Unsupervised RGB-D image segmentation using joint clustering and region merging

Md. Abul Hasnat
mdabul.hasnat@univ-st-etienne.fr
Olivier Alata
olivier.alata@univ-st-etienne.fr
Alain Trémeau
alain.tremeau@univ-st-etienne.fr

Hubert Curien Lab., UMR CNRS 5516,
Jean Monnet University, Saint Etienne, France.

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 (*naïve 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:

$$g(\mathbf{x}_i|\Theta_k) = \sum_{j=1}^{k} \pi_{j,k} f_g(\mathbf{x}_i^C|\mu_{j,k}^C, \Sigma_{j,k}^C) f_g(\mathbf{x}_i^P|\mu_{j,k}^P, \Sigma_{j,k}^P) f_w\left(\mathbf{x}_i^N|\mu_{j,k}^N, \kappa_{j,k}^N\right)$$

Here $\mathbf{x}_i = \{\mathbf{x}_i^C, \mathbf{x}_i^P, \mathbf{x}_i^N\}$ is the feature vector of the ith pixel with $i = 1,...,M$. Superscripts denote: C - color, P - 3D position and N - normal. $\Theta_k = \{\pi_{j,k}, \mu_{j,k}^C, \Sigma_{j,k}^C, \mu_{j,k}^P, \Sigma_{j,k}^P, \mu_{j,k}^N, \kappa_{j,k}^N\}_{j=1...k}$ denotes the set of model parameters where $\pi_{j,k}$ is the prior probability, $\mu_{j,k}$ is the mean, $\Sigma_{j,k}$ is the variance-covariance matrix and $\kappa_{j,k}$ is the concentration of the jth component. $f_g(.)$ and $f_w(.)$ 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.

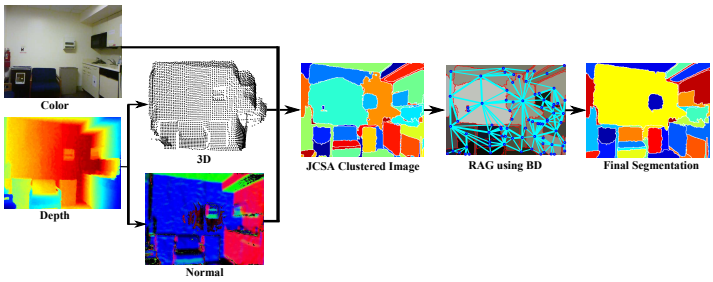

Figure 1: Work flow of the proposed RGB-D segmentation method.

We develop a Joint Color-Spatial-Axial (JCSA) clustering method to cluster pixels w.r.t. our image model. We exploit *Bregman Soft Clustering (BSC)* [1] method which has been effectively employed for mixture models based on exponential family of distributions. Compared to the traditional Expectation Maximization algorithms, BSC provides additional benefits: (a) it considers *Bregman Divergence* that generalizes a large number of distortion functions [1]; (b) simplifies computationally expensive Maximization step and (c) is applicable to mixed data type. Details of the JCSA clustering method is presented in the paper.

In an unsupervised setting the true number of segments are unknown. Therefore using JCSA, we cluster image features with an assumption of maximum number of clusters ($k = k_{max}$). Such an assumption often causes an over-segmentation of the image.

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 κ_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:

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

where κ_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_r is the threshold associated with the plane 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 (*VoI*); (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, κ_p, th_b, th_d). 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, VoI 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 ≈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.

[1] Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, and Joydeep Ghosh. Clustering with bregman divergences. *The Journal of Machine Learning Research*, 6:1705–1749, 2005.

[2] Carlo Dal Mutto, Pietro Zanuttigh, and Guido M Cortelazzo. Fusion of geometry and color information for scene segmentation. *IEEE Journal of Selected Topics in Signal Processing*, 6(5):505–521, 2012.

[3] Richard Nock and Frank Nielsen. Statistical region merging. *IEEE TPAMI*, 26(11):1452–1458, 2004.

[4] Xiaofeng Ren, Liefeng Bo, and Dieter Fox. Rgb-(d) scene labeling: Features and algorithms. In *CVPR*, pages 2759–2766. IEEE, 2012.

[5] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *Computer Vision–ECCV 2012*, pages 746–760. Springer, 2012.

[6] Suvrit Sra and Dmitrii Karp. The multivariate watson distribution: Maximum-likelihood estimation and other aspects. *J Multivar Anal*, 114:256 – 269, 2013.