Weighted Dissociated Dipoles for Evolutive Learning

Xavier BARÓ & Jordi VITRIÀ
Centre de Visió per Computador
Departament de Ciències de la Computació
Universitat Autònoma de Barcelona
{xbaro, jordi}@cvc.uab.cat

Abstract. The complexity of any learning task depends as in the learning method as on finding a good representation of the data. In the concrete case of object recognition in computer vision, the representation of the images is one of the most important decisions in the design step. As a starting point, in this work we use the representation based on Haar-like filters, a biological inspired feature based on local intensity differences. From this commonly used representation, we jump to the dissociated dipoles, another biological plausible representation which also includes non-local comparisons. After analyzing the benefits of both representations, we present a more general representation which brings together all the good properties of Haar-like and dissociated dipoles representations. All these feature sets are tested with an evolutionary learning algorithm over different object recognition problems. Besides, an extended statistically study of these results is performed in order to verify the relevance of these huge feature spaces applied to different object recognition problems.

Keywords. Evolutive learning, Dissociated dipoles, Haar-like features, Adaboost, Object Recognition, Friedman statistic

1. Introduction

Object recognition is one of the most challenging problems in the computer vision field. Given an image, the goal is to determine whether or not the image contains an instance of an object category or not. In the literature there are two main approaches to deal with the object recognition problem: Holistic methods and heuristical local methods.
Holistic methods use the whole image or a region of interest to perform object identification. These systems are typically based on Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) or on some form of artificial neural net. Although these methods have been used in a broad set of computer vision problems, there are still some problems that cannot be easily solved using this type of approaches. Complex backgrounds, partial object occlusions, severe lighting changes or changes in the scale and point of view represent a problem if they were to be faced under a holistic approach. An alternative are the heuristic local appearance methods or feature based methods, which provide a richer description of the image. In contrast to holistic methods that are problem independent, in this case we should select the type of features which bet-

ter adapts to the problem. We can choose from a wide variety of features, such as the fragments-based representation approach of Ullman [1], the gradient orientation based SIFT [2] or the Haar-like features used by Viola [3].

This paper is related to he feature based methods, namely, our starting point is the Haar-like feature model. Haar-like filters allow a powerful representation of an image using local contrast differences. This representation demonstrated to be a robust description to be applied over object recognition problems, specially in the case of face detection [3]. In addition, the evaluation of these filters using the integral image has a low computationally cost, being potentially useful for real-time applications.

The Haar-like filters are local descriptors, in the sense that they only compare adjacent regions. In [4], Sinha presents the dissociated dipoles, a non-local representation based on region contrast differences. The evaluation is done in the same way that in the case of the previous ones, with the only difference that now the regions do not have the adjacency constraint. In fact, some type of Haar-like features can be represented via dissociated dipoles.

Comparing Haar-like filters and dissociated dipoles, we can see that they share some desired properties as the robustness in front of the noise (they are integral based features) and severe illumination changes (they use region differences, not directly the intensity value). In contrast, the Haar-like have some filters to detect lines that the dipoles cannot simulate, and the dissociated dipoles have the non-local ability which Haar-like approach cannot perform.

In order to collect the good properties of both feature sets, a variant of dissociated dipoles is presented and evaluated. Using weights over the dissociated dipoles, we can represent most of the Haar-like features, obtaining a richer feature space that combines the benefits of both feature sets.

The evaluation of any of the above feature sets consists of the subtraction of the value of all the negative regions from the value of positive regions. Finally, the difference between the regions mean intensity is used to decide to which class a given image belongs.

At this point, two different approaches can be applied: Qualitative or quantitative. Although the most extended approach is the quantitative one used by Viola [3], which consists of finding the best threshold value to make a decision, recent works have demonstrated that qualitative approaches based only on the sign of the difference are more robust in front of noise and illumination changes [5].

All the previous feature sets have in common a high dimensionality, which difficult the application of the classical learning methods. In the case of Haar-like approaches, it is solved by scaling the samples to a small training window where the number of possible features is computationally feasible. In the case of dissociated dipoles, the original image is repeatedly filtered and subsampled to create different levels of the image pyramid. Then, a point in the deeper levels of the pyramid corresponds to the mean value of a region in the upper levels of the pyramid. Thus, we can only consider the relation between points. In this work we use an evolutionary version of Adaboost that is able to work with huge feature spaces, such as the presented before.

This paper is organized as follows: Section 2 describes the features of the three types of features and their properties. Section 3 introduces the qualitative approach to evaluate the features, and section 4 presents the learning strategy. Finally in section 5 we compare and analyze the behavior of each type of feature set using the evolutive approach.

2. Feature set

Selecting a good feature set is crucial to design a robust object recognition system. In this work we use region based features, which consist of differences between the mean values of each region. These features are robust in front of the noise and severe light conditions. In addition, using the integral image (fig. 1), the value of each region is calculated with just 4 accesses.

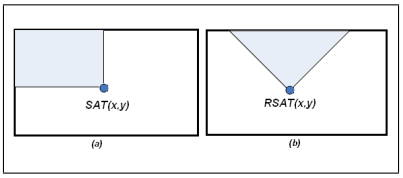

Figure 1. Integral images. Each point contains the sum value of the gray region. *a* Integral image *b* 45^o rotated integral image

2.1. Haar-like features

This type of features is a discretized version of the Haar wavelets. Viola used this type of features in their real-time face detector [3] because they are easy to be computed by means of the integral image and in addition they are robust in front of noise and severe illumination changes. The original set of Haar-like features types was extended by Lienhart in [6], adding a rotated version of the original types, and demonstrating that the performance of a classifier is related to the size of the feature space. The extended set of features is shown in fig. 2. The feature set is composed by all the possible configurations of position and scale inside a training window, therefore, learning a classifier using this feature set becomes unfeasible for large window sizes.

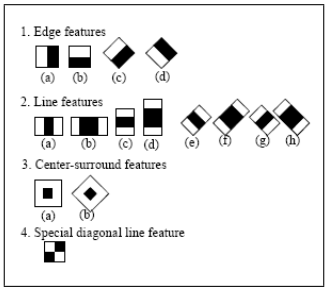

Figure 2. Extended haar-like features.

2.2. Dissociated dipoles

The sticks of dissociated dipoles where defined by Shina in [4]. In these works, the authors perform a set of physical experiments to demonstrate that as in the case of Haar-like features, the dissociated dipoles are a biological plausible type of features. This is a more general feature set, which compares the mean illuminance values of two regions,

the so called excitatory and the inhibitory dipoles (see 3). From a computational point of view, the evaluation of a dissociated dipole has the same cost as the evaluation of a Haar-like feature. They also share the robustness in front of noise and illuminance changes. In fact, the two regions Haar-like features (see edge features in fig. 2) can be represented by means of the dissociated dipoles. The feature set is composed by all the possible sizes and positions of each one of the two regions. Any exhaustive search (i.e Adaboost approach) over this feature set is unfeasible. An application based on scale-image approach can be found in [7].

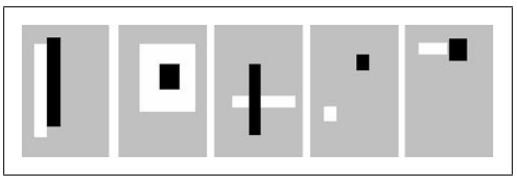

Figure 3. Dissociated dipoles.

2.3. Weighted Dissociated dipoles

Weighted dissociated dipoles are a more general definition of the dissociated dipoles, where each region has an associated weight $W \in \{1, 2\}$. With this simple modification, we can represent all the edge, line and center-surrounding types of the Haar-like feature set. As in the case of the normal dissociated dipoles, the huge dimension difficults the use of this type of features by classical approaches.

3. Ordinal Features

The term of ordinal features is related to the use of the sign instead of directly the value of the feature. In [5], a face detection approach is presented using only the sign of region intensity differences. They demonstrate that removing the magnitude of the difference, the model becomes more stable to illumination changes and image degradation.

4. Evolutive Adaboost

The huge dimensionality of all the explained feature sets is a problem for the classical learning approaches. In [8], an evolutive variation of the classical Adaboost used by Viola in [3] is presented. They propose to change the exhaustive search over the feature space performed by the weak learner by a genetic algorithm (see fig. 4). As a result, learning a classifier is formulated in terms of an optimization problem, where we want to find the parameters of the feature that minimizes the classification error.

To use that approach, we first must define the parametrization of each type of features. Let's denote $R_k = (x, y, W, H)$ the parametrization of the region R_k where the point (x, y) is the upper-left corner, and W and H the weight and height respectively. A dissociated dipole is defined by means of the excitatory and inhibitory regions, and thus we need at least 8 parameters. Analogously, the weighted dissociated dipoles, will be

Given: $(x_1, y_1), ..., (x_m, y_m)$ where $x_i \in X, y_i \in Y = \{-1, +1\}$
Initialize $D_1(i) = 1/m$
For $t = 1, .., T$:

Step 1. Use a genetic algorithm to minimize:

$$\epsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i]$$

the given solution is taken as the hypothesis h_t

Step 2. Get weak hypothesis $h_t : X \mapsto \{-1, +1\}$ with error ϵ_t.

Step 3. Choose $\alpha_t = \frac{1}{2} \ln \left(\frac{1-\epsilon_t}{\epsilon_t} \right)$

Step 4. Update:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & if \quad h_t(x_i) = y_i \\ e^{\alpha_t} & if \quad h_t(x_i) \neq y_i \end{cases}$$

$$= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Step 5. Output the final hypothesis:

$$H(x) = sign \left(\sum_{t=1}^{T} \alpha_t h_t(x) \right)$$

Figure 4. The evolutive Discrete Adaboost

defined in the same way, but now we need to add two extra parameters $W^+ \in \{1, 2\}$ and $W^- \in \{1, 2\}$ which correspond to the weight of the excitatory and inhibitory regions. Thus, we need at least 10 parameters to describe the weighted dissociated dipoles. Finally, using the restrictions in position and size of the Haar-like features, with only one of the regions and the type of Haar-like feature we can describe all the feature set. Therefore, with just 5 parameters we can describe all that features.
It is important to notice the differences in the descriptor vector, because high dimensions difficult the task of the evolutive algorithm, which must learn a large number of parameters.

5. Results

Once all the feature sets and the evolutive learning approach are exposed, we evaluate the performance of the classifier using the three types of features. First, we describe the data and methodology used to evaluate the performance. Finally, a statistical study of the obtained results is performed in order to analyze the effect of using the different feature sets.

5.1. Performance evaluation

We evaluate the performance of the classifiers over the following tasks:

Face detection: The first task is to learn a face detector. We use the MIT-CBCL face database [9] with a random selection of 1.000 face images and 3.000 non-face images, learning a classifier to distinguish between both classes.

Traffic sign detection: A traffic sign detector must be able to distinct when a given image contains or not an instance of a traffic sign. The experiment is performed using real images acquired in the context of a mobile mapping project provided by the ICC[1]. The database consists on 1.000 images containing a traffic sign and 3.000 background images.

Pedestrian detection: In this case, a detector is trained to identify instances of pedestrians. We use the INRIA Person Dataset[2], with 2.924 images divided into 924 pedestrian instances and 2.000 background images.

Cars detection: This problem consists of detecting instances of a car in urban scenes. We use the UIUC cars database [10], with a total of 1.050 images containing 550 instances of lateral views of different cars and 500 of background images.

Text detection: This task consist on detect text regions in a given image. We use the text location dataset from the *7th International Conference on Document Analysis and Recognition (ICDAR03)*[3].

To perform the experiments we fix the maximum number of iterations in the Adaboost algorithm to 200, with a maximum of 50.000 evaluations for the genetic algorithm. The evaluation is carried out using a stratified 10-fold cross validation with a confidence interval at 95% (assuming normal distribution over the error), and the results are shown in figure 5.
The weighted dipoles outperform the other types of features in all the problems used, obtaining good performance rates in all of them. The next step is to check if the observed differences are statistically significant.

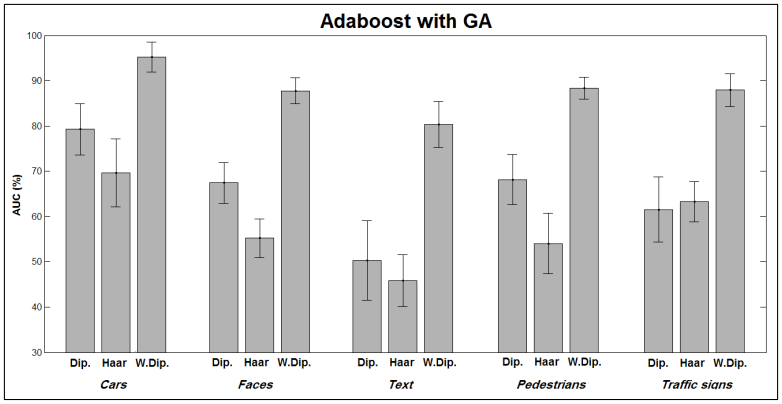

Figure 5. AUC value and confidence intervals.

[1] Institut Cartogràfic de Catalunya. www.icc.es
[2] pascal.inrialpes.fr/data/human/
[3] algoval.essex.ac.uk/icdar/TextLocating.html

5.2. Statistical analysis

In [11], Demšar performs a study of the validation schemes used in the works published in the International Conferences on Machine Learning between 1999 and 2003, pointing up the main validation errors and wrong assumptions. As a result of his work, Demšar describes a methodology to compare a set of methods over different data sets. In this section, we use that methodology to find out statistical significant differences between the use of the different feature sets used in the present paper. The techniques applied and the numerical results for the statistical study are shown in table 1.
Let r_i^j be the rank of the j-th of *k* algorithms on the i-th of *N* data sets. The Friedman test compares the average ranks of algorithms, $R_j = \frac{1}{N}\sum_i r_i^j$. Under the null-hypothesis, which states that all the feature sets are equivalent, and so their average ranks R_j are equal, the Friedman statistic

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right] = 8.4 \tag{1}$$

is distributed according to χ_F^2 with $k-1$ degrees of freedom when N and k are big enough. For a small number of algorithms and data sets, exact critical values have been computed. Iman and Davenport [12] showed that FriedmanŠs χ_F^2 is undesirably conservative and derived a better statistic

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} = \frac{4\times\chi_F^2}{10-\chi_F^2} = 21 \tag{2}$$

which is distributed according to the F-distribution with $k-1=2$ and $(k-1)(N-1) = 2\times 4 = 8$ degrees of freedom. The critical value of F(2,8) for $\alpha = 0.05$ is 4.4590, which is smaller than F_F, so we reject the null-hypothesis. To reject the null-hypothesis indicates that the algorithms are not statistically equivalents.
As the null-hypothesis is rejected, we can proceed with a post-hoc test. In our case, as no algorithm is singled out for comparisons, we use the Nemenyi test for pairwise comparisons. The performance of the two classifiers is significantly different if the corresponding average ranks differ by at least the critical difference

$$CD = q_\alpha\sqrt{\frac{k(k+1)}{6N}} = 1.48 \tag{3}$$

where critical values q_α are based on the Studentized range statistic divided by $\sqrt{2}$. Using averaged ranks in table 1 we calculate all the pair-wise differences. Comparing those differences with the critical value, we can conclude that the Weighted dissociated dipoles are significantly better than the Haar-like features ($2.6-1.0 = 1.6 > 1.48$), but we can say nothing about the Dissociated dipoles ($2.8-1.0 = 1.8 < 1.48$).
The conclusion of this study is that we can affirm (with $\rho = 0.05$) that the Weighted dissociated dipoles are better than the Haar-like features.

6. Conclusions and future work

Although the tests must be extended to a larger set of databases, the weighted dissociated dipoles demonstrated to perform good in combination with the evolutive strategy. As a

Table 1. Results obtained in the experiments and the obtained rank (AUC)

	Feature Set		
Data Set	**Dipoles**	**Haar-like**	**Weighted dipoles**
Face Det.	$67.47\% \pm 4.52(2.0)$	$55.22\% \pm 4.23(3.0)$	$87.74\% \pm 2.85(1.0)$
Traffic Signs Det.	$61.53\% \pm 7.17(3.0)$	$63.30\% \pm 4.41(2.0)$	$87.92\% \pm 3.61(1.0)$
Pedestrians Det.	$68.16\% \pm 5.57(2.0)$	$54.00\% \pm 6.68(3.0)$	$88.40\% \pm 2.40(1.0)$
Cars Det.	$79.27\% \pm 5.64(2.0)$	$69.65\% \pm 7.54(3.0)$	$95.21\% \pm 3.28(1.0)$
Text Det.	$50.33\% \pm 8.75(2.0)$	$45.84\% \pm 5.75(3.0)$	$80.35\% \pm 5.08(1.0)$
Average Rank	2.20	2.80	1.00

feature work, we plan to extend the study to other evolutive strategies, as the Evolutive Algorithms based on Probabilistic Models (EAPM), which seem that can represent better the relations between the parameters that configure the features.

Acknowledgements

This work has been partially supported by MCYT grant TIC2006-15308-C01, Spain. This has been developed in a project in collaboration with the "Institut Cartogràfic de Catalunya" under the supervision of Maria Pla.

References

[1] S. Ullman and E. Sali, "Object classification using a fragment-based representation," in *BMVC '00: Proceedings of the First IEEE International Workshop on Biologically Motivated Computer Vision*. London, UK: Springer-Verlag, 2000, pp. 73–87.

[2] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. of the International Conference on Computer Vision ICCV, Corfu*, 1999, pp. 1150–1157. [Online]. Available: citeseer.ist.psu.edu/lowe99object.html

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. I–511–I–518.

[4] B. Balas and P.Sinha, "Dissociated dipoles: Image representation via non-local comparisons," Annual meeting of the Vision Sciences Society, Sarasota, FL., 2003.

[5] K. Thoresz and P. Sinha, "Qualitative representations for recognition," *Journal of Vision*, vol. 1, no. 3, pp. 298–298, 12 2001. [Online]. Available: http://journalofvision.org/1/3/298/

[6] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proceedings of the International Conference on Image Processing*. Rochester, USA: IEEE, September 2002, pp. 900–903.

[7] F. Smeraldi, "Ranklets: orientation selective non-parametric features applied to face detection," in *Proceedings. 16th International Conference on Pattern Recognition*, vol. 3, 2002, pp. 379– 382.

[8] X. Baró and J. Vitrià, "Real-time object detection using an evolutionary boosting strategy," *Frontiers in Artificial Intelligence and Applications / Artificial intelligence Research and Development, IOS Press, Amsterdam*, pp. 9–18, October 2006.

[9] "MIT-CBCL face database." [Online]. Available: cbcl.mit.edu/projects/cbcl/software-datasets/FaceData1Readme.html

[10] S. Agarwal, A. Awan, and D. Roth, "UIUC cars database." [Online]. Available: l2r.cs.uiuc.edu/ cogcomp/Data/Car/

[11] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *JMLR*, vol. 7, January 2006. [Online]. Available: http://jmlr.csail.mit.edu/papers/v7/demsar06a.html

[12] R. L. Iman and J. M. Davenport, "Approximations of the critical region of the friedman statistic," in *Communications in Statistics*, 1980, pp. 571–595.