Learning Robust Color Name Models from Web Images

Boris Schauerte, Rainer Stiefelhagen

Computer Vision for Human Computer Interaction, Karlsruhe Institute of Technology, Germany

Contributions

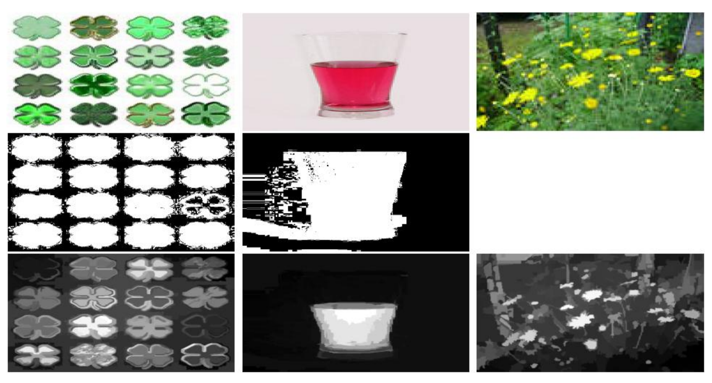
- State-of-the-art color naming with an emphasis on reducing “unnatural” naming mistakes
- ~14% improvement towards assigning natural, human-like color names
- Combines salient object detection, KLD outlier reduction, and supervised latent Dirichlet allocation

Motivation: Why learn color terms? What is an “unnatural” mistake?

Why: Reliable color naming and recognition is necessary for natural human-computer/-robot interaction, because it is one of the most common attributes used to communicate and reference objects.

Unnatural: Color terms have fuzzy boundaries and thus, e.g., describing the same object as “yellow” or “orange” may be equally appropriate. However, describing the object as “blue” would be a “unnatural” mistake that humans hardly ever make.

(Salient) Object Detection



Suppress non-relevant, distracting image content

a) Detect and *remove the image’s “border”*
b) Use *salient object detection* to highlight the potentially most relevant image part

Outlier Reduction

Remove outliers in the training data (see figure at the top, 3rd row)

1) Estimate a simple, *initial color model and calculate the Kullback-Leibler divergence* (KLD) between each image and the initial model

$$d^{zd}_{\mathrm{KLD}}(P'(\cdot|z), P(\cdot|d)) = \sum_w P(w|d) \ln \frac{P(w|d)}{P'(w|z)}$$

$$\text{with} \quad P'(w|z) = \frac{1}{N} \sum_{d \in D, l_d = z} P(w|d) \,.$$

2) Rank the images and *select a subset with the lowest KLD ratio*

$$R^{zd}_{\mathrm{KLD}} = \frac{d^{zd}_{\mathrm{KLD}}}{\min_{z' \neq z} d^{z'd}_{\mathrm{KLD}}}$$

z: term, w: word, d: document, l_d: label of document d

Learning and Classification

Representation: Treat image regions as documents. Color *histograms form a bag-of-pixel representation*, in which the histogram entries are words in the document.

Learning: Use multi-class *supervised latent Dirichlet allocation (SLDA)* to …

1) learn the *“topics” as latent feature* for classification
2) learn to use the histogram band *topic assignments for classification*

I.e., learn topics that are predictive for the class labels in combination with class coefficients for each topic.

Classification: *softmax regression* on basis of the topic assignments, class coefficients, and topic frequencies.

Evaluation

Training Data Set: Google-512

512 images from Google’s image search for each of the 11 basic English color terms

Test Data Set: eBay+

440 segmented eBay images of objects (4 object classes, 10 evaluation images for each of the 11 basic color terms); plus, for each image, the color term labels of 5 human subjects

	Cars	Pott.	Shoes	Dress	Total	Dist*
Ours	73.63	80.90	91.82	90.00	**84.09**	**0.50**
$\mathbf{X^2}$ **rank**	73.63	79.01	92.73	88.18	83.41	0.57
PLSA	71.82	83.64	92.73	86.36	83.64	0.73
Human	92.73	87.82	90.18	91.99	90.64	

* Dist: distance between the confusion matrices of the human observers and the trained model

Contact

{schauerte, rainer.stiefelhagen}@kit.edu

Computer Vision for Human Computer Interaction

http://cvhci.anthropomatik.kit.edu