Detection of thrown objects using ToF cameras

Serban Oprisescu, Laura Florea, Elena Ovreiu
Image Processing and Analysis Laboratory, LAPI,
University Politehnica from Bucuresti
Bucuresti, Romania
soprisescu [adt] imag.pub.ro

***Abstract*—This paper deals with the issue of thrown or dropped object detection as an application of the 3D time-of-flight (ToF) camera. The use of distance information provided by the ToF camera makes the segmentation step straight forward. Then, instead of high computational motion detection and estimation algorithms, mainly a labeling and tracking procedure is used. This approach provides real time low complexity object detection and trajectory estimation. The method is validated on 50 video sequences taken with the ToF camera. (*Abstract*)**

Keywords—3D ToF camera; image processing; thrown objects detection (key words)

I. INTRODUCTION

The detection of thrown objects is of interest mainly in the field of video surveillance or sports video analysis. Detecting potentially dangerous or illegal human behavior may be very useful in prison environment or in other person monitoring applications [1].

In this paper we will investigate the possibility to detect thrown objects using an acquisition system with a 3D time-of-flight (ToF) camera. More specifically we will focus on scenes where exists a single person that performs one of the two actions: throwing or dropping an object.

A. Related work

In the literature many solutions for the general problem of object tracking (either detection or trajectory estimation) have been proposed. General surveys on the topic may be followed in [2] and [1]. In the current paper we are interested in the specific movement described by throwing an object. According to the classification presented in [1], it falls under the category of point-like object (the object is small and can be represented by its centroid) with limited velocity variation (the direction and speed of the object do not change rapidly).

We note that most of the current techniques for thrown object detection are based on the motion analysis of video data captured with usual 2D surveillance cameras. For instance in [3] a technique for the detection of hand thrown objects in a video sequence is presented. The authors firstly identify areas of fast motion that fit certain size, compactness and speed criteria, and then use the expectation maximization algorithm to detect objects on parabolic trajectories (the normal trajectory expected for a thrown object) over a short time window. The overall detection rate is 74%.

The work has been co-funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/89/1.5/S/62557..

In [4] and [5] an application of throwing and catching objects is proposed as a replacement of normal transportation systems in a production flow. The motivation lies in the possible increase of parts assembly speed which will lead to higher productivity. In [5] the hit point of the thrown object is computed using trajectory data collected by several distance sensors. The authors use cylindrical objects with known geometrical parameters, propose a model of ideal throw trajectory and then validate their model with two high speed cameras and two precise laser sensors. The lasers are able to collect points on the surface of the flying cylinder with an accuracy of 0.02mm.

Our method does not need motion detection or other high computational techniques; it mainly relies on range image segmentation and object labeling, which is real time and with very low computational cost.

The paper is structured as follows: section II describes the proposed method, section III presents the obtained experimental results and section IV contains the conclusions and further work.

II. PROPOSED METHOD

The proposed method consists of three major steps: frame acquisition with ToF camera and pre-processing, content segmentation and movement analysis.

A. ToF Camera and data acquisition

The ToF camera is a camera system capable of estimating the distance to subject using the known speed of light, and measuring the time-of-flight of a light signal between the camera and the subject for each point of the image [6]. The ToF camera delivers simultaneously grey-level images and depth images of the scene and it is an "active illumination" system because the camera has its own illumination unit (a matrix of infrared LEDs) which emits modulated light (usually at a frequency of 20MHz). As shown in Fig. 1, the distance image is computed in each pixel using the phase shift between the emitted light wave and the received one (the farther the object, the bigger will be the phase shift). Since its introduction many practical applications have been developed around its functionality [7].

We note that nevertheless, the ToF principle has the disadvantage of the phase ambiguity problem: the phase shift can be only between 0 and 2π. A phase shift of 2π corresponds

978-1-4799-1494-4/13/$31.00 ©2013 IEEE

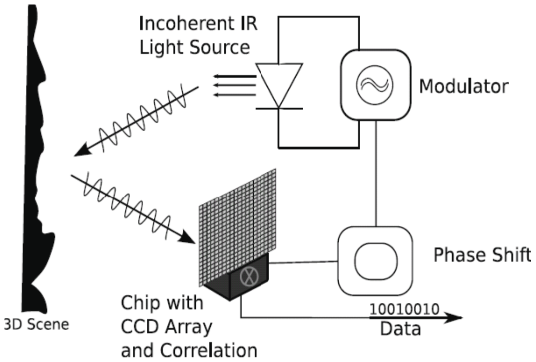

Fig. 1. The principle of ToF camera distance measurement

to the maximum measured distance, which for 20 MHz is of 7.5 meters.

To gather data for the hereby proposed experiments, we used the SR-3000 ToF camera that can achieve maximum 30FPS. The ToF camera delivers one distance frame after four exposures (four samples are needed to compute the distance). The acquisition speed can be adjusted by setting the integration time, which is the time of one exposure and typically is between 200 µs to 51.2 ms. But, the shorter the integration time, the higher will be the noise within the image (and the lower the confidence in the measured distance, especially for darker objects). Intensive testing with various different integration times, pointed out that a value of 7ms, as being the lowest value that could provide sufficient confidence in the measured distance. At this value, the acquisition speed is 22FPS, which proved to be enough for our goals. A typical visual example of the acquisition principle is shown in Fig. 1.

A visual overview of the proposed algorithm may be followed in Fig. 2.

In Fig. 3 a) an example of ToF intensity frame is shown, and in Fig. 3 b) the corresponding distance image is presented. The ToF camera delivers a 16 bit distance image; in Fig. 3 b) close objects appear darker and far objects appear brighter.

The noise due to the small integration time is typically present (as it can be noticed within both images of Fig. 3); thus a median filtering was needed as a pre-processing operation.

B. Frame segmentation

The first step in the image analysis chain, namely the segmentation aims at the extraction of the person's silhouette from the given frame.

While traditional image segmentation techniques may be used, here we make use of the specifics of the ToF acquisition. Thus the silhouette is extracted by simply imposing some thresholds on the distance image provided by the ToF camera. Basically the threshold is set to the minimum distance between the camera and the closest room background object (wall, desk etc.); for instance in our experiments T < 5 meters.

As it had been showed that very far dark objects produce distance errors [5], and in order to increase the confidence in the obtained segmented image, we also imposed a threshold on

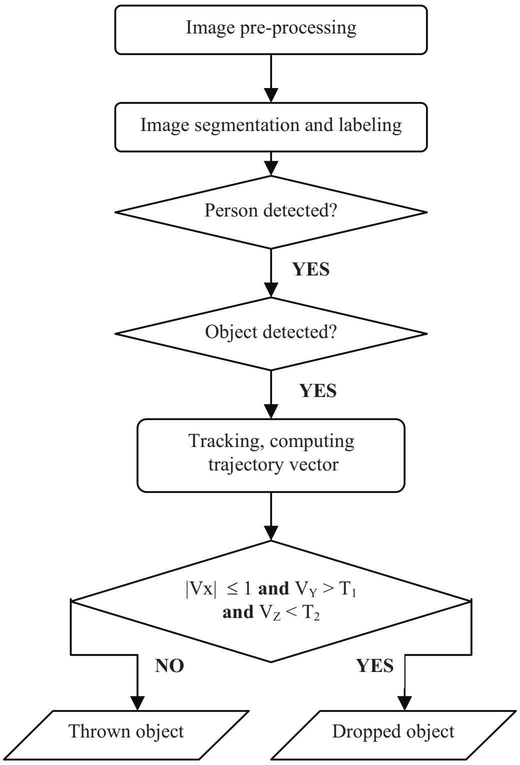

Fig. 2. Object throwing/dropping detection algorithm

the amplitude image. An example of the resulting segmented image is shown in Fig. 4 a). The person is separated from the background and all further analysis is performed only inside this region.

Subsequently we validate the scene with an intrusion detection test; using the segmentation result we see if the scene is empty (without any person), case when the algorithm simply waits for the next frame coming from the ToF camera, and then repeats the segmentation.

The next step in our thrown object detection algorithm consists of labeling the segmented image. The purpose of this step is to detect if the object has been thrown or not. Before throwing an object by a single person within the scene (a room for instance), the result of labeling would be a single label corresponding to the person's silhouette. But, if the person begins throwing something, in the moment when the thrown object separates form the person's hand, one will obtain two labels (spatially separated regions): one for the object and the other for the person. This situation is presented in Fig. 4 b).

Hence, the object throwing event can be triggered by the occurrence of a new small region separating from a much larger region which is the person's silhouette; "large and small" may be easily assigned by counting each region area. In fact the number of pixels within the object validate if it truly represents a thrown object (by some simple perspective rules and knowing the distance to the object from the range image one can set the thresholds for object size). Experimentally, we set the threshold for object size between 30 and 500 pixels.

Fig. 3. a) ToF intensity frame; b) ToF distance frame

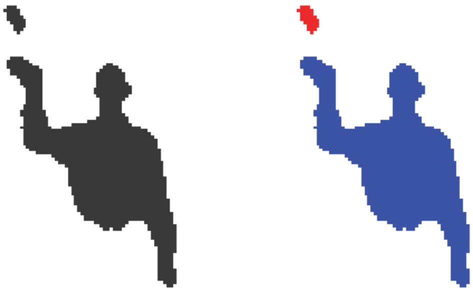

Fig. 4. a) Segmented image retrieved from the example showed in Fig 3.b; b) Labeled image

C. Trajectory analysis

Once we have triggered the throwing event, the algorithm falls in the third phase, namely tracking the object along the video, recording and analyzing its trajectory in (X, Y, Z) coordinates.

The estimation of the object trajectory direction may be easily performed by computing the vector tangent to the beginning of the trajectory.

A particularly interesting situation is obtained if the throwing speed is rather large. In this circumstance the object's image seems to split in two, as one can see in Fig. 5. This effect is caused by the fact that the ToF camera uses four exposures in order to compute one distance frame. Hence, very fast moving objects will be captured at different positions within the four frames. Anyway, this does not represent a problem for our algorithm, it is like two objects were thrown, and their trajectory is the same.

D. Dropping detection

Besides the detection of thrown objects, an interesting application is the detection of dropped objects.

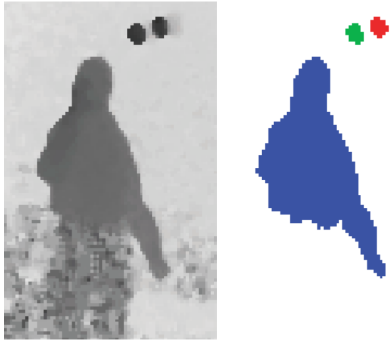

Fig. 5. Throwing with higher speed (distance and labeled image)

The detection principle is the same: a small size compact region of pixels (the object) is detaching from a much bigger compact area (the person's silhouette). The dropped object is also tracked on (X, Y, Z), its trajectory is recorded, and the vector tangent to the beginning of the trajectory is computed.

In Fig. 6 we show an example of a distance frame (to the left) and the corresponding segmented and labeled image (to the right), from an object dropping video sequence.

The proposed algorithm is able to discriminate between a thrown object and a dropped object by simply imposing some thresholds on the above computed tangent vector. In Fig. 7 the trajectory of a thrown object is plotted with a continuous line, and the trajectory of a dropped object is plotted using a dotted line. The variation of the thrown object's coordinates is important along all X, Y and Z axes. The trajectory of the dropped object is perpendicular to the X axis, exhibits an important variation on the Y axis, and insignificant variation along the Z axis. Thus, one can decide if the object is dropped if V_X (the projection of the trajectory vector on the X axis) is almost zero (in our algorithm we imposed $|V_X| \leq 1$), if V_Y is above a threshold (ex: $|V_Y| > 10$) and if V_Z is below a chosen threshold (ex: $|V_Z| < 25$).

III. Experimental Results

We tested our algorithm on 50 video sequences captured with the ToF camera. The database consists of 25 sequences containing object throwing actions (from different angles, throwing directions and speeds) and 25 sequences containing object dropping actions. Each video sequence began with some random motion performed by the subject, then one object throw / drop action was performed.

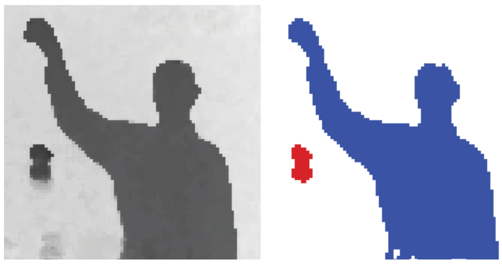

Fig. 6. Dropping an object (distance and labeled image)

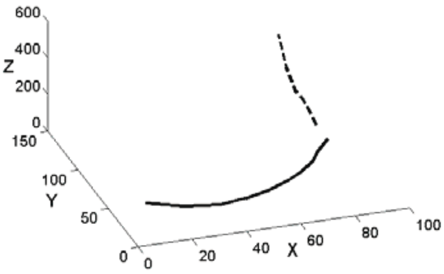

Fig. 7. Thrown objects trajectories

The detection results are depicted in the confusion matrix from Table I and the overall detection rates in Table II. For object throwing, one sequence was not detected because the person was farther away from the camera, and the object was too small. To cope with such situations, one could decrease the threshold for object size, but there is the risk that noise could trigger the object detection. The two dropped objects were misclassified as thrown objects due to the high noise present within the distance frame (the object's computed trajectory was modified by the noise). The acquisition conditions were very poor for the work with the ToF camera: small room (the background was too close), very dark objects in the background (for instance computer screens) which give high distance measurement errors etc. Thus, the detection rates could be substantially increased with better acquisition conditions. Also, some morphological filtering applied in the pre-processing step could improve the obtained results. With good acquisition conditions, the method provides robust results on all distance range of the ToF camera (0.5m ÷ 7.5m).

A. Duration

As for computation speed, we tested our algorithm on an i3 M370 2.4GHz CPU, and 50 frames were processed in about 0.237 seconds. Thus, the algorithm processing speed is around 210 FPS, which is almost 10 times faster than the acquisition speed of 22 FPS.

IV. Conclusions

We presented a real time, low complexity algorithm for the detection of object throwing or dropping based on a ToF camera. The proposed method relies on the 3D information of the scene given by the ToF camera.

TABLE I. Confusion Matrix

	Throw	**Drop**	**Undetected**
Throw	**24**	0	1
Drop	2	**23**	0

TABLE II. Detection rate

Action type	Throw	Drop
Detection rate (%)	96	92

The main image processing tasks are: distance image segmentation and labeling, object detection, object tracking along its trajectory. The decision between the two possible actions: object throwing or object dropping is taken using the computed trajectory vector. The overall decision rate is above 92%, and can be improved with better acquisition conditions in order to reduce the noise within the distance image. We have identified as the main limitations of the method the noise of image and the influence of very dark objects within the background.

As further work, the extension to multiple persons object throwing detection is considered. Also, the algorithm can be easily adapted to other 3D image sensors like the Kinect camera.

References

[1] A. Yilmaz, O. Javed, and M. Shah, Object tracking: A survey, ACM Computing Surveys, vol. 38, issue 4, 2006.

[2] W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Transactions on Systems,Man and Cybernetics, vol. 34, no. 3, pp. 334 - 352, 2004.

[3] E. Ribnick, S. Atev, N. Papanikolopoulos, O. Masoud and R. Voyles, "Detection of thrown objects in indoor and outdoor scenes," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2007, pp.979-984, 2007.

[4] H. Frank, N. Wellerdick-Wojtasik, B. Hagebeuker, G. Novak and S. Mahlknecht, "Throwing Objects -- A bio-inspired Approach for the Transportation of Parts," IEEE International Conference on Robotics and Biomimetics (ROBIO), pp.91-96, 17-20 Dec. 2006.

[5] T. Frank, U. Janoske and C. Schroedter, "Detection of position and orientation of flying cylinder shaped objects by distance sensors," IEEE International Conference on Mechatronics (ICM), 2011, pp.591-596, 2011.

[6] S.B. Gokturk, H. Yalcin, C. Bamji, A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions, Conference on Computer Vision and Pattern Recognition Workshop, 2004 (CVPRW '04), vol. 3, pp.35-43, 2004

[7] A. Kolb, E. Barth, R. Koch and R. Larsen, Time-of-flight sensors in computer graphics, Eurographics, State of the Art Reports, pp. 119-134, 2009.

[8] S. Oprisescu, D. Falie, M. Ciuc and V. Buzuloiu, “Measurements with ToF Cameras and their necessary corrections”, International Symposium on Signals, Circuits and Systems (ISSCS) 2007, vol.1, pp. 221-224, 13-14 July 2007.