Appendix (of the paper "Similarity by Composition")

I. Proof of Claim1 in Section 2.3 of the paper

Claim 1. Upper and lower bounds on GES:

$$\max_S \{\, logP(S|H_{ref}) \;+ \sum_{R_i \in S} LES(R_i|H_{ref}) \,\} \;\leq\; GES(Q|H_{ref}) \;\leq\; \sum_{q \in Q} PES(q|H_{ref})$$

Proof.

Lower Bound:
The lower bound is immediate:

$$GES(Q|H_{ref}) = \log \frac{P(Q|H_{ref})}{P(Q|H_0)} = \log \sum_S \frac{P(Q|S,H_{ref})P(S|H_{ref})}{P(Q|H_0)}$$

$$\geq \log \max_S \frac{P(Q|S,H_{ref})P(S|H_{ref})}{P(Q|H_0)} = \max_S \{\, logP(S|H_{ref}) + \sum_{R_i \in S} LES(R_i|H_{ref}) \,\}$$

Upper Bound:

$$GES(Q|H_{ref}) = \log \frac{P(Q|H_{ref})}{P(Q|H_0)} = \log \sum_S \frac{P(Q|S,H_{ref})P(S|H_{ref})}{P(Q|H_0)} \leq \log \max_S \frac{P(Q|S,H_{ref})}{P(Q|H_0)}$$

The last inequality is valid because the maximal element is always higher than the average one (average weighted by $P(S|H_{ref})$). Swapping log and max in the last expression we get

$$GES(Q|H_{ref}) \leq \max_S \log \frac{P(Q|S,H_{ref})}{P(Q|H_0)} = \max_S GES(Q|H_{ref},S) = \max_{S=R_1,..,R_k} \sum_{i=1}^{k} LES(R_i|H_{ref})$$

For every non-overlapping regions $R_1, ..., R_k$:

$$\sum_{i=1}^{k} LES(R_i|H_{ref}) = \sum_{i=1}^{k} \sum_{q \in R_i} \frac{LES(R_i|H_{ref})}{|R_i|} = \sum_{q \in Q} \frac{LES(R^q|H_{ref})}{|R^q|}$$

where R^q is the region in S which contains q (this is not necessarily the maximal region $R_{[q]}$ of q). Note that there is only one such region because the regions in S are disjoint. From the definition of $PES(q|H_{ref})$ as the maximal saving per point and $R_{[q]}$ as the region obtaining this saving we get:

$$\sum_{q \in Q} \frac{LES(R^q|H_{ref})}{|R^q|} \leq \sum_{q \in Q} \frac{LES(R_{[q]}|H_{ref})}{|R_{[q]}|} = \sum_{q \in Q} PES(q|H_{ref})$$

This applies for every segmentation (including the maximal segmentation). From the last three equations we get the upper bound on GES. □

II. Estimating the segmentation length $logP(S|H_{ref})$:

Computing the lower bound on GES requires estimation of $logP(S|H_{ref})$, which is also $-length(S|H_{ref})$. In our implementation, we assume that the "description length" of the segmentation $length(S) = \sum_{i=1}^{k} length(s_i) + const$, where s_i is the shape of the region R_i (including its position in Q), and $const$ is a constant overhead needed for specifying the number of regions in a segmentation S. For example, in images we computed $length(s_i)$ as the length of the chain code required to describe the perimeter of the region (plus the position in Q). In video it was the surface area of the region. Alternatively, we can estimate $length(s_i) = -log(P(s_i|H_{ref}))$ according to any given prior on shapes. To bound $const$, we assume that the number of regions in a segmentation is bounded by N (e.g., 1000). Thus, $const \leq logN$.