

Fast LSIS Profile Entropy Features for Robot Visual Self-Localization

Herve GLOTIN^{1,2}, Zhong-Qiu ZHAO^{1,3}, Emilie DUMONT¹

¹ Lab. Sciences de l'Information et des Systèmes LSIS, UMR USTV CNRS 6168 La Garde-France

² University of Sud Toulon-Var, France

³ School of Computer & Information, Hefei University of Technology, China

glotin@univ-tln.fr; zhongqiuzhao@gmail.com

Abstract

In the Robot Vision task, the participants are asked to answer where is the robot using its vision. The robot may be in 5 rooms (BO-One-person office, CR-Corridor, EO-Two-persons office, KT-Kitchen, PA-Printer Area). In order to train our models we structured the views of each room into several sub-classes: BO-inside, exit, enter; CR-enter, exit, leftstairs, nostairs, rightstairs; EO-enter, exit, inside; KT-enter, exit, cooking hearth, table, television; PA-cabinet, enter, exit. Then an SVM was constructed for each of these 19 sub-classes. After that, we combined the results of SVMs by maximizing to get the final decision. We run our classification models on the new Profile Entropy Features (PEF) that combines RGB color and texture, yielding to one hundred of dimension, and we compare them to generic Descriptor of Fourier (DF). We also made a fusion of the models on these 2 different features. So we got 3 runs. In our experiments, for each decision, we used only the current image, but we do not exploit continuity of the sequences. For this case, a total of 7 teams submitted runs. The official evaluation give for the SVM(PEF) run a score of 544, and for SVM(DF) run a score of -32, while their fusion a score of 509.5. Thus our result possesses the 5th rank over the seven. The experiments show that our SVM model works well with little training cost, and PEF feature works much better than DF feature. It could be concluded that PEF is quite efficient: it is very fast to be computed, with around 10 images computed per second on usual pentium, and less of 2 hours of training (compared to 60 hours for the best systems), but still give a competitive results.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2.3 [Database Management]: Visual Systems—*CBIR*

Keywords

Efficient visual features ; Fast Content Based Information Retrieval ; Profile Entropy Features ; LS-SVM ; Robot Vision ; Adaptive Localisation

1 Introduction

The task addresses the problem of topological localization of a mobile robot using visual information. Specifically, participants will be asked to determine the topological location of a robot based

on images acquired with a perspective camera mounted on a robot platform. The details of this task and dataset are shown in [10].

We manually classified the views of each room into several classes: BO-inside, exit, enter; CR-enter, exit, leftstairs, nostairs, rightstairs; EO-enter, exit, inside; KT-enter, exit, cooking hearth, table, television; PA-cabinet, enter, exit. Then an SVM was constructed for each small class, to distinguish this small-class from all the others and recognize if the robot is currently in this small-class or not. So we constructed 19 SVMs in total(one per class). After that, we combined the results of SVMs by maximizing to get the final decision. We run our classification models on PEF(Profile Entropy Feature) and DF(Descriptor of Fourier) features and compared them, and we also made a fusion of the models on these 2 different features. So we submitted 3 runs. In our experiments, for each decision, we used only the current image, but not exploit continuity of the sequences.

2 The Profile Entropy Features

We propose in [1, 2] a new feature equal to the pixel 'profile' entropy. A pixel profile can be a simple arithmetic mean in horizontal (or vertical) direction. The advantage of such feature is to combine raw shape and texture representations in a low cpu cost feature.

Let I be an image (or a part of) of $L(I)$ rows, and $C(I)$ columns. The PEF are computed on these normalized RGB channels : $l = (R + G + B)/3$, $r = R/l$, and $g = G/l$. We consider the profiles of the orthogonal projections of the pixels to the horizontal X axis, noted Π_X^{op} , and to the vertical Y axis (Π_Y^{op}), where op is a projection operator. This one is either the arithmetic mean of the pixels (noted Π^{Ari}), or their harmonic mean (noted Π^{Harm}), as illustrated in Fig.1 and Fig.2. Thus the length of a given profile is either $S = C(I)$ or $S = L(I)$.

Then, for each profile, we estimate its probability distribution function ($p\hat{d}f$) on N bins (where $N = round(\sqrt{S})$) as proposed in [3].

For each channel, and each operator op , we compute :

$\Phi_X^{op}(I) = p\hat{d}f(\Pi_X^{op}(I))$. Considering that the sources are ergodic, we set PEF_X component to the normalised entropy of this distribution :

$$PEF_X(I) = H(\Phi_X^{op}(I))/\log(N),$$

where N the number of bins of the considered distribution, and H the usual entropy function. We compute similar PEF on Y axis :

$$PEF_Y(I) = H(\Phi_Y^{op}(I))/\log(N).$$

We set a third PEF component to the entropy of the direct distribution of all the pixels in I , $p\hat{d}f(I)$:

$$PEF_B(I) = H(p\hat{d}f(I))/\log(N),$$

where $N = round(\sqrt{L(I) * C(I)})$ bins.

The whole PEF features are the concatenation of PEF_X , PEF_Y and PEF_B , with the usual mean and standard deviation of each channel of I .

The PEF are computed on three horizontal (noted '=') or vertical ('|||') equal segmented subimages, and on the whole image. For exemple, for a given operator, we have the whole image plus the three '|||' subimages, and for each of the 3 channels we have $PEF_{X,Y,B}$, plus their mean and variance, thus we have $4 * 3 * (3 + 2) = 60$ dimensions. We note '# ' the concatenation of '=' and '|||' PEF, without duplication.

3 Fast classification using Least Squares Support Vector Machines

In order to design fast image retrieval systems, we use the Least Squares Support Vector Machine (LS-SVM). The SVM [5] first maps the data into a higher dimensional input space by some kernel

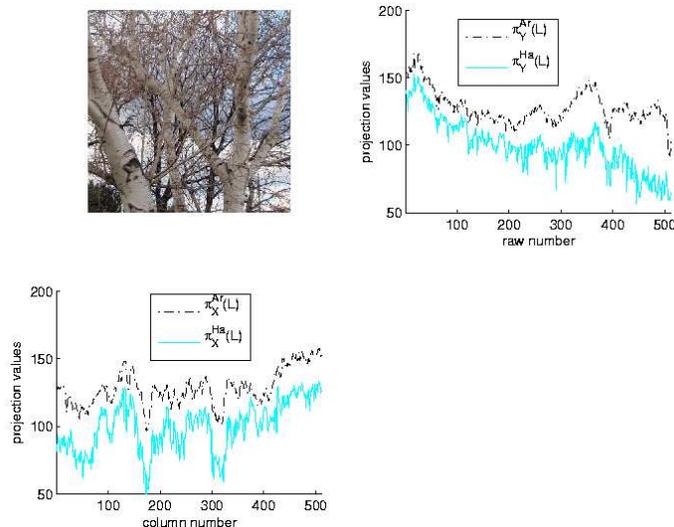


Figure 1: Horizontal X (Bottom) and, vertical Y (Top Right) profiles using arithmetic (-.-) and harmonic (-) projections of the luminance of an image of a tree. It shows clearly the difference between the two projections for this structured pattern.

functions, and then learns a separating hyperspace to maximize the margin. Currently, because of its good generalization capability, this technique has been widely applied in many areas such as face detection, image retrieval, and so on [6, 7]. The SVM is typically based on an ε -insensitive cost function, meaning that approximation errors smaller than ε will not increase the cost function value. This results in a quadratic convex optimization problem. So instead of using an ε -insensitive cost function, a quadratic cost function can be used. The least squares support vector machines (LS-SVM) [8] are reformulations to the standard SVMs which lead to solving linear KKT systems instead, which is quite computationally attractive. Thus, in all our experiments, we will use the LS-SVMlab1.5 (<http://www.esat.kuleuven.ac.be/sista/lssvmlab/>).

In our experiments, the RBF kernel

$$K(x_1 - x_2) = \exp(-|x_1 - x_2|^2 / \sigma^2)$$

is selected as the kernel function of our LS-SVM. So there is a corresponding parameter, σ , to be tuned. A large value of σ^2 indicates a stronger smoothing. Moreover, there is another parameter, γ , needing tuning to find the tradeoff between to stress minimizing of the complexity of the model and to stress good fitting of the training data points.

We set these two parameters as

$$\sigma^2 = [4 \ 25 \ 100 \ 400 \ 600 \ 800 \ 1000 \ 2000]$$

and

$$\gamma = [4 \ 8 \ 16 \ 32 \ 64 \ 128 \ 256 \ 512]$$

respectively. So hundred of SVMs were constructed for each SVM model, and then we selected the best SVM using the validation set.

4 Experimental Results

In this task, the participants are asked to answer 'where are the robots'. The robots may be in 5 rooms (BO-One-person office, CR-Corridor, EO-Two-persons office, KT-Kitchen, PA-Printer Area). We manually classified the views of each room into several classes: BO-inside, exit, en-

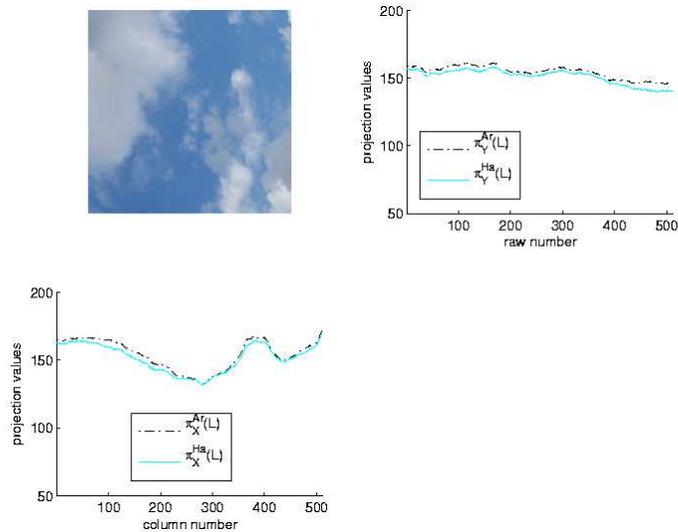


Figure 2: Similar to Fig.1 but for an image of the concept sky : arithmetic and harmonic profiles are similar.

Table 1: The submitted runs of ImageCLEF2009 Robot Vision Task

Rank	Run Tag	Method	Score	Training time
1	MIRG_UG	matched points; illumination filter	890.5	unknown
2	Idiap	CRFH+SIFT+PACT combined using G-DAS	793.0	unknown
3	UAIC	Full search using all frames	787.0	60h
4	CVIUI2R	LAB+histograms+Probabilistic SVM	784.0	unknown
5	LSIS_1	PEF+SVM	544.0	1h
6	SIMD	Hough+Canny+SIFT	511.0	60h
7	LSIS_3	Fusion of (PEF+SVM)(DF+SVM)	509.5	2h
8	MRIM	5x5 patches, Color SIFT, Language Model	456.5	unknown
9	LSIS_2	DF+SVM	-32.0	1h

ter; CR-enter, exit, leftstairs, nostairs, rightstairs; EO-enter, exit, inside; KT-enter, exit, cooking hearth, table, television; PA-cabinet, enter, exit (Figure 3). Then an SVM was constructed for each small class, to distinguish this small-class from all the others and recognize if the robot is currently in this small-class or not. So we constructed 19 SVMs in total. After that, we combined the results of SVMs by maximizing to get the final decision. We run our classification models on PEF(Profile Entropy Feature) and DF [9] features and compared them, and we also made a fusion of the models on these 2 different features. So we got 3 runs. In our experiments, for each definition, we used only the current image, but not exploit continuity of the sequences.

The evaluation was performed by the organizer. The following rules are used when calculating the score for a single test sequence: (1). +1.0 points for each correctly classified image Correct detection of an unknown room is treated the same way as correct classification. (2). -0.5 points for each misclassified image (3). 0.0 points for each image that was not classified (the algorithm refrained from the decision) In case several test sequences are used, the scores are calculated separately for each test sequence and then summarized.

As the evaluation, we got the SVM+PEF run with the score of 544, and the SVM+DF run with the score of -32, and the fusion run with the score of 509.5 (Table 4). The best result of us possesses the 9th rank. The experiments show that our SVM model works well with little training cost, and PEF feature works much better than DF feature. It could be concluded that PEF is efficient. Moreover, PEF is fast with around 10 images computed per second on usual pentium.

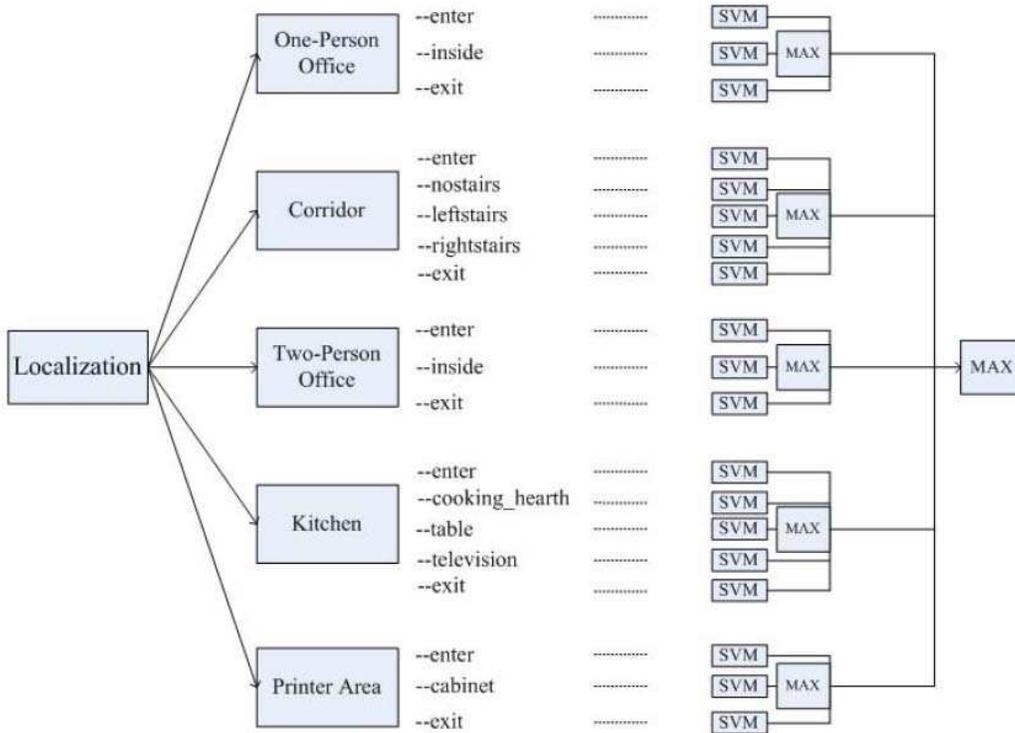


Figure 3: The framework of our system

5 Conclusion

Our SVM model on PEF works well, with medium performances, needing much less training time than other systems (around 50 times less). It could be concluded that our system with PEF feature is efficient for this task. We also noticed that the best result of those who used the continuity information, which is from SIMD group, attains a much higher score: 916.5. So in the future, we will combine in our system the continuity information through HMM optimisation, which shall result in a significant enhancement.

Acknowledgment

This work was supported by French National Agency of Research (ANR-06-MDCA-002) and Research Fund for the Doctoral Program of Higher Education of China (200803591024).

References

- [1] Glotin, H.: Information retrieval and robust perception for a scaled multi-structuration, Thesis for habilitation of research direction, University Sud Toulon-Var, Toulon (2007)
- [2] Glotin, H., Zhao, Z.Q., Ayache, S.: Efficient Image Concept Indexing by Harmonic & Arithmetic Profiles Entropy, 2009 IEEE International Conference on Image Processing, Cairo, Egypt, November 7-11, 2009 (2009)
- [3] Moddemeijer, R.: On estimation of entropy and mutual information of continuous distributions, *Signal Processing*, 16(3), 233–246 (1989)
- [4] Vapnik, V.: The nature of statistical learning theory. Springer-Verlag, New York (1995)

- [5] Vapnik, V.: Statistical learning theory. John Wiley, New York (1998)
- [6] Waring, C.A., Liu, X.: Face detection using spectral histograms and SVMs. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 35(3), 467–476 (2005)
- [7] Tong S., Edward, Chang: Support vector machine active learning for image retrieval. In *Proceedings of the ninth ACM international conference on Multimedia Ottawa, Canada*, pp. 107–118 (2001)
- [8] Suykens, J.A.K., Vandewalle, J.: Least Squares Support Vector Machine Classifiers *Neural Processing Letters*, 9, 293–300 (1999)
- [9] Smach, F., Lemaitre, C., Gauthier, J.P., Miteran, J., Atri, M.: Generalized Fourier Descriptors with Applications to Objects Recognition in SVM Context, 30, *J. Math Imaging Vis* 43–71 (2008)
- [10] Caputo B., Pronobis A., Jensfelt, P.: Overview of the CLEF 2009 robot vision track, *CLEF working notes 2009, Corfu, Greece* (2009)