

# Robust Tracking and Stereo Matching under Variable Illumination

Jingdan Zhang      Leonard McMillan  
UNC Chapel Hill  
Chapel Hill, NC 27599  
{zhangjd, mcmillan}@cs.unc.edu

Jingyi Yu  
University of Delaware  
Newark, DE 19716  
yu@cis.udel.edu

## Abstract

*Illumination inconsistencies cause serious problems for classical computer vision applications such as tracking and stereo matching. We present a new approach to model illumination variations using an Illumination Ratio Map (IRM). An IRM computes the intensity ratio of corresponding points in an image pair. We formulate IRM recovery as a Markov network, which assumes spatially varying illumination changes can be modeled as a locally smooth function with boundaries. We show that the IRM Markov network can be easily incorporated into low-level vision problems, such as tracking and stereo matching, by integrating IRM estimation with the optical flow field/disparity map solution process. This leads to a unified Markov network. We develop an iterative optimization algorithm based on Belief Propagation to efficiently recover the illumination ratio map and the optical field/disparity map at the same time. Experiments demonstrate that our methods are robust and reliable.*

## 1 Introduction

Illumination changes confound many computer vision problems. For instance, in classical dense two-frame stereo matching, the scene is assumed to be Lambertian and corresponding pixels have consistent intensities from different viewpoints. In the presence of illumination inconsistencies such as shadows that appear in one image but not the other, state-of-art stereo matching methods frequently break down in applications such as time-lapse stereo (Figure 3 and 4). Similar problems exist when tracking objects moving into and out of shadows or when the camera motion causes changes in gain and exposure between consecutive frames (Figure 5). In all these cases the same part of the scene is inconsistently illuminated. Conventional methods such as mean shift trackers [3] fail to reliably track desirable objects [6] on these occasions.

Ideally, each image could be decomposed into an illumination image and a reflectance image. State-of-art methods then can be applied only on the reflectance image.

The reflectance image is also one of the intrinsic images [19, 18, 12]. Most previous approaches for recovering intrinsic images require capturing a sequence of images under different illumination conditions with the same viewpoint. For outdoor scene, this usually means capturing multiple images at different times of day. These approaches are unsuitable for time sensitive applications such as tracking and stereo matching. Recently, single-image-based intrinsic image recovery methods have also been proposed [5, 18]. However, these techniques require prior knowledge of the camera response curve or training data and the resulting intrinsic images are generally not coherent frame-to-frame.

In this paper, we present a different approach to solve the illumination inconsistency problem. Instead of finding the illumination invariant map of the scene, we set out to find the illumination transformation between the images. Specifically, we calculate the ratio of the intensity for the each scene point over the two images. The resulting ratio image is called the Illumination Ratio Map (IRM). For static scenes, the IRM is simply the ratio between the reference and the target images [7, 2]. For dynamic scenes (e.g., moving objects or moving cameras), computing the IRM requires warping the target pixels onto the reference image. Jacob, Belhumeur, and Basri [7] have shown that the illumination ratios in static scenes tend to exhibit spatial smoothness, although abrupt changes may occur across the shadow and occlusion boundaries. A similar analysis applies to dynamic scenes. The spatial smoothness and discontinuity properties of IRMs are very similar to the common assumptions made for stereo disparity images. Recent results [15, 17] have shown that fields satisfying these properties can be robustly modeled as a Markov network. In this paper, we will use the Markov network model to analyze the IRM. We also show how to incorporate the IRM Markov network with the optical flow field for tracking, or with the disparity map for stereo matching. Using an iterative optimization algorithm based on Belief Propagation, we show how to efficiently recover the illumination ratio map and the optical field/disparity map simultaneously.

There are three major contributions in this paper. First, we formulate the illumination inconsistency problem us-

ing the illumination ratio map (IRM) and we analyze the smoothness and discontinuity property of the IRM. Second, we model the IRM as a Markov network and demonstrate how to integrate problem-specific information (e.g., the optical flow or the disparity map) into a unified Markov network. Finally, we provide an iterative optimization algorithm using Belief Propagation to simultaneously recover the IRM and the optical flow/disparity map.

The remaining of the paper is structured as follows. Section 2 discusses the background and the related work. Section 3 analyzes the property of the illumination ratio map and describes how to model it as a Markov network. Section 4 shows how to combine the optical flow field or disparity map with the IRM as a joint Markov network. Section 5 proposes an iterative Belief Propagation algorithm to solve the Markov network. Section 6 shows results on synthetic and real experiments for tracking and stereo matching. Section 7 concludes.

## 2 Background and Previous Work

Classical computer vision problems such as stereo matching and tracking assume that the scene objects maintain consistent appearance in nearby cameras or frames. Tremendous efforts have been focused on how to robustly handle issues that are largely unrelated to illumination such as textures, noise, and occlusion boundaries [13, 11, 14, 15]. Graph-cut based algorithms [1, 9, 10] and Belief Propagation methods [20, 4, 15, 16] have achieved highly accurate and stable reconstruction on illumination-consistent scenes. However, these algorithms are sensitive to illumination changes such as varying shadows or lightings. Normalized-cross correlation (NC) is often used to reduce these illumination incoherencies. However, because NC is a window-based approach, features smaller than the window size in the scene cannot be correctly recovered. Typical examples are shown in Figure 3 and 4. Illumination variations also create serious problems for tracking. For instance, when tracking objects moving inside and outside the shadow regions, state-of-art algorithms such as mean shift [3] fail [6]. Furthermore, illumination variations can also lead to dramatic changes in the camera gain and exposure settings. As a result, the overall intensity level in the neighboring frames can be inconsistent [8].

One way to tackle the illumination inconsistency problem is to compute the illumination invariance image, also called the intrinsic image, and then apply state-of-art methods on intrinsic images. Weiss [19] developed a maximum likelihood (ML) method to recover the intrinsic image from an image sequence captured at a fixed viewpoint under significantly different lighting conditions. Matsushita et al. [12] extended Weiss’s algorithm to derive time-varying intrinsic images and utilized an eigenspace for capturing the dominant illumination variations. However, in tracking or

stereo matching, capturing multiple images with varying illuminations for each frame is impractical. Single-image-based intrinsic image recovery algorithms have also been proposed. Finlayson et al. [5] proposed to use a single color image to recover the intrinsic image by calculating an angle for an “invariant direction” in a log-chromaticity space. Tappen et al. [18] used multiple cues and a trained classifier to estimate both shading and reflectance intrinsic images. Although these approaches achieve plausible results from a single image, it is unclear how to maintain spatial-temporal consistencies when applying these algorithms on to multiple images or video frames.

Alternatively, it is possible to estimate illumination variations instead of the illumination invariance images. Jacob et al [7, 2] proposed to use the ratio image on static scenes. They used the ratio images to distinguish whether the two images capture the same object. Recently, Freedman and Turek [6] proposed an illumination-insensitive tracking algorithm. They placed consistency constraints on the illumination model by assuming that the consistent scene intensities should propagate throughout all frames. In this paper, we also model the illumination variations. We model the illumination changes using an Illumination Ratio Map (IRM). An IRM computes the intensity ratio of the same scene point in two images. Jacob et al’s illumination ratio image is a special case of the IRM when the viewpoint of the two images does not change and the scene is static [7].

Before proceeding, we clarify the notation.  $\mathbf{I}\{I_R, I_T\}$  represents an image pair, where  $I_R$  is the reference and  $I_T$  is the target images,  $(x, y)$  represents a pixel,  $D$  represents the disparity map, and  $d$  corresponds to the disparity for each pixel in  $D$ .  $\Gamma$  represents the IRM and  $\gamma(x, y)$  is the illumination ratio for a specified pixel in  $\Gamma$ .

## 3 Illumination Ratio Map

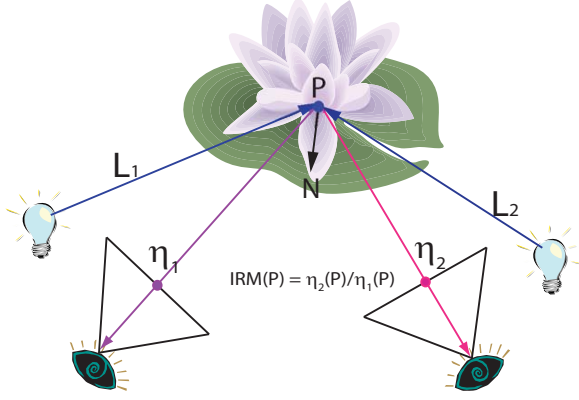
The Illumination Ratio Map (IRM) models the intensity variation or relative scale of corresponding scene points in two images. Consider a 3D point  $\dot{P}$  imaged at pixel  $(x, y)$ . Under Lambertian assumption, the intensity  $\eta(x, y)$  of  $\dot{P}$  can be computed as:

$$\begin{aligned} \eta(x, y) &= \sum_k \mu_k \rho(\dot{P}) (\hat{L}_k(\dot{P}) \cdot \hat{N}(\dot{P})) \\ &= \rho(\dot{P}) \sum_k \mu_k (\hat{L}_k(\dot{P}) \cdot \hat{N}(\dot{P})) \end{aligned} \quad (1)$$

where the scalar  $\mu_k$  is the magnitude of the light source  $k$  and  $\rho(\dot{P})$  is the reflectance (albedo) at  $\dot{P}$ .  $\hat{L}_k(\dot{P})$  is the normalized light direction for light source  $k$ , and  $\hat{N}(\dot{P})$  is the surface normal.

An IRM  $\Gamma$  computes the intensity ratio of each scene point  $\dot{P}$  between the target and the reference images, i.e.,

$$\gamma(x, y) = \frac{\rho(\dot{P}) \sum_k \mu_k (\hat{L}_k(\dot{P}) \cdot \hat{N}(\dot{P}))}{\rho(\dot{P}) \sum_{k'} \mu_{k'} (\hat{L}_{k'}(\dot{P}) \cdot \hat{N}(\dot{P}))}$$



**Figure 1.** The Illumination Ratio Map is computed as the intensity ratio of the corresponding pixels in a pair of images. It is a function of the surface normal and the lightings.

$$= \frac{\sum_k \mu_k(\hat{L}_k(\hat{P}) \cdot \hat{N}(\hat{P}))}{\sum_{k'} \mu_{k'}(\hat{L}_{k'}(\hat{P}) \cdot \hat{N}(\hat{P}))} \quad (2)$$

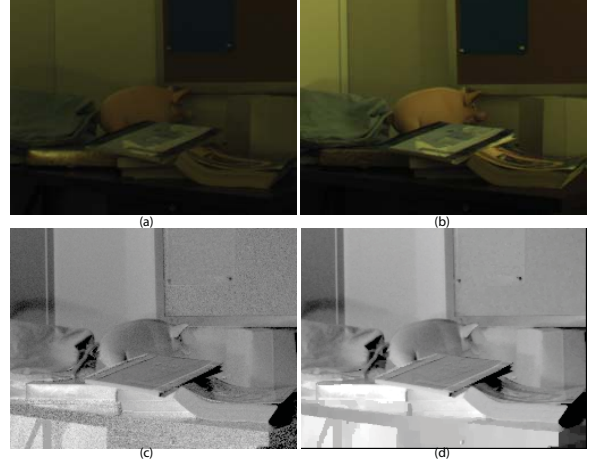
where  $k$  and  $k'$  correspond to different illuminations for  $I_T$  and  $I_R$ .

If  $I_R$  and  $I_T$  are captured at the same viewpoint, and scene is static, then the IRM  $\Gamma$  simply corresponds to  $\frac{I_T}{I_R}$ . If the camera position changes or any part of the scene moves, we need to finding the corresponding pixel in  $I_T$ , i.e.,

$$\gamma(x, y) = \frac{I_T(x + d^x, y + d^y)}{I_R(x, y)} \quad (3)$$

where  $D(d^x, d^y)$  corresponds to the disparity map or optical flow field. Our goal is to simultaneously recover  $\Gamma$  and  $D$  from images seen under differing illumination changes.

The Illumination Ratio Map (2) is independent of the surface reflectance  $\rho$ . This means an IRM should be insensitive to textures variations. The IRM is a smooth function of the surface normal. Since most of the objects are smooth in real scenes, the IRM maintains spatial smoothness in general. Discontinuities appear in the IRM at regions with abrupt normal changes (e.g., creases), across scene occlusion boundaries or shadow boundaries, or with moving objects. The discontinuities caused by normal changes are often moderate because the change in the viewing or lighting direction are usually limited. Occlusion boundaries are a major cause of IRM discontinuity. For instance, when the illumination direction changes across the two images, shadows abutting the occlusion boundaries may appear or disappear. As a result, sharp edges can appear near depth edges in the IRM. Shadow boundaries also change due to illumination variations. However, since most of the shadows are soft in real scenes, they often lead to smooth transition instead of discontinuities in the IRM. Finally, moving objects can also incur discontinuities. For instance, an object can



**Figure 2.** Illumination Ratio Map. (a) and (b) are captured with a fixed camera of a static scene under different lighting conditions. (c) shows the log ratio image by dividing (a) by (b). Notice the noise in the ratio image in the dark regions. (d) is the recovered IRM using Markov network and Belief Propagation. It is much smoother and cleaner because of the spatial smoothness prior.

move inside and outside the bright and dark regions (shadows) or cast new shadows onto the other parts of the scene. In practice, the illumination changes on the moving objects are usually smooth with an exception near depth discontinuities. Figure 2(c) and 4(c) illustrate typical Illumination Ratio Maps.

## 4 IRM Markov Network

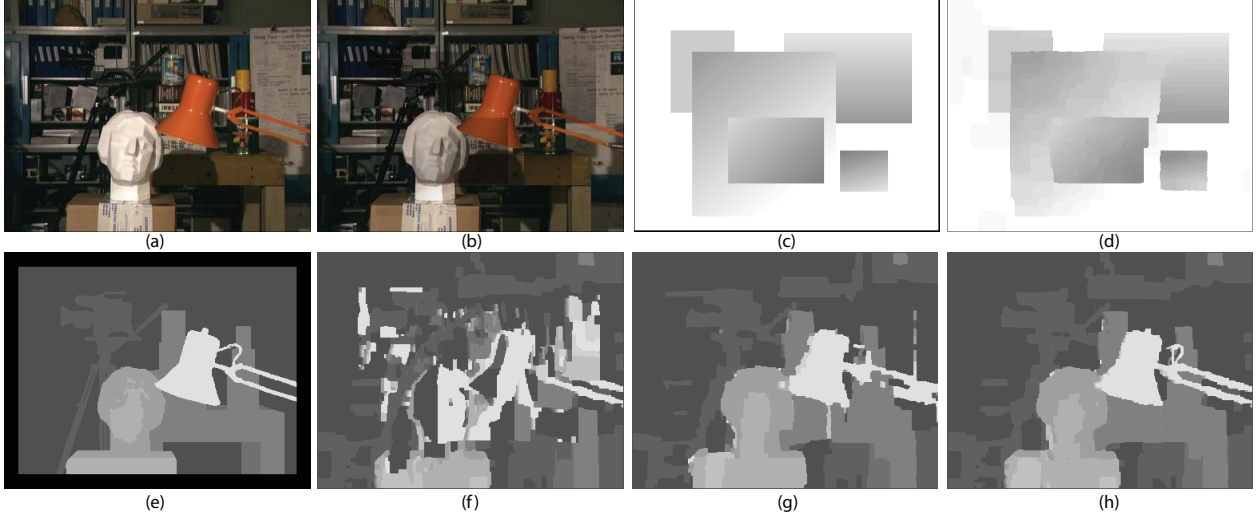
The spatial smoothness and occlusion discontinuity of the IRM are very similar to those seen in disparity maps from stereo matching. Previous researchers [20, 4, 15, 16, 17] have shown that fields satisfying such properties can be modeled as a probabilistic graphical model. Thus, we use a Markov network to model the IRM.

Each node in the graph corresponds to a pixel in the IRM and is assigned with an observed variable and a hidden variable. Given the reference and the target images  $\mathbf{I}\{I_R, I_T\}$ , our goal is to find a maximum a posteriori (MAP) Markov network of the IRM. This amounts to solving the following equation:

$$P(\Gamma|\mathbf{I}) \propto \prod_{s,t \in N} \Psi_{s,t}^\gamma(\gamma_s, \gamma_t) \prod_s \Phi_s(\gamma_s) \quad (4)$$

where  $\Phi_s$  is the evidence function describing how well the estimated IRM fits  $I_R$  and  $I_T$ .  $\Psi_{s,t}^\gamma$  is the compatibility function, which enforces the pairwise smoothness of  $\Gamma$ .

In the most basic setup where the reference and the target images are captured from the same viewpoint in a static scene, the IRM, in theory, could be computed as a simple



**Figure 3.** Stereo Matching on Synthetic Data. (a) is the left image in the Tsukuba stereo pair. (b) is the right image corrupted by a synthetic shadow mask (c). (c) Ground truth illumination ratio map (IRM). (d) The recovered IRM using our algorithm. (e) Ground truth disparity map. (f) Recovered disparity map using per pixel-based intensity difference (SD). (g) Recovered disparity map using normalized-cross correlation (NC). (h) Result of our algorithm.

ratio between the two images. However, IRM obtained this way can be very noisy due to the noise from the capturing process, as is shown in Figure 2(c). Instead, if we model the IRM as a Markov network and solve it using the Belief Propagation method (see Section 5), we are able to obtain much smoother and cleaner IRM as shown in Figure 2(d). In a more sophisticated setup such as stereo matching or tracking, neither the IRM nor the disparity map/optical flow field is unknown. We propose a unified framework to recover both the disparity map/optical flow field and the IRM at the same time.

#### 4.1 Disparity Markov Network

Stereo matching in presence of illumination changes is a challenging problem because both the disparity map  $D$  and illumination ratio map  $\Gamma$  are unknown. Here, we use a coupled Markov network to model this problem. Assume the two images have been rectified to have only horizontal parallax, we then assign the 1D disparity map to each node. Each node (pixel) in the Markov network, thus, has state  $(\gamma, d)$ , which corresponds to the illumination ratio map  $\Gamma$  and the disparity map  $D$ . We estimate the MAP of the joint probability given a image pair  $\mathbf{I}\{I_R, I_T\}$ . This can be formed as:

$$P(\Gamma, D|\mathbf{I}) \propto \prod_{s,t \in N} \Psi_{s,t}^d(d_s, d_t) \prod_{s,t \in N} \Psi_{s,t}^\gamma(\gamma_s, \gamma_t) \prod_s \Phi_s(\gamma_s, d_s, \mathbf{I}) \quad (5)$$

where the evidence function  $\Phi_s$  describes how well the illumination ratio  $\gamma_s$  and the disparity  $d_s$  fits the observation.

$\Psi_{s,t}^\gamma$  and  $\Psi_{s,t}^d$  encode the smooth prior of  $D$  and  $\Gamma$  separately.

We use intensity consistency to define the evidence function  $\Phi_s$  given hypothesized disparity map and illumination ratio map.

$$\Phi_s(\gamma_s, d_s, \mathbf{I}) = \exp(-E_c(\gamma_s, d_s, \mathbf{I})) \quad (6)$$

where  $E_c(\gamma_s, d_s, \mathbf{I})$  measures the consistencies for pixel  $(x_s, y_s)$  given  $\gamma_s$  and  $d_s$ .

$$E_c(\gamma_s, d_s, \mathbf{I}) = \min(\|\gamma_s \cdot I_R(x_s, y_s) - I_T(x_s + d_s, y_s)\|_2, T_c) \quad (7)$$

We use the L2-norm for computing  $E_c$ . The truncation threshold  $T_c$  is used to reduce noise and remove outliers.

Following the analysis in Section 3, we see that the IRM and the disparity map both satisfy the spatial smoothness property. Thus, we encode the piecewise smoothness property into the compatible functions as:

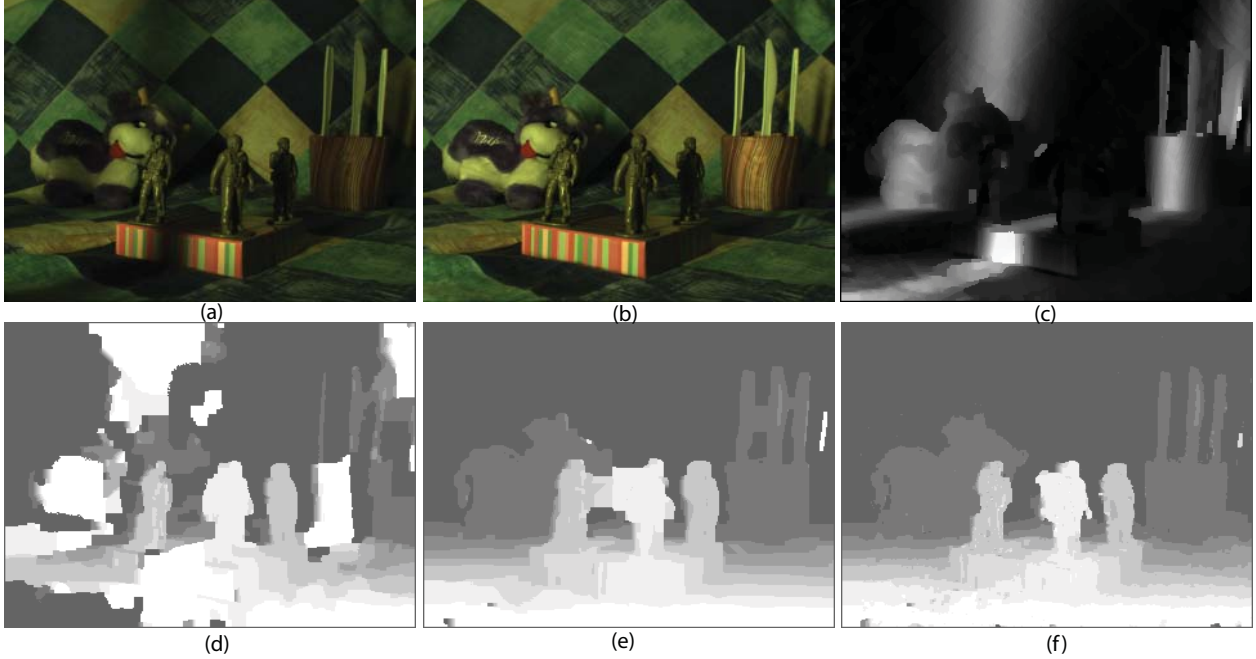
$$\Psi_s^d(d_s, d_t) = \exp(-E_d(d_s, d_t)) \quad (8)$$

$$\Psi_s^\gamma(\gamma_s, \gamma_t) = \exp(-E_\gamma(\gamma_s, \gamma_t)) \quad (9)$$

where  $E_d$  and  $E_\gamma$  are the truncated linear function:

$$\begin{aligned} E_d(d_s, d_t) &= \min(\lambda_d |d_s - d_t|, T_d) \\ E_\gamma(\gamma_s, \gamma_t) &= \min(\lambda_\gamma |\gamma_s - \gamma_t|, T_\gamma) \end{aligned} \quad (10)$$

where  $\lambda_d$  and  $\lambda_\gamma$  encode the rate of increase for the cost function, and  $T_d$  and  $T_\gamma$  are the truncation thresholds to model abrupt changes of the disparity map and the IRM near occlusion boundaries. The truncation function also allows us to implement an efficient Belief Propagation algorithm for solving the Markov network [4, 16].



**Figure 4.** Stereo Matching on real data. (a) and (b) are the left and the right stereo images of a real scene. In the left image, additional soft shadows were cast on the colored box, the background cloth, and the folks. (c) Estimated log domain IRM using our algorithm. (d) Estimated disparity map using per-pixel-based intensity difference. (e) Estimated disparity map using normalized-cross correlation. (f) Estimated disparity map using our algorithm.

The two compatible functions also differ in their parameters. Compared with the disparity map, the magnitude of the illumination ratio change is often much higher. In other words, the disparity map is smoother overall. We, therefore, set  $\lambda_d$  bigger than  $\lambda_\gamma$  to maintain stronger smooth priors.

## 4.2 Optical Flow Markov Network

When estimating the optical flow with the IRM, we formulate the problem in a similar way as the stereo matching. The only difference is that the displacement takes place at both the horizontal and vertical directions, i.e.,  $d_s = (d^x, d^y)$ . Therefore, the state space of each node is one dimensional higher than the disparity map Markov network. The evidence and the compatible function for the optical flow field are, therefore, modified as:

$$\begin{aligned}
 E_c(\gamma_s, d_s, \mathbf{I}) &= \\
 \min(\|\gamma_s \cdot I_R(x_s, y_s) - I_T(x_s + d_s^x, y_s + d_s^y)\|_2, T_c) \\
 E_d(d_s, d_t) &= \\
 \min(\lambda_d |d_s^x - d_t^x|, T_d) + \min(\lambda_d |d_s^y - d_t^y|, T_d)
 \end{aligned} \tag{11}$$

## 5 Optimization

In order to find the MAP of the joint Markov network, we use Belief propagation (BP) to iteratively optimize the IRM and the disparity map or optical-flow field. In BP, messages

are passed onto neighboring nodes to update the states of the hidden variables until it reaches MAP. We refer readers to [15, 17] for a detailed discussion on BP for stereo matching.

In the basic case when the two images are taken at the same viewpoint of a static scene, simply dividing, on a per pixel basis, the target image by the reference image gives a noisy IRM estimate, as shown in Figure 2(c). Instead, we use a single-hidden-variable IRM Markov network (4) to model the IRM. "Max-product" rule is used for computing the maximum joint posterior  $P(\Gamma|\mathbf{I})$ , where the message propagated from node  $s$  to  $t$  at the  $i^{th}$  iteration is computed as

$$M_{st}^i(\gamma_t) \leftarrow \kappa \max_{\gamma_s} \Phi_s(\gamma_s) \Psi_{s,t}^\gamma(\gamma_s, \gamma_t) \prod_{k \in N(s) \setminus t} m_{k,s}^{i-1}(\gamma_s) \tag{12}$$

The final belief of each node is determined by the local evidence function and the incoming messages from its neighbors:

$$b_s(\gamma_s) = \kappa \Phi_s(\gamma_s) \prod_{k \in N(s)} m_{k,s}(\gamma_s) \tag{13}$$

The result using BP for computing IRM of the static scene with fixed camera is shown in Figure 2(d).

### 5.1 Iterative BP

When the IRM is combined with the disparity map, the joint Markov network is two dimensional (or three dimen-

sional if combined with the optical flow). In theory, two or three dimensional BP algorithm could be directly applied to estimate the MAP of the joint Markov network. However, it is practically infeasible. In a normal situation, the possible range of illumination ratio is large, and we usually need more than 50 levels of illumination ratio to reduce the discretization error. We also use more than 10 levels of possible disparities for each pixel. This means that more than 500 states are needed to model the hidden variables for each pixel. The huge memory and computation requirements are prohibitive.

However, the joint Markov network (5) implies if one of the hidden variables, say the disparity, is known, then standard BP algorithm can be applied to calculate the MAP for the IRM, and vice versa. Therefore, we implement an iterative optimization algorithm to alternatively compute the MAP of the IRM and the disparity map Markov networks.

We start with estimating the initial disparity map. We use normalized cross-correlation (NC) to reduce the illumination inconsistencies and use it as the correspondence measure. We also modify the evidence function to measure only disparity consistencies, i.e.,

$$\Phi_s(d_s, \mathbf{I}) = \exp(-E_{corr}(d_s, \mathbf{I})) \quad (14)$$

where  $E_{corr}$  computes the normalized cross-correlation. We then solve the simplified disparity Markov network:

$$P(D|\mathbf{I}) \propto \prod_{s,t \in N} \Psi_{s,t}^d(d_s, d_t) \prod_s \Phi_s(d_s, \mathbf{I}) \quad (15)$$

After initial disparity estimation, we can iteratively estimate the IRM and the disparity map. When estimating the IRM given disparity, we omit the compatible function (8) that corresponds to disparity and use  $d_s = \hat{d}_s$  in equation (5), where  $\hat{d}_s$  is the estimated disparity from last iteration. Max product BP is again used to calculate the MAP of the IRM.

With known IRM, similar steps are taken to estimate the disparity map. In this case, the compatible function of the IRM (9) is omitted and we use  $\gamma_s = \hat{\gamma}_s$ , where  $\hat{\gamma}_s$  is the estimated IRM from the last iteration.

Our iterative optimization algorithm alternatively estimates the MAP of the IRM given a disparity map, and the MAP of the disparity map given an IRM. We implement a similar optimization algorithm in tracking except that we compute the MAP for both  $d^x$  and  $d^y$ . We apply a very efficient implementation [4] of BP to compute MAP at each iteration. We find it often sufficient to iterative 3 or 4 times to obtain satisfactory results in our experiment.

## 6 Results

In this section, we demonstrate our algorithm in tracking and stereo matching under variable illumination conditions.

So far, our algorithm assumes the surface albedo is independent of the light spectrum and, hence, the IRM can be computed as the ratio of the intensities. In practice, real materials often contain spectrum-dependent albedos particularly in color textured regions. In these cases, the IRM should also be computed in term of the spectrum IRM. To simplify this problem, we adopt a commonly used assumption [5] that the camera has narrow band responses for the  $r$ ,  $g$ , and  $b$  color channels. Under this assumption, we can compute the color IRM as the average of individual color IRM  $I^r$ ,  $I^g$ , and  $I^b$ .

$$\gamma_{color}(x, y) = \frac{\gamma^r(x, y) + \gamma^g(x, y) + \gamma^b(x, y)}{3} \quad (16)$$

where

$$\gamma^i(x, y) = \frac{I_T^i(x + d^x, y + d^y)}{I_R^i(x, y)} \quad i = r, g, b. \quad (17)$$

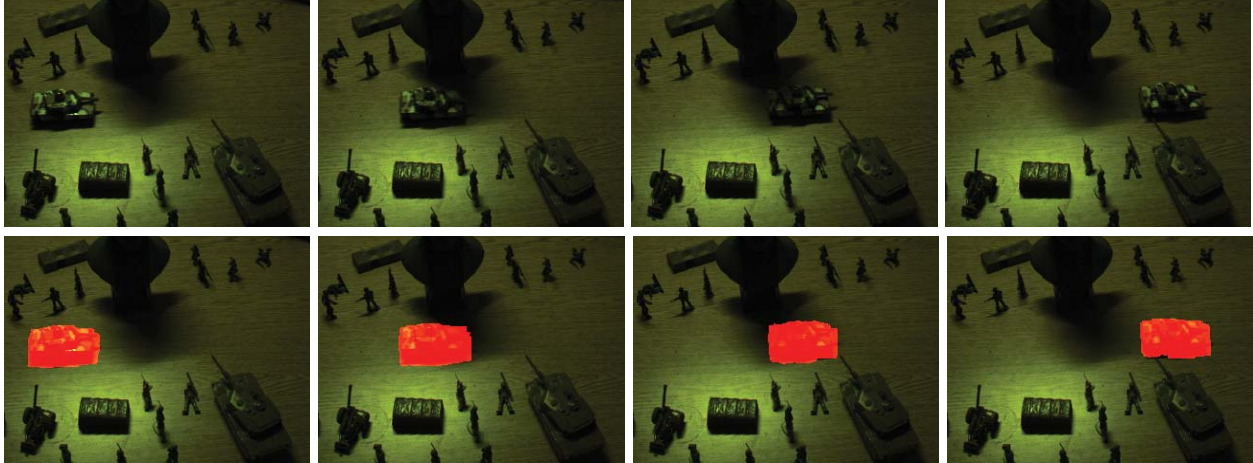
We find equation (16) is usually a good approximation. For example, in Figure 4, despite significant variations of the colored patterns on the background cloth, the recovered IRM is smooth and insensitive to the texture. In our experiment, we use equation (16) to estimate the IRM.

### 6.1 Stereo Matching

We first demonstrate our method using synthetic data. In Figure 3, we show the Tsukuba stereo image pair. The target (right) image is corrupted by modulation with a synthetic mask 3(c). Directly applying state-of-art stereo matching algorithms results in serious errors. In 3(f), we apply Belief Propagation using per-pixel-based intensity difference ( $SD$ ). The recovered disparity image captures details in unmasked regions but incur large errors in masked regions. To reduce the artifacts of inconsistent illumination, we then apply normalized-cross correlation ( $NC$ ) with Belief Propagation. The window-based approach incurs less error but the result is overly blocky. For example, it cannot capture fine features such as arm of the lamp. Figure 3(h) shows the results using our method. Our method is able to simultaneously capture fine details as well as maintain smoothness. Figure 3(c) and 3(d) compares the recovered IRM using our method and the ground truth.

We then tested our method on real scenes in Figure 4. The reference and the target images were captured using a Pointgrey Flea camera (Sony ICX204 sensor). In the left reference image, additional shadows were cast onto the colored box, the toy soldiers, and the plastic utensils. Figure 4(d) uses BP with  $SD$  and 4(e) with  $NC$ . Notice methods using  $SD$  are very sensitive to illumination changes and often incur large error in the illumination inconsistent regions.  $NC$  maintains an overall smoothness but is unable to recover fine details such as the background folks and the foreground





**Figure 5.** Tracking a toy tank across the shadow. Top row: frame 7, 34, 59, and 89 from the toy sequence. Bottom row: tracking results using our method. Our method maintains a secure focus on the toy tank through out the sequence.

toy soldiers. Furthermore, since additional shadows were cast between the left and the middle toy soldiers, *NC* merges the two with a uniform disparity. Our method 4(f) is able to robustly recover both the disparity map with fine details and the illumination ratio map.

Our result, in theory, can be further improved using the symmetric stereo algorithms proposed recently by Sun et al [16]. Authors of [16] proposed to include an additional binary process in the Markov network to model occlusions. However, doing so requires encoding even more states in the hidden variables. We intend to further investigate efficient ways to reduce the state space in the future.

## 6.2 Tracking

Next, we present results on tracking using our algorithm. The first sequence contains a toy car moving in and out of shadows. This is a challenging scene for conventional tracking algorithms [6]. To track the toy car, we compute the optical flow using the joint IRM/flow Markov network and integrate it over frames. In Figure 5, We highlight our tracking results in red. Our algorithm is able to reliably track the moving objects in the presence of both soft and strong shadows.

The second sequence shows the toy car moving under rapidly varying illumination. This is also a difficult scene for most algorithms. In the middle row of Figure 6, we show the tracking results of respective frames using our algorithm. Our method is insensitive to illumination variations.

## 6.3 Delighting

We can further reduce the illumination inconsistencies across the frames by delighting/relighting the video using the computed IRMs. For a static scene, delighting can be achieved by averaging the neighboring frames. For a dy-

amic scene, in theory, one can warp the moving parts and then compute the average intensity. However, averaging over the intensities is problematic because inaccurate optical flow calculation act like random noise, and blindly averaging them over the moving areas results in blurry images. Instead, we average the low-frequency illumination ratio maps over the frames.

Denote  $\gamma_{i \rightarrow k}$  to be the IRM of frame  $i$  with respect to  $k$ . The average IRM at frame  $k$   $\bar{\gamma}_k$  is computed over the neighboring  $2N + 1$  frames as:

$$\bar{\gamma}_k = \frac{\sum_{i=k-N}^{k+N} \gamma_{i \rightarrow k}}{2N + 1} \quad (18)$$

Notice,  $\gamma_{i \rightarrow k}$  accumulatively combines the IRMs as:

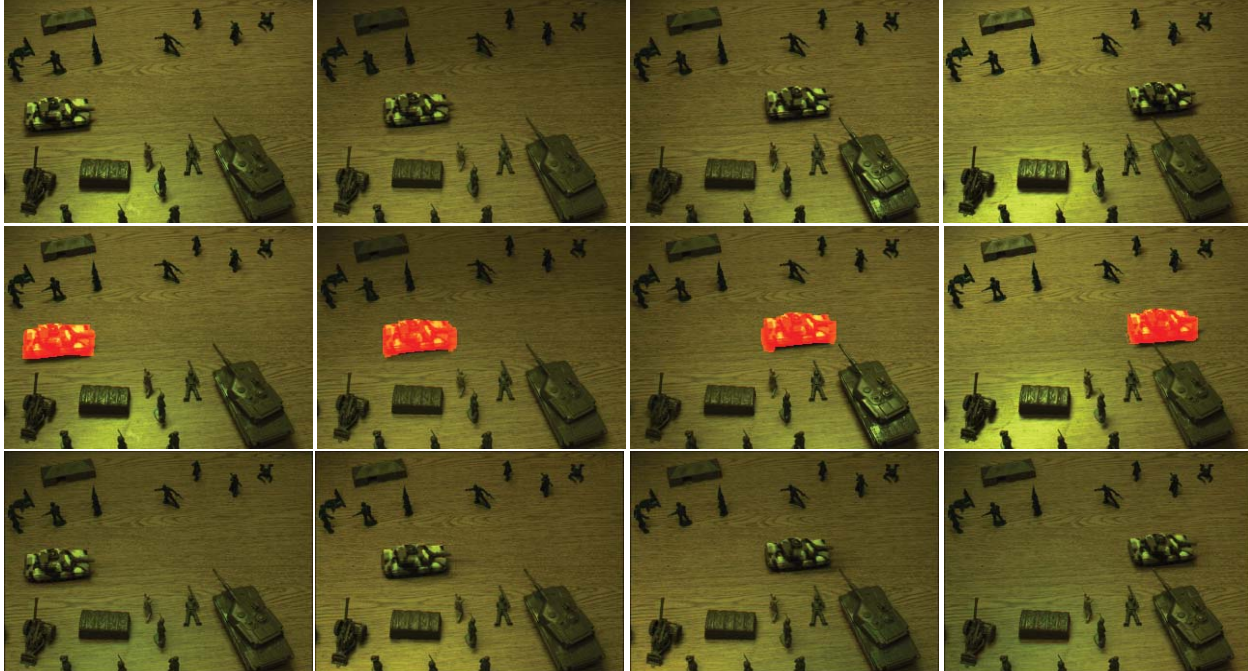
$$\gamma_{i \rightarrow k} = \begin{cases} \frac{1}{\gamma_i \gamma_{i+1} \cdots \gamma_{k-1}}, & i < k \\ 1, & i = k \\ \gamma_k \gamma_{k+1} \cdots \gamma_{i-1}. & i > k \end{cases} \quad (19)$$

where  $\gamma_i$  corresponds to the IRM of frame  $i$  with respect to frame  $i + 1$  after warping.

The bottom row in Figure 6 shows the delighted sequence using our algorithm. Despite the dramatic illumination changes between the frames, our algorithm is able to synthesize a new illumination consistent video.

## 7 Conclusions

We have presented a new approach to model illumination variations using an Illumination Ratio Map (IRM). An IRM computes the intensity ratio of corresponding points in an image pair. We have shown that an IRM maintains spatial smoothness and can be effectively modeled as a Markov network. We have demonstrated how to incorporate the IRM Markov network into low-level vision problems including tracking and stereo matching, by integrating IRM



**Figure 6.** Tracking a toy tank under variable illumination. Top row: frame 1, 20, 44, and 68 from the second toy sequence. Middle row: tracking results using our method. Bottom row: delighted frames.

estimation with the optical flow/disparity map solution process. This leads to a unified Markov network. We developed an iterative optimization algorithm based on Belief Propagation to efficiently recover the illumination ratio map and the optical field/disparity map at the same time. We have demonstrated our algorithm for both stereo matching and tracking on scenes with variable illumination. Experimental results show that our methods are robust and reliable.

## References

- [1] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222-1239, 2001.
- [2] H. F. Chen, P. N. Belhumeur, and D. W. Jacob. In Search of Illumination Invariants. *ICCV*, 254-261, 2000.
- [3] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 24(5):603-619, 2001.
- [4] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *CVPR*, Vol I:261-268, 2004.
- [5] G. Finlayson, S. Hordley, and M. Drew. Removing shadows from images. *ECCV*, 2002.
- [6] D. Freedman and M. W. Turek. Illumination-invariant tracking via graph cuts. *CVPR*, volume 2, pages 10-17, 2005.
- [7] D. W. Jacob, P. N. Belhumeur, and R. Basri. Comparing images under variable illumination. *CVPR*, 610-617, 1998.
- [8] S.J. Kim, M. Pollefeys. Radiometric Self-Alignment of Image Sequences. *CVPR*, 2004
- [9] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. *ICCV*, pages 508-515, 2001.
- [10] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. *ECCV*, Vol III:82-96, 2002.
- [11] D.Marr and T.A. Poggio. Cooperative computation of stereo disparity. *Science*, 194(4262):283-287, 1976.
- [12] Y. Matsushita, K. Nishino, K. Ikeuchi, and S. Masao. Illumination normalization with time-dependent intrinsic images for video surveillance. In *CVPR*, 2004.
- [13] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1):7-42, 2002.
- [14] C. Strecha, R. Fransens, and L. Van Gool. Wide-baseline stereo from multiple views: a probabilistic account. *CVPR*, Vol I:552-559, 2004.
- [15] J. Sun, H. Y. Shum, and N. N. Zheng. Stereo matching using belief propagation. *ECCV*, Vol II:510-524, 2002.
- [16] J. Sun, Y. Li, S. B. Kang, and H. Y. Shum. Symmetric stereo matching for occlusion handling. *CVPR*, Vol II: 399-406, 2005.
- [17] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. *ICCV*, Vol II:900-906, 2003.
- [18] M. F. Tappen, W. T. Freeman, E. H. Adelson. Recovering Intrinsic Images from a Single Image. *TPAMI*, vol. 27, no. 9, pp. 1459-1472, September, 2005.
- [19] Y. Weiss. Deriving intrinsic images from image sequences. *ICCV*, 68-75, 2001
- [20] Y. Weiss and W. T. Freeman. On the optimality of solutions of the max-product belief propagation algorithm in arbitrary graphs. *IEEE Transactions on Information Theory*. 47, 2, 723-735.