Synergistic Clustering of Image and Segment Descriptors for Unsupervised Scene Understanding – Supplementary Material

Daniel M. Steinberg, Oscar Pizarro, Stefan B. Williams
Australian Centre for Field Robotics, The University of Sydney
Sydney, NSW, 2006
{d.steinberg, o.pizarro, stefanw}@acfr.usyd.edu.au

Abstract

This document contains supplementary material for the ICCV submission entitled Synergistic Clustering of Image and Segment Descriptors for Unsupervised Scene Understanding. This material includes additional detail of the variational Bayes algorithm for learning the MCM, the "split-tally" model selection heuristic, and extra detail and results for the LabelMe and Robot experiments.

1. Free Energy Learning Objective for the Multiple-source Clustering Model

The fully factored joint distribution for the MCM is,

$$\begin{aligned} p(\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\eta}, \boldsymbol{\Psi}) = \prod_{k=1}^{K} \mathcal{N}\big(\boldsymbol{\mu}_k | \mathbf{m}, (\gamma\boldsymbol{\Lambda}_k)^{-1}\big) \mathcal{W}(\boldsymbol{\Lambda}_k | \boldsymbol{\Omega}, \rho) \times \prod_{t=1}^{T} \mathrm{Dir}(\boldsymbol{\beta}_t|\theta) \mathcal{N}\big(\boldsymbol{\eta}_t | \mathbf{h}, (\delta\boldsymbol{\Psi}_t)^{-1}\big) \mathcal{W}(\boldsymbol{\Psi}_t | \boldsymbol{\Phi}, \xi) \\ \times \prod_{j=1}^{J} \mathrm{GDir}(\boldsymbol{\pi}_j | a, b) \times \prod_{j=1}^{J}\prod_{i=1}^{I_j} \mathrm{Categ}(y_{ji}|\boldsymbol{\pi}_j) \times \prod_{j=1}^{J}\prod_{i=1}^{I_j}\prod_{t=1}^{T} \mathcal{N}\big(\mathbf{w}_{ji} | \boldsymbol{\eta}_t, \boldsymbol{\Psi}^{-1}\big)^{\mathbf{1}[y_{ji}=t]} \\ \times \prod_{j=1}^{J}\prod_{i=1}^{I_j}\prod_{t=1}^{T}\prod_{n=1}^{N_{ji}} \mathrm{Categ}(z_{jin}|\boldsymbol{\beta}_t)^{\mathbf{1}[y_{ji}=t]} \times \prod_{j=1}^{J}\prod_{i=1}^{I_j}\prod_{k=1}^{K}\prod_{n=1}^{N_{ji}} \mathcal{N}\big(\mathbf{x}_{jin} | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\big)^{\mathbf{1}[z_{jin}=k]} . \end{aligned} \tag{1}$$

Here $\mathbf{1}[\cdot]$ is an indicator function, and evaluates to 1 when the expression in the brackets is true, 0 otherwise. For notational simplicity we have dropped the subscript of the parameters to indicate the set of all parameters. The variational Bayes derivations begin by approximating the joint distribution with a family of factorised mean-field approximating distributions,

$$q(\mathbf{Y}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Psi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{k=1}^{K} q(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) \times \prod_{t=1}^{T} q(\boldsymbol{\eta}_t, \boldsymbol{\Psi}_t)\, q(\boldsymbol{\beta}_t) \times \prod_{j=1}^{J} q(\boldsymbol{\pi}_j) \prod_{i=1}^{I_j} q(y_{ji}) \prod_{n=1}^{N_{ji}} q(z_{jin}) \,. \tag{2}$$

Following [1], the free energy lower bound is,

$$
\begin{aligned}
\mathcal{F}[q(\mathbf{Y},\mathbf{Z})\,,q(\boldsymbol{\pi},\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\Psi},\boldsymbol{\mu},\boldsymbol{\Lambda})] = \sum_{t=1}^{T}\Bigg\{\mathbb{E}_{q_{\boldsymbol{\beta}}}\left[\log\frac{\mathrm{Dir}(\boldsymbol{\beta}_t|\theta)}{q(\boldsymbol{\beta}_t)}\right] + \mathbb{E}_{q_{\boldsymbol{\eta},\boldsymbol{\Psi}}}\left[\log\frac{\mathcal{N}\big(\boldsymbol{\eta}_t|\mathbf{h},(\delta\boldsymbol{\Psi}_t)^{-1}\big)\,\mathcal{W}(\boldsymbol{\Psi}_t|\boldsymbol{\Phi},\xi)}{q(\boldsymbol{\eta}_t,\boldsymbol{\Psi}_t)}\right]\Bigg\} \\
+\sum_{k=1}^{K}\mathbb{E}_{q_{\boldsymbol{\mu},\boldsymbol{\Lambda}}}\left[\log\frac{\mathcal{N}\big(\boldsymbol{\mu}_k|\mathbf{m},(\gamma\boldsymbol{\Lambda}_k)^{-1}\big)\,\mathcal{W}(\boldsymbol{\Lambda}_k|\boldsymbol{\Omega},\rho)}{q(\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k)}\right] + \sum_{j=1}^{J}\mathbb{E}_{q_{\boldsymbol{\pi}}}\left[\log\frac{\mathrm{GDir}(\boldsymbol{\pi}_j|a,b)}{q(\boldsymbol{\pi}_j)}\right] \\
+\sum_{j=1}^{J}\sum_{i=1}^{Ij}\mathbb{E}_q\left[\log\frac{\mathrm{Categ}(y_{ji}|\boldsymbol{\pi}_j)\prod_{t=1}^{T}\mathcal{N}(\mathbf{w}_{ji}|\boldsymbol{\eta}_t,\boldsymbol{\Psi}_t)^{\mathbf{1}[y_{ji}=t]}}{q(y_{ji})}\right] \\
+\sum_{j=1}^{J}\sum_{i=1}^{Ij}\sum_{n=1}^{N_{ji}}\mathbb{E}_q\left[\log\frac{\prod_{t=1}^{T}\mathrm{Categ}(z_{jin}|\boldsymbol{\beta}_t)^{\mathbf{1}[y_{ji}=t]}\prod_{k=1}^{K}\mathcal{N}(\mathbf{x}_{jin}|\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k)^{\mathbf{1}[z_{jin}=k]}}{q(z_{jin})}\right],
\end{aligned}
\tag{3}
$$

where the last two terms' expectations are with respect to all of the latent variables and parameters. These last two terms act like data-fitting objectives, and the first three terms act as model complexity penalties.

2. Normalisation Constants

The normalisation constants for the variational posterior label distributions (Equations 5 and 6 in the original paper) are as follows,

$$
\mathcal{Z}_{y_{ji}} = \sum_{t=1}^{T}\exp\Big\{\mathbb{E}_{q_\pi}[\log\pi_{jt}] + \sum_{k=1}^{K}\mathbb{E}_{q_\beta}[\log\beta_{tk}]\sum_{n=1}^{N_{ji}}q(z_{jin}=k) + \mathbb{E}_{q_{\boldsymbol{\eta},\boldsymbol{\Psi}}}\big[\log\mathcal{N}\big(\mathbf{w}_{ji}|\boldsymbol{\eta}_t,\boldsymbol{\Psi}_t^{-1}\big)\big]\Big\}. \tag{4}
$$

$$
\mathcal{Z}_{z_{jin}} = \sum_{k=1}^{K}\exp\Big\{\sum_{t=1}^{T}q(y_{ij}=t)\,\mathbb{E}_{q_\beta}[\log\beta_{tk}] + \mathbb{E}_{q_{\boldsymbol{\mu},\boldsymbol{\Lambda}}}\big[\log\mathcal{N}\big(\mathbf{x}_{jin}|\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k^{-1}\big)\big]\Big\}. \tag{5}
$$

3. Dirichlet Distribution

The Categorical distribution is most often used as the likelihood of the Dirichlet distribution in this paper,

$$
\mathrm{Categ}(z_{jin}|\boldsymbol{\beta}_t) = \prod_{k=1}^{K}\beta_{tk}^{\mathbf{1}[z_{jin}=k]}. \tag{6}
$$

Here $\mathbf{1}[\cdot]$ is an indicator function, and evaluates to 1 when the condition in the brackets is true, and 0 otherwise. The corresponding Dirichlet prior has the form,

$$
\mathrm{Dir}(\boldsymbol{\beta}_t|\theta) = \frac{\Gamma(K\cdot\theta)}{\Gamma(\theta)^K}\prod_{k=1}^{K}\beta_{tk}^{\theta-1}, \tag{7}
$$

here $\Gamma(\cdot)$ is a Gamma function. This is a symmetric Dirichlet prior, another a commonly used parameterisation is $\mathrm{Dir}(\boldsymbol{\beta}_t|\theta/K)$.

3.1. Expectations over the likelihood

The log Categorical expectation under a generalised Dirichlet is,

$$
\begin{aligned}
\mathbb{E}_{q_{\boldsymbol{\beta}}}[\log p(z_{jin}=k|\boldsymbol{\beta}_t)] &= \mathbb{E}_{q_\beta}[\log\beta_{tk}] \\
&= \Psi\Big(\tilde{\theta}_{tk}\Big) - \Psi\left(\sum_k\tilde{\theta}_{tk}\right),
\end{aligned}
\tag{8}
$$

where $\Psi(\cdot)$ is a Digamma function, and $\tilde{\theta}_{tk}$ is from Equation (9) in the paper.

3.2. Free energy expectations

The expectations of the model complexity penalty terms are,

$$\mathbb{E}_{q_{\boldsymbol{\beta}}}\left[\log \frac{\mathrm{Dir}(\boldsymbol{\beta}_t|\theta)}{q(\boldsymbol{\beta}_t)}\right] = \log\Gamma(K\cdot\theta) - \log\Gamma\left(\sum_{k=1}^{K}\tilde{\theta}_{tk}\right) + \sum_{k=1}^{K}\log\Gamma\left(\tilde{\theta}_{tk}\right) - K\log\Gamma(\theta) - \sum_{k=1}^{K}\left(\tilde{\theta}_{tk}-\theta\right)\mathbb{E}_{q_\beta}[\log\beta_{tk}]\,, \quad (9)$$

where $\mathbb{E}_{q_\beta}[\log\beta_{tk}]$ is from Equation 8.

4. Generalised Dirichlet Distribution

The Categorical distribution is most often used as the likelihood of the Generalised Dirichlet distribution in this paper, see Equation 6. The generalised Dirichlet prior on the mixture weights, $\mathrm{GDir}(\boldsymbol{\pi}_j|a,b)$, is similar to a truncated stick-breaking process [2],

$$\pi_{jt} = v_{jt}\prod_{s=1}^{t-1}(1-v_{js}), \quad v_{jt}\sim\begin{cases}\mathrm{Beta}(a,b) & \text{if } t<T\\ 1 & \text{if } t=T.\end{cases} \quad (10)$$

where $v_{jt}\in[0,1]$ are "stick-lengths" for each group, and $\mathrm{Beta}(\cdot)$ is a Beta distribution,

$$\mathrm{Beta}(v_{jt}|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}v_{jt}^{a-1}\left(1-v_{jt}\right)^{b-1}. \quad (11)$$

Here $\Gamma(\cdot)$ is a Gamma function.

4.1. Expectations over the likelihood

The log Categorical expectation under a generalised Dirichlet is,

$$\begin{aligned}\mathbb{E}_{q_{\boldsymbol{\pi}}}[\log p(z_{jin}=k|\boldsymbol{\pi}_j)] &= \mathbb{E}_{q_\pi}[\log\pi_{jt}]\\ &= \mathbb{E}_{q_v}[\log v_{jt}] + \sum_{s=1}^{t-1}\mathbb{E}_{q_v}[\log(1-v_{js})]\,,\end{aligned} \quad (12)$$

where,

$$\mathbb{E}_{q_v}[\log v_{jt}] = \begin{cases}\Psi(\tilde{a}_{jt}) - \Psi\left(\tilde{a}_{jt}+\tilde{b}_{jt}\right) & \text{if } t<T\\ 0 & \text{if } t=T,\end{cases} \quad (13)$$

and,

$$\mathbb{E}_{q_v}[\log(1-v_{jt})] = \Psi\left(\tilde{b}_{jt}\right) - \Psi\left(\tilde{a}_{jt}+\tilde{b}_{jt}\right) \quad \text{if } t<T. \quad (14)$$

Here $\Psi(\cdot)$ is a Digamma function and $\tilde{a}_{jt}$ and $\tilde{b}_{jt}$ are from Equations (7) and (8) in the paper.

4.2. Free energy expectations

The expectations of the model complexity penalty terms can be factorised,

$$\mathbb{E}_{q_{\boldsymbol{\pi}}}\left[\log\frac{\mathrm{GDir}(\boldsymbol{\pi}_j|a,b)}{q(\boldsymbol{\pi}_j)}\right] = \sum_{t=1}^{T-1}\mathbb{E}_{q_\pi}\left[\log\frac{p(\pi_{jt}|a,b)}{q(\pi_{jt})}\right], \quad (15)$$

where

$$\begin{aligned}\mathbb{E}_{q_\pi}\left[\log\frac{p(\pi_{jt}|a,b)}{q(\pi_{jt})}\right] = &-(\tilde{a}_{jt}-a)\,\mathbb{E}_{q_v}[\log v_{jt}] - \left(\tilde{b}_{jt}-b\right)\mathbb{E}_{q_v}[\log(1-v_{jt})] + \log\Gamma(\tilde{a}_{jt})\\ &-\log\Gamma(a) + \log\Gamma\left(\tilde{b}_{jt}\right) - \log\Gamma(b) - \log\Gamma\left(\tilde{a}_{jt}+\tilde{b}_{jt}\right) + \log\Gamma(a+b)\,.\end{aligned} \quad (16)$$

The free energy penalty term over the weights in Equation 16 only sums to $T-1$ (degrees of freedom).

5. Gaussian-Wishart Distribution

Gaussian distributions are often used to describe segment clusters[1] in this paper, which take the form,

$$\mathcal{N}\left(\mathbf{x}_{jin}|\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right) = \frac{|\boldsymbol{\Lambda}_k|^{1/2}}{(2\pi)^{D/2}} \exp\left\{-\frac{1}{2}\left(\mathbf{x}_{jin} - \boldsymbol{\mu}_k\right)^\top \boldsymbol{\Lambda}_k \left(\mathbf{x}_{jin} - \boldsymbol{\mu}_k\right)\right\}. \tag{17}$$

A Gaussian-Wishart prior is placed over the parameters,

$$\mathcal{N}\left(\boldsymbol{\mu}_k|\mathbf{m}, (\gamma\boldsymbol{\Lambda}_k)^{-1}\right) = \frac{|\gamma\boldsymbol{\Lambda}_k|^{1/2}}{(2\pi)^{D/2}} \exp\left\{-\frac{\gamma}{2}\left(\boldsymbol{\mu}_k - \mathbf{m}\right)^\top \boldsymbol{\Lambda}_k \left(\boldsymbol{\mu}_k - \mathbf{m}\right)\right\}, \tag{18}$$

$$\mathcal{W}(\boldsymbol{\Lambda}_k|\boldsymbol{\Omega}, \rho) = \frac{|\boldsymbol{\Lambda}_k|^{(\rho-D-1)/2}}{2^{\rho D/2}\,|\boldsymbol{\Omega}|^{\rho/2}\,\Gamma_D\left(\frac{\rho}{2}\right)} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}_k\right)\right\}, \tag{19}$$

where $\Gamma_D(\cdot)$ is a multivariate Gamma function,

$$\Gamma_D\left(\frac{\rho}{2}\right) = \pi^{D(D-1)/4}\prod_{d=1}^{D}\Gamma\left(\frac{\rho+1-d}{2}\right), \tag{20}$$

and $\Gamma(\cdot)$ is a Gamma function.

5.1. Expectations over the likelihood

The log Gaussian expectation under a Gaussian-Wishart prior is,

$$\mathbb{E}_{q_{\boldsymbol{\mu},\boldsymbol{\Lambda}}}\left[\log\mathcal{N}\left(\mathbf{x}_{jin}|\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}\right)\right] = \frac{1}{2}\mathbb{E}_{q_{\boldsymbol{\Lambda}}}[\log|\boldsymbol{\Lambda}_k|] - \frac{D}{2\tilde{\gamma}_k} - \frac{\tilde{\rho}_k}{2}\left(\mathbf{x}_{jin} - \tilde{\mathbf{m}}_k\right)^\top \tilde{\boldsymbol{\Omega}}_k \left(\mathbf{x}_{jin} - \tilde{\mathbf{m}}_k\right), \tag{21}$$

where

$$\mathbb{E}_{q_{\boldsymbol{\Lambda}}}[\log|\boldsymbol{\Lambda}_k|] = \sum_{d=1}^{D}\Psi\left(\frac{\tilde{\rho}_k+1-d}{2}\right) + D\log 2 + \log\left|\tilde{\boldsymbol{\Omega}}_k\right|, \tag{22}$$

and $\Psi(\cdot)$ is a Digamma function.

5.2. Variational updates

The variational posterior Gaussian-Wishart hyper-parameters are,

$$\tilde{\gamma}_k = \gamma + N_k, \tag{23}$$

$$\tilde{\mathbf{m}}_k = \frac{1}{\tilde{\gamma}_k}\left(\gamma\mathbf{m} + N_k\bar{\mathbf{x}}_k\right), \tag{24}$$

$$\tilde{\rho}_k = \rho + N_k, \tag{25}$$

$$\tilde{\boldsymbol{\Omega}}_k^{-1} = \boldsymbol{\Omega}^{-1} + N_k\mathbf{R}_k + \frac{\gamma N_k}{\tilde{\gamma}_k}(\bar{\mathbf{x}}_k - \mathbf{m})(\bar{\mathbf{x}}_k - \mathbf{m})^\top, \tag{26}$$

where

$$N_k = \sum_{j=1}^{J}\sum_{i=1}^{I_j}\sum_{n=1}^{N_{ji}} q(z_{jin} = k), \tag{27}$$

$$\bar{\mathbf{x}}_k = \frac{1}{N_k}\sum_{j=1}^{J}\sum_{i=1}^{I_j}\sum_{n=1}^{N_{ji}} q(z_{jin} = k)\,\mathbf{x}_{jin}, \tag{28}$$

$$\mathbf{R}_k = \frac{1}{N_k}\sum_{j=1}^{J}\sum_{i=1}^{I_j}\sum_{n=1}^{N_{ji}} q(z_{jin} = k)\,(\mathbf{x}_{jin} - \bar{\mathbf{x}}_k)(\mathbf{x}_{jin} - \bar{\mathbf{x}}_k)^\top. \tag{29}$$

Note that $\rho \geq D - 1$.

[1] These equations are almost the same for the image clusters, so they are omitted.

5.3. Free energy expectations

The expectations of the model complexity penalty terms are,

$$\mathbb{E}_{q_{\mu,\Lambda}}\left[\log \frac{\mathcal{N}\left(\boldsymbol{\mu}_k|\mathbf{m},(\gamma\boldsymbol{\Lambda}_k)^{\text{-}1}\right)\mathcal{W}(\boldsymbol{\Lambda}_k|\boldsymbol{\Omega},\rho)}{q(\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k)}\right] = -\frac{D}{2}\left(\frac{\gamma}{\tilde{\gamma}_k} - \log\frac{\gamma}{\tilde{\gamma}_k} - \tilde{\rho}_k - 1\right)$$
$$- \frac{\rho}{2}\left(\log|\boldsymbol{\Omega}| - \log|\tilde{\boldsymbol{\Omega}}_k|\right) - \frac{\tilde{\rho}_k}{2}\mathrm{Tr}\left(\boldsymbol{\Omega}^{\text{-}1}\tilde{\boldsymbol{\Omega}}_k\right) - \frac{\tilde{\rho}_k\gamma}{2}\left(\tilde{\mathbf{m}}_k - \mathbf{m}\right)^\top\tilde{\boldsymbol{\Omega}}_k\left(\tilde{\mathbf{m}}_k - \mathbf{m}\right)$$
$$- \sum_{d=1}^{D}\left(\frac{N_k}{2}\Psi\left(\frac{\tilde{\rho}_k+1-d}{2}\right) + \log\Gamma\left(\frac{\rho+1-d}{2}\right) - \log\Gamma\left(\frac{\tilde{\rho}_k+1-d}{2}\right)\right). \quad (30)$$

6. The "split-tally" heuristic for the MCM

The greedy splitting heuristic is based on two criteria. The first is the approximate free energy contribution of the segment cluster parameters and segment observations to be split. The second is how many split attempts have been tried for the segment cluster and not been accepted previously. The cluster split attempts are ordered by (a) least number of previous split attempts for the clusters, then (b) clusters with more free energy contribution. The first attempt that reduces model free energy is accepted. The approximate contribution to free energy is formulated from the heuristic,

$$\hat{\mathcal{F}}_k = \mathbb{E}_{q_{\mu,\Lambda}}\left[\log \frac{\mathcal{N}\left(\boldsymbol{\mu}_k|\mathbf{m},(\gamma\boldsymbol{\Lambda}_k)^{\text{-}1}\right)\mathcal{W}(\boldsymbol{\Lambda}_k|\boldsymbol{\Omega},\rho)}{q(\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k)}\right] + \sum_{j=1}^{J}\sum_{i=1}^{I_j}\sum_{n=1}^{N_{ji}} q(z_{jin}=k)\,\mathcal{L}_{z_{jin}=k} \quad (31)$$

where $\mathcal{L}_{z_{jin}=k}$ is the mixture likelihood of observation $\mathbf{x}_{jin}$ under segment cluster k (including the effect of the mixture weights). This likelihood is weighted by the observation's probabilistic membership to cluster k. For the MCM the exact form of this heuristic is,

$$\hat{\mathcal{F}}_k = \mathbb{E}_{q_{\mu,\Lambda}}\left[\log \frac{\mathcal{N}\left(\boldsymbol{\mu}_k|\mathbf{m},(\gamma\boldsymbol{\Lambda}_k)^{\text{-}1}\right)\mathcal{W}(\boldsymbol{\Lambda}_k|\boldsymbol{\Omega},\rho)}{q(\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k)}\right] + \sum_{j=1}^{J}\sum_{i=1}^{I_j}\sum_{n=1}^{N_{ji}} q(z_{jin}=k)\,\mathbb{E}_{q_{\mu,\Lambda}}\left[\log\mathcal{N}\left(\mathbf{x}_{jin}|\boldsymbol{\mu}_k,\boldsymbol{\Lambda}_k^{\text{-}1}\right)\right]. \quad (32)$$

A cluster weight term was not included in Equation 32 because a corresponding term of opposite sign existed in the last term in Equation 3, and adding it would nullify its effect in the overall model free energy.

The observations belonging to a Gaussian (with $q(z_{jin}=k) > 0.5$) are split in a direction perpendicular to its principal Eigenvector. This split is refined by iterating the VBE and VBM steps on only these observations. The algorithm is summarised in Algorithm 1. The expected model free energy, $\mathbb{E}[\mathcal{F}_{split,k}]$ is acquired by running variational Bayes for one iteration, with the new split, using all of the segment observations. To our knowledge this is the first time a split tally has been used in a cluster splitting heuristic. It was found to significantly reduce the run time of the algorithm and improve results over just using approximate free energy to guide the greedy search. This greedy cluster splitting heuristic often less than halved the run time of the total algorithm compared to the exhaustive cluster splitting heuristic. This speed-up was even more pronounced for the larger datasets. It also managed to maintain good clustering results compared to the exhaustive heuristic.

7. Extra Robotic Dataset Figures

Here we present some figures that accompany section 5.4 in the original paper. Figure 1 presents exemplar images of the nine hand-labelled AUV image classes. In Figure 2 an example of a run of the MCM is shown. Even after lens distortion correction, some of the segment clusters seem to be sensitive radial textual distortion and vignetting effects. This is especially obvious in the fourth image cluster (red) from the top. Unfortunately completely correcting for the air-lens-water refraction, and wavelength attenuation effects is still an open research question, and is beyond the scope of this work. Despite the relatively low NMI scores, the image clusters look very visually consistent, and the larger objects have been discovered by the segment clusters.

8. LabelMe Segment Classes

All LabelMe [3] segment labels with 4 or more instances were combined from over 300 into 22 classes;

Algorithm 1: The MCM greedy model selection heuristic

Data: Observations $\mathbf{W}, \mathbf{X}$

Result: Probabilistic assignments $q(\mathbf{Y})$ and $q(\mathbf{Z})$ and posterior hyper-parameters $\tilde{\mathbf{\Xi}}$

$\mathbf{\Xi} \leftarrow$ `CreatePriors()`;

$q(\mathbf{Y}) \leftarrow$ `RandomLabels(`$T_{trunc} = 30$`)`;

$q(\mathbf{Z}) \leftarrow \{\{\mathbf{1}\}_{i=1}^{I_j}\}_{j=1}^{J}$; // initialises with $K = 1$

splittally $\leftarrow \{0\}_{k=1}^{K}$;

repeat

  $q(\mathbf{Y})\,,q(\mathbf{Z})\,,\tilde{\mathbf{\Xi}}, \mathcal{F} \leftarrow$ `VarBayes(`$\mathbf{X}, q(\mathbf{Y})\,, q(\mathbf{Z})\,, \mathbf{\Xi}$`)`;

  splitorder $\leftarrow$ `GreedySorter(`$\mathbf{W}, \mathbf{X}, q(\mathbf{Z})\,, G,$ splittally`)`; // this is a sequence

  foreach $k \in$ splitorder **do**

    $\mathbf{X}_{split,k} \leftarrow \{\mathbf{x}_{jin} \in \mathbf{X} : q(z_{jin} = k) > 0.5\}$;

    $q(\mathbf{Z}_{split,k}) \leftarrow$`ClusterSplit(`$\mathbf{X}_{split,k}$`)`;

    $q(\mathbf{Z}_{split,k}) \leftarrow$`VarBayes(`$\mathbf{W}, \mathbf{X}_{split,k}, \{\mathbf{1}\}_{j=1}^{J}, q(\mathbf{Z}_{split,k}), \mathbf{\Xi}$`)` ; // refine

    $q(\mathbf{Z}_{aug,k}) \leftarrow$ `AugmentLabels(`$q(\mathbf{Z})\,, q(\mathbf{Z}_{split,k})$`)`; // add in split labels

    $\mathbb{E}[\mathcal{F}_{split,k}] \leftarrow$`VarBayes(`$\mathbf{W}, \mathbf{X}, q(\mathbf{Y})\,, q(\mathbf{Z}_{aug,k})\,, \mathbf{\Xi}$`)`; // 1 iteration

    if $\mathcal{F} > \mathbb{E}[\mathcal{F}_{split,k}]$ **then**

      $q(\mathbf{Z}) \leftarrow q(\mathbf{Z}_{aug,k})$;

      splittally$_k \leftarrow 0$;

      splittally$_{K+1} \leftarrow 0$;

      foundsplit $\leftarrow$ *true*;

      break;

    else

      splittally$_k \leftarrow$ splittally$_k + 1$;

      foundsplit $\leftarrow$ *false*;

until foundsplit $=$ *false*;

$q(\mathbf{Y}) \leftarrow$ `PruneEmptyClusters(`$q(\mathbf{Y})$`)`;

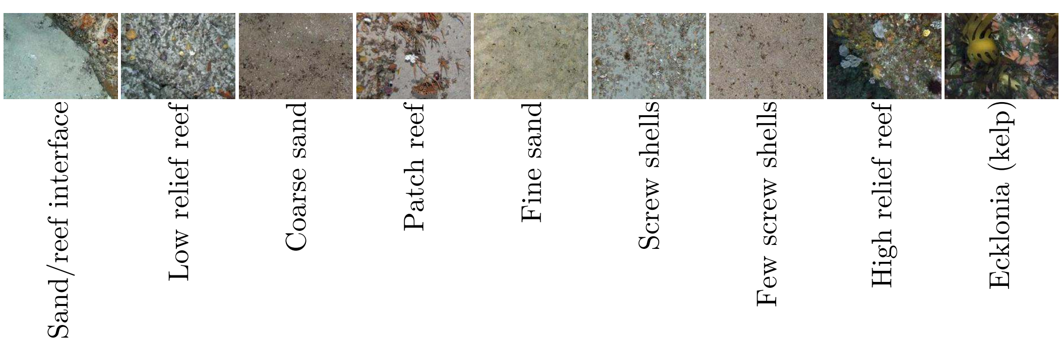
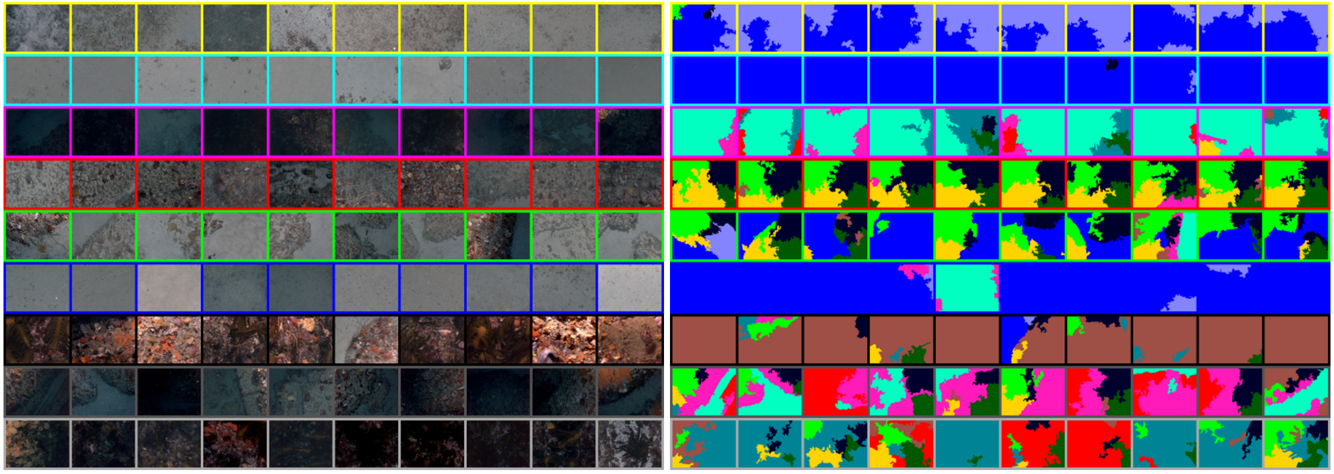


Figure 1. Exemplars of the nine hand-labelled AUV image classes.

1. person
2. vehicles
3. building
4. monument
5. grass/fields/garden
6. water



(a) Random image cluster examples

(b) Corresponding segment clusters

Figure 2. A random selection of images from all of the 9 image clusters found by the MCM (a). Also shown in (b) are the corresponding segment clusters (11 in total). In this run, the image clusters achieved a NMI score of 0.499, and the segments a NMI score of 0.325. The priors used were $C_{width,i} = 0.27$ and $C_{width,s} = 18$.

7. plants
8. animal
9. boat
10. mountain
11. snow
12. road
13. curb/sidewalk
14. sign
15. light
16. ground
17. rock
18. sand
19. cloud
20. sky
21. sun
22. desert

Many of the original labels were either synonyms, or spelling mistakes. We will make our tag file and code available at https://github.com/dsteinberg.

References

[1] M. J. Beal. *Variational algorithms for approximate Bayesian inference*. PhD thesis, University College London, 2003. 2

[2] H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001. 3

[3] B. C. Russell, A. Torralba, K. Murphy, and W. Freeman. LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77:157–173, 2008. 10.1007/s11263-007-0090-8. 5