CONTEXTUAL DETECTION OF DRAWN SYMBOLS IN OLD MAPS

Jonathan Guyomard, Nicolas Thome, Matthieu Cord, Thierry Artières

UPMC Univ Paris 6, LIP6, 4 place Jussieu, 75005 Paris, France

ABSTRACT

In this paper, we tackle the problem of detecting drawn symbols in old maps. We propose a novel approach that combines powerful low level descriptors to represent the local content of the objects, and contextual features to overcome the local analysis ambiguity. Our contribution is two-fold. Firstly, we propose a novel contextual feature adapted to our problem, where the context is integrated at two levels. In a close neighborhood, a local analysis is carried out to remove visual ambiguities between symbols. In a larger extent, co-occurrence statistics between classes are stored. Secondly, we propose an entire processing chain for learning and detection. The proposed method is evaluated on real french maps from the 18^{th} century. The experiments show the efficiency of the detection system, and validate the relevance of the proposed contextual feature to improve detection performances.

***Index Terms*—** Object Detection, Contextual Features, Pattern Recognition

1. INTRODUCTION

Object detection in images is an important task in computer vision and has a huge area of applications. In this paper, we are interested in detecting symbols in old french maps drawn by Cassini in the 18^{th} century. This research are part of the ANR project GeoPeuple, in partnership with EHESS/LaDéHis [1] and IGN/COGIT [2]. Each map has been drawn manually, and 181 maps have been produced to cover the whole french territory (Fig 1). In that context, automatic object detection is mandatory to prevent the manual annotation in such a huge database. In line with the GeoPeuple [3] project, this automatic detection may help the creation of an historical geodatabase useful for demographers to study the French population evolution from 1750 to present, and more specifically the relationships between landscape and demography during 2 centuries.

In this work, we focus on the detection of 8 different classes of symbols presented in Fig 2. In natural images, state of the art approaches for object detection are based on sliding window strategies [1], that scan the whole image and use learned classifier to decide if a given image region contains the object of interest. We propose here to evaluate this sliding window mechanism in our old map database, and report promising results for this first level of detection. However, the geometric structure of some classes of symbols can be very simple, like calvary or watermill, which basically look like crosses and circles. In these situations, the visual local information is not enough discriminant to accurately classify the symbol against the large number of similar shapes in the background or in other symbols. Therefore, incorporating contextual feature is necessary to improve detection performances.

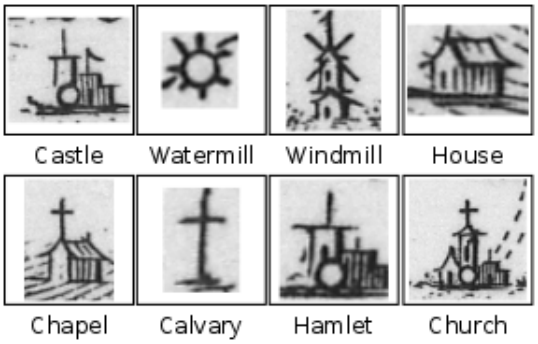


Fig. 2. Symbol examples for 8 classes. All the symbols are drawn by hand.

The literature about context modeling is vast, and a general review is outside the scope of the paper. Some methods [2, 3] use very local context based on pixel analysis. These methods are not directly connected to our problem, since we want to encode a contextual feature based on the output of an object detector. A good review about context for object detection is available in [4]. In a finer-grained analysis, the method on context for object detection can be classified, as proposed in [5], between Stuff-Thing, Stuff-Stuff, Thing-Thing. Another important criterion to distinguish methods is to consider if the context is defined a priori as in [6, 7, 8], or if it is learned from data [9, 10, 11]. We propose to learn a Thing-Thing contextual object detector. The approach the most connected to ours is the work of Felzenszwalb *et.al.* [10, 11]. We improve their contextual feature by incorporating spatial information, and use a second level of context encoding based on a co-occurrence statistics analysis.

In this paper, we present a new system to learn spatial contextual information and then use it for detection or to refine a visual classification. We propose a two-level system detailed in section 3: we use HOG features for the first level of the detection, and show that they offer a good description of the local appearance of old drawn map symbols. Then, the second level analyses the output of the first level classifier, extract contextual feature and provide a final decision to classify each image region. This process is detailed in section 2.

2. PROPOSED CONTEXTUAL DESCRIPTOR

As previously mentioned, running sliding window detectors independently inevitably leads to some misclassifications. Some obvious cases of visual local ambiguities are shown in Fig 3.a: for example, a calvary detector will happily fires at the top of churches. To solve this visual ambiguities, we propose to add, in a second detection

Acknowledgments: This research is funded by the ANR project GeoPeuple and map annotations are provided by IGN.

[1] École des Hautes Études en Sciences sociales

[2] Institut Gographique National

[3] http://geopeuple.ign.fr/

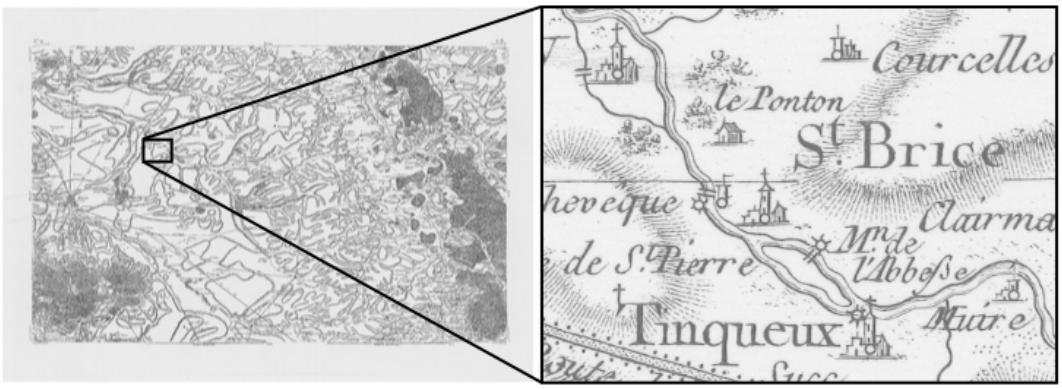
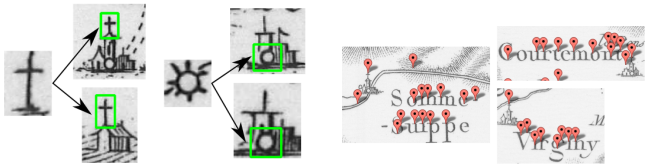


Fig. 1. Cassini digitized map with a zoom.

step, a contextual information. This is done by incorporating relative position information between classes in a very close neighborhood of the detection window, as described in section 2.1. Afterwards, we store statistics of co-occurrences between classes in a larger radius (section 2.2). As shown on Fig 3.b, there is always text near a symbol. Therefore, the fact that the calvary classifier often yields an alignment of good matching scores on text structure can be used as contextual cues. This spatial distribution around the first detection should help a lot to better identify symbols.



(a) (b)

Fig. 3. (a) Visual ambiguities between classes. (b) Calvary detection close to the considered symbol.

2.1. Small neighborhood: local context

Our extraction of the local context was inspired of the work of Felzenswalb in [10], but we enrich his contextual description with a position information.

We consider a region of the map where k classifiers $(C_1, ..., C_k)$ based on visual appearance were applied. The output of these classifiers give k sets of windows $(D_1, ..., D_k)$, where each window $d \in D_i$ is defined by its center $g = (g_x, g_y)$ and its classification score s_i. Let $w_j = (g_j, s_j), j \in [1, k]$ be a window that we want to classify and d_i^* be the highest scoring window of D_i in an area of radius d_{max} from g_w. We define our close context of a class i for w_j by:

$$n_i(w_j) = (\sigma(d_i^*), d_x, d_y)$$

where $\sigma(d_i^*)$ is defined by:

$$\sigma(d_i^*) = \begin{cases} d_i^*.s & \text{if } d_i^* \text{ exists} \\ -1 & \text{if no occurrence of the class } i \end{cases} \quad (1)$$

For the case $i = j$, d_i^* cannot overlap with w_j. d_x, d_y is a couple of values defining relative positions of d_i^* from w_j, normalized as $d_x, d_y \in [0, 1]$. The local context of w_j for the k classes is defined by the concatenation of close context of each class:

$$\mathcal{N}(w_j) = (s_j, n_1(w), ..., n_k(w))$$

2.2. Large neighborhood: global context

In a second time, we want to add the contextual information contained in a larger radius. We describe the more distant context by series of crowns, which are assigned for each statistic on occurrence of every class. Let $(R_1, ..., R_n)$ the n rings around the window w_j. The minimum radius of the first crown is d_{max}, and the maximum radius of the last crown is r_{max}. The thickness of each one is given by:

$$\frac{r_{max} - d_{max}}{n}$$

We define our large context, for a class i, of w_j by:

$$l_i(w_j, n) = \left(\sum_{d_i \in R_1} s_i, ..., \sum_{d_i \in R_n} s_i \right)$$

The global large context of w_j with k classes is defined by:

$$\mathcal{L}(w_j, n) = (l_1(w, n), ..., l_k(w, n))$$

Our final feature of a window w_j is then:

$$\mathcal{F}(w_j, n) = (\mathcal{N}(w_j), \mathcal{L}(w_j, n))$$

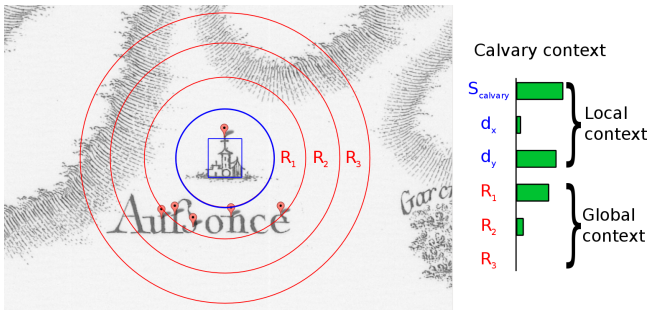


Fig. 4. Representation of the global context.

Our feature vector is L2 normalized. This descriptor possesses three hyper parameters. The maximum distance of the close context, that is also the minimum radius of the crowns d_{max}, the maximum radius of the global context r_{max} and the number of ring n.

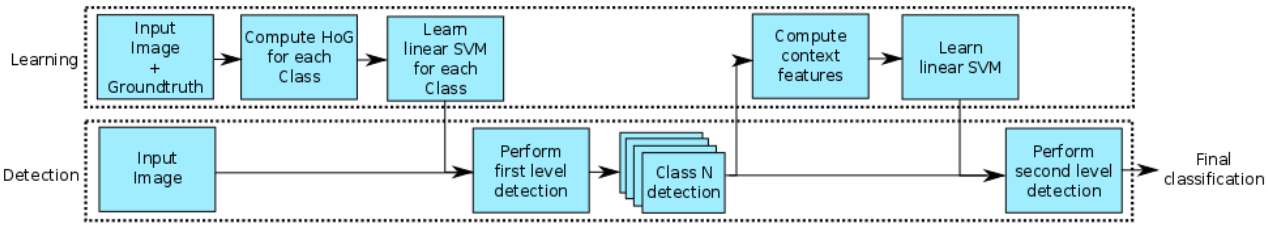


Fig. 5. Method's overview.

3. SYSTEM DETECTION FOR A WHOLE MAP

In this section, we present a whole and effective detection system using contextual feature. Our processing chain, shown on Fig 5, is divided into several parts. Learning and detection are performed in two steps.

The first step is based on an efficient descriptor, the Histograms of Oriented Gradients (HOG) [1], to describe the local appearance of symbols. This representation is fast and effective for all type of objects detection in complex images [4]. Therefore it seems relevant to use HOG for black and white symbols which are rich in gradients information. A database is made for each class, positive examples are the annotations and negatives examples are the windows that do not overlap them. When training our model for symbols detection, we have a very large number of negative examples. Thus we use a classical approach for data learning in computer vision, and we construct a hard negatives database with a subset of negative examples. We fed a SVM classifier for each class with these databases, that creates our first level of classification.

The second step is to form the contextual feature as described in the previous section. Outputs of the visual detection performed with HOG are the set of windows $(D_1, ..., D_k)$. Contextual classifier can be used in two different modes:

- Detection: perform the sliding window on the whole map, as the first level classification. By using contextual feature in this way, it allows to recover false negatives of the first level, at the risk of emergence of new false positives.
- Filtering: we don't take into consideration windows with a negative visual score. This mode is more faster, but it is impossible to recover false negatives.

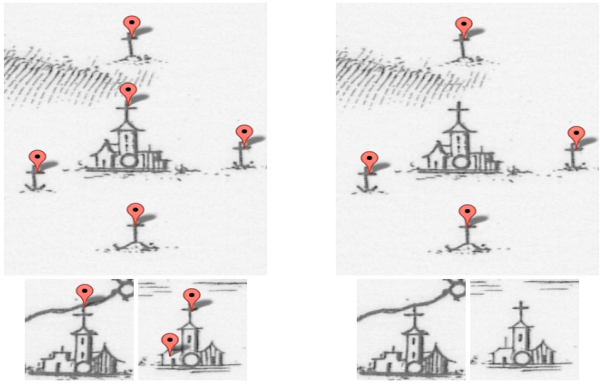
4. EXPERIMENTS AND RESULTS

An evaluation of our algorithms for the 181 Cassini old maps is proposed. They are digitized at 600 dpi, providing high resolution images with a size of about 25000 pixels by 17000 each. We carry out a quantitative assessment on one map for which we have annotations made by the National Geographic Institute (IGN). The groundtruth is manually labeled and the number of annotations is identified in the table 6.

Calvary	Church	Castle	Chapel	Watermill	House	Hamlet	Windmill
55	285	100	23	182	127	118	47

Fig. 6. Number of annotations.

Boxes have a size of 41 x 56 pixels for the smallest class (calvary), and 121 x 126 pixels for the largest class (church), it represents about one hundred million windows per class to classify during the detection with a step of 2 pixels.



(a) (b)

Fig. 7. Results obtained with resolved visual ambiguities, (a) calvary detection with HOG only, (b) calvary detection with context.

We evaluate four types of features: the histogram of oriented gradients, the feature used by Felzenswalb in [10], our close context descriptor and finally the full feature which combines local and global context. Our HoG descriptor follows the configuration of [1] with a window cut in $4 * 4$ cells, and gradients voting for a 9 bins histogram evenly space over $0°$ to $180°$. Then cells are grouped with $2 * 2$ blocks, and the final feature is L2-Norm. For the feature of [10], it is comparable to our close context description, without the position information, but scores are renormalize with a logistic function $\sigma(x) = 1/(1 + exp(-2x))$. For the close context description, we set the d_{max} to 70 pixels, and for the large context description, we set the number of crowns n to 3 and r_{max} to 500 pixels, which seems to give better results.

The map is split into three equal parts to make training, test and validation subsets. We perform the training of the two levels of classification on the same data. The validation database allows to adjust hyper parameters of the section 2.2. We apply the two classifiers to make a comparison between the two levels, visual and contextual, on the set of maps. As expected, see Fig 7.a, the visual classification for the class calvary have some ambiguities at the top of the church. After adding context information, Fig 7.b, our classifier successfully filters the false positives of the first level.

Shown in the table 8.a, the quantitative results of our methods in detection and filtering mode on one map. As we can see, HOG gives already good results with a mean average precision of 45% for highly unbalanced database between positives and negatives. Moreover, we notice that our contextual feature is more suited for filtering.

In columns detection and filtering of the table 8.a there are best results for a fixed set of parameters for each method. A lot of different parameter sets have been tested for each method (last column of 8.b) but only best results are finally kept for each class in filtering mode. In detection and filtering mode, with d_{max} fixed to the same value

		Detection			Filtering		
Classes	HOG	Felz [10]	Close	Large	Felz [10]	Close	Large
Calvary	40.8	43.2	41.2	43.7	40.2	36.1	34.6
Church	57.0	53.8	58.4	51.8	53.8	59.8	59.7
Castle	9.1	3.7	9.1	13.1	3.7	13.3	14.0
Chapel	17.1	14.9	17.5	14.4	15.0	41.6	41.6
Watermill	91.1	79.2	90.8	91.1	77.7	90.4	90.4
House	49.0	51.7	51.0	49.1	51.2	53.0	55.3
Hamlet	51.3	51.7	51.6	48.3	49.4	52.0	50.2
Windmill	51.7	45.1	51.6	52.0	45.1	46.5	54.3
MAP	45.8	42.9	46.4	45.4	42.0	49.0	50.0

(a)

	Max		
Classes	Felz [10]	Close	Large
Calvary	40.7	41.6	44.7
Church	56.7	59.8	59.9
Castle	15.1	13.3	14.4
Chapel	15.0	45.8	50.0
Watermill	85.3	91.2	91.1
House	51.2	53.0	56.0
Hamlet	51.6	53.1	52.9
Windmill	48.3	50.0	56.0
MAP	45.4	50.4	53.1

(b)

Fig. 8. (a) Comparative results between HOG and contextual descriptors with fixed hyper parameters (b) Best results with a different set of parameters for each class.

for the [10] and our close context, performances decrease for [10] when they increase for us. We assume that only presence information in a close neighborhood is not relevant in our case and brings more noise than information. The global context gives best results for a fixed set of parameters in filtering mode. By selecting best hyper parameters for each class, we can see a significant improvement of the performances. Our future work shall focus on making this task automatic. The Fig 9 summarizes results for hamlet category, and shows that the use of context can significantly improve the detection accuracy.

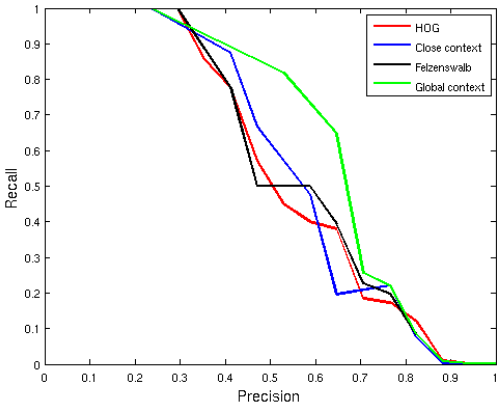


Fig. 9. Recall precision curve for the class hamlet.

5. CONCLUSION

We have proposed a complete system for symbol detection and recognition in old maps. Our approach is based a novel contextual feature modeling adapted to this specific type of detection problem. The spatial context is integrated at two levels around a detected symbol: in a small neighborhood to remove visual ambiguities between symbols, and in a larger area with co-occurrence statistics between classes. Our approach gives outstanding results on a large dataset extracted from Cassini maps, and outperforms state of the art methods on this type of detection problems. The main direction for future works concerns interactive learning [12, 13] that could help to better focus on difficult examples and boost the full annotation process of the 181 Cassini maps. Moreover our current method is based on HOG only for the visual classification, and combining visual features like in [14] would improve our results. Finally, to improve the global contextual description further investigation about including orientation information or using recent geographic data is definitely a promising research direction.

6. REFERENCES

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005, pp. 886–893.

[2] R. Perko and A. Leonardis, "Context awareness for object detection," in *Workshop of the Austrian Association for Pattern Recognition*, 2007.

[3] X. Wang, X. Bai, W. Liu, and LJ. Latecki, "Feature context for image classification and object detection," in *CVPR*, 2011.

[4] SK. Divvala, D. Hoiem, JH. Hays, A. Efros, and M. Hebert, "An empirical study of context in object detection," in *CVPR*, June 2009.

[5] G. Heitz and D. Koller, "Learning spatial context: Using stuff to find things," in *Proc. 10th European Conference on Computer Vision*, 2008.

[6] S. Kumar and M. Hebert, "A hierarchical field framework for unified context-based classification," in *ICCV*, 2005.

[7] A. Singhal, J. Luo, and W. Zhu, "Probabilistic spatial context models for scene content understanding," in *CVPR*, 2003.

[8] N. Thome, D. Merad, and S. Miguet, "Learning articulated appearance models for tracking humans: A spectral graph matching approach," *Signal Processing: Image Communication*, vol. 23, no. 10, pp. 769–787, 2008.

[9] C. Galleguillos, A. Rabinovich, and S. Belongie, "Object categorization using co-occurrence, location and appearance," 2008.

[10] P. Felzenszwalb, R. Girshick, D. Mcallester, and D. Ramanan, "Object detection with discriminatively trained part based models," *IEEE Transactions on PAMI*, vol. 32(9), 2010.

[11] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie, "Objects in context," in *ICCV*, 2007, pp. 1 –8.

[12] D. Gorisse, M. Cord, and F. Precioso, "Salsas: Sub-linear active learning strategy with approximate k-nn search," *Journal Pattern Recognition*, 2010.

[13] D. Gorisse, M. Cord, and F. Precioso, "Scalable active learning strategy for object category retrieval," in *Conf IEEE ICIP*, 2010.

[14] D. Picard, N. Thome, and M. Cord, "An efficient system for combining complementary kernels in complex visual categorization tasks," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 3877–3880.