Illumination Estimation and Cast Shadow Detection through a Higher-order Graphical Model

Alexandros Panagopoulos[1], Chaohui Wang[2,3], Dimitris Samaras[1] and Nikos Paragios[2,3]
[1]Image Analysis Lab, Computer Science Dept., Stony Brook University, NY, USA
[2]Laboratoire MAS, École Centrale Paris, Châtenay-Malabry, France
[3]Equipe GALEN, INRIA Saclay - Île-de-France, Orsay, France

Abstract

In this paper, we propose a novel framework to jointly recover the illumination environment and an estimate of the cast shadows in a scene from a single image, given coarse 3D geometry. We describe a higher-order Markov Random Field (MRF) illumination model, which combines low-level shadow evidence with high-level prior knowledge for the joint estimation of cast shadows and the illumination environment. First, a rough illumination estimate and the structure of the graphical model in the illumination space is determined through a voting procedure. Then, a higher order approach is considered where illumination sources are coupled with the observed image and the latent variables corresponding to the shadow detection. We examine two inference methods in order to effectively minimize the MRF energy of our model. Experimental evaluation shows that our approach is robust to rough knowledge of geometry and reflectance and inaccurate initial shadow estimates. We demonstrate the power of our MRF illumination model on various datasets and show that we can estimate the illumination in images of objects belonging to the same class using the same coarse 3D model to represent all instances of the class.

1. Introduction

Image formation is a function of three components: the 3D geometry of the scene, the reflectance properties of the present surfaces, and the distribution of lights. Much work has been done in estimating one or two of these components, assuming that the rest are known [18, 21, 23, 25, 26]. Illumination estimation methods often assume known geometry that is combined with strong assumptions about reflectance. In this work, we describe a method that relaxes these assumptions, based on the information contained in cast shadows. Cast shadows as a cue are relatively stable in the presence of large inaccuracies in knowledge of geometry and reflectance, compared to shading or specularities.

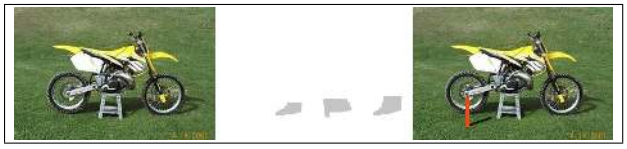

Figure 1. Our approach: from left to right, the original image; our shadow estimate; a sun dial rendered with the estimated illumination from our algorithm and overlayed on the image

In the computer vision community, there has been much research in extracting illumination from shading, specular reflection or shadows of objects. In [26], a small number of light source directions is detected using critical points, and [25] extends it to an image of an arbitrary object with known shape. In [23], a method is proposed for estimating the illumination distribution of a real scene from shadows, assuming known geometry illuminated by infinitely distant light sources, casting shadows onto a planar lambertian surface. In [7] illumination and reflectance are simultaneously estimated without the distant illumination assumption. In [27], a unified framework is proposed to estimate both distant and point light sources.

Prior art on illumination estimation using shadows cast on textured surfaces is limited. In [23], an extra image is necessary to deal with texture. In [18], a method is proposed that integrates multiple cues from shading, shadow, and specular reflections. [10] uses regularization by correlation to estimate illumination from shadows when texture is present, but requires extra user-specified information and assumes lambertian surface reflectance and known geometry. Recently, [19] proposed a method able to deal with inaccurate geometry and texture, but the shadow detection results when texture is present are limited. [14] proposed an approach that combines cues from the sky, cast shadows on the ground and surface brightness to estimate illumination of outdoor scenes with the sun as the single light source. Their method makes strong assumptions and is only applicable to daytime outdoor scenes.

There are several challenges when simultaneously estimating shadows and illumination. Solving the low level component that is the detection of cast shadows is not straightforward. The illumination estimation itself is also problematic due to the fact that measurements at the image level correspond to cumulative effects of all light sources leading to a very complex formulation. Shadow detection, in the absence of illumination estimation or knowledge of 3D geometry is a well studied problem. [22] uses invariant color features to segment cast shadows in still or moving images. In [3, 4], a set of illumination invariant features is proposed to detect and remove shadows from a single image, making several assumptions about the lights and the camera. Recently, [28] combined a number of cues in a complex method to recognize shadows in monochromatic images, while in [15], a learning approach is proposed to detect shadows in consumer-grade photographs, focusing on shadows on the ground.

While shadow detection can work at the image level, illumination estimation necessitates assumptions about geometry. In this paper, we propose a novel framework to recover the illumination environment of a scene and a rough cast shadow estimate from a single observed image, given coarse 3D geometry. Our main goal is to relax the necessary geometry assumptions so that simplistic approximations such as bounding boxes are enough to estimate illumination. Such approximate geometric information could be derived as part of more general scene understanding techniques, while enabling illumination estimation to be incorporated in the scene understanding loop; the obtained illumination information could be a crucial contextual prior in addressing various other scene understanding questions.

Graphical models can efficiently incorporate different cues within a unified framework [24]. To deal with the joint illumination and shadow estimation problem robustly in a flexible and extensible framework, we formulate it as an MRF model. All latent variables can then be simultaneously inferred by minimizing the MRF energy. To the best of our knowledge, this is the first time that scene photometry is addressed using an MRF model.

The MRF model we propose captures the interaction between geometry and light sources and combines it with image evidence of cast shadows, for joint estimation of cast shadows and illumination. The problem of shadow detection is well-posed in terms of the graph topology (graph nodes correspond to image pixels). On the other hand, illumination estimation implies a potential dependence between each pixel and all nodes representing the light sources, corresponding to higher-order cliques in the graph. At the same time, the number of light sources is unknown, resulting in unknown MRF topology, and the search space is continuous, complicating the use of discrete methods. The problem of inference in the presence of higher-order cliques has been given a lot of attention recently [9, 13]. Furthermore, we are able to reduce the search space and identify the MRF topology through an initial illumination estimate obtained using a voting algorithm. We then describe two methods to perform inference on this MRF model in the presence of higher-order cliques. We make the following *assumptions* (common in illumination modeling): the coarse 3D geometry is known, the illumination environment can be approximated by a set of distant light sources, and the reflectance of surfaces is roughly lambertian. For the extraction of shadows we utilize a recently proposed image cue [20]. It should be noted, however, that the proposed MRF model is flexible with respect to the shadow cues.

We evaluate our method on a set of images captured in a controlled environment, as well as on a set of car images from Flickr and images from the Motorbikes class of Caltech 101 [17]. Quantitative results are obtained on a synthetic dataset. Our results show that our method is robust enough to be able to use geometry consisting of bounding boxes or a common rough 3D model for a whole class of objects, while it can also be applied to scenes where some of our assumptions are violated.

This paper is organized as follows: Sec. 2 introduces the problem; Sec. 3 describes the MRF model to jointly estimate the shadows and illumination, while in Sec. 4 we discuss the inference process. Experimental results are presented in Sec. 5. Sec. 6 concludes the paper.

2. Problem Description

A commonly used set of assumptions, which we will use here, is that the surfaces in the scene exhibit lambertian reflectance, and that the scene is illuminated by point light sources at infinity, as well as some constant ambient illumination term. Under these assumptions, the outgoing radiance at a pixel i is given by:

$$L_o(\mathbf{p}) = \rho_{\mathbf{p}} \left(\alpha_0 + \sum_{i=1}^{N} V_{\mathbf{p}}(\mathbf{d}_i) \alpha_i \max\{\mathbf{d}_i \cdot \mathbf{n}_{\mathbf{p}}, 0\} \right), \tag{1}$$

where N is the number of light sources, $\rho_{\mathbf{p}}$ is the albedo at point $\mathbf{p}$, α_0 is the ambient intensity, $\alpha_i, i \in \{1, ..., N\}$ is the intensity of the i-th light source, $\mathbf{d}_i$ is the illumination direction of the i-th light source, and $V_{\mathbf{p}}(\mathbf{d}_i)$ is a visibility term for direction $\mathbf{d}_i$ at point $\mathbf{p}$, defined as:

$$V_{\mathbf{p}}(\mathbf{d}_j) = \begin{cases} 0, & \text{if ray from } \mathbf{p} \text{ along } \mathbf{d}_j \text{ intersects } \mathcal{G} \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

Assuming a simplified linear model for the camera sensors, we model the observed value at pixel (x, y) as:

$$I(x, y) = \kappa L_o(\mathbf{p}) + \epsilon, \tag{3}$$

where κ is an exposure parameter and ϵ is noise. Since we can only estimate light source intensities up to scale, we assume $\kappa = 1$.

We define illumination through parameters $\theta_{\mathcal{L}} = \{\alpha_0, \alpha_1, ..., \alpha_N, \mathbf{d}_1, ..., \mathbf{d}_N\}$, where $\mathbf{d}_i$ is the direction and α_i the intensity of light source i for $i \in \{1, N\}$, and α_0 is the ambient intensity. The information available to our method is a single color image $\mathbf{I}$ of the scene, an approximate 3D model of the geometry $\mathcal{G}$ of the scene and approximate camera parameters.

We assume we can obtain an initial cast shadow estimate from the input image (see Sec. 5.1). In such a shadow estimate the effects of albedo ρ are roughly factored out and the non-shadow pixels of $\mathbf{I}$ are masked out. Ideally, the value of each shadow pixel (x, y) in such a shadow image I_s would be the shading at that point due to the non-occluded light sources, given by:

$$I_s(x,y) = \left(\alpha_0 + \sum_{i=1}^{N} V_{\mathbf{p}}(\mathbf{d}_i)\alpha_i \max\{\mathbf{d}_i \cdot \mathbf{n}_{\mathbf{p}}, 0\}\right) + \epsilon, \tag{4}$$

where p is the 3D point where (x, y) projects to. In practice we can obtain a cast shadow cue $\hat{\mathbf{I}}_s$ which is a rough approximation of $\mathbf{I}_s$.

In the following sections we will present a model to jointly estimate the shadows $\mathbf{I}_s$ and the illumination parameters $\theta_{\mathcal{L}}$ from the approximate shadow cue $\hat{\mathbf{I}}_s$. In section 5.1 we present the shadow cue which we used to obtain our results.

3. Global MRF for Scene Photometry

In this section we describe the MRF model which models the creation of cast shadows, associating them with high-level information about geometry and the light sources. As mentioned earlier, the higher-order cliques in the model, the unknown MRF topology and the continuous search space complicate the problem. Therefore, we will first describe a method to reduce the search space and identify the MRF topology through an initial illumination estimate obtained using a voting algorithm.

3.1. Initializing the MRF Model

We use a greedy approach to get a rough estimate of illumination from the shadow cue $\mathbf{I}_s$, by the voting method in Algorithm 1. The idea is that, shadow pixels that are not explained from the discovered light sources vote for the occluded light directions. The pixels that are not in shadow vote for the directions that are not occluded. The set of all possible directions is evenly sampled by the nodes of a geodesic sphere [23]. After discovering a new light source direction, we estimate the associated intensity using the median of the values of pixels in the shadow of this new light source. The process of discovering new lights stops when the current discovered light does not have a significant contribution to the shadows in the scene. The results of the voting algorithm are used to initialize the MRF both in terms of topology and search space leading to more efficient use of discrete optimization. When available, the number of light sources can also be set manually.

Algorithm 1 Voting to initialize illumination estimate

Lights Set: $\mathcal{L} \leftarrow \varnothing$
Direction Set: $\mathcal{D} \leftarrow$ all the nodes of a unit geodesic sphere
Pixel Set: $\mathcal{P} \leftarrow$ all the pixels in the observed image
loop
  votes[$\mathbf{d}$] $\leftarrow 0, \forall \mathbf{d} \in \mathcal{D}$
  for all pixel $i \in \mathcal{P}$ **do**
    for all direction $\mathbf{d} \in \mathcal{D} \setminus \mathcal{L}$ **do**
      if $I_s(i) < \theta_S$ **and** $\forall \mathbf{d}' \in \mathcal{L}, V_i(\mathbf{d}') = 0$ **then**
        if $V_i(\mathbf{d}) = 1$ **then** votes[$\mathbf{d}$] $\leftarrow$ votes[$\mathbf{d}$] $+ 1$
      else
        if $V_i(\mathbf{d}) = 0$ **then** votes[$\mathbf{d}$] $\leftarrow$ votes[$\mathbf{d}$] $+ 1$
  $\mathbf{d}^* \leftarrow \arg\max_{\mathbf{d}}(votes[\mathbf{d}])$
  $\mathcal{P}_{\mathbf{d}^*} \leftarrow \{i | c_i(\mathbf{d}^*) = 1$ and $\forall \mathbf{d} \neq \mathbf{d}^*, c_i(\mathbf{d}) = 0\}$
  $\alpha_{\mathbf{d}^*} \leftarrow median \left\{ \frac{1 - I_s(i)}{\max\{-\mathbf{n}(\mathbf{p}(i)) \cdot \mathbf{d}^*, 0\}} \right\}_{i \in \mathcal{P}_{\mathbf{d}^*}}$
  if $\alpha_{\mathbf{d}^*} < \epsilon_\alpha$ **then**
    stop the loop
  $\mathcal{L} \leftarrow \mathcal{L} \cup (\mathbf{d}^*, \alpha_{\mathbf{d}^*})$

3.2. Markov Random Field Formulation

The proposed MRF consists of one node for each image pixel $i \in \mathcal{P}$ and one node for each light source $l \in \mathcal{L}$. Each pixel node and all the light nodes compose a high-order clique $c \in \mathcal{C}$. The 4-neighborhood system [1] composes the edge set $\mathcal{E}$ between pixels. The energy of our MRF model has the following form:

$$\begin{aligned} E(\mathbf{x}) \quad = \quad & \sum_{i \in \mathcal{P}} \phi_p(x_i) + \sum_{l \in \mathcal{L}} \phi_l(x_l) + \sum_{(i,j) \in \mathcal{E}} \psi_p(x_i, x_j) \\ & + \sum_{i \in \mathcal{P}} \psi_c(x_i, \mathbf{x}_{\mathcal{L}}), \end{aligned} \tag{5}$$

where $\phi_p(x_i)$ and $\phi_l(x_l)$ are the singleton potentials for pixel nodes and light nodes respectively, $\psi_p(x_i, x_j)$ is the pairwise potential defined on a pair of neighbor pixels, and $\psi_c(x_i, \mathbf{x}_{\mathcal{L}})$ is the high-order potential associating all lights in $\mathcal{L}$ and a pixel x_i.

The latent variable x_i for pixel node $i \in \mathcal{P}$ represents the intensity value for that pixel. We uniformly discretize the real intensity value $[0, 1]$ into N bins to get the candidate set $\mathcal{X}_i$ for x_i. The latent variable x_l for light node $l \in \mathcal{L}$ is composed of the intensity and the direction of the light. We sample the space in the vicinity of the light configuration obtained by the previous voting approach to get the candidate set $\mathcal{X}_l$ for x_l (see details later in this section).

3.2.1 Singleton Potentials for Pixel Nodes

This term encodes the similarity between the estimated intensity value and the shadow cue value and is defined as:

$$\phi_p(x_i) = w_s \min\{|x_i - I_s(i)|, t_p\}. \quad (6)$$

where an upper bound t_p for this cost term is used to avoid over-penalizing outliers and w_s is a positive weight coefficient (same for w_l, w_p and w_c below).

3.2.2 Singleton Potentials for Light Nodes

We use this term to favor illumination configurations generating shadow shapes that match observed shadow outlines by incorporating a shadow shape-matching prior into the MRF model.

We detect edges in the shadow cue image using a Sobel detector [5]. Let $\tau(i) \in [0, 2\pi)$ be the angle of the gradient at pixel i with the x-axis, and $\hat{\tau}(i) \in \{0, K-1\}$ a quantization of $\tau(i)$. For each possible direction $d \in \{0, K-1\}$, we compute a distance map υ_d which contains, for each pixel, the distance to the closest edge of orientation d (zero for pixels that lie on an edge of orientation d).

For pixel i with gradient angle $\tau(i)$, the distance function is computed by interpolating between the distance map values for the two closest quantized orientations:

$$dist_{\tau(i)}(i) = (1-\lambda) \cdot \upsilon_{\hat{\tau}(i)}(i) + \lambda \cdot \upsilon_{\hat{\tau}(i)+1}(i), \quad (7)$$

$$\lambda = \left\{ \frac{K \cdot \tau(i)}{2\pi} \right\}, \quad (8)$$

where $\{.\}$ indicates the fractional part. In our experiments, we chose $K = 4$.

We examine a configuration x_l of light l. The shape-matching prior expresses the quality of the match between the edges of the synthetic shadow $\mathcal{S}_l$ associated with x_l, given the geometry $\mathcal{G}$, and the observed edges in the shadow cue image:

$$\phi_l(x_l) = w_l \frac{1}{|\mathcal{E}_{\mathcal{S}_l}(x_l)|} \sum_{i \in \mathcal{P}_E(x_l)} dist_{\tau_{\mathcal{S}_l}(i)}(i), \quad (9)$$

where $\mathcal{E}_{\mathcal{S}_l}(x_l)$ is the set of all pixels that lie on edges of the shadow $\mathcal{S}_l$ generated by light label x_l and $\tau_{\mathcal{S}_l}(i)$ is the gradient angle of the synthetic shadow edge generated by x_l at pixel i. To determine the set of shadow edge pixels $\mathcal{E}_{\mathcal{S}_l}(x_l)$, we generate the shadow $\mathcal{S}_l$ created by light label x_l and the geometry $\mathcal{G}$ and then apply gaussian smoothing and the Sobel edge detector. The set $\mathcal{E}_{\mathcal{S}_l}(x_l)$ contains all pixels whose gradient magnitude is above θ_e.

Note that our MRF model is flexible with respect to both singleton terms and other singleton measures can be considered.

3.2.3 Pairwise Potentials

We adopt the well-known *Ising* prior to define the pairwise potential between a pair of neighboring pixels $(i, j) \in \mathcal{E}$ to favor neighbor pixels having the same value:

$$\psi_p(x_i, x_j) = \begin{cases} w_p & \text{if } x_i \neq x_j \\ 0 & \text{if } x_i = x_j \end{cases} \quad (10)$$

3.2.4 Higher-order Potentials

We use this term to impose consistency between the illumination configuration and the pixel intensity values.

Let $\mathcal{S}$ be the synthetic shadow, generated by light configuration $\mathbf{x}_\mathcal{L}$ and geometry $\mathcal{G}$. The intensity at pixel $i \in \mathcal{S}$, given a configuration $\mathbf{x}_\mathcal{L}$ of the lights, is:

$$s'_i(\mathbf{x}_\mathcal{L}) = \mathbf{x}^{\alpha_0} + \sum_{l \in \mathcal{L}} x_l^\alpha V_i(\mathbf{x}_l^{dir}) \max\{-\mathbf{x}_l^{dir} \cdot \mathbf{n}(i), 0\}, \quad (11)$$

where $\mathbf{x}^{\alpha_0}$ corresponds to the ambient intensity, x_l^α is the light intensity component of x_l, $\mathbf{x}_l^{dir}$ is the light direction component, $\mathbf{n}(i)$ is the normal at 3D point $\mathbf{p}$ imaged at pixel i and $V_i(\mathbf{x}_l^{dir}) \in \{0, 1\}$ is the visibility term for light direction $\mathbf{x}_l^{dir}$ at 3D point $\mathbf{p}$ (cf. Eq.2). For pixels $i \notin \mathcal{S}$, we set $s'_i(\mathbf{x}_\mathcal{L}) = 1$, according to the definition of our shadow cue $I_s(i)$. The clique potential is defined as:

$$\psi_c^{(1)}(x_i, \mathbf{x}_\mathcal{L}) = w_c \min\{(s'_i(\mathbf{x}_\mathcal{L}) - x_i)^2, t_c\}, \quad (12)$$

where t_c is also an upper bound to avoid over-penalizing outliers.

In cases where the geometry $\mathcal{G}$ is far from the real scene geometry, a light configuration that does not generate any visible shadows in the image might result to a lower MRF energy than the true light source. To avoid this degenerate case, we introduce the term $\psi_c^{(2)}(\mathbf{x}_\mathcal{L})$, which penalizes light configurations that do not generate any visible shadows in the image. The final form of the clique potential is:

$$\psi_c(x_i, \mathbf{x}_\mathcal{L}) = \psi_c^{(1)}(x_i, \mathbf{x}_\mathcal{L}) + \psi_c^{(2)}(\mathbf{x}_\mathcal{L}). \quad (13)$$

4. Inference

We can simultaneously estimate the cast shadows and the illumination through a minimization over the MRF's energy defined in Eq. 5:

$$\mathbf{x}^{opt} = \arg\min_{\mathbf{x}} E(\mathbf{x}) \quad (14)$$

This MRF model contains high-order cliques of size $|\mathcal{L}|+2$, which make energy minimization challenging.

The most straightforward manner to minimize the model energy is the high-order clique reduction technique proposed in [9], while a more promising alternative given the

complexity and the dimensionality of the problem is the dual decomposition [13].

[9] performs inference in a higher-order MRF with binary labels by reducing any pseudo-Boolean function to an equivalent quadratic one while keeping the minima of the resulting function the same as the original. Like [9], we employ the fusion-move [16] and QPBO [6, 12] algorithms to extend this method to deal with multi-label MRFs: During energy minimization, a number of iterations is performed, and for each iteration, the algorithm fuses the current labeling L_{cur} and a proposed labeling L_{prop} by minimizing a pseudo-Boolean energy [9].

However, this method in practice failed to provide good solutions. This can be explained by the complexity of the graph-structure, the number of labels (in particular with respect to the illumination variables) and the nature of pairwise and higher order interactions.

In order to address this failure and efficiently perform inference, we can split the minimization of the energy in Eq.5 in two stages [2]. If we assume that the light parameters are fixed, the high-order clique potentials in Eq.12 become singleton potentials of the form:

$$\psi_c^{(1)}(x_i|\mathcal{L}) = w_c \min\{(s_i'(\mathbf{x}_\mathcal{L}) - x_i)^2, t_c\}. \tag{15}$$

This way, for a fixed light configuration $\mathcal{L}$, we can compute the energy of the MRF model by rewriting the energy in Eq.5 as:

$$E(\mathbf{x}) = E_I(\mathbf{x}|\mathbf{x}_L) + E_L(\mathbf{x}_L), \tag{16}$$

where

$$E_I(\mathbf{x}|\mathbf{x}_L) = \sum_{i\in\mathcal{P}} \Big(\phi_p(x_i) + \psi_c^{(1)}(x_i|\mathcal{L})\Big) + \sum_{(i,j)\in\mathcal{E}} \psi_p(x_i, x_j) \tag{17}$$

is the energy of an MRF involving only pairwise potentials, given the light configuration $\mathcal{L}$, and

$$E_L(\mathbf{x}_L) = \sum_{l\in\mathcal{L}} \Big(\phi_l(x_l) + \psi_c^{(2)}(\mathbf{x}_\mathcal{L})\Big) \tag{18}$$

is the energy associated with the (fixed) light configuration $\mathcal{L}$. Given the light configuration $\mathcal{L}$, we can minimize the energy $E_I(\mathbf{x}|\mathbf{x}_L)$ using any of the various available inference algorithms for pairwise MRF models. For our experiments, we used the TRW-S belief propagation algorithm [11].

Furthermore, $\min_x\{E_I(\mathbf{x}|\mathbf{x}_L)\}$ changes with different light configurations, as shown in Fig.2. In order to minimize $E(\mathbf{x})$, we sample the light parameter space around the current estimate and we minimize the pairwise energy $E_I(\mathbf{x}|\mathbf{x}_L)$. We then compute the total MRF energy for a sample t of the light parameter space as

$$E^{(s)}(\mathbf{x}) = \min_x \{E_I(\mathbf{x}|\mathbf{x}_L) + E_L(\mathbf{x}_L)\}. \tag{19}$$

When we have multiple lights, in each iteration of the algorithm we change the parameters of only one of the light sources to a new guess. The final solution corresponds to the light parameter sample that generated the labeling with the lowest energy:

$$\mathbf{x}^{opt} = \arg\min_s E^{(s)}(\mathbf{x}). \tag{20}$$

This method is more tolerant to local minima in the model energy (which appear often in practice), while we also noticed that, compared to a more standard gradient descent method, it resulted in significantly less calls to evaluate energy $E_I(\mathbf{x}\}|\mathbf{x}_L)$, which are very costly. Results on convergence can be found in the supplemental materials.

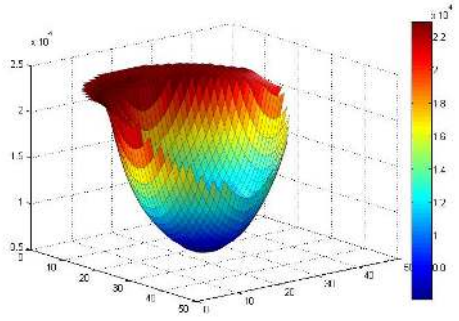

Figure 2. The model energy over possible directions of one light, for a simple synthetic scene.

4.1. Generating Proposals

Both approaches above utilize sampling of the solution space in order to generate proposals to minimize the MRF energy. Generating good guesses for these proposals is important in order to achieve fast convergence to a good solution. Here we describe how we generate proposals for each random variable class.

Light directions: We generate the proposed light source direction $\hat{\mathbf{x}}_l^{dir}$ by drawing a sample from a von Mises-Fisher distribution with mean direction $\dot{\mathbf{x}}_l^{dir}$ and concentration parameter κ_{sample}, where $\dot{\mathbf{x}}_l^{dir}$ is the estimated value from the previous iteration of the algorithm. The estimate from the voting algorithm is used for the first iteration. For our experiments, κ_{sample} was set to 200. Samples are drawn using the accept-reject algorithm.

Light intensities: For light source l, we compute a proposed intensity by adding a random offset (drawn from a normal distribution) to the current light source intensity estimate. We generate proposals for the ambient intensity $\hat{\mathbf{x}}^{\alpha_0}$ in the same way.

Pixel intensities: In the case of higher-order clique reduction, we also need to generate proposals for the pixel labels. The light proposal is kept fixed for N successive iterations, while each of the N pixel labels is proposed for every pixel node, after which a new light proposal is generated.

Figure 3. Bright channel: a. original image (from [3]); b. bright channel; c. confidence map; d. refined bright channel

5. Experimental Validation

In our discussion so far, we have assumed that some per-pixel estimate $\hat{\mathbf{I}}_s$ of the shadow/shading image $\mathbf{I}_s$ is available to be used as input in our MRF model. Before we proceed with the experimental validation, we will discuss the cues we used to estimate shadow intensity.

5.1. Shadow Cues

We use the *bright channel* cue [20], which is based on the following observations:

- The value of each of the color channels of the image has an upper limit which depends on the incoming radiance. This means that, if little light arrives at the 3D point corresponding to a given pixel, then all color channels will have low values.
- In most images, if we examine an arbitrary image patch, the albedo for at least some of the pixels in the patch will probably have a high value in at least one of the color channels.

From the above observations it is expected that, the maximum value of the r, g, b color channels in an image patch will be roughly proportional to the incoming radiance for many patches in the image. Therefore, the *bright channel*, I_{bright} for image $\mathbf{I}$ is defined in a way similar to [8]:

$$I_{bright}(i) = max_{c\in\{r,g,b\}}\left(max_{j\in\Omega(i)}(I^c(j))\right) \qquad (21)$$

where $I^c(j)$ is the value of color channel c for pixel j and $\Omega(i)$ is a rectangular patch centered at pixel i. As described in [20], the bright channel cue is computed in multiple scales, and confidence values are computed for each region based on hue differences across the region borders. The confidences are combined across scales, and then dark regions with low confidence are discarded (Fig.3).

5.2. Results

We evaluated our approach using images collected under controlled illumination conditions in the lab, as well as with real-world images of cars collected from Flickr, and the Motorbike images from Caltech 101 [17]. To visualize the estimated illumination, we rendered a synthetic vertical pole (sun dial) using the estimated light parameters and overlayed it to the original images.

We used 8 values for the pixel node labels, and performed 1000 iterations of our algorithm. The values we selected for the weights in our experiments were: $(w_s, w_l, w_p, w_c) = (8, 1, 1, 4)$. The upper bounds for the truncated potentials were selected to be $(t_p, t_c) = (0.5, 0.5)$.

5.2.1 Synthetic Dataset

We evaluated our method quantitatively on a set of synthetic images, rendered using a set of known light sources, selected randomly for each test image. We used area light sources of various sizes to evaluate our approach on both soft and hard shadows. The number of light sources varied from 1 to 3. We examined three different cases:

1. Accurate geometry: We estimated the illumination using the same 3D model used to render each image.
2. Approximate geometry: We estimated the illumination using a coarse 3D model that roughly approximated the original geometry by a bounding box and a ground plane.
3. Approximate geometry with noisy shadow input: We estimated the illumination using a coarse 3D model and a noisy initial shadow estimate. To obtain the latter, we added random dark patches to the rendered shadow. The reason is that, on one hand our methods are relatively insensitive to spatially-uniform random noise, and on the other, this kind of patch-based noise better emulates the errors in shadow estimation that happen in real data, which generally result in whole image regions erroneously identified as shadows.

For each estimated light source, we computed the difference in parameters from the true light source that was closest in direction to the estimated one. Table 5.2 shows the computed errors for light source direction and intensity, averaged over all images in the synthetic test set. We also compare the estimation accuracy with the results from the voting algorithm used to initialize the MRF model parameters. Our results demonstrate both the accuracy of our method and the robustness of the estimate with respect to large inaccuracies in the geometry and initial shadow estimate.

5.2.2 Real Datasets

We evaluated our approach further on images of the class "Motorbikes" of the Caltech 101 dataset [17]. For every image in this dataset we used *the same* rough 3D model representing an average motorbike, and *the same* average camera parameters across all images. We demonstrate that our algorithm can estimate the illumination effectively, despite the variations in geometry, pose and camera position

Figure 4. Results for the Motorbikes class of the Caltech101 dataset. A synthetic sun dial (orange) rendered with the estimated illumination is overlayed on each original image. A single 3D model capturing an average motorbike was used for the estimation in all instances, with the same average camera parameters.

	Exact geometry	Approx. geometry	Approx. geometry + noisy shadow input
Voting	12.30	12.86	28.81
MRF(without shape prior)	8.04	12.53	22.46
MRF(with shape prior)	6.77	13.38	15.44
MRF(HOCR [9])+shape prior	12.19	16.58	26.64

Table 1. Synthetic results: from left to right, we present the mean error in the estimated light directions on a synthetic dataset, using the exact geometry to do the illumination estimation; using geometry approximated by bounding boxes and a ground plane; and using approximate geometry and a noisy shadow input. Results are averaged over a random mix of images rendered with 1, 2 or 3 area light sources of random size. As expected, the benefit from the shadow shape-matching prior is largest in the case of noisy shadow input.

in each individual image. Results are shown in Fig.4. The results show that our approach is robust enough to estimate the illumination by using the same generic 3D model for all instances of a class of objects. In the most general case, an object detector can be used to recognize objects of known classes in an image, and then simple common class geometry can be used to estimate the illumination from that image, combined with a horizon line estimator, without any other information provided by the user for each image separately.

We also present results on images of cars collected from Flickr (Fig.5). The geometry in these cases was a bounding box corresponding to the car body and the ground plane. Camera parameters were matched manually. Results with both the Caltech 101 dataset and the images in Fig.5 were obtained assuming one light source in the scene. Despite our initial assumption of Lambertian reflectance, the results show that our algorithm can cope with the abundance of non-lambertian surfaces in these images.

Some cases where our algorithm fails are presented in Fig.6. The two main reasons for failure are that either there are large, dark areas that get mistaken for shadows or that the shadows are very dim and not well-defined, as is the case in cloudy days. In the case of the Caltech 101 dataset another reason was that, for a few images, the common 3D model we used was imaged at a significantly different position than the motorbike in the image.

6. Conclusions

In this paper, we introduced a higher-order MRF model to jointly estimate the illumination parameters and the cast shadows, where the joint modeling of the low-level evidence and the high-level prior knowledge within a single probabilistic model significantly improves the estimation performance. We presented results in various classes of scenes, demonstrating the power of our MRF illumination model. Our results with the Caltech 101 dataset show that we can estimate the illumination environment using the same geometry and even pose for a large class of scenes - aided by an object detector in more complex environments. In many cases, as with our results on car images from Flickr, a bounding box is enough to perform estimation. The experiments show that our approach is more general and more robust than previous approaches in illumination estimation and thus quite successful in real world images. Future work includes dual decomposition [13] for the optimization of the MRF. Furthermore, we are interested in incorporating our method in more general scene understanding tasks, *e.g.* refining our knowledge of scene geometry from the illumination estimates.

Acknowledgments: This work was partially supported by NIH grants 5R01EB7530-2, 1R01DA020949-01 and NSF grants CNS-0627645, IIS-0916286, CNS-0721701.

Figure 5. Results with images of cars collected from Flickr. Top row: the original image and a synthetic sun dial rendered with the estimated illumination; Bottom row: the final shadow labels. The geometry consists of the ground plane and a single bounding box for the car.

Figure 6. Some failure cases; top: the dark ground is mistakenly labeled as shadow; bottom: a very dim shadow and a dark ground patch result in an incorrect shadow and illumination estimate.

References

[1] Y. Boykov and G. F. Lea. Graph cuts and efficient n-d image segmentation. *IJCV*, 70(2):109–131, November 2006.

[2] M. Bray, P. Kohli, and P. H. S. Torr. Posecut: Simultaneous segmentation and 3d pose estimation of humans using dynamic graph-cuts. In *In ECCV*, pages 642–655, 2006.

[3] G. Finlayson, M. Drew, and C. Lu. Intrinsic images by entropy minimization. In *ECCV*, 2004.

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

[5] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., 2006.

[6] P. L. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 0-1 optimization. *Mathematical Programming*, 28:121–155, 1984.

[7] K. Hara, K. Nishino, and K. Ikeuchi. Light source position and reflectance estimation from a single view without the distant illumination assumption. *PAMI*, 27(4):493–505, 2005.

[8] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. In *CVPR*, 2009.

[9] H. Ishikawa. Higher-order clique reduction in binary graph cut. In *CVPR*, 2009.

[10] T. Kim and K. Hong. A practical approach for estimating illumination distribution from shadows using a single image. *IJIST*, 15(2):143–154, 2005.

[11] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.

[12] V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts-a review. *PAMI*, 29(7):1274–1279, 2007.

[13] N. Komodakis and N. Paragios. Beyond pairwise energies: Efficient optimization for higher-order mrfs. In *CVPR*, 2009.

[14] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Estimating natural illumination from a single outdoor image. In *ICCV*, 2009.

[15] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Detecting ground shadows in outdoor consumer photographs. In *European Conference on Computer Vision*, 2010.

[16] V. Lempitsky, C. Rother, , and A. Blake. Logcut - efficient graph cut optimization for markov random fields. In *ICCV*, 2007.

[17] F. Li, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *CVIU*, 106(1):59–70, April 2007.

[18] Y. Li, S. Lin, H. Lu, and H.-Y. Shum. Multiple-cue illumination estimation in textured scenes. In *ICCV*, 2003.

[19] A. Panagopoulos, D. Samaras, and N. Paragios. Robust shadow and illumination estimation using a mixture model. In *CVPR*, 2009.

[20] A. Panagopoulos, C. Wang, D. Samaras, and N. Paragios. Estimating shadows with the bright channel cue. In *CRICV 2010 (in conjuction with ECCV'10)*, 2010.

[21] R. Ramamoorthi, M. Koudelka, and P. Belhumeur. A fourier theory for cast shadows. *PAMI*, 27(2):288–295, 2005.

[22] E. Salvador, A. Cavallaro, and T. Ebrahimi. Cast shadow segmentation using invariant color features. *CVIU*, 95(2):238–259, 2004.

[23] I. Sato, Y. Sato, and K. Ikeuchi. Illumination from shadows. *PAMI*, 25(3):290–300, 2003.

[24] C. Wang, M. De la Gorce, and N. Paragios. Segmentation, ordering and multi-object tracking using graphical models. In *ICCV*, 2009.

[25] Y. Wang and D. Samaras. Estimation of multiple directional light sources for synthesis of augmented reality images. *Graph. Models*, 65(4):185–205, 2003.

[26] Y. Yang and A. Yuille. Sources from shading. In *CVPR*, 1991.

[27] W. Zhou and C. Kambhamettu. A unified framework for scene illuminant estimation. *IVC*, 26(3):415–429, 2008.

[28] J. Zhu, K. G. G. Samuel, S. Masood, and M. F. Tappen. Learning to recognize shadows in monochromatic natural images. In *CVPR*, 2010.