in a hardware/software prototype, it has served to show the effectiveness of the approach in real situations, even when in some experiments the scene is in a small scale; this does not gives less generality to the system, minimum adjustments have to be done for cover any other scene

The scene included objects in movement (people, animals, cars, etc), as well as static objects (trees, buildings, parked cars, etc), the system has operated in outdoor conditions. In order to carry out camera movement, we used one template for each position of the cameras.

6.1 Horizontal Movement

For the case of the horizontal movement of the camera, we use a sequence from PETS[5]; this sequence was taken with two cameras. The movement of the cameras is 5 pixels each frame. The cameras move from left to right and vice versa. For each camera we took some frames as templates in order to obtain the binary image. Once we have binary image, we used a size filter to clean the image. The result of the filter is, in almost all the cases, the tracked object. Fig. 4 shows the i-th template and Fig. 5 shows the i-th frame; the Fig. 6 shows the difference between the template and the frame (binary image).

In the binary image, we found the vertical axis and built the frame of discernment as explained in section 4.

In order to carry out the sensor fusion, the two sensors used in the system have to give information; if one of the sensors does not, the information from the other sensor is used as the output due to the inexistence of help from the other sensor.

Fig. 4. I-th template.

Fig. 5. I-th frame.

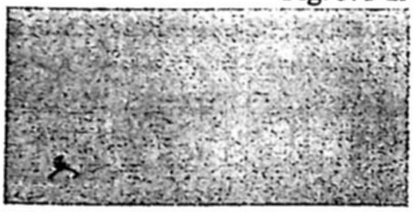


Fig. 6. Binary image.

The sensor fusion is done with the two frames and the output is an array of values; we take the largest value as the final output and the output tell us where the object is placed. As we know where the object is, we can evaluate the final result.

6.1 Circular Movement

In order to prove that the system works in other kind of scenes, we made a device that allows us to make a circular movement (Fig. 1.b). In this device, we placed two cameras placed 1.5m from the center of the scene and they have 90° of separation between them. In this experiment, we notice that the movement could serve to solve the problem of partial occlusion of the object.

6.2 Sensor Movement as a Solution of Object Occlusion

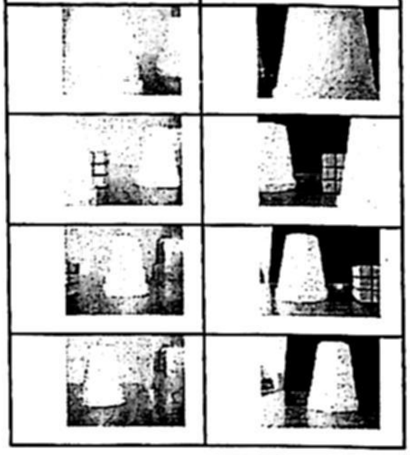
Such as a people moves the head in order to see an object occluded, the cameras movement of the sensors can be used to see an object that is partially occluded by another object. Notice that as the sensor moves, it sees a different FOV; the movement helps the sensor to see other interesting things; it is shown in table 1.

6.3 The Sensor Movement Increases the Evidence

The main difference between the work in [2] and our work is that the movement of the sensors contributes to obtain a larger evidence of the object; as the sensor moves, we make information fusion in each sensor; after 3 frames, the information that has been fusioned and accumulated in each sensor (accumulated evidence) is fusioned. The results show that the fusion of the accumulated evidence is better than just the only fusion in each frame.

Table 1. The movement of the cameras can be used to discover occluded objects.

Camera 1	Camera 2



6.4 Experimental Results

In the Table 2 we show that evidence grows with the accumulated fusion; it is due to the movement of the sensors: when they move, they can see a lager part of the object. For the time t and $t+1$ in Table 2 there are no information in the accumulated evidence, this is because the fusion between the sensors is not performed in these times; however in the time $t+2$ the evidence is larger than the single fusion in t, $t+1$ and $t+2$.

Table 2. Results of the accumulated sensor fusion versus the single sensor fusion.

	t	$t+1$	$t+2$
Single Fusion	[0.9106 0.0894]	[0.9819 0.0181]	[0.9106 0.0894]
Accumulated evidence	[0.000 0.000] (no sensor fusion)	[0.000 0.000] (no sensor fusion)	**[0.9982 0.0018] (Fusion)**

In table 3 the required time to calculate sensor fusion is shown. We can see that the accumulated fusion takes more time than the simple fusion; however the results are better with the accumulated fusion, respect to simple fusion.

We have compared the fusion based on DSmT versus an approach based on Bayesian fusion; the results are that the DSmT is more accurate than the Bayesian fusion due to the DSmT uses more information than the Bayesian approach. Figure 7(a) shows the regions were the object is and Figure 7(b) shows the results between the DSmT (red) and the Bayesian fusion (blue)

7 Conclusions

We have shown a novel approach to make mobile sensor fusion. The obtained results show that our proposal improves the evidence obtained respect to static sensors, due to the accumulation of evidence in terms of previous scenes. This is of especial interest when the objects are partially occluded respect to a fixed position of the sensors and therefore, the evidence generated by the fusion is almost null or totally null. Allowing the sensors to move, the system can discover partially or completely occluded objects, though they were static respect to the plane sensed.

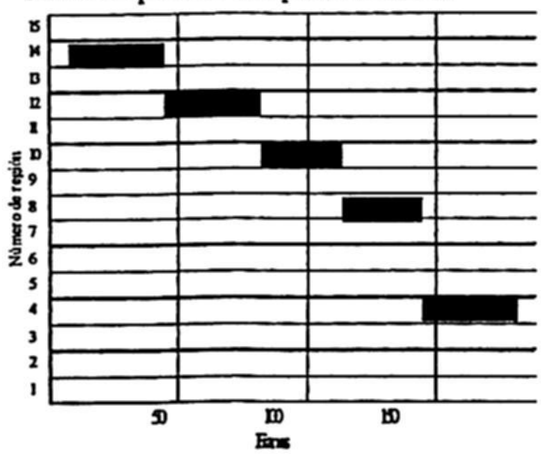


Figure 7 (a) Real object positions

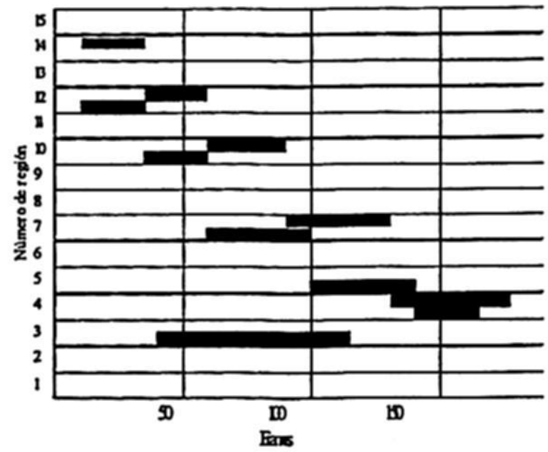


Figure 7 (b) Object position reported by DSmt (red) Bayesian fusion (blue)

References

1. Weiming Hu, Min Hu, et. Al: "Principal Axis-Based Correspondence between Multiple Cameras for People Tracking". IEEE Transactions on pattern analysis and machine inteligence April 2006
2. Esteban García (2008), "Fusión multicámara para seguimiento de objetos basada en Teoría evidencial". Tesis de maestría INAOE.
3. Floretin Smarandache, Jean Dezert: "An introduction to the DSm Theory for the combination of paradoxical, uncertain and imprecise sources of information". Online paper revised on August 1st 2006,
4. Aridiam Biswas, Pritwijit Guha et al: "Intrusion Detection and Tracking with Pan-Tilt Cameras". Second European conference on Computer Vision 1995.
5. PETS Repository: an Internet resource. Available from anonymous ftp: pets2001.cs.rdg.ac.uk/. University of Edinburgh..
6. Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, "Digital Image Processing using Matlab", Pearson Education.
7. Boudet, L., Midenet, "Pedestrian crossing detection based on evidential fusion of video-sensors", 2009 Transportation Research Part C: Emerging Technologies, Article in Press