How to Exploit Scene Constraints to Improve Object Categorization Algorithms for Industrial Applications?

Steven Puttemans and Toon Goedemé

EAVISE, Campus De Nayer, ESAT/PSI-VISICS, KU Leuven, Kasteelpark Arenberg 10, Heverlee, Belgium
{steven.puttemans, toon.goedeme}@lessius.eu

Abstract: State-of-the-art object categorization algorithms are designed to be heavily robust against scene variations like illumination changes, occlusions, scale changes, orientation and location differences, background clutter and object intra-class variability. However, in industrial machine vision applications where objects with variable appearance have to be detected, many of these variations are in fact constant and can be seen as constraints on the scene, which in turn can reduce the enormous search space for object instances. In this position paper we explore the possibility to fixate certain of these variations according to the application specific scene constraints and investigate the influence of these adaptations on three main aspects of object categorization algorithms: the amount of training data needed, the speed of the detection and the amount of false detections. Moreover, we propose steps to simplify the training process under such scene constraints.

1   INTRODUCTION

Object categorization has extended the principle of detecting objects with a known appearance towards detecting objects based on a general object class model that tries to contain all intra-class variability. For example, Figure 3 shows that multiple instances of the object class 'pedestrians', do have a lot of intra-class variability, like different clothing, size, poses, gestures, etc. This variability can be captured within a single class model, as in Figure 3.

In academic context these object categorization algorithms are tested on typical classes, see Figure 1, like bikes, chairs, airplanes, etc. and perform a search for object instances of these classes in very complex scenes like street views, an airport, a shop, etc.
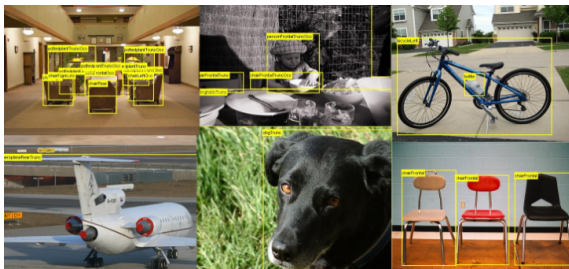
However, the actual needs of industrial applications (Figure 2) differ a lot from these circumstances. There we would like to detect object classes with a large intra-class variability in very controlled scenes. Examples of these specific industrial machine vision applications are counting micro-organisms in a closed lab environment, counting the amount of flower buds for orchid grading, picking of peppers from a conveyor belt or random bin picking. Especially natural grown products show this large intra-class variability and are frequently handled in very controlled production environments, with e.g. constant illumination and camera-object distance.

Figure 1: Examples of typical object categorization test classes used in academic research: chairs, bikes, airplanes, dogs, children ...

Figure 2: Examples of industrial object categorization applications: robot picking and object counting of natural products.

Table 1: Comparison of robustness against degrees of freedom of existing object categorization algorithms. (**Illumin.** = Illumination differences / **Locati.** = Location of objects / **Scale** = Scale changes / **Orient.** = Orientation of objects / **Occlu.** = Occlusions / **Clutt.** = Clutter in scene / **I.C.V.** = Intra-class variability)

Technique	Example	Degrees of freedom						
		Illumin.	Locati.	Scale	Orient.	Occlu.	Clutt.	I.C.V.
NCC - based pattern matching	(Lewis, 1995)	X	X	–	–	–	–	–
Edge - based pattern matching	(Hsieh et al., 1997)	X	X	X	X	–	–	–
Global moment invariants for recognition	(Mindru et al., 2004)	X	X	X	X	–	–	–
Object recognition with local keypoints	(Bay et al., 2006)	X	X	X	X	X	X	–
Object categorization algorithms	(Gall and Lempitsky, 2009)	X	X	X	–	X	X	X
Industrial Applications	–	**–**	**–**	**–**	**X**	**X**	**–**	**X**

Figure 3: Example of (LEFT) intra-class variability and (RIGHT) an object class model for detecting pedestrians.

To illustrate our case we refer to four state-of-the-art object categorization algorithms, each using their own kind of low-level features: (Viola and Jones, 2001) uses Haar-like wavelets, (Felzenszwalb et al., 2010) makes use of a Histogram of Oriented Gradients, (Leibe and Schiele, 2004) is based on local features combined with an implicit shape model, while (Gall and Lempitsky, 2009) uses random Hough forests.

Starting from these existing algorithms, the main goal of this research is to derive a complete new set of object categorization algorithms by using the information of the controlled industrial environments as constraints for the algorithms. Exploiting these constraints will lead to algorithms that will need less training data, perform faster up to real-time, have more accurate detections and reduce the amount of false detections.

The remaining part of this position paper is organized as follows: section 2 situates our approach in the evolution towards robust algorithms, section 3 discusses the different constraints that can be applied in industrial applications, section 4 explains how the constraints from section 3 can improve the complete training phase and finally section 5 summarizes the discussion in previous sections and concludes what we try to achieve with our research in progress.

2 EVOLUTION TOWARDS ROBUST ALGORITHMS

Nowadays, the main strive of object categorization algorithms is evolving towards algorithms that are as robust as possible. In real-life applications, for example detecting persons in a scene, any object detection algorithm has, next to the object variation itself, to take care of multiple variable elements like illumination changes, object scale changes, etc.

In this position paper we will focus on the scene variations that are presented in Table 1 as degrees of freedom. When applying this e.g. to detecting pedestrians in a street view scene, lighting depends on available sunlight. Occlusion happens when a pedestrian is partially covered by walking behind a car. Pedestrians occur in different scales and locations depending where in the street view they appear. The street view contains heaps of clutter like pets, shops, street signs, etc. which should not be classified as pedestrians.

In order to cope with all these scene variations, object categorization algorithms can follow two approaches: normalization or invariance. The first approach eliminates variations of a certain degree of freedom by a normalization of the input data towards a fixed value. By e.g. rescaling each search window to a standard scale, we do not need to provide training examples to the algorithm containing all possible scales. In the other approach the algorithm is made invariant to a certain influence. Removing illumination changes can be done by converting the image to an illumination invariant form, like a gradient image.

It is clear that making object categorization algorithms robust to all of these variations leads to very complex and computationally expensive algorithms. Taking into account the non-variability of certain degrees of freedom for a specific application area (Table 1) will surely be beneficial.

3 USE OF SCENE CONSTRAINTS

Looking at the machine vision applications at hand, random object picking and counting, we notice that many of the degrees of freedom discussed in Table 1 can actually be fixed, resulting in specific constraints. Using this knowledge we reckon two approaches are possible. The first approach would be taking existing state-of-the-art object categorization algorithms and using the constraints to reduce the amount of training samples needed, because less scene variation needs to be included. A second, more advanced, approach adapts the existing algorithms in order to remove the internal functionality that handles these scene variations.

We believe that these scene specific constraints will reduce the detection time and increase the performance of these object categorisation algorithms. The next subsections discuss which scene variations can be constrained in these industrial applications and how we can benefit from doing so. Subsection 3.1. discusses illumination changes, 3.2. scale changes and localization and 3.3 orientation.

3.1 Illumination changes

In most industrial vision applications, illumination of the scene is highly controlled and therefore constant. If we apply this constraint, all operations needed to make object categorization algorithms illumination invariant, can be simply removed. No longer will we use grayscale gradient images but we will explore the possibility to use the very RGB values instead. Doing so we will regain valuable colour information and we will reduce the amount of input images, needed to cover illumination variance. (Abdel-Hakim and Farag, 2006) already proves that using colour knowledge for object detection can improve detection rates.

3.2 Scale changes and localization

Scale and location of objects that need to be detected in the scene are known when the position of the camera is fixed above the conveyor belt containing the objects of interest. This reduces the region of interest where object instances could occur and it reduces the amount of scales that can occur inside this region of interest to a single scale. The algorithm in (Viola and Jones, 2001), which uses a sliding window approach, applies different scales of the initial feature detectors, repeating this for every single window. Applying both constraints leads to a single scale search in a selected region of interest, and thus a much smaller search space.

By first determining regions that actually change compared to the background, using background subtraction algorithms, we can define possible regions where objects can be found, since these objects aren't static in our applications. This can also help to improve localization and reduces again the large search space of object candidates.

3.3 Orientation

In random bin picking applications, as well as in counting object applications, the orientation of object class instances can be unknown, unlike in state-of-the-art object categorization applications such as pedestrian detection, where pedestrians always appear upright. The classic solution to train an orientation invariant object detector is to use a large amount of training samples which contain all possible orientations of the actual object class.

We suggest to apply a structural change in the classic approach, by first searching for the dominant orientation in the selected region and by applying a transformation towards a single trained orientation.

An illustration of such an approach is our previous work in (Van Beeck et al., 2012), where pedestrian detection in the blind spot camera of trucks is addressed. Both the knowledge of orientation and scale at certain positions are used as constraints, helping to obtain a single detection at a fixed scale and orientation, yielding a $10 \times$ speed-up.

4 SIMPLIFYING THE TRAINING PROCESS

Next to using scene constraints for improving object categorization presented in section 3, another improvement of the existing algorithms is simplifying the complex training process. Especially the preprocessing annotation phase, where all training images are formatted and processed in a specific way, is a very expensive and time consuming labour. Next to that, the actual training phase used to train an effective classifier can be improved by using the scene constraints of our specific industrial applications.

Subsection 4.1. discusses the improvements we plan in the actual training phase, while subsection 4.2. focusses on the changes made in the annotation phase.

4.1 The Actual Training Phase

A scene constraint on the object categorization algorithm should be the motivation of only trying to train a specific scale, orientation, ... and thus removing the

positive training samples covering all other options of these scene variations. This will lead to a greatly reduced positive training sample set.

Training object categorization algorithms also requires negative training samples, containing clutter and random elements not related to the object class. Looking at the constrained scene, we are sure we can drastically lower the amount of negatives training samples needed, maybe even to a single negative training example, e.g. an empty conveyor belt.

Another industrially relevant aspect is how much training examples are in fact really needed to obtain a robust classifier. None of the papers describing the state-of-the-art object categorization algorithms define a way of determining the exact size of the positive and negative training sample set in order to include as many object variations possible.

4.2 The Annotation Phase

Since this preprocessing phase is actually the most time-consuming part, due to multiple training prerequisites of training samples, it is interesting to improve this phase also. Traditionally many thousands of training samples are required, of which each sample needs to be manually formatted correctly. Formatting exist of grayscale conversion and adding a region of interest for each object instance. Furthermore, the latest techniques (like (Leibe and Schiele, 2004) and (Gall and Lempitsky, 2009)) also model the relation of different parts towards an object centre, requiring the centre of each object instance to be defined also.

We plan to evolve from a fully supervised annotation phase towards a semi-supervised. To accomplish this we will use a selected set of manually annotated training samples in order to create a first basic classifier, which in turn will be used to perform object detections in the remaining training samples. Detections will be accepted based on a confidence score, giving an indication of how certain the algorithm actually detected an object. We then reduce the manual annotation work towards simply accepting or removing suggested selections by this basic classifier. If the above basic classifier however would not succeed in achieving a decent performance rate, then a possible solution is to look deeper into combining our approach which existing machine learning algorithms like boosting and online learning.

A more detailed study of the approach mentioned above will show how much we can actually reduce the number of manually labelled training samples in comparison to the complete training sample set.

5 CONCLUSION

Many state-of-the-art object categorization techniques try to become robust against widely varying parameters like illumination and scale changes in the scene. By doing so they become complex to handle and result in algorithms that have a good detection rate but a large computational time.

However in many real-life industrial machine vision applications, a lot of those changing parameters are actually constant, except for the intra-class variability which is still one of the greatest challenges of today's object categorization algorithms. Industrial applications are in need of object categorization algorithms that can handle this intra-class variability, but at the same time perform real-time processing.

By making use of the known scene and translating this into constraints for the algorithm we are sure that we can adopt existing approaches and create a new set of fast real-time processing object categorization algorithms that can answer to the needs of this specific application area of industrial applications like random object picking and object counting.

REFERENCES

Abdel-Hakim, A. and Farag, A. (2006). Csift: A sift descriptor with color invariant characteristics. In *CVPR*, pages 1978–1983.

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. *ECCV*, pages 404–417.

Felzenszwalb, P., Girshick, R., and McAllester, D. (2010). Cascade object detection with deformable part models. In *CVPR*, pages 2241–2248.

Gall, J. and Lempitsky, V. (2009). Class-specific hough forests for object detection. In *CVPR*, pages 1022–1029.

Hsieh, J., Liao, H., Fan, K., Ko, M., and Hung, Y. (1997). Image registration using a new edge-based approach. *CVIU*, pages 112–130.

Leibe, B. and Schiele, B. (2004). Scale-invariant object categorization using a scale-adaptive mean-shift search. *Pattern Recognition*, pages 145–153.

Lewis, J. (1995). Fast normalized cross-correlation. In *Vision interface*, volume 10, pages 120–123.

Mindru, F., Tuytelaars, T., Gool, L., and Moons, T. (2004). Moment invariants for recognition under changing viewpoint and illumination. *CVIU*, 94:3–27.

Van Beeck, K., Goedemé, T., and Tuytelaars, T. (2012). A warping window approach to real-time vision-based pedestrian detection in a truck's blind spot zone. In *ICINCO*, volume 2, pages 561–568.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *CVPR*, pages I–511.