Multisubjects Tracking by Time-of-Flight Camera

Piercarlo Dondi[1], Luca Lombardi[1], and Luigi Cinque[2]

[1] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy
[2] Department of Computer Science, Sapienza University of Rome, Via Salaria 113, Roma, Italy
{piercarlo.dondi,luca.lombardi}@unipv.it, cinque@di.uniroma1.it

Abstract. Time-of-Flight cameras are the state of art sensors for a fast detection of depth data in a scene. This kind of sensors can be very useful for tracking, in particular in indoor ambient, since, using light in near-infrared spectrum, they are less affected by abrupt change in illumination. In this paper we propose a new method for the tracking of multiple subjects based on Kalman filter. The first step of our solution is a ToF based foreground segmentation, that retrieves all significant clusters in the scene, followed by a robust tracking system able to correctly handle occlusions and possible merging between clusters.

Keywords: Tracking, Time-of-Flight camera.

1 Introduction

From their introduction in 2003 [1], Time-of-Flight (ToF) cameras are quickly become the state of art sensors for achieving a real-time (from 20 fps of the older models to 54 fps of the newer ones) depth measurement of a scene. ToF cameras do not need reference points or external illumination sources, they work emitting light in near-infrared spectrum and distances are estimated according to the time spent by the reflected light to come back from objects to the sensor. This kind of active sensors have trigged the interest of many researchers, in various fields, such as 3D object reconstruction, human-computer interaction, tracking, augmented reality or also medicine and bio-informatics.

In this paper we focus on the problem of real-time tracking of multiple subjects. Our approach is based on the use of a Kalman filter. In particular we extend the standard Kalman filter of the OpenCV library, with some automatic methods specifically designed to associate, as well as possible, the detected clusters with the respective trackers. The tracker works on the clusters retrieved by a fast foreground segmentation that exploit the particular kind of data provided by a ToF camera (depth data and intensity of reflected light).

A preliminary implementation of this solution was presented in [2], then the results achievable by a more complete version was presented in [3]. In this paper

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 692–701, 2013.

we introduce new improvements such as a more precise initial association between clusters and trackers (that guarantees a correct tracking of both near and far clusters) and the handle of clusters merging due to occasional imprecision in the segmentation (e.g. subjects too close). From the computational point of view, the current implementation is more efficient than the original one and the new features does not influences the overall performance of the system.

The paper is organized as follow: section 2 supplies a brief overview of the sensor; section 3 provides the state of the art of tracking method based on ToF cameras; section 4 describes the segmentation method; section 5 analyzes in details the tracking procedure; section 6 shows the experimental results; section 7, at last, draws some conclusions.

2 Time-of-Flight Camera Overview

Time-of-Flight cameras are active imaging sensors that can provide distance measures of an ambient using laser light in near-infrared spectrum. There are two main technologies: pulsed light and modulated light. In the first case a coherent wavefront (similar to a "light wall") hits the targets and then the distances are measured analyzing the deformation in the reflected "wall". In the second case, currently the most widespread technology, the camera emits a modulated light and the depth information are gained by phase delay detection.

Respect to other depth measuring systems, such as stereo cameras or laser scanners, ToF cameras supply some advantages: they can work in real-time, the depth data are directly provided by the sensors without complex additional computations; they do not need external light (the illumination is self-provided); they can operate in any kind of scenario without external reference points or colors contrast; the shape of the objects does not influence the measures.

On the other hand there are also some disadvantages that must be considered for a better use of these devices. ToF cameras have still a limited resolution (e.g. 176x144 of Mesa SR4000 or 200x200 of PMD CamCube 3.0) and they are affected by different kinds of noise: the "flying pixels" due to areas with abrupt changes in depth (e.g. the corners of an object); the "motion artifacts", measurement errors proportional to the speed of moving subjects; the noise cause by sunlight that can significantly alter the result and limit the applicability of these sensors to indoor use. Finally the precision of the measures strictly depends on the reflectivity of the objects, if it is too high it can saturate the sensor, while if it is too low the object can be not correctly detected [4].

All the experiments in this work are been made with a modulated light cam, the Swissranger SR3000. This ToF camera is able to supply simultaneously two images per frame: a distance map and the map of the intensity of the reflected light. Both of them have QCIF resolution (176x144 pixels) and a color depth of 16 bits. The 55 active leds emit in the near infrared around 850nm with a frequency of 20Mhz, a value that guarantees a nominal range without ambiguities of 7.5m. The depth accuracy goes from a few centimeters to millimeter in optimal conditions. The distance accuracy depends on distance range, signal intensity

and the background illumination. The field of view (FOV) is about 47.5×39.6 degrees [5]. The camera has been used at 18-20 fps in order to maintain a good ratio between noise and real-time capability: the noise depends directly on the frame rate.

3 State of Art

A lot of approaches have been used for obtaining a good ToF-based tracking. In [6] the retrieved clusters are projected on the ground plane for creating a so called "flat-map", then an Expectation Maximization algorithm has been applied to that map. A method for multiple people tracking based on Shape from Silhouettes (SfS) is proposed in [7]: it appears robust but greatly limited by the high numbers of the cams needed (six in the proposed solution) and by the small dimension of the room. In [8] instead a traditional Kalman filter has been used with the camera placed to provide a top-down view of the scene. This particular position simplifies the tracking problem, but significantly decreases the visible area, moreover almost all the details of the detected subjects are lost.

All these methods segment the scene and retrieve the clusters by a background subtraction algorithm. This solution, even if wide diffused, suffers of known problems such as ghosts appearing at changes in background objects or absorption of still people . The generation of the model can also be computationally expensive, especially if it needs a high resolution or a dynamic adaptation to ambient changes (such as light variations).

An alternative approach, based only on the analysis of depth data, can be found in [9]. The described algorithm allows the detection and the classification of objects in the scene studying the probability density function (pdf) of depth image and its histogram distribution. Then an appropriate distance metric, based on the integrated square error between the pdfs, is used to recognize the clusters through consecutive frames.

A particular category of solutions involves the combination of traditional RGB cameras with a ToF one. In [10] is proposed an approach that exploits two particle filter-based visual trackers, one for each stream data type (RGB and depth). Depending from the scene the system uses the one that guarantees the better performances (generally RGB for outdoor and depth for indoor). In [11], instead, the fusion of color and depth data is adopted for compensating the respective weaknesses of the two type of sensors. More specifically the segmentation and the tracking are achieved using a well designed method based on mean shift algorithm.

4 Segmentation

The proposed ToF-based segmentation method provides different advantages: no need of preprocessing operation (no learning phase); no need of a priori knowledge of the environment (the system is therefore robust to background changes); the shape of objects has no influence on the results.

The main steps are summarized in Fig. 1: firstly a thresholding based on values in the intensity map is applied to the distance map; then a region growing algorithm, that starts from seeds planted in the peaks of the intensity, is executed on the filtered distance map to produce separate clusters corresponding to foreground objects.

Fig. 1. The main steps of foreground segmentation from the ToF data input to the full body extraction

In the intensity map, foreground objects are brighter than those in background (they received more light), so, the reflectance data can be successfully used as a mask on the depth map to reduce the area of investigation on which the region growing will be applied (Eq. 2). The region growing is applied on the distance map and not directly on the intensity one, due to the greater stability of depth information (intensity is enough precise for a preliminary thresholding, but it varies too much due to different reflection properties of the objects framed). The best seeds for region growing are found applying an opportune intensity threshold (λ_{seed}), estimated using the Otsu's method, a well known thresholding algorithm based on image histogram.

The similarity measure S between a cluster pixel x and a neighboring pixel y is defined as follow:

$$S(x, y) = |\mu_x - D_y| \tag{1}$$

where D_y is the distance value of pixel y and μ_x is a local parameter related to the mean distance value around x (Eq. 3). The lower S is, the more similar pixels are. When a seed is planted, μ_x is initialized to D_x. Considering a 4-connected neighborhood, a pixel x belonging to a cluster C absorbs a neighbor y according to the following conditions:

$$\{x \in C,\ S(x, y) < \theta,\ I_y \geq \lambda,\ D_y < \delta\} \rightarrow \{y \in C\} \tag{2}$$

λ is the intensity threshold, proportional to λ_{seed}, dynamically calculated for each frame (Otsu's threshold λ_{seed} proves to be very effective to find the peak values of the intensity image, but it has turned out to be too strict for the thresholding required in this phase); θ is a constant parameter, experimentally estimated, related to clusters separation; δ is an optional parameter used for excluding a priori all points beyond a fixed distance. It quickly reduces the search area and can be very useful in those applications in which the maximum operating distance is known a priori or if the shot is made too close to a wall.

When a neighbor y of seed x is absorbed, the average distance value μ_y is computed in an incremental manner as follows:

$$\mu_y = \frac{\mu_x * \alpha + D_y}{\alpha + 1} \tag{3}$$

where α is a learning factor of the local mean of D. If pixel y has exactly α neighbors in the cluster, and if the mean of D in this neighbor is exactly μ_x, then μ_y becomes the mean of D when y is added to the cluster.

Every region grows excluding the just analyzed pixels from successive steps. The process is iterated for all seeds in descending intensity order. Regions too small are discarded for removing the possible false positive areas over the threshold (for example little surfaces with high reflectivity).

The performance of the method, varying its parameters, has been analyzed in [12]: all the tests show a correctness (the ratio between true positive and the sum of true positive and false positive) between 94% and 97% and completeness (the ratio between true positive and the sum of true positive and false negative) between 92% and 96%.

5 Tracking

The multi-subjects tracking method presented in this work is based on the use of Kalman filter. As mentioned in section 3, this solution for ToF based tracking was first studied, with good results, by [8]. Respect to that paper the proposed implementation adopts a more general approach, the camera has not been placed to provide a top-down view, but a frontal view. The purpose of this choice is to obtain a more versatile solution, with a major visible area combined with a better angle of view. Our tracking method can also correctly handle the occlusions and the merging between clusters.

5.1 Kalman Tracker

A Kalman filter is characterized by a six dimensions state, (x, y, z, v_x, v_y, v_z) which refers to the position of the assigned cluster: (x, y, z) represent the centroid, while (v_x, v_y, v_z) is the velocity vector. All these parameters are expressed in image coordinates, since the SR3000 supplies data already organized in 3D Cartesian coordinates. When the tracking starts, at each detected cluster is assigned a new Kalman tracker, that it is initialized with the current position of the centroid of the real cluster. At time *t*, each Kalman predicts the most probable position of its correspondent cluster at the next frame. So, at time *t+1*, the predicted coordinates of each Kalman will be compared with the real positions of the centroids of each clusters retrieved, in order to find again the previous objects. The association between measured clusters and Kalman trackers is evaluated by minimum square euclidean distance between their centroids. In particular each Kalman tracker connects itself with the nearest object that is not yet been assigned to another closer Kalman (Eq. 5). The Kalman and

the cluster associated in this way are excluded by the respective set of possible candidates; the system iterates until all the visible clusters are connected with a Kalman. Since sometimes the ToF noises (especially motion artifacts) can generate false small clusters that last less than a second, a Kalman tracker is assigned to a new cluster only if it is in scene at least from 5-7 frames.

According to Kalman filter behavior, after the association, the prediction is corrected with the real measurement in order to refine the state estimation.

For reducing to minimum the errors each Kalman tracker searches correspondences only in a limited spherical area. The set P_c of all the cluster that are enough near at least to one Kalman tracker is defined as follow:

$$\left\{\sqrt{d_e(c,k)} < \alpha\right\} \rightarrow \{c \in P_c\} \tag{4}$$

where $d_e(c,k)$ is the euclidean distance between the coordinates of retrieved cluster c and of Kalman tracker k; α is an association threshold experimentally defined considering the limited field of view and the limited resolution of the ToF camera.

The assignment procedure can be summarized by the following equation:

$$\left(\sqrt{d_e(c,k)} = \min\left[\sqrt{d_e(x,y)}\right], \forall x \in P_c, \forall y \in K\right) \rightarrow (c\ ass.\ to\ k) \tag{5}$$

where K is the set of all active Kalman trackers, and $(c\ ass.\ to\ k)$ indicates that cluster c has been recognized as the cluster assigned to Kalman k at previous frame.

After all these steps, if one or more retrieved clusters are not associated to any Kalman, they are considered as new objects entered in scene, so an equivalent number of Kalman trackers are initialized. Otherwise, if one or more active Kalman trackers are not associated, probably there is an occlusion. In this case the Kalman maintains all the precedent data and tries to estimate the most probable path of the disappeared cluster using its last detected movements; the research area is also doubled in order to compensate estimation errors. If the cluster reappears shortly (within 30-40 frames) in a position closed to the predicted one it is immediately recognized and reassigned to its precedent tracker. On the contrary, if the cluster does not reappear in a fixed time, it is considered out of scene, so its Kalman is reset and can be reassigned to a new subject.

The described behaviour of a Kalman tracker can be defined by three possible states: 0, not assigned; 1, assigned to a visible cluster; 2, assigned to a not visible cluster probably occluded. The possible changes between these states are showed in Fig. 2 with a 2D example: the crosses are the retrieved clusters, the arrows are their directions, the points are the Kalman trackers and the circles are their search areas. Figure 2(a) presents a typical situation with two moving clusters assigned to two Kalman trackers. When one cluster is occluded by the other (Fig. 2(b)), its correspondent Kalman goes in state 2 (note the research area increased). From this situation there are two possible exits: the cluster reappears near to the predicted position, so it is associated again to its precedent Kalman that returns in state 1 (Fig. 2(c)); the cluster does not reappear, so the Kalman

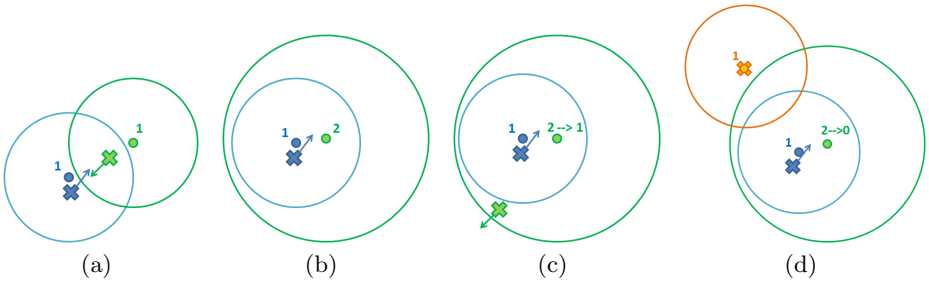

Fig. 2. 2D representation of Kalman trackers behaviour: (a) standard situation, two clusters in movement (the crosses) assigned to two Kalmans (the points), no occlusions; (b) one cluster occludes the other; (c) the occluded cluster reappears and it is assigned again to its precedent Kalman tracker; (d) the occluded cluster does not come back, a new cluster enters in scene and it is assigned to a new Kalman tracker

comes back to state 0 (Fig. 2(d)). At the same time, if a cluster out of all the research areas comes in the scene, it will be assigned to a new Kalman tracker (Fig. 2(d)). Note as in Fig. (Fig. 2(c)) the green tracker is associated to the green cross even if the blue one is nearest; this happens according with Eq. 5 because the blue tracker is closer to blue cluster, so the green tracker associates itself with the second closer object inside its search area.

5.2 Feedback for Smart Seeding

The Kalman predictions can be actively used for increasing the precision of the seeding and so correcting cluster detection mistakes. Considering a case in which there are two subjects. If one of them gets too close to the sensor (Fig. 3, left) its intensity values grow too much so all the seeds will be concentrated on it. As a consequence the faraway cluster is excluded from region growing and disappeared in the final results (Fig. 3, top right), even if it has been correctly included, after thresholding, in the filtered range image (Fig. 3, middle).

This issue can be overcome using the predictions of Kalman trackers as additional input for the seeding phase [3]. At time t new seeds are planted in pixels around to all centroids coordinates predicted at time *t-1*. This smart seeding allows the concurrent detection and tracking of middleground and foreground objects (Fig. 3, bottom right).

However such solution does not work well if a new subject enters in scene when the first one is close to the camera. In this case the new subject is initially undetected because he has not an assigned Kalman tracker. In this situation we plant seeds also in distance peaks of the filtered range image. This approach is not so precise as Kalman seeding, because there is not a direct correlation between distance and presence of a subject, but it is fast and simple and it can correctly handle such kind of issue. Moreover adding the distance seeding is not computationally expensive, so it can be successfully combined with the Kalman seeding for retrieving new entering clusters.

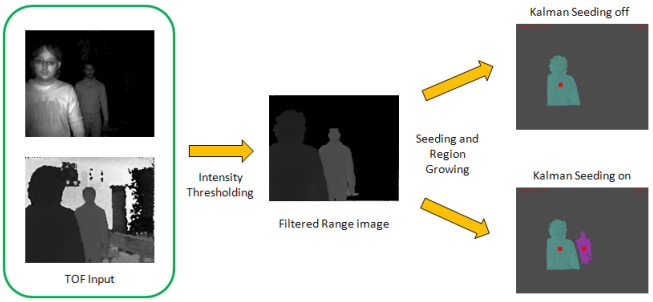

Fig. 3. Segmentation without and with Kalman seeding

5.3 Control for Clusters Merging

A final improvement involves the incorrect clusters merging. The region growing is able to correctly divide near clusters if their Z coordinates are enough different (Fig. 4 left). However, if two clusters have distance values too much similar, they will be wrongly fused in a single cluster (Fig. 4 center). This is not a common situation and, even when occurs, it is usually of short duration (also a little movement can vary enough the Z) – so, most of the times, it can be correctly handled as a traditional occlusion. For those cases in which the fusion takes too much time, and can confuse the tracker, a simple control can be adopted.

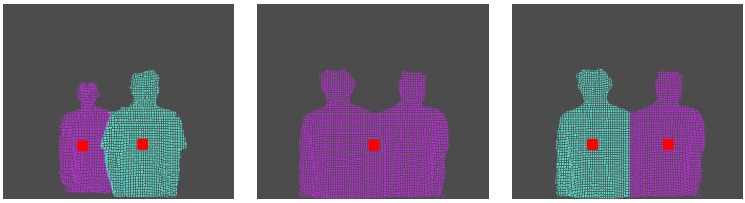

Fig. 4. Cluster merging correction: (left) two correctly separated clusters; (center) wrong merging of two clusters in one; (right) clusters subdivision after corretion

At time *t*, after the segmentation, the system saves the current positions, the size and the number of the current clusters. Different axes of vertical symmetry are traced between each near clusters. At time *t+1*, if there is a reduction in the number of clusters and there is a cluster with a size consistent with the sum of two or more previous clusters, a merging may have occurred (Fig. 5.3). In that case if there is one axis that passes through this cluster, that axis is chosen as lines of separation for splitting the cluster in two parts (Fig. 4 right). When we track people, another practical index of a clusters merging is the presence of a cluster with two heads (that can be recognized with a standard face detection system). The obtained subdivision is only an approximation not much precise,

in particular for the shape of the split clusters; however the positions of the centroids are still quite accurate, so they can be used for maintaining a correct tracking. This correction should be used mainly when the moving area is very reduced (i.e. the clusters are very close) and when it is crucial not to confuse the subjects tracked even for a moment.

6 Experimental Results

A series of evaluation tests have been made to prove that the tracking system is able to manage the concurrent movements of multiple subjects and is also robust to the occlusions. The only possible sources of errors are moving objects with abrupt changes of direction or new clusters that appear closed to a just active Kalman. Figure 5 shows some frames of an example sequence. The colored spheres on the top of the three subjects are the markers of Kalman trackers, note how the correspondence between clusters and trackers is always maintained.

Fig. 5. Tracking sequence of three moving subjects with multiple occlusions

The test starts with three people placed at different distances. Then the first one crosses the stage until exiting from the framed area. Clusters temporarily occluded (Fig. 5, second image) correctly maintain their "labels" when they are again visible (Fig. 5, third image). A similar event happens when the second subject occludes the third one (Fig. 5, fourth and fifth images); in this case, also, the first subject re-enters in scene in a position closed to the exit one and after a short time, so it is correctly recognized.

Even if there is not a theoretical limit to the number of clusters that can be followed at the same time, the limited resolution of the SR3000 reduces the useful number for a correct working to a maximum of 3-4 objects. When the number is bigger, the risk to fill all the field of view of the camera with clusters all closed to each other, with a consequent increase of possible sources of errors, is very high.

7 Conclusions

This paper presents a new robust approach to the multi-subjects tracking, based on Kalman filter, that does not need any a-priori information about ambient or clusters. Due to the use of a ToF camera our system can work in any indoor

scenario, in particular without controlled illumination sources. The algorithm allows the concurrent tracking of moving subject, correctly handling occlusions or accidentally clusters merging. A flexible seeding system that uses Kalman, depth and intensity data guarantees a fast detection of people placed at difference distance from the camera. Now we are studying how to improve the description of the clusters in order to recognize reentering objects after a longer time than the actual few seconds.

References

1. Oggier, T., Lehmann, M., Kaufmann, R., Schweizer, M., Richter, M., Metzler, P., Lang, G., Lustenberger, F., Blanc, N.: An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger). In: Proceeding of the SPIE, vol. 5249, pp. 634–645 (2003)
2. Bianchi, L., Gatti, R., Lombardi, L., Lombardi, P.: Tracking without Background Model for Time-of-Flight Cameras. In: Wada, T., Huang, F., Lin, S. (eds.) PSIVT 2009. LNCS, vol. 5414, pp. 726–739. Springer, Heidelberg (2009)
3. Dondi, P., Lombardi, L.: Fast Real-Time Segmentation and Tracking of Multiple Subjects by Time-of-Flight Camera. In: 6th International Conference on Computer Vision Theory and Applications (VISAPP 2011), pp. 582–587 (2011)
4. Kolb, A., Barth, E., Koch, R., Larsen, R.: Time-of-Flight Cameras in Computer Graphics. Journal of Computer Graphics Forum 29, 141–159 (2010)
5. CSEM: SwissRanger SR-3000 Manual, Mesa Imaging (2006)
6. Hansen, D.W., Hansen, M.S., Kirschmeyer, M., Larsen, R., Silvestre, D.: Cluster tracking with Time-of-Flight cameras. In: Proceedings of Computer Vision and Pattern Recognition Workshops (CVPRW 2008), pp. 1–6. IEEE Computer Society (2008)
7. Guomundsson, S.A., Larsen, R., Aanaes, H., Pardas, M., Casas, J.R.: TOF imaging in Smart room environments towards improved people tracking. In: Proceedings of Computer Vision and Pattern Recognition Workshops (CVPRW 2008), IEEE Computer Society (2008)
8. Bevilacqua, A., Di Stefano, L., Azzari, P.: People Tracking Using a Time-of-Flight Depth Sensor. In: Proceedings of the AVSS 2006, Video and Signal Based Surveillance, p. 89. IEEE Computer Society (2006)
9. Parvizi, E., Jonathan Wu, Q.M.: Multiple Object Tracking Based on Adaptive Depth Segmentation. In: Proceedings of Canadian Conference of Computer and Robot Vision, pp. 273–277. IEEE Computer Society (2008)
10. Sabeti, L., Parvizi, E., Jonathan Wu, Q.M.: Visual Tracking Using Color Cameras and Time-of-Flight Range Imaging Sensors. Journal of Multimedia 3(2), 28–36 (2008)
11. Bleiweiss, A., Werman, M.: Fusing Time-of-Flight Depth and Color for Real-Time Segmentation and Tracking. In: Kolb, A., Koch, R. (eds.) Dyn3D 2009. LNCS, vol. 5742, pp. 58–69. Springer, Heidelberg (2009)
12. Bianchi, L., Dondi, P., Gatti, R., Lombardi, L., Lombardi, P.: Evaluation of a foreground segmentation algorithm for 3D camera sensors. In: Foggia, P., Sansone, C., Vento, M. (eds.) ICIAP 2009. LNCS, vol. 5716, pp. 797–806. Springer, Heidelberg (2009)