Recent advances in imaging sensors, such as Kinect, provide access to the synchronized depth with color, called RGB-D image. Numerous researches [2, 4] have shown that the use of depth as an additional feature improves accuracy of scene segmentation. However, it remains an important issue - what is the best way to fuse color and geometry in an unsupervised manner? We focus on this issue and propose a solution.

In this paper, we propose an unsupervised method for indoor RGB-D image segmentation and analysis. The proposed method combines a clustering method with a region merging method. First, it identifies the possible image regions using clustering w.r.t. a statistical image generation model. Then, it merges regions based on planar statistics.

We consider a statistical image generation model in order to fuse color and shape (3D and surface normal) features. The model assumes that the features are independently (naive Bayes assumption) issued from a finite mixture of multivariate Gaussian (for color and 3D) and a multivariate Watson distribution [6] (for surface normal). Mathematically, such a model with k components has the following form:

$$g(x_i|\Theta_k) = \sum_{j=1}^{k} \pi_{jk} f_{\Sigma_j}^C(x_i^C|\mu_{jk}, \Sigma_j) f_{\Sigma_j}^P(x_i^P|\mu_{jk}, \Sigma_j) f_{\Sigma_j}^\kappa(x_i^\kappa|\mu_{jk}, \Sigma_j)$$

where $x_i = (x_i^C, x_i^P, x_i^\kappa)$ is the feature vector of the $i$th pixel with $i = 1, ..., M$. Superscripts denote: C - color, P - 3D position and N - normal. $\Theta_k = \{\pi_{jk}, \mu_{jk}, \Sigma_j, \kappa_{jk}\}$, $j=1...k$ denotes the set of model parameters where $\pi_{jk}$ is the prior probability, $\mu_{jk}$ is the mean, $\Sigma_j$ is the variance-covariance matrix and $\kappa_{jk}$ is the concentration of the $j$th component. $f_{\Sigma_j}^C(.)$ and $f_{\Sigma_j}^P(.)$ are the density functions of the multivariate Gaussian distribution and the multivariate Watson [6] distribution respectively.

Fig. 1 illustrates the work flow of our RGB-D segmentation method that consists of two tasks: (1) cluster features and (2) merge regions. The first task performs a joint color-spatial-axial clustering and generates a set of regions. The second task performs a refinement on the set with the aim to merge regions which are susceptible to be over-segmented.

In order to tackle the over-segmentation issue mentioned above, we develop a statistical region merging method. It exploits planar property, which is related to the parameters (µ and κ) of the Watson distribution associated with each region. Our method first builds a region adjacency graph $G = (V,E)$. Each node $v_i \in V$ consists of concentration $\kappa_i$ of the surface normals of its corresponding region. Each edge $e_{ij}$ consists of two weights: $w_d$, based on statistical dissimilarity and $w_b$, based on boundary strength between adjacent nodes $v_i$ and $v_j$. Then, following the standard region merging methods [3], we define a region merging predicate as:

$$P_{ij} = \begin{cases} 
    \text{true}, & \text{if (a) } \kappa_i > \kappa_p \text{ and } w_d(v_i,v_j) < th_d \text{ and } w_b(v_i,v_j) < th_b \text{ and } \text{(c) planar outlier ratio} > th_c; \\
    \text{false}, & \text{otherwise.}
\end{cases}$$

where $\kappa_p$ is the threshold to define the planar property of a region. $th_d$ and $th_b$ are the thresholds associated with the distance weight $w_d$ and boundary weight $w_b$. $th_c$ is the threshold associated with the planar outlier ratio. The details of these thresholds are discussed in the paper. The region merging order sorts the adjacent regions that should be evaluated and merged sequentially.

Our proposed method is called JCSA-RM (joint color-spatial-axial clustering and region merging). We evaluate JCSA-RM on the benchmark image database NYUD2 [5] which consists of 1449 indoor RGB-D images with ground-truth segmentation. We evaluate its performance using five standard benchmarks: (1) Probability Rand Index (PRI); (2) Variation of Information (VI); (3) Boundary Displacement Error (BDE); (4) Ground Truth Region Covering (GTRC) and (5) Boundary based F-Measure (BFM).

First, we study the sensitivity of JCSA-RM w.r.t. the parameters ($k$, $\kappa_p$, $th_d$, $th_b$). Then, we compare JCSA-RM with several unsupervised RGB-D segmentation methods. Among them, RGB-D extension of OWT-UCM [4] (UCM-RGBD) method is the most competitive method. Results (presented in the paper) show that JCSA-RM performs best in PRI, Vol and GTRC and comparable in BDE and BFM. We compared these two competitive methods based on computation time and observe that JCSA-RM (MATLAB) is $\approx 3$ times faster than UCM-RGBD (C++).

JCSA-RM is an unsupervised RGB-D image segmentation method. It is comparable with the state of the art methods and it needs less computation time. It opens interesting perspectives to fuse color and geometry in an unsupervised manner. We foresee several possible extensions, such as: more complex image model and clustering with additional features, region merging with additional hypothesis based on color.