\( \mathcal{K} \) set of all locations
\( \mathcal{L} \) set of all pixels
\( \mathcal{Z} \) set of all possible depth values
\( X_k \) binary occupancy variable for a location \( k \)
\( Z_i \) observed depth value variable at pixel \( i \)
\( z^\infty \) special value for when no depth is observed
\( M_{ki} \) segmentation mask at pixel \( i \) for location \( k \)
\( S_k \) object silhouette for \( k \)-th location
\( |S_k| \) number of pixels in the \( k \)-th silhouette
\( \langle \cdot \rangle_p \) expectation wrt a distribution \( p \)
\( \sigma(x) \) sigmoid function \( (1 + e^{-x})^{-1} \)
\( \rho_k \) approximate posterior \( Q(X_k = 1) \)
\( \pi^\infty \) probability of observing \( z^\infty \)
\( \pi^\circ \) probability of observing an outlier
\( \Delta_{l,i} \) \( \log \theta_{l,i}(z), l \in \mathcal{K} \cup \{\text{bg}\} \)
\( \tau_{ki} \) \( \prod_{l<k, i \in M_l}(1 - \rho_l) \)

Table 1: Notations

We know that \( X_k \) is a Bernoulli variable \( (Q(X_k = 1) + Q(X_k = 0) = 1) \), and keeping in mind Eq. 7 from the paper:

\[
Q(X_k = 1) = \frac{\exp(\langle \log P(Z, X, M) | X_k = 1 \rangle)}{\tilde{Z}_{X_k}}
\]
\[
Q(X_k = 0) = \frac{\exp(\langle \log P(Z, X, H) | X_k = 0 \rangle)}{\tilde{Z}_{X_k}}
\]

which allows us to find the normalizing factor \( \tilde{Z}_{X_k} \) and get the following update on \( \rho_k = Q(X_k = 1) \) (Eq. 8 in the paper):

\[
\rho_k = \sigma(\langle \log P(Z, X, M) | X_k = \xi \rangle)_{Q(X/X_k)}, \xi \in \{0, 1\}
\]

Let’s take a closer look at the conditional expectation of the joint \( \langle \log P(Z, X, M) | X_k = \xi \rangle_{Q(X/X_k)}, \xi \in \{0, 1\} \) (omitting \( M \) for clarity):
\[ \langle \log P(Z, X \mid X_k = \xi) \rangle_{Q(X \mid X_k)} = \]
\[ \langle \log P(Z \mid X)P(X) \mid X_k = \xi \rangle_{Q(X \mid X_k)} = \]
\[ \sum_{i \in S_k} \langle \log P(Z_i \mid X) \mid X_k = \xi \rangle + \sum_{l \in K} \langle \log P(X_l) \mid X_k = \xi \rangle \]

where we used an assumption of conditional independence between pixels, and also assumption of independence of occupancies a-priori.

We now need to evaluate the expectation of the observation likelihood \( \log P(Z_i \mid X) \) conditioned on \( X_k = \xi \) for \( \xi \in \{0, 1\} \). Under our generative model, each pixel is either generated by one of the silhouettes, or by the background. If it was generated by some silhouette \( S_l \), then, first, all the silhouettes that are closer to the camera should be absent (which happens with probability \( \tau_{l-1,i} \)), and, second, the silhouette itself should be present (which happens independently with probability \( \rho_l \)). Otherwise, all the silhouettes are absent (probability \( \tau_{|K|,i} \)), and pixel was generated by the background. Now, let’s write down the expected log-likelihood for observing some value \( z \) at pixel \( i \in L \) (omitting segmentation masks for clarity):

\[ \langle \log P(Z_i = z \mid X) \rangle_{Q(X)} = \sum_{l \in K} \tau_{l-1,i} \rho_l \log \theta_{l,i}(z) + \tau_{|K|,i} \log \theta_{bg,i}(z) \]

When conditioned on \( X_k = 1 \) the expectation will be:

\[ \sum_{l < k} \tau_{l-1,i} \rho_l \log \theta_{l,i}(z) + \tau_{k-1,i} \log \theta_{k,i}(z) \]

And conditioned when on \( X_k = 0 \):

\[ \sum_{l < k} \tau_{l-1,i} \rho_l \log \theta_{l,i}(z) + \sum_{l > k} \tau_{l-1,i} \rho_l \log \theta_{l,i}(z) + \tau_{|K|,i} \log \theta_{bg,i}(z) \]

\[ (1 - \rho_k) \]

Now, one can evaluate \( \langle \log P(Z, X \mid X_k = \xi) \rangle_{Q(X \mid X_k)} \) (expectations for the priors is trivial, see e.g. Fleuret’09). If one substitutes these into Eq. 8, one will get Eq. 10:

\[ \rho_k = \sigma(\log \frac{\epsilon}{1-\epsilon} + \)
\[ \sum_{i \in M_k} \tau_{k-1,i} \log \theta_{k,i} - \]
\[ \sum_{i \in M_k} \frac{1}{1-\rho_k} (\sum_{l > k, i \in M_l} \tau_{l-1,i} \rho_l \log \theta_{l,i} + \tau_{|K|,i} \log \theta_{bg,i}) \]