Measuring Human Probabilistic Segmentation Maps

Jonathan Vacher[1] (jonathan.vacher@einsteinmed.org), Pascal Mamassian[2], Ruben Coen-Cagli[1]
[1] Albert Einstein College of Medicine, New-York, US; [2] École Normal Supérieure, Paris, France.

Summary Visual segmentation is a core function of biological vision, key to adaptive behavior in complex environments. Foundational work identified Gestalt principles of segmentation, *e.g.* grouping by similarity, proximity and good continuation [1], and revealed that visual cortical neurons are sensitive to those cues [2]. Early models [3] inspired by the feedforward cortical architecture described texture-based human segmentation as the process of comparing the summary statistics of low-level visual features across space. Indeed the summary statistics representation is the most prominent model of naturalistic texture perception [4], yet it has been challenged precisely because it does not fully capture the influence of segmentation [5].

Here we consider the alternative view that, due to image ambiguity and sensory noise, perceptual segmentation requires probabilistic inference. This view is consistent with reports that humans combine multiple segmentation cues near-optimally in artificial displays [6, 7], and that Gestalt laws reflect optimization to natural image statistics [8, 9, 10]. The probabilistic approach is also widespread in computer vision algorithms for unsupervised segmentation, but has not been used to model perceptual segmentation.

We present new experiments that for the first time measure perceptual segmentation maps and their variability, allowing us to test the probabilistic inference hypothesis, and compare it quantitatively to summary statistics models. We use composite textures, with segments characterized by different statistical relations between features. Optimal probabilistic inference assigns pixels to segments by evaluating which of those relations better explains the observed features (generative model), as opposed to comparing summary statistics at different locations (feature discrimination). We find the generative model best captures our data, and perceptual variability reflects image uncertainty beyond sensory noise. We also demonstrate the approach on natural images, which will allow testing more sophisticated segmentation algorithms.

Our results provide a normative explanation of human perceptual segmentation as probabilistic inference, and demonstrate a novel framework to study perceptual segmentation of natural images.

Experimental Methods Participants were presented with images on a calibrated VPiXX monitor at 53 cm viewing distance. To ensure that all participants performed the same task, prior to the start of a session they were instructed that the image needed to be decomposed into a given number of segments K. A trial starts with presentation of one image (size 8×8 deg) at the center of the screen for 3 s, followed by M sequences (Fig. 1a) in which participants are asked to report with a button press if a pair of cued locations of the image belongs to the same or different segments. The collected data allow to recover segment probabilistic maps and the segmentation map as shown in figure 1b (see the method below).

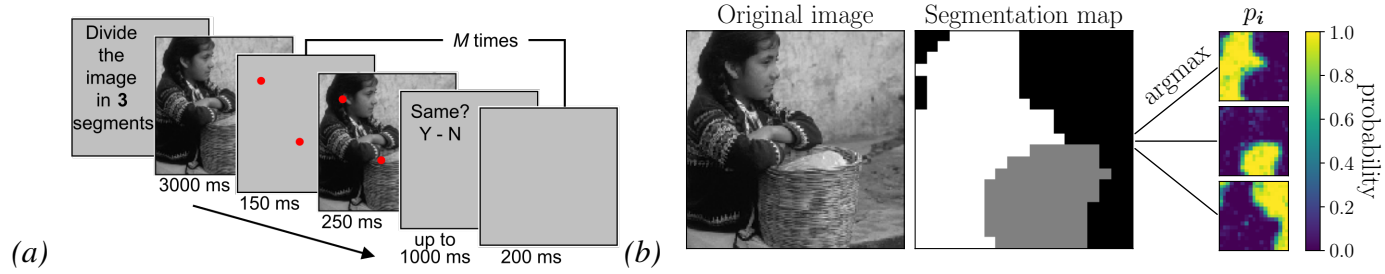

Figure 1: (a) Experiment trial layout. (b) Segmentation map of a natural image and probability of assignment to each segment, obtained with our protocol.

Segmentation map reconstruction and models Each pair $(\mathbf{i},\mathbf{j}) \in \mathscr{P}$ of tested locations lie on a $N \times N$ grid and are tested N_t times. Stimuli are assumed to have K segments. The N_t participant responses for each pair $r_{\mathbf{i},\mathbf{j}}^{(n_t)}$ are independent samples from a Bernoulli distribution with parameter $p_{\mathbf{i},\mathbf{j}}(\Theta)$ that depends on some model parameters Θ. To recover the models parameters we solve the following regularized non-linear least square regression

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{(\mathbf{i},\mathbf{j})\in\mathscr{P}} \|k_{\mathbf{i},\mathbf{j}} - p_{\mathbf{i},\mathbf{j}}(\Theta)\|^2 + \lambda R_1(\Theta) + \mu R_2(\Theta) \quad (1)$$

where $k_{\mathbf{i},\mathbf{j}}$ is the responses average, R_1 and R_2 are regularizing functions (ℓ^2-norm and ℓ^2-norm of discrete gradient) that reduce over-fitting in models *(ii)* and *(iii)* described below. When $\lambda = \mu = 0$, the optimization problem (1) is equivalent to maximum likelihood optimization. We consider three models:

(i) a non-parametric model that only assumes the existence of underlying probability maps for the segments

$$p_{\mathbf{i},\mathbf{j}}(\Theta) = \alpha + (1-2\alpha)\langle p_{\mathbf{i}}, p_{\mathbf{j}}\rangle \quad \text{with} \quad \Theta = ((p_{\mathbf{i}})_{\mathbf{i}}, \alpha), \quad \text{(NP)}$$

where $\langle , \rangle$ denotes the dot product, $p_{\mathbf{i}k}$ is probability that pixel $\mathbf{i}$ belong to segment k with $\sum_k p_{\mathbf{i}k} = 1$ (see an example of probability maps on the right of figure 1b) and α is the lapse rate.

(ii) a generative model that further assumes the probability maps are obtained via probabilistic inference from noisy observations

$$p_{\mathbf{i},\mathbf{j}}(\Theta) = \alpha + (1-2\alpha)\langle p(x_{\mathbf{i}}|\Lambda), p(x_{\mathbf{j}}|\Lambda)\rangle \quad \text{(GM)}$$

with $\Theta = (\Lambda, \alpha)$, where $(x_{\mathbf{i}})_{\mathbf{i}}$ are oriented wavelet feature vectors and $p_k(x|\Lambda) = \frac{|\Lambda_k|\exp\left(-0.5x^{\mathrm{T}}\Lambda_k x\right)}{\sum_i |\Lambda_i|\exp\left(-0.5x^{\mathrm{T}}\Lambda_i x\right)}$ with $\Lambda_k = (\Sigma_k + \sigma_0 \mathrm{I})^{-1}$. The matrix Σ_k is the feature covariance (assumed diagonal in our setting) and σ_0 is the feature noise.

(iii) a feature discrimination model that does not require any underlying probability maps, and only assumes that local features are directly compared

$$p_{\mathbf{i},\mathbf{j}}(\Theta) = \alpha + (1-2\alpha) S_{\sigma,\mu}\left(\cos_W(x_{\mathbf{i}}, x_{\mathbf{j}})\right) \quad \text{(FD)}$$

with $\Theta = (W, \mu, \sigma, \alpha)$, where $S_{\sigma,\mu}(u) = \Phi\left(\left(\log\left(\frac{u}{1-u}\right) - \mu\right)/\sigma\right)$ with Φ being the sigmoid function, (μ, σ) the point of subjective equality and the threshold; and where $\cos_W$ is a cosine similarity index between the weighted component product of feature vectors $x_{\mathbf{i}}$ and $x_{\mathbf{j}}$.

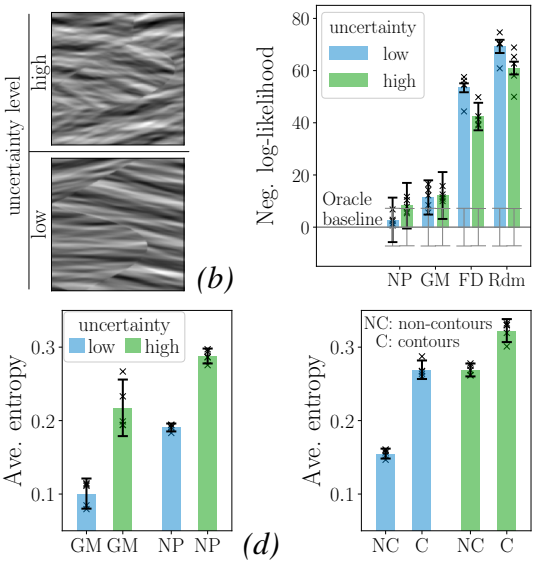

Figure 2: (a) High and low uncertainty stimuli. (b) Fit quality (cross-val. negative log-likelihood, lower is better). Rdm: chance level. (c) Average entropy. (d) Average entropy for contour and non-contour areas obtained with NP model. Error bars: 99.7% conf. interval.

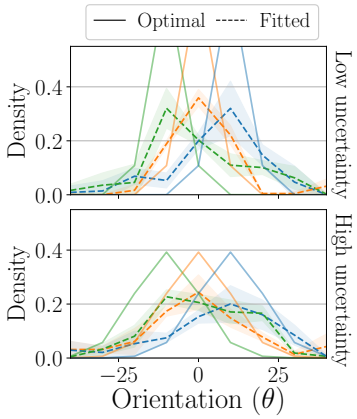

Figure 3: Ground truth covariances from the stimuli (optimal) and the fitted covariances of GM.

Results We fitted the three models on human data (4 participants) with low- and high uncertainty stimuli (Fig. 2a) and measured fit quality using the cross-validated negative log-likelihood. Figure 2b shows that probabilistic inference based on the internal model allowing for recognition and classification (GM) explains the data significantly better than a direct feature comparison allowing for discrimination (FD). We then tested whether the variability of human segmentation correlates with image uncertainty, as expected in the probabilistic framework. We quantified the total measured variability by the entropy of the segment probability maps recovered from NP, and the portion of measured variability that is due to the intrinsic image uncertainty using the maps recovered from GM. The difference between the two accounts for other factors including measurement noise and differences between participants. First, figure 2c shows that variability was significantly larger for more uncertain stimuli (NP), and the increase was largely explained by the increase in image uncertainty (GM). Second, the spatial distribution of variability was not uniform, but rather concentrated around contours (Fig. 2d), and the effect was stronger for low uncertainty stimuli (blue) where contours are more spatially localized.

Lastly, to quantify how closely the subjects approximated the optimal observer, we compared the fitted covariances $\hat{\Sigma}_k$ (*i.e.* the internal representation of the average participant) to the ground truth covariances of the stimuli (*i.e.* the optimal observer). The participants covariances were narrower for low-uncertainty than for high-uncertainty stimuli, and qualitatively followed the ground truth (Fig. 3) despite being broader overall.

Our work offers strong evidence that human segmentation is probabilistic, and demonstrates a new protocol that will allow studying natural image segmentation.

References **[1]** J. Wagemans et al. *Psychological bulletin* (2012). **[2]** P. R. Roelfsema. *Annu. Rev. Neurosci.* (2006). **[3]** M. S. Landy and J. R. Bergen. *Vision research* (1991). **[4]** J. Freeman and E. P. Simoncelli. *Nature neuroscience* (2011). **[5]** T. S. Wallis et al. *eLife* (2019). **[6]** M. S. Landy and H. Kojima. *JOSA A* (2001). **[7]** P. M. Claessens and J. Wagemans. *Journal of Vision* (2008). **[8]** W. S. Geisler et al. *Vision research* (2001). **[9]** M. Sigman et al. *Proceedings of the National Academy of Sciences* (2001). **[10]** J. H. Elder and R. M. Goldberg. *Journal of Vision* (2002).