Recognition of people's positioning by cooperative mobile robots for human groups steering

Edgar Martinez†, Akihisa Ohya†‡, Shin'ichi Yuta†
†Intelligent Robot Laboratory, University of Tsukuba, Japan
‡PRESTO, JST
www.roboken.esys.tsukuba.ac.jp

Abstract

In this research we attempt to build a multi cooperative mobile robot system able to perform the process of steering people in indoors. When a target-group is integrated by several persons, steering task is more difficult to accomplish, specially when only a robot is intended to be used. Since that, the problem can be overcomed if we consider that a cooperative mobile robot system's performance is higher for steering multiple humans than only one, and will greatly improve the performance of the tasks. In this paper, we detail a novel method for multiple human localization which has been developed by using multi homogeneous mobile robots, equipped with trinocular stereo vision. As a first step of this research the method includes spatial noise data filtering, multi sensor data fusion and clustering based segmentation. In addition, some experimental results are shown to verify the feasibility of the method.

1 Introduction

Oftentimes in different public and private institutions, visitors are guided for recognition of such places, meanwhile information and assistance about those physical facilities are provided. In order to take the role of human guides, a multi cooperative mobile robot system able to steer those groups is being developed in our lab. Such tasks imply the implementation of multi robots planning and robots formation[1], human tracking-motion[2], conduction and gathering of people as well.

Interaction between humans and robots may occur in a variety of ways, and are deployed in everyday human environments. Interactive navigational tasks, such as leading, following, intercepting, and avoiding people, require the ability to track human motion. This paper mainly focus on estimating position of several guided people by fusion of stereo range data from a multi cooperative mobile robot system.

2 Problem specification

Guided tours, chiefly means a walking, displacement or moving from one point to a target location by employing steering tasks, which essentially involve a mechanism to flock people. Our motivation is to achieve some steering tasks which involve crowding while navigation, by using multi-mobile robots capable to gather together a determined number of people as well as maneuvering them by a specific navigation course. See figure 1.

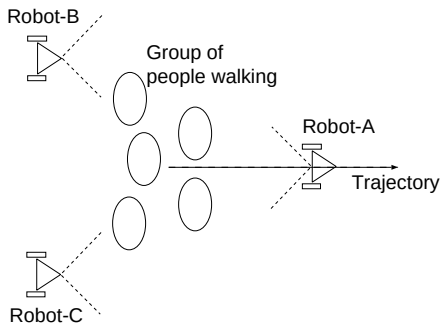

Figure 1: Human group steering by multiple mobile robots.

Our system is a set of homogeneous mobile robots, which directs the course by using global map based navigation, and by using a cooperative framework they are able to localize each member of the group.

A mobile robot likely will not perceive fully each member of the whole group in a given moment, since crowded people occlude themselves some members, losing temporaly their presence. Thus, given this fact multiple cooperative mobile robots can solve the problem of people localization including as well motion tracking while conducting their course.

For steering groups of people some elements are necessary such as multi people localization, multiple robot localization, planning and self-formation strategies as well as flocking tasks. From which this paper mainly focus on the method for localization of multiple people as early laboratory result's explanation. Likewise, for accomplishing this first goal, stereo vision is being used for environment perception.

3 Stereo vision based range data

Stereo vision based range data has several favorable reasons in this development. Basically it facilitates:(1)3D spatial coordinates, (2)object segmentation. Although some of those features can be computed by using other methods or sensors, stereo vision facilitates the problem of segmentation, since stereo images can provide a ranged image, it makes easier to compute people segmentation[3].

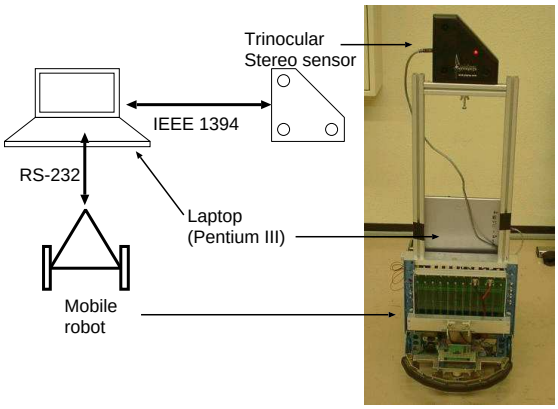

Figure 2: Equipment configuration, robot's note-PC and the trinocular stereo vision sensor.

Each mobile robot is equipped with a Digiclops[4] trinocular stereo vision sensor(figure 2). It provides disparity maps, color and gray scale images in 3 different resolutions such as 640x480, 320x240 and 160x120 pixels. Sensor data are acquired by an IEEE 1394 bus communication[4]. Likewise, the Note-PC has a 900Mhz Pentium-III running under Linux Momonga 2.4. Results are computed in real time with the lowest resolution. Figure 3 depicts a diagram of a mobile robot for sensing a person(left), its disparity map and the image upon which calcultations were done(figure 4).

In order to have a solid reference for points discrimination, the first step was to fit adequately the sensor's angles. The sensor's center was fitted in such way that the origen in XY-space(image's center) is approximately at the middle of people body at 100cm of height, see details of figure 3, the rectangle representing a human is intersected by a dotted line, taken as a reference for filtering process which will be explained in later sections.

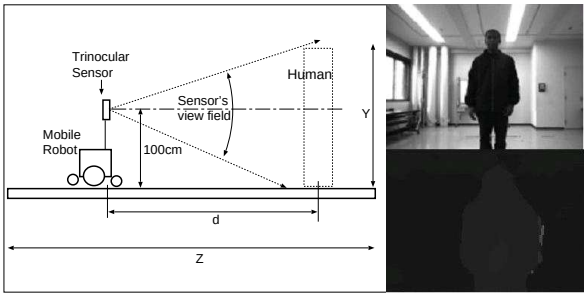

Figure 3: Left:Side-view of robot configuration for sensing; Up-right:Image, Down-right:Disparity map.

In addition, in order to calculate depth images with the best features when disparity map computation is performed, some parameters were carefully determined for our purpose, considering indoor environments. Thus, to determine the parameters, object's were measured at no more than 5m far away from the stereo sensor. On the base of that, stereo parameters were determined by using a software, which enabled us to modify and choose manually the best results for the environmental conditions. Some of those stereo parameters are: stereo mask established in 15 pixels, disparity range was fitted within a range of 2-65 gray level values, edge mask's size 11, and also subpixel interpolation function was enabled to improve the range of disparities values(16 bits), having more density points. Moreover, the algorithm for establishing correlation is the Sum of Absolute Differences(SAD)[4], and results are obtained in real time.

As a first approach for ranging 3D data points, earlier experiments were developed inside an empty area of approximately 8x7m. In order to avoid high rate of noise as much as possible, due to occlusion, constrast, brightness, and similarly, to avoid a high density of data ranged from other objects. The results of such experiment are depicted in figure 4, which those given results show the top and side view of the 3D points computed from the disparity map(figure 3). Sensor location was established at (0,0).

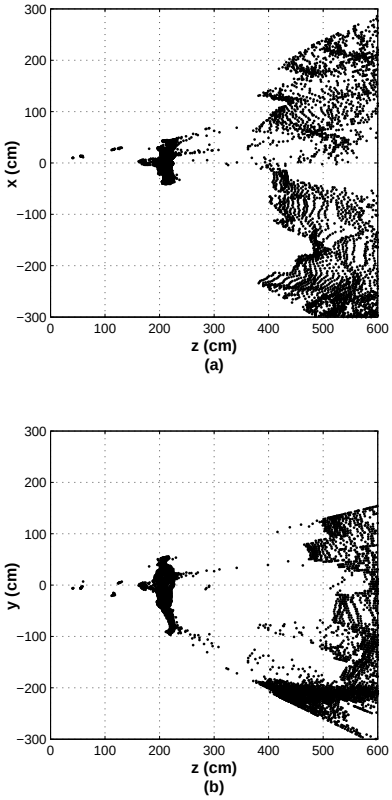

Figure 4: A human located at 2m away from sensor; (a)Top view, (b)Side view.

4 Multi people localization

In this section, we describe the condition of an experiment needed for multi human localization established upon multi sensor data fusion.

In figure 5, the conditions of an experiment are depicted, the obtained results were performed in a stationary environment just for the purpose of the performance of multi localization task. Basically, 2 rows of people were lined up, 3 and 2 people at front and at back sides respectively, between robot A and robots B, C(left, right also respectively). The reason why several mobile robots are used for multi human localization, is because the following:(1)To overcome the problem of total and partial occlusion due to multiple people. (2)Several robots can guide and control together the group.

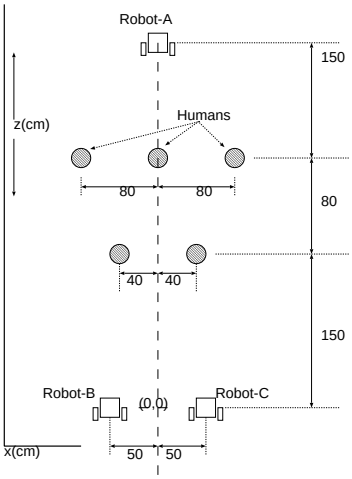

Figure 5: Multi people(5) localization by 3 mobile robots

In order to detect multiple people location by using a single robot, its position and orientation must be arranged strategically, because in other way such robot can not match correctly the whole members of the group from a given angle. But on the other hand, by using multi mobile robots, they can be instrumented with simple capabilities than the required for one robot. In addition, even if each robot can only perceive a partial number of people from different locations, the problem becomes easier to overcome since localized people data from each robot is fused into a cooperative framework. In this experiment the figure 6 is the result of a stereo image compounded by 3 overlapped images (rigth, left and top) from robot-A.

Figure 6: Stereo images from robot A

In figure 6, back side people is difficult to perceive as a whole from robot-A. For such reason robot-A can only sense small amounts of data specially from people of the back side row because of partial occlusion produced from the members themselves.

On the other hand robots B and C, were able to perceive only 4 members of the group each one. It means that one member of the first line in formation was hided because occlusion from members of the back side row.

5 Data Processing

By means of a cooperative framework, we have developed a method, which is compounded by 6 main stages: (1)A 3D-points calculation task, (2)a pre-filtering process task, (3)a noise reduction spatial filter, (4)a 2D points tranformation routine, (5)a data clustering based segmentation task and (6)a multi human localization calculation. Although figure 7 just shows the data flow until (5), also in this paper we will discuss about (6) and its results.

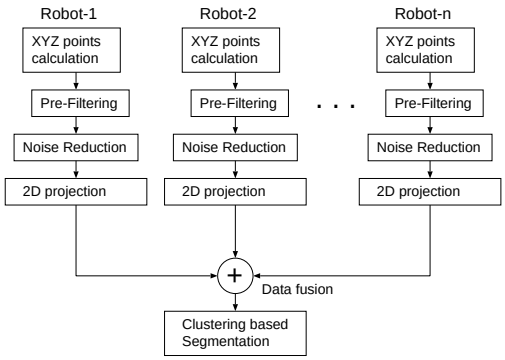

Figure 7: Diagram of data processing

Considered as a part of data processing, data filtering is one of the more important elements of the core of this method. Likewise, filtering basically is divided into 3 phases:pre-filtering, noise reduction and 2D transformation, which this last one is also a routine for points reduction.

In filtering process we mainly aim to reduce the noise produced by light conditions and partial occlusion. Next subsections will describe and show some experimental results. Figure 8 shows the original sensor data from the mobile robot-A on which filtering process will be performed. Let us compare figure 8 with figure 5, where this last is the experimental result based on stereo vision from robot-A location at (0,380). All graphs have been arranged into a global coordinate system.

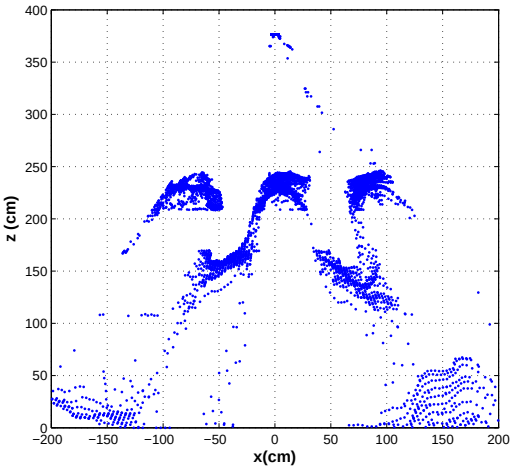

Figure 8: Top view, original sensor data from robot-A(0,380)

5.1 Pre-filtering

The first consideration in our method was to reduce unneeded 3D points. We have called pre-filtering to this

first stage, because basically is a routine of 3D points discrimination upon 2 thresholds. It has been assumed that some points over the Y-coordinate can be discriminated to reduce a large amount of data, which undoubtely do not belong to the group of 3D points of the ranged people. For instance, floor or ceiling and even mismatched points(over ceiling and under floor) are zones from which sensor generates a considerable large amount of data, and are not useful for people localization(see figure 4-(b)). In our method for human localization, people's body-appearance information is unneeded, at least for this purpose and only people cartesian coordinates in the XZ-space calculation are needed.

Thus, we considered that points falling between body's shoulders and knees, have more probability to belong to the set of people's ranged data. The XY coordinate center(0,0) is fitted at the center of the image.

Moreover, after applying a spatial filter, data size was reduced approximately in 50% or more, for all experiments done. Thus, $\vec{D}_j^y = \{y_i \in V^{xyz}/th_1 \leq \vec{y}_i^y \leq th_2\}$. Where V^{xyz} is the original data from sensor, y_i are valid Y-points. $th_{1,2}$ represent both thresholds(knees and shoulders), and $\vec{D}$ is the vector of points that likely are part of people's bodies.

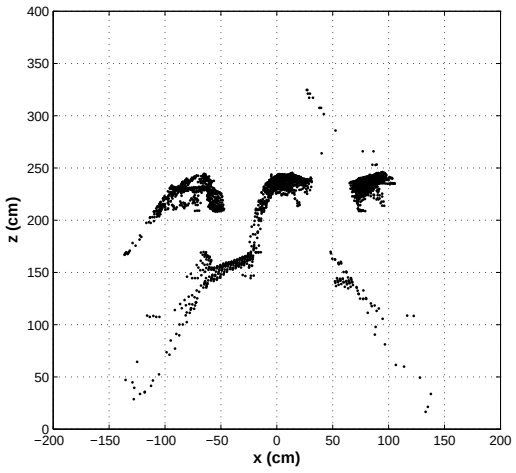

Figure 9: Results after data pre-filtering from robot-A$(0, 380)$.

Let's compare the results of figure 9 with figure 8 and it can be noticed the difference of quantity and data distribution. In this graph and in the following figures data has been arranged into a global coordinate system.

5.2 Noise reduction

For human segmentation, noise reduction is an essentiall task in order to avoid undesired small segmented objects without meaning. Since ranged data from sensor are not a perfect noiseless data model, thus we must deal with this problem by implementing a spatial noise reduction method. As a difference from human's points, noise is considered as small 3D spaces containing a low density and poor uniformity distribution of ranged 3D-points, where on the other side people's ranged bodies hold a high rate of density. Based on this simple consideration, we faced up this problem by implementing a 2D spatial filter, based on a XZ square window(10cm) and performed over the pre-filtered data. The filter-window discriminates sets of points upon a threshold of their number in the cell. This last takes effect in an XZ-space and its total height is the whole Y-space.

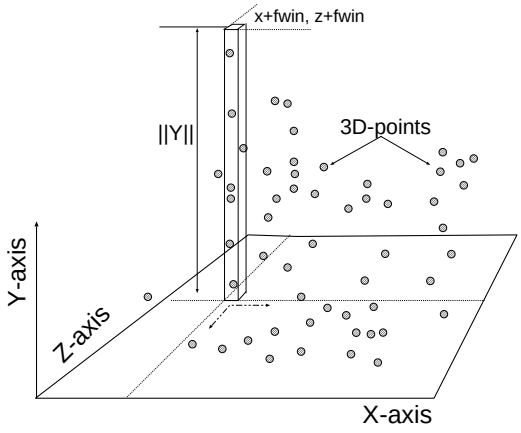

Figure 10: Square window spatial filter taking action in 3D points space.

The area of filtering is given by the size of the square window $fwin$ of 10cm and threshold of 20 points for our experiment.

With this algorithm in most experiments, data were reduced between 10% and 25%, it depends on the cell's size. Results from robot-A has been plotted in figure 11.

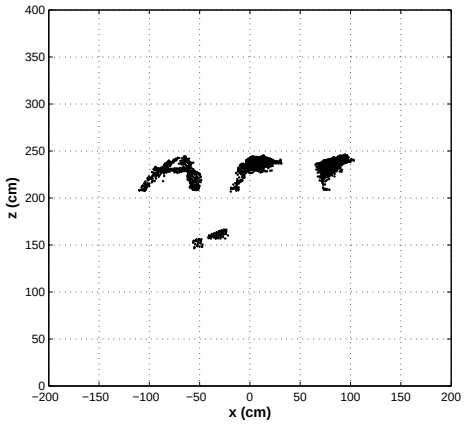

Figure 11: Noise reduction results(robot-A)

5.3 2D point transformation

At this stage, 2D points are just needed for human localization, because of since such task implies the usage of only an XZ coordinate for each people. Thus, from each 3D point in (x,y,z), only (x,z) information is necesary to be obtained. In the same time with this task we eliminate redundancy of points, which means that a amount of sensor data was ranged on the same point.

By tranforming XYZ points into a space of cells in the XZ space, those cells are basically a grid where each unique cell contains a different number of XZ points, in such way that is easier and faster to compute small number of cells than a large number of points, see figure 12. Besides, time computation was improved considerably.

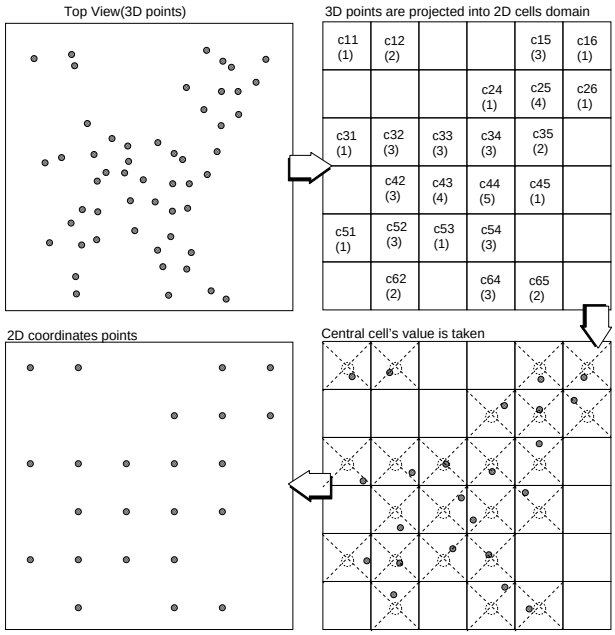

Figure 12: Point reduction process, cell size is 5cm

Classifying each XZ-point into their equivalent cell. Each cell will have a determined number of points, and those points are labeled or referenced with their respective number of cell. In such manner that from now calculus are done just by the cell's address. The sense of this technique is for reducing number of points by selecting only the central point among the whole group in the cell. From here, only a 2D point is considered and it will appeare at every cell size distance in the XZ space. Then, our results are depicted by the figure 13.

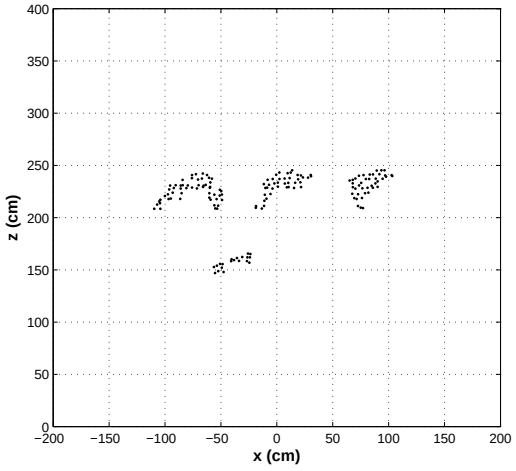

Figure 13: Top view, point reduction(robot-A)

Practically, 98% of the data was reduced without losing the 2D occupancy data grid of each sensed object.

6 Data fusion and clustering

Once data was filtered, robots A, B and C must sharing their data all together as a fusion into the same XZ global space. Figure 14 depicts the data fusion after previous filtering process.

The reason why data fusion is just performed at this stage, is because after finishied the three main phases of filtering, the 2D-point's number is reduced while keeping the projection of human shape over a 2D space, which only this information is useful for human segmentation.

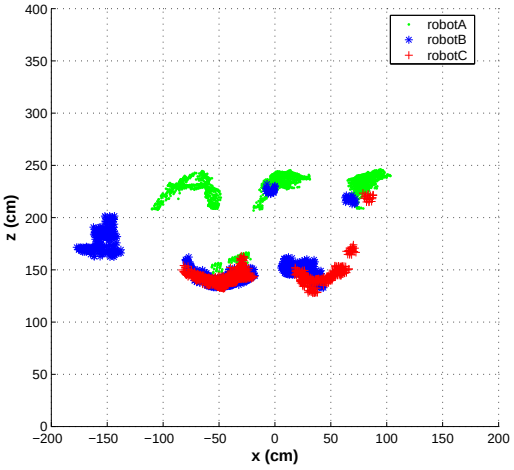

Figure 14: Data fusion from robots A, B and C.

However, as we can notice in figure 13, sometimes depending on people location, human ranged data can be lost partially due to acclusion, producing as result a very poor ranged points density of some people, whereby the filtering could erased them completely.

Therefore, by fusing the data from all robot's positions, each human shape projection can be restored as in figure 14. In fact, very cleaned data avoid that two different objects can be mixed confusing them as a bigger single object. Subsequently by making a correspondence of each 2D-object projection, clustering based segmentation is remaining to be performed.

The purpose of any clustering technique is to evolve a $K \times n$ partition matrix of data set $V (V = \{x_1, x_2, ..., x_n\})$ in R^N, representing its partitioning into a number, say K, of clusters[5]. Thus, given a vector containing the already filtered and reduced points (x,z) in $\vec{B}$.

Now let's establish a distance threshold denoted by $dth = 15$cm, which is the limit distance between 2 points enough close to consider that both belong to the same cluster. Then $j \in \mathbb{R}$, is the j^{th} cluster of points.

$$\vec{D}_l = \sqrt{(\vec{x}_i - \vec{V}^x)^2 + (\vec{z}_i - \vec{V}^z)^2}|_{l=1}^{||V||}$$

$\vec{D}$ is the distance between the current point (x_i, z_i) and the rest of the vector(points) with length $||\vec{V}||$. The valid points $\vec{P}$ are those points which their distances respect to (x_i, z_i) is less or equal than dth.

$$\vec{P} = \left\{ {1, \vec{D}_i \leq dth \atop 0, other} \right\}$$

Sub-cluster's addresses are represented by $\vec{OB}$, which keeps some subsets of data, which have not been clustered yet. There are basically 3 types of data subsets:(1)Points belonging to the current cluster j but

not classified yet. (2)Points, which are close to a pre-classified clusters and other unclassified points. (3)When points are close only to other pre-classified clusters. Let's see the following relation.

$$\vec{G}_1 = \left\{ \vec{OB}_i \ni \vec{P}_i \wedge \vec{OB}_i > 0 \right\} \cup \left\{ \vec{OB}_i \ni \vec{P}_i \right\}$$

$\vec{G}_1$ contains valid sub-clusters and valid points. In order to complete the clusters, the points still non-classified, such that those points and the previously classified in $\vec{G}_1$ are

$$\forall 1 \leq i \leq ||\vec{V}|| \Rightarrow \vec{G} = \vec{G}_1 \cup \left\{ \vec{OB}_i + (i \cup \vec{P}_i) \right\}$$

Now, the whole points have been classified as a new cluster, in the k^{th} object, now called Kob.

$$Kob = min_{g \in \vec{G}}$$

Then, the new cluster is updated as

$$\vec{OB}_{\vec{G}_k} = Kob$$

Once clustering algorithm has been applied, objects have been segmented and can be addressed by means of a determined numeric value automatically assigned during the process of clustering. Results are depicted in figure 15, clustered objects are the previous result of fused data from robots A, B and C.

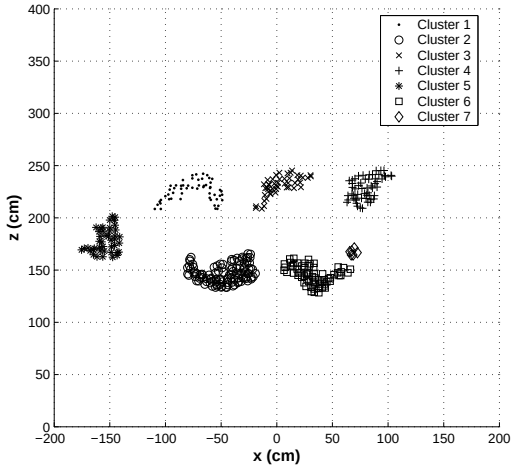

Figure 15: Clustering based segmentation of data from robot-A, B and C.

7 People localization

Since each object has been segmented, people localization can result a relatively simple task just by calculating cluster's centroid. Figure 16 shows how humans and other objects have been successfully localized. From data fusion, specifically measured data of robot-B, which it was able to ranged a table projected and classified as cluster 5 in figure 15. Meanwhile cluster 7 represents a fragmented portion of the cluster 6, which it does not really take importance for human localization, due to the fact that by means of statistical information such as variance, etc., and 3D geometrical measures, those objects are easily discriminated.

In addition, all segmented objects have been projected as ellipses in order to approximate them to human shapes as top-view's shapes. We considered for this process, some parameters such as cluster's centriod(x,z), width, depth and angle of data distribution in order to provide people's orientation.

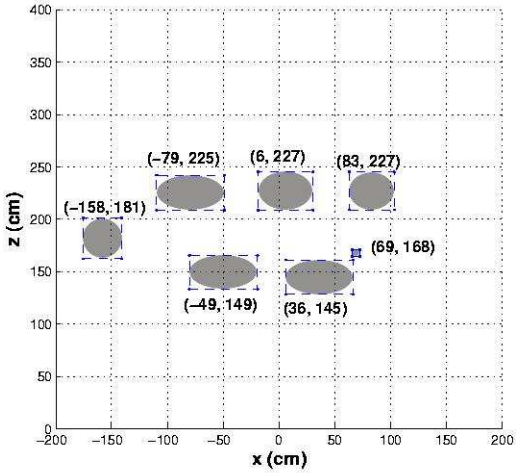

Figure 16: Detected objects from multi robot fusion data

8 Conclusions

A first step for steering groups of people has been developed and its results in a stationary environment have been showed. Figure 16 is the result of the experiment's configuration depicted in figure 5, where the mean error for centroid location is about $\pm$5cm, which is enough accurate for our purpose.

The method has resulted feasible for multi people localization in a cooperative framework with multi-mobile robots. Our next goal is attempting to localize multiple people in a dynamical fashion, trying to overcome the challenge of simultaneously multi robot's sensor data calibration and multi robot self localization.

References

[1] Kahng A.B., Meng F., et. al. Cooperative mobile robotics: Antecedents and directions. In *Intelligent robots and Systems, IROS'95*, pages 226-234, 1995.

[2] Whitaker W., Montemerlo M., Thrun S. "Conditional particle filters for simultaneous mobile robot localization and people-tracking". In *"Proceedings of the 2002 IEEE International Conference on Robotics and Automation"*, pages 695-701, May 2002.

[3] David Beymer and Kurt Konolige. real tiem tracking of multiple people using continuous detection. In *IEEE Proceedings International Conference on Computer Vision(ICCV'99)*, September 1999.

[4] *Installation guide and camera control API command reference*. Point Grey Research, 2000.

[5] Bandyopadhyay S., Maulik U., Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 24(12), pages 1650-1654, December 2002.