Minimum Description Length Characterization of Low End Color Cameras

ADRIAN BURIAN
Nokia
Technology Platforms
Visiokatu 4, Tampere
FINLAND

MIHAELA CIRLUGEA
Technical University of Cluj-Napoca
Bases of Electronics Department
Baritiu 26-28
ROMANIA

Abstract: In this paper, a new Minimum Description Length (MDL) approach for the characterization of a mobile phone's color camera is presented. The use of high-order polynomials, Fourier sine series, and artificial neural networks (ANN) for solving this problem are compared and contrasted. The MDL formalism is used for determining the stochastic complexity of polynomial and Fourier sine models for the characterization of a Nokia N90 mobile phone camera. A quantitative evaluation of their performances, as well as for using an ANN, is provided.

Key–Words: Minimum Description Length, High-Order Polynomial, Artificial Neural Network, Imaging Mobile Phone.

1 Introduction

Digital cameras are becoming increasingly important image acquisition tools to realize image and color processing. The accurate handling of the color characteristics of the obtained images is a difficult task, due to the fact that the *RGB* signals generated by a digital camera are device-dependent. Different digital cameras will produce different *RGB* responses for the same captured scene. Furthermore, while digital cameras bring simplicity in handling image capture, satisfying customer expectations is difficult as well. This is due to the fact that a camera captures the physical values of the light, while human observers are perceiving the result of processing of their visual systems.

The proliferation of camera phone devices in the consumer market has led to an increased need to transfer images among different storing/manipulating/displaying mediums without loss of color fidelity. Furthermore, the picture quality of camera phones is improving substantially enough that it is expected that these phones will begin to replace the low-end digital cameras [1]. The quality in even low-end camera phones will be enough to compete with low-end digital cameras. Already it is considered that, e.g., the two mega-pixels Nokia N90 with Carl Zeiss lens is a digital camera replacement. We must also specify that this is high-end phone, more costly than a high-end digital camera with five mega-pixels; so it is generally considered that the high-end market for digital cameras is safe.

We can consider that the images captured by a camera are depending mainly on three factors: the characteristics of the used camera, the illumination of the captured image, and on the actual color content of the scene. A common solution to obtain high-fidelity cross-media color reproduction is to characterize each device in terms of CIE tristimulus values [2]. Using appropriate characterization procedures, it is possible to convert the camera *RGB* values to CIE *XYZ* values, and then to convert back the *XYZ* values to *RGB* ones on another medium, e.g., a monitor.

In this paper, the MDL formalism will be used to compare the characterization of the digital camera of a Nokia N90 mobile phone, using polynomial transforms and Fourier sine transforms. We mention here that the MDL formalism was previously used in color processing for spatial segmentation of color images [3, 4]. In addition, ANN are used and compared with the previous two approaches for camera characterization. In [5] the Levenberg-Marquardt (LM) optimization method was used for training a fully connected multi-layer perceptron network to derive mappings between the camera responses and tristimulus values. An Agfa digital StudioCam camera, a three-chip CCD device with 8-bit resolution for each channel and 4500 $\times$ 3648 pixel spatial resolution was used. The ANN contained three input units to receive the camera responses, three output units to output the tristimulus values and a single hidden layer. The number of units in the hidden layer was varied to be 3, 5, 10, 18, 27 or 40. The conclusion from [5] was that the optimum hidden layer has 18 units, and the obtained neural network has almost identical results with the more traditional technique of polynomial transforms. We considered that adding also neural networks to our study can be of interest; this is because for high-end cameras the camera responses exhibit an approximately linear relationship with the mean reflectance, or Y tristimulus value of the grey sample, while for low-end cameras this relationship is nonlinear. Our purpose in the case of ANN solution is again to obtain the smallest complexity, so we decided to use the method from [6] for training. This training method uses MDL as stopping criterion in order to avoid over-fitting and has a feed-forward neural network architecture with a single hidden layer.

2 Camera Characterization

For the characterization of cameras, linear transforms are considered fundamental [7, 8]. The camera characterization is the relationship between device coordinates (*RGB*) and some-device independent color space, such as CIE *XYZ*.

The polynomial approach appears to be a common method for obtaining the *XYZ* tristimulus values. This approach was also used in [8] for comparing a high-end digital camera with a low-end digital camera, in terms of accuracy of colorimetric characterization and "What You See Is What You Get" color texture simulation. Their conclusion was that the high-end digital camera clearly outperformed the low-end digital camera used in tests, in terms of texture simulation. The high-end camera produced mostly acceptable and good matches while the low-end camera produced mostly bad matches. We can also drawn the conclusion that polynomial transforms do not have good performances for low-end cameras, and we consider this to be in connection with their stronger nonlinear characteristics when compared to high-end cameras.

The most common and efficient method for characterizing a digital camera is to use a chart containing a set of colors of known tristimulus values. These charts include neutral patches that may be used to linearize the camera *RGB* responses, and colored patches that may be used for camera characterization to CIE *XYZ* values.

3 Minimum Description Length

We start this section by defining the *complexity* of a given model $\mathcal{M}$ as

$$\mathcal{C}_n(\mathcal{M}) := \log \sum_{x^n \in \mathcal{X}^n} P(x^n \mid \hat{\theta}(x^n)), \qquad (1)$$

where $\bar{P}$ is a probability distribution on $\mathcal{X}^n$ ($\bar{P}$ is not necessarily in $\mathcal{M}$).

To get a first idea of why $\mathcal{C}_n$ is called model complexity, we note that the more sequences x^n with large $P(x^n \mid \hat{\theta}(x^n))$, the larger $\mathcal{C}_n(\mathcal{M})$. In other words, the more sequences that can be fit well by an element of $\mathcal{M}$, the larger $\mathcal{M}$'s complexity.

MDL tells us to pick the model $\mathcal{M}^{(j)}$ maximizing the *normalized* maximum likelihood (NML) $\bar{P}_{\text{NML}}(D \mid \mathcal{M}^{(j)})$, or, equivalently, minimizing

$$-\log \bar{P}_{\text{NML}}(D \mid \mathcal{M}^{(j)}) = -\log P(D \mid \hat{\theta}^{(j)}(D)) + \mathcal{C}_n(\mathcal{M}^{(j)}) \qquad (2)$$

From a coding theoretic point of view, we associate with each $\mathcal{M}^{(j)}$ a code with lengths $\bar{P}_{\text{NML}}(\cdot \mid \mathcal{M}^{(j)})$, and we pick the model minimizing the codelength of the data. The codelength $-\log \bar{P}_{\text{NML}}(D \mid \mathcal{M}^{(j)})$ has been called the *stochastic complexity of the data D relative to model* $\mathcal{M}^{(j)}$ [9], whereas $\mathcal{C}_n(\mathcal{M}^{(j)})$ is called the *parametric complexity* or *model cost* of $\mathcal{M}^{(j)}$. We have already indicated that $\mathcal{C}_n(\mathcal{M}^{(j)})$ measures something like the 'complexity' of model $\mathcal{M}^{(j)}$. On the other hand, $-\log P(D \mid \hat{\theta}^{(j)}(D))$ is minus the maximized log-likelihood of the data, so it measures something like (minus) fit or *error* – in the linear regression case it can be directly related to the mean squared error. Thus, (2) embodies a trade-off between lack of fit (measured by minus log-likelihood) and complexity (measured by $\mathcal{C}_n(\mathcal{M}^{(j)})$). The *confidence* in the decision is given by the codelength difference

$$\left| -\log \bar{P}_{\text{NML}}(D \mid \mathcal{M}^{(1)}) - [-\log \bar{P}_{\text{NML}}(D \mid \mathcal{M}^{(2)})] \right|.$$

In general, $-\log \bar{P}_{\text{NML}}(D \mid \mathcal{M})$ can only be evaluated numerically, except the case when $\mathcal{M}$ is the Gaussian family. In many cases even numerical evaluation is computationally problematic.

We will use MDL for modeling our data, in order to compare three models for camera characterization. Taking into consideration the polynomial transforms used in [5] and computational simplicity issues we considered the following models:

1. Polynomial model

$$\mathcal{M}_1 = a_0 + a_1 R + a_2 G + a_3 B + a_4 RG + a_5 GB + \qquad (3)$$

$$a_6 RB + a_7 R^2 + a_8 G^2 + a_9 B^2 + a_{10} R^2 GB +$$

$$a_{11} RG^2 B + a_{12} RGB^2 + \epsilon.$$

2. Fourier sine model (we considered camera response functions to be even)

$$\mathcal{M}_2 = a_0 + a_1 \sin R + a_2 \sin G + a_3 \sin B + \qquad (4)$$

$$a_4 \sin RG + a_5 \sin GB + a_6 \sin RB + a_7 \sin 2R +$$

$$a_8 \sin 2G + a_9 \sin 2B + a_{10} \sin 2RG +$$
$$a_{11} \sin 2GB + a_{12} \sin 2RB + \epsilon.$$

3. Neural network model. We will not include this model in the current computation, but we will give the results of using this model in Section 4. We just specify here that we start with a minimum size model, selectively add the needed neurons and prune the less fit neurons.

In order to estimate the optimal model by using the MDL principle we need to compute the stochastic complexity. The MDL principle states that the best model/model class among a collection of tentatively suggested ones is the one that gives the smallest stochastic complexity to the given data. In fact, by using the MDL principle we are looking for an optimal trade-off between the model complexity, which in our case is given by the number of coefficients and goodness-of-fit. The model complexity is increasing with the number of coefficients because the more coefficients $\mathcal{M}_1$ or $\mathcal{M}_2$ has, the more bits we need to describe it. Both models: polynomial model $\mathcal{M}_1$ and sine Fourier series model $\mathcal{M}_2$ will be used to compress the description of data points. The *RGB* values are regarded as given so we do not have to encode them.

When dealing with color images, which are multi-compo-nent images, a common problem is how to exploit the information present in various components. We have used a multi-dimensional Gaussian probability like in [4], to model the residual noise ϵ, se we considered the following density function:

$$P_{X,Y,Z}(\epsilon, \mathcal{M}) = \tag{5}$$
$$\frac{q^d}{(\sqrt{2\pi})^d \sqrt{\Sigma}} \exp\left[-\frac{(\{X,Y,Z\} - \mathcal{M})^2}{2\Sigma^2}\right],$$

where Σ represents the model of correlation between components and the amplitude of the noise.

Encoding the model $\mathcal{M}$ means to encode its parameters a_i for $i = 0 - 12$. This encoding will give us the complexity term. The overall optimum depends on both the degree of the polynomial and the precision with which the parameter values are encoded. Normally, we expect that the squared error to decrease as the order of the selected model is increasing, but the complexity of the model is increasing. We need to compute the maximum likelihood of data for the corresponding density function. The system formed by the following equations is obtained:

$$\frac{\partial \log P}{\partial a_i} = 0, \tag{6}$$

for $i = 0, \ldots, 12$. The NML density function is considered:

$$\hat{f}(y^n, \gamma) = \frac{f(y^n, \gamma, \hat{\beta}, \hat{\Sigma})}{\int\limits_{Y(\tau_0, R)} f(z^n, \gamma, \hat{\beta}, \hat{\Sigma}) dz^n}, \tag{7}$$

where $y = \{X, Y, Z\}$ and

$$Y(\Sigma_0, R) = \left\{ z^n : \hat{\Sigma}(z^n) \geq \Sigma_0, \hat{\beta}'(z^n)\Sigma\hat{\beta}(z^n) \leq R \right\}. \tag{8}$$

The selection criteria is given by:

$$\min_{\gamma \in \Omega} \left\{ (n-k) \ln \frac{n\hat{\Sigma}}{n-k} + k \ln(n\hat{R}) - \right. \tag{9}$$
$$\left. \ln \frac{n}{n-k} - (k+1) \ln k \right.$$

where k denotes the number of elements in γ and $\hat{R} = \frac{1}{n} \hat{\beta}'(s's)\hat{\beta}$ with $s = \{R, G, B\}$. The model that minimizes the above expression is the one that fits the data best. We will prefer the model class with the smallest stochastic complexity with respect to that model class.

4 Experimental Results

For the results presented in this extended abstract we have used the Macbeth ColorChecker chart which contains a number of 24 patches. 160 training samples and 80 test samples were captured using this chart with a two mega-pixels Nokia N90 camera phone. As suggested in [5], we turned the automatic white balance off, in order to obtain an effective camera characterization. The training samples were used to train the neural network and to determine the stochastic complexity using the MDL formalism for polynomial and Fourier sine models.

i	7	8	9	10	11
$\mathcal{M}_1$	-170.2	-211.4	-264.4	-260.3	-254.2
$\mathcal{M}_2$	-163.5	-232.8	-268.9	-263.4	-234.3

Table 1. Stochastic complexity of $\mathcal{M}_1$ and $\mathcal{M}_2$.

In Table 1 the results for the two models $\mathcal{M}_1$ and $\mathcal{M}_2$ are presented. On the first line the number of free parameters a_i, with i ranging from 7 to 11, is given. The results for values smaller than 7 or bigger than 11 are increasing, so we consider them not to be of interest. For space economy we decided not to display them. On the second line the results for the polynomial model $\mathcal{M}_1$, and on the third line the ones for the Fourier sine model $\mathcal{M}_2$, are given. From these results we conclude that for the both models the best model order is 9. The Fourier series model $\mathcal{M}_2$ has the smallest stochastic complexity, so it is the best one for the given data.

For the neural network model, we started with 3 input and output units, and 1 hidden unit. The neural network was trained by adding units in the hidden layer using a procedure with a MDL stopping criterion similar with the one presented in [6]. The resulted neural network contained about 26 hidden units. The presented results are average values from training the neural network models 10 times.

Model	$\mathcal{M}_1$	$\mathcal{M}_2$	ANN
Median CIELAB error	1.78	1.22	1.31
Maximum CIELAB error	35.4	32.8	44.2

Table 2. Performance of tested models.

The obtained models were then tested using the test samples. The color errors between measured and estimated tristimulus values were computed using the CIELAB color difference formula. The maximum and median CIELAB errors for the used 3 models are given in Table 2. We note that better performances are obtained by the models with a stronger nonlinear character.

5 Conclusions

In this paper, the MDL formalism has been used to study the low-end digital camera characterization of a Nokia imaging phone. Three different models have been considered: polynomial transforms, Fourier sine transforms, and neural networks. Better performances were obtained for the models with higher nonlinear characteristics. The abilities of camera characterization models based on Fourier sine and neural networks are slightly similar and better than those of the polynomial model.

References:

[1] Red Herring, "Photo phones hurting cameras," *http://www.redherring.com/*, August 2005.

[2] M.C. Stone, "Representing colors as three numbers [color graphics]," *IEEE Computer Graphics and Applications*, vol. 25, no. 4, pp. 78–85, July-August 2005.

[3] F. Galland, N. Bertaux, and P. Refregier, "Minimum description length synthetic aperture radar image segmentation," *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 995–1006, September 2003.

[4] S. Pateux, "Spatial segmentation of color images according to the mdl formalism," *Proceedings of the International Conference on Image Processing ICIP*, vol. 2, pp. 554–557, September 2000.

[5] V. Cheung, S. Westland, D. Connah, and C. Ripamonti, "A comparative study of the characterisation of colour cameras by means of neural networks and polynomial transforms," *Coloration Technology*, vol. 120, no. 1, pp. 19–25, 2004.

[6] Yu Ning Lai and Shiu Yin Yuen, "Successive-least-squares error algorithm on minimum description length neural networks for time series prediction," *Proceedings of the 17th International Conference on Pattern Recognition ICPR*, vol. 4, pp. 609–612, August 2004.

[7] T.L.V. Cheung and S. Westland, "Colour camera characterisation using artificial neural networks," *Proceedings of the 10th Color Imaging Conference*, pp. 117–120, 2002.

[8] Hong Guowei, Han Bing, and M.R. Luo, "Colorimetric characterisation of low-end digital camera and its application for on-screen texture visualisation," *Proceedings of the International Conference on Image Processing ICIP*, vol. 1, pp. 741–744, September 2000.

[9] J. Rissanen, "Strong optimality of the normalized ml models as universal codes and information in data," *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 1712–1717, June 2001.