Searching for A Thing

Smeulders, A.W.M.; Tao, R.

Published in:
ICMR'17

DOI:
10.1145/3078971.3079006

Searching for A Thing

Arnold W.M. Smeulders
a.w.m.smeulders@uva.nl

Ran Tao
r.tao@uva.nl

ABSTRACT

For humans, one picture usually suffices to identify an object of search. *I am looking for this little girl, have you seen her?* or *Do you have such another one?* are two ways to specify a target even to someone who has never seen the object of search before. Searching from one example in digital multimedia retrieval is a hard problem. From the one example one needs to derive an accurate estimate of all accidental variations in the target picture as well as the structural variation of the target in all other potential pictures. From the one example one needs to derive an accurate estimate of all accidental variations the target instance might have.

We argue that with only one reliable example, locality is at the core of the instance search algorithms. What we need is to emphasize locality at all steps in the algorithm [1]. For each image in the dataset, first one needs to gather evidence from a local area in the image and compare the evidence from that region with the target. Then one needs to search in feature space in the region around the feature points of the target for evidence in the current local region to explain all evidence in view. And, finally one needs to emphasize the similarity function to give preference to evidence which is close to appearances of the target. With this emphasis on locality one is especially successful when searching for flat 2D geometrically rigid layouts like logos and buildings. It also holds for scene recognition, who are effectively flat and one-sided as well due to their large depth.

To generalize to searching for rigid 3D-objects from one example image the variability in view point variations is so big that one needs to learn about other viewpoints in general. General knowledge can come from known object-types, which will limit the search to objects of the same type, prior to verifying in the resulting local neighborhood for instances of the one given example [2]. Then, also the attributes within the category need to be similar to the attributes of the target. We have learned viewpoint invariant attributes from clean shoe images automatically, that is without labelling the attributes a priori [2]. Most of these attributes appear to correspond to components of shoes such as high heels, boots, open toes and so on. With the offline-learned viewpoint-invariant category type and subtype (attributes) classification, the search for the target in the wild is surprisingly successful from one example [2].

To generalize further, there is no other route than first learning the variances and invariances off-line. Where in [1] objects are essentially 2D-flat, and in [2] objects are rigid 3D, the next boundary to cross is searching for articulated objects. Then, non-rigid objects are a hurdle yet to take, followed by searching for objects for which the category or attributes have not been seen before. Current research is on learning more about object variations in general [3].

Searching things from one example is not over yet, but there is a reward waiting in the sense that one example search is useful in other branches of retrieval.

KEYWORDS

Instance search; content based image retrieval

ACM Reference format:

Arnold W.M. Smeulders, Ran Tao. Int Conf Multimedia Retrieval 2017.

ICMR '17, June 06-09, 2017, Bucharest, Romania

ACM ISBN 978-1-4503-4701-3/17/06.
http://dx.doi.org/10.1145/3078971.3079006

REFERENCES

[1] R Tao, E Gavves, CGM Snoek, and AWM Smeulders, Locality in generic instance search from one example. In CVPR, 2014.
[2] R Tao, AWM Smeulders, and SF Chang, Attributes and categories for generic instance search from one example. In CVPR, 2015.
[3] R Tao, E Gavves, and AWM Smeulders, Siamese instance search for tracking. In CVPR, 2016.