A Probabilistic Model of the Categorical Association between Colors

Jason Chuang (Stanford University)
Maureen Stone (StoneSoup Consulting)
Pat Hanrahan (Stanford University)

Color Imaging Conference
November 12, 2008
Overview

• **Introduction**
  – Color names
  – Previous work
  – Goals

• Categorical association
  – Probabilistic framework
  – Results

• Summary
Why Color Names?

• Affect our perception

• Improve communication
Naming Studies

- Earlier studies
  - For 19 languages, only 1 speaker (Berlin & Kay)
  - 10 speakers for English (Benavente)
  - 25-30 speakers for 110 languages (World Color Survey)
More Naming Data

- Online surveys
  - 238 speakers in English (Dolores Labs)
  - 4000 speakers in 20 languages (HP Naming Experiment)
Color Name Models

“Do two colors have the same name?”

“Which color has a more consistent name?”
Color Name Models

- Preserving color names
Color Name Models

- Colors with consistent names
Previous Work

- **Single term**
  - Focus plus boundary
  - Simple partitioning

- **Prototypes**
  - Focus plus distribution
  - Linear combination of names

- **Non-parametric models**
  - Statistical representation
  - Requires sufficient data
Single Terms

Partition the color space into name regions

- Illustrating basic color terms (Berlin & Kay 1969)
- Lin et al. 2000, Chang et al. 2004, Kelly & Judd, ...
Previous Work

• Single term
  – Focus plus boundary
  – Simple partitioning

• Prototypes
  – Focus plus distribution
  – Linear combination of names

• Non-parametric models
  – Statistical representation
  – Requires sufficient data
• Focus plus parameterized distribution
• Linear combination of a list of names
  – Gaussians (Motomura 2001)
  – Sigmoid-Gaussians (Benavente et al. 2002)
Previous Work

- **Single term**
  - Focus plus boundary
  - Simple partitioning

- **Prototypes**
  - Centroid plus distribution
  - Linear combination of names

- **Non-parametric models**
  - Statistical representation
  - Requires sufficient data
Non-Parametric Models

- Don’t assume particular distribution
- Represent each term by a histogram
  - Benavente et al. 2006
  - Moroney 2007
Issues

• Color name regions
  – Different shapes and sizes
  – Want to preserve detailed shapes of the region

• Color vocabulary
  – More than 11 basic color terms
  – Emerging terms, multilingual data
  – Want to determine important color words from data, instead of using a pre-determined list
  – Robust when rare words are included
Our Goals

- Non-parametric model
- Inclusion of all color words, possibly from multiple languages
- Support a rich set of computational tools
Overview

- Introduction
  - Color naming and categories
  - Previous work
  - Goals
- **Categorical association**
  - Probabilistic framework
  - Results
- Summary
Categorical Association

• **Notations**
  – Colors and Words

• Probabilistic Framework
  – “Examples” of colors
  – Color saliency

• Results
  – World Color Survey
  – DoloresLabs Dataset
  – Comparison
Notation

• Color naming data consists of two variables
  – Colors $C = \{ \text{red}, \text{orange}, \text{purple}, \text{green}, \text{blue}, \text{black} \}$
  – Words $W = \{ \text{“aqua”}, \text{“black”}, \text{“blue”}, \text{“brown”}, \text{“chartreuse”}, \text{“cyan”}, \ldots \}$
Relationships

• Two types of relationships
  – Given a color $\text{□}$, what are the likely words applied to the color?

\[ P(W | C=\text{□}) \]

  – Give a word “blue”, what are the likely colors the word refers to?

\[ P(C | W=\text{“blue”}) \]
Categorical Association

• Notations
  – Colors and Words

• Probabilistic Framework
  – “Example” colors
  – Color saliency

• Results
  – World Color Survey
  – DoloresLabs Dataset
  – Comparison
Categorical Association

- Represent a color by other “example” colors
Categorical Association

- Represent a color by other “example” colors

![Diagram showing categorical association with colors like Aqua, Black, Blue, Brown, Chartreuse, Cyan, etc., and their conditional probabilities P(W|C=) and P(C|W=“Aqua”).]
Categorical Association

• Represent a color by other “example” colors
Categorical Association

• Represent a color by other “example” colors

Sum up distributions, weighted by the frequency of the word.
Categorical Association

• Represent a color by other “example” colors

\[ P(C | c=\textcolor{blue}{\text{color}}) = \sum_w P(C | w)P(w | c=\textcolor{blue}{\text{color}}) \]

• Sum contribution from all color words, weighted by frequency of word
Categorical Association

- Represent a color by other “example” colors.

$P(C|c)$
Color Saliency

• Entropy = “Uncertainty”

\[ \text{Saliency} = - \frac{1}{ \text{Entropy} } \]

Low Entropy
Weakly associated with many colors

High Entropy
Strongly associated with a few colors

High Saliency

Low Saliency

• Saliency = \(-H(C|c=\text{Few Colors}) = -P(C|c) \log( P(C|c) )\)
Categorical Association

- **Notations**
  - Colors and Words

- **Probabilistic Framework**
  - “Example” colors
  - Color saliency

- **Results**
  - World Color Survey
  - DoloresLabs Dataset
  - Comparison
Color Saliency on Munsell Surface

- Based on 6 languages from World Color Survey
- How does saliency match known color foci?
Color Saliency on Munsell Surface

• Based on 6 languages from World Color Survey
• How does saliency match known color foci?
Correspondence with Known Foci

- English basic color foci observed by
  - Orange = Berlin and Kay (1969)
  - Orange = Sturges and Whitfield (1995)
Correspondence with Known Foci

• English basic color foci observed by
  – ● = Berlin and Kay (1969)
Correspondence with Known Foci

- English basic color foci observed by
  - ● = Berlin and Kay (1969)
  - □ = Sturges and Whitfield (1995)
• English basic color foci observed by
  – ⬤ = Berlin and Kay (1969)
DoloresLabs Naming Data

- Unconstrained online naming survey
  - 10,000 colors from 238 speakers
  - 1,966 distinct responses (raw ASCII strings)
  - 1,740 distinct phrases (spellings & punctuations)
  - 302 distinct words

- Non-uniformly sampled from RGB space
Color Saliency in sRGB Gamut

- Plotted in IPT space (assuming sRGB)
Color Saliency in sRGB Gamut

- Plotted in IPT space (assuming sRGB)

IPT space at 8 levels of brightness
Clusters of Salient Basic Colors

- Brown
- Red
- Blue
Clusters of Salient Non-Basic Colors

- Brightness = 0.2
- Brightness = 0.3
- Brightness = 0.4
- Brightness = 0.5
- Brightness = 0.6
- Brightness = 0.7
- Brightness = 0.8
- Brightness = 0.9

Cyan
Lavender
Clusters of Salient Non-Basic Colors

Boynton & Olson (1990)

Cyan

Lavender

Kay & McDaniel (1978)
Color Saliency in sRGB Gamut
Comparison of Munsell/sRGB

Munsell value = 2.0
IPT brightness = 0.2317

Munsell value = 3.0
IPT brightness = 0.3179

Munsell value = 4.0
IPT brightness = 0.4100

Munsell value = 5.0
IPT brightness = 0.5051

Munsell value = 6.0
IPT brightness = 0.6032

Munsell value = 7.0
IPT brightness = 0.7041

Munsell value = 8.0
IPT brightness = 0.8106

Munsell value = 9.0
IPT brightness = 0.9337
Comparison of Munsell/sRGB
Comparison of Munsell/sRGB

Munsell value = 2.0  
IPT brightness = 0.2317

Munsell value = 3.0  
IPT brightness = 0.3179

Munsell value = 4.0  
IPT brightness = 0.4100

Munsell value = 5.0  
IPT brightness = 0.5051

Munsell value = 6.0  
IPT brightness = 0.6032

Munsell value = 7.0  
IPT brightness = 0.7041

Munsell value = 8.0  
IPT brightness = 0.8106

Munsell value = 9.0  
IPT brightness = 0.9337

sRGB Gamut
Comparison of Munsell/sRGB

Munsell value = 2.0
IPT brightness = 0.2317

Munsell value = 3.0
IPT brightness = 0.3179

Munsell value = 4.0
IPT brightness = 0.4100

Munsell value = 5.0
IPT brightness = 0.5051

Munsell value = 6.0
IPT brightness = 0.6032

Munsell value = 7.0
IPT brightness = 0.7041

Munsell value = 8.0
IPT brightness = 0.8106

Munsell value = 9.0
IPT brightness = 0.9337
Comparison at Munsell Value=7

Munsell Value = 7
IPT Brightness = 0.7041
Comparison at Munsell Value=7

Salient yellow on reflective Surface

Salient yellow on sRGB display

Munsell Value = 7
IPT Brightness = 0.7041
Saliency Differs for Munsell/sRGB

- Munsell value = 2.0
  IPT brightness = 0.2317

- Munsell value = 3.0
  IPT brightness = 0.3179

- Munsell value = 4.0
  IPT brightness = 0.4100

- Munsell value = 5.0
  IPT brightness = 0.5051

- Munsell value = 6.0
  IPT brightness = 0.6032

- Munsell value = 7.0
  IPT brightness = 0.7041

- Munsell value = 8.0
  IPT brightness = 0.8106

- Munsell value = 9.0
  IPT brightness = 0.9337
Overview

• Introduction
  – Color naming and categories
  – Previous work
  – Goals

• Categorical association
  – Probabilistic framework
  – Results

• Summary
Summary

• Categorical association between colors
  – Non-parametric model
    • Captured details and was robust on a large noisy dataset
  – Inclusion of all color words
    • Merged cross-linguistic data
    • Identified contributions from emerging non-basic terms
  – Computational tools
    • Saliency based on entropy
    • Saliency on Munsell surface and for sRGB Gamut