Video Frame naive-hddCRP hddCRP

Figure 5: Video segments discovered by the naive-hddCRP and hddCRP. *Left to right:* A representative video frame, segments discovered by the naive-hddCRP and the corresponding segments discovered by hddCRP. Dashed red horizontal lines indicate different segments, and numbers indicate video frame numbers.

2013), and against the hCRP (Teh et al., 2006). For a controlled comparison, all three CRP models use identical likelihoods and hyperparameters. We use the MIT human annotated video dataset (Liu et al., 2008), which contains 9 human annotated videos, to quantitatively measure segmentation performance. We benchmark performance using the first 10 frames of each sequence.

Figure 6 summarizes this experiment. For HGVS the displayed segmentations were produced at 90 percent of the highest hierarchy level, which appears to produce the best visual and quantitative results. For the hddCRP variants, the segmentations correspond to the MAP sample of five MCMC chains, each run for 400 iterations[3]. We decided to run the samplers for 400 iterations based on the results shown in Figure 4, where a large majority of the pseudo-Gibbs chains converged within the first 300 iterations.

The Rand index was computed by treating the entire video sequence as one spatio-temporal block. This penalizes spatially coherent, but temporally inaccurate, segmentations that exhibit frequent "label switching" between frames. HGVS operates on pixels rather than superpixels and consequently produces finer-scale segmentations. However, these segmentations exhibit large segmentation errors (for instance, the neck and face regions get merged with the background in the second sequence). The hddCRP produces more coherent segmentations and in terms of Rand index, outperforms HGVS on all but one video sequence. The hddCRP also performs substantially better than the hCRP which ignores both superpixel and segment-level correlations; "bag of feature" assumptions are insufficient for this task. The gains over the naive-hddCRP appear to be more modest. However, a closer inspection (Figure 5) reveals that the hddCRP segments are visually cleaner and more coherent. Additionally, naive-hddCRP often falsely merges visually similar but distinctly moving objects together, while the hddCRP recognizes them as distinct segments. The videos in our dataset have large background regions with no significant motion. Both the hddCRP and the naive-hddCRP models tend to agree on such regions, while disagreeing on smaller foreground objects with distinct motions. Large regions dominate the Rand index, which explains the similar global performance by that metric.

[3]Roughly 6 hours on a 2.3 GHz intel core i7.

4.2 DISCOURSE SEGMENTATION

Next we consider the problem of discourse segmentation. Given a collection of documents, the goal is to partition each document into a sequence of topically coherent non-overlapping discourse fragments. Previous work by Riedl & Biemann (2012) found that sharing information across documents tends to produce better segmentations, motivating the development of several text segmentation algorithms that exploit document relationships.

We conducted experiments on the *wikielements* dataset (Chen et al., 2009), which consists of 118 Wikipedia articles (at paragraph resolution) describing chemical elements. Although not explicitly made available in the dataset, each article corresponds to a chemical element that is characterized by its chemical properties and has a unique location in the periodic table. Our distance-dependent models are capable of exploiting this additional information to produce better discourse segmentations. As an illustration, consider the alternative problem of clustering articles. Figure 7 illustrates such a clustering where we leverage element properties by defining distances between documents as the Manhattan distance between corresponding element locations in the periodic table. The discovered clustering corresponds well with known element groupings. Discourse segmentation requires clustering the paragraphs describing documents, instead of the documents themselves. Nonetheless, we find that exploiting the periodic table location of each document's element leads to noticeable performance gains.

We compare two versions of the hddCRP to the naive-hddCRP and hCRP. To encourage topic contiguity, naive-hddCCRP and hddCRP allowed paragraphs to either link to themselves or to other paragraphs immediately preceding or succeeding them. We experimented with two affinity functions to capture the intuitions that similar documents tend to present similar topics in similar orders, and that clusters are more likely to be shared among articles about similar elements. The first function (hddCRP1) biased clusters of paragraphs to connect to those that occur at similar locations within other documents. Further, clusters were constrained to connect to only those that were contained in documents with lower atomic numbers. A second variant (hddCRP2) modeled distances between articles using the