

SIFTing the Relevant from the Irrelevant: Automatically Detecting Objects in Training Images

Edmond Zhang, Michael Mayo
 Department of Computer Science
 The University of Waikato
 Hamilton, New Zealand
 {ez1, mmayo}@cs.waikato.ac.nz

Abstract—Many state-of-the-art object recognition systems rely on identifying the location of objects in images, in order to better learn its visual attributes. In this paper, we propose four simple yet powerful hybrid ROI detection methods (combining both local and global features), based on frequently occurring keypoints. We show that our methods demonstrate competitive performance in two different types of datasets, the Caltech101 dataset and the GRAZ-02 dataset, where the pairs of keypoint bounding box method achieved the best accuracies overall.

Keywords—Image Processing; SIFT Keypoints; Image Recognition and Categorization; ROI Detection

I. INTRODUCTION

Generic object categorization is a challenging problem in computer vision. Given an arbitrary image, we would like to classify it according to the objects that can be detected and recognised – a task that is so natural and effortless for the human visual system, but proven to be difficult for current computer vision algorithms. This is mainly due to variability, and the need to generalize across variations in the appearance of objects belonging to the same class [11][12]. More specifically, this paper focuses on the problem of determining the region-of-interest (ROI) from images. We argue that by ‘homing in’ on the object of interest, we will be able to better describe and learn its visual attributes, while eliminating background noise from the images in which the objects are detected. Figure 1 depicts an example of ROI detection.



Figure 1. The entire image is used to extract visual features in A. Only ROI are used to extract visual features in B.

Recently, many approaches using machine learning techniques have been proposed, such as [13]. Broadly speaking,

the machine learning approach requires conversion of an image to feature vector. This task of generating feature vectors can be done either with local or global image features. The global approach is easy to implement and inexpensive to compute. However, one inherent disadvantage is that it is susceptible to local and global variation (e.g. changes in viewpoints or illumination). Local features, on the other hand, provide a better foundation to handle local and global variability such as various forms of transformation (e.g. affine, scale, rotation). However, one of the fundamental drawbacks of the local approach is that in order to include spatial relationships between local features, they must be modelled explicitly [14][15].

Although many important recent studies [12][16][17][18] have focused on the local approach, in this paper, we propose a ‘hybrid’ approach that is based on both local and global features. More specifically, local features are used to determine the frequently occurring visual attributes for object classes, before aggregating statistical information over not the entire image but rather a specific subregion that is detected, the region-of-interest (ROI). The reasoning for this approach is two fold. Firstly, we believe that the category of an image can be described reliably by low-dimensional global features, as demonstrated in [19], where spectral and coarsely localised information is used to provide a meaningful description of the image and its semantic category. Secondly, unlike local features, global features are inexpensive to compute, which is essential for large datasets (e.g. Caltech101 [7]).

The motivation of this research is that detecting ROIs in training images should increase accuracy of classifiers built with those ROIs rather than entire image. The reasoning is that for all the images belonging to the same object class, a spatially localized subset of visual features will occur in most of the images. For example, features from wheels and tyres from images belong to a car object class will tend to occur in one region of an image. Background clutter and image noise will also occur in all images, but they will be different across different images.

In order to locate informative image features, we developed three methods for automatically detecting ROIs from

training images. Unlike the popular sliding window object localization methods, our methods only need to deal with a small number of frequent keypoints in each of the images. Our method builds on the theory that frequently occurring keypoints are unique and informative, in representing specific object classes.

The rest of this paper is organized in the following order. In the background section, we will discuss the sliding frame approaches, along with some background on how most frequent keypoints are determined. In the algorithms for ROI detection section, we present our methods for detecting ROI. We then show the results of the two datasets used for our experiments, in the Experiments section. In the discussion section, we explain the results of the experiments, followed by the conclusion section.

II. BACKGROUND

This section is divided up into four parts. In the first part, we will discuss the popular sliding window approach for detection ROI. We then discuss the SIFT keypoints in the second part of this section. This will be followed by an explanation of how frequent keypoints are produced. Finally, the last part of this section briefly examines the PHoG descriptor.

A. Sliding Window for Object Localization

Object localization with bounding boxes, based on the sliding window technique has been popular recently [20][21][22][23]. Broadly speaking, the sliding window method works by first dividing an image into smaller patches, then treating object localization as localized detection. In other words, a classifier is applied to all sub images within an image and takes the maximum of the classification score as indication for the presence of an object in this region [24].

One inherent disadvantage of this approach is the significant increase in computational cost, because of the large number of candidate subimages. Moreover, the number of subimages grow in the order of n^4 for images of size $n \times n$, which is computationally too expensive when dealing with large datasets.

Lampert *et al*, in [24], stated that in order to speed the search, several heuristics methods have been proposed. These approaches can be grouped into two categories. The first category [7][11][19], consists of reducing the number of necessary function evaluations by searching only over a coarse grid of possible rectangle locations and by allowing only limited shapes and sizes as candidates. The second category [3][6] utilizes local optimization methods by applying them locally. In that, promising regions in the image are identified then maximized by a discrete gradient ascent procedure from there. However, these speed-up approaches severely limited localization robustness in order to achieve acceptable speed.

B. SIFT Keypoints

The SIFT (Scale Invariant Feature Transform) descriptor is based on grey scale representation of images. SIFT features are essentially local histograms of gradient directions computed over different parts of the keypoint region. It computes the gradient vector for each pixel in the keypoint's 4×4 pixel neighbourhood and builds a normalised orientation histogram (default bin size is 8) of gradient directions. The size of the feature vector is thus 128 (that being $4 \times 4 \times 8$) attributes. Orientation invariance is attained by estimating the dominant orientation of the local image patch and is normalized to enhance invariance to changes in illumination. The process of generating keypoints can be summed up in two steps: feature detection and feature extraction.

1) *Feature Detection*: Keypoint detection can be simply seen as the process of locating salient points and/or regions from images, in order to produce useful local image descriptors. Keypoints in [11] are defined as the special points that survive longest when gradually blurring the image in scale space. Rosin [25] also pointed out that image lifetime is an important selection criterion for interest points as well as local image features such as wiggleness, spatial width, and phase congruency. Other important attributes of a reliable and meaningful interest point are that it must be invariant to image transformation, such as scale, rotation, and affine transform, in addition to perspective transformation, illumination and brightness variations.

2) *Feature Extraction*: Local descriptors are computed on the region around each keypoint detected by the keypoint detection process. There are several sources of information at the local level that local descriptors can be based on, such as gradient orientation, size, and point of origin. Lowe, in [1] defined Local Invariant Feature Descriptors as the local image gradients measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination. Our definition of local invariant feature descriptors is defined simply as the fingerprint of images, where these fingerprints are essential for the task of object recognition. Generally, the primary role of descriptors is to identify and extract local features around the keypoints within the image. The extracted features must be highly distinctive, have a low probability of mismatch, should be tolerant to image noise, changes in illumination, uniform scaling, rotation, and minor changes in viewing direction.

C. The PHoG Descriptor

The PHoG descriptor [6] is the spatial representation of the HoG descriptor [4] based on the spatial pyramid scheme proposed by Lazebnik *et al* [5]. The HoG descriptor, in essence, represents local shapes by histograms of edge orientations gradients within an image subregion quantized into M bins. Each bin in the histogram represents the

number of edges that have orientations within a certain angular range. This representation can be compared to the traditional bags of words; where each visual word is a quantization on edge orientations.

Based on the shape presentation of the HoG descriptor, spatial properties of images can be better captured by combining the spatial pyramid scheme. In that, spatial pyramids repeatedly subdivide an image, computing all features repeatedly for all progressively smaller sub-images. The first image is always the global image, and then the image is divided into 2×2 sub-images, and features are computed from each of those. The image may then be further subdivided, this time 4×4 regions, and so on. For a spatial pyramid with l levels, the maximum granularity will be a division of an image into $2^{l1} \times 2^{l-1}$ sub-images. Figure 2 illustrates a division of a beaver image into sub-images for an $l = 3$ pyramid, then the PHoG is computed [6].

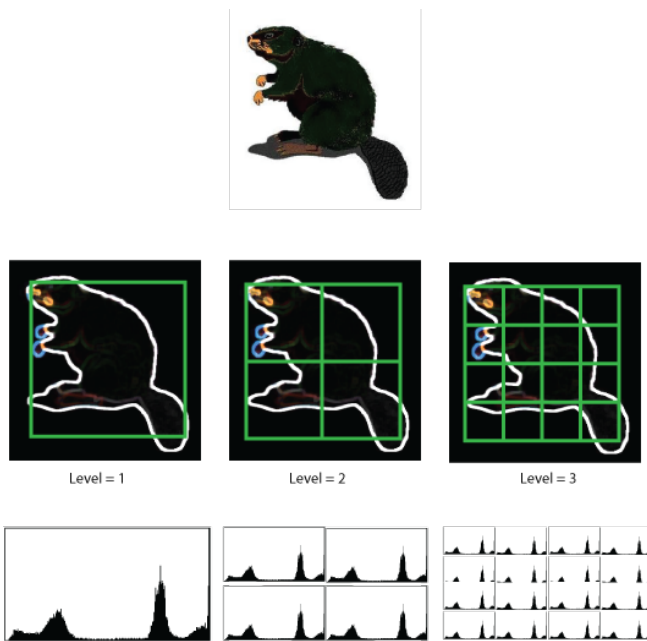


Figure 2. The PHoG descriptor.

III. ALGORITHMS FOR ROI DETECTION

In this section we will discuss the algorithms we developed for detecting Region-of-interest (ROI), from training images, in order to improve recognition performance. Our four algorithms, single keypoint patch selection (Algorithm A), single keypoint bounding box (Algorithm B), pairs of keypoint patch selection (Algorithm C), and pairs of keypoint bounding box (Algorithm D), rely on the use of frequent keypoints for locating the ROI from images.

A. Frequent Keypoint Selection

A large portion of features (keypoints) detected and described by descriptors such as the SIFT descriptor [1] are

not useful, as these mainly consist of background clutter and image noise. Using all the keypoints of the whole image for classification can lead to a very high computational complexity and severely hamper recognition performance. In order to select only the most informative image features for classification, k-means [3] clustering was used to determine the most frequent image features.

The simple iterative square-error partition method, k-means was used for grouping similar keypoints into clusters. This algorithm proceeds by iterated assignments of keypoints to their closest cluster centres and re-computation of the cluster centres. One of the shortcomings of the k-means clustering algorithm is that the parameter K , for determining the initial number of clusters, must be set manually.

Because SIFT keypoints are high dimensional (128 dimensions), direct 1-to-1, single digit comparison between cluster centres and keypoints is not possible. Consequently, we have tried several different high dimensional distance functions, and we found that the X^2 distance [3] is the fastest and also has the highest recall rate (Equation (1) depicts the X^2 distance function).

$$d_{X^2}(H, P) = \sum_{i=0..127} \frac{(h_i - m_i)^2}{m_i} \quad (1)$$

where:

$$m_i = \frac{h_i + p_i}{2}$$

In Equation 1, the 128-dimensional vectors H and P represent two keypoints, and h_i and p_i are the i^{th} element or bin of each keypoint respectively. If the two features are identical, then the X^2 distance between them is 0. However, the chance of finding two identical keypoint matches is extremely low.

Once the ROI is located from images, the pyramid representation of the Histogram of gradients (PHoG) is used to describe that region of the image. Attributes extracted from the descriptors are combined to form feature vectors - where attributes are simply concatenated together one next to each other. This process takes advantage of the abstraction provided by the feature vector representation of input data that enables the use of numerous domain-independent classifiers.

B. Single Keypoint Image Patch Selection

For our first method, we first divide the image into smaller patches. Each image patch is then tested for the number of single or pairs of keypoints existing within it. If the number of keypoints is greater than the parameter X , then the patch is considered to be informative and will be kept. Otherwise, the image patch will be discarded by blanking out. Figure 3 illustrates an example of image patch selection method.

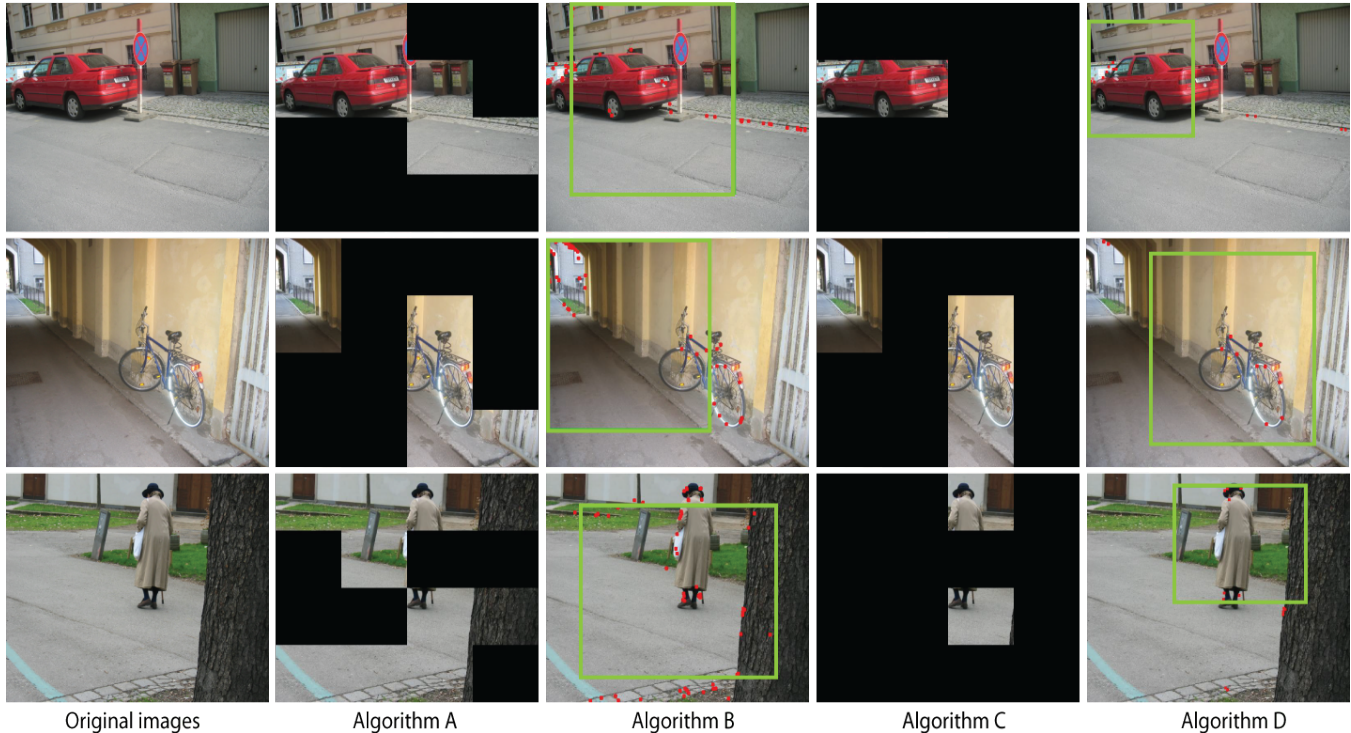


Figure 3. Example of single keypoint patch selection (Algorithm A), single keypoint bounding box (Algorithm B), pairs of keypoint patch selection (Algorithm C), and Pairs of keypoint bounding box (Algorithm D).

C. Single Keypoint Bounding Box

The second method creates a bounding box around the ROI. Everything inside of the box is kept for feature extraction, while everything else will be discarded. Figure 3 depict examples of the single keypoint bounding box method.

The placement of the bounding box is determined by the centre-of-mass of all the frequent keypoints found in the image. This approach is made possible due to the fact that the majority of the frequent keypoints are either on or near the object of interest. Figure 4 illustrates an example comparison between frequent keypoints and all keypoints.

Initially, the size of the bounding box is fixed at 60% of the original image size. However, because the size of objects varies greatly across different images, we developed a simple method depicted in Algorithm 1 to automate the bounding box size selection. The available sizes are 15%, 30% and 60% of the original image size.

See Figure 5 for an example comparison between fixed and variable bounding boxes.

D. Pairs of Keypoint Patch Selection

The third method is similar to the first method. Namely, the image is divided into smaller patches and the total number of keypoints exist in the patch determines whether that patch is kept or discarded. The only different to the first method is that pairs of frequent keypoints are used, instead of single frequent keypoints.

Algorithm 1 The single keypoint bounding box method.

Input: An image and its set of frequent keypoints

- 1: Start with the smallest bounding box, which is at 15% of the original image size. Place the box around the centre-of-mass of the frequent keypoints.
- 2: **if** the number of keypoints found inside of this area is greater than or equal to 90% of the total number of keypoints found in the image **then**
- 3: The current bounding box will be the final bounding box. Exit Algorithm.
- 4: **else**
- 5: Increase the size of bounding box.
- 6: **end if**
- 7: Repeat step 2 until no more bigger bounding boxes to test.

Output: An image with a bounding box

Zhang and Mayo [2] argued that single keypoints are not unique enough to represent object classes by themselves independently. However, pairs of frequently occurring keypoints are more distinctive and efficient in representing object classes, due to the spatial properties they carry. Instead of computing the pairs of frequent keypoints from all available keypoints, only frequent keypoints are used. This approach significantly reduces background clutter and image



Figure 4. All the keypoints are displayed on the top image, while only the frequent keypoints are displayed on the bottom image. (Images best viewed in colour)

noise. Algorithm 2 shows the steps in frequent keypoint pairs generation.

E. Pairs of Keypoint Bounding Box

Our final method is similar to the second method, in that, a bounding box of variable size is placed on the image. The only difference is that pairs of frequent keypoints are used to determining the placement of the bounding box, instead of the single keypoints.

IV. EVALUATION

We have experimented our ROI detection methods on two benchmark datasets: GRAZ-02 [8] and Caltech101 [7]. We report the experiment setup and results in this section. Multi-class classification is done with quadratic SVM classifier and the SMO learning algorithm, with default parameters as specified in WEKA V. 3.5.5 [9]. The number of clusters

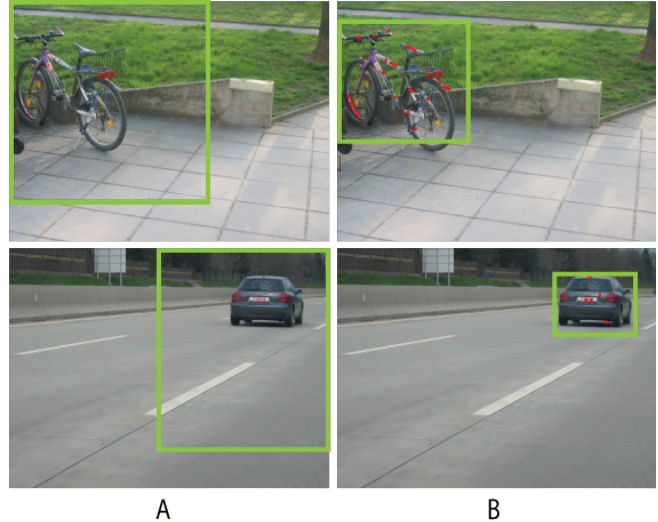


Figure 5. The difference between manual(A, 60%) and automatic(B) bounding box size selection.

Algorithm 2 The pairs of frequent keypoints discovery method.

Input: An image and its set of frequent keypoints

- 1: Traverse through all frequent keypoints extracted from each training image.
- 2: Determine all the frequent keypoints that are within a pre-defined radius of the currently selected frequent keypoint.
- 3: Generate unique pairs of keypoint patterns from the set of selected keypoints from step 2
- 4: Keypoint pairs are then ranked from the most frequent to less frequent.
- 5: Only the top X pairs are selected for each object class.

Output: An image with a bounding box

is set at 300, and only the top 5 clusters are tagged as frequently occurring features. We then select only the top N number ($N = 50$) of most frequent patterns from each object class because they are the most distinctive and informative. We experimented with various different radius sizes ranging from 10 pixels to the entire image; we found that the radius of 50 pixels is the best for both the datasets. Figure 3-D illustrates the example output produced by this method. For Algorithm 1 and 3, the parameter X, which is the threshold for the number of keypoints per image patch, is set at 3.

In order to see the effects of ROI detection in terms of overall performance, for each of the datasets, we first experimented with the PHoG descriptor on the whole image, before applying the PHoG descriptor on ROI only. ROI detection is applied on training images only. Test images are not altered in any manner.

This section is divided into three parts: Datasets, Exper-

iments and Discussion. In the Datasets section, we briefly present the two datasets used for our evaluations. In the Experiments section, we explain the experimental setup. Finally, we discuss our results in the Discuss section.

A. Datasets

1) *Caltech101*: The first dataset we experimented with was the Caltech101 dataset. This is probably one of the most diverse datasets in the research community. However, this is an easy dataset for detecting the ROI because the majority of objects are located in the middle of the image. Each object class contains between 31 and 800 images. The resolution for most of the images is about 300 by 300 pixels. Figure 6 illustrates some examples of this dataset.

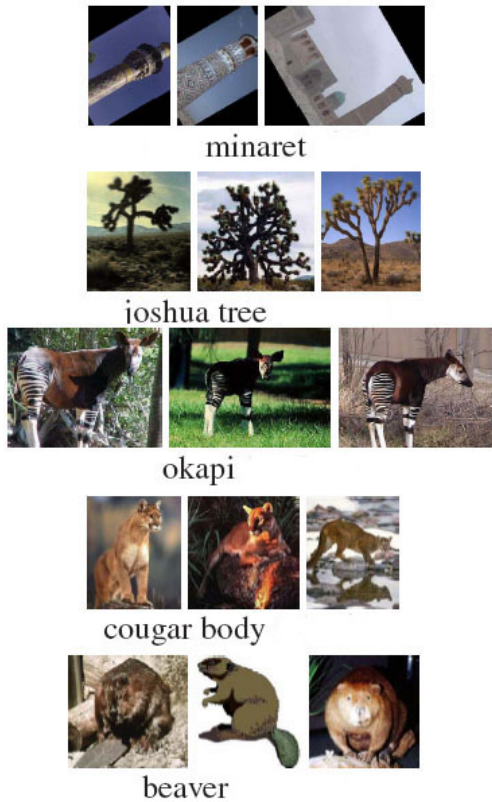


Figure 6. The caltech101 dataset.

2) *GRAZ-02*: The second dataset we experimented on was the GRAZ-02 dataset. The dataset contains four categories: Bike, Person, Cars and background as counterexamples. Meaning it contains no bike, person or cars. This dataset is much more complex than the Caltech101 dataset in terms of intra-class variation, such as illumination, scale, pose, viewing angle, occlusion, and clutter. Figure 7 illustrates some examples of this dataset.

B. Experiments

All four of our algorithms were experimented on the two datasets. For the Caltech101 dataset, we follow the



Figure 7. The GRAZ-02 dataset.

experimental setup of J. Zhang et al. [10]. Specifically, 30 images per class are used for training and the rest are tagged as test images. Experiments are repeated 10 times with randomly selected training and test images. J. Zhang et al. achieved the state-of-the-art for this dataset, which is at 66.2% for 30 training images. Our best method obtained 61.34% with 30 training images. See Table I for results produced from our methods.

For the GRAZ-02 dataset, we follow the experimental setup of Opelt [8]. Namely, we took a training set consisting of 150 images of the object category as positive images and 150 of the counter-class as negative images. The tests were carried out on 300 images half belonging to the category and half not. Table II shows the categorization results measured in ROC-equal-error rates of various specific ROI detection techniques.

C. Discussion

Results obtained from experiments on the two datasets were comparable to those of the original authors. Consistently, the bounding box methods obtained better results than the patch selection method. The main reason for the difference in the two performances, we believe, is that with the patch selection method, often only parts of the objects are selected because some parts of the object do not contain any keypoints. However, for the bounding box methods, we were able to capture the object as a whole more frequently. It has also shown that the pairs of keypoints bounding box method was able to home in on the object of interest more accurately than the single keypoint bounding box method, resulting in better overall performances.

The Caltech101 dataset is considerably easier for the problem of ROI detection. In that, objects of interest are nearly all located in the centre of the image and occupying large portion of the image. The GRAZ-02 dataset, however,

Table I
RESULTS FOR THE CALTECH101 DATASET.

Zhang et al.	No ROI	Algorithm A	Algorithm B	Algorithm C	Algorithm D
66.2	50.18±1.6	59.23±2.6	61.45±3.2	58.45±1.9	61.34±2.3

Table II
RESULTS FOR THE GRAZ-02 DATASET.

	Opelt et al.	No ROI	Algorithm A	Algorithm B	Algorithm C	Algorithm D
Bike	76.5	82.1	85.2	90.6	84.1	91.4
Cars	81	72	73.3	78.7	71.3	81.2
Person	70.2	65.2	65.2	67.3	65.3	69.8

are more challenging because objects are of various scales, different viewing angles, and the presence of occlusion, clutter and noise.

Recognition performances were obtained for both the datasets, with and without ROI detection, all using the same PHoG global image descriptor. The inclusion of the ROI detection procedure improved the overall performance. For the Caltech101 dataset, the difference was about 10%. Additionally, for the Bike class in the GRAZ-02 dataset, a significant 15% increase was realised.

Better recognition performances could be gained by using more descriptors, such as the PHoW descriptor, as shown in [6]. We will investigate these in future work.

V. CONCLUSION

In this paper, we proposed four algorithms for detection ROI for better object categorization performance. Our methods were experimented on two of the most popular datasets, one easy and one hard, with promising results.

Unlike the popular sliding window approaches, where classifiers have to be evaluated over a large set of candidate subimages, our methods rely only on frequently occurring keypoints for locating the ROI.

REFERENCES

- [1] Lowe, D. G. 1999. Object Recognition from Local Scale-Invariant Features. In Proceedings of the international Conference on Computer Vision-Volume 2 - Volume 2 (September 20 - 25, 1999). ICCV. IEEE Computer Society, Washington, DC, 1150.
- [2] Zhang, E. and Mayo, M. (2008). Pattern discovery for object categorization. In Proceeding of 23rd International Conference Image and Vision Computing New Zealand 2008(IVCNZ 2008).
- [3] O. Duda, P.E. Hart, D.G. Stork, Pattern classification, John Wiley & Sons, 2000.
- [4] Dalal, N. and Triggs, B. 2005. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cvpr'05) - Volume 1 - Volume 01 (June 20 - 26, 2005). CVPR. IEEE Computer Society, Washington, DC, 886-893. DOI= <http://dx.doi.org/10.1109/CVPR.2005.177>
- [5] Lazebnik, S., Schmid, C., and Ponce, J. 2006. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (June 17 - 22, 2006). CVPR. IEEE Computer Society, Washington, DC, 2169-2178. DOI=<http://dx.doi.org/10.1109/CVPR.2006.68>
- [6] A. Bosch, A. Zisserman, and X. Munoz. Image classification using random forests and ferns. In Proceedings of the 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007.
- [7] Wang, G., Zhang, Y., and Fei-Fei, L. 2006. Using Dependent Regions for Object Categorization in a Generative Framework. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (June 17 - 22, 2006). CVPR. IEEE Computer Society, Washington, DC, 1597-1604. DOI=<http://dx.doi.org/10.1109/CVPR.2006.324>
- [8] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer. Weak hypotheses and boosting for generic object detection and recognition. In Proc. ECCV, pages 7184, 2004.
- [9] Ian H. Witten, Eibe Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
- [10] Zhang, J., Marszaek, M., Lazebnik, S., and Schmid, C. 2007. Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study. Int. J. Comput. Vision 73, 2 (Jun. 2007), 213-238. DOI=<http://dx.doi.org/10.1007/s11263-006-9794-4>
- [11] Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. Intl. J. of Comp. Vision 43 (2001) 2944
- [12] C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csürka. Visual categorization with bags of keypoints. In ECCV International Workshop on Statistical Learning in Computer Vision, 2004. 4
- [13] Ilkay Ulusoy and Christopher M. Bishop. Generative versus discriminative methods for object recognition. In CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2, pages 258265, Washington, DC, USA, 2005. IEEE Computer Society.

- [14] R. Fergus, P. Perona, A. Zisserman, O. P. U. K. Object class recognition by unsupervised scale-invariant learning, In CVPR, 2003. 264271
- [15] Daniel Keysers, Christian Gollan and Hermann Ney. Classification of Medical Images Using Non-Linear Distortion Models. In *Bildverarbeitung für die Medizin*, 2004, 366 – 370, Springer-Verlag
- [16] Lazebnik, S., Schmid, C., Ponce, J. 2006. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (June 17 - 22, 2006)*. CVPR. IEEE Computer Society, Washington, DC, 2169-2178. DOI=<http://dx.doi.org/10.1109/CVPR.2006.68>
- [17] Wang, G., Zhang, Y., Fei-Fei, L. 2006. Using Dependent Regions for Object Categorization in a Generative Framework. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (June 17 - 22, 2006)*. CVPR. IEEE Computer Society, Washington, DC, 1597-1604. DOI=<http://dx.doi.org/10.1109/CVPR.2006.324>
- [18] Zhang, J., Marszaek, M., Lazebnik, S., Schmid, C. 2007. Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study. *Int. J. Comput. Vision* 73, 2 (Jun. 2007), 213-238. DOI=<http://dx.doi.org/10.1007/s11263-006-9794-4>
- [19] Oliva, A., Torralba, A. 2001. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *Int. J. Comput. Vision* 42, 3 (May. 2001), 145-175. DOI=<http://dx.doi.org/10.1023/A:1011139631724>
- [20] Bosch, A., Zisserman, A., Munoz, X. , Representing shape with a spatial pyramid kernel, *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, 401–408, ACM Press,2007
- [21] O. Chum, A. Zisserman. An Exemplar Model for Learning Object Classes. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, 2007.
- [22] N. Dalal, B. Triggs. Histograms of Oriented Gradients for Human Detection. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 886 - 893, 2005.
- [23] V. Ferrari, L. Fevrier, F. Jurie, C. Schmid. Groups of Adjacent Contour Segments for Object Detection. *PAMI*, 30:3651, 2008
- [24] Christoph H. Lampert, Matthew B. Blaschko, Thomas Hoffmann: "Beyond Sliding Windows: Object Localization by Efficient Subwindow Search", *IEEE Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, 2008
- [25] Paul L. Rosin. Edges: saliency measures and automatic thresholding. *Machine Vision and Applications*, 9(4):139159, 1997.