On Spatial Quantization of Color Images*

Jens Ketterer, Jan Puzicha, Marcus Held,
Martin Fischer, Joachim M. Buhmann, and Dieter Fellner

Institut für Informatik III,
Rheinische Friedrich–Wilhelms–Universität
Römerstraße 164, D-53117 Bonn, Germany
email:{ketterer,jan,held,jb,dieter}@cs.uni-bonn.de
WWW:http://www-dbv.cs.uni-bonn.de

Abstract. Image quantization and dithering are fundamental image processing problems in computer vision and graphics. Both steps are generally performed sequentially and, in most cases, independent of each other. Color quantization with a pixel-wise defined distortion measure and the dithering process with its local neighborhood typically optimize different quality criteria or, frequently, follow a heuristic approach without reference to any quality measure.
In this paper we propose a new model to *simultaneously* quantize and dither color images. The method is based on a rigorous cost–function approach which optimizes a quality criterion derived from a simplified model of human perception. Optimizations are performed by an efficient multiscale procedure which substantially alleviates the computational burden.
The quality criterion and the optimization algorithms are evaluated on a representative set of artificial and real–world images thereby showing a significant image quality improvement over standard color reduction approaches.

1 Introduction

True color images typically contain up to 16 million different colors. One of the basic tasks of low level image processing consists of reducing the number of colors with minimal visual distortion. Such a coarse graining of colors is important for fast image manipulation on a reduced *color palette*, for computational efficiency in subsequent image processing and computer vision applications, and for image coding and transmission. Furthermore, many image display and printing devices provide only a limited number of colors. The representation problem for colors aggravates when many images are displayed simultaneously resulting in a palette size of 256 or even substantially less colors assigned to each image. Several techniques have been proposed for color reduction, most of which obey a two–step scheme:

* A more detailed report is found in [16]. This work has been supported by the German Research Foundation (DFG) under grant #BU 914/3–1, by the German Israel Foundation for Science and Research Development (GIF) under grant #1-0403-001.06/95 and by the Federal Ministry for Education, Science and Technology (BMBF #01 M 3021 A/4).

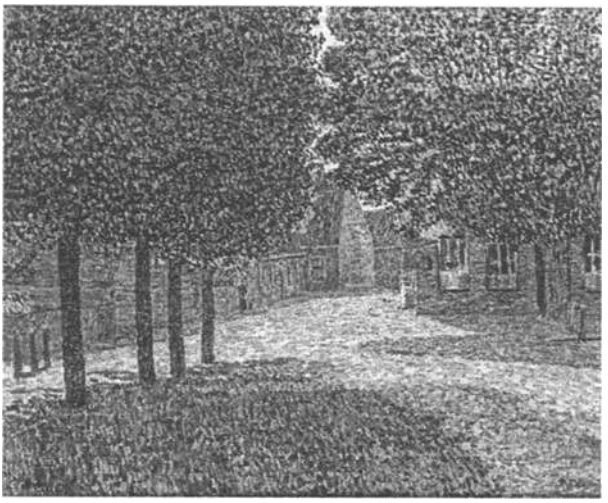

Fig. 1. Pointillism artwork: Paul Baum (1859–1928): Place à St. Anna, Hollande, 1905.

1. Initially, a color set is selected by minimizing a proper *pixel distortion* error. Examples are the popular median–cut quantizer [1] and the application of a variety of clustering methods like LBG [2] or agglomerative clustering [3]. In fact, any suitable clustering approach could be used, see [4] for an overview. It is a characteristic fact for all clustering approaches, that they neglect spatial, i. e. contextual, information.
2. Several types of degradation appear in the quantized image due to the limited number of colors. The most severe is the appearance of contouring artifacts in uniform regions. *Dithering* methods as a subsequent processing step provide a way to address this problem by exploiting a property of the human visual system. Human beings perceive high frequency spatial variations in color without preferred direction as a *uniform* color, which can be understood as an averaging superposition. Impressionistic painters from the French school of pointillism have exploited this effect in a spectacular way as illustrated in Fig. 1. Therefore, additional illusionary colors can be created by spatial mixing. In a common dithering technique called *error diffusion* the quantization error is spread to neighboring pixels, i.e. the distortion at neighboring pixels is biased in opposite direction. Several error diffusion filters have been proposed [5, 6].

Quantization and dithering are generally performed sequentially. Therefore, quantization and dithering usually optimize different quality criteria. While clustering distortion measures like the well–known K–means cost function are exclusively pixel–based, dithering techniques rely on spatial information and distribute the distortion error among neighboring pixels. Joint quantization and dithering approaches have been considered only occasionally on a heuristic ad–hoc basis [7].

In this paper we propose a rigorous cost–function based approach, which simultaneously performs quantization and dithering of color images in a joint

error minimizing algorithmic step, which we refer to as *spatial quantization*. The presented cost function extends the K–means criterion to a spatially weighted distortion measure. It, thereby, incorporates dithering techniques into the quantization cost–function. The strengths of both, quantization and dithering, are combined in a rigorous fashion leading to a significant improvement in image display quality.

From our point of view a successful optimization approach consists of two conceptually well–separated parts:

1. A good quality criterion, which appropriately *models* the information processing task: For color reduction the global minima of the cost function should correspond to psychophysically pleasing image reproductions.
2. An efficient optimization algorithm to minimize the proposed cost function.

We develop two closely related optimization algorithms for the spatial clustering criterion, (i) an extension of the **I**terative **C**onditional **M**ode (ICM) algorithm, which is similar in spirit to K–means clustering and (ii) a *deterministic annealing* variant along the lines of the maximum entropy quantization [8].

While ICM is fast, it is a purely local optimization procedure. **D**eterministic **A**nnealing (DA) as a homotopy method inspired by statistical physics is closely related to simulated annealing and has been empirically shown to be fast, yet to possess global optimization properties in the sense, that optimal or near optimal solutions are found and bad local minima are avoided. We extend the convergence proof for assignment problems given in [9] to the case of additional continuous parameters and derive efficient mean-field equations for spatial clustering.

Both ICM and DA are significantly accelerated by applying *multiscale optimization* techniques [10]. Multiscale optimization can be understood as the minimization of the original cost function over a properly reduced search space and has been shown to substantially accelerate the algorithms for other clustering cost functions [11]. For spatial quantization the corresponding cost functions on coarse image grids are derived.

In Sect. 2 we discuss color spaces and present the novel dithered color quantization cost function. Sect. 3 is dedicated to optimization methods. Multiscale expressions are derived and both DA and ICM are discussed in detail. Results are presented in Sect. 4 followed by a short conclusion.

2 Combining Quantization and Error Diffusion

2.1 Perception model

The spatial frequency response of the human vision system performs poorly at high frequencies due to the finite resolution of the human eye and the physical limitations of display devices. Thus, additional imaginary colors can be generated by *digital half-toning*. To simplify the model of imaginary colors the chosen color space should represent perceived superposition of colors as linear superpositions. A color space with this property is called a *uniform* or *linear color space*. The

commonly used RGB color space and its linear derivates do not define linear color spaces. In contrast, the CIE Lab color space [12] represents differences in color by the Euclidean metric in a psychophysically meaningful way, although it is well known to suffer from some minor defects [13]. To overcome this shortness, new non-Euclidean metrics were proposed. Despite these facts, in this paper RGB and CIE Lab color spaces with the Euclidean metric are used for computational simplicity and to provide a general interface for other color spaces.

Having defined a linear color space U, we propose to model human perception as a blurring operation, i.e. a convolution of the image with a Gaussian kernel to ensure locality. To obtain a formulation for a discrete image grid we introduce a neighborhood system $\mathcal{N}_i$ defined as the spatial support for a Gaussian kernel with standard deviation σ centered at pixel i. We define the weights w_{ij} for neighborhood pixels as

$$w_{ij} \sim \exp\left(-\frac{(i-j)^2}{\sigma^2}\right), \quad \sum_{j \in \mathcal{N}_i} w_{ij} = 1 \ . \tag{1}$$

The perceived color $\boldsymbol{c}_i \in U$ at location i for a given image is modeled as

$$\boldsymbol{c}_i(\boldsymbol{x}) = \sum_{j \in \mathcal{N}_i} w_{ij} \boldsymbol{x}_j \ , \tag{2}$$

where $\boldsymbol{x}_j \in U$ denotes the pixel value at location j.

2.2 Spatial Quantization Cost Function

To define the spatial quantization cost function we first introduce a set or palette of colors, which are denoted by a vector of prototype variables $\boldsymbol{Y} = (\boldsymbol{y}_\nu^t)_{\nu=1,\dots,K}$, $\boldsymbol{y}_\nu \in U \subset \mathbb{R}^3$, where U is again the linear color space defined above and the superscript t denotes the transpose of a vector. A quantization is then defined as an assignment of pixel colors $\boldsymbol{x}_i$ to prototypical colors $\boldsymbol{y}_\nu$, which is formalized by Boolean assignment variables $M_{i\nu} \in \{0,1\}$. $M_{i\nu} = 1(0)$ denotes that the image site $\boldsymbol{x}_i$ is (is not) quantized to color $\boldsymbol{y}_\nu$. All assignments are summarized in terms of a Boolean assignment matrix $\boldsymbol{M} \in \mathcal{M}$, where

$$\mathcal{M} = \left\{ \boldsymbol{M} \in \{0,1\}^{N \times K} : \sum_{\nu=1}^{K} M_{i\nu} = 1, \ 1 \leq i \leq N \right\} \ . \tag{3}$$

As a cost function for faithful color reproduction we employ the pixel-wise distance between the perceived image before and after quantization. For a linear color space the Euclidean norm is the natural choice yielding costs[1]

$$\mathcal{H}(\boldsymbol{M}, \boldsymbol{Y}) = \sum_{i=1}^{N} \left\| \left(\sum_{j \in \mathcal{N}_i} w_{ij} \boldsymbol{x}_j \right) - \left(\sum_{j \in \mathcal{N}_i} \sum_{\nu=1}^{K} M_{j\nu} w_{ij} \boldsymbol{y}_\nu \right) \right\|^2 \ . \tag{4}$$

[1] The classical K–means cost function is obtained for $w_{ij} = \delta_{ij}$ and can be understood as the special case of our model with a blur free perception model.

The task of spatial quantization is then defined as a search for a parameter set $(\boldsymbol{M}, \boldsymbol{Y})$ which minimizes (4) [2]. In contrast to the K–means cost function, which is linear with respect to discrete and continuous parameter set, (4) is quadratic. Standard quantization techniques are computationally efficient, whereas our color quantization approach achieves substantially better modeling quality at higher computational costs. The reader should be aware at this point that changes in the model of human perception only result in a different cost function, when the perception model is no longer linear w.r.t. the input image. Thus (4) covers a broad range of possible perception models.

The cost function (4) can be rewritten as a quadratic form in the assignment variables and in the continuous variables $\boldsymbol{Y}$. For this purpose, we introduce a new, enlarged neighborhood

$$\tilde{\mathcal{N}}_i = \{k : \exists j : j \in \mathcal{N}_i \wedge j \in \mathcal{N}_k\} \quad . \tag{5}$$

Abbreviate

$$b_{ij} = \sum_{\substack{k \in \mathcal{N}_i \\ \wedge k \in \mathcal{N}_j}} w_{ik} w_{jk}, \quad j \in \tilde{\mathcal{N}}_i, \quad \boldsymbol{a}_i = -2 \sum_{j \in \tilde{\mathcal{N}}_i} b_{ij} \boldsymbol{x}_j \tag{6}$$

and note that b_{ij} is translation and rotation invariant and depends only on $|i-j|$ for our specific perception model (1). An equivalent expression for (4) is given by

$$\mathcal{H}(\boldsymbol{M}, \boldsymbol{Y}) = \sum_{i=1}^{N} \sum_{j \in \tilde{\mathcal{N}}_i} \sum_{\nu=1}^{K} \sum_{\alpha=1}^{K} b_{ij} M_{j\nu} M_{i\alpha} \boldsymbol{y}_\nu^t \boldsymbol{y}_\alpha + \sum_{i=1}^{N} \sum_{\nu=1}^{K} M_{i\nu} \boldsymbol{a}_i^t \boldsymbol{y}_\nu \quad , \tag{7}$$

which makes the quadratic nature of $\mathcal{H}$ explicit. Constant terms without influence on either the assignments or the prototypes have been discarded.

3 Optimization for Spatial Quantization

3.1 Multiscale Optimization

The statistics of natural images support the assumption that colors are distributed homogeneously in images, i. e. pixels adjacent to each other contain with high probability similar colors. This fact can be exploited to significantly accelerate the optimization process by minimizing the criterion over a suitable nested sequence of subspaces in a coarse to fine manner. Each of these subspaces is spanned by a greatly reduced number of optimization variables. In contrast to most multi-resolution optimization schemes the identical cost function is optimized at all grids, solely the variable configuration space is reduced.

[2] It is also possible to (partially) fix a set of prototypes, e.g. a set of available colors, and to optimize the assignments of pixels to colors alone. Therefore, this cost function provides also a new method for dithering an image given a fixed color table.

This strategy is formalized by the concept of *multiscale optimization* [10] and it leads in essence to cost functions redefined on a coarsened version of the original image. Formally, we denote by $\mathcal{S}^0 = \mathcal{S}$ the original set of sites and we assume that a set of grid sites $\mathcal{S}^l = \{1, \ldots, N^l\}$ is given for each coarse grid level l. Throughout this section coarse/fine grid entities are denoted by upper/lower case letters. Define a coarsening map C_l on the sets of indices:

$$C_l : \mathcal{S}^l \to \mathcal{S}^{l+1}, \qquad i \mapsto I = C_l(i) \ , \tag{8}$$

where each fine grid point is linked to a single coarse grid point. Typically $\mathcal{S}^0 = \mathcal{S}$ corresponds to the set of pixel sites and $\mathcal{S}^{l+1}$ is obtained by sub-sampling $\mathcal{S}^l$ by a factor of 2 in each direction, i. e. 4 sites are combined into a single coarse site by the two-dimensional index operation $(I, J) = (\lfloor i/2 \rfloor, \lfloor j/2 \rfloor)$. Since this operation is a many-to-one map the inverse C_l^{-1} is a subset of the fine grid sites, $C_l^{-1}(I) \subset \mathcal{S}_l$. Multiscale optimization proceeds not by coarsening the image, but by *coarsening the variable space.* Each coarse grid is associated with a reduced set of optimization variables $\boldsymbol{M}^l \in \mathcal{M}^l$,

$$\mathcal{M}^l = \left\{ \left(M^l_{I\nu}\right)_{\substack{I=1,\ldots,N^l \\ \nu=1,\ldots,K}} : M^l_{I\nu} \in \{0,1\} \right\} \ . \tag{9}$$

Thus K Boolean variables $M^l_{I\nu}$ are attached to each grid point I denoting whether the set of respective pixels is assigned to color $\boldsymbol{y}_\nu$. Coarsened cost functions at level $l+1$ are defined recursively by proper restriction of the optimization space at level l:

$$\mathcal{H}^{l+1}(\boldsymbol{M}^{l+1} \in \mathcal{M}^{l+1}, \boldsymbol{Y}) := \mathcal{H}^l(\boldsymbol{M}^l \in \tilde{\mathcal{M}}^l : M^l_{i\nu} = M^{l+1}_{C_l(i)\nu}, \boldsymbol{Y}) \ , \tag{10}$$

where

$$\tilde{\mathcal{M}}^l = \left\{ \boldsymbol{M}^l \in \mathcal{M}^l : M^l_{i\nu} = M^l_{j\nu} \ \text{ for } \ C_l(i) = C_l(j) \right\} \tag{11}$$

denotes the subspace $\tilde{\mathcal{M}}^l \subset \mathcal{M}^l$ with identical assignments for sites with the same coarse grid point. Now introduce a coarse grid neighborhood by

$$\tilde{\mathcal{N}}^{l+1}_I = \left\{ J : \exists i \in C_l^{-1}(I), j \in C_l^{-1}(J) : j \in \tilde{\mathcal{N}}^l_i \right\} \ . \tag{12}$$

We recursively define

$$b^{l+1}_{IJ} = \sum_{i \in C_l^{-1}(I)} \sum_{\substack{j \in C_l^{-1}(J) \\ \wedge j \in \tilde{\mathcal{N}}_i}} b^l_{ij}, \qquad \boldsymbol{a}^{l+1}_I = \sum_{i \in C_l^{-1}(I)} \boldsymbol{a}^l_i \ . \tag{13}$$

For the spatial quantization cost function (7) the following coarse grid cost functions are obtained applying definition (10):

$$\mathcal{H}^l(\boldsymbol{M}^l, \boldsymbol{Y}) = \sum_{I=1}^{N^l} \sum_{J \in \tilde{\mathcal{N}}^l_I} \sum_{\nu=1}^{K} \sum_{\alpha=1}^{K} b^l_{IJ} M^l_{I\nu} M^l_{J\alpha} \boldsymbol{y}^t_\nu \boldsymbol{y}_\alpha + \sum_{I=1}^{N^l} \sum_{\nu=1}^{K} M^l_{I\nu} \left(\boldsymbol{a}^l_I\right)^t \boldsymbol{y}_\nu \ . \tag{14}$$

Note, that $\mathcal{H}^l$ has the same functional form as $\mathcal{H}^0 = \mathcal{H}$ and, therefore, an optimization algorithm developed for $\mathcal{H}$ is applicable to any coarse grid cost function $\mathcal{H}^l$ without change.

3.2 Deterministic Annealing

A common way to optimize mixed combinatorial optimization problems such as $\mathcal{H}(\boldsymbol{M}, \boldsymbol{Y})$ is by an alternating minimization scheme, i.e. to optimize with respect to the discrete parameters keeping the continuous parameters fixed and then to optimize the continuous parameters for a fixed discrete set. This twofold optimization is iterated until a predefined convergence criterion is fulfilled.

For fixed $\boldsymbol{Y}$ the cost function $\mathcal{H}(\boldsymbol{M}, \boldsymbol{Y}) = \mathcal{H}(\boldsymbol{M})$ expresses a purely combinatorial problem, which can be efficiently tackled by the class of *annealing methods*. The basic idea of annealing methods is to treat the unknown Boolean variables as random variables, to introduce a scale parameter T, often called the computational temperature, and to calculate equilibrium Gibbs averages of assignments $M_{i\nu}$. This estimate can be achieved either by Monte Carlo sampling or (at least approximately) by analytical methods. The temperature T is gradually lowered and for $T \to 0$ a solution of the combinatorial optimization problem is obtained. In Deterministic Annealing (DA) approaches [8, 14] the stochasticity is incorporated in a *probabilistic formulation* of the optimization problem and a *deterministic optimization* is performed over a *probabilistic state space* $\mathcal{Q}(\mathcal{M})$. The minimization is carried out over the space of factorial distributions

$$\mathcal{Q} = \left\{ Q(\boldsymbol{M}) := \prod_{i=1}^{N} \sum_{\nu=1}^{K} M_{i\nu} m_{i\nu},\ \forall \boldsymbol{M} \in \mathcal{M} \right\} \tag{15}$$

to guarantee computational tractability and efficiency. For factorial distributions random variables at different sites are independent, i.e. $\langle M_{i\nu} M_{j\alpha} \rangle_Q = \langle M_{i\nu} \rangle_Q \langle M_{j\alpha} \rangle_Q$ is valid for $i \neq j$. Furthermore, the distribution parameters $m_{i\nu}$ equal the variable expectations, $m_{i\nu} = \langle M_{i\nu} \rangle_Q$. The original cost function $\mathcal{H}$ over $\mathcal{M}$ is embedded in the family of smoothed cost functions over $\mathcal{Q}$

$$\mathcal{F}_T(Q) = \langle \mathcal{H} \rangle_Q - TS(Q) \ , \tag{16}$$

where S denotes the entropy. $\mathcal{F}_T$ is called *generalized free energy* referring to its usage in statistical physics and it is convex over $\mathcal{Q}$ for sufficiently large T. Thus $\mathcal{F}_T$ can be understood as a *coarse* version of $\mathcal{H}$ while T takes the role of a *scale parameter* in optimization space. Denote by $Q^{i\alpha}$ the matrix obtained by replacing the i–th row of $Q = (m_{i\nu})$ with the unit vector $\boldsymbol{e}_\alpha$. The fundamental relationship [9]

$$m_{i\nu} = \frac{\exp\left(-\langle \mathcal{H} \rangle_{Q^{i\nu}} / T\right)}{\sum_{\alpha=1}^{K} \exp\left(-\langle \mathcal{H} \rangle_{Q^{i\alpha}} / T\right)} \tag{17}$$

is valid for distributions Q minimizing (16). A transcendental system of equations is obtained for the $m_{i\nu}$ from (17), which is solved by an asynchronous, iterative

and convergent fixed-point relaxation. DA can be understood as a continuation method [14], since the (unique) minimum at high temperature is tracked while gradually lowering ('annealing') T. The local ICM–algorithm is obtained in the $T = 0$ limit, where $m_{i\nu} \in \{0, 1\}$ and $m_{i\nu} = M_{i\nu} = 1$, iff $\nu = \arg\max_\alpha \langle \mathcal{H} \rangle_{Q^{i\alpha}}$.

We now turn to the derivation of mean-field equations for the general spatial clustering cost function. Both ICM and DA are implemented using an efficient bookkeeping scheme. The so–called meanfields $\langle \mathcal{H} \rangle_{Q^{i\nu}}$ are given by

$$\langle \mathcal{H} \rangle_{Q^{i\nu}} = \boldsymbol{y}_\nu^t \left(\boldsymbol{p}_i + b_{ii} \boldsymbol{y}_\nu \right) \tag{18}$$

with the bookkeeping entities

$$\boldsymbol{p}_i = 2 \sum_{\substack{j \in \tilde{\mathcal{N}}_i \\ j \neq i}} b_{ij} \sum_{\alpha=1}^{K} m_{j\alpha} \boldsymbol{y}_\alpha + \boldsymbol{a}_i \ . \tag{19}$$

For fixed Boolean variables $\boldsymbol{M}$ the optimization of $\mathcal{H}(\boldsymbol{M}, \boldsymbol{Y}) = \mathcal{H}(\boldsymbol{Y})$ yields a simple matrix equation. Let $\boldsymbol{S} = (s_{\nu\alpha}) \in \mathbb{R}^K \times \mathbb{R}^K$ and $\boldsymbol{R} = (\boldsymbol{r}_\nu^t) \in \mathbb{R}^K \times \mathbb{R}^3$, where

$$s_{\nu\alpha} = \sum_{i=1}^{N} \sum_{j \in \tilde{\mathcal{N}}_i} b_{ij} m_{i\nu} m_{j\alpha}, \qquad \boldsymbol{r}_\nu = \sum_{i=1}^{N} m_{i\nu} \boldsymbol{a}_i \ , \tag{20}$$

then the generalized free energy (16) is equivalent to the quadratic form

$$\mathcal{F}_T(Q) \equiv \langle \mathcal{H}(\boldsymbol{Y}) \rangle_Q = \sum_{\nu=1}^{K} \sum_{\alpha=1}^{K} s_{\nu\alpha} \boldsymbol{y}_\alpha^t \boldsymbol{y}_\nu + \sum_{\nu=1}^{K} \boldsymbol{r}_\nu^t \boldsymbol{y}_\nu \ . \tag{21}$$

Therefore the optimal $\boldsymbol{Y}$ is given by

$$\boldsymbol{Y} = -(2\boldsymbol{S})^{-1} \boldsymbol{R} \ . \tag{22}$$

For fixed T the overall alternating minimization scheme can be proven to converge to a local minimum of (16). In the proof given in [15] for arbitrary assignment problems the free energy (16) plays the role of a Lyapunov function. Extending this proof to mixed combinatorial problems it is straight forward to show that (22) also decreases the free energy in each step, see [16] for details. For large T the inverse of $\boldsymbol{S}$ becomes singular and the system of linear equations is solved e.g. by a conjugate gradient method or the Moore–Penrose inverse to avoid numerical instabilities.

3.3 Multiscale Annealing

Algorithms like K–means or LBG efficiently minimize $\mathcal{H}$ by *splitting techniques* to obtain successive solutions for a growing number of clusters. For ICM we adopt an idea from K–means clustering by splitting clusters with high distortion. One of the key advantages of the DA approach is the *inherent splitting*

strategy. Clusters degenerate at high temperature and they successively split at *bifurcations* or *phase transitions* when T is lowered [8]. Therefore, at a specific temperature scale an easily measurable *effective number* K_T of clusters is visible. Since the number of effective data points available drastically reduces at coarser resolution levels, splitting strategy and coarse–to–fine optimization should be interleaved. Thus for a given resolution level l we anneal until K_T exceeds a certain maximal number of clusters $K^l_{\max}$ for some temperature T^*. After prolongation to level $(l-1)$ the DA optimization is continued at temperature T^*. This scheme has been introduced as *multiscale annealing* in [11], where also the question of choosing the maximal number of clusters for a given resolution has been addressed in a statistical learning theory context. We adopt this approach by choosing $K^l_{\max} \sim N^l/\log N^l$ and by determining the proportionality factor on an empirical basis. The complete algorithm is summarized in Algorithm I.

Algorithm I

INPUT $w_{ij}, \boldsymbol{x}_i, K$
INITIALIZE $m_{i\nu}^{l_{\max}}$ randomly, temperature $T = T_{\text{start}}$, $l = l_{\max}$;
COMPUTE $\boldsymbol{a}_i$, b_{ij} according to (6), $\boldsymbol{a}_I^l$, b_{IJ}^l according to (13).
WHILE $(T > T_{\text{FINAL}})$ and $(l \geq 0)$ (multiscale annealing)
    REPEAT
        REPEAT
            generate a permutation $\pi \in S_N$
            FOR $i = 1, \ldots, N^l$
                update all $m^l_{\pi(i)\nu}$ according to (17)
        UNTIL converged
        UPDATE $\boldsymbol{Y}$ according to (22)
    UNTIL converged
    IF $K_T > K^l_{\max}$ PROPAGATE $m^l_{I\nu}$ to $m^{l-1}_{i\nu}$, SET $l = l-1$
    ELSE SET $T = \eta \cdot T$, $0 < \eta < 1$
SET $M^l_{I\nu} = \delta_{\nu, \arg\max_\alpha m_{I\alpha}}$ (winner-takes-all rule)
WHILE $(l \geq 0)$ ($T = 0$ ICM-optimization)
    PROPAGATE $M^l_{I\nu}$ to $M^{l-1}_{i\nu}$, SET $l = l-1$
    REPEAT
        UPDATE $M^l_{i\nu}$ by ICM until converged
        UPDATE $\boldsymbol{Y}$ according to (22)
    UNTIL converged

4 Results

The proposed spatial quantization algorithms are compared with several standard color image quantization methods, which all employ both quantization and dithering. As median cut quantizer [1] in conjunction with the Floyd–Steinberg (FS) dithering algorithm ppmquant found in the PPM tools by Jef Poskanzer is used. The octree quantization [17] implementation is based upon C code published in Dr. Dobbs Journal. As the most simple alternative we have applied a

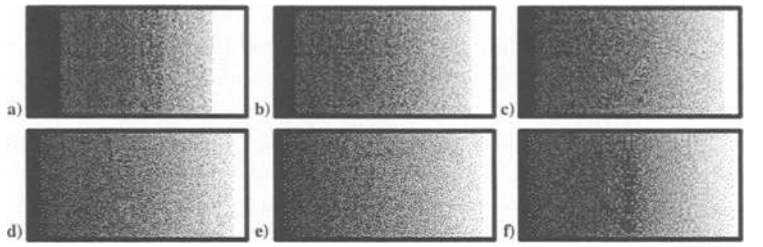

Fig. 2. Image quantization on a smooth transition of grey levels (256 different grey values): (a) to (e) spatial quantization with neighborhood sizes from 3×3 to 11×11 (from left to right), (f) Floyd–Steinberg (FS) dithering.

	Pool	Tea Pot	Mandrill	Rose	Golden Gate
Spatial quantization	0.29 (0.81) 72 s	0.39 (1.04) 115 s	1.54 (5.33) 213 s	0.97 (2.89) 19 s	0.92 (2.74) 174 s
median cut + FS	1.36 (1.95) 2 s	0.73 (1.27) 4 s	2.24 (4.36) 5 s	1.59 (2.95) 1 s	1.45 (2.66) 6 s
octree + FS	1.37 (2.36) 3 s	1.00 (1.86) 5 s	6.00 (8.76) 4 s	4.95 (6.67) <1 s	3.77 (5.13) 5 s
bit-cut + FS	9.79 (25.25) 1 s	18.98 (39.05) 1 s	27.1 (48.05) 1 s	22.02 (42.63) <1 s	28.45 (46.61) 1 s

Table 1. Comparison of different color image quantization methods for quantization to 64 colors. First row: Quality measured in terms of (4) on a scale $[0, 100]$ (average deviation). In brackets the quality measured by pixel-wise squared difference (K–means) is given. Second row: run–time in seconds (Pentium Pro 200MHz). ICM has been used for optimization in spatial quantization. Every quantization is performed in the RGB color space. The neighborhood size for our approach has been set to 5×5.

simple bit–cutting for quantization followed again by a Floyd–Steinberg dithering procedure [6]. A representative set of images has been chosen for comparison and evaluation, which are depicted in Fig. 4.

To examine the dithering properties of the novel cost function independently from the built–in quantization several runs on an artificial image with smooth transition of grey values as depicted in Fig. 2 have been carried out. To identify the role of the neighborhood size for the quality of our dithering approach is the main purpose of this experiment. The available two colors were fixed as black & white and only the assignments of pixels to the given color values were optimized.

It has to be noticed that the (subjective) dithering quality grows with neighborhood size. Starting with a neighborhood size of 5×5 a smooth transition between grey values is obtained, while the smaller neighborhood of 3×3 suffers from its limited variability to distribute black and white pixels and generates artificial 'edge' structures. In contrast, the Floyd–Steinberg algorithm introduces significant visual distortions by edge effects and over–regular patterns and it is not capable to generate a smooth transition of grey values. In Fig. 5 the full spatial quantization approach is compared to median cut / Floyd–Steinberg with respect to quality for different number of colors. The "Pool" image was selected for its large range of colors. Especially the billiard balls exhibit a smooth

	Pool	Tea Pot	Mandrill	Rose	Golden Gate
Multiscale optimization	0.62	0.83	2.38	1.71	1.64
	119 s	180 s	225 s	20 s	242 s
Single scale optimization	1.02	0.95	2.75	1.87	1.82
	259 s	373 s	666 s	65 s	429 s

Table 2. Comparison of multiscale and single scale methods. First row: Quality measured in terms of (4). Second row: run–time in seconds measured on a Pentium Pro 200MHz. The images were quantized to 16 colors with a neighborhood size of 7×7.

transition from dark to bright primary colors. It can be seen that the proposed algorithm is able to distribute the small number of available colors in a more efficient way than ppmquant. Notice especially the lack of any yellow color in the image quantized by ppmquant. This unsatisfactory behavior is caused by the large size of the green area in the original image. The median cut quantizer assigns too much resources, i. e. color prototypes, to green color values. This behavior is basically caused by the fact that median cut creates clusters with approximately equal size instead of rigorously optimizing a distortion measure. Depending on the number of desired colors only four to one prototypes are left for all other colors in the image. To represent the large range of leftover colors the center color was taken which is some greyish color[3].

In Fig. 6 the "Mandrill" image quantized to 8 colors is depicted. It is possible to reduce the number of colors from 171877 with only minor perceptive defects, since many illusionary colors are created by dithering. This result is illustrated by the magnification of the monkey's eye in Fig. 6 (b) and (c), which demonstrates that very different colors can be used to create a highly similar visual perception.

Fig. 7 studies the role of the chosen color space, in which the spatial quantization is performed. Quantization in CIE Lab space improves the rendering of bright colors by reducing the number of darker colors. At low luminance the human vision system looses its ability to distinguish chromaticity. For the "Tea Pot" image, reduced to 16 colors, just two (the RGB image reserves four) dark colors were needed to render the shadows, all other dark colors are imaginary. Instead more yellowish colors were allocated to reproduce all reflections and highlights correctly.

The quality and performance results for all images are summarized in Tab. 1. In absence of a better, psychophysically defined distortion measure the quality according to the K–means criterion and the novel spatial quantization cost function (4) are reported, although we are convinced that (4) better reflects visual distortion. As expected, according to its own cost function spatial quantization outperforms all other methods significantly. But even according to the K–means criterion it produces better results, which can be explained by the fact that the heuristic dithering procedures tend to increase the K–means distortion costs drastically.

The quality of spatial clustering is improved by better optimization techniques. In Fig. 3 quantization results of the image "Rose" to 4 colors are de-

[3] A more complete comparison with other color reduction schemes is found in [16].

picted. The costs obtained according to (4) by the deterministic annealing (DA) algorithm are lower than the cost of the ICM solution. The resulting images clearly indicate that better minimization techniques can result in an improved perception quality of the color reduced image. In the ICM solution some smaller details (especially in the lower right corner) vanish. For a more detailed discussion of the deterministic annealing approach for color quantization see [16].

Fig. 3 also indicates that color reduction down to 4 colors still results in recognizable images, which might be of interest for generating iconized images in multimedia applications. Also, some widespread operating systems restrict icons to resolutions of 32×32 with a maximum of 12 colors.

On the other hand it has to be stressed that the computational complexity of spatial quantization increases significantly in comparison to the other methods, see again Tab. 1. It is therefore of particular importance to design efficient optimization algorithms such as the multiscale approach. As seen in Tab. 2 multiscale optimization accelerates the optimization by a factor 2–5. Fig. 8 illustrates the typical progress in multiscale optimization. According to (13) colors in coarse images are replaced by local averaging, thus coarse grid colors correspond to perceived colors. At the same time the neighborhood smoothing kernel sharpens according to (13). Therefore, spatial optimization on coarser grids has a tendency to suppress dithering of colors. Thus a bias towards local minima with homogeneous colors is introduced. On the other hand it is well–known that multiscale coarsening leads to an *implicit smoothing* of the energy landscape [10] and therefore avoids bad local minima. This effect is confirmed by the results in Tab. 2.

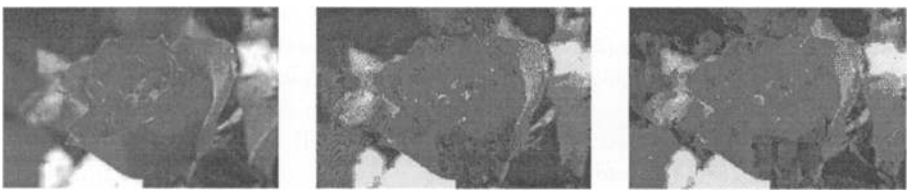

Fig. 3. Original image "Rose" (left), spatial quantization with ICM using 4 colors (center), spatial quantization with DA using 4 colors (right).

5 Conclusion

We have presented a novel approach to simultaneous color image quantization and dithering on the basis of a novel quality measure. This criterion incorporates error diffusion in the clustering cost function and it is motivated by a model of the human visual system. An efficient deterministic annealing algorithm has been developed and an extension to multiscale optimization has been derived to accelerate the algorithmic process.

Using a rigorous cost function approach it is possible to evaluate the modeling quality independently of the optimization procedure. The algorithm has been shown to yield a significant improvement in quality compared to alternative approaches on a large set of images. The results are especially impressive

for small color palettes, where standard quantization schemes completely fail. Moreover, it is possible to incorporate a (partially) predefined color palette in the optimization process.

In the derivation of the cost function a simplified model of human perception has been used. More elaborated models with emphasis on color constancy and feature preservation will be discussed elsewhere.

References

1. P. Heckbert, "Color image quantization for frame buffer displays," *Computer Graphics*, vol. 16, no. 3, pp. 297–307, 1982.
2. G. Braudaway, "A procedure for optimum choice of a small number of colors from a large palette for color imaging," in *Electron Imaging'87*, 1987.
3. Z. Xiang and G. Joy, "Color image quantization by agglomerative clustering," *IEEE Computer Graphics and Applications*, vol. 14, no. 3, pp. 44–48, 1994.
4. A. Jain and R. Dubes, *Algorithms for Clustering Data*. Prentice Hall, 1988.
5. R. Ulichney, "Dithering with blue noise," *Proceedings of the IEEE*, vol. 76, pp. 56–79, 1988.
6. R. Floyd and L. Steinberg, "An adaptive algorithm for spatial greyscale," in *Proc. SID, Vol. 17, No. 2*, Wiley, 1976.
7. L. Akarûn, D. Özdemir, and Ö. Yalcun, "Joint quantization and dithering of color images," in *Proceedings of the International Conference on Image Processing (ICIP'96)*, pp. 557–560, 1996.
8. K. Rose, E. Gurewitz, and G. Fox, "A deterministic annealing approach to clustering," *Pattern Recognition Letters*, vol. 11, pp. 589–594, 1990.
9. T. Hofmann, J. Puzicha, and J. Buhmann, "Deterministic annealing for unsupervised texture segmentation," in *Proc. of the EMMCVPR'97*, LNCS 1223, pp. 213–228, 1997.
10. F. Heitz, P. Perez, and P. Bouthemy, "Multiscale minimization of global energy functions in some visual recovery problems," *CVGIP: Image Understanding*, vol. 59, no. 1, pp. 125–134, 1994.
11. J. Puzicha and J. Buhmann, "Multiscale annealing for real–time unsupervised texture segmentation," Tech. Rep. IAI–97–4, Institut für Informatik III (a short version appeared in: Proc. ICCV'98, pp. 267–273), 1997.
12. C. I. de L'Eclairage, "Colorimetry." CIE Pub. 15.2 2nd ed., 1986.
13. D. Alman, "Industrial color difference evaluation," *Color Res. Appl. 18 137-139*, 1993.
14. G. Bilbro, W. Snyder, S. Garnier, and J. Gault, "Mean field annealing: A formalism for constructing GNC–like algorithms," *IEEE Transactions on Neural Networks*, vol. 3, no. 1, 1992.
15. T. Hofmann, J. Puzicha, and J. Buhmann, "A deterministic annealing framework for textured image segmentation," Tech. Rep. IAI-TR-96-2, Institut für Informatik III, 1996.
16. J. Puzicha, M. Held, J. Ketterer, J. Buhmann, and D. Fellner, "On spatial quantization of color images," Tech. Rep. IAI–TR–98–1, Department for Computer Science III, University Bonn, 1998.
17. M. Gervauz and W. Purgathofer, "A simple method for color quantization: Octree quantization," in *Graphic Gems*, pp. 287–293, Academic Press, New York, 1990.

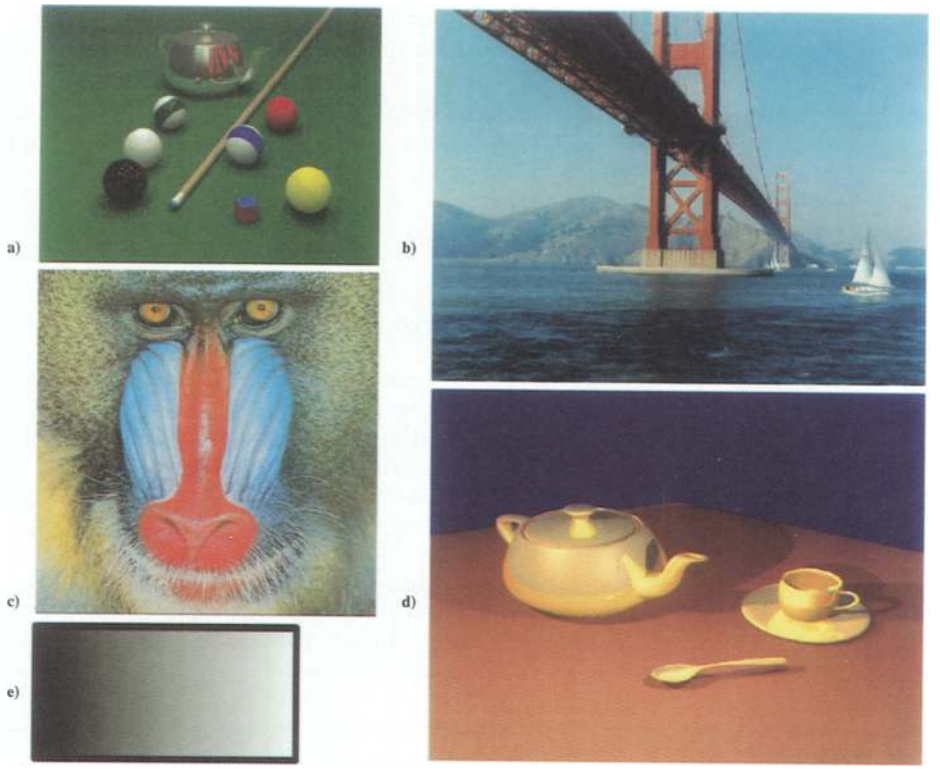

Fig. 4. Original images used for the different experiments: (a) "Pool", (b) "Golden Gate", (c) "Mandrill", (d) "Tea Pot" and (e) grey wedge.

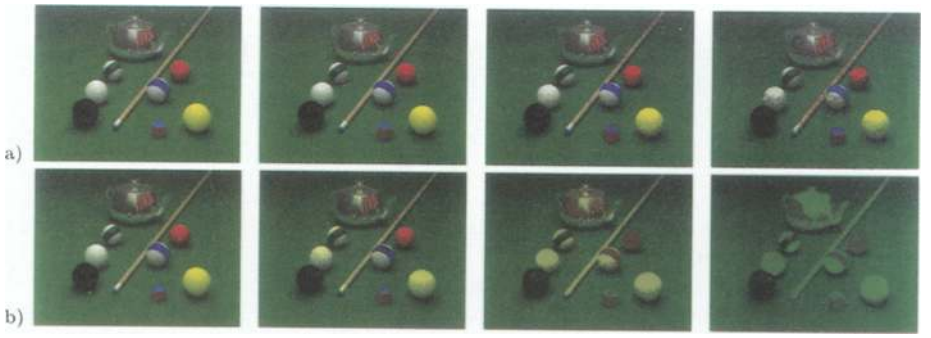

Fig. 5. Image quantization with different numbers of colors (from left to right: 256, 64, 32 and 16 colors): (a) spatial image quantization, (b) median cut quantization and Floyd–Steinberg (FS) dithering.

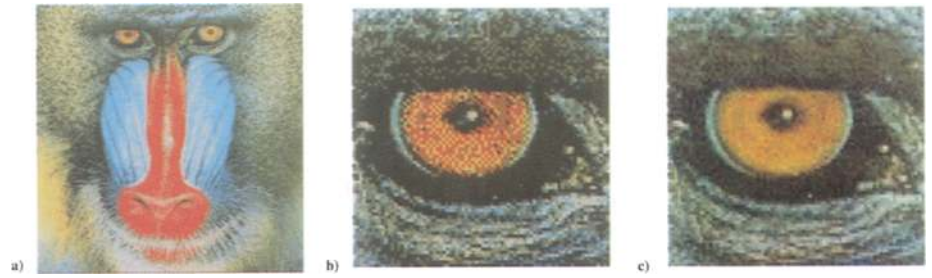

Fig. 6. Spatial quantization of the image "Mandrill": (a) Image with 8 colors, (b) detail of 8 color image and (c) same sub image from the original image with 171877 colors.

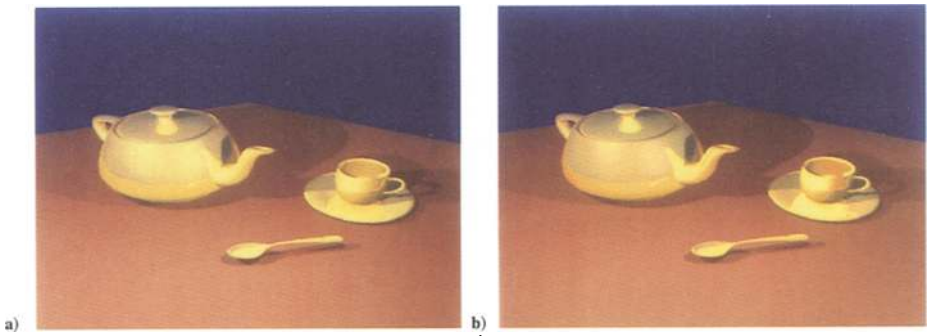

Fig. 7. Spatial image quantization using 16 colors in different color spaces ("Tea Pot" image): (a) RGB color space and (b) CIE Lab color space.

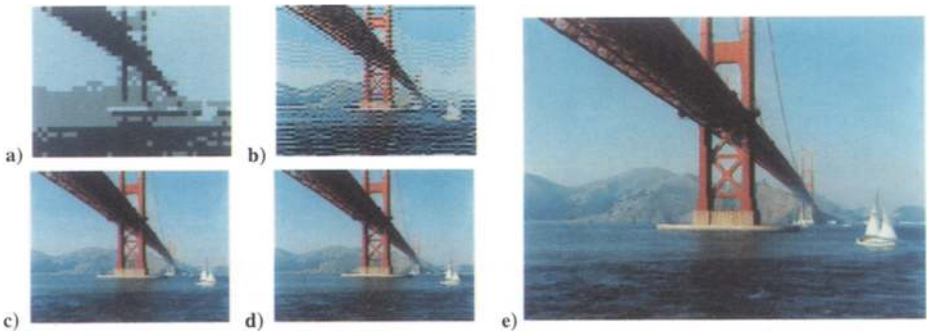

Fig. 8. Multi-scale optimization of the "Golden Gate". (a) to (d) intermediate coarse scale results ((a) 16 colors and (b) to (d) 32 colors), (e) resulting image (32 colors).