Color in Image & Video Processing Applications

Theo Gevers
Joost van de Weijer

Image/Video Applications

Image Segmentation:

Video retrieval:

human segmentation

video sequences
Why use Color?

- photometric invariance
- discriminative power
- saliency detection

PART I (low-level)
Joost van de Weijer

1. Reflection Models
   - Dichromatic reflection model
   - Color Spaces
2. Color Differential Structure
   - Color Edges
   - Photometric Invariant Edge Detection
3. Saliency and Color Boosting
   - Itti and Koch model
   - Color boosted
4. Color Constancy
   - At the pixel
   - Low-level
   - High-level

PART II (high-level)
Theo Gevers

1. Interest point detection
   - Harris Laplace
   - Color boosted
2. Descriptors
   - SIFT
   - Extension to color
3. Object recognition (VOC/TRECVID)
   - Dense and point sampling
   - Code book generation
   - Results
4. Applications
   - Tracking in video
   - Object replacement
   - Emotion recognition
   - Head pose estimation
Reflection Models

Electromagnetic radiation spectrum

\[ \log_{10}(\lambda(m)) \]

- Radio
- TV
- Radar
- Microwaves
- Infrared
- Visible light
- Ultraviolet
- X rays
- Gamma rays

\[ \lambda(m) \quad \lambda(nm) \]

- \(10^{-6}\) to 1000
- \(10^{-7}\) to 100
visible light spectrum

![Visible Light Spectrum](http://askabiologist.asu.edu/research/seecolor/atable.html)

nanometers $\iff 1nm = 10^{-9} m$

---

**Surface reflectance**

$e(\lambda) \rightarrow \{R, G, B\} \rightarrow s(\lambda)$

$e(\lambda)$

$\{R, G, B\}$

$s(\lambda)$
**Reflecting materials**

*Body Reflectance*
Diffuse reflection, isotropic reflection. The spectral distribution depends on colorants.

*Surface Reflectance*
Specular reflection. The reflection angle is similar to the incident angle. Its spectral distribution depends on the illuminant.

\[ f_b(\lambda, \Theta) = m_b(\Theta)c_b(\lambda) \]
\[ f_s(\lambda, \Theta) = m_s(\Theta)c_s(\lambda) \approx m_s(\Theta)h \]

**Dichromatic reflection model:**

\[ f(\lambda, \Theta) = f_b(\lambda, \Theta) + f_s(\lambda, \Theta) \]

- \( f_b(\lambda, \Theta) \): Reflected light by the object body. It depends on the pigments used to colour the object and it’s the one that makes the object look coloured. (Diffuse reflectance)

- \( f_s(\lambda, \Theta) \): Reflected light from the surface. It has a SPD nearly the same as the incident light. (Specular or regular reflectance)

\( \Theta \): Angles that depend on light source position, observer and surface

The spectral and geometrical terms can be separated:

\[ f(\lambda, \Theta) = m_b(\Theta)c_b(\lambda) + m_s(\Theta)c_s(\lambda) \]
\[ f = m_b c_b + m_s c_s \]
Dichromatic Reflection Model

dichromatic model for matte surfaces:
\[ \mathbf{f} = m_b \mathbf{c}_b \]

Dichromatic Reflection Model

dichromatic model for specular surfaces:
\[ \mathbf{f} = m_b \mathbf{c}_b + m_s \mathbf{c}_s \]
Dichromatic Reflection Model

\[ f(\lambda, \Theta) = m_b(\Theta)c_b(\lambda) + m_s(\Theta)c_s(\lambda) \]

- we want to describe the object independent of scene accidental events:
  - shadow - a change of \( m_b(\Theta) \)
  - shading - a change of \( m_s(\Theta) \)
  - viewpoint/orientation object - a change of \( m_b(\Theta) \) and \( m_s(\Theta) \)
  - specularities - a change of \( m_s(\Theta)c_s(\lambda) \)
- the description should only be dependent on \( c_b(\lambda) \)

---

**color spaces: normalized RGB**

- normalized RGB is given by:
  \[ \{r, g, b\} = \left\{ \frac{R}{R + G + B}, \frac{G}{R + G + B}, \frac{B}{R + G + B} \right\} \]

- invariant for shadow and shading variations (matte surfaces):
  \[ r = \frac{R}{R + G + B} = \frac{m_b c_R^b}{m_b c_R^b + m_b c_G^b + m_b c_B^b} = \frac{c_r^b}{c_r^b + c_g^b + c_b^b} \]

---

slide credit: R. Baldrich
**color spaces: hue-saturation-intensity**

- defined as:
  
  $$\text{hue} = \arctan \left( \sqrt{3} \frac{(R - G)}{(R + G - 2B)} \right)$$

  $$\text{sat} = \sqrt{\frac{1}{3} \left( R^2 + G^2 + B^2 - RG - RB - GB \right)}$$

  $$i = \frac{R + G + B}{\sqrt{3}}$$

- hue is invariant for shading variations and specularities under white light:

  $$\text{hue} = \arctan \left( \frac{\sqrt{3} \left( c_R^b + c_s^s - c_G^b \right)}{m^b \left( c_R^b + c_s^s + c_G^b + \frac{2c_R^b - c_G^b}{2} \right)} \right)$$

*Take care of instabilities*

- when working in different color spaces always take instabilities into account!

- Error propagation is a convenient tool for instability evaluation:

Suppose that $u, ..., w$ are measured with corresponding uncertainties $\sigma_u, ..., \sigma_w$ to compute function $q(u, ..., w)$.

The predicted uncertainty is defined by:

$$\sigma_q = \sqrt{\left( \frac{\partial q}{\partial u} \sigma_u \right)^2 + \cdots + \left( \frac{\partial q}{\partial w} \sigma_w \right)^2}$$
Take care of instabilities

- when working in different color spaces always take instabilities into account!

- Error propagation is a convenient tool for instability evaluation:

\[
\text{Ex. 1} \quad \text{hue} = \arctan \left( \frac{\sqrt{3} (R-G)}{R+G-2B} \right) \quad \rightarrow \quad (\partial \text{hue})^2 = \left( \frac{1}{\text{sat}^2} \partial^2 R \right) \quad \text{(assuming } \partial^2 R = \partial^2 G = \partial^2 B) \]

\[
(\partial \text{hue})^2 = \left( \frac{\partial \text{hue}}{\partial R} \right)^2 \partial^2 R + \left( \frac{\partial \text{hue}}{\partial G} \right)^2 \partial^2 G + \left( \frac{\partial \text{hue}}{\partial B} \right)^2 \partial^2 B
\]
references: photometric invariants


overview

**PART I (low-level)**
Joost van de Weijer

1. **Reflection Models**
   - Dichromatic reflection model
   - Color Spaces
2. **Color Differential Structure**
   - Color Edges
   - Photometric Invariant Edge Detection
3. **Saliency and Color Boosting**
   - Itti and Koch model
   - Color boosted
4. **Color Constancy**
   - At the pixel
   - Low-level
   - High-level

**PART II (high-level)**
Theo Gevers

1. **Interest point detection**
   - Harris Laplace
   - Color boosted
2. **Descriptors**
   - SIFT
   - Extension to color
3. **Object recognition (VOC/TRECVID)**
   - Dense and point sampling
   - Code book generation
   - Results
4. **Applications**
   - Tracking in video
   - Object replacement
   - Emotion recognition
   - Head pose estimation
Color Differential Structure

1. How do we combine the differential structure of the various color channels?
2. How do we incorporate color invariance theory into the measurements of the differential structure while maintaining robustness?
isoluminance

luminance gradient: isoluminant edges are not detected.

Color Feature Detection
from luminance to color

vector: \[ R_x + G_x = 0 \]

![2-channel test-image]

\[
\text{tensor:} \quad \begin{pmatrix} R_x^2 & R_x R_y \\ R_x R_y & R_y^2 \end{pmatrix} + \begin{pmatrix} G_x^2 & G_x G_y \\ G_x G_y & G_y^2 \end{pmatrix} = \begin{pmatrix} R_x^2 + G_x^2 & R_x R_y + G_x G_y \\ R_x R_y + G_x G_y & R_y^2 + G_y^2 \end{pmatrix}
\]


feature detection in oriented patterns

more tensor-based features:
- Harris corner points
- symmetry points
- (star and circle structures)
- optical flow
- orientation estimation
- curvature estimation
- ...

oriented texture

traditional orientation estimation:
\[ \theta = \arctan \left( \frac{f_y}{f_x} \right) \rightarrow \bar{\theta} = \arctan \left( \frac{f_y}{f_x} \right) \]

tensor-based orientation estimation:
\[ \theta = \arctan \left( \frac{2 f_x f_y}{f_x^2 - f_y^2} \right) \rightarrow \bar{\theta} = \arctan \left( \frac{2 f_x f_y}{f_x^2 - f_y^2} \right) \]
1. How do we combine the differential structure of the various color channels?
2. How do we incorporate color invariance theory into the measurements of the differential structure while maintaining robustness?

Photometric Invariant Edge Detection

- we differ between three types of edges
  1. material edge
  2. shadow/shading edge
  3. specular edge
- assumptions:
  1. white illumination
  2. neutral interface reflection
  3. shadows are not colored.
Computation of quasi-invariance

nonlinear transformation

IMAGE → FULL INVARIANT → FULL INVARIANT DERIVATIVE

linear operation

DERIVATIVES → QUASI- INVARIANT DERIVATIVE

Shadow-Shading-Specular Quasi-Invariant

spherical coordinates  opponent colors  hue-saturation-intensity

shading variant  specular variant  shading-specular variant

shading invariant  specular invariant  shading-specular invariant
**spherical coordinates**

- For matte surfaces: \( f = m^b c^b \)
- all shadow-shading variation is in the radial direction

\[
f_x = \begin{pmatrix} R_x \\ G_x \\ B_x \end{pmatrix} \xrightarrow{spherical} \begin{pmatrix} r_x \\ r \phi_x \\ \sin \varphi \theta_x \end{pmatrix} = \begin{pmatrix} r_x \\ 0 \\ \sin \varphi \theta_x \end{pmatrix} \rightarrow c_x = \begin{pmatrix} 0 \\ \varphi_x \\ \sin \varphi \theta_x \end{pmatrix}
\]

**hue-saturation-intensity**

- For specular surfaces: \( f = m^b c^b + m^s c^s \)
- there is no specular-shadow-shading variation in the hue-direction.

\[
f_x = \begin{pmatrix} R_x \\ G_x \\ B_x \end{pmatrix} \xrightarrow{hsi} \begin{pmatrix} s h_x \\ s_x \\ i_x \end{pmatrix} = \begin{pmatrix} 0 \\ s_x \\ i_x \end{pmatrix} \rightarrow h_x = \begin{pmatrix} h_x \\ 0 \end{pmatrix}
\]
**Invariant Edge Detection Applications**

- Colo Feature Extraction
- Multi Image Applications
  - image retrieval

**Color Feature Detection**

- Single Image Applications
  - snakes
  - feature extraction

**Instabilities**

- Shadow-shading invariance:
  \[
  \lim_{{\{R,G,B\} \to 0}}
  \]

- Specular-shadow-shading invariance:
  \[
  \lim_{{\{R,G,B\} \to \alpha \{1,1,1\}}}
  \]
Edge Detection

- experiments conducted on pantone color set (1012) which is used to compose 500,000 edges.

- edge detection is based on the maximum response path of the derivative energy.

- edges are tested on
  - edge displacement.
  - percentage of missed edges.

<table>
<thead>
<tr>
<th>shadow-shading:</th>
<th>( \Delta )</th>
<th>( \varepsilon )</th>
</tr>
</thead>
<tbody>
<tr>
<td>full</td>
<td>0.21</td>
<td>2.0</td>
</tr>
<tr>
<td>quasi</td>
<td>0.043</td>
<td>0.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>specular-shadow-shading:</th>
<th>( \Delta )</th>
<th>( \varepsilon )</th>
</tr>
</thead>
<tbody>
<tr>
<td>full</td>
<td>0.85</td>
<td>9.8</td>
</tr>
<tr>
<td>quasi</td>
<td>0.35</td>
<td>5.8</td>
</tr>
</tbody>
</table>

• Conclusion: Quasi invariants more than half the edge displacement, and have higher discriminative power.
experiments: canny edge detection

luminance-gradient  RGB-gradient

shadow-shading  shadow-shading-specular
quasi-invariant  quasi-invariant
Edge Classification

shadow edges

specular edges

Edge Classification
Edge Classification

- red - object edge
- green - shading/shadow edge
- Blue - specular edge

Photometric Invariant Corner Detection

- Harris corner detector combined with the quasi-invariants allows for photometric invariant corner detection

RGB  shadow-shading  specular-shadow-shading
experiments: Hough transform

RGB-gradient

shadow-shading-specular quasi-invariant

references: color differential structure

PART I (low-level)
Joost van de Weijer

1. Reflection Models
   • Dichromatic reflection model
   • Color Spaces
2. Color Differential Structure
   • Color Edges
   • Photometric Invariant Edge Detection
3. Saliency and Color Boosting
   • Itti and Koch model
   • Color boosted
4. Color Constancy
   • At the pixel
   • Low-level
   • High-level

PART II (high-level)
Theo Gevers

1. Interest point detection
   • Harris Laplace
   • Color boosted
2. Descriptors
   • SIFT
   • Extension to color
3. Object recognition
   (VOC/TRECVID)
   • Dense and point sampling
   • Code book generation
   • Results
4. Applications
   • Tracking in video
   • Object replacement
   • Emotion recognition
   • Head pose estimation

Color Salient Features
Saliency Detection

- Goal: direct our gaze rapidly towards objects of interest in our environment.
- Visual attention is known to be driven by both bottom up (image based) and top-down (task based) cues.
- Bottom-up saliency uses simple visual attributes such as intensity, contrast, color opponency, orientation, direction and velocity of motion.
- What matters is feature contrast rather than absolute feature strength (as in center surround systems).

overview approach

Computational Modeling of Visual Attention


black-white focus of detectors

luminance-based points  color-based points
color distinctiveness

- the information content of an event, $v$, is equal to:

$$I(v) = -\log(p(v)) = -\log(p(f)p(f_x)p(f_y))$$

$$v = (R \ G \ B \ R_x \ G_x \ B_x \ R_y \ G_y \ B_y)$$

- equation differential-based salient point detectors:

$$H(f_x, f_y)$$

\[Color\ Boosting\ Saliency: \quad p(f_x) = p(f'_x) \iff |g(f_x)| = |g(f'_x)|\]

statistics of color images:

- The statistics of $f_x$ is computed by looking of the 40.000 images of the Corel database.

- Isosalient surfaces can be approximated by aligned ellipsoids in decorrelated color spaces.
statistics of color images:

Color Boosting Saliency: \( p(f_x) = p(f'_x) \leftrightarrow g(f_x) = |g(f'_x)| \)

Color Boosting function: \( g(f_x) = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} h(f_x) \)

saliency points

input car-image

RGB-based (first 20 points)  saliency boosting (first 4 points)
generality approach: global optimal regions

RGB gradient

color boosting

experiment: quantitative analysis

Quantitative evaluation of color boosting on a retrieval experiment.

- Nister database: around 10,000 images
- detector: DoG (color boosted)
- descriptor: SIFT+hue
The do's and dont's of Color Features

1. Take care in combining different channels:
   Tensor-based features solve the opposing vector problem.

2. Look at what kind of photometric invariance your problem needs:
   
   * Do not take derivatives of circular color spaces.
   * Compute first derivatives, then color space transform.

   Quasi-invariants are more stable for feature detection.

3. When working with invariance take instabilities into account.
   Use error analysis to find certainty measures for your invariants.

4. When considering photometric invariance always also take discriminative power into account.

5. From information theory an optimal color space for salient feature detection can be derived.

6. Color information is highly corrupted in compressed data. In compression (jpeg, mpeg) chrominance is subsampled.
PART I (low-level)
Joost van de Weijer

1. Reflection Models
   • Dichromatic reflection model
   • Color Spaces
2. Color Differential Structure
   • Color Edges
   • Photometric Invariant Edge Detection
3. Saliency and Color Boosting
   • Itti and Koch model
   • Color boosted
4. Color Constancy
   • At the pixel
   • Low-level
   • High-level

PART II (high-level)
Theo Gevers

1. Interest point detection
   • Harris Laplace
   • Color boosted
2. Descriptors
   • SIFT
   • Extension to color
3. Object recognition (VOC/TRECVID)
   • Dense and point sampling
   • Code book generation
   • Results
4. Applications
   • Tracking in video
   • Object replacement
   • Emotion recognition
   • Head pose estimation

Color Constancy at a Pixel
problem statement

How do we recognize colors to be the same under varying light sources?

color constancy: the ability to recognize colors of objects invariant of the color of the light source.

Colour constancy algorithms

Invariant Normalizations
color constancy at a pixel

Assumptions:

1. Lambertian model:
   - linear relation pixel values and intensity light.
   - no specularities and interreflections.
2. perfectly narrow-band sensors (Dirac delta functions).
3. the illuminants are Planckian.

However, the final algorithm is shown to be robust to deviations from the assumptions.
### Dirac delta functions

\[ p_k = \int_\omega e(\lambda) c_b(\lambda) s_k(\lambda) d\lambda \]

**assumption: Dirac sensors**

\[ p_k = \int_\omega e(\lambda) c_b(\lambda) q_k \delta(\lambda - \lambda_k) d\lambda \]

\[ p_k = e(\lambda_k) c_b(\lambda_k) q_k \]

### Planckian illuminants

**Planck’s law** of black body radiation states the spectral intensity of electromagnetic radiation from a black body at temperature \( T \) as a function of wavelength:

\[
E(\lambda, T) = \frac{c_1}{\lambda^5} e^{\frac{c_2}{T\lambda}}
\]

**Wien’s approx:**

\[
E(\lambda, T) = \frac{c_1}{\lambda^5} e^{\frac{c_2}{T\lambda}}
\]

The **Planckian locus** is the path that the color of a black body as the blackbody temperature changes.
**Planckian illuminants**

**Planck's** law of black body radiation states the spectral intensity of electromagnetic radiation from a black body at temperature $T$ as a function of wavelength:

$$ E(\lambda, T) = \frac{c_1}{\lambda^5} e^{-\frac{c_2}{T\lambda}} $$

**Wien's approx:**

The **Planckian locus** is the path that the color of a black body as the blackbody temperature changes.

**Daylight illuminants** can be approximated by Planckian illuminants.

( indoor illuminants to some extend

- 2500K Household light bulbs
- 3000K Studio lights, photo floods
- 4000K Clear flashbulbs
- 5000K Typical daylight; electronic flash )

**Color constancy at a pixel**

Consider the logarithm of the chromaticity coordinates:

$$ \chi_j = \log \left( \frac{p_k}{p_p} \right) = \log \left( \frac{c_1}{\lambda_k^5} e^{-\frac{c_2}{T\lambda_k}} c_b(\lambda_k) q_k \right) $$

$$ \chi = s + \frac{1}{T} e $$

$$ \chi_j = \log \left( \frac{s_k}{s_p} \right) + \frac{1}{T} \left( e_k - e_p \right) $$

$e_k \equiv -\frac{c_2}{\lambda_k}$

depends on surface color

depends on illuminant color $s_k = \lambda_k^{-5} c_b(\lambda_k) q_k$
color constancy at a pixel - examples

examples log chromaticity plots:

Macbeth Color Checker  HP912 Digital Still Camera  Nikon D-100

illuminant invariant direction axis

illuminant invariant representation

Every pixel can be represented in a illuminant invariant representation!

illuminant variant axis (camera dependent)

images source: Eli Arbel

examples illuminant invariant

Since shadows are a change in illuminant these representation are shadow free.
shadow detection

Comparison of the edge maps of the original and the shadow invariant image allows for shadow detection.

examples:

- sky and sun light
- sky light
- removal of colored shadow
- shading is not effected
references:


Gamut Mapping
“In real-world images, for a given illuminant, one observes only a limited number of different colors.”

Solux 4700K  Solux 4700K + Roscolux filter  Sylvania Warm White Fluorescent

Gamut mapping algorithm:
• Obtain input image.
Gamut mapping algorithm:
• Obtain input image.
• Compute gamut from image.
Gamut mapping algorithm:

- Obtain input image.
- Compute gamut from image.
- Determine feasible set of mappings from input gamut to canonical gamut.
- Apply some estimator, to select one mapping from this set.

Use mapping on input image to recover the corrected image, or on canonical illuminant to estimate the color of the unknown illuminant.
Color Constancy from Color Derivatives

Color Constancy

color constancy: the ability to recognize colors of objects invariant of the color of the light source.

Grey world hypothesis: the average reflectance in a scene is grey.

White patch hypothesis: the highest value in the image is white.

Grey-world: \( \sum_{m=1}^{M} f_i(x) \propto c \)

white-patch: \( \left( \sum_{m=1}^{M} (f_i(x))^\alpha \right)^{\frac{1}{\alpha}} \propto c \)

Shades of Grey hypothesis: the n-Minkowsky norm based average of a scene is achromatic.

- unifies Grey-World and White Patch: \( e^p \approx \sqrt[p]{\int |f(x)|^p \, dx} \)
**Color Constancy**

color constancy: the ability to recognize colors of objects invariant of the color of the light source.

Grey world hypothesis: the average reflectance in a scene is grey.

White patch hypothesis: the highest value in the image is white.

Grey edge hypothesis: the average edge in a scene is grey.

**Generalization I:** the L-norm:

\[
\left( \sum_{m=1}^{M} \left( f_i(x) \right)^k \right)^{\frac{1}{k}} \propto c
\]

**Generalization II:** L-norm + differentiation order:

\[
\left( \sum_{i=1}^{M} \left| \frac{\partial^n f_i(x)}{\partial x^n} \right|^p \right)^{\frac{1}{p}} \propto c
\]

**Color Constancy in 4 lines of matlab code!**

```matlab
function Illuminant=GreyEdgeCC(im,mink,sigma,dif)
    im = gauss_derivative(im,sigma,dif);
    im = reshape(im,size(im,1)*size(im,2),3);
    Illuminant= 1./power( sum( power( im, mink ) ), 1/mink ) ;
    Illuminant = Illuminant./norm(Illuminant);
```
general color constancy framework

Low-level color constancy:

\[
\left( \sum_{i=1}^{M} \left( \frac{\partial^n f_i(x)}{\partial x^n} \right)^p \right)^{\frac{1}{p}} \propto c
\]

- \( n = 0, p = 1 \) (grey-world)
- \( n = 0, p = \infty \) (white-patch)
- \( n = 0, p = k \) (shades-of-gray)
- \( n = 1, p = 1 \) (grey-edge)

G. Finlayson, E. Trezzi, “Shades of gray and colour constancy”, CIC 2004

Color Constancy: experiment

- test set: 23 objects under 11 illuminants (Computational Vision Lab: Simon Fraser)
- angular error = \( \cos(\hat{e} \cdot e) \)
Color Constancy: experiment

- real-world data set (F. Ciurea and B. Funt: Vision Lab - Simon Fraser)
Color Constancy: experiment

• real-world data set (F. Ciurea and B. Funt: Vision Lab - Simon Fraser)

<table>
<thead>
<tr>
<th></th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grey-World</td>
<td>7.3</td>
</tr>
<tr>
<td>White-Patch</td>
<td>6.7</td>
</tr>
<tr>
<td>General Grey-World</td>
<td>4.7</td>
</tr>
<tr>
<td>Grey-Edge</td>
<td>4.1</td>
</tr>
<tr>
<td>2nd order Grey-Edge</td>
<td>4.3</td>
</tr>
</tbody>
</table>

“In real-world images, for a given illuminant, one observes only a limited number of different colored edges.”

Solux 4700K  Solux 4700K + Roscolux filter  Sylvania Warm White Fluorescent

A. Gijsenij, T. Gevers, J. van de Weijer, “Edge-Based Color Constancy”, IJCV 2010
Experiments (real-world images)

Some examples:

<table>
<thead>
<tr>
<th>Original</th>
<th>Ideal</th>
<th>Derivative-based</th>
<th>Regular Gamut</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Original Image" /></td>
<td><img src="image2" alt="Ideal Image" /></td>
<td><img src="image3" alt="Derivative-based Image" /></td>
<td><img src="image4" alt="Regular Gamut Image" /></td>
</tr>
<tr>
<td><img src="image5" alt="Original Image" /></td>
<td><img src="image6" alt="Ideal Image" /></td>
<td><img src="image7" alt="Derivative-based Image" /></td>
<td><img src="image8" alt="Regular Gamut Image" /></td>
</tr>
</tbody>
</table>

How do you choose the best cc-algorithm?

High-Level Color Constancy
Could it be that different scenes prefer different color constancy methods?

Geusebroek and Smeulders (2005) – Weibulls

Examples:

Distribution of edge responses follows Weibull distribution.

Two parameters:

- $\beta$ – Contrast of the image. A higher value indicates more contrast.
- $\gamma$ – Grain size. A higher value indicates more fine textures.
Color Constancy – Selection

Postsupervised Prototype Classification:
Compute Weibull-parameters for all images

Partition weibull-parameters using \( k \)-means

slide credit: Arjan Gijsenij
Color Constancy – Selection

Postsupervised Prototype Classification:
Compute Weibull-parameters for all images
Partition weibull-parameters using $k$-means
Label cluster centers according to the minimum mean angular error

Build 1-NN Classifier on these cluster centers

slide credit: Arjan Gijsenij
Experiments

Data set consisting of 11000+ images

The true illuminants are known (ground truth)

Grey sphere is masked during experiments

Performance measure → angular error:

\[ \cos^{-1}(\hat{e}_l \cdot \hat{e}_e) \]

slide credit: Arjan Gijsenij

Experiments – Results

<table>
<thead>
<tr>
<th>Original</th>
<th>Ideal</th>
<th>Selection</th>
<th>White-Patch</th>
<th>Grey-World</th>
</tr>
</thead>
</table>

slide credit: Arjan Gijsenij
## Experiments – Performance

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grey-World</td>
<td>7.9°</td>
<td>7.0°</td>
</tr>
<tr>
<td>White-Patch</td>
<td>6.8°</td>
<td>5.3°</td>
</tr>
<tr>
<td>General Grey-World</td>
<td>6.2°</td>
<td>5.3°</td>
</tr>
<tr>
<td>1\textsuperscript{st}-Order Grey-Edge</td>
<td>6.2°</td>
<td>5.2°</td>
</tr>
<tr>
<td>2\textsuperscript{nd}-Order Grey-Edge</td>
<td>6.1°</td>
<td>5.2°</td>
</tr>
<tr>
<td>Gamut mapping</td>
<td>8.5°</td>
<td>6.8°</td>
</tr>
<tr>
<td>Color-by-Correlation</td>
<td>6.4°</td>
<td>5.2°</td>
</tr>
</tbody>
</table>

slide credit: Arjan Gijsenij

## Experiments – Performance

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>2\textsuperscript{nd}-Order Grey-Edge (baseline)</td>
<td>6.1°</td>
<td>5.2°</td>
</tr>
<tr>
<td>Selection – 5 methods</td>
<td>5.7° (-7%)</td>
<td>4.7° (-10%)</td>
</tr>
<tr>
<td>Combining – 5 methods</td>
<td>5.6° (-8%)</td>
<td>4.6° (-12%)</td>
</tr>
<tr>
<td>Combining – 75 methods</td>
<td>5.0° (-18%)</td>
<td>3.7° (-29%)</td>
</tr>
</tbody>
</table>

slide credit: Arjan Gijsenij
Color Constancy from High-Level Visual Information

**Problem Statement**

How do we recognize colors to be the same under varying light sources?

Color constancy: the ability to recognize colors of objects invariant of the color of the light source.
computational color constancy

- Grey-World
  Buchsbaum, 1980
- Gamut Mapping
  Forsyth, 1990
- White-Patch
  Land, 1976
- Color-by-Correlation
  Finlayson, 2001

bottom-up approaches!

---

top-down color constancy

psychophysical motivation:

**problem statement**

How do we recognize colors to be the same under varying light sources?

**color constancy**: the ability to recognize colors of objects invariant of the color of the light source.

How can we apply high-level visual information for computational color constancy?

**overview our approach**

- **input image**
- **cast bottom-up hypotheses**
- **compute semantic likelihood for all images, and select most likely.**
- **output image**
plsa-based image segmentation

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

**Image representation**
- dense extraction of 20x20 pixel patches on 10x10 pixel grid
- each patch described by discretized features, the words.
  - texture: SIFT (750 visual words, k-means)
  - color: hue (100 visual words, k-means)
  - position: patch location indicated by cell in an 8x8 grid

An image is modeled as a mixture of semantic topics:

\[
p(w | d) = \sum_z p(w | z) p(z | d)
\]

The \( p(w^m | z) \) can either be learned supervised or unsupervised. We assume them to be learned from images taken under a white illuminant.

**likelihood image**

\[
p(d) = \prod_w p(w | d)
\]
### Plsa-based Image Segmentation

#### Supervised Learning

- **P(w|cow)**
- **P(w|grass)**

**P(w|z)**

\[ p(w|d) = \sum_z p(w|z) p(z|d) \]

**Using EM:** \( p(z|d) = \{0.6, 0.4\} \)

**Test Image**

**Semantic Image Segmentation**

---

#### Unsupervised Learning

\[ p(w|c_1) \]

\[ p(w|c_2) \]

\[ p(w|z) \]

\[ p(w|d) = \sum_z p(w|z) p(z|d) \]

**Using EM:** \( p(z|d) = \{0.6, 0.4\} \)

**Test Image**

**Semantic Image Segmentation**
casting hypotheses: bottom-up

Low-level color constancy:

\[ \left( \sum_{i=1}^{M} \left| \frac{\partial^n f_i(x)}{\partial x^n} \right|^P \right)^{\frac{1}{P}} \propto c \]

- \( n = 0, p = 1 \)
- \( n = 0, p = 2 \)
- \( n = 0, p = k \)
- \( n = 1, p = 1 \)

We will use \( n=\{0,1,2\} \) and \( p=\{2,12\} \) to cast a total of 6 bottom-up hypotheses.

G. Finlayson, E. Trezzi, “Shades of gray and colour constancy”, CIC 2004
J. van de Weijer, T. Gevers “Edge-Based Color Constancy”, IEEE TIP 2007
casting hypotheses: top-down

apply PLSA based on texture and position to assign pixels to classes

cast one illuminant hypothesis for each detected class

green grass hypothesis: the average reflectance of a semantic class in an image is equal to the average of the semantic class in the train-set

bottom-up hypotheses

compute semantic likelihood for all hypotheses, and select most likely

Data Set contains both indoor and outdoor scenes from a wide variety of locations (150 training, 150 testing)

Topic-word distributions are learned unsupervised on the texture and position cue (color is ignored in training).

**experiment: illuminant estimation**

*results* in angular error:

<table>
<thead>
<tr>
<th></th>
<th>standard color constancy</th>
<th>high-level selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no cc</td>
<td>worst BU</td>
</tr>
<tr>
<td>indoor</td>
<td>12.8</td>
<td>12.3</td>
</tr>
<tr>
<td>outdoor</td>
<td>5.5</td>
<td>7.4</td>
</tr>
</tbody>
</table>

**experiment: semantic segmentation**

**Data Set** training: labelled images of Microsoft Research Cambridge (MSRC) set, together with ten images collected from Google Image for each class. Traning: 350 images. Test : 36 images.

**Topic-word distributions** are learned supervised.

**Classes:** building, grass, tree, cow, sheep, sky, water, face and road.

experiment: pixel classification

results pixel classification in %:

<table>
<thead>
<tr>
<th></th>
<th>standard color constancy</th>
<th>high-level selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>no cc</td>
<td>worst BU</td>
<td>best BU</td>
</tr>
<tr>
<td>39.6</td>
<td>41.4</td>
<td>52.2</td>
</tr>
</tbody>
</table>

Summary Color Constancy

• The Planckian locus describes natural light illuminants.

• Color constancy at the pixel allows for shadow removal.

• The general grey-world algorithm generalizes a set of low-level color constancy algorithms, including white patch, grey-world, grey-edge, and shades –of-grey.

\[
\left( \sum_{i=1}^{M} \frac{\partial^n f_i(x)}{\partial x^n} \right)^\frac{1}{p} \propto c
\]

• Top-down information improves both color constancy performance and semantic segmentation results.
references: color constancy


acknowledgements:

Amsterdam ISLA : Theo Gevers, Jan-Mark Geusebroek, Arnold Smeulders, Arjan Gijsenij.
INRIA Rhone-Alpes : Cordelia Schmid, Jakob Verbeek, Marcin Marszalek.
CVC Barcelona : Maria Vanrell, Ramon Baldrich, Juan Toledo.
Color Naming

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.

The English language has 11 basic color terms.

Development color names in languages:

- white
- green
- red
- blue
- brown
- yellow
- purple
- pink
- orange
- grey

Images retrieved with Google image

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

false positives

Images retrieved with Google image
learning color names

Labeled input images:

LAB-histogram representation:

PLSA-bg

Color name distribution:

learning color names

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

Ebay user: “Find me all yellow cars?”

Result automatic labeling pixels:
Example: classification soccer data

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

<table>
<thead>
<tr>
<th>color</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>black</td>
<td>19</td>
</tr>
<tr>
<td>blue</td>
<td>12</td>
</tr>
<tr>
<td>brown</td>
<td>23</td>
</tr>
<tr>
<td>grey</td>
<td>19</td>
</tr>
<tr>
<td>green</td>
<td>10</td>
</tr>
<tr>
<td>orange</td>
<td>2</td>
</tr>
<tr>
<td>pink</td>
<td>2</td>
</tr>
<tr>
<td>purple</td>
<td>2</td>
</tr>
<tr>
<td>red</td>
<td>4</td>
</tr>
<tr>
<td>white</td>
<td>6</td>
</tr>
<tr>
<td>yellow</td>
<td>1</td>
</tr>
</tbody>
</table>

Statistics based on 40,000 Corel images.

- when using photometric invariance always consider discriminative power.

Example: classification soccer data

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

<table>
<thead>
<tr>
<th>color</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>black</td>
<td>19</td>
</tr>
<tr>
<td>blue</td>
<td>12</td>
</tr>
<tr>
<td>brown</td>
<td>23</td>
</tr>
<tr>
<td>grey</td>
<td>19</td>
</tr>
<tr>
<td>green</td>
<td>10</td>
</tr>
<tr>
<td>orange</td>
<td>2</td>
</tr>
<tr>
<td>pink</td>
<td>2</td>
</tr>
<tr>
<td>purple</td>
<td>2</td>
</tr>
<tr>
<td>red</td>
<td>4</td>
</tr>
<tr>
<td>white</td>
<td>6</td>
</tr>
<tr>
<td>yellow</td>
<td>1</td>
</tr>
</tbody>
</table>

Statistics based on 40,000 Corel images.

- when using photometric invariance always consider discriminative power.
Results flower data set:

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

<table>
<thead>
<tr>
<th>dataset</th>
<th>flower</th>
</tr>
</thead>
<tbody>
<tr>
<td>method</td>
<td>color</td>
</tr>
<tr>
<td>HSV-SIFT</td>
<td>78</td>
</tr>
<tr>
<td>hue</td>
<td>40</td>
</tr>
<tr>
<td>opponent</td>
<td>39</td>
</tr>
<tr>
<td>color names</td>
<td>57</td>
</tr>
</tbody>
</table>

references: color naming

Blur Robust and Color Constant Image description

problem statement

How do we recognize colors to be the same under varying light sources?

\[
\begin{pmatrix}
R' \\
G' \\
B'
\end{pmatrix} = \begin{pmatrix}
\alpha & 0 & 0 \\
0 & \beta & 0 \\
0 & 0 & \gamma
\end{pmatrix}\begin{pmatrix}
R \\
G \\
B
\end{pmatrix}
\]

color constancy: the ability to recognize colors of objects invariant of the color of the light source.

Change of illuminant can be modeled by the \textit{diagonal model}.
Color Constant Derivatives

- A color constant representation of a single color patch is impossible.
- The difference between two color patches can be represented invariant to the color illuminant.

<table>
<thead>
<tr>
<th>Funt and Finlayson:</th>
<th>Gevers and Smeulders:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mondrian-world:</strong> $f(x) = m^b c^b (x)e$</td>
<td><strong>3D-world:</strong> $f(x) = m^b (x)e^b (x)e$</td>
</tr>
<tr>
<td>$p = \frac{R^1}{R^2} = \frac{m^b c_1^R e^R}{m^b c_2^R e^R} = \frac{c_1^R}{c_2^R}$</td>
<td>$m = \frac{R^1 G^2}{R^2 G^1} = \frac{m_1^b c_1^R e^R}{m_2^b c_2^R e^R} = \frac{c_1^R c_2^G}{c_2^R c_1^G}$</td>
</tr>
<tr>
<td>$\ln p = \ln \frac{R^1}{R^2} = \ln R^1 - \ln R^2 = \frac{R^1}{c^b_1} \ln R$</td>
<td>$\ln m = \ln \frac{R^1 G^2}{R^2 G^1} = \ln \frac{R^1}{G^1} - \ln \frac{R^2}{G^2} = \frac{R^1}{c^b_1} \ln R$</td>
</tr>
<tr>
<td>$m^b$</td>
<td>$m_1^b$</td>
</tr>
</tbody>
</table>
Why is this a problem?

- Image blur is a frequently encountered phenomenon.
- Possible causes are: out-of-focus, relative motion between camera and object, and aberrations of the optical system.

Obtaining Invariance to Image Blur

- A color constant representation of a single color patch is impossible.
- The difference between two color patches can be represented invariant to the color illuminant.

\[
\text{Funt and Finlayson:} \\
\text{Mondrian-world:} \quad f(x) = m^b e^h (x) \\
P = \frac{R^1}{R^2} = \frac{m^b c_1^R e^R}{m^b c_2^R e^R} = \frac{c_1^R}{c_2^R} \\
\ln p = \ln \frac{R^1}{R^2} = \ln R^1 - \ln R^2 = \frac{\partial}{\partial x} \ln R
\]

Consider a blurred image: \( R' = R \otimes G^\sigma \)

\[
\frac{\partial}{\partial x} \ln R = \frac{R_x^\sigma}{R_y^\sigma} \quad \frac{\partial}{\partial y} \ln R' = \frac{R_x^\sigma}{R_y^{\sigma + 2\sigma}} \\
\text{On the edge the following holds:} \\
R_x^\gamma = R_y^\gamma \quad \gamma' = C(\sigma_j) R_x^\gamma \\
\text{Robustness with respect to blur is obtained by:} \\
\phi_p^1 = \arctan \left( \frac{R, G}{G, R} \right) \quad \phi_p^1 = \arctan \left( \frac{G, B}{B, G} \right)
\]
Retrieval Experiment I

- Twenty different objects were captured under 11 different object orientations and 11 different light sources (Simon Fraser).
- We compare the retrieval results of the color constant description with the color constant and blur robust description.
- Error given in Normalized Average Rank (NAR).

<table>
<thead>
<tr>
<th>rank</th>
<th>1</th>
<th>2</th>
<th>&gt;2</th>
<th>ANAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>180</td>
<td>5</td>
<td>15</td>
<td>0.010</td>
</tr>
<tr>
<td>(\varphi_p)</td>
<td>169</td>
<td>17</td>
<td>14</td>
<td>0.012</td>
</tr>
<tr>
<td>m</td>
<td>155</td>
<td>22</td>
<td>23</td>
<td>0.024</td>
</tr>
<tr>
<td>(\varphi_m)</td>
<td>115</td>
<td>23</td>
<td>65</td>
<td>0.049</td>
</tr>
</tbody>
</table>
Retrieval Experiment I

- Twenty pairs of images with varying image blur.
- We compare the retrieval results of the color constant description with the color constant and blur robust description.

<table>
<thead>
<tr>
<th>rank</th>
<th>1</th>
<th>2</th>
<th>&gt;2</th>
<th>ANAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>7</td>
<td>2</td>
<td>11</td>
<td>0.365</td>
</tr>
<tr>
<td>$\varphi_p$</td>
<td>16</td>
<td>3</td>
<td>1</td>
<td>0.018</td>
</tr>
<tr>
<td>m</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td>0.303</td>
</tr>
<tr>
<td>$\varphi_m$</td>
<td>13</td>
<td>1</td>
<td>6</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Extra slides
Dichromatic Reflection Model in Chromaticity Representation

Chromaticities:
\[ \{r, g, b\} = \left\{ \frac{R}{R + G + B}, \frac{G}{R + G + B}, \frac{B}{R + G + B} \right\} \]

experiments

no boosting
The salient space is highly photometrically invariant.
**color distinctiveness**

- color distinctiveness is measured by the information content:

\[
I(v) = - \sum \log(p(v))
\]

<table>
<thead>
<tr>
<th>20 points</th>
<th>100 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>incr(%)</td>
<td>decr(%)</td>
</tr>
<tr>
<td>incr</td>
<td>decr</td>
</tr>
<tr>
<td>opponent</td>
<td>63</td>
</tr>
<tr>
<td>spherical</td>
<td>49</td>
</tr>
<tr>
<td>HSI</td>
<td>57</td>
</tr>
</tbody>
</table>

**Dichromatic Model**

- dichromatic model:

\[
F = e(m^bC^b + m^sC^s)
\]

body + specular

- intensity illuminant

- first order photometric structure:

\[
F_x = \{R_x, G_x, B_x\} = m^bC^b_x + \left(e_x m^b + em^b_x\right)C^b + em^i_x C^i
\]

material + (shadow+shading) + specular
experiments

- visual inspection

- evaluation criteria:
  1. repeatability: salient point detection should be stable under varying viewing conditions.
  2. distinctiveness: salient point should focus on events with a low probability of occurrence.
  3. optimal: is the transformation optimal?

The do’s and dont’s of Color Features

1. Take care in combining different channels:
   Tensor-based features solve the opposing vector problem.
2. Look at what kind of photometric invariance your problem needs:
   Quasi-invariants are more stable for feature detection
   *Do not take derivatives of circular color spaces.*
   *Compute first derivatives, then color space transform.*
3. When working with invariance take instabilities into account:
   Use error analysis to find certainty measures for your invariants.
4. When considering photometric invariance always also take discriminative power into account.
5. In case of image derivative-based descriptors take be aware of blur.
   Divisions of derivative based descriptors are often robust to blur.
6. From information theory an optimal color space for salient feature detection can be derived.
7. Color information is highly corrupted in compressed data. In compression (jpeg, mpeg) chrominance is subsampled.
Colour constancy algorithms

Invariant Normalizations

experiments: robust optical flow estimation