Optimal Object Categorization under Application Specific Conditions

Steven Puttemans and Toon Goedemé
EAVISE, KU Leuven, Campus De Nayer, Jan De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium
ESAT/PSI-VISICS, KU Leuven, Kasteelpark Arenberg 10, Heverlee, Belgium

1 STAGE OF THE RESEARCH

The main focus of this PhD research is to create a universal object categorization framework that uses the knowledge of application specific scene and object variation to reach detection rates up to 99.9%. This very high detection rate is one of the many requirements of industrial applications, before the industry will even consider using object categorization techniques. Currently the PhD research has been running for one year and has initially focussed on analyzing existing state-of-the-art object categorization algorithms like (Viola and Jones, 2001; Gall and Lempitsky, 2013; Dollár et al., 2009; Felzenszwalb et al., 2010). Besides that, scene and object variation were used to apply pre- and postprocessing on the actual detection output, to reduce the occurance of false positive detections. The next step will be to actually create a new universal object categorization framework based on the experience gathered during the first year of research, using the selected technique of (Dollár et al., 2009) as a backbone for further research.

2 RESEARCH PROBLEM

The focus of this research lies in industrial computer vision applications that want to perform object detection on object classes with a high intra-class variability. This means that objects have varying size, color, texture, orientation, ... Examples of these specific industrial cases can be seen in Figure 1. These day-to-day industrial applications, such as product inspection, counting and robot picking, are in desperate need of robust, fast and accurate object detection techniques which reach detection rates of 99.9% or higher. However, current state-of-the-art object categorization techniques only guarantee a detection rate of ±85% when performing *in the wild* detections (Dollár et al., 2010). In order to reach a higher detection rate, the algorithms impose very strict restrictions on the actual application environment, e.g. a constant and uniform lighting source, a large contrast between objects and background, a constant object size and color, ... Compared to these more complex object categorization algorithms, classic thresholding based segmentation techniques require all of these restrictions to even guarantee a good detection result and are thus unable to cope with variation in the input data.

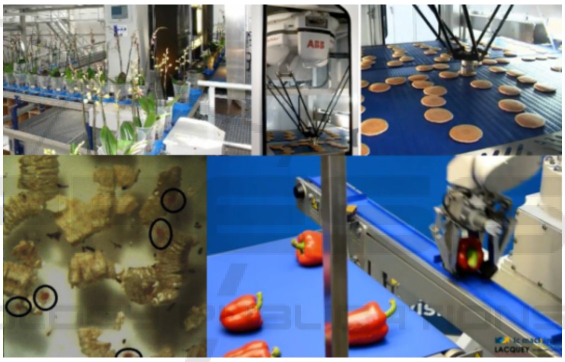
Figure 1: Examples of industrial object categorization applications: robot picking and object counting of natural products. [checking flower quality, picking pancakes, counting micro-organisms, picking peppers]

Looking at the state-of-the-art object categorization techniques, we see that the evolution of these techniques is driven by *in the wild* object detection (see Table 1). The main goal exists in coping with as many variation as possible, achieving a high detection rate in very complex scenery. However, specific industrial applications easily introduce many constraints, due to the application specific setup of the scenery and the objects. Exploiting that knowledge can lead to smarter and better object categorization techniques. For example, when detecting apples on a transportation system, many parameters like the location, background and camera position are known. Current object categorization techniques don't use this information because they do not expect this kind of known variation. However exploiting this information will lead to a new universal object detection framework that yields high and accurate detection rates, based on the scenery specific knowledge.

Puttemans S. and Goedemé T..
Optimal Object Categorization under Application Specific Conditions.

Table 1: Evolution in robustness of object recognition and object detection techniques trying to cope with object and scene variation as mentioned in (Puttemans and Goedemé, 2013) (**[1]** Illumination differences / **[2]** Location of objects / **[3]** Scale changes / **[4]** Orientation of objects / **[5]** Occlusions / **[6]** Clutter in scene / **[7]** Intra-class variability).

Technique	**Example**	**Degrees of freedom**						
		1	**2**	**3**	**4**	**5**	**6**	**7**
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	(Felzenszwalb et al., 2010)	X	X	X	–	X	X	X
Industrial Applications	–	**–**	**–**	**–**	**X**	**X**	**–**	**X**

3 STATE OF THE ART

Object detection is a widly spread research topic, with large interest in current state-of-the-art object categorization techniques. (Dollár et al., 2009) suggested a framework based on integral channel features, where all object characteristics are captured into feature descriptions which are then used as a large pool of training data in a boosting process (Freund et al., 1999). In contrast to the original boosted cascade of weak classifiers approach, suggested by (Viola and Jones, 2001), this technique incorporates multiple sources of information to guarantee a higher detection rate and less false positive detections.

In the following years of research, this technique has been a backbone for many well performing object detection techniques, mainly for into the wild detections of pedestrians (Benenson et al., 2012; Benenson et al., 2013; Dollár et al., 2010) and traffic signs (Mathias et al., 2013). All these recently developped techniques profit from the fact that the integral channel features framework allows to integrate extra application-specific knowledge like stereo vision information, knowledge of camera position, ground plane assumption, ... to obtain higher detection rates. The concept of using application specific scene constraints to improve these state-of-the-art object categorization techniques was introduced in (Puttemans and Goedemé, 2013). The paper suggests using the knowledge of the application specific scene and object conditions as constraints to improve the detection rate, to remove false positive detections and to drastically reduce the number of manual annotations needed for the training of an effective object model.

Aside from effectively using the scene and object variation information to create a more accurate application specific object detector, the PhD research will focus on reducing the amount of time needed for manually annotating gigantic databases of positive and negative training images. This will be done using the technique of active learning, on which a lot of recent research was performed (Li and Guo, 2013; Kapoor et al., 2007). This research clearly shows that integrating multiple sources of information into an active learning strategy can help to isolate the large problem of outliers giving reason to include the wrong examples.

4 OUTLINE OF OBJECTIVES

During this PhD existing state-of-the-art object categorization algoritms will be reshaped into a single universal semi-automatic object categorization framework for industrial object detection, which exploits the knowledge of application specific object and scene variation to guarantee high detection rates. Exploiting this knowledge will enable three objectives, each focussing on another aspect of object detection that is important for the industry.

1. **A High Detection Rate of 99.9% or Even Higher.** Classic techniques reach detection rates of 85% during *in the wild* detections, but for industrial applications a rate of 99.9% and higher is required. By integrating the knowledge of the object and scene variation, the suggested approach will manage to reach this high demands. Using the framework of (Dollár et al., 2009) as a backbone for the universal object categorization framework that will be created, these characteristics will be used to include new feature channels to the model training process, focussing on this specific object and scene variation.
2. **A Minimal Manual Input During the Training of an Object Model.** Classic techniques demand many thousands of manual annotations during the collection of training data. By using an innovative active learning strategy, which again uses the knowledge of application specific scene and

object variation, the number of manual annotations will be reduced to a much smaller number of input images. By iteratively annotating only a small part of the trainingset and using that to train a temporary detector based on the already annotated images, the algorithm will decide which new examples will actually lead to a higher detection rate, only offer those for a new annotation phase and omit the others.

3. **A Faster and More Optimized Algorithm.** By adding all of this extra functionality, resulting in multiple new feature channels, into a new framework, a large portion of extra processing is added. Based on the fact that the original algorithm is already time consuming and computational expensive, the resulting framework will most likely be slower than current state-of-the-art techniques. However, by applying CPU and GPU optimalizations wherever possible, the aim of the PhD is to still provide a framework that can supply real time processing.

The use of all this application specific knowledge from the scene and the object, with the aim of reaching higher detection rates, is not a new concept. Some approaches already use pre- and postprocessing steps to remove false positive detections based on application specific knowledge that can be gathered together with the training images. For example, (Benenson et al., 2012), use the knowledge of a stereo vision setup and ground plane assumption, to reduce the area where pedestrian candidates are looked for. This PhD research however will take it one step further and will try to integrate all this knowledge into the actual object categorization framework. This leads to several advantages over the pre- and postprocessing approaches:

- There will be no need for manual defining or capturing features that are interesting for this pre- and postprocessing steps.
- Multiple features will be supplied as a large package to the framework. The underlying boosting algorithm will then decide which features are actually interesting to use for model training.
- The algorithm can seperate the input data better than human perception based on combination of features.
- Each possible scene and object variation will be transformed into a new feature channel, in order to capture as much variation as possible. Once a channel is defined, it can be automatically recalculated for every possible application.

Besides not being able to reach top level detection rates, state-of-the-art object categorization techniques face the existence of false positive detections. These detections are classified by the object detector model as actual objects, because they contain enough discriminating features. However they are no actual objects in the supplied data. By adding a larger set of feature channels to the framework, and thus integrating a larger knowledge of scene and object variation during the training phase, the resulting framework will effectively reduce the amount of false positive detections.

5 METHODOLOGY

In order to ensure a systematic approach, the overal research problem of the PhD is divided into a set of subproblems, which can be solved one by one in an order of gradual increase in complexity, in order to guarantee the best results possible. Section 5.1 will discuss the integration of the application specific scene and object variation during the model training process, by highlighting different variation aspects of possible applications and how they will be integrated as feature channels. Section 5.2 will illustrate how the use of an innovative active learning strategy can help out with reducing the time consuming job of manual annotation. Finally section 5.3 will discuss how the resulting framework can be optimized using CPU and GPU optimalizations wherever possible.

5.1 Integration of Scene and Object Variation During Model Training

Different properties of application specific scene and object variation allow to design a batch of new feature channels in a smart way, that can be used for a universal object categorization approach. During training the generation of as many extra feature channels (see Figure 3) as possible is stimulated, in order to capture as many variation and knowledge of the application as possible from the image data. This is no problem, since the boosting algorithm of the training will use all these features to determine which feature channels capture the most variation, in order to prune channels away and only keep the most descriptive feature channels. This immediately ensures that the algorithm won't become extremely slow during the actual detection phase because of the feature channel generation. By integrating all these extra feature channels into the actual object model training process, a better universal and more accurate object categorization framework will be supplied, which works very application specific to reach the highest performance and detection rate possible.

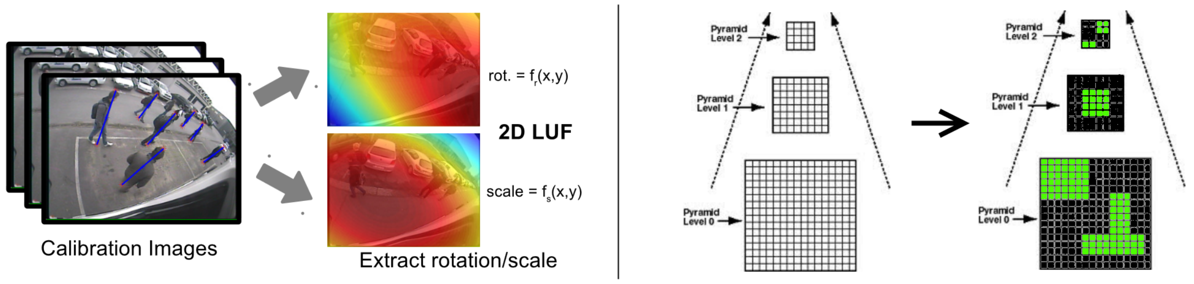

Figure 2: [Left] Example of a scale-location-rotation lookup function for pedestrians in a fixed and lens deformed camera setup [Right] Example of a fragmented scale space pyramid.

In subsection 5.1.1 the influence of the object scale and position in the image will be discussed. Subsection 5.1.2 discusses the influence of lighting, color and texture. Subsection 5.1.3 addresses the influence of background clutter and occlusion. Finally subsection 5.1.4 will handle the object rotation and orientation knowledge.

5.1.1 Influence of Object Scale and Position

In state-of-the-art object categorization an object model is trained, by rescaling all provided training images towards a fixed scale, which results into a single fixed scale model. Using a sliding window approach, with the window size equal to the size of the resulting model, an object detection is performed at each image position. However, there is only a limited number of applications that have fixed scale objects. In order to detect objects of different scales in all those other applications, an image scale space pyramid is generated. In this scale space pyramid the original image is down- and upsampled and used with the single scale model. This will generate the possibility to detect objects at different scales, depending on the amount of scales that are tested. The larger the pyramid, the more scales that will be tested but the longer the actual detection phase will take. Reducing this scale space pyramid effectively is a hot research topic. (Dollár et al., 2010) interpolates between several predefined images scales, while the detector of (Benenson et al., 2012) uses an approach that interpolates between different trained scales of the object model. These multiscale approaches are frequently used because the exact range of object scales is unknown beforehand in many applications.

However, many industrial applications have the advantage that the position of the complete camera setup is fixed and known beforehand (e.g. a camera mounted above a conveyor belt). Taking this knowledge into account, the scale and position of the objects can actually be computed and described fairly easy as seen in Figure 2). Using this information, new feature channels can be created. Based on manual annotation information, a 2D probability distribution can be produced over the image giving a relation between the scale and the position of the object in the image. (Van Beeck et al., 2012) discusses a warping window technique that uses a lookup function defining a fixed rotation and a fixed scale for each position in the image. However reducing the detection to a single scale for each position limits the intra-class variability that object categorization wants to maintain. To be sure this is not a problem, instead of using a fixed scale, a probability distribution of possible scales for each position can be modelled. The use of these distribution functions can lead to a serious reduction of the scale space pyramid, resulting in a fragmented scale space pyramid, as seen in Figure 2. This fragmented scale space pyramid can again be used as a seperate feature channel for object model training.

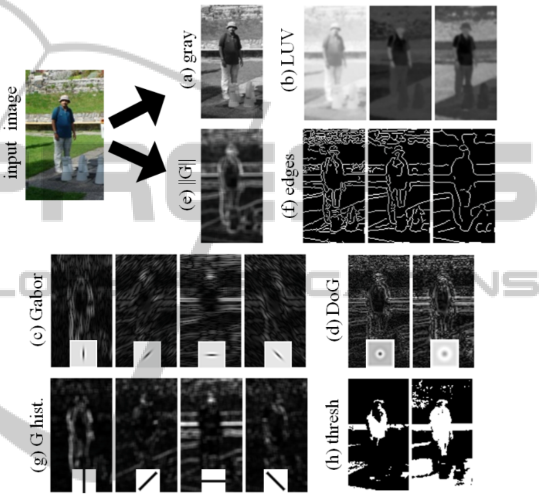

Figure 3: Example of different image channels used in the integral channel features approach of (Dollár et al., 2009). (a) Grayscale image (b) LUV color space (c) Gabor orientation filters (d) Difference of Gaussians (e) Gradient magnitude (f) Edge detector (g) Gradient histogram (h) Thresholded image.

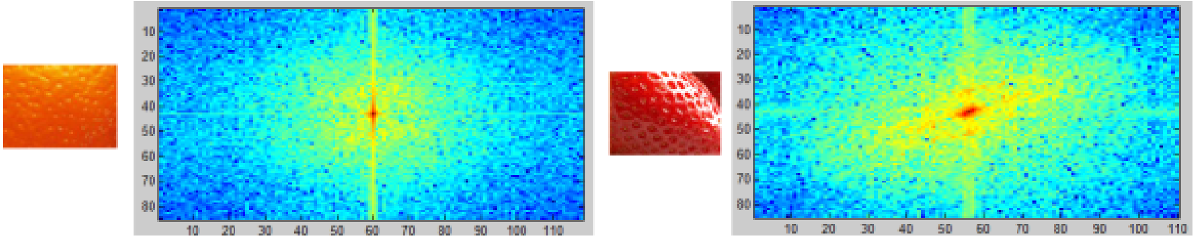

Figure 4: Texture variation based on the Fourier powerspectrum of an orange and a strawberry.

5.1.2 Influence of Lighting, Color and Texture

State-of-the-art object categorization ensures a certain robustness by making training samples and new input images invariant for color and lighting variations. To do so they use a color invariant image form, like a histogram of oriented gradient representation. Another possible approach is to use Haar-like features, like suggested by (Viola and Jones, 2001). Making the images invariant to lighting and color has a twofold reason. First of all the color variation in academic application is too large (e.g. the colors of clothing in pedestrian detection). On the other hand the color is too much influenced by the variation in lighting conditions. Therefore, academic applications try to remove as many of this variation as possible by applying techniques like histogram equalization and the use of gradient images.

By choosing a color and light invariant image form, all the information from the RGB color spectrum is lost which is in fact quite usefull in the industrial applications suggested by this PhD research. In many of these applications a uniform and constant lighting is used, leading to fixed color values. This information cannot be simply ignored when detecting objects with specific color properties like strawberries. The advantage of adding this color information has already been proven in (Dollár et al., 2009), where color information of the HSV and LUV space is added to optain a better and more robust pedestrian detector.

Besides focussing on the color information, it can be interesting to focus on multispectral color data. It is possible that objects cannot be seperated in the visual RGB color spectrum, but that there are higher multispectral frequency resolutions that make the seperations of objects and background rather easy. Academic research (Conaire et al., 2006; Yu et al., 2006; Shackelford and Davis, 2003) has already shown great interest in these multispectral approaches, where most of the applications are located in remote sensing and mobile mapping.

Another parameter that is not widely spread for object categorization is the use of relevant texture information in the training objects. Texture can be described as a unique returning pattern of gradients, which will almost never occur in the background information. In order te derive these patterns from the input data, techniques like Fourier transformations (Cant et al., 2013) (see Figure 4) and Gabor filters (Riaz et al., 2013) are used. These transformations show which frequencies are periodically returning in the image to define application and object specific textures.

5.1.3 Influence of Background Clutter and Occlusion

State-of-the-art object categorization approaches always attempt to detect objects *in the wild* which means that it can occur in every kind of situation, leading to an infinite number of possible background types, ... In order to build a detector that is robust to all this scene background variation, an enormous amount of negative images samples is needed during model training. This is required to try to model the background variation for correct classification and to ensure that the actual object model will not train background information. Besides that, it is necessary to collect as much positive examples as possible in those varying environments. Doing so ensures that only object features get selected that describe the object unrelated to the background behind it. This variation in the background is referred to as clutter.

Many industrial applications however have a known background, or at least a background with minimal variation. Combined with occlusion, where the object is partially or completely covered, clutter seems to happen much less frequent than in *in the wild* detection tasks. Take for example the taco's on the conveyor belt in Figure 5. The conveyor belt is moving and changes maybe slightly, but it stays quite constant during processing. Making a good model of that background information, can help to form an extra feature channel defining foreground and background information.

Other cases, like the picking of pears, will have much more variation in background, and will not give the possibility to simply aplly foreground-background

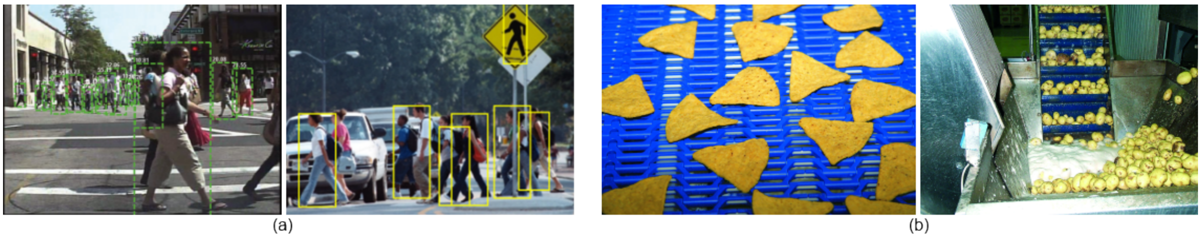

Figure 5: Example of background variation and occlusion in (a) academic cases and (b) industrial cases.

segmentation (see Figure 6).

A technique that is widely used for this kind of information is foreground-background segmentation, like in (Yeh et al., 2013). This technique helps us identify regions in the image that can be classified as foreground and thus regions of interest for possible object detections. The masks created by this segmentation can be applied as an extra feature channel. Using a dynamic adapting background model (Hammami et al., 2013), the application specific background will be modelled and a likelihood map of a region belonging to the foreground will be created. These are referred to as heat maps.

Due to the context of application specific algorithms, one can state that the only negative images that need to be used as negative training samples, are images that contain the possible backgrounds. This leads to the conclusion that many case specific object models can be reduced to having a very limited amount of negative training images, based on the applications scene and background variation, maybe even reducing the negative training images to a single image, if a static background occurs.

5.1.4 Influence of Rotation and Orientation

Most state-of-the-art object categorization approaches, e.g. detecting pedestrians, assume that there is no rotation of the actual object, since pedestrians always appear more or less upright. However this is not always the case, like shown in (Van Beeck et al., 2012), where pedestrians occur in other orientations due to the lens deformation and the birdseye viewpoint of the camera input.

Figure 6: Example of pear fruit in an orchard, where more background clutter and occlusion occurs.

Many industrial applications however contain different object orientations, which leads to problems when having a fixed orientation object model. Adding all possible orientations to the actual training data for a single model, will lead to a model that is less descriptive and which will generate tons of extra false positive detections. A second approach is to test all possible orientations, by taking a fixed angle step, rotating the input image and then trying the trained single orientation model. Once a detection is found, it can be coupled to the currect angle and then used to rotate the detection bounding box, like discussed in (Mittal et al., 2011). However, in order to reach realtime performance using this approach, a lot of GPU optimalizations will be needed, since the process of rotating and performing a detection on each patch is computationally intensive. A possible third approach trains a model for each orientation, as suggested in (Huang et al., 2005). However, this will lead to an increase of false positive detections.

The currently used approaches to cope with different orientations do not seem to be the best approaches possible. In this PhD research we want to create an automated orientation normalization step, where each patch is first put through a series of orientation filters that determine the orientation of the current patch and then rotates this patch towards a standard model orientation. A possible approach is the dominant gradient approach as illustrated in Figure 7. However, preliminary test results have shown that this approach doesn't work in every case. Therefore a combination of multiple orientation defining techniques will be suggested in our framework. Other techniques that can be included into this approach are eigenvalues of the covariance matrix (Kurz et al., 2013), calculating the geometric moments of a colour channel of the image (Leiva-Valenzuela and Aguilera, 2013) or even defining the primary axis of an ellipse fitted to foreground-background segmentation data (Ascenzi, 2013).

Our suggested orientation normalization filter will

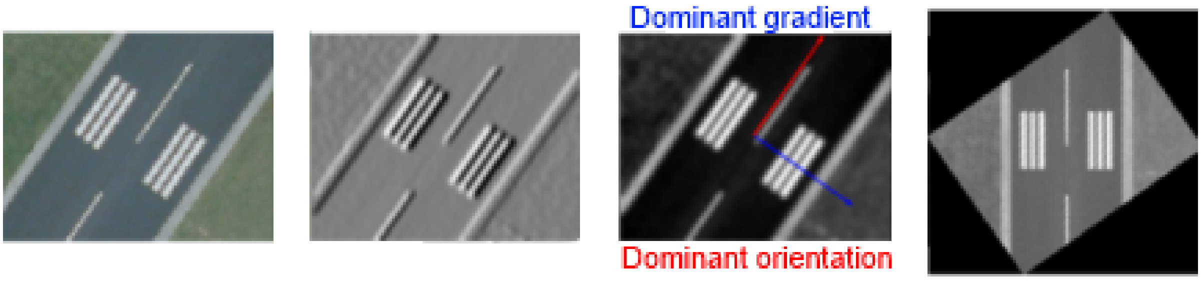

Figure 7: Example of rotation normalization using a dominant gradient technique. From left to right: original image (road marking), gradient image, dominant orientation and rotation corrected image.

use the combination of multiple orientation features to decide which one is the best candidate to actually define the patch orientation. In order to create this extra filter, all manual positive annotations are given an extra parameter, which is the object orientation of the training sample. From that data a mapping function is learned to define a pre-filter that can output a general orientation for any given window. Part of this general idea, where the definition of the orientation is seperated from the actual detection phase, is suggested in (Villamizar et al., 2010).

5.2 Innovative Active Learning Strategy for Minimal Manual Input

Limited scene and object variation can be used to put restrictions on the detector, by supplying extra feature channels to the algorithm framework, as previously explained. However, we will take it one step further. The same information will be used to optimize the complete training process and to drastically reduce the actual amount of training data that is needed for a robust detector. For state-of-the-art object categorization algorithms, the most important way to obtain a detector with a high detection rate is increasing the amount of positive and negative training samples enormously. The idea behind it is simple, if you add a lot of extra images, you are bound to have those specific examples that lie close to the decision boundary and that are actually needed to make an even better detector. However, since several industrial applications have a smaller range of variation, it should be possible to create an active learning strategy based on this limited scene and object variation, that succeeds in getting a high detection rate with as less examples as possible, by using the variation knowledge to look for those specific examples close to the decision boundary.

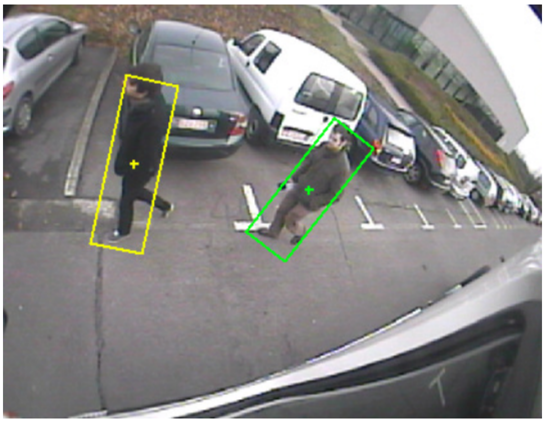

Figure 8: Example of viewpoint and lens deformation, changing the natural orientation of objects. (Van Beeck et al., 2012).

Like described in the conclusion of (Mathias et al., 2013), using immense numbers of training samples is currently the only way to reaching the highest possible detection rates. Since all these images need to be manually annotated, which is very time consuming job, this extra training data is a large extra cost for industrial applications. Knowing that the industry wants to focus more and more on flexible automatization of several processes, this extra effort to reach high detection rates is a large downside to current object categorization techniques, since companies do not have the time to invest all this manual annotation work. The industry wants to retrieve a robust object model as fast as possible, in order to start using the detector in the actual detection process.

5.2.1 Quantization of Existing Scene and Object Variation

In order to guarantee that the suggested active learning approach will work, it is necessary to have a good quantization of the actual variation in object and scene. These measurements are needed to define if new samples are interesting enough to add as extra training data. The main focus is to define how much intra-class variation there is, compared to the amount of variation in the background. Many of these variations, like scale, position, color, ... can be expressed

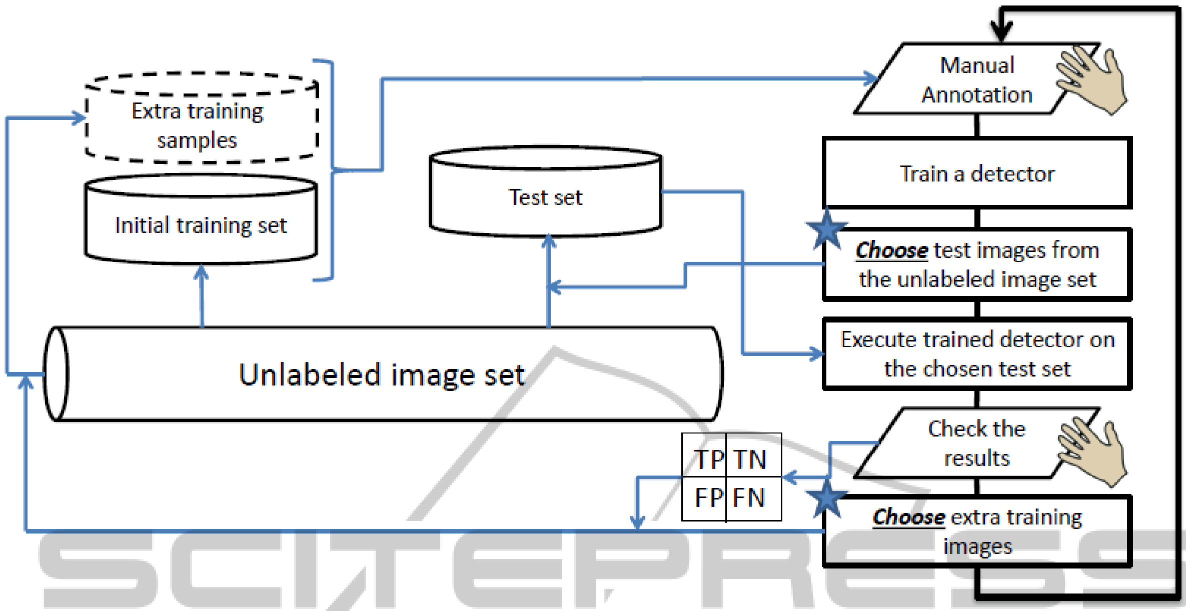

Figure 9: Workflow of the suggested active learning strategy. [☝ = manual input, ★ = knowledge of scene and object variation is used, TP = true positive detection, TN = true negative detection, FP = false positive detection, FN = false negative detection].

by using a simple 1D probability distribution over all different training samples. However, some variations are a lot harder to quantize correctly. If it is important to guarantee the intra-class variability, then it can even be extended to a 2D probability distribution, to allow multiple values for a single point in the distribution. However, features like texture and background variation cannot be modelled with a simple 1D probability distribution. A main part of the PhD research will thus go into investigating this specific problem and trying to come up with good quantizations for all these scene and object variations.

5.2.2 Active Learning During Object Model Training

Initial tests have shown that it is possible to build robust object detectors by using only a very limited set of data, as long as the training data is chosen based on application specific knowledge. However, figuring out which examples are actually needed, sometimes turns out to be more time consuming than just simply labeling large batches of training data, if the process is not automated. Therefore we suggest using an active learning strategy which should make the actual training phase more simple and more interactive. Eventually the algorithm optimizes two aspects: first being a minimal manual intervention and second an as high as possible detection rate. This research will be the first of its kind to integrate the object and scene variation into the actual active learning process, combining many sources of scene and object specific knowledge to select new samples, that can then be annotated in a smart and interactive way.

Figure 9 shows how the suggested active learning strategy based on application specific scene and object variation should look like. As a start a limited set of training data should be selected from a large database of unlabeled images. Since capturing many input images is not the problem in most cases, the largest problem lies in annoting the complete set, which is very time consuming. Once this initial set of data is selected, they are given to the user for annotation and a temporarily object model is trained using this limited set of samples. After the training a set of test images is smartly selected from the database using the scene and object variations that are available. By counting the true positives, false positives, true negatives and false negatives, the detector performance is validated on this test data, by manually supervising the output of the initial detector. Based on this output and the knowledge of the variation distributions in the current images, an extra set of training images is selected cleverly. The pure manual annotation is now splitted into a part where the operator needs to annotate a small set of images, but after the detection step, needs to validate the detections in order to compute the correctness of the detection output. This process is iteratively repeated until the desired detection rate is reached and a final object model is trained.

The above described innovative active learning strategy will yield the possibility to make a well fundamented guess on how many positive and negative training samples there will actually be needed to reach a predefined detection rate. In doing this, the approach will drastically reduce the amount of manual annotations that need to be provided, since it will only propose to annotate new samples that actually improve the detector. Training images that describe frequently occuring situations, and are thus being classified as objects with a high certainty are not interesting in this case. On the contrary, it will be more interesting trying to select those positive and negative training samples that lie very close to the decision boundary, in order to make sure that the boundary will be more stable, more supported by good examples and thus leading to higher detection rates.

It is important to mention that classic active learning strategies are often quite sensitive to outliers (Aggarwal, 2013) that get selected in the learning process and that lead to overfitting of the training data. However by adding multiple sources of information, being different application specific scene and object variations, the problem of single outliers can be countered, since their influence on the overal data distribution will be minimal. The suggested approach will filter out these outliers quite effectively, making sure that the resulting detector model will not overfit to the actual training set.

5.3 CPU and GPU Optimalization Towards a Realtime Object Categorization Algorithm

Once the universal object categorization framework, combined with an innovative active learning strategy, will be finished it will produce a better and more accurate detection system for industrial applications and in general, for all applications where the variation in scene and/or object is somehow limited. However expanding a framework to cope with all these application specific scene and object variations will lead to more internal functionality. This will result in a computationally more expensive and thus a slower running algorithm.

Since real time processing is essential for most industrial applications, this problem cannot be simply ignored. The longer the training of a specific object model takes, the more time a company invests in configuration and not in the actual detection process that generates a cash flow. This is why during this PhD research each step of the processing will be optimized using CPU and GPU optimalization. Classical approaches like parallelization and the use of multicore CPU's can improve the process (De Smedt et al., 2013), while the influence of general purpose graphical processing units (GPGPU) will also be investigated. The CUDA language will be used to implement these GPU optimalizations, but the possibility of using OpenCL will be considered.

6 EXPECTED OUTCOME

At the end of this PhD research a complete new innovative object categorization framework will be available that uses industrial application specific object and scene constraints, in order to obtain an accurate and high detection rate of 99.9% or higher. The result will be a stimulation for the industry to actively use this technology for robust object detection. The research will lead to new insights in general for object detection techniques. If this is proved to be successfull, the same approach will be introduced in other frameworks like the deformable parts model of (Felzenszwalb et al., 2010), to reach higher performances without increasing the number of training examples.

ACKNOWLEDGEMENTS

This work is supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT) via the IWT-TETRA project *TOBCAT: Industrial Applications of Object Categorization Techniques*.

REFERENCES

Aggarwal, C. C. (2013). Supervised outlier detection. In *Outlier Analysis*, pages 169–198. Springer.

Ascenzi, M.-G. (2013). Determining orientation of cilia in connective tissue. US Patent 8,345,946.

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

Benenson, R., Mathias, M., Timofte, R., and Van Gool, L. (2012). Pedestrian detection at 100 frames per second. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2903–2910. IEEE.

Benenson, R., Mathias, M., Tuytelaars, T., and Van Gool, L. (2013). Seeking the strongest rigid detector. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.

Cant, R., Langensiepen, C. S., and Rhodes, D. (2013). Fourier texture filtering. In *UKSim*, pages 123–128.

Conaire, C. O., O'Connor, N. E., Cooke, E., and Smeaton, A. F. (2006). Multispectral object segmentation and retrieval in surveillance video. In *Image Processing,* *2006 IEEE International Conference on*, pages 2381–2384. IEEE.

De Smedt, F., Van Beeck, K., Tuytelaars, T., and Goedemé, T. (2013). Pedestrian detection at warp speed: Exceeding 500 detections per second. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.

Dollár, P., Belongie, S., and Perona, P. (2010). The fastest pedestrian detector in the west. In *BMVC*, volume 2-3, page 7.

Dollár, P., Tu, Z., Perona, P., and Belongie, S. (2009). Integral channel features. In *BMVC*, volume 2-4, page 5.

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

Freund, Y., Schapire, R., and Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612.

Gall, J. and Lempitsky, V. (2013). Class-specific hough forests for object detection. In *Decision Forests for Computer Vision and Medical Image Analysis*, pages 143–157. Springer.

Hammami, M., Jarraya, S. K., and Ben-Abdallah, H. (2013). On line background modeling for moving object segmentation in dynamic scenes. *Multimedia Tools and Applications*, pages 1–28.

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

Huang, C., Ai, H., Li, Y., and Lao, S. (2005). Vector boosting for rotation invariant multi-view face detection. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 446–453. IEEE.

Kapoor, A., Grauman, K., Urtasun, R., and Darrell, T. (2007). Active learning with gaussian processes for object categorization. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE.

Kurz, G., Gilitschenski, I., Julier, S., and Hanebeck, U. D. (2013). Recursive estimation of orientation based on the bingham distribution. *arXiv preprint arXiv:1304.8019*.

Leiva-Valenzuela, G. A. and Aguilera, J. M. (2013). Automatic detection of orientation and diseases in blueberries using image analysis to improve their postharvest storage quality. *Food Control*.

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

Li, X. and Guo, Y. (2013). Adaptive active learning for image classification. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Mathias, M., Timofte, R., Benenson, R., and Gool, L. (2013). Traffic sign recognition-how far are we from the solution? In *Proceedings of IEEE International Joint Conference on Neural Networks*.

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

Mittal, A., Zisserman, A., and Torr, P. (2011). Hand detection using multiple proposals. *BMVC 2011*.

Puttemans, S. and Goedemé, T. (2013). How to exploit scene constraints to improve object categorization algorithms for industrial applications? In *Proceedings of the international conference on computer vision theory and applications (VISAPP 2013)*, volume 1, pages 827–830.

Riaz, F., Hassan, A., Rehman, S., and Qamar, U. (2013). Texture classification using rotation-and scale-invariant gabor texture features. *IEEE Signal Processing Letters*.

Shackelford, A. K. and Davis, C. H. (2003). A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. *Geoscience and Remote Sensing, IEEE Transactions on*, 41(10):2354–2363.

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.

Villamizar, M., Moreno-Noguer, F., Andrade-Cetto, J., and Sanfeliu, A. (2010). Efficient rotation invariant object detection using boosted random ferns. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1038–1045. IEEE.

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

Yeh, C.-H., Lin, C.-Y., Muchtar, K., and Kang, L.-W. (2013). Real-time background modeling based on a multi-level texture description. *Information Sciences*.

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., and Schirokauer, D. (2006). Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing*, 72(7):799.