the identity transformation, because on average we expect mesh faces to undergo small deformations. For the noise covariance prior, we set the degrees of freedom $n_0 = 5$, a value which makes the prior variance nearly as large as possible while ensuring that the mean remains finite. The expected part variance S_0 captures the degree of non-rigidity which we expect parts to demonstrate, as well as noise from the mesh alignment process. The correspondence error in our human meshes is approximately 0.01m; allowing for some part non-rigidity, we set $\sigma = 0.015$m and $S_0 = \sigma^2 \times \mathbf{I}_{3\times3}$. K is a precision matrix set to $K = \sigma^2 \times \text{diag}(1, 1, 1, 0.1)$.The Kronecker product of K^{-1} and S_0 governs the covariance of the distribution on A. Our settings make this nearly identity for most components, but the translation components of A have variance which is an order of magnitude larger, so that the expected scale of the translation parameters matches that of the mesh coordinates.

In our experiments, we ran the mesh-ddcrp sampler for 200 iterations from each of five random initializations, and selected the most probable posterior sample. The computational cost of a Gibbs iteration scales linearly with the number of meshes; our unoptimized Matlab implementation required around 10 hours to analyze 56 human meshes.

4.2 Baseline Segmentation Methods

We compare the mesh-ddcrp model to three competing methods. The first is a modified agglomerative clustering technique [16] which enforces spatial contiguity of the faces within each part. At initialization, each face is deemed to be its own part. Adjacent parts on the mesh are then merged based on the squared error in describing their motion by affine transformations. Only adjacent parts are considered in these merge steps, so that parts remain spatially connected.

Our second baseline is based on a publicly available implementation of spectral clustering methods [17], a popular approach which has been previously used for mesh segmentation [18]. We compare to an affinity matrix specifically designed to cluster faces with similar motions [19]. The affinity between two mesh faces u, v is defined as $C_{uv} = \exp\{-\frac{\sigma_{uv}+\sqrt{m_{uv}}}{S^2}\}$, where $m_{uv} = \frac{1}{J^2}\sum_j \delta_{uvj}$, δ_{uvj} is the Euclidean distance between u and v in pose j, $\sigma_{uv} = \sqrt{\frac{1}{J}\sum_j(\delta_{uvj} - \bar{\delta}_{uv})^2}$ is the corresponding standard deviation, and $S = \frac{1}{M}\sum_{u,v}\sigma_{uv} + \sqrt{m_{uv}}$ for all M pairs of faces u, v.

For the agglomerative and spectral clustering approaches, the number of parts must be externally specified; we experimented with $K = 5, 10, 15, 20, 25, 30$ parts. We also consider a Bayesian nonparametric baseline which replaces the ddCRP prior over mesh partitions with a standard CRP prior. The resulting *mesh-crp* model may estimate the number of parts, but doesn't model mesh structure or enforce part contiguity. The expected number of parts under the CRP prior is roughly $\alpha \log N$; we set $\alpha = 2$ so that the expected number of mesh-crp parts is similar to the number of parts discovered by the mesh-ddcrp. To exploit bilateral symmetry, for all methods we only segment the right half of each mesh. The resulting segmentation is then reflected onto the left half.

4.3 Part Discovery and Motion Prediction

We first consider the synthetic Tosca dataset [20], and separately analyze the Centaur (six poses) and Horse (eight poses) meshes. These meshes contain about 31,000 and 38,000 triangular faces, respectively. Figure 3 displays the segmentations of the Tosca meshes inferred by mesh-ddcrp. The inferred parts largely correspond to groups of mesh faces which undergo similar transformations.

Figure 4 displays the results produced by the ddCRP, as well as our baseline methods, on the human mesh data. Qualitatively, the segmentations produced by mesh-ddcrp correspond to our intuitions about the body. Note that in addition to capturing the head and limbs, the segmentation successfully segregates distinctly moving small regions such as knees, elbows, shoulders, biceps, and triceps. In all, the mesh-ddcrp detects 20 distinctly moving parts for one half of the body.

We now introduce a quantitative measure of segmentation quality: segmentations are evaluated by their ability to explain the articulations of test meshes with novel shapes and poses. Given a collection of T test meshes Y_t with corresponding reference meshes X_{b_t}, and a candidate segmentation into K parts, we compute

$$\mathcal{E} = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K} ||Y_{tk} - A^*_{tk} X_{b_t k}||_2. \tag{12}$$