We show that across six data sets, performance measures and for a broad range segmentation scale from fine to coarse, the performance of ISCRA is superior to that of current state-of-the-art methods. To our knowledge, it is by far the most comprehensive evaluation of the leading approaches to superpixel extraction, and in general to image segmentation. Some typical examples for ISCRA compared to leading other methods are shown in Figure 1.

2. Background

There is a rich body of work on edge or boundary detection in computer vision, with the state of the art represented by the gPb boundary detector and its variants [2]. However, a boundary map may not correspond to a valid segmentation, since it may not provide closed contours. Below we review related work that produces explicit partition of image into regions. Note that we leave out of our discussion here the extensive recent literature on "semantic segmentation", i.e., category-level labeling of every pixel in the image. In this paper we are only concerned with methods that are agnostic about any category-level reasoning.

For the purpose of our discussion, we will define two segmentation regimes. The *superpixel regime* corresponds to more than 50 segments per image. In this regime the purpose of oversegmentation is mostly to reduce complexity of representation, without sacrificing future segmentation accuracy. Therefore, the natural notion of *scale* for this regime is the number of segments k; typical values of k are in the hundreds for a moderately sized image.

Superpixels, introduced in [17], have become a staple in vision research. Broadly, methods that produce superpixels can be grouped into graph-based [18, 6, 14, 22], clustering of pixels such as SLIC [1] and MeanShift [4], and curve evolution such as Turbopixels [13] and SEEDS[21].

In contrast, the *large segment* regime produces fewer than 100 segments. The appropriate number of segments in this regime depends on the content of the image, and specifying k is not natural. Instead, the precise meaning of scale and the way it is controled varies between segmentation methods, as described below.

In OWT-UCM [2] the oriented watershed transform on gPb boundaries is followed by greedy merging of regions, resulting in weighted boundary map such that thresholding it at any level produces a valid segmentation; the value of the threshold controls the scale. Throughout the merging process, OWT-UCM uses the same set of weights on various features throughout the process, and thus despite the greedy iterative nature of the merging, this is in a sense a single stage process. In contrast, ISCRA uses a large cascade of stages, with weightes learned per stage. We show in Section 5 that this leads to performance better than that of OWT-UCM.

The agglomerative merging segmentation algorithm in [11], like ISCRA, starts with a fine oversegmentation, learns a boundary probability model, applies it to merge regions until the estimated probability of merging is below a threshold. The classifier is then retrained, and applied again; their implementation includes four such stages, therefore defining four segmentation scales. There is a number of differences, however: while in ISCRA we use asymmetric loss and a universal threshold of $\frac{1}{2}$, in [11] the loss is symmetric, but the threshold is tuned in ad-hoc fashion. Consequently, in ISCRA we are able to learn many more stages (60 vs. four), producing a more gradual and accurate merging process. We show empirically that this contributes significantly to performance.

Higher Order Correlation Clustering (HOCC) [12] also starts with fine segmentation, but instead of a greedy merging applies a "single-shot" partition over the superpixel graph. The scale in HOCC is controled by specifying explicitly the number of regions, which may be a disadvantage when the user would like a more adaptive scale definition like in OWT-UCM or in ISCRA.

Finally, a recently proposed method called SCALPEL [24] shares many similarities with our work. In SCALPEL, a region is "grown" by applying a cascade of greedy merging steps to an initial over-segmentation. Similarly to ISCRA, the order of merging in each step is determined by learning weights that reflect importance of features at different scales. However, SCALPEL relies heavily on class-specific shape priors for known object classes, and its objective is to produce a set of class-specific region proposals, that can be used in semantic image labeling. This is in contrast to our work, which is agnostic with respect to categorization of regions. Thus the two methods, while sharing many of the ideas, are not directly comparable.

An important question about any segmentation algorithm is how it handles multiple scales. Specifically, it is often desirable to produce a *hierarchy*, in which regions obtained at a finer scale are necessarily subregions of the regions at a coarser scale. This is the case with OWT-UCM, the agglomeration algorithm of [11], and with ISCRA, but generally not with the graph-based algorithms like [12, 18]. This is also not the case with superpixel algorithms like SLIC, ERS, SEEDS or Turbopixels.

Although the definitions above of the two regimes are somewhat arbitrary, we adopt them for convenience. Still, it seems clear that the objectives in partitioning an image into 1000 segments vs. just ten are different, and it dictates different evaluation protocol for these cases. Some algorithms, including ISCRA, are competitive in both regimes, as demonstrated in Section 5.