1 Probabilistic Generative Mixture Model

The proposed static structure is modeled by a probabilistic generative mixture model. Three states are introduced to describe the cases that the input depth samples may occupy, each of which has a distribution as

- **State-I: Fitting the static structure** $p(d_x | Z_x, m_x^I = 1) = \mathcal{N}(d_x | Z_x, \xi_x^2)$;
- **State-F: Forward Outliers** $p(d_x | Z_x, m_x^F = 1) = \mathcal{U}_f(d_x^F | Z_x) = U_f \cdot 1_{|d_x^F < Z_x|}$;
- **State-B: Backward Outliers** $p(d_x | Z_x, m_x^B = 1) = \mathcal{U}_b(d_x^B | Z_x) = U_b \cdot 1_{|d_x^B > Z_x|}$.

For the purpose to combine all the three states into a united model and describe the overall likelihood that the input depth samples fit the current static structure, we use a mixture model similar to the Gaussian Mixture Model [1]. Together with prior distributions of the hidden variable $m_x$ and the static structure $Z_x$, we can further estimate the posterior with respect to $Z_x$ to infer the most possible static structure given the input depth samples, and the posterior with respect to $m_x$ to indicate the states that the input depth samples belong to.

1.1 Likelihood

The likelihood of the input depth sample $d_x^I$ with respect to the depth value of the static structure $Z_x$ and the hidden variable for the state indicator $m_x$ is

$$
p(d_x^I | m_x, Z_x) = \mathcal{N}(d_x^I | Z_x, \xi_x^2)^{m_x^I} \times \mathcal{U}_f(d_x^F | Z_x)^{m_x^F} \times \mathcal{U}_b(d_x^B | Z_x)^{m_x^B},
$$

which switches among these states by setting one specific $m_x^k = 1, k \in \Psi = \{I, F, B\}$ and the rest as 0s.

1.2 Prior Distributions

Given the likelihood as well as suitable prior distributions, we will have a tractable joint distribution. Thus the choices of the priors are essential to ensure tractable and efficient estimation of the joint distribution as well as the posteriors.

To be compatible with the likelihood in Sec. 1.1, we also introduce a Gaussian distribution for $Z_x$ as

$$
p(Z_x) = \mathcal{N}(Z_x | \mu_x, \sigma_x^2).
$$

The prior for $m_x$ needs to cope with the switching property that $m_x$ offers. Thus we employ the categorical distribution since it outputs a probability $\omega_x^k$ when a state $m_x^k$ is activated

$$
p(m_x | \omega_x) = \text{Cat}(m_x | \omega_x) = \prod_{k \in \Psi} (\omega_x^k)^{m_x^k}, \quad \text{given} \quad \sum_{k \in \Psi} \omega_x^k = 1.
$$

This distribution otherwise introduce an additional parameter $\omega_x$, which also needs an explicit distribution [1]. We apply the Dirichlet distribution as

$$
p(\omega_x) = \text{Dir}(\omega_x | \alpha_x^I, \alpha_x^F, \alpha_x^B) = \text{Dir}(\omega_x | \alpha_x), \quad \text{given} \quad \alpha_x^k \geq 0, k \in \Psi.
$$

The reason to introduce $p(\omega_x)$ is that we want to model the chance that one state may happen so that we can judge the reliability of the estimated static structure. Furthermore, given a prior distribution for $\omega_x$, we can further estimate the posterior with respect to $\omega_x$ when a series of data come into the model.

---

1 Please refer to the journal paper to find the definition of symbols.
1.3 Joint Distribution

Given the input depth sample $d^t_x$, the joint distribution can be written as

$$p(d^t_x; Z_x, m_x; \omega_x; \mathcal{P}_x) = \int p(d^t_x|Z_x, m_x; \xi_x)p(Z_x; \mu_x, \sigma_x)p(m_x|\omega_x)p(\omega_x; \alpha_x),$$

(5)

in which the parameter set is $\mathcal{P}_x = \{\xi_x, \mu_x, \sigma_x, \alpha_x\}$. By marginalizing the hidden variable, we can have a joint distribution that only contains two variables: the depth value $Z_x$ and the chance of each state $\omega_x$, as well as the observation $d^t_x$ and the parameters $\mathcal{P}_x$. It will result in a distribution as

$$p(d^t_x, Z_x, \omega_x; \mathcal{P}_x) = \int \omega_x N(d^t_x|Z_x, \alpha_x^2) + \omega_x B \int U_f \left( \int d^t_x | Z_x \right) p(Z_x; \mu_x, \sigma_x)p(\omega_x; \alpha_x),$$

(6)

which is a weighted combination of the three state densities multiplied with the prior distributions of $Z_x$ and $\omega_x$.

1.4 Data Evidence

The data confidence $p(d^t_x; \mathcal{P}_x)$ is simply calculated by marginalizing the variables $Z_x$ and $\omega_x$ a step further as

$$p(d^t_x; \mathcal{P}_x) = \mathcal{P}_x \int_{Z_x} \int_{\omega_x} p(d^t_x, Z_x, \omega_x; \mathcal{P}_x) dZ_x d\omega_x =$$

$$\frac{1}{\sum_{k \in \wp} \alpha^k_x} \left\{ \alpha^0_x N(d^t_x|\mu_x, \xi^2_x + \sigma^2_x) + \left( \alpha^B_x \mu - \alpha^F_x U_f \right) \Phi \left( \frac{d^t_x - \mu_x}{\sigma_x} \right) + \alpha^F_x U_f \right\}.$$  

(7)

1.5 Posteriors with First-order Markov Chain

In this paper, we want to estimate the posterior under an online fashion, it means the posterior is estimated frame by frame, with new data to sequentially increase the confidence of the static structure.

$$p(Z_x, \omega_x|D^t_x; \mathcal{P}_x) = \frac{1}{p(d^t_x|D^t_x; \mathcal{P}_x)} \int_{Z_x} \int_{\omega_x} p(d^t_x, Z_x, \omega_x; \mathcal{P}_x)p(Z_x, \omega_x|D^t_x; \mathcal{P}_x) dZ_x d\omega_x$$

(8)

The posterior with respect to the hidden variable $m_x$ indicates the distributions of the states that the input depth sample may occupy, it is similar as equation (8).

$$p(m_x|D^t_x; \mathcal{P}_x) = \frac{1}{p(d^t_x|D^t_x; \mathcal{P}_x)} \int_{Z_x} \int_{\omega_x} p(d^t_x, m_x, Z_x; \mathcal{P}_x)p(m_x|\omega_x)p(Z_x, \omega_x|D^t_x; \mathcal{P}_x) dZ_x d\omega_x$$

(9)

The posteriors seem complex and are not easy to estimate, thus we employ the variational approximation so that the posterior $p(Z_x, \omega_x|D^t_x; \mathcal{P}_x)$ can be factorized into the product of an independent Gaussian distribution $q^t(Z_x)$ and an independent Dirichlet distribution $q^t(\omega_x)$ with suitable parameters. The posterior $p(m_x|D^t_x; \mathcal{P}_x)$ can also be rewritten by substituting $p(Z_x, \omega_x|D^t_x; \mathcal{P}_x)$ with the approximated posterior $q^t(Z_x, \omega_x)$.

2 Derivations of the Results in Variational Approximation

In this section, we show the detailed derivations of the results present in Section III-B. For brevity, we omit the related superscripts and subscripts of parameters and variables as follows:

$$\{d^t_x, Z_x, \omega_x\} \implies \{d, Z, \omega\}$$

$$\{\mu^{t-1}_x, \sigma^{t-1}_x, d^t_x, \xi^t_x, \xi^t_x\} \implies \{\mu, \sigma, \mu_{new}, \sigma_{new, \omega}, \xi\}$$

$$\{\alpha^{t-1}_x, \alpha^{t-1}_x, B^{t-1}_x, \alpha^{t+1}_x, \alpha^{t+1}_x, \sum_{k \in \wp} \alpha^{k,t}_x, \sum_{k \in \wp} \alpha^{k,t}_x\} \implies \{\alpha_1, \alpha_2, \alpha_3, \alpha_1^{new}, \alpha_2^{new}, \alpha_3^{new}, \alpha_0, \alpha_0^{new}\}.$$ 

2.1 Approximated Joint Distribution $Q(Z, \omega, d)$

Incorporating the properties of Gaussian and Dirichlet distribution, the approximated joint distribution is in detail a mixture of several products of Gaussian and Dirichlet distributions.

$$Q(Z, \omega, d) = \int p(d, Z, \omega)^{1-t} (Z, \omega) =$$

$$\left( \omega^t N(d|Z, \xi^2) + \omega^t U_b(d|Z) + \omega^F U_f(d|Z) \right) N(Z|\mu, \sigma^2) \text{Dir}(\omega|\alpha_1, \alpha_2, \alpha_3)$$

$$= \frac{\alpha_1}{\alpha_0} N \left( d; \mu, \sigma^2 \right) N \left( Z; \frac{\xi^2 \mu + \sigma^2 d}{\xi^2 + \sigma^2}, \frac{\xi^2 \sigma^2}{\xi^2 + \sigma^2} \right) \text{Dir}(\omega|\alpha_1 + 1, \alpha_2, \alpha_3)$$

$$+ \frac{\alpha_2}{\alpha_0} U_f(d|Z) N \left( Z; \mu, \sigma^2 \right) \text{Dir}(\omega|\alpha_1, \alpha_2 + 1, \alpha_3) + \frac{\alpha_3}{\alpha_0} U_b(d|Z) N \left( Z; \mu, \sigma^2 \right) \text{Dir}(\omega|\alpha_1, \alpha_2, \alpha_3 + 1).$$

(10)
2.2 Approximated Joint Distribution \( Q(Z, d) \) and \( Q(\omega, d) \)

It is easy to calculate the moments related to \( Z \) and \( \omega \) respectively by estimating the moments of the approximated posterior \( Q(Z|d) \) and \( Q(\omega|d) \). Specifically, we need to calculate the joint distribution with respect to \( Z \) and \( d \) as

\[
Q(Z, d) = \int \omega Q(Z, \omega, d) d\omega = \frac{\alpha_1}{\alpha_0} N(d|\mu, \xi^2 + \sigma^2) \mathcal{N}(Z|\xi^2 \mu + \sigma^2 d, \xi^2 + \sigma^2) + \frac{\alpha_2}{\alpha_0} U_f(d|Z) \mathcal{N}(Z|\mu, \sigma^2) + \frac{\alpha_3}{\alpha_0} U_b(d|Z) \mathcal{N}(Z|\mu, \sigma^2),
\]

and the joint distribution with respect to \( \omega \) and \( d \) as

\[
Q(\omega, d) = \int \omega Q(\omega, d) d\omega = \frac{1}{\alpha_0} \left\{ \alpha_1 N(d|\mu, \xi^2 + \sigma^2) \text{Dir}(\omega|\alpha_1 + 1, \alpha_2, \alpha_3) + \alpha_3 U_f((d - \mu)/\sigma) \text{Dir}(\omega|\alpha_1, \alpha_2, \alpha_3 + 1) + \alpha_2 U_f(1 - \Phi((d - \mu)/\sigma)) \text{Dir}(\omega|\alpha_1, \alpha_2 + 1, \alpha_3) \right\}.
\]

2.3 Approximated Data Evidence \( q^t(d) \)

Similarly, the approximated data evidence is presented below:

\[
q^t(d) = \int Z Q(Z, \omega, d) dZ = \frac{\alpha_1}{\alpha_0} N(d|\mu, \xi^2 + \sigma^2) + \frac{\alpha_2}{\alpha_0} U_f \left[ \mu (1 - \Phi((d - \mu)/\sigma)) + \sigma^2 N(d|\mu, \sigma^2) \right] + \alpha_3 U_b \left[ \mu \Phi((d - \mu)/\sigma) - \sigma^2 N(d|\mu, \sigma^2) \right],
\]

which is also analytic as long as the parameters are known. The posteriors \( Q(Z|d) \) and \( Q(\omega|d) \) are calculated accordingly by dividing the joint distribution \( Q(Z, d) \) and \( Q(\omega, d) \) by the data evidence \( q^t(d) \).

2.4 Parameter Updating for \( q^t(Z) \)

The parameter estimation for \( q^t(Z) \) is to match first and second moments between \( q^t(Z) \) and \( Q(Z|d) \). The first moment between \( q^t(Z) \) and \( Q(Z|d) \) is

\[
\mu_{\text{new}} = \mathbb{E}_{Q(Z|d)}[Z] = \frac{1}{q^t(d) \alpha_0} \left\{ \alpha_1 N(d|\mu, \xi^2 + \sigma^2) \frac{\xi^2 \mu + \sigma^2 d}{\xi^2 + \sigma^2} + \alpha_2 U_f \left[ \mu (1 - \Phi((d - \mu)/\sigma)) + \sigma^2 N(d|\mu, \sigma^2) \right] + \alpha_3 U_b \left[ \mu \Phi((d - \mu)/\sigma) - \sigma^2 N(d|\mu, \sigma^2) \right] \right\},
\]

which can be further written as

\[
\mu_{\text{new}} = \frac{1}{q^t(d) \alpha_0} \left\{ \alpha_1 N(d|\mu, \xi^2 + \sigma^2) \frac{\xi^2 \mu + \sigma^2 d}{\xi^2 + \sigma^2} + \alpha_2 U_f \mu + (\alpha_3 U_b - \alpha_2 U_f) \left[ \mu \Phi((d - \mu)/\sigma) - \sigma^2 N(d|\mu, \sigma^2) \right] \right\}.
\]

The second moment is under a similar fashion as

\[
\mu_{\text{new}}^2 + \sigma_{\text{new}}^2 = \mathbb{E}_{Q(Z|d)}[Z^2] = \frac{1}{q^t(d) \alpha_0} \left\{ \alpha_1 N(d|\mu, \xi^2 + \sigma^2) \left[ \left( \frac{\xi^2 \mu + \sigma^2 d}{\xi^2 + \sigma^2} \right)^2 + \frac{\xi^2 \sigma^2}{\xi^2 + \sigma^2} \right] + \alpha_2 U_f \left[ \mu^2 + \sigma^2 \right] + (\alpha_3 U_b - \alpha_2 U_f) \left[ \mu^2 + \sigma^2 \Phi((d - \mu)/\sigma) - (d + \mu) \sigma^2 N(d|\mu, \sigma^2) \right] \right\}.
\]

2.5 Parameter Updating for \( q^t(\omega) \)

The parameters \( \alpha_k^{\text{new}}, k \in \{1, 2, 3\} \) are calculated by introducing new variables \( m_i \) and \( m_i^{(2)}, i = 1, 2, 3 \), which defines the first moments and the second moments with respect to \( \omega \) for \( q^t(\omega) \) [2]. The first moments are calculated according to the property of the Dirichlet distribution as

\[
m_1 = \frac{\alpha_1^{\text{new}}}{\alpha_0^{\text{new}}} = \mathbb{E}_{Q(\omega|d)}[\omega] = \frac{\alpha_1}{\alpha_0 q^t(d)} N(d|\mu, \xi^2 + \sigma^2) \frac{\alpha_1 + 1}{\alpha_0 + 1} + \frac{\alpha_2}{\alpha_0 q^t(d)} U_f (1 - \Phi((d - \mu)/\sigma)) \frac{\alpha_1}{\alpha_0 + 1} + \frac{\alpha_3}{\alpha_0 q^t(d)} U_b \Phi((d - \mu)/\sigma) \frac{\alpha_1}{\alpha_0 + 1},
\]

\[
m_2 = \frac{\alpha_2^{\text{new}}}{\alpha_0^{\text{new}}} = \mathbb{E}_{Q(\omega|d)}[\omega^2] = \frac{\alpha_1}{\alpha_0 q^t(d)} N(d|\mu, \xi^2 + \sigma^2) \frac{\alpha_1 + 1}{\alpha_0 + 1} + \frac{\alpha_2}{\alpha_0 q^t(d)} U_f (1 - \Phi((d - \mu)/\sigma)) \frac{\alpha_2}{\alpha_0 + 1} + \frac{\alpha_3}{\alpha_0 q^t(d)} U_b \Phi((d - \mu)/\sigma) \frac{\alpha_2}{\alpha_0 + 1},
\]

\[
m_3 = \frac{\alpha_3^{\text{new}}}{\alpha_0^{\text{new}}} = \mathbb{E}_{Q(\omega|d)}[\omega^3] = \frac{\alpha_1}{\alpha_0 q^t(d)} N(d|\mu, \xi^2 + \sigma^2) \frac{\alpha_3}{\alpha_0 + 1} + \frac{\alpha_2}{\alpha_0 q^t(d)} U_f (1 - \Phi((d - \mu)/\sigma)) \frac{\alpha_3}{\alpha_0 + 1} + \frac{\alpha_3}{\alpha_0 q^t(d)} U_b \Phi((d - \mu)/\sigma) \frac{\alpha_3 + 1}{\alpha_0 + 1}.
\]
The second moments are calculated as follows:

\[ m_1^{(2)} = \text{E}_Q(\omega^1 | d^1) = \alpha_0^{\text{new}}(\alpha_0^{\text{new}} + 1) = \frac{\alpha_1}{\alpha_0 q^2(d)} N(\mu, \xi^2 + \sigma^2) \frac{(\alpha_1 + 1)(\alpha_1 + 2)}{(\alpha_0 + 1)(\alpha_0 + 2)} + \frac{\alpha_2}{\alpha_0 q^2(d)} U_f (1 - \Phi((d - \mu)/\sigma)) \frac{\alpha_1(\alpha_1 + 1)}{(\alpha_0 + 1)(\alpha_0 + 2)}. \]

\[ m_2^{(2)} = \text{E}_Q(\omega^2 | d^2) = \frac{\alpha_0^{\text{new}}(\alpha_0^{\text{new}} + 1)}{\alpha_0^{\text{new}}(\alpha_0^{\text{new}} + 1)} = \frac{\alpha_1}{\alpha_0 q^2(d)} N(\mu, \xi^2 + \sigma^2) \frac{\alpha_2(\alpha_2 + 1)}{(\alpha_0 + 1)(\alpha_0 + 2)} + \frac{\alpha_2}{\alpha_0 q^2(d)} U_f (1 - \Phi((d - \mu)/\sigma)) \frac{\alpha_2(\alpha_2 + 1)}{(\alpha_0 + 1)(\alpha_0 + 2)}. \]

\[ m_3^{(2)} = \text{E}_Q(\omega^3 | d^3) = \frac{\alpha_0^{\text{new}}(\alpha_0^{\text{new}} + 1)}{\alpha_0^{\text{new}}(\alpha_0^{\text{new}} + 1)} = \frac{\alpha_1}{\alpha_0 q^2(d)} N(\mu, \xi^2 + \sigma^2) \frac{\alpha_3(\alpha_3 + 1)}{(\alpha_0 + 1)(\alpha_0 + 2)} + \frac{\alpha_3}{\alpha_0 q^2(d)} U_f (1 - \Phi((d - \mu)/\sigma)) \frac{\alpha_3(\alpha_3 + 1)}{(\alpha_0 + 1)(\alpha_0 + 2)}. \]

The parameters are thus estimated with the help of introduced variables as

\[ \alpha_0^\text{new} = \frac{\sum_{i=1}^3 m_i - m_i^{(2)}}{\sum_{i=1}^3 m_i^{(2)} - m_i^{(2)}}, \quad \alpha_i^\text{new} = \alpha_0^\text{new} m_i, \quad i = 1, 2, 3. \]

### 2.6 Approximated Posterior \( q^t(m_x | d_x^t) \)

Similarly, the approximated posterior with respect to each state is

- **State-I**: Fitting the static structure \( q^t(m_x = I | d_x^t) = \alpha_x^I N(d_x^t | \mu_x^I, \xi^2 + (\sigma_x^I)^2)/ (q^t(d_x^t) \sum_{k \in \Psi} \alpha_k^x) \);  
- **State-F**: Forward Outliers \( q^t(m_x = F | d_x^t) = \alpha_x^F U_f (1 - \Phi((d_x^t - \mu_x^{I-1})/\sigma_x^{I-1}))/ (q^t(d_x^t) \sum_{k \in \Psi} \alpha_k^x) \);  
- **State-B**: Backward Outliers \( q^t(m_x = B | d_x^t) = \alpha_x^B U_f ((d_x^t - \mu_x^{I-1})/\sigma_x^{I-1}))/ (q^t(d_x^t) \sum_{k \in \Psi} \alpha_k^x) \).

### 3 The Choice of Depth Noise Standard Deviation

#### 3.1 Depth Map from Stereo or Kinect

Since depth map obtained by Stereo or Kinect is actually estimated via the disparity estimation technique, in which case the conversion between depth and disparity is

\[ \frac{d^{\text{disp}}}{B} = \frac{f}{d} \Rightarrow d = \frac{f B}{d^{\text{disp}}}. \]

\( d^{\text{disp}} \) is the disparity and \( d \) is the depth. \( f \) is the focal length of the camera, \( B \) is the baseline between stereo sensors.

The noise and outliers in the depth map are originated at the errors in the disparity map. Assume the Gaussian noise and uniform outliers in the disparity map, we try to find their characteristics in the corresponding depth map. Define a universal Gaussian noise standard deviation \( \sigma_n^{\text{disp}} \) in the disparity map, it results in a noise disparity value \( d_n^{\text{disp}} \) from a mean \( \mu_n^{\text{disp}} \). Converting the noisy disparity value into the depth, we have

\[ d_n = \frac{f B}{\mu_n} = \frac{f B}{\mu_n + (d_n^{\text{disp}} - \mu_n^{\text{disp}})} = \frac{f B}{\mu_n} \frac{1}{1 + (d_n^{\text{disp}} - \mu_n^{\text{disp}})/\mu_n^{\text{disp}}} \approx \frac{f B}{\mu_n} \left(1 + \frac{d_n^{\text{disp}} - \mu_n^{\text{disp}}}{\mu_n^{\text{disp}}} \right) = 2 \mu_n - \mu_n^{\text{disp}} \frac{d_n^{\text{disp}}}{\mu_n^{\text{disp}}}. \]

Here the mean \( \mu_n = f B/\mu_n^{\text{disp}} \). It needs one constraint \(|\mu_n^{\text{disp}} - d_n^{\text{disp}}| < \mu_n^{\text{disp}} \), which can be satisfied in a general setting. Thus the mean value for \( d_n \) is \( \mathbb{E}[d_n] = 2 \mu_n - \mu_n^{\text{disp}} \frac{\mu_n}{\mu_n^{\text{disp}}} = \mu_n \), its variance is

\[ \sigma_n^2 = \mathbb{E}[(d_n - \mu_n)^2] = \frac{\mu_n}{(\mu_n)^2} \mathbb{E}[(\mu_n^{\text{disp}} - d_n^{\text{disp}})^2] = \left(\frac{\mu_n^{\text{disp}}}{\mu_n}ight)^2 \left(\frac{\mu_n}{\mu_n^{\text{disp}}} \sigma_n^{\text{disp}}\right)^2 \Rightarrow \sigma_n = \sigma_n^{\text{disp}} \frac{\mu_n}{f B}. \]

The outliers in the depth map is still modeled by uniform distribution.

Therefore, to better model the static structure estimation, we set the depth noise standard deviation \( \xi_n \propto (d_x^t)^2/f B \), which is a function of the depth sample \( d_x^t \). Samples with larger depth values will require larger standard deviations to fit its noise.
3.2 Depth Map from Other Sources

For depth map obtained by other sources, the noise standard deviation $\xi = \sigma$ is a constant over the image domain. If the property of the systematic error for a depth sensor is available, the standard deviation $\xi$ can be modeled more specifically.

References
