Evaluating the CIE 1931 Color Matching Functions

Mark Shaw
Munsell Color Science Laboratory

Thesis Advisor : Mark Fairchild
Presentation Overview

- Aim and Objectives
- Metameric Data Sets
- Computational Analysis
- Deriving an Optimized Set of Weighing Functions
- CRT Simulation
- Conclusions
Aim and Objectives

Aim
- To test the accuracy of the CIE Color Matching Functions
- Benchmark performance against other weighting functions

Objectives
- Perform computational analysis
- Derive an optimized set of weighing functions
- Develop a CRT Simulation
Background

- CIE 1931 Color Matching Functions
  - Constructed from relative color matching data of Wright (1928-9), and Guild (1931)
  - Assumptions
    - Tri-chromacy
    - $V\lambda$ is a linear combination of the 3 weighing functions
    - Grassmann’s Laws
Background

- **Stiles Verification**
  - An investigation encouraged by a CIE committee
  - Concern specifically focussed on
    - Use of $V_\lambda$
    - Discrepancies between perceived and measured colors
    - Applicability to large visual fields
  - Stiles report showed significant differences, but not large enough to warrant a change in the standard data.
Background

- Vision Research
  - Many scientists have questioned the work of the CIE
  - Vos and Walraven (1970)
  - Smith and Pokorny (1971)
  - Stockman, MacLeod and Johnson (1983)
  - Stockman and Sharpe (1998)
Thesis Foundation

- Evaluate the CIE 1931 Color Matching Functions
  - Are the claims valid or insignificant
  - What kind of a difference will a new set of functions make

- Metameric color matching data
  - Test weighting functions with ‘visual matches’
  - Using multiple observers
● Derive an Optimized Set of Weighting Functions
  - Can one improve on the CIE
  - How does observer metamerism affect the results

● CRT Simulation
  - Side by side comparison
Metameric Data Sets

- 3 Data Sets Used
  - Alfvin, 1995
  - Shaw, 1998
  - Shaw, 1999

- Underlying Concept
  - An ideal set of weighing functions will yield a minimal average color difference over all matches
Alfvin Data Set

- Cross media color match
- 7 color centers
- 20 Observers
CRT - Hardcopy Match

Color matches are inherently metameric

Evaluating the CIE 1931 Color Matching Functions
Shaw ‘98 Data Set

- ACS VCS additive color mixture device
- Munsell N5 gray
- 2 primary sets - RGB and BYP
- 6 Observers
Shaw ‘99 Data Set

- Identical experimental setup to Shaw ‘98
- Matching Fujix neutral gray
- 4 Observers
- 2 Primary Sets - RGB and GYP
- 10 matches per observer, per primary set
Color Space Comparison

- How can one compare the accuracy of different sets of weighting functions?
  - What metric should one use?
  - In which color space?
  - How can one compare the results?
Tristimulus Rotation

- Rotate each set of weighting functions to CIE approximation
- Use a linear transformation matrix

\[
\begin{bmatrix}
  x_{400} & y_{400} & z_{400} \\
  \vdots & \vdots & \vdots \\
  x_{700} & y_{700} & z_{700}
\end{bmatrix}
\rightarrow
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  a_{31} & a_{32} & a_{33}
\end{bmatrix}
\begin{bmatrix}
  r_{400} & g_{400} & b_{400} \\
  \vdots & \vdots & \vdots \\
  r_{700} & g_{700} & b_{700}
\end{bmatrix}_{\text{Stiles}}
\]
Tristimulus Rotation

Evaluating the CIE 1931 Color Matching Functions
Evaluating the CIE 1931 Color Matching Functions
Tristimulus Calculation and CIELAB

- Tristimulus values
  - Rotated weighing functions
  - Spectral radiance data

- CIELAB Values
  - XnYnZn reference set at 5 times radiance of gray
  - $\Delta L_{ab}$, $\Delta a_{ab}$, $\Delta b_{ab}$ for each pair
  - Color difference using $\Delta E_{ab}$ and $\Delta E_{94}$
Statistical Analysis

- **Students t-test**
  - Evaluate the distribution of error
  - Applied to $\Delta L$, $\Delta a$, $\Delta b$

- **Multivariate 95% Confidence Ellipse**
  - Evaluate multivariate case
  - Defined by sample variances and covariances of $\Delta L$, $\Delta a$, $\Delta b$
## Results - Computational Analysis

### CIE 1931 Color Matching Functions

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>4.560</td>
<td>3.018</td>
<td>22.300</td>
<td>0.223</td>
</tr>
<tr>
<td>DE 94</td>
<td>3.412</td>
<td>2.409</td>
<td>13.470</td>
<td>0.000</td>
</tr>
<tr>
<td>DL</td>
<td>0.332</td>
<td>2.696</td>
<td>13.410</td>
<td>-13.100</td>
</tr>
<tr>
<td>Da</td>
<td>-0.754</td>
<td>2.696</td>
<td>15.440</td>
<td>-8.384</td>
</tr>
<tr>
<td>Db</td>
<td>0.681</td>
<td>3.260</td>
<td>18.030</td>
<td>-21.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Critical Value</th>
<th>Test Statistic</th>
<th>P Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL</td>
<td>1.965</td>
<td>2.668</td>
<td>0.004</td>
<td>Reject Null</td>
</tr>
<tr>
<td>Da</td>
<td>1.965</td>
<td>-4.935</td>
<td>0.000</td>
<td>Reject Null</td>
</tr>
<tr>
<td>Db</td>
<td>1.965</td>
<td>4.516</td>
<td>0.000</td>
<td>Reject Null</td>
</tr>
</tbody>
</table>

### CIE 1964 Color Matching Functions

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>4.017</td>
<td>2.970</td>
<td>22.720</td>
<td>0.429</td>
</tr>
<tr>
<td>DE 94</td>
<td>2.818</td>
<td>2.087</td>
<td>13.780</td>
<td>0.000</td>
</tr>
<tr>
<td>DL</td>
<td>0.260</td>
<td>2.712</td>
<td>13.650</td>
<td>-13.260</td>
</tr>
<tr>
<td>Da</td>
<td>-0.029</td>
<td>2.712</td>
<td>12.270</td>
<td>-8.756</td>
</tr>
<tr>
<td>Db</td>
<td>0.321</td>
<td>3.143</td>
<td>18.250</td>
<td>-22.170</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Critical Value</th>
<th>Test Statistic</th>
<th>P Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL</td>
<td>1.965</td>
<td>2.070</td>
<td>0.020</td>
<td>Reject Null</td>
</tr>
<tr>
<td>Da</td>
<td>1.965</td>
<td>-0.224</td>
<td>0.411</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>Db</td>
<td>1.965</td>
<td>2.208</td>
<td>0.014</td>
<td>Reject Null</td>
</tr>
</tbody>
</table>

### Stiles and Burch Color Matching Functions

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>4.142</td>
<td>2.911</td>
<td>23.200</td>
<td>0.287</td>
</tr>
<tr>
<td>DE 94</td>
<td>3.007</td>
<td>2.133</td>
<td>13.640</td>
<td>0.216</td>
</tr>
<tr>
<td>DL</td>
<td>0.295</td>
<td>2.700</td>
<td>13.580</td>
<td>-13.160</td>
</tr>
<tr>
<td>Da</td>
<td>-0.537</td>
<td>2.700</td>
<td>14.450</td>
<td>-8.557</td>
</tr>
<tr>
<td>Db</td>
<td>-0.065</td>
<td>3.137</td>
<td>17.620</td>
<td>-22.870</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Critical Value</th>
<th>Test Statistic</th>
<th>P Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL</td>
<td>1.965</td>
<td>2.367</td>
<td>0.009</td>
<td>Reject Null</td>
</tr>
<tr>
<td>Da</td>
<td>1.965</td>
<td>-4.071</td>
<td>0.000</td>
<td>Reject Null</td>
</tr>
<tr>
<td>Db</td>
<td>1.965</td>
<td>-0.451</td>
<td>0.326</td>
<td>Fail toReject</td>
</tr>
</tbody>
</table>

Evaluating the CIE 1931 Color Matching Functions
# Results - Computational Analysis

## Demarco, Smith and Pokorny Cone Fundamentals

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>4.364</td>
<td>2.939</td>
<td>23.210</td>
<td>0.350</td>
</tr>
<tr>
<td>DE 94:</td>
<td>3.240</td>
<td>2.225</td>
<td>13.480</td>
<td>0.000</td>
</tr>
<tr>
<td>DL</td>
<td>0.350</td>
<td>2.697</td>
<td>13.410</td>
<td>-13.040</td>
</tr>
<tr>
<td>Da</td>
<td>-0.255</td>
<td>2.697</td>
<td>15.930</td>
<td>-8.107</td>
</tr>
<tr>
<td>Db</td>
<td>-0.232</td>
<td>3.179</td>
<td>16.890</td>
<td>-22.800</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Critical Value</th>
<th>Test Statistic</th>
<th>P Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL</td>
<td>1.965</td>
<td>2.810</td>
<td>0.003</td>
<td>Reject Null</td>
</tr>
<tr>
<td>Da</td>
<td>1.965</td>
<td>-1.737</td>
<td>0.042</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>Db</td>
<td>1.965</td>
<td>-1.580</td>
<td>0.057</td>
<td>Fail to Reject</td>
</tr>
</tbody>
</table>

## Stockman, MacLeod and Johnson Cone Fundamentals

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>4.142</td>
<td>2.909</td>
<td>23.180</td>
<td>0.288</td>
</tr>
<tr>
<td>DE 94:</td>
<td>2.997</td>
<td>2.136</td>
<td>13.640</td>
<td>0.000</td>
</tr>
<tr>
<td>DL</td>
<td>0.296</td>
<td>2.700</td>
<td>13.580</td>
<td>-13.160</td>
</tr>
<tr>
<td>Da</td>
<td>-0.540</td>
<td>2.700</td>
<td>14.450</td>
<td>-8.568</td>
</tr>
<tr>
<td>Db</td>
<td>-0.056</td>
<td>3.136</td>
<td>17.620</td>
<td>-22.840</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Critical Value</th>
<th>Test Statistic</th>
<th>P Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL</td>
<td>1.965</td>
<td>2.358</td>
<td>0.009</td>
<td>Reject Null</td>
</tr>
<tr>
<td>Da</td>
<td>1.965</td>
<td>-4.089</td>
<td>0.000</td>
<td>Reject Null</td>
</tr>
<tr>
<td>Db</td>
<td>1.965</td>
<td>-0.383</td>
<td>0.351</td>
<td>Fail to Reject</td>
</tr>
</tbody>
</table>

## Vos and Walraven Cone Fundamentals

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std Dev</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>4.365</td>
<td>2.939</td>
<td>23.220</td>
<td>0.350</td>
</tr>
<tr>
<td>DE 94:</td>
<td>3.242</td>
<td>2.222</td>
<td>13.480</td>
<td>0.350</td>
</tr>
<tr>
<td>DL</td>
<td>0.351</td>
<td>2.697</td>
<td>13.410</td>
<td>-13.040</td>
</tr>
<tr>
<td>Da</td>
<td>-0.249</td>
<td>2.697</td>
<td>15.940</td>
<td>-8.104</td>
</tr>
<tr>
<td>Db</td>
<td>-0.238</td>
<td>3.178</td>
<td>16.880</td>
<td>-22.810</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Critical Value</th>
<th>Test Statistic</th>
<th>P Value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL</td>
<td>1.965</td>
<td>2.813</td>
<td>0.003</td>
<td>Reject Null</td>
</tr>
<tr>
<td>Da</td>
<td>1.965</td>
<td>-1.694</td>
<td>0.045</td>
<td>Fail to Reject</td>
</tr>
<tr>
<td>Db</td>
<td>1.965</td>
<td>-1.620</td>
<td>0.053</td>
<td>Fail to Reject</td>
</tr>
</tbody>
</table>
95% Confidence Ellipses - $\Delta a \Delta b$

CIE 1931 Color Matching Functions

CIE 1964 Color Matching Functions

Evaluating the CIE 1931 Color Matching Functions
95% Confidence Ellipses - $\Delta L\Delta a$

CIE 1931 Color Matching Functions

CIE 1964 Color Matching Functions

Evaluating the CIE 1931 Color Matching Functions
95% Confidence Ellipses - $\Delta L \Delta b$

CIE 1931 Color Matching Functions

CIE 1964 Color Matching Functions

Evaluating the CIE 1931 Color Matching Functions
Summary of Results

- **Combined Data Sets**
  - CIE 1931  
    - Average $\Delta E_{ab}$ 4.56  
    - Std Dev 3.02  
    - Average $\Delta E_{94}$ 3.41  
    - Std Dev 2.41  
  - CIE 1964  
    - Average $\Delta E_{ab}$ 4.02  
    - Std Dev 2.97  
    - Average $\Delta E_{94}$ 2.82  
    - Std Dev 2.09

- Tests indicate CIE 1931 is not centered at 0

- Every set of weighting functions is offset in at least one dimension
Deriving an Optimized Set of Weighting Functions

- Assume the data is truly metameric
- Optimal set - $0\Delta E_{ab}$ Average
- Observer metamerism
- How much low can $\Delta E_{ab}$ go?
- Must maintain the integrity of the CMFs
Various Approaches Tried

- Linear Regression
- Linear Regression + Cubic Splines
- Unconstrained Non-linear - Powell
- Constrained Non-linear - Levenberg-Marquardt
- Monte Carlo
- Newton's Variation
Important Attributes

- Optimized functions must be smooth
- Output functions must be ‘realistic’
- Optimize over combined data sets
Best Method

- Constrained Non-Linear

\[ \bar{x}_{\text{optimum}} = \sum_{j=1}^{6} w_j C_j \quad \rightarrow \quad w_j = \text{Arg min} \Delta E_{\text{rotated}} \]

where

- \( \bar{x}_{\text{optimum}} \) = Optimized weighting function
- \( w_j \) = \( j \)th Weight
- \( C_j \) = \( j \)th Rotated Weighting Function

constraint

\[ w_1 = 1 - w_2 - w_3 - w_4 - w_5 - w_6 \]
Constrained Non-linear

1. Estimate Weights
2. Create Functions
3. Rotate Functions
   - Reference Spectra
   - Sample Spectra
4. Calculate XYZs
5. Calculate Lab’s
6. Calculate $\Delta E_{ab}$
7. Re-estimate Weights

Evaluating the CIE 1931 Color Matching Functions
Optimized Weighting Functions

- **Average**
  - $\Delta E_{ab}$ 3.92
  - $\Delta E_{94}$ 2.78

- **Shift in peak sensitivity of $\bar{z}$ function**

- **Change in $\bar{x}$ below 460nm**
Optimization Summary

- 14 optimization techniques were tried
- Improvement was only $0.1 \Delta E_{ab}$

<table>
<thead>
<tr>
<th>CIE 1964 Color Matching</th>
<th>Shaw and Fairchild Color Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td><strong>Std Dev</strong></td>
<td><strong>Std Dev</strong></td>
</tr>
<tr>
<td>DE</td>
<td>4.017</td>
</tr>
<tr>
<td>DE 94:</td>
<td>2.818</td>
</tr>
<tr>
<td>DL</td>
<td>0.260</td>
</tr>
<tr>
<td>Da</td>
<td>-0.029</td>
</tr>
<tr>
<td>Db</td>
<td>0.321</td>
</tr>
</tbody>
</table>

Evaluating the CIE 1931 Color Matching Functions
CRT Simulation

● Purpose
  - Simulate the differences between each set of weighting functions

● Calibration
  - GOG Model
  - Spectrally Calibrated
CRT Modeling

Determining Flare XYZ to RGB transform

GOG Model

XYZ\textsubscript{CIE31}

XYZ\textsubscript{CIE64}

XYZ\textsubscript{Stiles}

XYZ\textsubscript{Demarco}

XYZ\textsubscript{Stockman}

XYZ\textsubscript{Vos}

XYZ\textsubscript{Shaw}

RGB\textsubscript{CIE31}

RGB\textsubscript{CIE64}

RGB\textsubscript{Stiles}

RGB\textsubscript{Demarco}

RGB\textsubscript{Stockman}

RGB\textsubscript{Vos}

RGB\textsubscript{Shaw}

Evaluating the CIE 1931 Color Matching Functions
CRT Simulation

Viewing Booth

PhotoResearch

CRT

Evaluating the CIE 1931 Color Matching Functions
CRT Simulation

Evaluating the CIE 1931 Color Matching Functions
Conclusions

- The performance of all sets of functions were similar
- CIE 1964 10° functions performed best
- CIE 1931 2° functions performed worst
- Only 0.5ΔE_{ab} difference
- Optimized functions only attained 0.1ΔE_{ab} improvement
Conclusions

- Simulation exemplifies differences
- Observer Metamerism is an important consideration
  - Less emphasis on accuracy of weighing functions
  - Observer variance is far greater
Acknowledgements

- Mark Fairchild
  - Thesis Advisor

- Noboru Ohta
  - Computational Advisor

- Munsell Laboratory Staff and Students
  - Their support and encouragement to get this done on time