Comparing Improved Versions of 'K-Means' and 'Subtractive' Clustering in a Tracking Application

Marta Marrón Romera[1], Miguel Angel Sotelo Vázquez[1], Juan Carlos García García[1], Johanna Brodfelt[1],

[1] Electronics Department, University of Alcala,
Edificio Politécnico, Campus Universitario Alcalá de Henares, 28871, Madrid, Spain
{marta,sotelo,jcarlos}@depeca.uah.es

Abstract. Deterministic and probabilistic clustering algorithms are compared in this paper in terms of accuracy, robustness and efficiency. 3D data points extracted from a stereo-vision system have to be clustered for a tracking application in which a particle filter is the kernel of the estimation task. 'K-Means' and 'Subtractive' algorithms have been enriched with a validation process in order improve its functionality in the tracking task. Comparisons and conclusions of the clustering results both in a stand-alone process and in the proposed tracking task are shown in the paper.

Keywords: clustering, probabilistic-deterministic, particle-filters, tracking.

1 Introduction

Clustering algorithms are used in a large amount of applications. In artificial vision, these processes are particularly useful as compress visual data generating classes that contain more accurate and robust environmental information.

In the robotic application developed by the authors in [1], 3D information obtained from a stereo-vision system is used as input vector in a probabilistic estimation process, in an obstacle tracking and detection task. In this case, clustered data increases the reliability of the estimator.

In this paper, two different clustering algorithms are tested and compared in order to find the most adequate solution for the tracking task exposed.

2 Clustering algorithms

As mentioned two different clustering algorithms are compared. They are described in the following paragraphs:

- *Modified 'K-Means':* The deterministic solution presented by MacQueen, and afterwards improved ([2]), is modified to classify a variable number of clusters. A functional diagram of this clustering proposal is presented in Fig. 1, in which the functionality included to achieve an adaptive behaviour of the algorithm to

the number o
speed, in its r
looking for c
centroid from

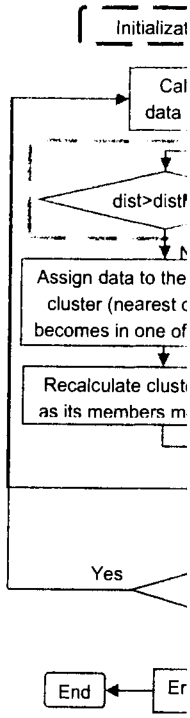

Fig. 1. Flowcharts of
describe 'K-Means' a
improvements includ
desired operation.

- *Modified 'Subt*
the forward on
used in this wo
and therefore a
found in the de
information is
functionality is
each cluster m
rate. This proce
in the dashed bl

All described im
algorithms. To inc

the number of clusters is marked with a dashed line. To improve the algorithm speed, in its recursive performance, the 'K-Means' segmentation process starts looking for clusters near a predicted value that is calculated for each class centroid from the previous execution of the algorithm (denoted with dashed line).

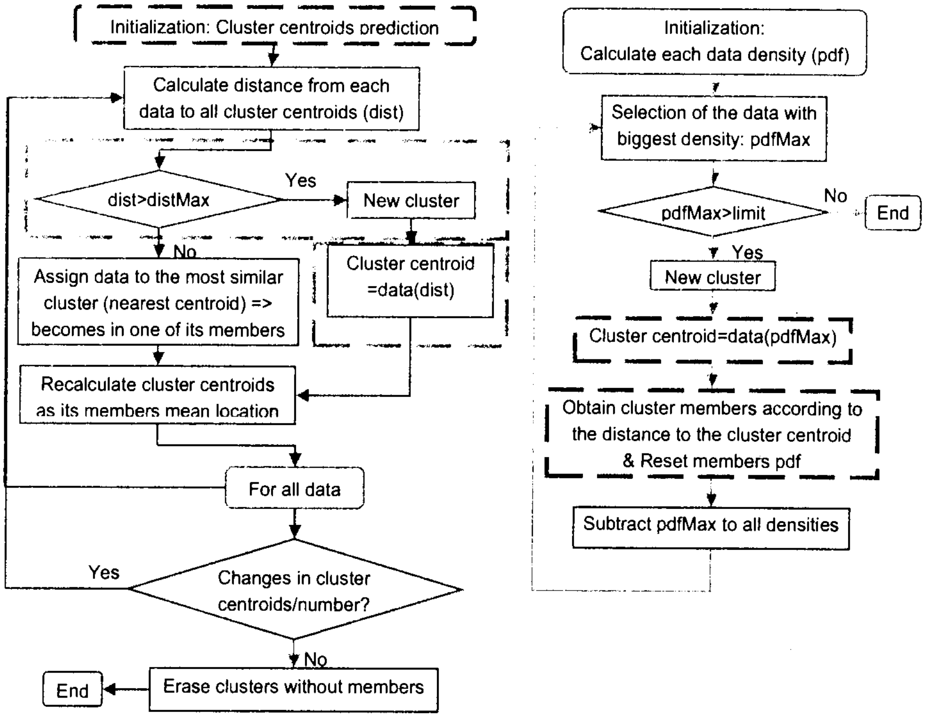

Fig. 1. Flowcharts of the clustering processes compared in this work. Left and right diagrams describe 'K-Means' and 'Subtractive' functionality respectively. Dashed lines remark the main improvements included in the standard version of both algorithms in order to achieve the desired operation.

- *Modified 'Subtractive':* A fuzzy solution based on the Mountain algorithm, and the forward one exposed in [3] by Chiu. The functionality of the fuzzy clustering used in this work is described in Fig. 1. The 'Subtractive' algorithm is unlinked, and therefore a process to obtain the centroid and the members for each cluster found in the density function has to be included in the standard algorithm if this information is needed, as it is the case of the present application. This functionality is shown with dashed lines in Fig. 1. Moreover, the probability of each cluster members is reset in order to decrease the cluster duplication error rate. This procedure is also included in the designed algorithm as it can be notice in the dashed blocks.

All described improvements increase the efficiency and accuracy of the clustering algorithms. To increase their robustness, a validation process is added to both clustering algorithms. This procedure is used when a cluster appears or seems to disappear due to noisy measurements or outliers. Fig. 2 shows the functionality of this algorithm. Two parameters are used in the validation process referred:

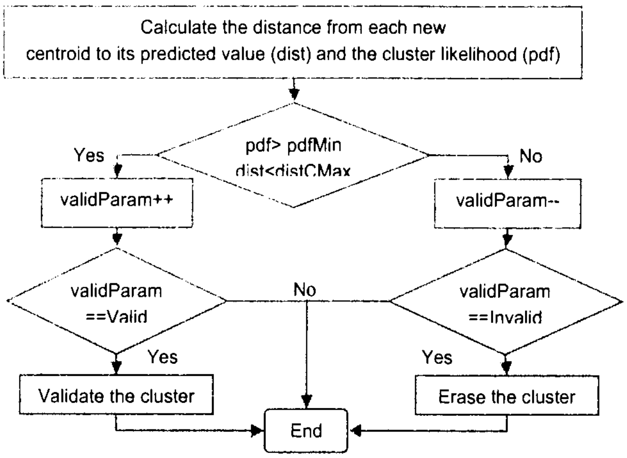

Fig. 2. Flowchart showing the validation process functionality used to create and erase clusters. This process is used to increase the robustness of the segmentation algorithms.

- *Distance between the estimated and the extracted cluster centroid.* The clusters centroid estimation already used with the commented functionality in the 'K-Means' is also developed for the 'Subtractive' one and used to compare the clustering results from two consecutive iterations, in order to obtain a confidence value for them.
- *Cluster likelihood.* The clusters inherent information of probability in 'Subtractive' algorithm is calculated also in 'K-Means', as a function of the measurements agglomeration in each cluster. This is also used as a validation parameter or confidence value in the cluster validation process.

3 Comparison

In order to extract comparative conclusions about the accuracy, robustness and efficiency of the clustering methods presented, they have been run with different data sets containing 3D points obtained from the environment in the tracking task mentioned.

From these experiments, many conclusions can be extracted. Here only some of the most important are resumed:

- 'Subtractive' clustering behaves better with noisy measurements than 'K-Means'.
- 'K-Means' algorithm is less time consuming than 'Subtractive' one (a mean execution time of 1.5ms for 'K-Means' versus 17ms for 'Subtractive'). The

execu
centr
- 'K-M
with
- The
reset
- The
outlie

The execu
extracted in
algorithms a
All tests
of B&W car
order to extr
Both algo
1, in order t
the tracking
shown in Fig

Fig. 3. Three
Data points ar

Acknowled
(CICYT: DF

Reference

1. Marrón M
particle fi
Proceedin
Systems (
2. Alsabti K.,
the First
Processing
3. Chiu S.L.:
& Fuzzy S

execution time in 'K-Means' is decreased almost in a 50% if predicted centroids are used in the segmentation.

- 'K-Means' algorithm shows a higher reliability than the 'Subtractive' one, with a mean error rate of 4.7% against 3.9%, respectively.
- The duplication error is decreased in 'Subtractive' algorithm in a 12% if the reset process included as an improvement in the algorithm is used.
- The validation process rejects almost the 90% of noisy measurements and outliers in both clustering algorithms.

The execution time is environmental dependent and the values shown have been extracted in a complex situation with more than 5 items moving in the scene. The algorithms are run in a PII standard computer at 1.9GHz.

All tests have been developed in real-time in a Pioneer 2AT robot in which a pair of B&W cameras has been statically mounted and a stereo-vision process is used in order to extract the 3D data points used in the segmentation.

Both algorithms have been used in the probabilistic tracker mentioned in paragraph 1, in order to improve its reliability. The functionality of the clustering algorithms in the tracking process is analyzed in [1], and an example of the results achieved is shown in Fig. 3.

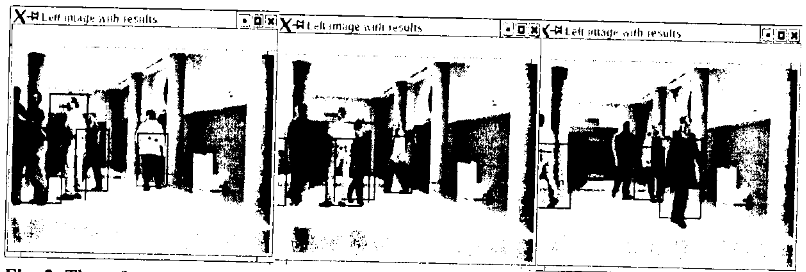

Fig. 3. Three frames in a real tracking sequence show clustering results in the tracking process. Data points are plotted in green and clusters red boxes.

Acknowledgments. This work has been financed by the Spanish administration (CICYT: DPI2005-07980-C03-02).

References

1. Marrón M., Sotelo M.A., García J.C., Fernandez D., Pizarro D.: XPFCP: An extended particle filter for tracking multiple and dynamic objects in complex environments. In: Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS05), ISBN: 0-7803-9252-3, pp. 234-239, Edmonton, August 2005.
2. Alsabti K., Ranka S., Singh V.: An efficient k-means clustering algorithm. In: Proceedings of the First Workshop on High Performance Data Mining at the International Parallel Processing Symposium (IPPS98), Orlando, 1998.
3. Chiu S.L.: Fuzzy model identification based on cluster estimation. In: Journal of Intelligent & Fuzzy Systems, Vol. 2, nº 3, pp. 267-278, September 1994.