CENTER FOR
MACHINE PERCEPTION

CZECH TECHNICAL
UNIVERSITY

RESEARCH REPORT

ISSN 1213-2365

Joint Image GMM and Shading MAP Estimation

(Version 0.8)

Alexander Shekhovtsov
Václav Hlaváč

shekhovt@cmp.felk.cvut.cz hlavac@fel.cvut.cz

Czech Technical University in Prague

Report number K333–35/10,
CTU–CMP–2010–03

January 29, 2010

This work was supported by projects EU ICT-215078 DIPLECS, and CZ 1M0567 CAK.

Research Reports of CMP, Czech Technical University in Prague, No. 3, 2010

Published by

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Abstract

We consider a simple statistical model of the image, in which the image is represented as a sum of two parts: one part is explained by an i.i.d. color Gaussian mixture and the other part is a (piecewise-)smoothly varying grayscale shading function. The smoothness is ensured by a quadratic (Tikhonov) or total variation regularization. We derive an EM algorithm to estimate simultaneously the parameters of the mixture model and the shading. Our algorithms solve for shading and mean parameters of the mixture model jointly for both kinds of the regularization.

1 Introduction

This work proposes a novel model to estimate shading in natural images. We assume some areas of the image are brighter or darker because of the illumination/shadows. We refer to this effect as shading. The shading must be simple a priori – either smooth or piecewise-smooth function over the image domain. The image with the shading removed is then explained by a Gaussian mixture model (GMM) with unknown parameters. We pose the estimation problem as maximizing the image likelihood jointly in shading and GMM parameters. This maximum a posteriori problem (MAP) corresponds to the optimization problem, where smoothness of the shading is imposed via regularization.

Expectation Maximization. Estimating GMM parameters is commonly done by the EM algorithm. Suppose, however, that there are other unknowns to be estimated (shading). Having the GMM parameters fixed, estimating the best shading also could be done by some standard method (e.g. gradient descend). Especially, if the shading is discretized, one could solve for optimal shading using discrete optimization methods. Under pairwise convex regularization this subproblem can be solved exactly. It is then very easy to design an algorithm alternating between the two estimates in order to archive the best joint estimate of GMM and shading. However, this algorithm would be unstable: each new iteration, depending on the slight adjustment of the shading, the GMM estimate may end up in a different local optimum.

Contribution. We showed that in the considered model it is possible to optimize the likelihood w.r.t. shading and mean parameters of the GMM jointly, which makes the algorithms less prone to getting stuck in local minima. We derived a simple algorithm for the case of the quadratic regularization on the shading, which monotonously improves the objective. For the case of Total Variation (TV) regularization, we derived algorithms which follow existing TV methods, but estimate GMM mean parameters jointly with the shading and put these algorithms in the EM loop.

Related Work. This work was mostly inspired by [11], where the model included segmentation and shading. The problem of recovering shading is posed there in two forms: MAP and Bayesian estimation with additive quadratic risk function. They considered grayscale images and the non-parametric color model. However within the learning/inference framework, color images and GMM appearance model can be handled similarly, which is mentioned in the further work [6]. In the case of MAP estimate, the estimate of the shading is alternated with the estimate of appearance parameters, which is inferior to our method. In the case of Bayesian estimate, shading and GMM mixture hidden variables are sampled jointly.

A lot of previous work, e.g. [7], [5], exploit illumination-invariant characteristics of the color image such as chromaticity. The shading could be recovered from edges present in the image but absent in the illumination-invariant representation. Such edges are considered to be produced by shadows and they are integrated to reconstruct the shading. In contrast, our model can estimate soft shadows and remains meaningful also for grayscale images.

A variational formulation of the problem was proposed in [8]. However, it seems to be too simplistic, and the results shown in further work [4] and also in [12] are not convincing.

Our problem of decomposing the image into shading and a general color mixture components is closely related to several other decompositions. Image denoising can be viewed as decomposition into signal and noise parts. A Gaussian noise model would correspond to the special case of a single-component GMM. Structure-texture decomposition [9], [1] seeks for piecewise smooth structure and possibly largely oscillating texture components. For grayscale images, the shading in our model corresponds to the structure in [1] (and the same regularization is imposed). The GMM log likelihood term can be viewed as a parametrized regularization on the texture component.

Discrete and Continuous methods. Model [11] is based on the discrete formulation where values of the shading are discretized. In this work, we take advantage of the continuous formulation. For the quadratic regularization, we can differentiate the objective and solve for shading and GMM means from linear equations. In this case, it does not matter whether we consider image domain as continuous or discrete, the derivation would be the same. However, for TV regularization we use certain integral identities (following the prior work) to derive the algorithm, which would not be possible with the discretized domain. Hence we keep everything continuous and work with the calculus of variations.

Notation. Let $\Omega \subset \mathbb{R}^2$ be a continuous image domain. For a function $u\colon \Omega \to \mathbb{R}^p$, its value in a point $s \in \Omega$ is denoted as u_s and belongs to $\mathbb{R}^p$. Euclidean norm is denoted by $||\cdot||$. For a function $u\colon \Omega \to \mathbb{R}^2$, the point-wise norm $|u|\colon \Omega \to \mathbb{R}$ is defined as $|u|_s = ||u_s||$.

2 Model

2.1 Additive Shading

Let $I\colon \Omega \to \mathbb{R}^3$ be the RGB color image. Let $h\colon \Omega \to \mathbb{R}$ be unknown shading (lighting) for the image I. We define the likelihood of observing the image I given the shading h as

$$p(I|h;\theta) = \exp\Big\{\int_\Omega \log p(I_s|h_s;\theta)\mathrm{d}s\Big\}, \tag{1}$$

where $p(I_s|h_s;\theta)$ is the likelihood of observing color I_s given the shading h_s in the point s. In the case of discrete domain and pixel colors i.i.d. from $p(I_s|h_s;\theta)$, their joint probability would have been equal to

$$\prod_s p(I_s|h_s;\theta) = \exp\sum_s \log p(I_s|h_s;\theta). \tag{2}$$

The model (1) is a continuous analogue of (2).

In the additive shading model, we define "unshaded" image U as

$$U(h)_s = I_s - h_s\mathbf{1}_3, \tag{3}$$

where $\mathbf{1}_3 = (1\,1\,1)^\mathsf{T}$. Hence the shading h is the amount of white light added to the unshaded image to produce the observed image. Whereas this formula does not correspond to a physical law (in particular, unshaded intensities may become negative), it should be understood as a local approximation to the effect we want to model.

The unshaded image color in a point s is explained by a Gaussian mixture model:

$$p(I_s|h_s;\theta) = \sum_k \pi_k G_{\Sigma_k}(I_s - h_s\mathbf{1}_3 - \mu_k), \tag{4}$$

where $G_{\Sigma_k}(x-\mu_k)$ is a Gaussian pdf with mean μ_k and covariance Σ_k, and $\theta = (\pi_k,\, \mu_k,\, \Sigma_k)_{k=1}^K$ is the vector of all parameters of the mixture model, where $\pi_k \in [0,\ 1]$, $\mu_k \in \mathbb{R}^3$, $\Sigma_k \in \mathbb{R}^{3\times 3}$.

The main assumption about the shading is that it should be a smoothly varying function. We consider two priors (regularizations) enforcing smoothness:

$$p(h) \propto \exp\Big\{-\frac{\gamma}{2}\int_\Omega |\nabla h|^2 \mathrm{d}s\Big\} \tag{5}$$

and

$$p(h) \propto \exp\Big\{-\frac{\gamma}{2}\int_\Omega |\nabla h| \mathrm{d}s\Big\}. \tag{6}$$

Both functions assign low probability to those shadings which have strong gradients. The first choice allows easier optimization, whereas the second one (Total Variation) better fits to model piecewise smooth functions, which are more appropriate to describe shading of 3D scenes. Note that parameter γ has a different meaning for the two cases and it depends on the scale of h.

Our goal is to estimate shading h and the unknown parameters θ. We choose the joint MAP estimate $(h^{\mathrm{MAP}}, \theta^{\mathrm{MAP}})$, maximizing the joint likelihood, $p(I,h;\theta) = p(I|h;\theta)p(h)$. We cal this a MAP estimate by the following argument. Let θ be a random variable with a uniform prior[1], then maximizing joint likelihood $p(I,h,\theta) \propto p(I,h;\theta)$ is equivalent to maximizing the conditional likelihood $p(h,\theta|I) = p(I,h,\theta)/p(I)$. Arguably, one might prefer to have a maximum likelihood estimate for θ, $\theta^{\mathrm{ML}} = \arg\max p(I;\theta) = \arg\max \int_h p(I,h;\theta)$. However, this problem is more difficult and is not addressed here. Taking the logarithm, the MAP problem can be written as maximizing $E(h,\theta) =$

$$\int_\Omega \log \sum_k \pi_k G_{\Sigma_k}(U(h)_s - \mu_k)\mathrm{d}s - \frac{\gamma}{2}\int_\Omega |\nabla h|^\rho \mathrm{d}s. \tag{7}$$

We refer to case $\rho = 2$ as quadratic regularization model, and case $\rho = 1$ as TV model.

Remarks. In the special case, $K = 1$, $\rho = 1$, with fixed Σ_1, the model becomes efficiently a TV denoising model. Suppose, we are working with a grayscale image, then the mean parameter μ_1 of the single Gaussian component can be excluded from the model. This is because the shading already includes an arbitrary constant offset not penalized by the regularization. This case is well-studied, so in our experiments we have chosen images, such that GMM does not degenerate to a single Gaussian.

[1] This would be an *improper* prior. However, a proper prior, such as normal distribution with very large variance would efficiently give the same result.

3 Optimization

3.1 Quadratic Regularization

Consider maximizing (7) with ρ=2. To derive the EM algorithm, we introduce numbers $\alpha_{k|s} > 0$, $s \in \Omega$, $k = 1 \dots K$ such that $\sum_k \alpha_{k|s} = 1$ for all s. It follows from the generalized inequality of arithmetic and geometric means for positive numbers $q(s,k)$ that

$$\log \sum_k q(s,k) \geq \sum_k \alpha_{k|s} \log \frac{q(s,k)}{\alpha_{k|s}}, \tag{8}$$

where we let $q(s,k) = \pi_k G_{\Sigma_k}(U(h)_s - \mu_k)$. Then

$$E(h,\theta) \geq \int_\Omega \Big(\sum_k \alpha_{k|s} \log \pi_k G_{\Sigma_k}(U(h)_s - \mu_k) - \sum_k \alpha_{k|s} \log \alpha_{k|s} \Big) \mathrm{d}s - \frac{\gamma}{2} \int_\Omega |\nabla h|^\rho \mathrm{d}s = E(h,\theta,\alpha). \tag{9}$$

(We refer to [10] regarding this interpretation of the EM algorithm.) In (9) and further on we will use the convention that E with different arguments will denote different functionals. When α is fixed, it is easy to differentiate and optimize $E(h,\theta,\alpha)$ in (h,θ) (M-step) and when h and θ are fixed, the inequality in (9) can be tightened by maximizing $E(h,\theta,\alpha)$ over α subject to the above constraints (E-step). Thus the problem reads

$$\max_{h,\theta,\alpha} E(h,\theta,\alpha). \tag{10}$$

The E-step is the same as for the usual Gaussian mixture model:

$$\alpha_{k|s} = \frac{\pi_k G_{\Sigma_k}(U(h)_s - \mu_k)}{\sum_k \pi_k G_{\Sigma_k}(U(h)_s - \mu_k)}. \tag{11}$$

When α is fixed, optimizing in π_k subject to the constraint $\sum_k \pi_k = 1$ gives the update

$$\pi_k = \int_\Omega \alpha_{k|s} \mathrm{d}s \Big/ \int_\Omega 1 \mathrm{d}s, \tag{12}$$

where we used that $\sum_k \alpha_{k|s} = 1$ for all s. To optimize in the other variables, recall that

$$\log G_{\Sigma_k}(U(h)_s - \mu_k) = -\frac{3}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_k| - \frac{1}{2} \big(U(h)_s - \mu_k\big)^\mathsf{T} \Sigma_k^{-1} \big(U(h)_s - \mu_k\big). \tag{13}$$

We can optimize w.r.t. shading h and means μ simultaneously. The corresponding first order necessary optimality conditions are:

$$\begin{aligned} 0 &= \nabla_h E \\ 0 &= \nabla_\mu E, \end{aligned} \tag{14}$$

where $\nabla_h E$, $\nabla_\mu E$ are Frechet derivatives. Standard derivation (Appendix A) gives

$$0 = \sum_k \alpha_{k|s} \mathbf{1}_3^\mathsf{T} \Sigma_k^{-1} (I_s - h_s \mathbf{1}_3 - \mu_k) + \gamma \Delta h_s \quad \forall s \tag{15}$$

$$\mathbf{0}_3 = \int_\Omega \alpha_{k|s} \Sigma_k^{-1} (I_s - h_s \mathbf{1}_3 - \mu_k) \mathrm{d}s \quad \forall k. \tag{16}$$

The system (15)-(16) is a system of linear equations in h, μ. While the equations are differential in h, in the discretization they become just a sparse linear system and can be solved numerically with standard methods.

Optimizing w.r.t. covariance matrices Σ_k under fixed remaining variables gives the update

$$\Sigma_k = \frac{\int_\Omega \alpha_{k|s} \big(I_s - h_s \mathbf{1}_3 - \mu_k\big)\big(I_s - h_s \mathbf{1}_3 - \mu_k\big)^\mathsf{T} \mathrm{d}s}{\int_\Omega \alpha_{k|s} \mathrm{d}s}. \tag{17}$$

To prevent appearance of degenerate components we further make sure the smallest eigenvalue of Σ_k is not smaller than some σ_0. It is achieved by projecting Σ_k on the constraint $\min \mathrm{eig}(\Sigma_k) \geq \sigma_0$. Let $\Sigma_k = U \mathrm{diag}(\lambda) U^\mathsf{T}$, then the projection is given by

$$\tilde{\Sigma}_k = U \mathrm{diag}(\max(\lambda, \sigma_0)) U^\mathsf{T}. \tag{18}$$

Discretization. The discretization is built as follows. Domain Ω is quantized into 2D unit grid. Symbols $\int_\Omega \mathrm{d}s$ are to be replaced with $\sum_{s \in \Omega}$. The first derivative is approximated as forward difference: $\frac{\partial h}{\partial x} \approx h_{s+(1,0)} - h_s$. The second derivative is approximated as $\frac{\partial^2 h}{\partial x^2} \approx h_{s-(1,0)} + h_{s+(1,0)} - 2h_s$. The exact choice of the discretization and boundary conditions is more important for the TV model, in which case we followed recommendations of [13].

Let us summarize the algorithm for the total squared variation.

Algorithm 1

Input: image I, initial GMM $(\pi_k, \mu_k, \Sigma_k \,|\, k = 1 \dots K)$
Initialize $h = 0$
Repeat
  Update α via (11)
  Update π via (12)
  Update h, μ via solving linear system (15)(16)
  Update Σ_k via (17), (18).

The variables and updates should be understood as discretized. The algorithm is performing alternating maximization of the objective $E(h, \theta, \alpha)$ and hence each step is guaranteed to make a monotonous improvement to the objective.

3.2 Total Variation

Consider maximizing (7) with ρ=1. We first proceed as in the previous case. We introduce expectation variables α and inequality (9) holds. Optimization w.r.t. variables α, π, Σ is performend by the update equations (11), (12), (17). We then focus on optimizing (7) w.r.t. shading h and means μ. The derivative of the TV regularization $\int_\Omega |(\nabla h)_s| \mathrm{d}s$ is the non-linear expression (Appendix A)

$$-\nabla \cdot \frac{\nabla h}{|\nabla h|}. \tag{19}$$

It is not straightforward how to perform optimization w.r.t. h or (h, μ) jointly. We tested two algorithms described in [13]. Our derivation introduces optimization w.r.t. additional parameters

μ into these algorithms. The first one, referred to as primal fixed point iteration [14] is obtained as follows.

Primal TV Update. Let h_0 be the current estimate of the optimal h, then (19) is approximated with

$$-\nabla \cdot \frac{\nabla h}{|\nabla h_0| + \varepsilon} = -\nabla \cdot (a \nabla h), \tag{20}$$

which is linear in h and approximates (19) when h is close to h_0 and ε is small. The system of equations in (h, μ) becomes

$$\begin{cases} 0 = \sum_k \alpha_{k|s} \mathbf{1}_3^\mathsf{T} \Sigma_k^{-1} (I_s - h_s \mathbf{1}_3 - \mu_k) + \frac{\gamma}{2} (\nabla \cdot (a \nabla h))_s \\ \text{equation (16)} \end{cases} \tag{21}$$

The method is then summarized by the following algorithm:

Primal TV improve

Input: current h, a, α, $\{\pi_k, \Sigma_k \mid k = 1 \dots K\}$
Repeat
Update u and μ via solving (21).
Update a:

$$a_s = \frac{1}{|(\nabla h)_s| + \varepsilon} \tag{22}$$

Output: h, a, μ.

Should this algorithm converge, h will be optimal to the energy with the regularization $-\frac{\gamma}{2} \frac{|\nabla h|^2}{|\nabla h| + \varepsilon}$, which is close to the desired $-\frac{\gamma}{2} |\nabla h|$. However the algorithm is not monotonous and is not guaranteed to converge. The full optimization of (9) is summarized as Alg. 2P in the end of the section.

Dual TV Update. The dual algorithm [2, 3, 13] is as follows. The difficult term $|\nabla h|$ is represented as the maximization problem:

$$||(\nabla h)_s|| = \max_{\substack{u_s \\ \text{s.t.} \quad ||u_s|| \le 1}} u_s \cdot (\nabla h)_s. \tag{23}$$

Introducing additional unknown $u \colon \Omega \to \mathbb{R}^2$, the maximization of (9) can be written as

$$\max_{h, \theta, \alpha} \min_{\substack{u \\ |u| \le \mathbf{1}}} E(h, \theta, \alpha, u), \tag{24}$$

with

$$E(h, \theta, \alpha, u) = \int_\Omega \Big(\dots \Big) \mathrm{d}s - \frac{\gamma}{2} \int_\Omega u_s \cdot (\nabla h)_s \mathrm{d}s, \tag{25}$$

where the bracket $(\dots)$ is the same as in (9). Noting that $E(h, \theta, \alpha, u)$ is convex in u, concave in (h, μ) and the constraints on u are convex, we can swap in (24) max over (h, μ) and min over u:

$$(24) = \max_{\alpha, \pi, \Sigma} \min_{\substack{u \\ |u| \le \mathbf{1}}} \max_{h, \mu} E(h, (\pi, \mu, \Sigma), \alpha, u). \tag{26}$$

Our goal is to solve explicitly the inner maximization problem and to represent (26) as

$$\max_{\alpha,\pi,\Sigma} \min_{\substack{u \\ |u|\leq \mathbf{1}}} E((\pi,\Sigma),\alpha,u). \tag{27}$$

The necessary optimality condition w.r.t. h, $\nabla_h E = 0$, expresses as

$$0 = \sum_k \alpha_{k|s} \mathbf{1}_3^\mathsf{T} \Sigma_k^{-1}(I_s - h_s \mathbf{1}_3 - \mu_k) + \frac{\gamma}{2}(\nabla \cdot u)_s \quad \forall s, \tag{28}$$

where we used that $\int_\Omega u \cdot \nabla h \mathrm{d}s = -\int_\Omega (\nabla \cdot u) h \mathrm{d}s$ (see Appendix A).

Optimality conditions w.r.t. μ are given by (16). Equations (28) and (16) are linear in (h,μ) and, as we shall see, it is not difficult to derive a closed-form solution. Consider the minimization problem in (27). Optimality conditions w.r.t. u are complicated by the constraint $|u| \leq \mathbf{1}$. Following [13] we will use the projected gradient descend on u. From (25), we have for the gradient

$$(\nabla_u E)_s = -\tfrac{\gamma}{2}(\nabla h)_s \quad \forall s. \tag{29}$$

We will solve for (h,μ) from (28),(16) and by substituting into (29) find the expression for $\nabla_u E$ as a function of u. From (28), we have

$$h_s = \frac{\sum_k \alpha_{k|s} \mathbf{1}_3^\mathsf{T} \Sigma_k^{-1}(I_s - \mu_k) + \frac{\gamma}{2}(\nabla \cdot u)_s}{\sum_k \alpha_{k|s} \mathbf{1}_3^\mathsf{T} \Sigma_k^{-1} \mathbf{1}_3}. \tag{30}$$

Let us abbreviate this expression as

$$h_s = J_s - \sum_k c_{k,s}^\mathsf{T} \mu_k + \lambda_s (\nabla \cdot u)_s, \tag{31}$$

where $J\colon \Omega \to \mathbb{R}$, $c_k\colon \Omega \to \mathbb{R}^3$, $\lambda\colon \Omega \to \mathbb{R}$ are appropriate functions of s as for (31) to match (30). Substituting h into optimality condition w.r.t. μ (16) we have for all k:

$$\mathbf{0}_3 = \int_\Omega \alpha_{k|s} \Sigma_k^{-1} \Big(I_s - \mathbf{1}_3 \big(J_s + \lambda_s (\nabla \cdot u)_s \big) \Big) \mathrm{d}s + \int_\Omega \alpha_{k|s} \Sigma_k^{-1} \Big(\mathbf{1}_3 \sum_{k'} c_{s,k'}^\mathsf{T} \mu_{k'} - \mu_k \Big) \mathrm{d}s \tag{32}$$

This is a system of linear equations in μ, let us write it in the form

$$A\mu = b(u), \tag{33}$$

where μ is the concatenated vector $\{\mu_k^l \,|\, k = 1 \dots K,\ l = 1 \dots 3\}$, and $A \in \mathbb{R}^{3K \times 3K}$ and $b(u) \in \mathbb{R}^{3K}$ are given by

$$\begin{aligned} A_{(l,k),(l',k')} &= \textstyle\int_\Omega \alpha_{k|s} \Big[(\Sigma_k^{-1} \mathbf{1}_3)_l c_{s,k'}^{l'} - (\Sigma_k^{-1})_{l,l'} \delta_{\{k=k'\}} \Big] \mathrm{d}s \\ b(u)_{(l,k)} &= -\textstyle\int_\Omega \alpha_{k|s} \Big[\Sigma_k^{-1} \Big(I_s - \mathbf{1}_3 \big(J_s + \lambda_s (\nabla \cdot u)_s \big) \Big) \Big]_l \mathrm{d}s \end{aligned} \tag{34}$$

Substituting $\mu = A^{-1} b(u)$ into (31) and then (31) into (29) we obtain for the gradient w.r.t. u: $(\nabla_u E)_s =$

$$-\frac{\gamma}{2} \Big[(\nabla J)_s - \sum_k (\nabla c_k)_s^\mathsf{T} \big(A^{-1} b(u) \big)_k + (\nabla(\lambda(\nabla \cdot u)))_s \Big]. \tag{35}$$

The projected gradient descend on u is then as follows:

Dual TV improve

Input: current u, α, $\{\pi_k, \Sigma_k \,|\, k = 1 \dots K\}$

Compute J, c, λ as for (31):

$$\begin{aligned}
&\sigma_k = \Sigma_k^{-1} \mathbf{1}_3 \\
&\varsigma_k = \mathbf{1}_3^{\mathsf{T}} \sigma_k \\
&\lambda_s = \tfrac{\gamma}{2} \big(\textstyle\sum_k \alpha_{k|s} \varsigma_k \big)^{-1} \\
&c_{k,s}^l = \tfrac{2}{\gamma} \lambda_s \alpha_{k|s} \sigma_k^l \\
&J_s = \textstyle\sum_{k,l} c_{k,s}^l I_s^l
\end{aligned} \tag{36}$$

Compute A:

$$A_{(k,l)(k',l')} = \frac{2}{\gamma} \sigma_k^l \sigma_{k'}^{l'} \int_\Omega \alpha_{k|s} \lambda_s \alpha_{k'|s} \mathrm{d}s - \delta_{\{k=k'\}} (\Sigma_k^{-1})_{l,l'} \int_\Omega \alpha_{k|s} \mathrm{d}s \tag{37}$$

Compute b_0:

$$b_{0_{k,l}} = - \textstyle\int_\Omega \alpha_{k|s} \Sigma_k^{-1} \big(I_s - \mathbf{1}_3 J_s \big) \mathrm{d}s \tag{38}$$

Repeat

Compute

$$b(u)_{k,l} = b_{0_{k,l}} + \sigma_k^l \textstyle\int_\Omega \alpha_{k|s} \lambda_s (\nabla \cdot u)_s \mathrm{d}s \tag{39}$$

Set $\mu = A^{-1} b(u)$

Compute $g = \nabla_u E$ according to (35)

Set $u = u - \tau g$

Project u: $u_s = \frac{u_s}{\max(1, ||u_s||)} \; \forall s$

Output: u, μ and h recovered via (31).

This algorithm is guaranteed to converge to a global optimum of the inner minimization problem in (27). Let us give now the corresponding algorithm for the full problem (24) using primal or dual updates.

Algorithm 2P (resp. **D**)

Input: Image I, initial GMM $\{\pi_k, \mu_k, \Sigma_k \,|\, k = 1 \dots K\}$

Repeat

Update α as (11)

Update π_k as (12)

Update h, u, μ by **Primal TV improve**
(resp. **Dual TV improve**)

Update Σ_k by (17)

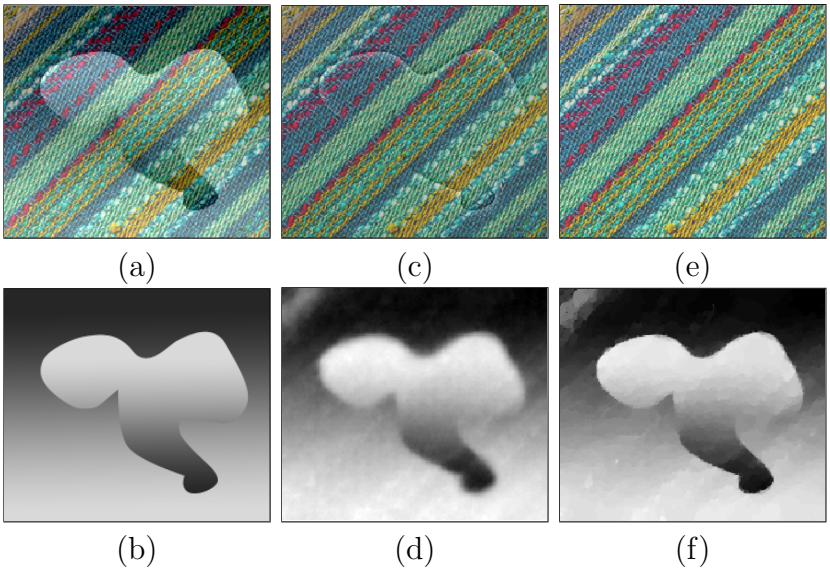

Figure 1: (a) A texture image with the added shading. (b) the shading. Results of the quadratic model showing estimated shading in (d) and reconstructed unshaded image in (c). Results of the TV model optimized with primal updates are similar to (c), (d). Results for the dual-based TV method initialized with (d) are shown in (e), (f).

Remarks. All the considered algorithms are at best to output a local maximum, because GMM estimation is a non-convex problem by itself. Alg. 1 is monotonous and is guaranteed to converge. Alg. 2P is not monotonous and is not guaranteed to converge. It is very sensitive to the value of ε: it affects the approximation of TV regularization and convergence of the algorithm, but also it affects which local optimum of the joint GMM shading problem is found.

Alg. 2D only becomes monotonous when the inner minimization by "Dual TV improve" is solved to the optimality. This inner minimization however, has linear convergence and may require long time to converge to a sufficient accuracy. We got successful results with Alg. 2D when it was starting from a good initial estimate by Alg. 1 or Alg. 2P. When an estimate of the shading h is given, the initial estimate for u can be obtained as

$$u_s = \begin{cases} \frac{(\nabla h)_s}{||(\nabla h)_s||}, & ||(\nabla h)_s|| > 0 \\ \mathbf{0}_2, & ||(\nabla h)_s|| = 0. \end{cases} \tag{40}$$

This expression make sure that $|u| \leq \mathbf{1}$ and that $u \cdot \nabla h = |\nabla h|$.

4 Experiments

We present a few encouraging results. We note, however, that the algorithms are prune to local optima and are sensitive to the parameters. We set the number of Gaussian components, K, to 10. It turned out that parameter σ_0 should be set to a rather high value, we used $\sigma_0 = 0.1$. It makes Gaussian components smoother so there is a smooth transition in expectations α between

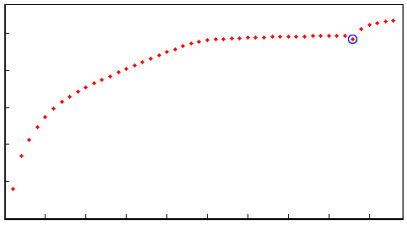

Figure 2: Progress of Alg. 1 on the texture unshading instance and progress of Alg. 2D continuing on the instance (starting on from the blue circle). x-axis show iterations and y-axis show the EM objective (9).

the components and therefore the full EM procedure is less likely to get stuck in a suboptimal point. This suggests that we could as well have used fixed $\Sigma_k = \sigma_0 I_3$. However, the covariances estimated in the experiments often exceeded σ_0, so we consider them still a useful part of the model.

Clearly, the strength of the regularization γ is a meaningful parameter to adjust for each particular use case. It should correspond to the scale and smoothness of the shading to be estimated. We adjusted this parameter for each experiment and each model $\rho = 1, 2$ individually as to obtain visually the best results.

Texture unshading. As a simple test we prepared the following artificial example. We took a real world texture and added to it a known piecewise-continuous smooth shading (Fig. 1 a,b). This experiment is similar to [11, fig.1], only we took a texture instead of i.i.d. Gaussian noise. The colorful texture allows to test a general case where Gaussian mixture model does not degenerate to a single Gaussian. Fig. 1 shows the results: as expected, quadratic model gives larger error around discontinuities of the shading, and the error is removed by the TV model.

Collor Illusion. The second experiment is on the famous color illusion picture shown in Fig. 3 a. Square B in the picture is percepted as white, however it is absolutely identical in color to square A, which is percepted as black. The challenge is whether we can undo this effect by estimating the proper shading. Fig. 3 b-g shows the results of the proposed methods. Inspection of the colors shows that color difference between white squares is much smaller than between A and B in all three results.

Some more results are shown in Fig. 4, the image is taken from [12].

5 Conclusion

We presented a simple model incorporating image shading with quadratic and total variation regularizations. The model with TV regularization seems to perform better when there are sharp boundaries of the objects or sharp shadows (the shading is piecewise-smooth). However, we faced a difficulty in the optimization: the best algorithm, Alg. 2D, needs to solve an inner minimization problem, which takes lots of iterations to converge. We know how to maximize the objective w.r.t. one group of variables (α, π, Σ) and minimize it w.r.t. other (u). We would be interested to find a fixed point in both of the groups (a saddle point problem). However, when the inner minimization problem is not solved to the optimality and only few iterations are taken, the

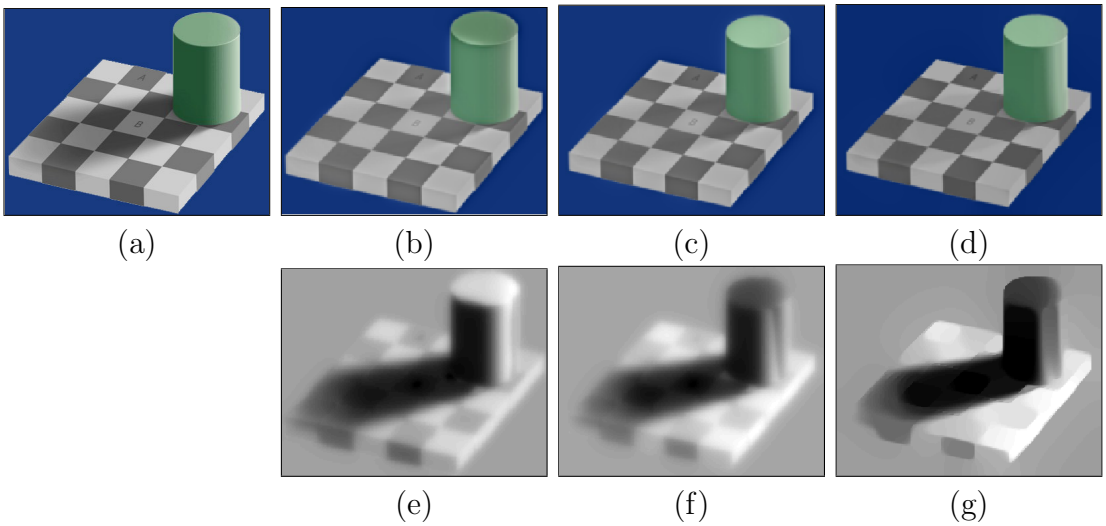

Figure 3: (a) Input image. (b), (c) results of the quadratic model (Alg. 1). (d), (e) Results of the TV model (Alg. 2P). (f), (g) results of the TV model (Alg. 2D) initialized with (e).

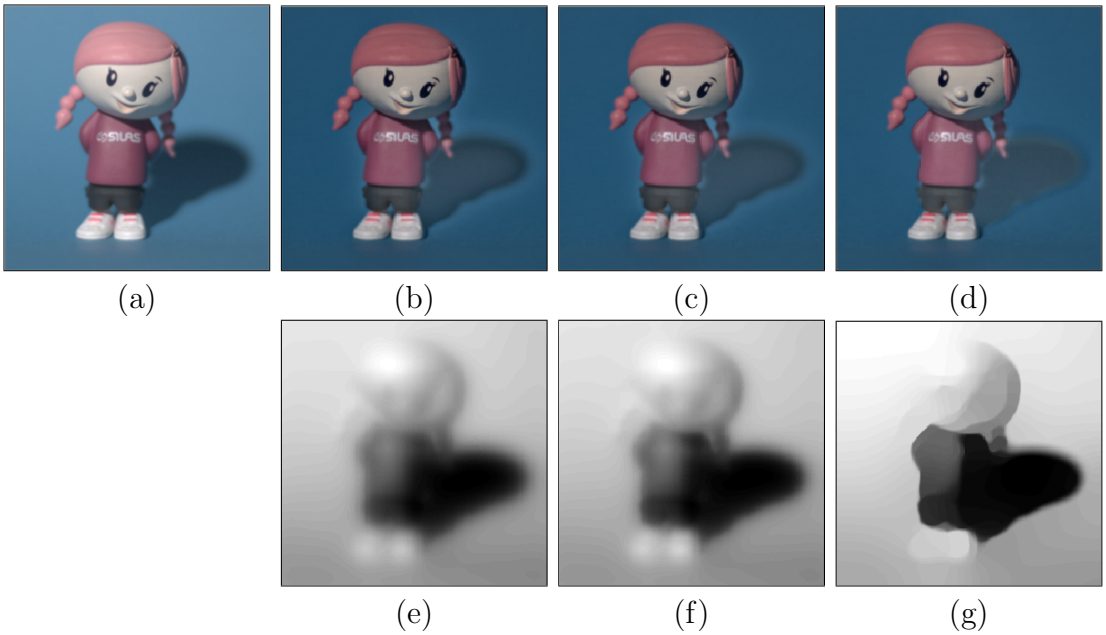

Figure 4: (a) Input image. (b), (c) results of the quadratic model (Alg. 1). (d), (e) Results of the TV model (Alg. 2P). (f), (g) results of the TV model (Alg. 2D) initialized with (e).

algorithm may oscillate. In the preliminary experiments, we were not able to obtain good results with this saddle point version. This seem to be a generic problem for parameter estimation within dual algorithms.

Let us also note that in order to recover better shading, a segmentation should be introduced into the model, like in [11]. Different objects in the image could then have different GMM models

and shadings, and the sharp boundary would be ensured by the segmentation. We expect that given the correct segmentation, shadings for the individual segments would be more accurate. For this purpose, the quadratic model may turn out more appropriate.

It is also questionable, whether the additive model is an appropriate approximation for shading estimation. For example, in Fig. 4f it is seen, that the background and the shadow are different in color, and this difference cannot be fully described by the white light.

Appendix A: Vector and Variation Calculus

Let $\Omega \subset \mathbb{R}^2$ be a compact set. Let $\mathcal{X}$ be the space of continuously differentiable functions $f\colon \Omega \to \mathbb{R}$ such that $f(\partial\Omega) = \mathbf{0}$ and f and ∇f are square integrable. The inner product of $f \in \mathcal{X}$ and $g \in \mathcal{X}$ is the integral

$$\langle f, g\rangle = \int_\Omega f_s g_s \mathrm{d}s. \tag{41}$$

A vector $f \in \mathcal{X}^p$ will denote a multi-valued function such that $f_s \in \mathbb{R}^p$.

Vector Calculus. Let $f \in \mathcal{X}^2$, $g \in \mathcal{X}$. *Divergence* $(\nabla \cdot f)$ is defined by $\nabla \cdot f = \frac{\partial f^1}{\partial x} + \frac{\partial f^2}{\partial y}$. *Laplace operator* Δ is defined by $\Delta g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}$. The following properties hold:

$$\nabla \cdot (gV) = (\nabla g) \cdot V + g \nabla \cdot V \tag{42}$$

$$\Delta g = \nabla \cdot (\nabla g) \tag{43}$$

Divergence theorem. Let $F \in C^1(\Omega)^2$ be a continuously differentiable vector field.
Then $\int_\Omega (\nabla \cdot F)_s \mathrm{d}s = -\int_{\partial\Omega} F \cdot n \,\mathrm{d}l$, where n is unit boundary normal vector.
By letting $F = fg$, $f \in C^1(\Omega)^2$, $g \in \mathcal{X}$ we get the following corollary:

$$\int_\Omega f_s \cdot (\nabla g)_s \mathrm{d}s + \int_\Omega (\nabla \cdot f)_s g_s \mathrm{d}s = -\int_{\partial\Omega} g f \cdot n \,\mathrm{d}l = 0, \tag{44}$$

where the last equality holds because $g(\partial\Omega) = \mathbf{0}$.

Frechet Derivative. Let $E\colon \mathcal{X} \to \mathbb{R}$. The Frechet derivative of E, denoted ∇E, is a function $\Omega \to \mathbb{R}$ satisfying the relation

$$E(f + \varepsilon\xi) = E(f) + \varepsilon\langle \nabla E, \xi\rangle + O(\varepsilon^2) \quad \forall \xi \in \mathcal{X} \tag{45}$$

Let $E(f) = \int_\Omega L(f_s, (\nabla f)_s)\mathrm{d}s$, where $L \in \mathcal{C}^2(\mathbb{R}^2)$. Then,

$$\nabla_f E = L_1' - \nabla \cdot L_2', \tag{46}$$

where L_1' is the partial derivative of L w.r.t. first argument and $L_2' = \begin{pmatrix} \frac{\partial L}{\partial f_x'} \\ \frac{\partial L}{\partial f_y'} \end{pmatrix}$ is partial derivative of L w.r.t. second argument.

Proof.

$$
\begin{aligned}
E(f+\varepsilon\xi) &= \int_\Omega L(f_s+\varepsilon\xi_s, [\nabla(f+\varepsilon\xi)]_s)\mathrm{d}s = \int_\Omega L(f_s) + \varepsilon\big[(L_1')_s\xi_s + (L_2')_s\cdot(\nabla\xi)_s\big] + O(\varepsilon^2)\mathrm{d}s \\
&= \int_\Omega L(f_s)\mathrm{d}s + \varepsilon\langle L_1', \xi\rangle + \varepsilon\int_\Omega (L_2')_s\cdot(\nabla\xi)_s\mathrm{d}s + O(\varepsilon^2) \\
&= \int_\Omega L(f_s)\mathrm{d}s + \varepsilon\langle L_1', \xi\rangle - \varepsilon\langle \nabla\cdot L_2', \xi\rangle + O(\varepsilon^2),
\end{aligned}
\tag{47}
$$

where we applied identity (44) with $f = L_2'$ and $g = \xi$. □

Let us derive Frechet derivatieves of several standard functionals.

$$
\frac{1}{2}\nabla\int_\Omega ||(\nabla h)_s||^2\mathrm{d}s = -\Delta h. \tag{48}
$$

$$
\nabla\int_\Omega ||(\nabla h)_s||\mathrm{d}s = \nabla\cdot\frac{\nabla h}{|\nabla h|}. \tag{49}
$$

$$
\frac{1}{2}\nabla\int_\Omega a_s||(\nabla h)_s||^2\mathrm{d}s = -a\Delta h - (\nabla a)\cdot(\nabla h). \tag{50}
$$

Proof.

(48): Let $L(h_s, (\nabla h)_s) = \frac{1}{2}||(\nabla h)_s||^2 = \frac{1}{2}((h_x')_s^2 + (h_y')_s^2)$. We have $(L_2')_s = \begin{pmatrix}(h_x')_s\\(h_y')_s\end{pmatrix}$, and by (46) it is $\nabla_h E = -\nabla\cdot\nabla h = -\Delta h$.

(49): Let $L(h_s, (\nabla h)_s) = ||(\nabla h)_s|| = \sqrt{(h_x')_s^2 + (h_y')_s^2}$. We have $(L_2')_s = \frac{-1}{2||(\nabla h)_s||}\begin{pmatrix}2(h_x')_s\\2(h_y')_s\end{pmatrix}$. By (46) it is $\nabla_h E = \nabla\cdot\frac{\nabla h}{|\nabla h|}$.

(50): Let $L(h_s, (\nabla h)_s) = \frac{1}{2}a_s||(\nabla h)_s||^2$. We have $L_2' = a\nabla h$, and hence $\nabla_h E = -\nabla\cdot(a\nabla h) = -a\nabla\cdot(\nabla h) - (\nabla a)\cdot(\nabla h) = -a\Delta h - (\nabla a)\cdot(\nabla h)$.

□

Acknowlegement

Authors would like to thank the reviewers and M. Navara for their comments.

References

[1] J.F. Aujol, G. Gilboa, T. Chan, and S.J. Osher. Structure-texture image decomposition: Modeling, algorithms, and parameter selection. *IJCV*, 67(1):111–136, April 2006. 2

[2] J. Carter. *Dual Methods for Total Variation-Based Image Restoration.* PhD thesis, UCLA, Los Angeles, CA, USA, 2001. 6

[3] A. Chambolle. Total variation minimization and a class of binary mrf models. In *EMMCVPR*, pages 136–152, 2005. 6

[4] M. Elad, R. Kimmel, D. Shaked, and R. Keshet. Reduced complexity retinex algorithm via the variational approach. *JVCIR*, 14(4):369–388, December 2003. 2

[5] G.D. Finlayson, S.D. Hordley, C. Lu, and M.S. Drew. On the removal of shadows from images. *PAMI*, 28(1):59–68, January 2006. 2

[6] Boris Flach and Dmitrij Schlesinger. Combining shape priors and mrf-segmentation. In *SSPR & SPR*, pages 177–186. Springer-Verlag, 2008. 1

[7] Brian V. Funt, Mark S. Drew, and Michael Brockington. Recovering shading from color images. In *ECCV*, pages 124–132. Springer-Verlag, 1992. 2

[8] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel. A variational framework for retinex. *IJCV*, 52(1):7–23, April 2003. 2

[9] Yves Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*. American Mathematical Society, Boston, MA, USA, 2001. 2

[10] T. Minka. Expectation-maximization as lower bound maximization, 1998. 4

[11] D. Schlesinger and B. Flach. A probabilistic segmentation scheme. In Gerhard Rigoll, editor, *Pattern Recognition*, volume 5096 of *LNCS*, pages 183–192. Springer, 2008. 1, 2, 10, 11

[12] H. Takahashi, T. Saito, and T. Komatsu. Variational retinex algorithm with its application to a high-quality chroma key. In *ICIP*, pages 977–980, 2006. 2, 10

[13] M. Unger, T. Pock, and H. Bischof. Interactive globally optimal image segmentation. Technical Report ICG-TR-08/02, Inst. for Computer Graphics and Vision Graz University of Technology, Austria, July 2008. 5, 6, 7

[14] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput*, 17:227–238. 5