Automatic color constancy algorithm selection and combination

Pattern Recognition,
Vol. 43, No. 3, 2010
S. Bianco, G. Ciocca, C. Cusano, and R. Schettini

Presented by Woo-Heon Jang

School of Electrical Engineering and Computer Science
Kyungpook National Univ.
Abstract

◆ Proposed method
  – Selection of best algorithm for given image
    • Decision of several trees on heterogeneous features
      – Representing features
        » Image content in terms of low level visual properties
      – Trained trees for selection
        » Minimization of expected error in illuminant estimation
    • Estimating illuminant as weighted sum of different algorithms estimation
Aim of computational color constancy

- Estimation of illuminant color in acquired scene
  - Color invariant
    - Image data without estimating scene illuminant
  - Illuminant estimation
    - Image data with estimating scene illuminant
- Assumptions about statistical properties
  - Expected illuminant
  - Object reflectance in scene
- Hordley’s method
  - Improving performance of color constancy algorithm
    - Additional measurement at time of image capture
    - Combining algorithms
  - Illuminant estimation by using high level visual information on image
Automatic color correction

- Content based image analysis
  - Schroder and Moser’ method
    - Classification of image into signal-oriented generic classes
    - Class-specific application of set of color correction algorithm
      » White patch algorithm
      » Gray world algorithm
    - Consideration of class-specific reliability of each algorithm
    - Analysis of hierarchical Bayesian image content
      » Consistency of feature extraction
      » Unsupervised clustering
• Gasparini and Schettini’s method
  – Adaptive mixture of algorithm
    » White balance
    » Gray world algorithm
  – Prevention of mistaken removal of intrinsic color
    » Temporary removal from analyzed image

• Van de Weijer et al.
  – High-level visual information for improving illuminant estimation
  – Mixture of semantic classes
  – Computing set of possible illuminant estimation
  – Best of semantic composition of selected image
  – Approach of outdoor images better than other tested algorithms
- Classification algorithm into indoor and outdoor
  - Optimally tuned algorithm parameters
    - Indoor class
    - Outdoor class
    - Both class
  - Improved illuminant estimation of semantic class
  - Using of semantic information on image content
    - Improved illuminant estimation
      » Accurate choice
• Gihsenij and Gerves’s method
  – Using of natural image statistic
    » Identification of most important characteristic of color images
    » Achieve selection of color constancy algorithm
    » Achieve combination of color constancy algorithm
  – Using Weibull parameterization on first order derivative filter
    » Direction for image characteristic
  – Using -means algorithm
    » Clustering parameter
  – Combination of best-suited color constancy algorithm
    » Each cluster
  – Application of 75 different instantiation of color constancy algorithms
– Proposed method

  • Automatic extraction of illuminant estimation algorithm
    – Set of visual features
  • Evaluation of illuminant estimation strategies
    – Response of trees on single algorithm for first selection
    – Using of small set of simple
Proposed framework

◆ Framework

– Image values for Lamvertian surface

\[ \rho(x, y) = \int_{\omega} I(\lambda) S(x, y, \lambda) C(\lambda) d\lambda \] (1)

where \( \rho(x, y) \) is sensor response
\( I(\lambda) \) is illuminant spectral power distribution,
\( S(\lambda) \) is surface spectral reflectance,
\( C(\lambda) \) is sensor spectral sensitivities,
\( \omega \) is the wavelength range of visible light spectrum,
\( C(\lambda) \) and \( \rho \) are three-component vectors.

– Goal of color constancy

• Estimating color of scene illuminant

\[ I = \int_{\omega} I(\lambda) C(\lambda) d\lambda \] (2)
– Proposed classification
  • Improving performance of color constancy algorithm
  • Selection of image category
– Van de Weijer er al.
  • Unified variety of algorithm
  • Approximation of illuminant color by implementing instantiation

\[ I(n, p, \sigma) = \frac{1}{k} \left( \iint |\nabla^n \rho_\sigma(x, y)|^p \, dx \, dy \right)^{1/p} \] (3)

where \( n \) is the order of derivative
\( p \) is the Minkowski norm
\( \rho_\sigma(x, y) = \rho(x, y) \otimes G_\sigma(x, y) \) is the convolution of image with Gaussian filter \( G_\sigma(x, y) \) with scale parameter \( \sigma \), and
\( k \) is a constant to chosen such that the illuminant color \( I \) has unit length.
– Generated four algorithm by three variables
  
  • Gray world algorithm
    – Achromatic color by average reflectance in scene
    – Generated setting \((n, p, \sigma) = (0,1,0)\)
  
  • White point algorithm
    – Achromatic color by maximum reflectance in scene
    – Generated setting \((n, p, \sigma) = (0,\infty,0)\)
  
  • Gray edge algorithm
    – Achromatic color by \(p\)-th Minkowski norm of first order derivative in scene
    – Generated setting \((n, p, \sigma) = (1, p, \sigma)\)
• Second order gray edge algorithm
  – Achromatic color by $p$-th Minkowski norm of first order derivative in scene
  – Generated setting $(n, p, \sigma) = (2, p, \sigma)$
• Do nothing algorithm
  – Same estimation of illuminant color in every image
  – Generated setting $(n, p, \sigma) = (0, 0, 0)$
- CART methodology
  - Selecting algorithm with given image
  - Classifier in forest vote for illuminant estimation algorithm
  - Consideration of error cost
    - Erroneous algorithm selection by training classifier
      » Minimization of expected error in illuminant estimation
  - Estimation of illuminant
    - Weighted sum of estimating algorithm
      » Using vote of classifier in forest
      » Selected classification
Algorithm selection by decision forests

Classification tree

- Advantages of suitable classification tree
  - Trained classification by arbitrary high number of classes
    - Consideration of any set of illuminant estimation algorithm
  - Approval of feature vectors
    - Composition of several heterogeneous features
  - Information about correlation
    - Errors of different algorithms in selection algorithm
    - Priori probability
    - Misclassification cost
CART methodology

- Classification and regression trees
- Effective method for image classification
- Recursively partitioning set of feature vector
  - \( T = \{x_1, \ldots, x_N\} \)
  - Corresponding correction class
    - \( \{y_1, \ldots, y_N\}, \ y_j \in \{1, \ldots, K\} \)
  - Impurity function
    - Measuring diversity of associated classes
    - Basis of estimated distribution of classes in set
    - Gini diversity index
• Gini diversity index

\[ i_{\text{Gini}}(P_1, \ldots, P_K) = 1 - \sum_{j=1}^{k} P_j^2 \]

• Computing decrease in impurity

\[ \Delta I(j, \tau) = i(T) - i(T_L)P_L - i(T_R)P_R \tag{4} \]

where \( P_L \) and \( P_R \) are the resubstitution estimates of probabilities, \( T_L \) and \( T_R \) are subsets, and \( i(T) \) is impurity function to resubstitution estimates of distribution of classes of elements of set.

• Misclassification error

\[ \hat{y} = \arg \max P(y = j|L) \tag{5} \]

where \( L \) is terminal node.
• Cost-complexity criterion
  – Elimination of overfitting
    \[ R_\alpha (T) = R(T) + \alpha |T| \]  
    where \( R(T) \) is the probability of misclassification estimated on training set, \( |T| \) is the size of tree, and \( \alpha \) is a parameter which weights prediction errors and complexity of tree.

• Pruning process
  – Improved accuracy of trees
  – Low misclassification error
    » Instability of training procedure
– Classifier
  • Selection for illuminant estimation algorithm
    – Basis of image content
    – Final step algorithm
      » Applying algorithm of lowest estimation error on image
– Straightforward application of CART training process
  • Generation of poor results
    – Existence of excellent algorithm
    – Creation of high variability by non-optimal choice
– Acquiring algorithm of two algorithms for best choice

\[
c(k|h) = \frac{\sum_{j:y_j = h} e_j^{(k)} - e_j^{(h)}}{|\{j : y_j = h\}|}, \quad h, k \in \{1, \ldots, K\}
\]  

(7)

where \( e_j^{(k)} \) is the error of the \( k \)-th algorithm on the \( j \)-th training sample, and 
\( c(k|h) \) is expected cost caused by choice of algorithm \( k \) when 
algorithm \( h \) is the best choice.

– Misclassification costs

• Influence between pruning and label assignment

\[
\hat{y} = \arg\min_{j \in \{1, \ldots, K\}} \sum_{h=1}^K c(j|h) P(y = h|L)
\]  

(8)
– Average cost of decision of tree

\[ R(T) = \sum_{L \in \tilde{T}} \left( \min_{j \in \{1, \ldots, K\}} \sum_{h=1}^{K} c(j|h) P(y = h|L) \right) P(L) \]  \hspace{1cm} (9)

where \( \tilde{T} \) is the set of leaves of \( T \), and \( P(L) \) is the resubstitution estimate of probability that a case falls in the leaf \( L \).

– Twoing criterion

• Dividing set of classes into two micro-class
• Decrease of classified effort between similar classes

\[ \Delta I_{\text{twoing}}(j, \tau) = \frac{P_L P_R}{4} \left[ \sum_{k=1}^{K} \left| P(y = k|T_L) - P(y = k|T_R) \right| \right]^2 \]  \hspace{1cm} (10)
Image features

- Considered groups of low level features
  - General purpose features
    - Needlessness of capture characteristic of image
    - Color histogram
    - Edge direction histogram
    - Statistics on wavelet coefficients
    - Color moment
• Problem-dependent features
  – Capture properties of images
    » Improving performance of task under consideration
  – Number of colors
  – Clipped color components
  – Cast indexes

**Table 1.** Summary of the features used to describe the images.

<table>
<thead>
<tr>
<th>Name</th>
<th>No. components</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>YCbCr color moments</td>
<td>6</td>
<td>Color</td>
</tr>
<tr>
<td>RGB color histogram</td>
<td>27</td>
<td>Color</td>
</tr>
<tr>
<td>Number of colors</td>
<td>1</td>
<td>Color</td>
</tr>
<tr>
<td>Cast indexes</td>
<td>2</td>
<td>Color</td>
</tr>
<tr>
<td>Color clipping</td>
<td>8</td>
<td>Color</td>
</tr>
<tr>
<td>Edge magnitude histogram</td>
<td>5</td>
<td>Edges</td>
</tr>
<tr>
<td>Edge direction histogram</td>
<td>18</td>
<td>Edges</td>
</tr>
<tr>
<td>Wavelet statistics</td>
<td>20</td>
<td>Texture</td>
</tr>
</tbody>
</table>
Edge direction histogram

- Edge
  - Clue of depicted subject in image
- Determination of edge structure in image
- Distinction between different image classes
- Computing edge
  - Gaussian filter with $\sigma = 1$ in luminance image in direction
- Computing edge orientation at edge position

$$\theta(x, y) = \arctan \left( \frac{G_y(x, y)}{G_R(x, y)} \right)$$  \hspace{1cm} (11)

- Quantized orientations into 18 bins
• Feature of edge

**Fig. 1.** Examples of images of man-made structures.

**Fig. 2.** Examples of images of natural scenes.
– Edge strengths
  • Computing histogram of edge magnitudes
    – Estimating relevance of edges
  • Different edge strength in image

Fig. 3. Examples of images with weak(left) and strong(right) edge magnitudes.
– Color histogram
  • Representing color distribution of image
  • Quantized RGB color space by dividing color axis into three intervals
– Wavelet statistics
  • Obtained information by wavelet decomposition
    – Textures
    – Structures

**Fig. 4.** A three-iteration Daubechies wavelet decomposition.
– YCbCr color moments
  • Separation of luminance component from chrominance component
  • Color transformation

\[
\begin{pmatrix}
Y \\
Cb \\
Cr
\end{pmatrix} = \begin{pmatrix}
16 \\
-37.95 \\
112.44
\end{pmatrix} + \begin{pmatrix}
65.74 & 129.06 & 25.06 \\
-74.50 & 112.44 \\
-94.15 & -18.29
\end{pmatrix} \begin{pmatrix}
r \\
g \\
b
\end{pmatrix}
\] (12)

– Number of colors
  • Related color range of image
  • Gray world assumption
  • Maximum number of different color
    – \(2^6 \times 2^6 \times 2^6 = 262144\)
• Images with different colors

**Fig. 5.** Left image: 10782 colors with average color (122, 123, 121)
Right image: 13882 colors with average color (107, 106, 110).

**Fig. 6.** Left image: 5380 colors with average color (150, 97, 47)
Right image: 7538 colors with average color (96, 126, 150).
– Clipped color components
  • Extent of highly saturated color pixels
  • Representing histogram of probability density distribution

– Cast indexes
  • Identifying presence of relevant cast with image
  • Modification of original formulation
    – Color space representation from CIELAB to YCbCr
    – CbCr plane
      » Means
      » Variance
      » Using color equivalence circle center of means
      » Radius of variance
    – Estimation of color distribution from neutral axis
Fig. 7. Example of an image with a strong color cast. The equivalence circle is compact and far from the neutral axis.

Fig. 8. Example of an image without a strong color cast. The equivalence circle is compact and far from the neutral axis.


**Experiment results**

- **Performance evaluation**
  - Error measure
    - Intensity independent
    - Angle between illuminant color and estimation color
      \[
      e_{ANG} = \arccos \left( \frac{\rho_w^T \hat{\rho}_w}{\|\rho_w^T\| \|\hat{\rho}_w\|} \right)
      \]
      where \( \rho_w \) is illuminant color, and \( \hat{\rho}_w \) is algorithm’s estimated color.
    - Median angular error
      - Comparison of two color constancy algorithms
Tuning of the color constancy algorithms
  – Consideration of two color constancy algorithms
    • Gray edge algorithm
      – \((p, \sigma) = (1.10, 1.08)\)
    • Second order gray edge algorithm
      – \((p, \sigma) = (1.55, 1.83)\)
    • Using pattern search method
      – Direct search method for nonlinear optimization

\[ x^+ = x_k + \Delta_k d_k \]

where \(x_k\) is current iterate, \(d_k\) is directions, and \(\Delta_k\) is parameters.
Decision forest training and evaluation

- Priori probability for each algorithm at best choice
- Misclassification costs

**Table 2.** A priori probabilities, corresponding to five illuminant estimation algorithms, estimated on images of training set.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DN</td>
<td>0.33</td>
</tr>
<tr>
<td>GW</td>
<td>0.34</td>
</tr>
<tr>
<td>WP</td>
<td>0.04</td>
</tr>
<tr>
<td>GE1</td>
<td>0.12</td>
</tr>
<tr>
<td>GE2</td>
<td>0.17</td>
</tr>
</tbody>
</table>

**Table 3.** Matrix of the misclassification costs estimated on the images of the training set (7).

<table>
<thead>
<tr>
<th>Best algorithm</th>
<th>Predicted algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DN</td>
</tr>
<tr>
<td>DN</td>
<td>0.00</td>
</tr>
<tr>
<td>GW</td>
<td>8.43</td>
</tr>
<tr>
<td>WP</td>
<td>0.50</td>
</tr>
<tr>
<td>GE1</td>
<td>2.80</td>
</tr>
<tr>
<td>GE2</td>
<td>2.86</td>
</tr>
</tbody>
</table>
**Fig. 9.** Histogram of the occurrences of the features in the splits of the trained trees.

**Table 4.** Confusion matrix of the decision forest used for algorithm selection, estimated on the images of the test set.
**Fig. 10.** Distribution of the rank of the algorithm selected by the decision forest on the images of the test set.

**Fig. 11.** Distribution of the difference in angular error between the algorithms selected by the decision forest and the best choice for each image of the test set.
Table 5. Summary of the results obtained on the test set by the Classification-based Algorithm Selection (CAS) strategy.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Median</th>
<th>Mean</th>
<th>WSTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DN</td>
<td>6.05</td>
<td>8.07</td>
<td>0</td>
</tr>
<tr>
<td>GW</td>
<td>5.95</td>
<td>7.27</td>
<td>0</td>
</tr>
<tr>
<td>WP</td>
<td>5.48</td>
<td>7.45</td>
<td>2</td>
</tr>
<tr>
<td>GE1</td>
<td>4.47</td>
<td>5.84</td>
<td>4</td>
</tr>
<tr>
<td>GE2</td>
<td>4.65</td>
<td>6.23</td>
<td>3</td>
</tr>
<tr>
<td>CAS</td>
<td>3.21</td>
<td>4.76</td>
<td>6</td>
</tr>
<tr>
<td>Semantic</td>
<td>3.54</td>
<td>4.89</td>
<td>5</td>
</tr>
<tr>
<td>Ideal classifier</td>
<td>2.31</td>
<td>3.27</td>
<td>–</td>
</tr>
</tbody>
</table>

The best score for each column are reported in bold.

Table 6. Summary of the results obtained on the test set by the Classification-based Algorithm Selection (CAS) strategy.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Median</th>
<th>Mean</th>
<th>WSTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVG</td>
<td>4.66</td>
<td>5.99</td>
<td>0</td>
</tr>
<tr>
<td>LMS</td>
<td>4.12</td>
<td>5.29</td>
<td>2</td>
</tr>
<tr>
<td>N2M</td>
<td>4.79</td>
<td>5.82</td>
<td>0</td>
</tr>
<tr>
<td>CAC</td>
<td>3.04</td>
<td>4.46</td>
<td>3</td>
</tr>
</tbody>
</table>

The best score for each column are reported in bold.
**Fig. 12.** Average angular error obtained by the five illuminant estimation algorithms on the images of the set, as a function of the number of votes received by the trees of the decision forest.
Conclusions

- Automatic illuminant estimation
  - Selection of simple algorithms
  - Combination of simple algorithms
  - Improvement of illuminant estimation accuracy
    - Trained decision forest
      - Identification of best algorithm for given image
    - Classification based algorithm selection strategy
    - Classification based algorithm combination strategy