LaTeX Author Guidelines for CVPR Proceedings

Anonymous CVPR submission

Paper ID ****

Abstract

Proposal for zimo's project.

1. Constructing Functional Maps from Object Proposals

The goal of this sub-project is to study how to optimize a functional map between a pair of images from object proposals (and their correspondences) computed on these images. Specifically, suppose we are given a pair of images I_1, I_2. Suppose for each image I_i, we have computed a set of object proposals $\mathcal{S}_1 = \{s_1 \in I_1\}$. By comparing descriptors of object proposals, we have obtained a set of nearest-neighbor correspondences between these object proposals $\mathcal{C}_{12} \in \mathcal{S}_1 \times \mathcal{S}_2$, namely, for each object proposal $s_1 \in \mathcal{S}_1$, we link it to the object proposal $s_2 \in \mathcal{S}_2$ so that

$$s_2 = \underset{\arg\min}{s \in \mathcal{S}_2} \|d(s_1) - d(s)\|.$$

Each correspondence (s_1, s_2) receives a weight $w_{(s_1,s_2)} \in [0, 1]$ based on

- the saliency score of each proposal
- the difference in the descriptors.
- the difference in object bounding boxes (if the descriptor are computed after normalizing objects). For example, it is unlikely for a 5x5 box to be matched to a 200x200 box on the other image. Even they both represent the same object, we tend to avoid the case here since later we will use intermediate images to link them: 5x5 to 20x20 on one image to 40x40 on the third image......

The following are some comments are setting this up:

- The object proposals are computed using some off-the-shelf generators (e.g., geodesic object proposals [1]). We favor the methods that generate object proposal boundaries in addition to object proposal bounding boxes. Note that each object proposal method has many parameters. These parameters should be tuned to minimize the number of necessary object proposals per image.
- The object proposal descriptor (preferred to be HOG) should be computed in a smart way. For example, we can enlarge the bounding box a bit to include more contextual information (e.g, 100x100 to 200x200). As another example, we may remove the background before computing the descriptor. The descriptors computed in different ways may be combined together. The goal is to improve the quality of the resulting descriptors.
- We may set up a threshold on the difference between object proposal descriptors. This helps since many object proposals are small and do not have good matches on the other image.

OK, now we are ready to do the real stuff. Suppose we have pre-computed a functional space $\mathcal{F}_1$ and $\mathcal{F}_2$ on each image, e.g., using the procedure described in Fan's ICCV 2013 paper. Denote $\mathbf{f}_{s_1}$ as the projection of the segment indicator function of s_1 on to $\mathcal{F}_1$. We seek to optimize a functional map $X : \mathcal{F}_1 \rightarrow \mathcal{F}_2$ by solving the following optimization problem:

$$\begin{aligned} \underset{X}{\text{minimize}} \sum_{c=(s_1,s_2)\in\mathcal{C}} & \Big(w_{(s_1,s_2)} \|X\mathbf{f}_{s_1} - \alpha_c \mathbf{f}_{s_2}\|_0 + \lambda(\alpha_c - 1)^2 \Big) \\ & + \mu \sum_{i,j} (\lambda_i(I_1) - \lambda_j(I_2))^2 X_{ij}^2. \end{aligned} \quad (1)$$

The terms are explained as follows:

- The data term $X\mathbf{f}_{s_1} - \alpha_c \mathbf{f}_{s_2}$ is measures in the terms of the L^0-norm. This is to account for the noise in the objects correspondences.
- We put α_c in front of each $\mathbf{f}_{s_2}$ to account for variations in object-sizes and the shrinking effect caused by the regularization term. Note that we are using truncated-basis, making it hard to separate inliers and outliers

by their residuals. For example, $X\mathbf{f}_{s_1} - \alpha_c \mathbf{f}_{s_2}$ would not be zero even if (s_1, s_2) is a perfect match. Yet including α_c can alleviate this.

- Zimo proposed to measure $\frac{X\mathbf{f}_{s_1}}{\|X\mathbf{f}_{s_1}\|} - \mathbf{f}_{s_2}$. This is a good idea, but it is hard to implement.
- the regularization term is defined after normalizing the eigenvalues.

For optimization, we can use the following alternating scheme. Initialize with $w^2_{(s_1,s_2)} = 1, (s_1, s_2) \in \mathcal{C}$. Solve the following optimization problem to derive X:

$$\underset{X}{\text{minimize}} \sum_{c=(s_1,s_2)\in\mathcal{C}} \Big(w_{(s_1,s_2)} w^2_{(s_1,s_2)} \|X\mathbf{f}_{s_1} - \alpha_c \mathbf{f}_{s_2}\|^2 + \lambda(\alpha_c - 1)^2 \Big) + \mu \sum_{i,j} (\lambda_i(I_1) - \lambda_j(I_2))^2 X^2_{ij}. \quad (2)$$

After that, update the weights

$$w^2_{(s_1,s_2)} = \frac{\sigma^2}{\sigma^2 + \|X\mathbf{f}_{s_1} - \alpha_c \mathbf{f}_{s_2}\|^2}.$$

Remarks. The parameters should be tuned carefully, understanding the tradeoffs in depth.

In theory, the proposed technique can only tolerate '50% of noise'[1]. So we need to understand the regime of image pairs where this technique works.

2. Extension to Multiple Images

3. Joint Optimization of Functional Maps and Basis

References

[1] P. Krähenbühl and V. Koltun. Geodesic object proposals. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*, pages 725–739, 2014. 1

[1]This is not a rigorous statement because we pre-weight the correspondences