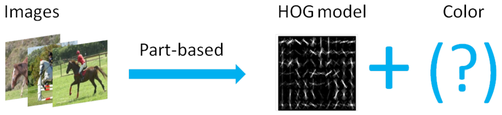
Color Attributes for Object Detection

Fahad Shahbaz Khan[1,*], Rao Muhammad Anwer[2,*], Joost Van de Weijer[2], Andrew D. Bagdanov[2,3], Maria Vanrell[2], Antonio M. López[2]

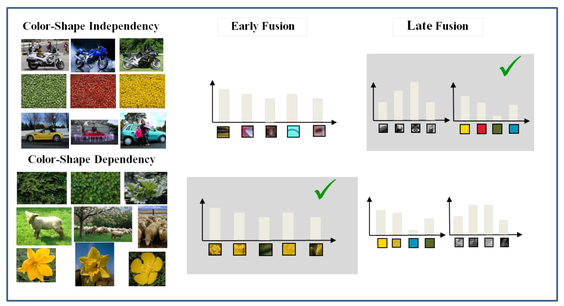
[1]Computer Vision Laboratory, Linköping University, Sweden
[2]Computer Vision Center,Universitat Autonoma de Barcelona,Spain
[3] Media Integration and Communication Center, University of Florence, Italy

Problem

Goal: Augmenting existing intensity based detectors with color information.



Early vs Late



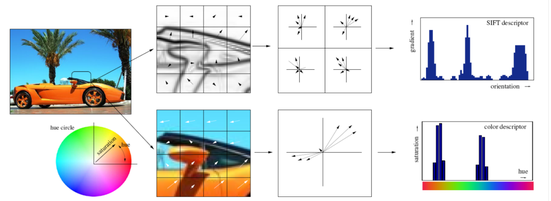
It is known that late fusion obtains better results in higher level pyramid representations.

Method	Level	Sports	Butterflies
Early fusion	1	80.6	79.6
Late fusion	1	81.6	81.9
Early fusion	2	80.8	81.7
Late fusion	2	82.7	84.4
Early fusion	3	82.7	83.3
Late fusion	3	**84.4**	**87.9**

Cartoon Character Detection

18 Classes: The Simpsons, Tom, Jerry, Fred, Barney, Sylvester, Mickymouse, Donaldduck, Tweety, Coyote, Roadrunner, Buggs, Daffy, Shaggy and Scooby.
Images: 586(train 304, testing 282).
Source: Google.

Color Descriptors



HUE descriptor (HUE): cells are represented by a histogram:

$$hue = \arctan\left(\frac{O1}{O2}\right) \rightarrow s = \sqrt{O1^2 + O2^2} \quad (1)$$

Opponent descriptor (OPP): cells are represented by a histogram over the opponent angle:

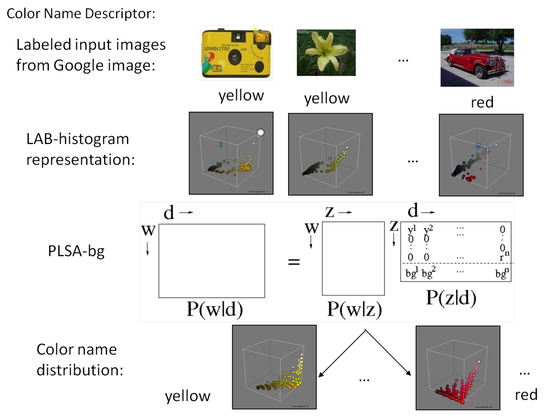
$$ang^O_{\mathbf{x}} = \arctan\left(\frac{O1_{\mathbf{x}}}{O2_{\mathbf{x}}}\right) \rightarrow s = \sqrt{O1^2_x + O2^2_x} \quad (2)$$

Color names: are linguistic color labels which human assign to colors in the world. In this work we use the mapping learned from Google images [Van de Weijer, TIP09].

$$CN = \{p(cn1|R), p(cn2|R), ..., p(cn11|R)\} \quad (3)$$

with

$$p(cn_i|R) = \frac{1}{p}\sum_{x\in R} p(cn_i|\mathbf{f}(x)) \quad (4)$$



Properties:
Compactness: only an 11D histogram for each cell is computed.
Invariance: possess a degree of photometric invariance.
Discriminative power: separate bins represent the achromatic colors: black, grey, and white.

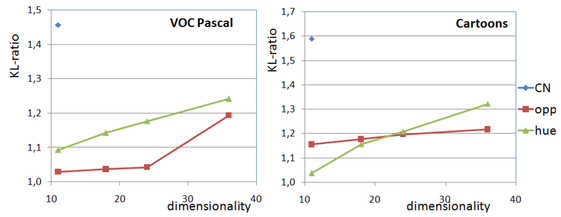
Color Descriptor Selection

Evaluation criteria: To avoid laborious cross validation, we use the KL-ratio to compare the discriminative power. A high KL-ratio reflects a more discriminative descriptor.

$$\text{KL-ratio} = \frac{\sum_{k\in C^m} \min_{j\notin C^m} KL(p_j, p_k)}{\sum_{k\in C^m} \min_{i\in C^m, i\neq k} KL(p_j, p_k)} \quad (5)$$

where

$$KL(p_i, p_j) = \sum_{x=1}^{N} p_i(x) \log \frac{p_i(x)}{p_j(x)} \quad (6)$$



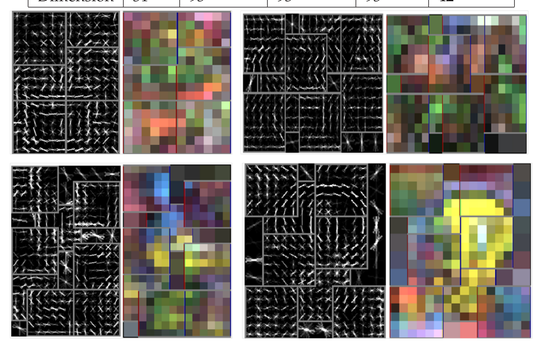
Conclusion: We select color names since it has more discriminative power while being compact.

Coloring Object Detection

Part-Based: We extend the HOG feature with color attributes within the part-based framework [Felzenswalb, PAMI11].

$$C_i = [HOG_i, CN_i] \quad (7)$$

Feature	HOG	OPPHOG	RGBHOG	C-HOG	CN-HOG
Dimension	31	93	93	93	**42**



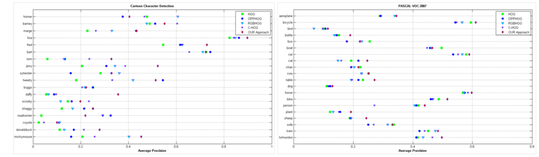
ESS-Based: We add a separate color vocabulary to the ESS framework [Lampert CVPR08]. Color and shape are combined with late fusion.

Cartoon Character Detection:

	SIFT	CN-SIFT	C-SIFT	OPPSIFT
mean AP	8.8	**12.9**	10.3	9.3

Experimental Validation

PASCAL VOC2007: no early fusion method improves over standard HOG.
Cartoon: Our approach provides a gain of 14% over standard HOG.



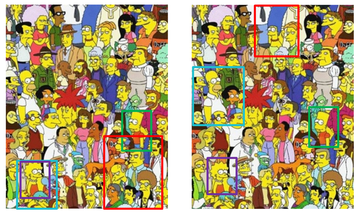
PASCAL VOC: Our approach improves on 15 out of 20 object categories compare to baseline.

Method	mean AP(VOC07)	mean AP(VOC09)
HOG	32.3	28.0
Best 2007	23.3	-
UCI	27.1	-
UOCTTI	-	27.9
LEO	29.6	-
Oxford-MKL	32.1	27.7
LBP-HOG	34.3	21.9
Our Approach	**34.8**	**29.4**

Method	mean AP (Cartoon)	mean AP (VOC07)
HOG	27.6	32.3
OPPHOG	34.2	30.6
RGBHOG	35.3	31.3
C-HOG	35.2	30.1
Our Approach	**41.7**	**34.8**

Cartoon Detection:

Conclusions



1. We propose the use of color attributes as explicit color representation.
2. We introduce a new dataset of cartoon character images.
3. Early fusion based approaches yield inferior results for object detection. Our approach achieves state-of-the-art on PASCAL VOC 2007, 2009 and cartoon datasets despite its simplicity.