An approach to consistently model and characterize potential object candidates presented in non-static scenes. Three principal procedures support our method:

i) the segmentation of the captured range images into 3D clusters or blobs, by which we obtain a first gross impression of the spatial structure of the scene,

ii) the maintenance and reliability of the map, which are obtained through the fusion of the captured and mapped data to which we assign a degree of existence (confidence value),

iii) the visual motion estimation of potential object candidates, through the combination of the texture and 3D-spatial information, allows not only to update the state of the actors and perceive their changes in a scene, but also to maintain and refine their individual 3D structures over time.