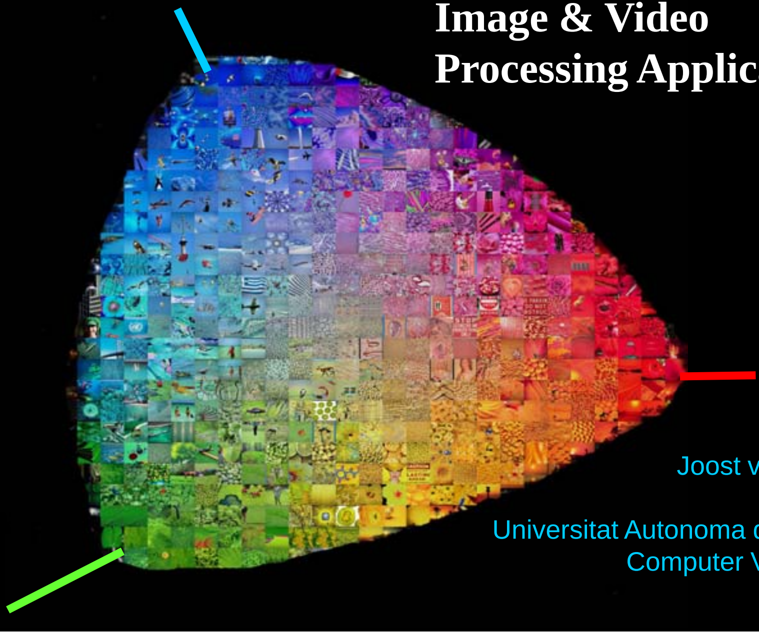
Color in Image & Video Processing Applications

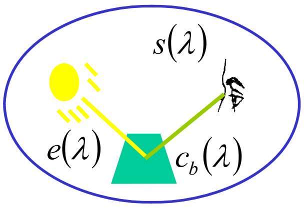


DAGM 2010
Darmstadt

Joost van de Weijer

Universitat Autonoma de Barcelona
Computer Vision Center

overview



THEORY

PART I: features
- color spaces
- color feature detection
- saliency detection.

PART II: color constancy
- bottom-up color constancy
- top-down color constancy
- color constant features

APPLICATIONS

IMAGE CLASSIFICATION
- combining color and shape

IMAGE SEGMENTATION
- deviations from the reflection model

OBJECT RECOLORING
- complex reflection models

Application I: Image Classification

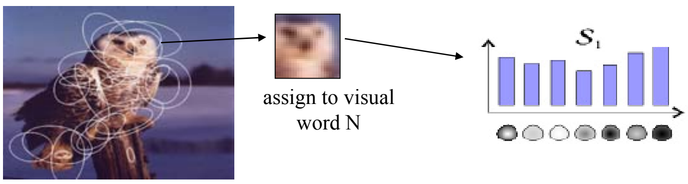
- robust color features
- color saliency
- how to combine features

Coloring Color Feature Extraction, ECCV 2006.
Top-down color-attention for object recognition, ICCV 09.

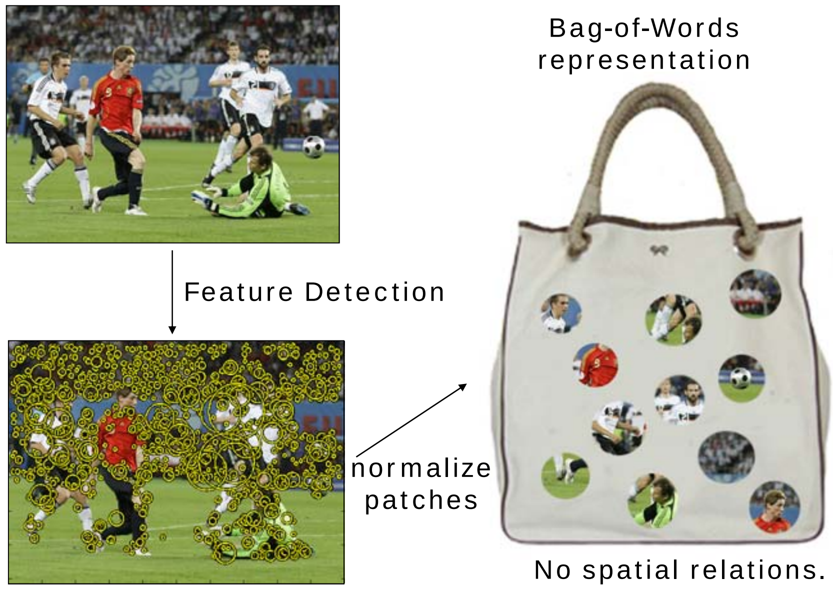
from images to frequency histogram

•Compute visual words:
- detect local regions from a set of images.
- describe every local region by a descriptor
  - texture
  - color
- cluster all descriptors into visual words

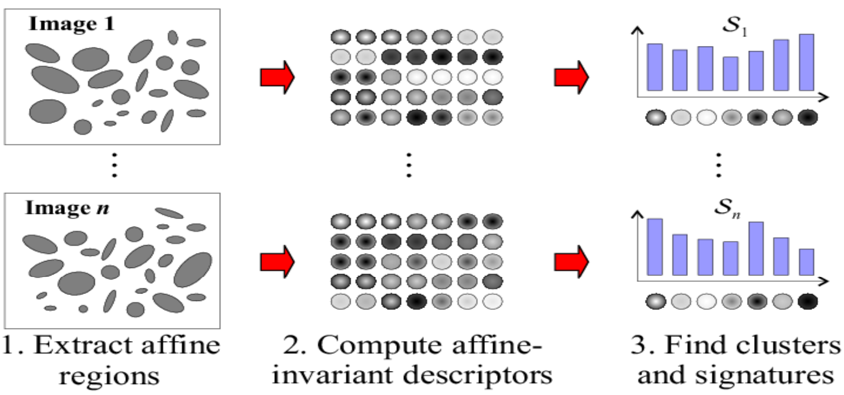
Given a new image:
- detect local regions from a set of image.
- assign every region to its nearest visual word.
- compute visual word-image histogram

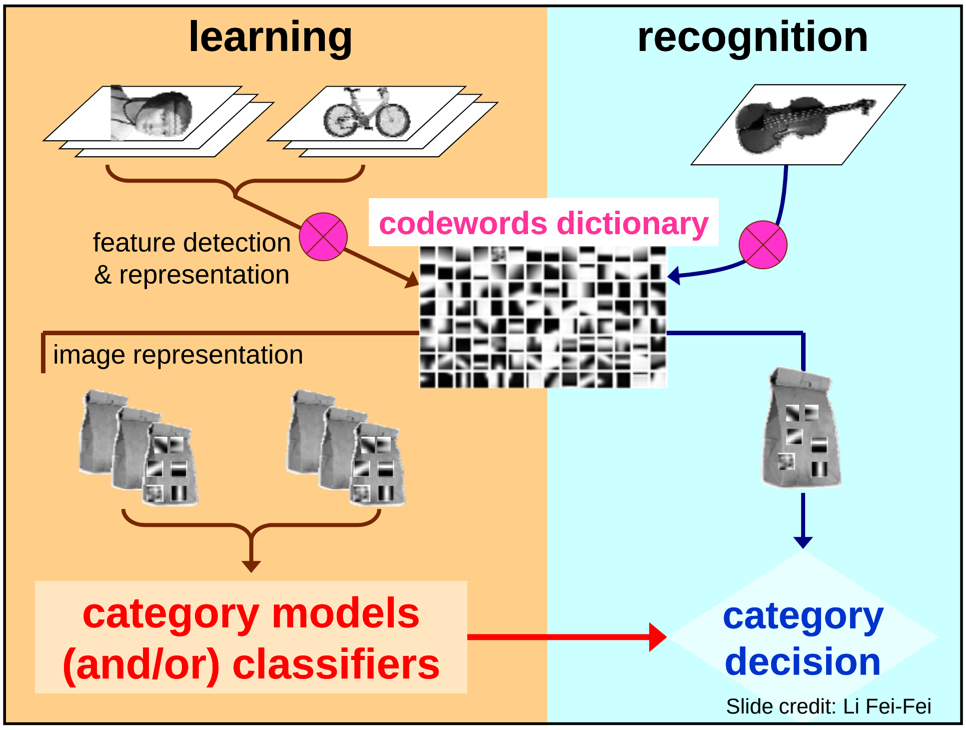


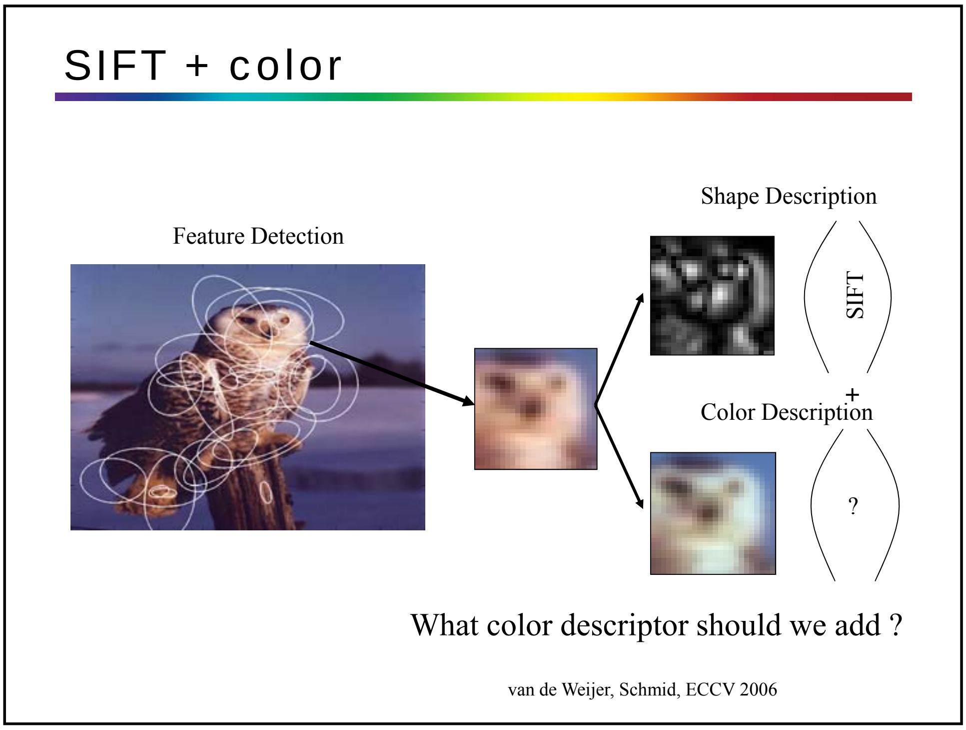
Bag of Visual Words representation



Bag of Visual Words



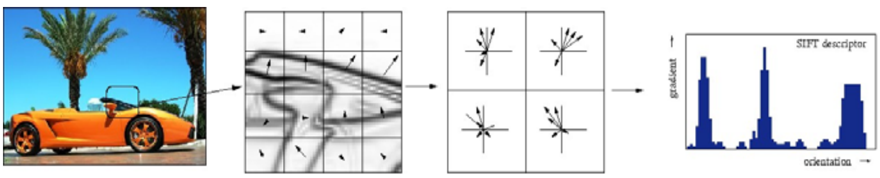

learning
recognition
codewords dictionary
feature detection
& representation
image representation
category models
(and/or) classifiers
category
decision
Slide credit: Li Fei-Fei


SIFT + color
Feature Detection
Shape Description
SIFT
+
Color Description
?
What color descriptor should we add ?
van de Weijer, Schmid, ECCV 2006

Color Descriptors -Hue

Shape descriptor

- **SIFT :**



- **robust update:** the update can also be derived from error analysis:

$$\varphi = \arctan\left(\frac{f_y}{f_x}\right) \Rightarrow \partial\varphi = \frac{1}{\sqrt{f_x^2 + f_y^2}} = \frac{1}{f_w}$$

Robust hue descriptor

HSI:

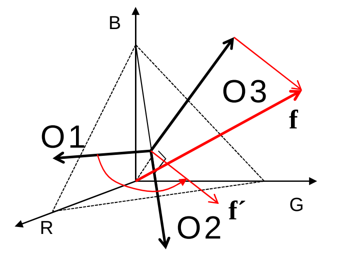
Orthonormal opponent space:

$$\begin{pmatrix} O1 \\ O2 \\ O3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

$$H = \arctan\left(\frac{O1}{O2}\right)$$

$$S = \sqrt{O1^2 + O2^2}$$

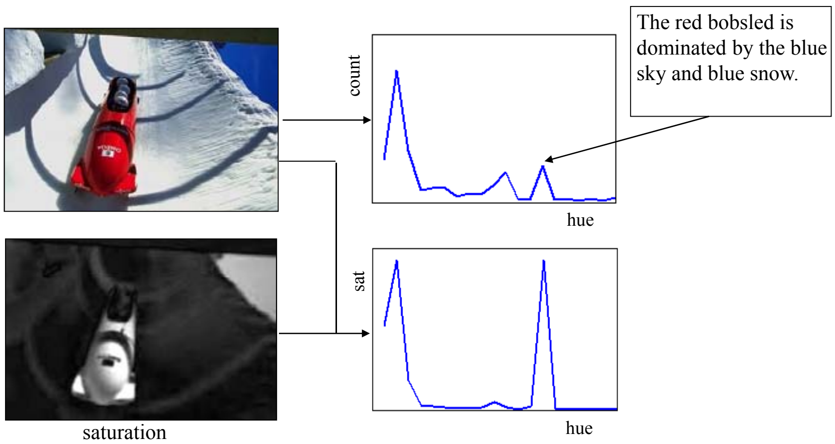
$$I = O3$$



Projection of **f** on **O3** is the *intensity*. Let **f**´ the projection on the plane O1-O2, its length is the *satutation* and its angle the *hue*.

- hue is invariant for specularities and shadow-shading effects.

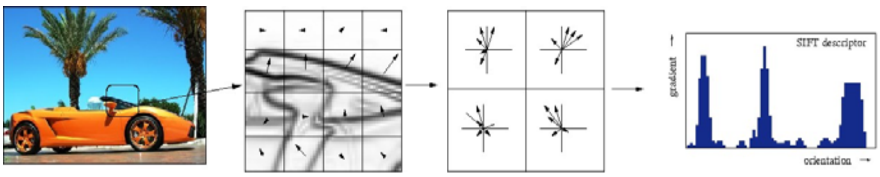
Robust hue descriptor



saturation

Feature descriptor

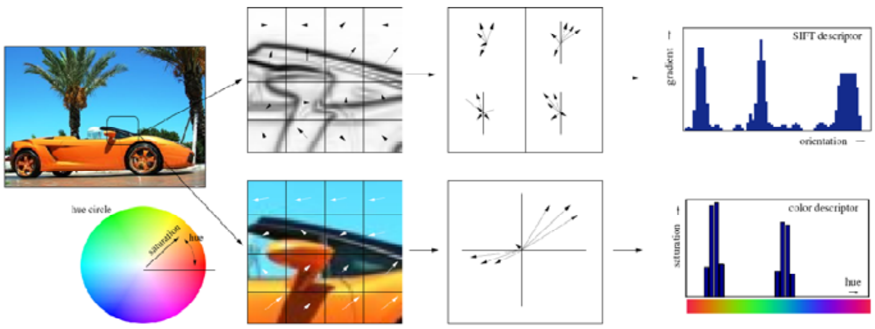
- **SIFT :**



- **robust update:** the update can also be derived from error analysis:

$$\varphi = \arctan\left(\frac{f_y}{f_x}\right) \Rightarrow \partial\varphi = \frac{1}{\sqrt{f_x^2 + f_y^2}} = \frac{1}{f_w}$$

Photometric Invariant Color Description



zero-order:

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

first-order:

$$ang_x^O = \arctan\left(\frac{O1_x}{O2_x}\right) \rightarrow O_x = \sqrt{O1_x^2 + O2_x^2}$$

Color Descriptors –Color Names

learning color names

task: Object colors in many images are often not explicitly labeled. Can we label these image automatically with color names ?

Ebay user: “Find me all yellow cars ?”

learning color names

From linguistic studies it is known that the development of color names follows a similar pattern for all languages.

Development color names in languages:

white
black
red
green
yellow
blue
brown
purple
pink
orange
grey

The english language has 11 basic color terms.

learning color names

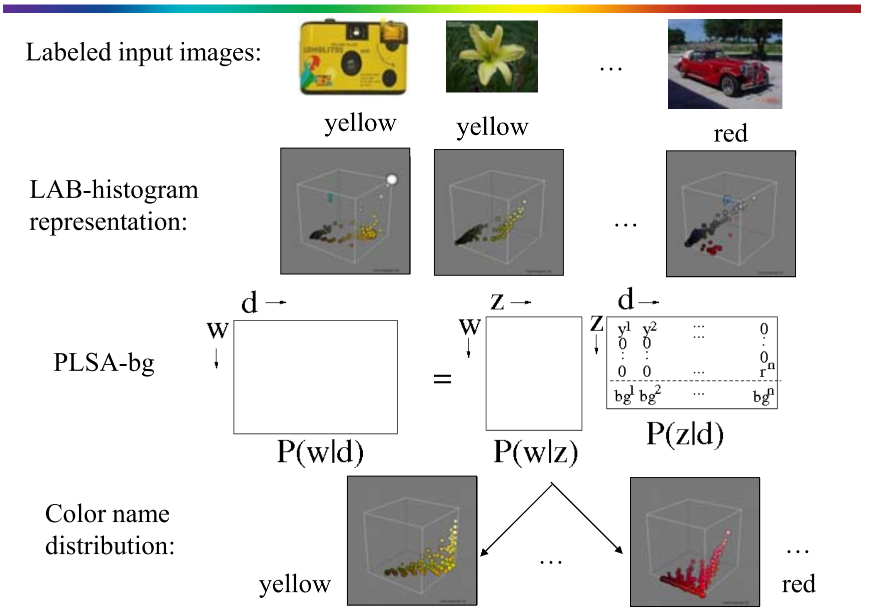
• Use google image to assemble a set of weekly labeled images.

black blue brown green orange purple white

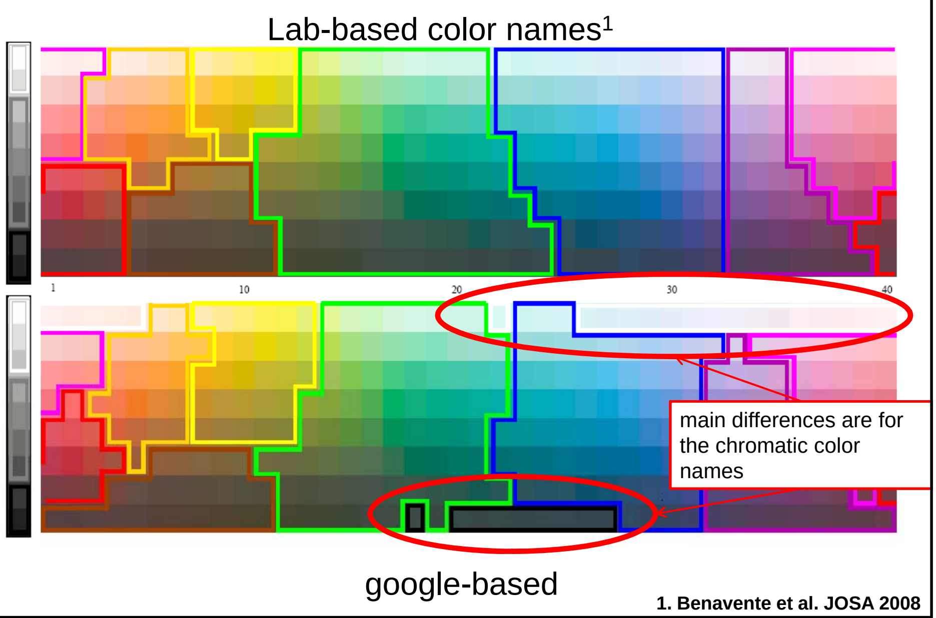
false positives

Images retrieved with Google image

learning color names



learning color names



retrieval of color names

EER	cars	shoes	dresses	pottery	overall
lab[1]	91	97	97	92	94.0
google	93	99	99	94	96.4

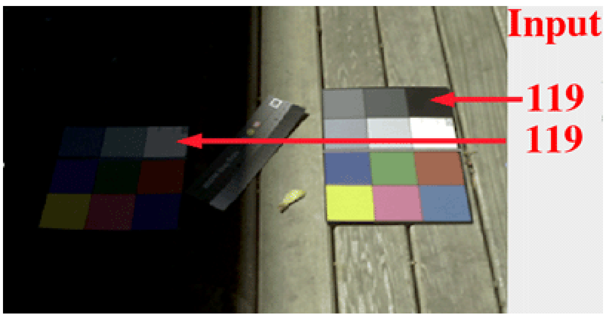
Ebay data set of 4 categories: shoes, cars, dresses, and pottery.

1.Menagaz, Eurasip 2007

retrieval of color names

EER	cars	shoes	dresses	pottery	overall
lab[1]	91	97	97	92	94.0
google	93	99	99	94	96.4

Errors are mainly due to absence of lightness estimation, which is a very little studied problem in computer vision.



Images courtesy John McCann

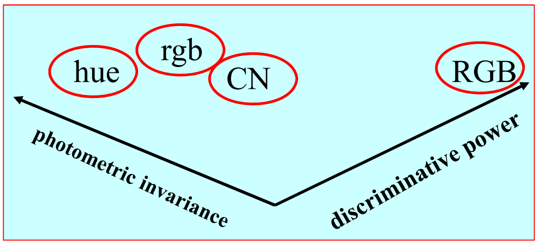
Color name descriptor

• Achromatic colors are very abundant in the world, about 45 % (67 % with brown) .

black	blue	brown	grey	green	orange	pink	purple	red	white	yellow
19	12	23	19	10	2	2	2	4	6	1

statistics based 40.000 corel images.

• when using photometric invariance always consider discriminative power.



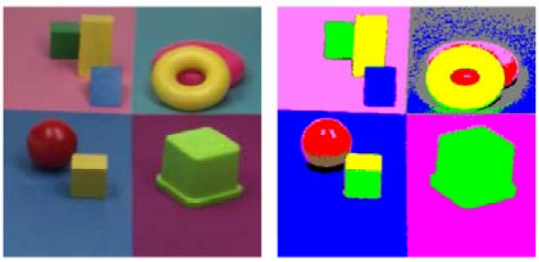
Color name descriptor

• Achromatic colors are very abundant in the world, about 45 % (more than 60 % with brown) .

black	blue	brown	grey	green	orange	pink	purple	red	white	yellow
19	12	23	19	10	2	2	2	4	6	1

statistics based 40.000 corel images.

• when using photometric invariance always consider discriminative power.



Color name descriptor

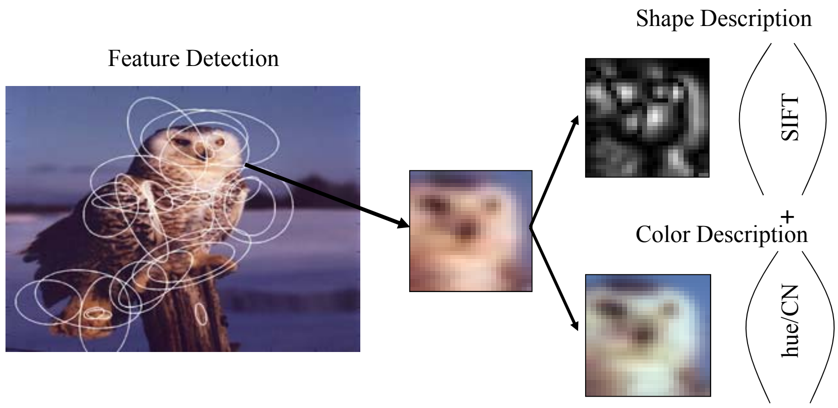
• test color names for image classification on a flower data set of 1360 images over 17 classes.

dataset		flower	
method	color	shape	color & shape
HSV-SIFT	-	-	78
hue	40	65	79
opponent	39	65	79
color names	57	65	81

references: color naming

- B. Berlin, P. Kay. *Basic Color terms: their universality and evolution.* Berkeley: University of California, 1969.
- A. Mojsilovic. *A computational model for color naming and describing color composition of images.* IEEE TIP 14(5), 2005.
- K. Yanai, K. Barnard, Image region entropy: a measure of *visualness* of web images associated with on concept, ACMMM 2005.
- R. Benavente, M.Vanrell, R. Baldrich. *Parametric fuzzy sets for automatic color naming*, JOSA 25(10), 2008.
- G. Menegaz, A. L. Troter, J. Sequeira, and J. M. Boi, "A discrete model for color naming," EURASIP Journal on Advances in Signal Processing, vol. 2007, 2007.
- J. van de Weijer, Cordelia Schmid, Jakob Verbeek, Diane Larlus. *Learning Color Names for Real-World Applications*. IEEE TIP 2009.

SIFT + color



van de Weijer, Schmid, ECCV 2006

Results soccer data set:

	Shape	Color	Shape & Color
Hue	58	75	84
color names	58	86	89

van de Weijer, Schmid, ECCV 2006

Results flower data set:

• test color names for image classification on a flower data set of 1360 images over 17 classes.

dataset		flower	
method	color	shape	color & shape
HSV-SIFT	-	-	78
hue	40	65	79
opponent	39	65	79
color names	57	65	81

van de Weijer, Schmid, ECCV 2006

Results: Pascal Challenge 2007

	INRIA (genetic)	INRIA (flat)	XRCE	TKK	QMUL (lspch)	QMUL (hsls)
aeroplane	0.775	0.748	0.723	0.714	0.716	0.706
bicycle	0.636	0.625	0.575	0.517	0.550	0.548
bird	0.561	0.512	0.532	0.485	0.411	0.357
boat	0.719	0.694	0.689	0.634	0.655	0.645
bottle	0.331	0.292	0.285	0.273	0.272	0.278
bus	0.606	0.604	0.575	0.499	0.511	0.511
car	0.780	0.763	0.754	0.701	0.722	0.714
cat	0.588	0.576	0.503	0.512	0.551	0.540
chair	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	6
cow	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	6
dining tal	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4
dog	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9
horse	0.775	0.765	0.757	0.726	0.715	0.715
motorbike	0.640	0.623	0.585	0.602	0.579	0.554
person	0.859	0.845	0.840	0.822	0.808	0.806
potted plant	0.363	0.353	0.326	0.317	0.156	0.158
sheep	0.447	0.413	0.397	0.301	0.333	0.358
sofa	0.506	0.501	0.509	0.392	0.419	0.415
train	0.792	0.776	0.751	0.711	0.765	0.731
tv/monitor	0.532	0.493	0.495	0.410	0.459	0.455
average	0.594	0.575	0.556	0.517	0.512	0.503

Concatenation seems not the best method to combine color and shape.

SIFT: 53.3 %
SIFT+hue : 54 %

Marszalek Pascal 2007

plsa-based image segmentation

• We use Probabilistic Latent Semantic Analysis (pLSA) to compute the semantic likelihood of an image.

An image is modeled as a mixture of semantic topics:

$$p(w|d) = \sum_z p(w|z)\, p(z|d)$$

image-specific mixture proportions

visual word   image   semantic topics

$$p(w|z) = \prod_{m=1}^{M} p(w^m|z)$$

{texture, color, position}

likelihood image $p(d) = \prod_w p(w|d)$

results segmentation

• color and shape obtain almost equal results.
• the combination of color (56%) and shape (61%) improves results significantly (75 %).

	Building	Grass	Tree	Cow	Sky	Aeroplane	Face	Car	Bicycle	Average
S	51.1 (12.1)	74.0 (10.8)	68.1 (15.8)	59.0 (15.3)	59.2 (6.4)	52.1 (16.5)	52.5 (12.9)	59.4 (14.9)	76.3 (5.8)	61.3 (3.1)
C	50.4 (13.8)	77.6 (12.8)	46.8 (24.6)	51.0 (17.4)	81.2 (10.5)	20.8 (13.9)	77.2 (13.1)	58.5 (15.3)	38.0 (17.3)	55.7 (5.3)
P	0.0 (0.0)	86.6 (5.5)	0.0 (0.0)	0.0 (0.0)	68.9 (6.0)	3.5 (6.3)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	17.7 (1.0)
SC	66.6 (10.3)	84.0 (8.8)	59.5 (18.9)	74.8 (16.6)	89.4 (3.5)	74.8 (9.0)	80.7 (8.3)	73.9 (9.3)	73.0 (8.1)	75.2 (3.4)
SP	58.0 (8.9)	76.1 (7.3)	**62.6** (19.8)	74.0 (11.1)	81.0 (4.3)	69.4 (14.4)	55.9 (12.6)	69.3 (11.9)	**76.7** (5.1)	69.2 (3.6)
CP	60.5 (13.5)	80.2 (12.3)	38.5 (22.4)	57.2 (20.7)	89.5 (6.1)	48.4 (13.3)	76.6 (10.9)	63.6 (14.1)	34.9 (13.0)	61.0 (4.5)
SCP	**70.5** (9.1)	**88.3** (7.9)	62.5 (15.3)	**77.8** (15.4)	**93.5** (3.0)	**86.7** (6.7)	**82.5** (7.5)	**76.2** (8.7)	71.3 (8.7)	**78.8** (3.5)

J. Verbeek, B. Triggs
Region classification with Markov field aspect models. CVPR 07.

Combining Color and Shape in BOW

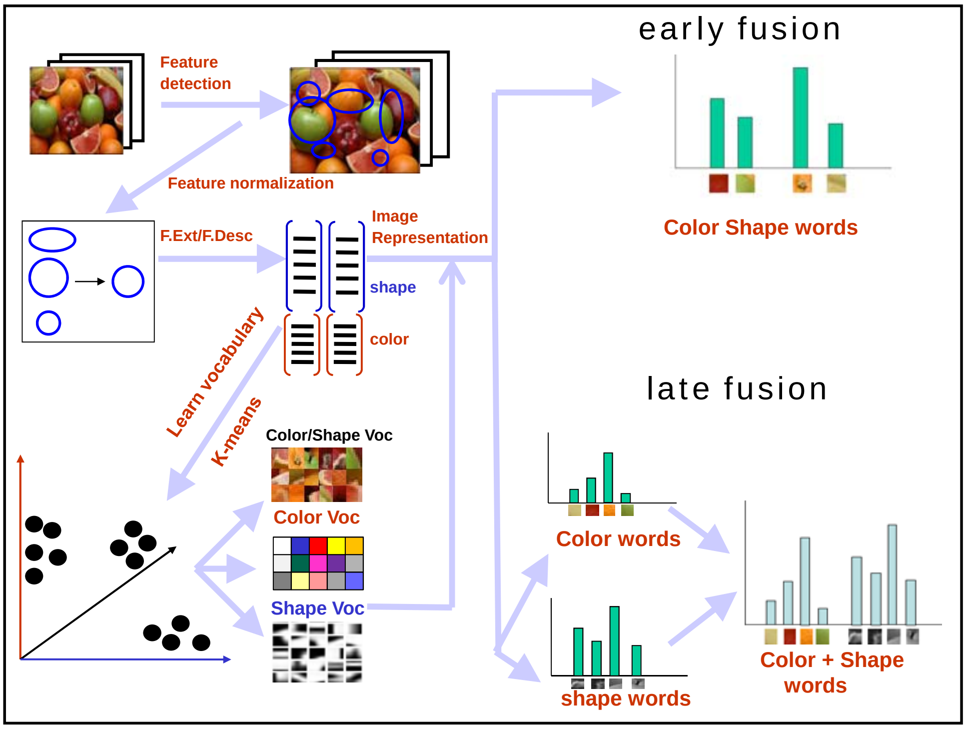
F. Shahbaz Khan, Joost van de Weijer, Maria Vanrell
Top-down color-attention for object recognition, ICCV 09.

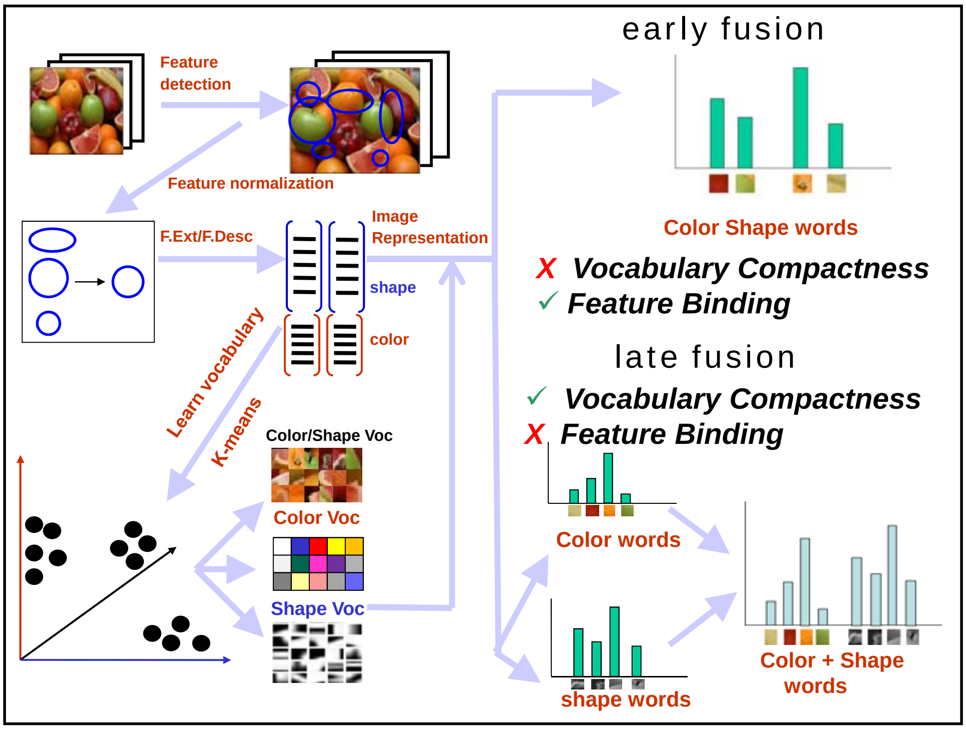
feature combination

desired properties:

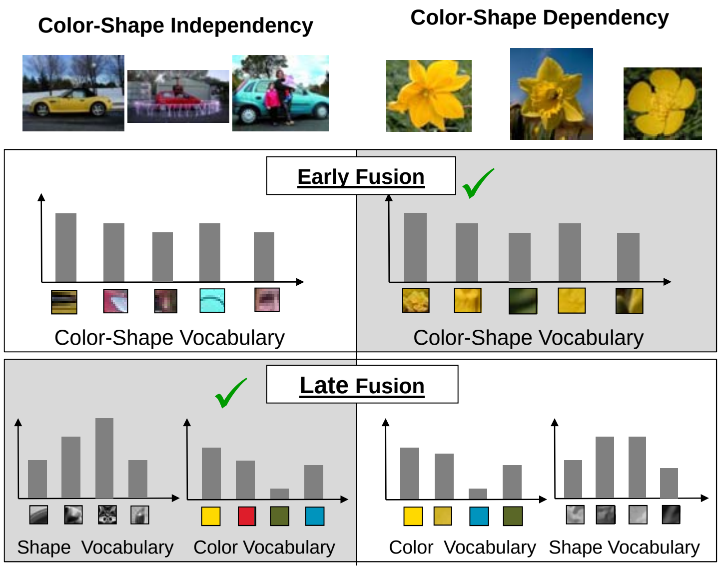
• ***vocabulary compactness*** is the property of having a separate vocabulary for each of the different cues.

• ***feature binding*** involves combining information from different cues at the local level (not at the image level).


early fusion
Feature detection
Feature normalization
F.Ext/F.Desc
Image Representation
shape
color
Learn vocabulary
K-means
Color/Shape Voc
Color Voc
Shape Voc
Color Shape words
late fusion
Color words
shape words
Color + Shape words

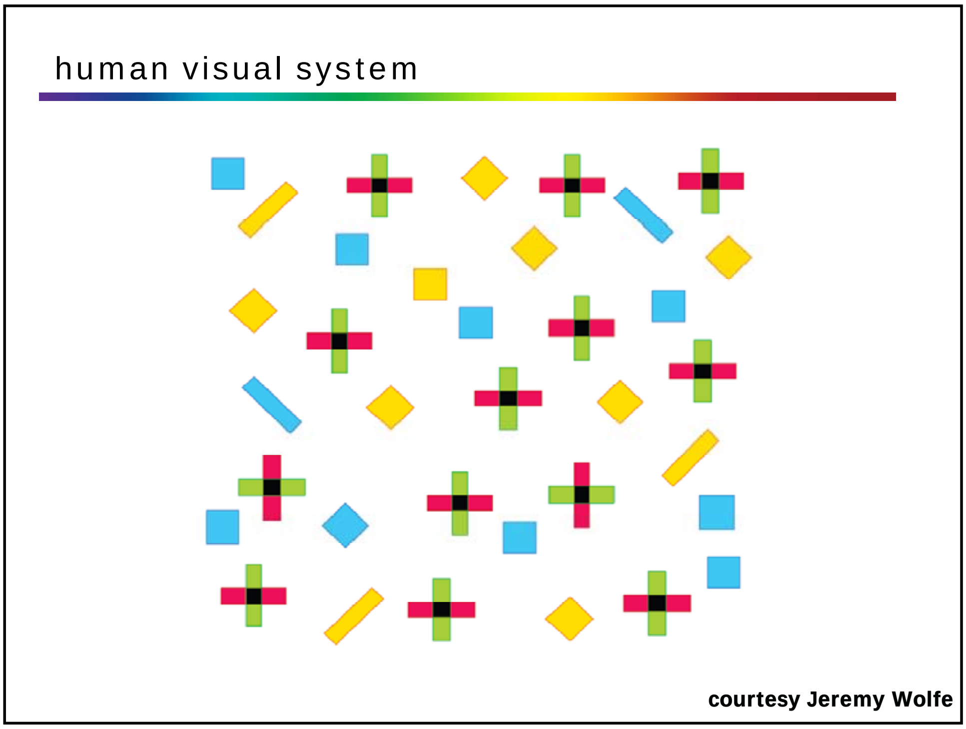

early fusion
Feature detection
Feature normalization
F.Ext/F.Desc
Image Representation
shape
color
Learn vocabulary
K-means
Color/Shape Voc
Color Voc
Shape Voc
Color Shape words
X Vocabulary Compactness
✓ Feature Binding
late fusion
✓ Vocabulary Compactness
X Feature Binding
Color words
shape words
Color + Shape words

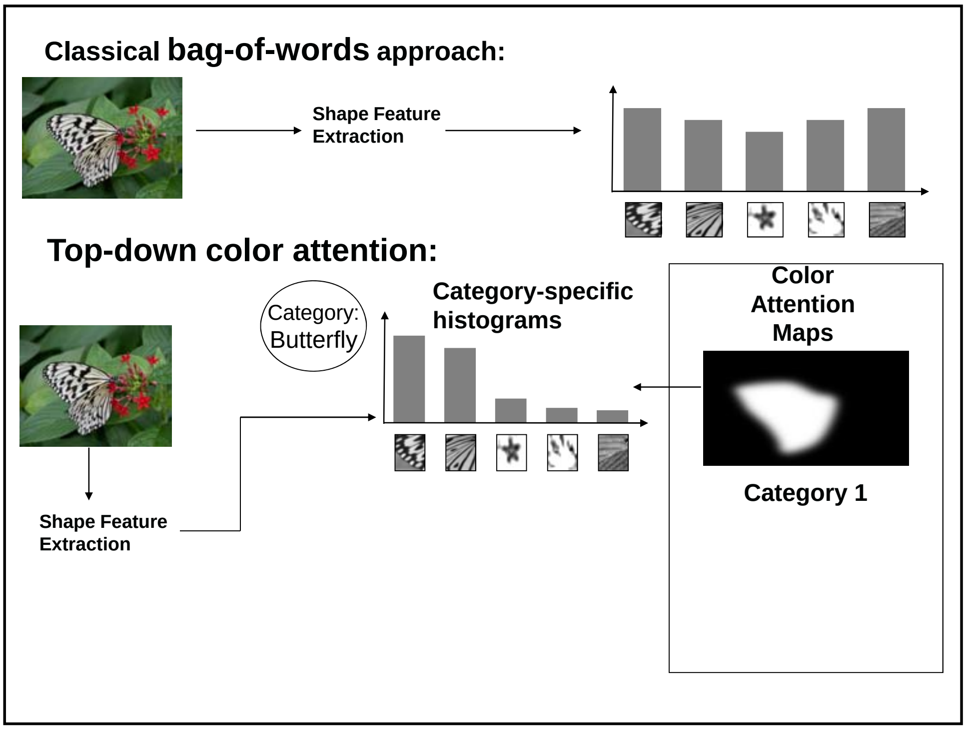
early vs. late fusion

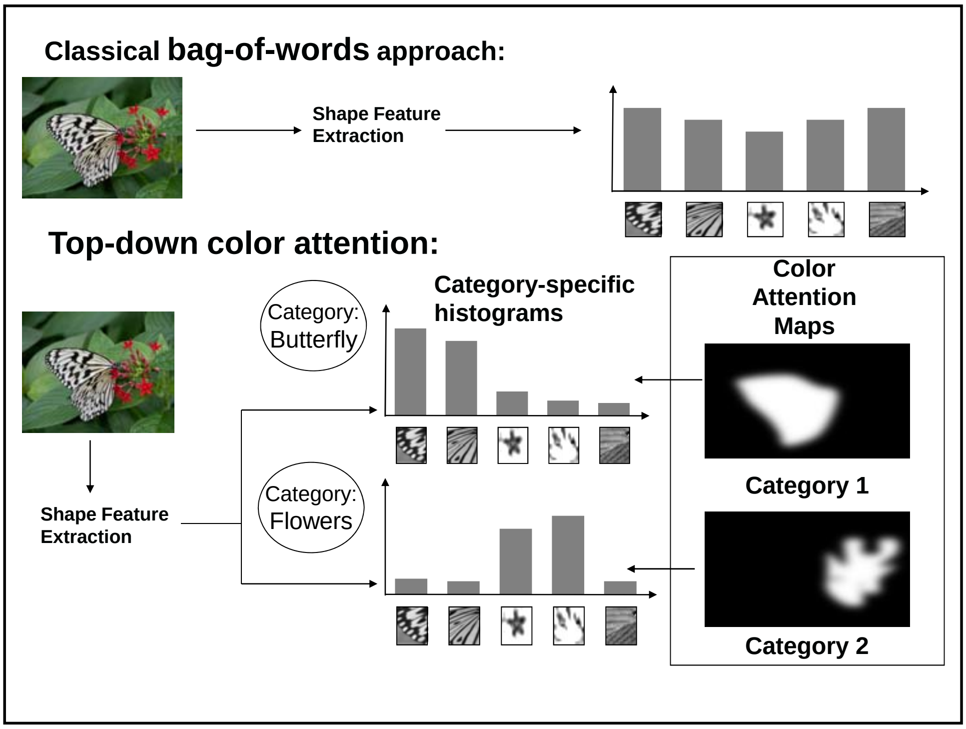


human visual system

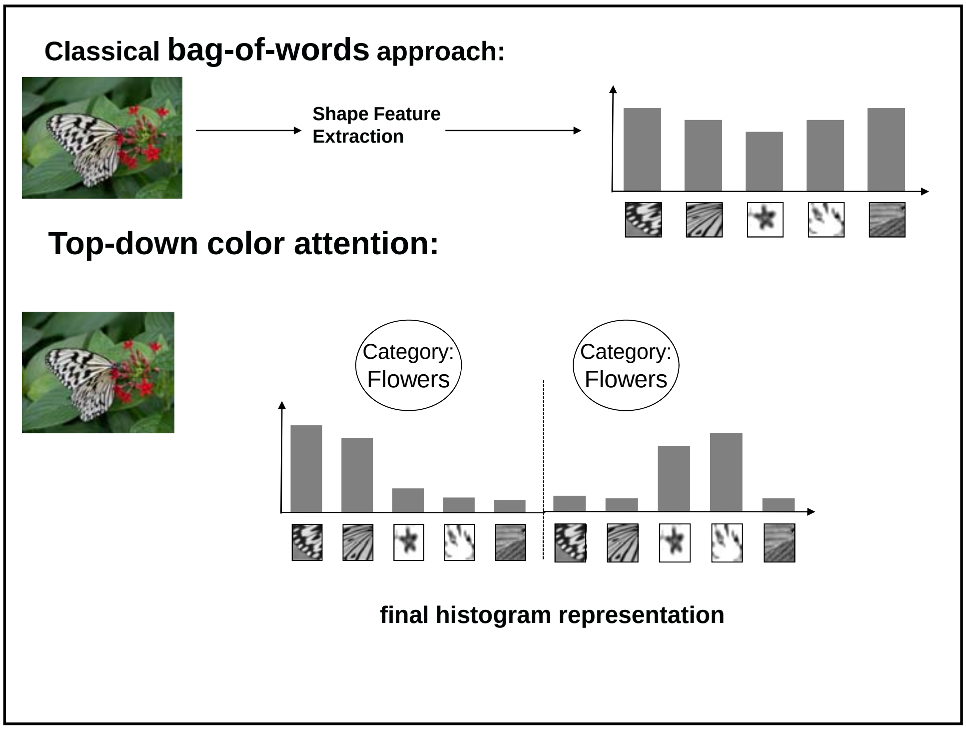
- different visual cues such as color and shape are processed in parallel (contrary to most computer vision approaches)
- the binding of the cues is done in the presence of visual attention.
- visual attention has both a bottom-up and top-down task-driven component.


human visual system
courtesy Jeremy Wolfe


Classical bag-of-words approach:
Shape Feature Extraction
Top-down color attention:
Category: Butterfly
Category-specific histograms
Color Attention Maps
Category 1
Shape Feature Extraction



Classical bag-of-words approach:
Shape Feature Extraction
Top-down color attention:
Category-specific histograms
Color Attention Maps
Category: Butterfly
Category: Flowers
Shape Feature Extraction
Category 1
Category 2



Classical bag-of-words approach:
Shape Feature Extraction
Top-down color attention:
Category: Flowers
Category: Flowers
final histogram representation

Top-down color attention:

Standard Bag-of-Words: $n\left(\mathrm{w}^s \middle| I\right) = \sum_{j=1}^{M} \delta\left(w_j^s, \mathrm{w}^s\right)$

Top-down CA: $n\left(w^s \middle| I, class\right) = \sum_{j=1}^{M} p\left(class \middle| w_j^c\right) \delta\left(w_j^s, w^s\right)$

Where $p\left(class \middle| w^c\right) \propto p\left(w^c \middle| class\right) p\left(class\right)$

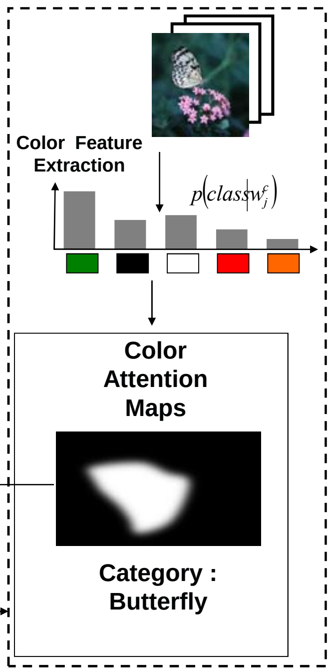
$$p\left(w^c \mid class\right) \propto \sum_{I^{class}} \sum_{j=1}^{M} \delta\left(w_j^s, w^c\right)$$

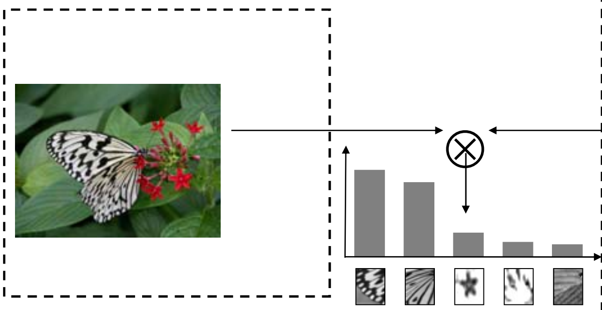
Top-down color attention:

- Color is not explicitly coded, but is present in the relative height of the shape-words for the various classes.

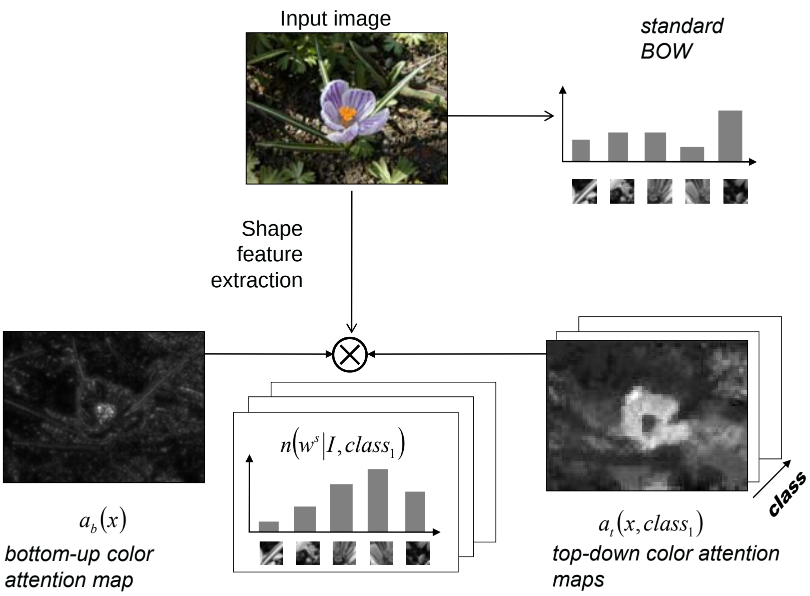
✓ Feature Binding
✓ Vocabulary Compactness

- Summing the histograms for all classes results in standard BOW.



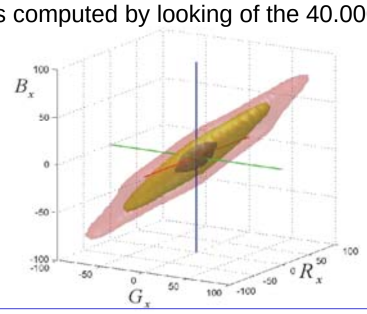


Bottom-up and Top-down Color Attention



bottom-up color attention:

- The statistics of $\mathbf{f}_x$ is computed by looking of the 40.000 images of the Corel database.



color boosting:

$$\mathrm{N} = \overline{\mathrm{f_x}\left(\mathrm{f_x}\right)^{\mathrm{t}}} = \begin{pmatrix} \overline{R_x R_x} & \overline{R_x G_x} & \overline{R_x B_x} \\ \overline{R_x G_x} & \overline{G_x G_x} & \overline{G_x B_x} \\ \overline{R_x B_x} & \overline{G_x B_x} & \overline{B_x B_x} \end{pmatrix}$$

$$\overline{R_x R_x} = \sum_{\mathbf{x} \in X^i} R_x(\mathbf{x})\, R_x(\mathbf{x}),$$

$$\mathrm{N} = \mathrm{U}\Lambda\Lambda\mathrm{U}^{\mathrm{t}}$$

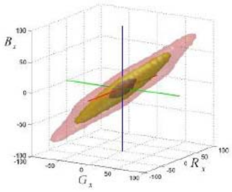
$$\mathrm{g}\left(\mathrm{f_x}\right) = \Lambda^{-1}\mathrm{U}^{\mathrm{t}}\mathrm{f_x}$$

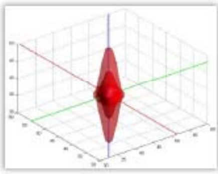
J. van de Weijer, Th. Gevers, A. Bagdanov, Boosting color saliency in image feature detection, IEEE PAMI 2006.

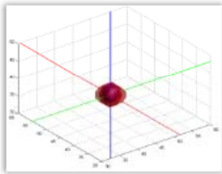
bottom-up color attention:

decorrelation

whitening







derivatives
RGB color space

derivatives
opponent color space

derivatives
color boosted space

color boosting:

$$\mathbf{N} = \overline{\mathbf{f_x}\left(\mathbf{f_x}\right)^{\mathbf{t}}} = \begin{pmatrix} \overline{R_\mathbf{x}R_\mathbf{x}} & \overline{R_\mathbf{x}G_\mathbf{x}} & \overline{R_\mathbf{x}B_\mathbf{x}} \\ \overline{R_\mathbf{x}G_\mathbf{x}} & \overline{G_\mathbf{x}G_\mathbf{x}} & \overline{G_\mathbf{x}B_\mathbf{x}} \\ \overline{R_\mathbf{x}B_\mathbf{x}} & \overline{G_\mathbf{x}B_\mathbf{x}} & \overline{B_\mathbf{x}B_\mathbf{x}} \end{pmatrix}$$

$$\mathbf{N} = \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^{\mathbf{t}}$$

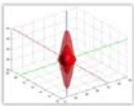
$$\overline{R_\mathbf{x}R_\mathbf{x}} = \sum_{\mathbf{x}\in X^i} R_\mathbf{x}\left(\mathbf{x}\right) R_\mathbf{x}\left(\mathbf{x}\right),$$

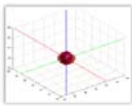
$$\mathbf{g}\left(\mathbf{f_x}\right) = \mathbf{\Lambda}^{-1}\mathbf{U}^{\mathbf{t}}\mathbf{f_x}$$

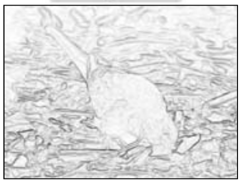
J. van de Weijer, Th. Gevers, A. Bagdanov, Boosting color saliency in image feature detection, IEEE PAMI 2006.

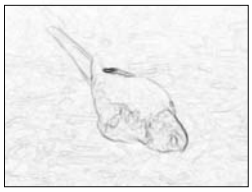
bottom-up color attention:

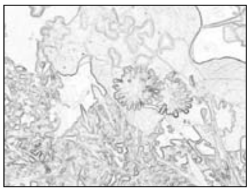
examples:

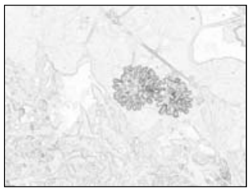












input image

color edges

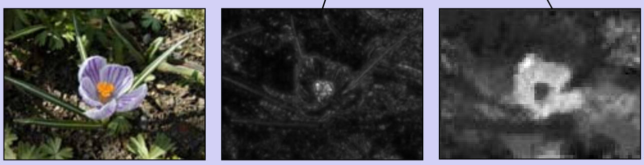
color boosted edges
bottom-up attention

Bottom-up and top-down color attention

Standard Bag-of-Words: $n\left(\mathrm{w}^s \middle| I\right) = \sum_{j=1}^{M} \delta\left(w_j^s, \mathrm{w}^s\right)$

Color Attention: $n\left(w^s \middle| I, class\right) = \sum_{j=1}^{M} a\left(x_j, class\right) \delta\left(w_j^s, w^s\right)$

$$a\left(x_j, class\right) = a^b\left(x_j\right) a^t\left(x_j, class\right)$$



bottom-up top-down

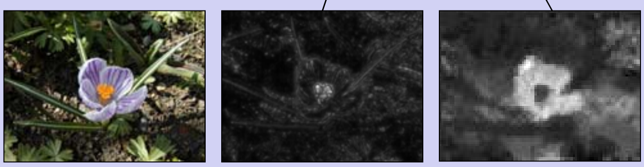
Bottom-up and top-down color attention

Standard Bag-of-Words: $n\left(\mathrm{w}^s \middle| I\right) = \sum_{j=1}^{M} \delta\left(w_j^s, \mathrm{w}^s\right)$

Color Attention: $n\left(w^s \middle| I, class\right) = \sum_{j=1}^{M} a\left(x_j, class\right)^{\gamma} \delta\left(w_j^s, w^s\right)$

$$a\left(x_j, class\right) = a^b\left(x_j\right)^{\beta} a^t\left(x_j, class\right)^{1-\beta}$$

γ weight color vs. shape
β weight BU vs. TD attention



bottom-up top-down

Experiments

Experimental Setup BOW

Image classification:

Color predominance Soccer data set

Color and shape parity Flower data set

Shape predominance Pascal Voc 2007 / 2009

Comparison:

Early Fusion, Late Fusion

OpponentSIFT, WSIFT

Feature Detection:

DOG detector (Soccer and flower data set)

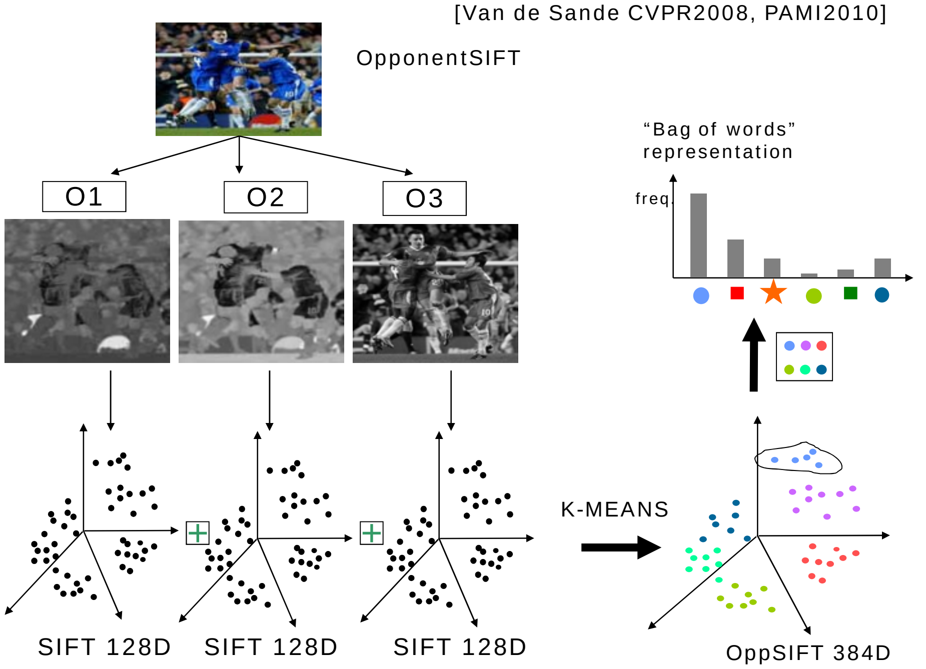
Multiscale Grid, Harris-Laplace and DoG (pascal 2007)

Feature Extraction:

SIFT (Shape), Color Names(Color), Hue (Color)

Learning:

Intersection kernel



Color Predominance

Soccer Data set:

Recognize the soccer team present in the image
Most dicriminative cue is player's outfit
Seven classes (40 images per class)
Training: 25 images per class, Testing: 15 images per class

Color Predominance

Method	Voc Size	Score
SIFT	400	50
Early Fusion	1200	88
Late Fusion	400+300	86
C-SIFT	1200	72
OpponentSIFT	1200	82
TD(SIFT,CN)	400,300	88
TD(SIFT,HUE)	400,300	82
TD(SIFT,{CN,HUE})	400,{300,300}	94
CA(SIFT,CN)	400,300	91
CA(SIFT,HUE)	400,300	88
CA(SIFT,{CN,HUE})	400,{300,300}	96

Best result reported 89 (Van de Weijer ICIP 07)

Color and Shape Parity

Flower Data set:

Recognize the flower-species in the image
Both color and shape are vital
Seventeen classes (80 images per class)
Training: 60 images per class, Testing: 20 images per class

Color and Shape Parity

Method	Voc Size	Score
SIFT	1200	63
Early Fusion	2000	85
Late Fusion	1200+300	84
C-SIFT	2000	77
OpponentSIFT	2000	83
TD(SIFT,CN)	1200,300	86
TD(SIFT,HUE)	1200,300	86
TD(SIFT,{CN,HUE})	1200,{300,300}	87
CA(SIFT,CN)	1200,300	90
CA(SIFT,HUE)	1200,300	89
CA(SIFT,{CN,HUE})	1200,{300,300}	91

best reported 89% by Xie CVPR 2010

Shape Predominance

Pascal Voc 2007:

Recognize objects from number of object classes in realistic scenes

Shape is the dominant cue

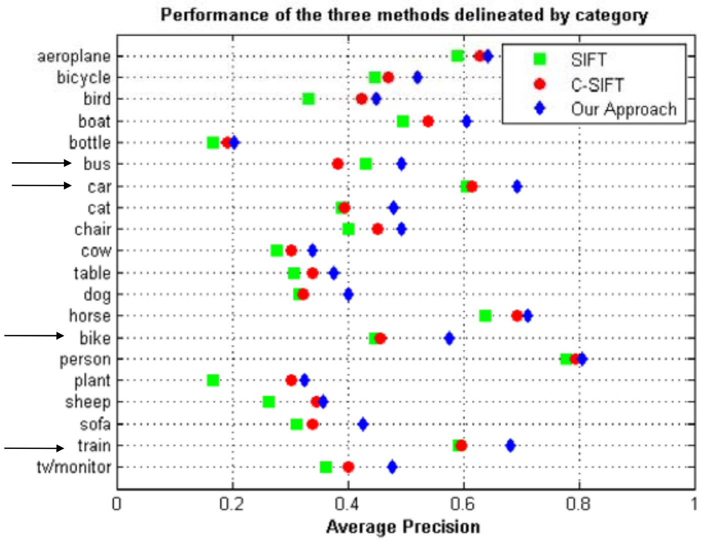
Twenty classes (9963 images), Training: 5011, Test: 4952

Animal: bird, cat, cow, dog, horse, sheep
Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
Indoor: bottle, chair, dining table, potted plant, sofa, tv
Person: person

Shape Predominance

Method	Voc Size	Mean AP
SIFT	4000	53.7
TD(SIFT,CN)	4000,500	56.8
TD(SIFT,HUE)	4000,300	56.6
TD(SIFT,{CN,HUE})	4000,{500,300}	57.5
CA(SIFT,CN)	4000,500	57.5
CA(SIFT,HUE)	4000,300	57.0
CA(SIFT,{CN,HUE})	4000,{500,300}	**58.0**

Shape Predominance



PASCAL 2009

Pipeline overview

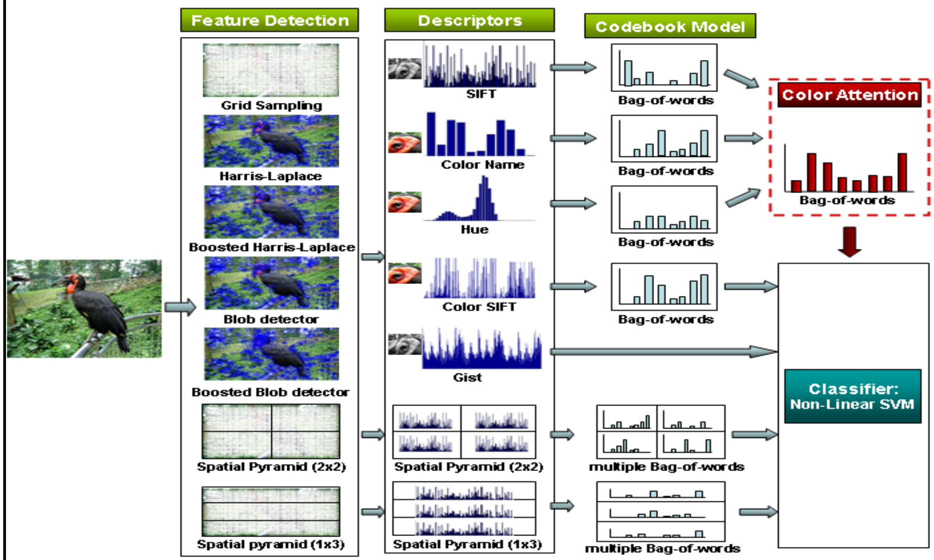
- Feature Extraction:
- SIFT [Lowe IJCV04]
- HUE [Van de Weijer ECCV06]
- Color Names [Van de Weijer CVPR07]
- Gist [Torrelba IJCV03]
- Color SIFT [Van de sande CVPR08]

- Codebook Construction:
- Kmeans Vocabulary with compression [Vedaldi ECCV08]

- Assignment:
- Soft Assignment [VanGemert ECCV08]

- Spatial Pyramids:
- 1x1 (Whole Image) , 2x2 (Image Quarters) [Lazebnik CVPR06], 1x3 (Horizontal Bars)

Pipeline overview



Classification Results

	Map on 2009 Val
SIFT	**51.0**
Color Attention	**56.2** + 5 %
Color Attention +ColorSIFT + GIST+Pyramids	**59.4** + 3 %

Classification + Detection

	Map on 2009 test
CVC-Flat (Classification)	**60.2**
CVC-Flat+ESS+HOG (Classification+Detection)	**61.1**

+ 0.9 %

[Harzallah, Schmid ICCV 09]

Conclusions

• We presented a method to combine color and shape information. Color is used as an attention cue to modulate the shape features.

• The attention maps are computed from *task-specific top-down* color attention and *image statistics based bottom-up* color attention.

• Method combines the advantages of early and late fusion:

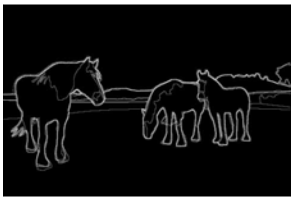
Vocabulary Compactness

Feature Binding

• The approach is shown to outperform both early and late fusion on several data sets.

Application II: Color Image Segmentation

Deviations from the dichromatic reflection model:
- jpeg compression
- unknown gamma

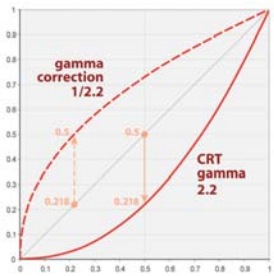


Describing Reflectances for Colour Segmentation Robust to Shadows, Highlights, and Textures, E. Vazquez et al. PAMI 2010.

gamma correction

- The dichromatic reflection model is only valid for *linear images.*
- *Gamma correction needs to be applied before processing:*

$$\mathbf{f}_{out} = \mathbf{f}_{in}^{\gamma}$$



- When gamma is unknown one often assumes sRGB with $\gamma = 2.2$
- Many computer vision operators are designed on non-linear images and gamma correction could deteriorate results (e.g. SIFT descriptor).

gamma correction

uncorrected – linear image

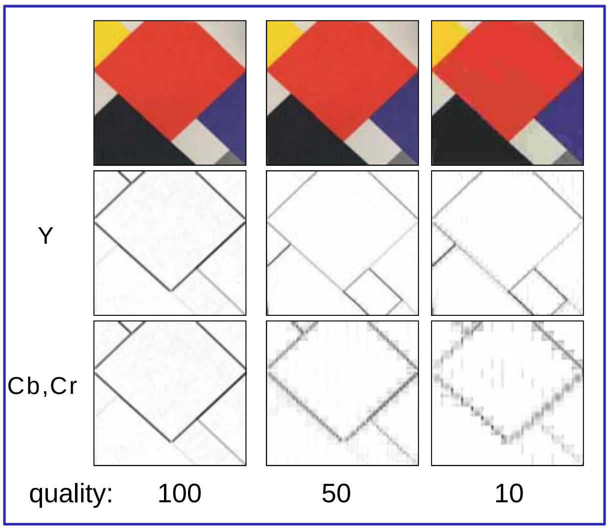
gamma correction

gamma corrected

jpeg compression

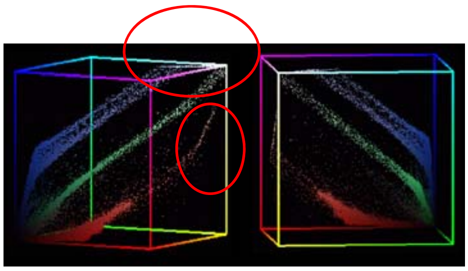
The human visual system is more sensitive for brightness than color changes. For this reason compression algorithms apply *chroma subsampling*.

This influences all color representations: *hue*, *saturation*, *opponent colors*.



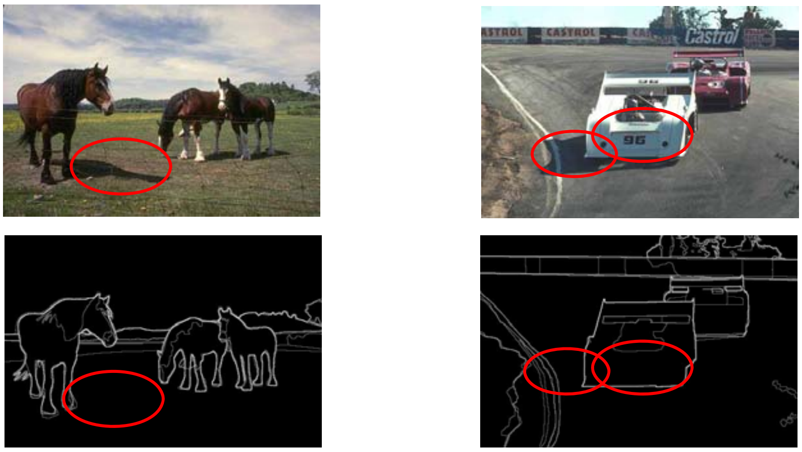
specularity clipping

- Often specularities are clipped, invalidating the reflection model for these values.
- Advisable to reject clipped RGB values (>250) before processing.
- Some information is still left after clipping and might be used [Werman 2010, CGIV 2010]



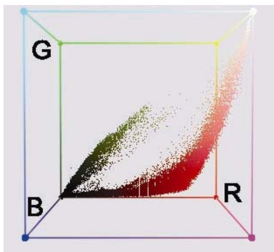
introduction

Problem statement: segmentation often fails in the presence of shadows and highlights



Source :The Berkeley Segmentation Dataset and Benchmark [Martin01]

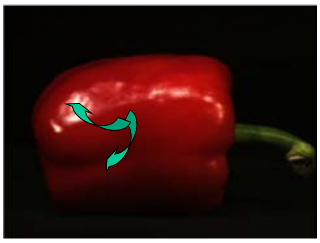
Dichromatic model

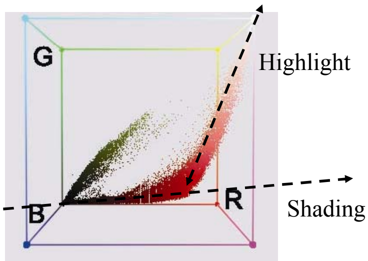


Distributions formed by a single-colored object have a physically determined shape in colour histogram-space.

Dichromatic model

Shape of the distribution is described by dichromatic reflection model.





Dichromatic reflection model

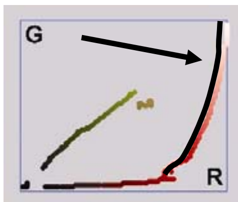
$$\mathbf{f}(\mathbf{x}) = m^b(\mathbf{x})\,\mathbf{c}^b + m^i(\mathbf{x})\,\mathbf{c}^i$$

Dichromatic model

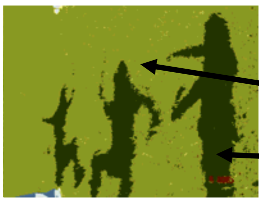
Problems:

- Non-linearities (*gamma, compressions, camera aberrations, interreflections)*

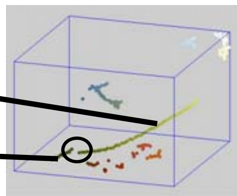
Non linearity



- Gaps (abrupt geometrical variations)

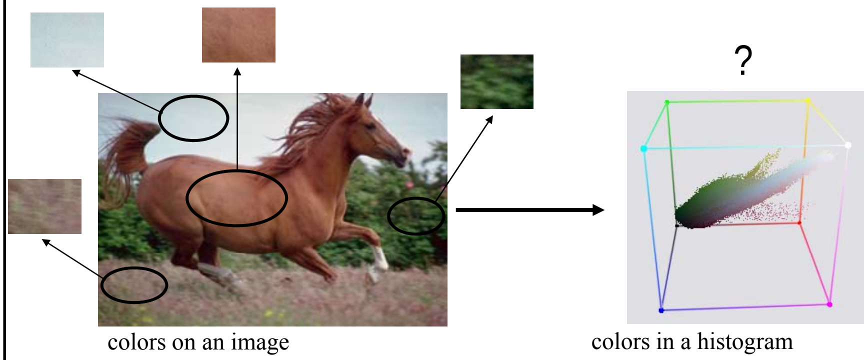


Gaps

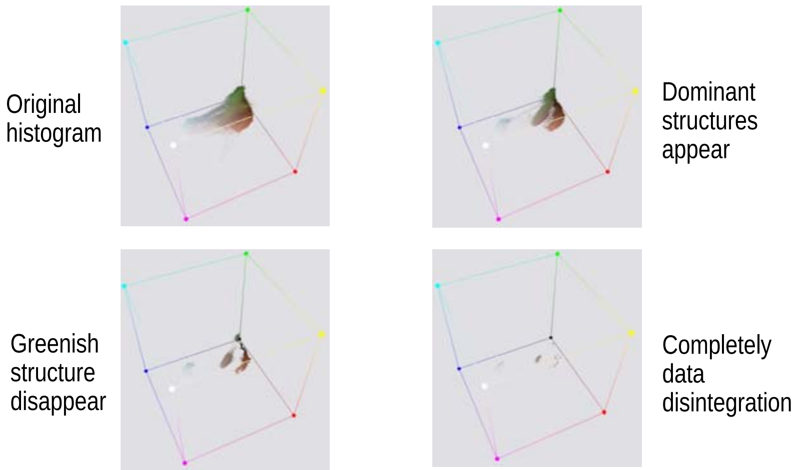


Color histogram representation

An example on a complex scene



Color histogram representation



Ridge Based Distribution Analysis (RAD)

The method is designed to avoid the main shortcomings of the dichromatic model.

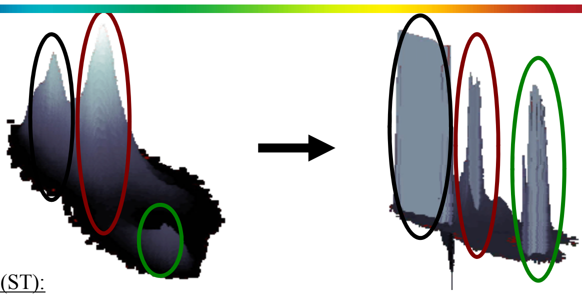
- Creaseness computation: to avoid gap problem

- Ridge extraction: to avoid limitations of linear assumption.

Ridge extraction

Distribution

3D view



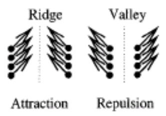
MLSEC-ST operator [Lopez00]

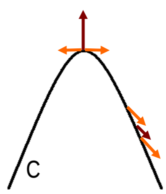
Creaseness

3D view

Structure Tensor (ST):

Computes the dominant gradient orientation in a neighbourhood of size proportional to a Gaussian with standard deviation σ_d



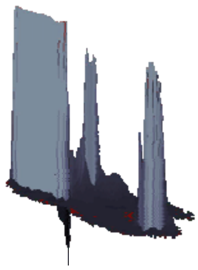


Vectors in opposite directions reinforce one another, whereas those regions with neither attraction nor repulsion remains unchanged.

Ridge extraction

Distribution

3D view



MLSEC-ST operator [Lopez00]

Creaseness

3D view

MLSEC:

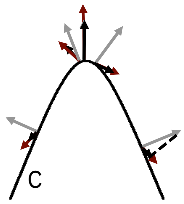
Computes the creaseness values as the **divergence of the dominant orientations** previously computed. A grade of concavity/convexity of ST, in a neighbourhood of size proportional to σ_i

$$\bar{\kappa}_d = -\mathrm{div}(\bar{\mathbf{w}}) = -\frac{d}{r}\sum_{k=1}^{r} \bar{\mathbf{w}}_k^t \cdot \mathbf{n}_k,$$

Normal vector

Normalization factor

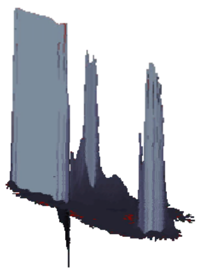
Dominant gradient orientation



Ridge extraction

Distribution

3D view



MLSEC-ST operator [Lopez00]

Creaseness

3D view

Distribution

2D view

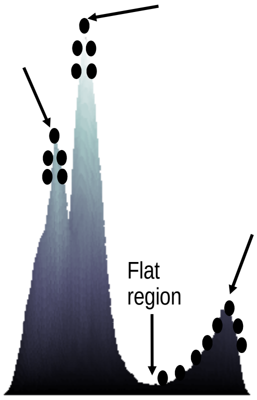
Creaseness

2D view

Ridge extraction

Ridge extraction algorithm:

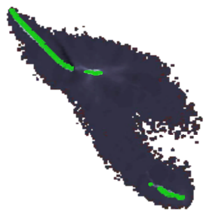
1. Find a local maxima. $\mathbf{m_i}$
2. Add this local maxima to the list of ridge points. $\mathbf{R=\{m_i\}}$
3. Remove m_i from the original distribution.
4. If we reach a flat area stop. Else, go to 1.



RAD clustering

Creaseness

2D view

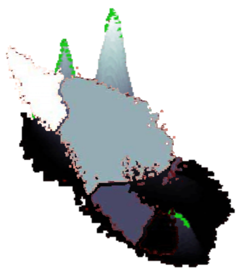


Ridges found

2D view

Ridges in
distribution

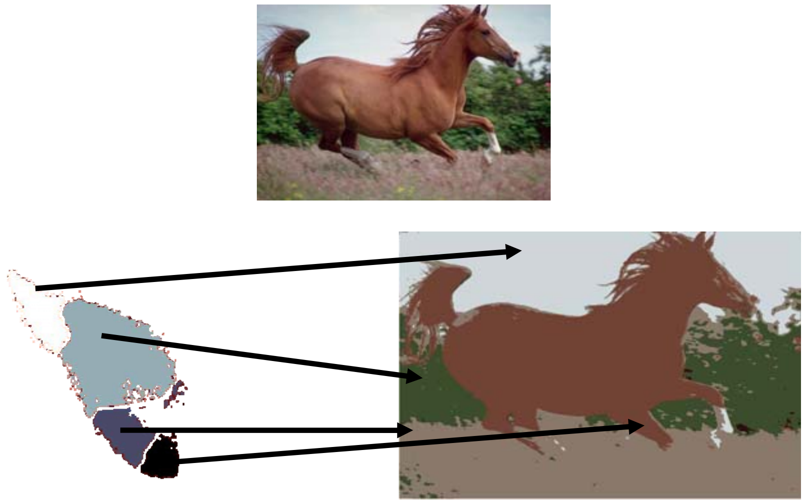
2D view



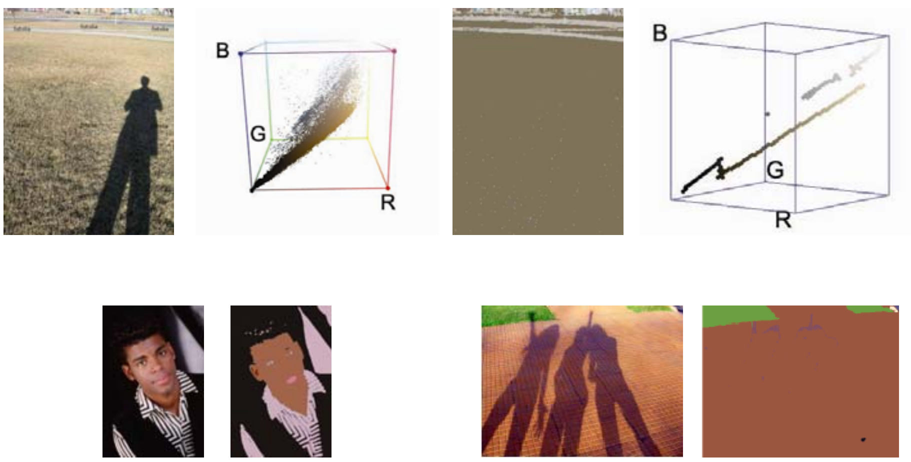
Clustering
2D view

Ridges in
distribution
3D view

RAD: results



RAD: results



RAD: results

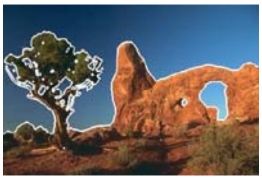
RAD: (σ_i, σ_d) - From a soft undersegmentation to a soft oversegmentation.

Original Image

RAD: set #1
6.04 s.

RAD: set #2
5.99 s.

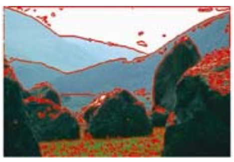
RAD: set #7
6.44 s.

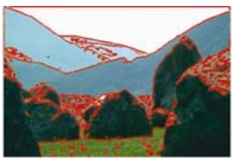


RAD: set #10
6.35 s.

RAD: results

Original image





RAD 1 RAD 2 RAD 3 Human segmentation

Performance evaluation

- Comparison with Mean Shift (MS)
  - MS is a well-known and widely used segmentation method.
  - MS and RAD are feature based methods which look for structures in the histogram.
- Qualitative evaluation on Berkeley image DB
- Quantitative evaluation using GCE [Martin01]

Comparing RAD and MS

RAD: set #1 RAD: set #2 RAD: set #3

Human Segmentation

MS: set #1 MS: set #2 MS: set #3

Comparing RAD and MS


Original image

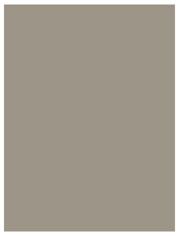

RAD


MS


Original image


RAD


MS


MS

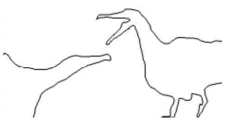
Quantitative comparison

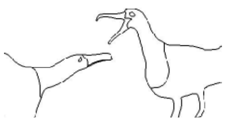
Quantitative comparison using Global Constancy Error (GCE) proposed by Martin *et al.*

(Berkeley Segmentation Dataset and Benchmark 01)

- GCE takes care of the refinement between different segmentations.


Original image


Human segmentation #1


Human segmentation #2


Human segmentation #3

Quantitative comparison

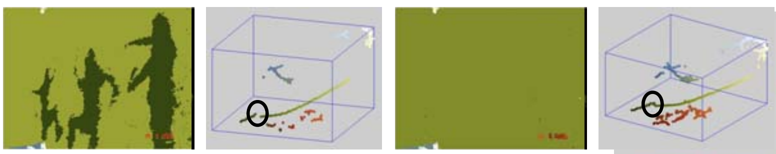
Global Constancy Error for several state-of the-art methods:

	GCE
Human segmentation	0.080
RAD(our approach)	0.204
Seed positioning (Micusık, ECCV06)	0.209
Affinity functions (Fowlkes CVPR03)	0.214
Mean shift (Comaniciu, PAMI02)	0.259
nCuts (Shi – PAMI00)	0.336

Eduard Vazuez et al. PAMI 2010

Conclusions

- RAD is a physics-based feature space segmentation method that extracts the Ridges formed by a dominant colour in an image.
- Overcomes limitations of the dichromatic model:



Non-linearity: solved
Gaps: solved with some limitations

Application III: Object recoloring

Extensions of dichromatic reflection model:
- ambient light
- multiple illuminants
- interreflections

Problems with the Dichromatic Reflection Models

Problems with Shafer's DRM:

- The model does not include interreflections.
- It is only valid for a single illuminant
- Shafer's original model represented ambient illumination as a constant (Later work dropped the ambient term because cameras couldn't measure it)

$$\mathbf{f} = m^b \mathbf{c_b} + m^s \mathbf{c_s} + \mathbf{c_a}$$

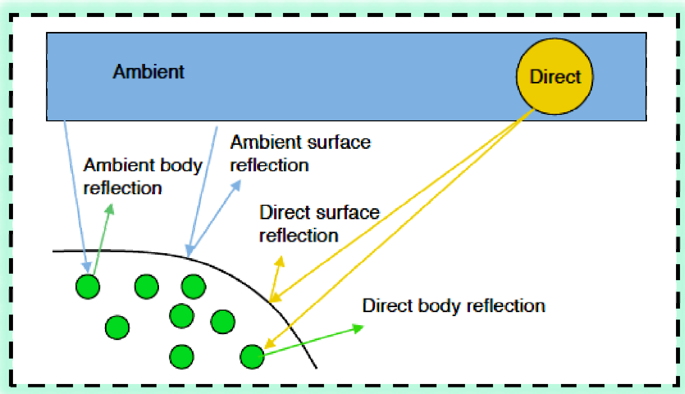
ambient lighting

• Maxwell et all. [CVPR 2006] propose the BIDR: bi-illuminant dichromatic reflection model.

Maxwell et all. [CVPR 2006]

BIDR model

Real-world Objects exhibit body and surface reflection under both direct and ambient illumination.



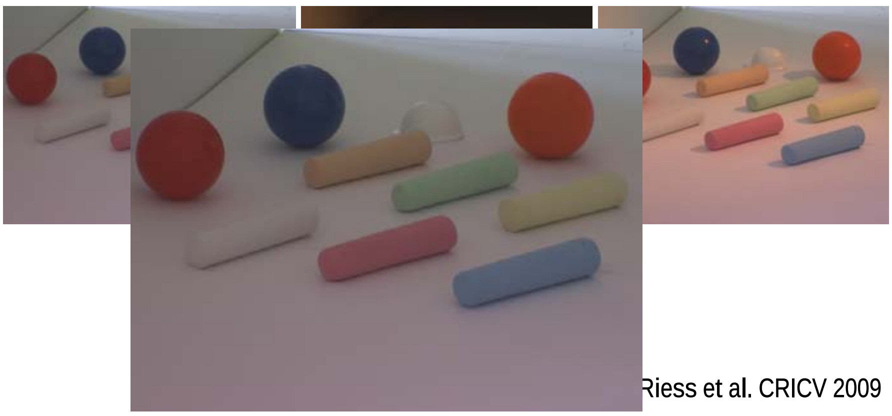
Maxwell et all. [CVPR 2006]

Ambient light example:

Under the Ambient light

Under the Direct light

Ambient & Direct



Riess et al. CRICV 2009

BIDR model

Bi-illuminant dichromatic reflection model [BIDR] model includes four terms:

$$\left.\begin{aligned} \mathbf{I}(\theta_e,\phi_e,\lambda) = {} & m_b(\theta_e,\phi_e,\theta_i,\phi_i)\mathbf{C}_b(\lambda) l_d(\theta_L,\phi_L,\lambda) + \\ & m_s(\theta_e,\phi_e,\theta_i,\phi_i)\mathbf{C}_s(\lambda) l_d(\theta_L,\phi_L,\lambda) + \end{aligned}\right\} \text{Dominant}$$

$$\text{Ambient} \left\{\begin{aligned} & \mathbf{C}_b(\lambda)\int_{\theta_i,\phi_i} m_b(\theta_e,\phi_e,\theta_i,\phi_i) l_a(\theta_i,\phi_i,\lambda) d\theta_i,\phi_i + \\ & \mathbf{C}_s(\lambda)\int_{\theta_i,\phi_i} m_s(\theta_e,\phi_e,\theta_i,\phi_i) l_a(\theta_i,\phi_i,\lambda) d\theta_i,\phi_i \end{aligned}\right.$$

I =image measurement
(θ_x, ϕ_x)=direction to the local surface normal
l_d=direct illuminant color and magnitude
l_a=ambient illuminant color and magnitude

$$I = \gamma_b F c_b l_d + \gamma_s F c_s l_d + F c_b M_{ab} + F c_s M_{as}$$

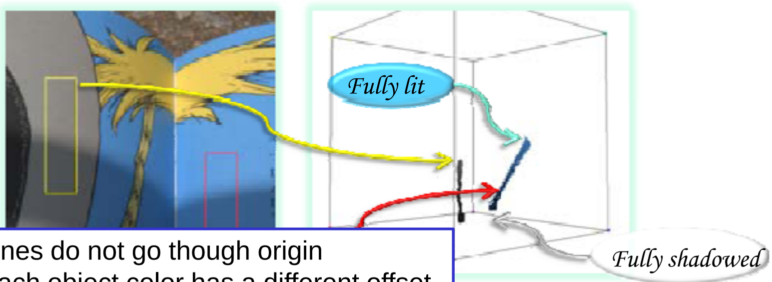
Y_b ,Y_s=percent of direct illuminant visible
F= sensor response of camera
M_{ab}=magnitude of ambient body reflection over hemisphere
M_{as}=magnitude of ambient surface reflection over hemisphere

BIDR model : matter reflection

Consider body reflection components (matte surface):

$$I = \gamma_b F c_b l_d + F c_b M_{ab}$$

- First term defines a line as the direct illuminant changes, second term defines an offset from the origin
- If the ambient and direct illumination are not the same



1. Lines do not go though origin
2. Each object color has a different offset.

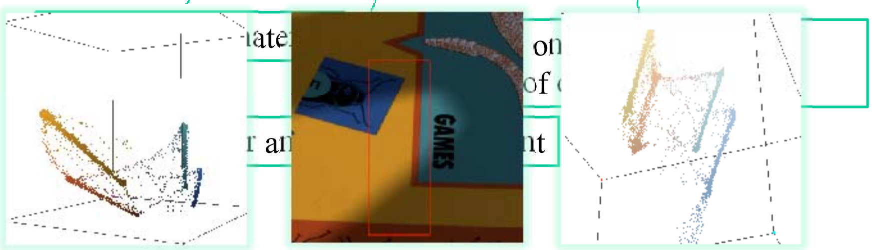
Maxwell et all. [CVPR 2006]

Log chromaticity space

$$\mathbf{f} = \gamma_b \mathbf{F}\,\mathbf{c}_\mathrm{b}\mathbf{l}_\mathrm{d} + \mathrm{F}\,\mathbf{c}_\mathrm{b}\mathbf{M}_{ab}$$

$$\log \mathbf{f} = \log\left(\mathbf{F}\,\mathbf{c}_\mathrm{b}\mathbf{M}_{ab}\left(\gamma_b \mathbf{l}_\mathrm{d} / \mathbf{M}_{ab} + 1\right)\right)$$

$$\log \mathbf{f} = \log\left(\mathbf{F}\,\mathbf{c}_\mathrm{b}\right) + \log\left(\mathbf{M}_{ab}\right) + \log\left(\gamma_b / S + 1\right)$$

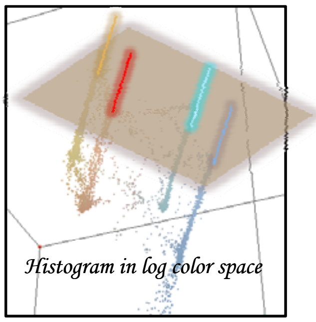


Histogram in RGB space — *Original Image* — *Histogram in log color space*

Maxwell et all. [CVPR 2006]

Illuminantion invariant color space:

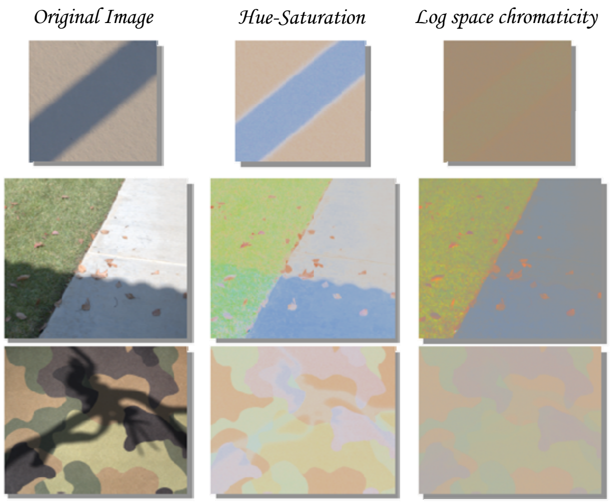
- The plane perpendicular to the curves is illumination invariant color space for realistic illumination.
- Log space provides real 2D chromaticity coordinates on a plane for RGB images
- No a priori assumptions about the color of the direct or ambient illuminants.



Histogram in log color space

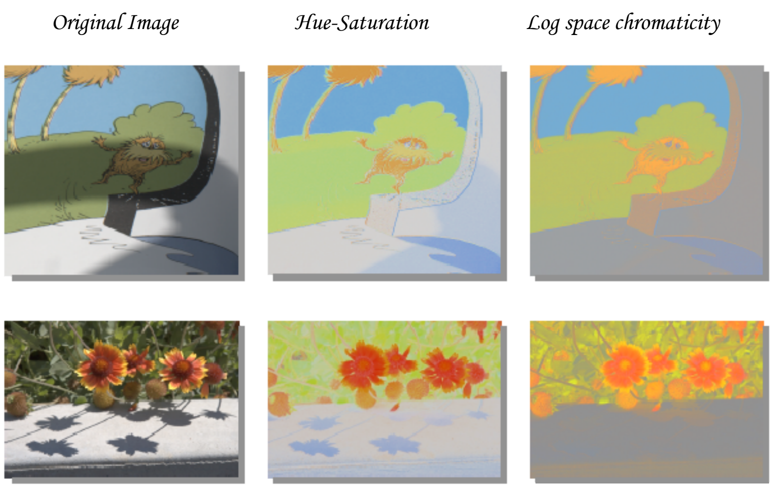
Maxwell et all. [CVPR 2006]

Results



Maxwell et all. [CVPR 2006]

Results



Maxwell et all. [CVPR 2006]

Conclusions

- The BIDR model describes the appearance of materials interacting with a direct and an ambient illumination via body and surface reflection.
- The model is the basis for a new illumination invariant 2-D chromaticity space for a direct and ambient illuminant pair with differing spectra.

Maxwell et all. [CVPR 2006]

References

- Bruce A. Maxwell, Richard M. Friedhoff, Casey A. Smith: A bi-illuminant dichromatic reflection model for understanding images. CVPR 2008
- G. D. Finlayson, M. S. Drew, and L. Cheng. Intrinsic images by entropy minimization. In T. Pajdla and J. Matas, editors, Proc. of European Conf. on Computer Vision, LNCS 3023, pages 582–595, 2004.
- S. A. Shafer. Using color to separate reflection components. Color Research Applications, 10:210–218, 1985.
- Christian Riess, Johannes Jordan and Elli Angelopoulou, A Common Framework for Ambient Illumination in the Dichromatic Reflectance Model, CRICV 2009

Application III: Object recoloring

Extensions of dichromatic reflection model:
- multiple illuminants
- interreflections

NOTE: Slides on recoloring have been removed because the work has not been published yet.

Open Research Topics in Color Research for Computer Vision

Color Feature Description

• Learning class specific color descriptors. One size fits all strategy might fail.

variable color category

constant color category

• Improved methods to combine multiple cues: color, texture, shape, and appearance.

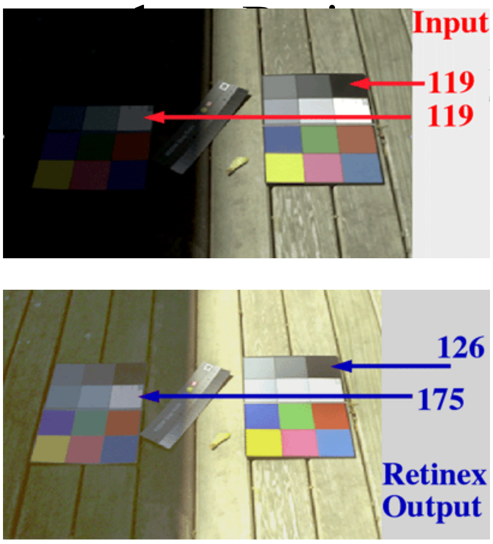
Color Constancy

• How to apply color constancy algorithms in images with multiple illuminants ? No benchmark data available.

• Perceptual Error Measures for Color Constancy.

A. Gijsenij et al. “A perceptual Analysis of Distance Measures for Color Constancy Algorithms”, JOSA 2009

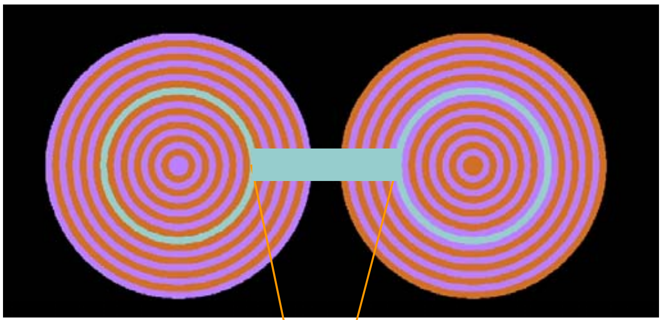
Brightness estimation



Images courtesy John McCann

Human Color Perception

How to compute perceived colors ?



perceived colors:

slide credit: Xavi Otazu

Acknowledgements

- Theo Gevers
- Cordelia Schmid
- Maria Vanrell
- Ramon Baldrich
- Arjan Gijsenij
- Eduard Vazquez
- Fahad Shahbaz
- Shida Beigpour
- David Rojas

Questions ?