

Image annotation : which approach for realistic databases ?



N. Hervé and N. Boujemaa



INRIA Rocquencourt, IMEDIA Project, BP 105 Rocquencourt
78153 Le Chesnay Cedex France

Abstract

We describe an efficient approach to image annotation that ranked first on the recent scene categorization track of the ImageEVAL benchmark. We show how homogeneous global image descriptors combined with a pool of Support Vector Machines achieve very good results. We also used this approach on several well known object recognition databases to emphasize two main aspects of this research domain : the importance of contextual information in object recognition and the unsuitability of many standard databases for this task.

Our approach

We use 6 generic global low-level descriptors. They are structurally homogeneous (they are all histograms) and therefore can be used simultaneously.

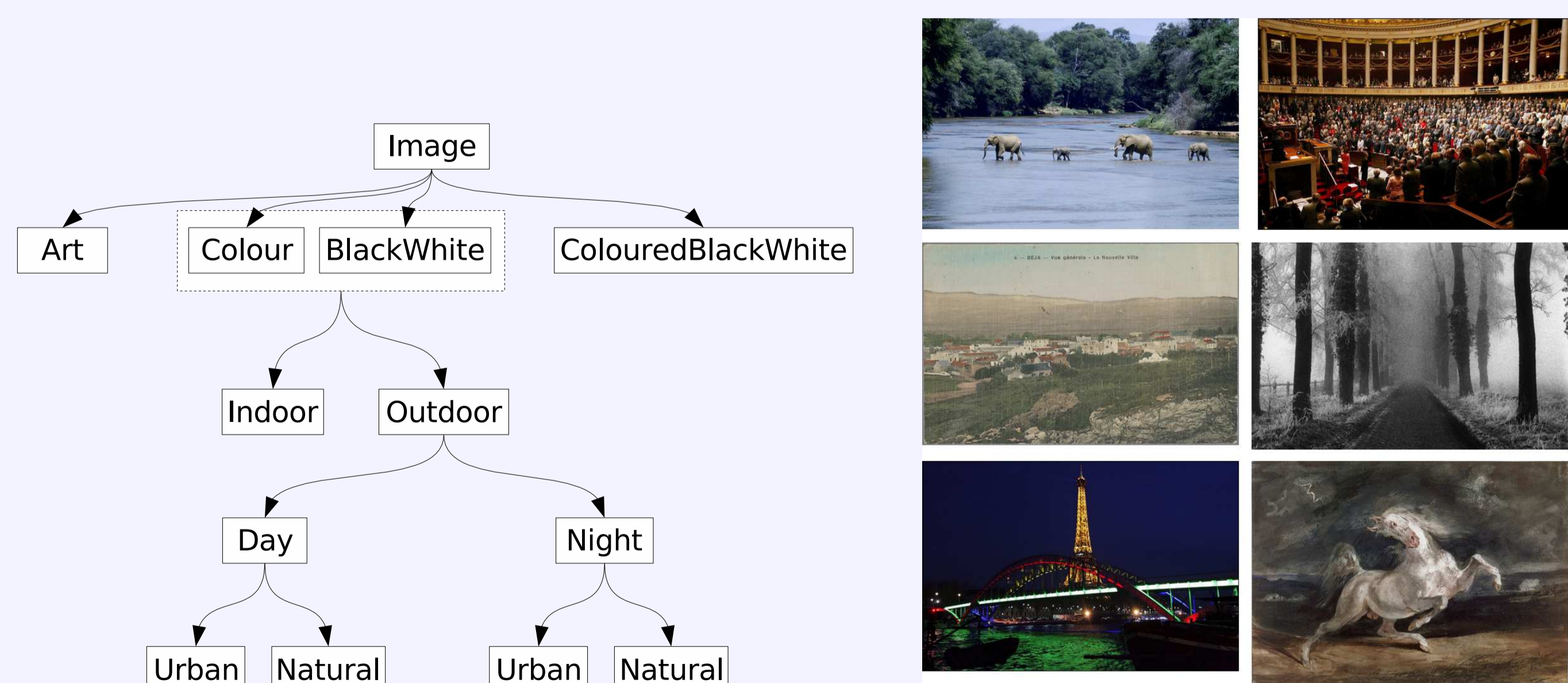
Descriptor	Type	Size
HSV histogram	Color	120
Laplacian weighted histogram	Color + Shape	216
Probability weighted histogram	Color + Texture	216
Fourier histogram	Texture	64
Hough histogram	Shape	49
Local edge orientation histogram	Shape	32

We choose an early fusion approach with soft-margin Support Vector Machine as learning algorithms. We produce pools of SVMs to predict the different concepts. We tested both parametric and non-parametric kernels. The results are slightly equivalent, but the non-parametric kernels have the huge advantage of being much more quicker to train.

Kernel	
Laplace	$k(x,y) = e^{-\sigma \sum_i x_i - y_i }$
RBF	$k(x,y) = e^{-\sigma \sum_i (x_i - y_i)^2}$
Triangular	$k(x,y) = \sum_i x_i - y_i $
GHI	$k(x,y) = \sum_i \min(x_i , y_i)$

We choose the simplest approach that assimilates the score of the SVM decision function in a confidence level.

ImageEval attribute extraction benchmark



Semantics to extract and learning database samples

IMEDIA ranked first on 6 teams.

Run	MAP	Run	MAP
imedia - parametric kernel	0.6784	anonymous	0.4931
imedia - non-parametric kernel	0.6556	anonymous	0.3676
cea	0.5771	anonymous	0.3141
etis	0.4912	anonymous	0.1985
anonymous	0.4907		

Other evaluation databases

Corel2000 - Scene categorization

Algorithm	Result
Our approach - 5 desc.	83.7
Our approach - HSV only	71.6
Chen - MILES	68.7

ImageEVAL object recognition task

Run	MAP	Run	MAP
imedia - local	0.2242	imedia	0.1545
imedia - local	0.2111	anonymous	0.1506
etis	0.1974	cea	0.1493
imedia - global	0.1777	anonymous	0.14
imedia	0.1733		

Caltech4

Algorithm	Plane	Car	Face	Motorbike
Our approach	99.2	100	98.6	98.8
Zhang J.	98.8	98.3	100	98.5
Willamowski	97.1	98.6	99.3	98.0

Xerox7

Algorithm	Result
Our approach	92.5
Zhang J.	94.3

Pascal VOC2005-1

Algorithm	Bike	Car	Motorbike	People
Our approach	88.7	92.2	95.8	86.9
Best published score	93.0	96.1	97.7	91.7

Pascal VOC2005-2

Algorithm	Bike	Car	Motorbike	People
Our approach	57.9	66.3	64.8	69.2
Zhang J.	68.1	74.1	79.7	75.3

Caltech101

Algorithm	30 im./class	15 im./class
Our approach	39.6	32.7
Zhang H.	66.23	59.08

Conclusion

We obtained good results on a scene categorization task with our approach involving global descriptors and pools of SVMs. We believe that the techniques used in this context are now mature enough to be implemented in real applications and could help end-users.

For the object recognition task, databases such as Corel, Caltech4, Xerox7 and VOC2005-1 should now clearly be abandoned for testing local approaches as simple global methods achieve equivalent accuracy. This implies that these databases are not problematic ones. We do not argue that object recognition problem is solved by using global descriptors. We do think, however, it is important to measure the suitability of the databases used by researchers to evaluate their methods. A global approach is a good way to obtain baseline results that provide pieces of information on the difficulty of the task and the database. The importance of contextual visual information has also been shown for this type of database.

We believe that the use of realistic databases such as ImageEVAL should now be standard. We will face very challenging problems that meet end-users scenarios.