Class-specific 3D Localization using Constellations of Object Parts

Mukta Prasad[1]
mprasad@ee.ethz.ch
Jan Knopp[2]
jan.knopp@esat.kuleuven.be
Luc Van Gool[1]
vangool@vision.ee.ethz.ch

[1] ETH Zürich,
Zürich, Switzerland
[2] Katholic University of Leuven,
Leuven, Belgium

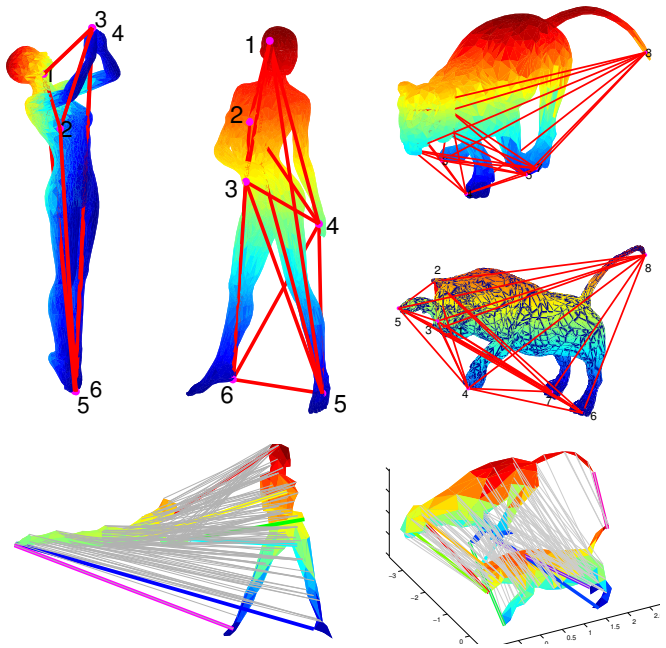

Figure 1: **First look:** Fully connected graphs of object part nodes are learnt and inferred for 2 models each of "Victoria" and "Lion" (TOSCA). Subsequently, correspondences can be estimated between the 3D models (coloured arbitrarily for 3D visualization).

We address the problem of learning class-specific, deformable, 3D part-based structure for object localization, on fully connected (FC) part-based graphs, along the lines of Pictorial Structures (PS) [1], Constellation Models [2] and ISM [3]. Further the above object part localization is used to find dense correspondences between class 3D models. Our results show improvement over state of the art and promise for application in more complex 3D processing tasks such as part retrieval, pose estimation, scene understanding and recognition.

Problem statement: We have a set of class-specific 3D shape instances $S_m, m \in \{1 \cdots M\}$, denoted as a set of V_m vertices v_m and edges ε_m. The appearance of S_m at the v^{th} vertex is given by the Heat Kernel Signature descriptor $\mathbf{s}_{mv}$. If each S_m has P parts located at L_{mp}, the aim is to learn the part configuration as a graphical model (governed by parameter set Θ), in an inference framework. The object parts, governed by appearance parameters $\alpha = \{\alpha_p\}$, $p \in \{1 \ldots P\}$, define the nodes of a graph $G \subset F$, (F: fully connected graph over P nodes), whose edge parameters are $\gamma = \{\gamma_{ij} | G_{ij} \in G\}$. $\Theta = \{G, \alpha, \gamma\}$ is ideally learnt from the training shapes and configurations in a Maximum-Likelihood framework. For query scene S^*, we can find the MAP estimate of the object part layout $L^* = \text{argmax}_L\, p(L|S^*, \Theta)$ as:

$$L^* \approx \underset{L}{\text{argmax}}\, p(S^*|L,\alpha)p(L|\gamma,G),$$

$$\text{where, } p(S^*|L,\alpha) = \prod_{p=1}^{P} p(\mathbf{s}_{L_p}|L_p,\alpha_p), \quad p(L|\gamma,G) \propto \prod_{\{i,j\}} \psi(L_i,L_j;\gamma_{ij}).$$

Tree-based Pictorial Structures (PS) model: assumes a tree-structured G and the parameters can be separated and learnt efficiently using the following:

$$\alpha_p = \underset{\alpha_p}{\text{argmax}} \prod_m p(S_{mL_p}|\alpha_p), \tag{1}$$

$$G = \underset{G \subset F}{\text{argmax}} \prod_{m,\{i,j\} \in G} \psi(L_{mi}, L_{mj}; \gamma_{ij}), \tag{2}$$

$$\gamma_{ij} = \underset{\gamma_{ij}}{\text{argmax}} \prod_m \psi(L_{mi}, L_{mj}; \gamma_{ij}) \text{ for} \tag{3}$$

Fully connected (FC) constellation-based model: We upgrade the tree-structured G to the fully connected (FC) graph F (similar to [2]), to explore its power, especially in promoting repulsion between leaf node locations (*e.g.* two "cat" paw parts should be found on separate paws of a query cat).

Inference results: The unary and pairwise terms of (1, 3) can be defined as energies/costs for a given shape as:

$$p(S_{mL_p}|\alpha_p) \propto \exp\left(-\phi(S_{mL_p};\alpha_p)\right)) \propto \mathcal{N}(\mathbf{s}_{mL_p}|\alpha_{\mu_p}, \alpha_{\Sigma_p}), \tag{4}$$

$$\psi(L_{mi}, L_{mj}; \gamma_{ij}, \{ij\} \in G) \propto \mathcal{N}(\text{dist}(L_{mi}, L_{mj})|\gamma_{\mu_{ij}}, \gamma_{\Sigma_{ij}}). \tag{5}$$

Both the PS and FC models are solved using TRW-S [4]. The approximate inference of the more complex FC model performs better, while simultaneously telling us how close we are to the global optimum, than the exact optimization on the simpler tree-based PS one. We significantly outperform both PS and Implicit Shape Model based benchmarks qualitatively and quantitatively, on full and "punctured" versions of the TOSCA dataset.

Extension to dense correspondence estimation: The task is to assign the appropriate vertex label $L_i \in \{1 \ldots V_n\}$ of S_n to each vertex $i \in \{1 \ldots V_m\}$ of S_m. The graph G is replaced with $\{v_m, \varepsilon_m\}$ of shape S_m. Similar to (4), the unary term evaluates how well $\mathbf{s}_{nj}$ (shape S_n at vertex j) matches $\mathbf{s}_{mi}$. The pairwise cost encourages labels $L_i, L_{i'}$ in S_n to have a similar relative distance as neighbouring nodes i, i' in S_m. Sparse correspondences from the FC object part localization, are used as hard constraints to seed the dense correspondence optimization. Part p localized to i in S_m and k in S_n, gives us a vertex correspondence i, k. In summary, the unary and pairwise energies can be written as:

$$\phi(L_i;\lambda) = \begin{cases} \infty, & (L_i \neq k)\,\&\,(i,k \text{ in corr.}) \\ \lambda(\mathbf{s}_{nL_i} - \mathbf{s}_{mi})^2, & \text{otherwise} \end{cases}$$

$$-\log(\psi(L_i, L_{i'}|\lambda)) = \begin{cases} \infty, & (L_i \neq k)\,\&\,(i,k \text{ in corr.}) \\ (\text{dist}_n(L_i, L_{i'}) - \text{dist}_m(i,i'))^2, & \text{otherwise} \end{cases}$$

Again, this energy can be optimized using [4]. λ is empirically set. With with meshes of ~ 3000 nodes and label range of ~ 3000, this is computationally daunting. Thus, we perform this correspondence establishment hierarchically; the mesh resolution is increased in steps of ~ 300 vertices. The result of the previous stage is used as fixed correspondences in the subsequent stage. Highly unlikely labels are pruned. Thus a relatively simple problem is solved at each step to give robust correspondences.

[1] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. *IJCV*, 61(1):55–79, 2005.

[2] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. CVPR*, volume 2, pages 264–271, Jun 2003. URL `http://www.robots.ox.ac.uk/~vgg`.

[3] J. Knopp, M. Prasad, G. Willems, R. Timofte, and L. Van Gool. Hough transform and 3d surf for robust three dimensional classification. In *Proc. ECCV*, 2010.

[4] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE PAMI*, 28(10):1568 –1583, Oct 2006. ISSN 0162-8828. doi: 10.1109/TPAMI.2006.200.