Model-driven Sketch Reconstruction with Structure-oriented Retrieval - Supplemental Material

Lei Li[1] Zhe Huang[2] Changqing Zou[3,4*] Chiew-Lan Tai[1] Rynson W.H. Lau[2] Hao Zhang[3] Ping Tan[3] Hongbo Fu[2]
[1]Hong Kong UST [2]City University of Hong Kong [3]Simon Fraser University [4]Hengyang Normal University

In this supplemental material, we provide more details of our structure-oriented shape retrieval and sketch reconstruction methods.

1 Shape Retrieval

1.1 Pairwise Cost

In Section 3.1 of the main text, we defined a pairwise cost between two parts v_g^i and v_m^j. The precise definition of the cost is:

$$E(v_g^i, v_m^j) = w_d E_d(\hat{v}_g^i, \hat{v}_m^j) + w_s E_s(\hat{v}_g^i, \hat{v}_m^j), \tag{1}$$

where E_d is the distance cost, E_s is the similarity cost, w_d and w_s are the weights, and $\hat{v}_g^i$, $\hat{v}_m^j$ are the best-fit ellipses (obtained via Principal Component Analysis) of v_g^i and v_m^j. As mentioned in the main text, we compare the best-fit ellipses instead of the original shapes since we would like to focus on comparing the positions and sizes of the semantic parts.

E_d is defined as the Euclidean distance between the centers of $\hat{v}_g^i$ and $\hat{v}_m^j$, and E_s is defined as the average distance between the elliptic contours of $\hat{v}_g^i$ and $\hat{v}_m^j$ after aligning their centers. Precisely, we have

$$E_s(\hat{v}_g^i, \hat{v}_m^j) = \sum_{\mathbf{p}^k \in \hat{v}_m^j} D(\hat{v}_g^i, \mathbf{p}^k)/N, \tag{2}$$

where $\mathbf{p}^k$ is a point on the contour of $\hat{v}_m^j$ and N is the total number of such points. $D(\hat{v}_g^i, \mathbf{p}^k)$ is the Euclidean distance from $\mathbf{p}^k$ to the closest point in $\hat{v}_g^i$. One could have defined the dissimilarity of the ellipses $\hat{v}_g^i$ and $\hat{v}_m^j$ by comparing their semimajor axes, semiminor axes and tilts, but this will lead to the tricky problem of weighting the three factors, which we would like to avoid.

To balance the effects of the two cost terms, the weights are computed as the reciprocal of the average cost of the respective terms. That is, after E_d and E_s are computed for all pairs of (v_g^i, v_m^j), we have $w_d = 1/\overline{E_d}$ and $w_s = 1/\overline{E_s}$, where $\overline{E_d}$ and $\overline{E_s}$ are the mean values of E_d and E_s.

1.2 Part Merging

In Section 3.1 of the main text, we presented a part merging cost for structure comparison:

$$\begin{aligned} E_g(v_g^i, V_m^i) =& U(v_g^i, V_m^i) \sum ||p(v_m^a) - p(v_m^b)||_2, \\ & \forall (v_m^a, v_m^b) \in MST(V_m^i), \end{aligned} \tag{3}$$

where the binary indicator function $U(v_g^i, V_m^i)$ indicates whether v_g^i is *dissimilar* to the elements in V_m^i combined (as the union of all the constituent strokes). It is computed by measuring the dissimilarity of v_g^i an V_m^i, where they are first represented by their best fit ellipses $\hat{v}_g^i$ and $\hat{V}_m^i$ respectively, and then compared using the following function, with their centers aligned beforehand:

$$D(\hat{v}_g^i, \hat{V}_m^i) = 1 - \frac{area(\hat{v}_g^i \cap \hat{V}_m^i)}{area(\hat{v}_g^i \cup \hat{V}_m^i)}, \tag{4}$$

where $\hat{v}_g^i \cap \hat{V}_m^i$ and $\hat{v}_g^i \cup \hat{V}_m^i$ are the intersection and the union of the ellipses $\hat{v}_g^i$ and $\hat{V}_m^i$. In our implementation, the ellipses are approximated by polygons, so polygon boolean operations are employed here. Note that although Eq. (2) can also be used to compare ellipses, it is not adopted in Eq. (4) since it computes an unnormalized distance, while Eq. (4) outputs a value between 0 and 1, which makes it easy to define $U(v_g^i, V_m^i)$ by a threshold value, independent of the sizes of the ellipses:

$$U(v_g^i, V_m^i) = \begin{cases} 1, & D(\hat{v}_g^i, \hat{V}_m^i) > 0.5; \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

If multiple parts in the sketch are mapped to a single part in the model, that is, $V_g^i = \{v_g^1, v_g^2, ..v_g^k\} \subset V_g$ mapping to $v_m^i \in V_m$, the cost is computed a bit differently. Specifically, $U(v_m^i, V_g^i)$ is always 1, meaning the cost is always applied, even if V_g^i and v_m^i might be similar, since we assume the segmentation of the database model is generally no coarser than the sketch segmentation. It is rarely necessary to merge sketch parts, as such, the merging is always penalized.

2 Sketch Reconstruction

In the following, we give the precise definitions of the cost terms E_{pr}, E_{pl} and E_{pj} used in the optimization process in Section 4.2 of the main text.

Proximity cost E_{pr}. We define E_{pr} as

$$E_{pr} = \frac{1}{3|\Gamma|} \sum_{(\mathbf{u},\mathbf{w}) \in \Gamma} \|\overline{\mathbf{u}} - \overline{\mathbf{w}}\|^2, \tag{6}$$

where $\overline{\mathbf{u}}$ and $\overline{\mathbf{w}}$ correspond to the unknown 3D positions of $\mathbf{u}$ and $\mathbf{w}$ respectively in $\overline{Q}^i$s.

Plane cost E_{pl}. We further divide the parts in Δ_1 into two subclasses based on the types of the selected planes: Parts in Δ_{1a} have planes with normals respecting $\mathbf{v}_m$, that is, the angular difference between the plane normal and $\mathbf{v}_m$ is less than a threshold value, while parts in Δ_{1b} have planes with normals disrespecting $\mathbf{v}_m$. We require $|\Delta_{1a}| > 0$ since the initial point positions, through back-projection, of the parts in Δ_{1a} are more reliable and we would like to treat them as anchoring parts to guide the deformations of other sketch parts. We represent the candidate plane as $\overline{p}^i = (\mathbf{n}^i, d^i)$, where $\mathbf{n}^i$ is the normal of the plane and d^i is the distance to the origin. If $\Delta^i \in \Delta_{1b}$, we treat d^i as an unknown variable to allow translations from the original position. The plane cost is given by

$$E_{pl} = \frac{1}{N_{pl}} \sum_{\Delta^i \in \Delta_1} \sum_{\overline{\mathbf{u}} \in \overline{Q}^i} \|\overline{\mathbf{u}} \cdot \mathbf{n}^i - d^i\|^2, \tag{7}$$

*Corresponding author: Changqing Zou (aaronzou1125@gmail.com)

where $N_{pl} = \sum_{\Delta^i \in \Delta_1} |\overline{Q}^i|$.

Projection cost E_{pj}. Under the orthographic view, we represent the projection cost as

$$E_{pj} = \frac{1}{N_{pj}} \sum_i \sum_{\overline{\mathbf{u}} \in \widetilde{v}_g^i} \|\Pi(\overline{\mathbf{u}}) - \Pi(\mathbf{u})\|^2, \tag{8}$$

where $\Pi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $N_{pj} = 2\sum_i |\widetilde{v}_g^i|$ and point $\overline{\mathbf{u}} \in \widetilde{v}_g^i$ corresponds to $\mathbf{u} \in v_g^i$. Note that due to the user's imprecise drawing, where the views of different parts may not be consistent, we do not assign a high weight to this term and only treat it as a regularization term.