

Local Feature Based Unsupervised Alignment of Object Class Images

Jukka Lankinen

<http://personal.lut.fi/users/Jukka.Lankinen>

Joni-Kristian Kamarainen

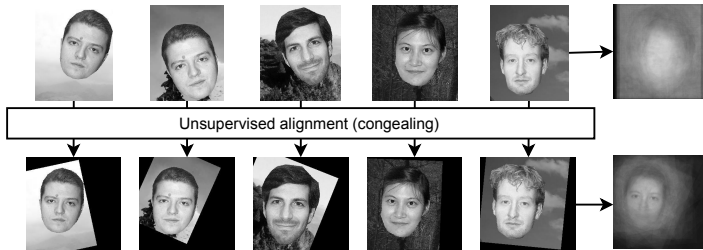
<http://personal.lut.fi/users/joni.kamarainen>

Machine Vision and Pattern Recognition Laboratory

Lappeenranta University of Technology (LUT)

Kouvola Unit, Finland

Our method can be used to automatically align a set of images containing similar objects

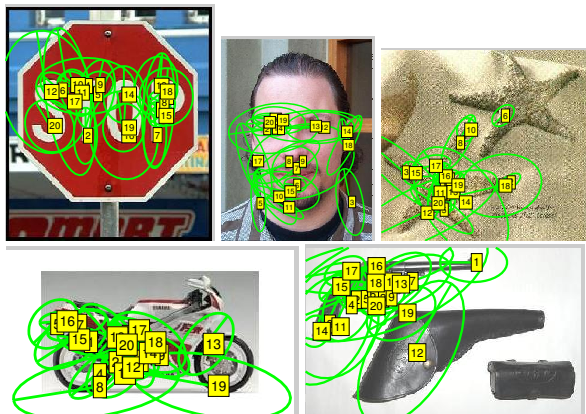


Alignment of objects is a predominant problem in visual object categorisation (VOC). State-of-the-art part-based VOC methods try to automatically learn object parts and their spatial variation, which is difficult for objects in arbitrary poses. A straightforward solution is to annotate images with a set of “object landmarks”, but due to laborious work required, less supervised methods are preferred. Effective semi-supervised VOC methods have been introduced, but none of them explicitly define an alignment procedure or study its effect to overall VOC performance.

Unsupervised alignment has been recognised as its own problem referred to as “spatial image congealing” and a number of congealing methods have been proposed. These methods are mainly seminal work to Learned-Miller [3, 4] extending and improving the original algorithm. The main drawback of the congealing methods is that they are iterative optimisation methods operating on pixel-level and thus require at least moderate initial alignment to converge.

Our approach [2] deviates from the congealing works by the fact that we utilise local features instead of pixel level processing, i.e. feature-based congealing. Our solution is more similar to those used in the part-based VOC methods, but we explicitly define the alignment algorithm and measure its performance.

Our method selects one image as a “seed” and then searches for a sub-set of seed’s landmarks which match and appear in similar spatial configuration in other images



The alignment algorithm basically has four steps: Local feature extraction, feature matching, spatial scoring and finally the image alignment. For the extraction and matching our method can utilise any available combination of local feature detectors and descriptors, such as Hessian-Affine and SIFT in this work.

In the spatial scoring step, scores of seed features which match under a selected transformation, such as 2D homography, between other images are iteratively accumulated. For point correspondences homography transformation is computed using standard linear methods for isometry

and similarity. For a number of iterations we randomly select the minimum number of correspondences (two for isometry and similarity, and three for affinity), estimate homography, transform image points to the seed, and accumulate scores of the features matching within a pre-set distance limit. Justification to this random approach yields from the random sample consensus (RANSAC) robust estimation. Robust estimation is needed since there are lots of false matches and only a few seed features match to a particular class example.

The spatial scoring algorithm outputs the best K landmarks in the seed based on the top scores. The top scoring seed features represent landmarks which have been independently verified by other features in a similar configuration in other images. With the automatically selected seed landmarks the alignment procedure is similar to robust stereo and baseline matching.

Our method outperforms a state-of-the-art congealing method for difficult examples in Caltech-101 and its randomised version r-Caltech-101

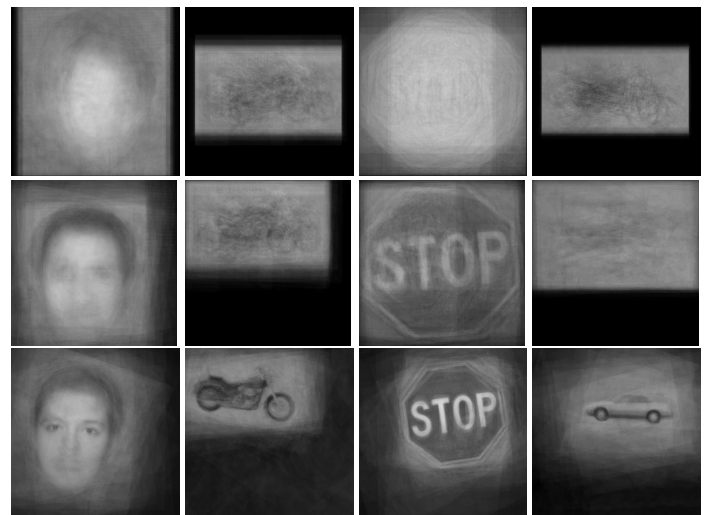


Figure 1: Qualitative alignment results for r-Caltech-101: unaligned average (top), Huang et al. [1] (middle), and our method [2] (bottom).

Both qualitative and quantitative results in our experiments verified that the proposed feature-based congealing approach finds object class specific landmarks and successfully aligns a set of object class images. The most important remaining research question is: how to select the best seed automatically? In this work, multiple seeds were used and the best selected manually based on the average images. However, the optimal seed selection is an intriguing research question as it is also analog to the problem of selecting the most representative class example within a given set of visual object class images.

- [1] G.B. Huang, V. Jain, and E. Learned-Miller. Unsupervised joint alignment of complex images. In *ICCV*, 2007.
- [2] J. Lankinen and J.-K. Kamarainen. Local feature based unsupervised alignment of object class images. In *British Machine Vision Conference (BMVC2011)*, Dundee, UK, 2011.
- [3] E.G. Learned-Miller. Data driven image models through continuous joint alignment. *IEEE PAMI*, 28(2):236–250, 2006.
- [4] E. Miller, N. Matsakis, and P. Viola. Learning from one example through shared densities of transforms. In *CVPR*, 2000.