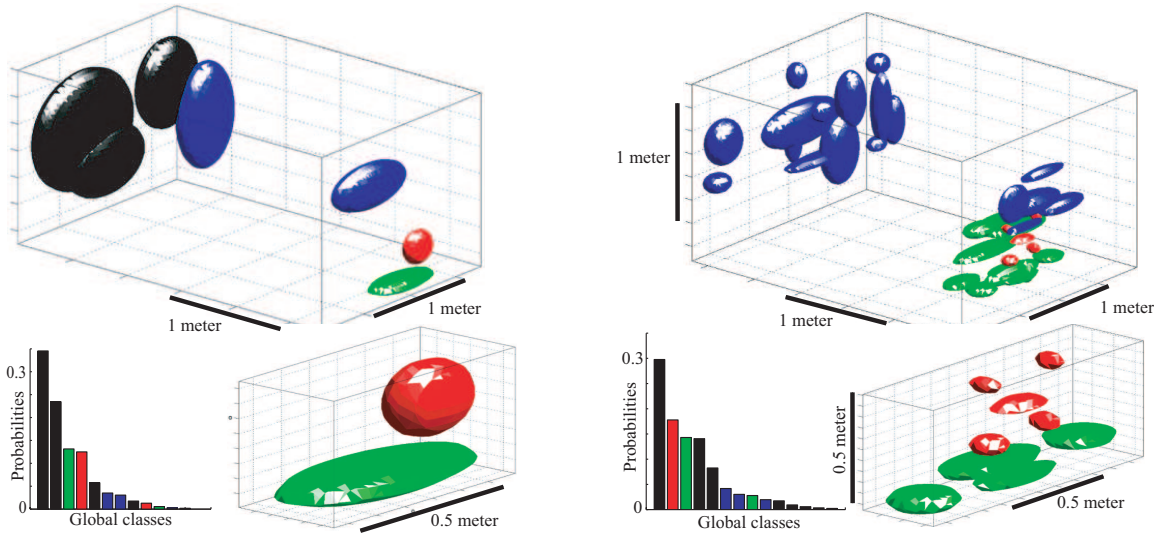

Figure 4. Visual object categories learned from stereo images of office scenes containing computer screens (red), desks (green), bookshelves (blue), and background clutter (black). Covariance ellipses model 3D part geometry, and are positioned at their mean transformed location. Bar charts show posterior probabilities for all instantiated global categories. *Left:* Single part TDP, as in Sec. 3.2. We show the seven visual categories with highest posterior probability (top), and a close–up view of the screen and desk models (bottom). *Right:* Multiple part TDP, as in Sec. 3.3. For clarity, we show the most likely parts (those generating 85% of observed features) for the five most frequent non–background categories (top). The close–up view shows a five–part screen model, and a four–part desk model (bottom).

Note that transformed parts whose mean is farther from the projection ray are given lower overall weight ω_{tk}. To evaluate the likelihood of new object instances $\bar{t}$, we integrate over potential transformations $\rho_{j\bar{t}}$, and evaluate eq. (18) with an appropriately inflated 3D covariance.

The final term of eq. (15) is the depth likelihood corresponding to stereo–based disparity matches. For monocular images, we jointly resample $(t_{ji}, k_{ji}, u^z_{ji})$ by using the prior clustering bias of eqs. (16, 17), and appearance likelihood, to reweight the Gaussian mixture of eq. (18). For stereo training images, we evaluate the likelihood learned in Sec. 2.3 on a uniformly spaced grid determined by the largest expected scene geometry. We then evaluate eq. (18) on the same grid for each candidate instance and part, and resample from that discrete distribution. Given Z depths, and T_j object instances with (on average) K parts, this resampling step requires $\mathcal{O}(ZT_jK)$ operations.

4.2. Inferring Object Categories

In the second phase of each Gibbs sampling iteration, we fix feature depths u^z and object assignments $\mathbf{t}$, and consider potential reinterpretations of each instance t using a new global object category o_{jt}. Because parts and transformations are defined with respect to particular categories, blocked resampling of $(o_{jt}, \rho_{jt}, \{k_{ji} \mid t_{ji} = t\})$ is necessary. Suppose first that $o_{jt} = \ell$ is fixed. Given ρ_{jt}, part assignments k_{ji} are conditionally independent:

$$p\left(k_{ji} = k \mid w_{ji}, u_{ji}, t_{ji} = t, o_{jt} = \ell, \mathbf{k}_{\setminus ji}, \mathbf{t}_{\setminus ji}, \mathbf{o}_{\setminus jt}\right) \propto p\left(k \mid \mathbf{k}_{\setminus ji}, \mathbf{t}, \mathbf{o}\right) \eta_{\ell k}(w_{ji}) \mathcal{N}\left(u_{ji}; \mu_{\ell k}, \Lambda_{\ell k}\right) \quad (19)$$

Here, the first term is as in eq. (17). Alternatively, given fixed part assignments ρ_{jt} has a Gaussian posterior:

$$p\left(\rho_{jt} \mid o_{jt} = \ell, \{k_{ji}, u_{ji} \mid t_{ji} = t\}\right) \propto \mathcal{N}(\rho_{jt}; \phi_\ell) \prod_{k=1}^{K_\ell} \prod_{i \mid k_{ji}=k} \mathcal{N}\left(u_{ji} - \rho_{jt}; \mu_{\ell k}, \Lambda_{\ell k}\right) \quad (20)$$

The Gaussian transformation prior $\mathcal{N}(\phi_\ell)$ is specific to the visual category (see eq. (14)), while the posterior mean and covariance follow standard equations [4, 14]. Note that our use of continuous, Gaussian position densities avoids an expensive discretization of 3D world coordinates.

For each candidate visual category o_{jt}, we first perform a small number of auxiliary Gibbs sampling iterations using eqs. (19, 20). Given the resulting transformations, the part assignments of eq. (19) may be directly marginalized to compute the likelihood of o_{jt}. The stick–breaking construction of eq. (14) also induces a clustering prior:

$$p\left(o_{jt} \mid \mathbf{o}_{\setminus jt}\right) \propto \sum_{\ell=1}^{L} M_\ell^{-t} \delta(o_{jt}, \ell) + \gamma \delta(o_{jt}, \bar{\ell}) \quad (21)$$

Here, M_ℓ^{-t} denotes the number of object *instances* assigned to the L current categories, and $\bar{\ell}$ indicates a new visual category. Combining these terms, we resample o_{jt}, and conditionally choose $(\rho_{jt}, \{k_{ji} \mid t_{ji} = t\})$ via eqs. (19, 20).

4.3. Inferring Part and Transformation Parameters

The preceding sections assumed fixed values for the parameters $\theta_{\ell k} = (\eta_{\ell k}, \mu_{\ell k}, \Lambda_{\ell k})$ defining part appearance and position, as well as category–specific transformation