*Technical Report DU-CS-10-05*

*Department of Computer Science, Drexel University*

# Bayesian Defogging

**Ko Nishino** **Louis Kratz** **Stephen Lombardi**

Department of Computer Science

Drexel University

3141 Chestnut Street

Philadelphia, PA 19104, USA

E-mail: kon@drexel.edu

**Abstract**

Atmospheric conditions induced by suspended particles, such as fog and haze, severely alter the scene appearance. Restoring the true scene appearance from a single observation made in such bad weather conditions remains a challenging task due to the inherent ambiguity that arises in the image formation process. In this paper, we introduce a novel Bayesian probabilistic method that jointly estimates the scene albedo and depth from a single foggy image by fully leveraging their latent statistical structures. Our key idea is to model the image with a factorial Markov random field in which the scene albedo and depth are two statistically independent latent layers and to jointly estimate them. We show that we may exploit natural image and depth statistics as priors on these hidden layers and estimate the scene albedo and depth with a canonical expectation maximization algorithm with alternating minimization. We experimentally evaluate the effectiveness of our method on a number of synthetic and real foggy images. The results demonstrate that the method achieves accurate factorization even on challenging scenes for past methods that only constrain and estimate one of the latent variables.

## I. INTRODUCTION

Seeing through varying weather conditions remains a challenge for computer vision systems to operate in the real world. Among the many possible weather conditions, stationary atmospheric effects induced by suspended particles such as aerosols and water droplets, i.e., haze and fog, are of particular interest as they cause non-trivial degradations to the captured images. Removing such unwanted visual effects, often referred to as "dehazing" or "defogging," has strong implications in many scene understanding applications. For instance, it can help retain visibility when autonomously navigating a robot or a vehicle, alleviate the contrast loss that makes object recognition and detection harder, and most important provide means for analyzing the true scene appearance regardless of the weather condition.

Restoring the true scene appearance, i.e., recovering how the scene would have looked like on a clear day, from a single weather-degraded image, however, remains a challenging problem. As shown by Narasimhan and Nayar (2002), the foggy day scene appearance is the end result of attenuated scene color and transmitted airlight whose extents depend on the scene distance to the camera (depth). As a result, an inherent bilinearity between scene albedo and depth underlies the weather-degraded image that renders the inverse problem fundamentally ill-posed, as we will later make explicit. Past approaches tackle this problem by either assuming that one of the two unknowns –usually the scene depth– is known or rely on empirical observations to constrain

Fig. 1. An example foggy image of a pumpkin patch (reprinted from (Fattal, 2008b)). The scene appearance is nonlinearly distorted due to scattering of scene reflection and airlight by suspended atmospheric particles. Restoring the true scene appearance, i.e., "defogging" the image, is challenging due to the inherent bilinear ambiguity of scene albedo and depth in the underlying image formation process.

the scene color estimates, e.g., find a clear-day image with locally maximal contrast. In these approaches, the scene depth is known or is rather a byproduct computed once the clear-day image is estimated.

In this paper, we derive a novel *joint estimation* method based on a Bayesian formulation to factorize a single foggy image into its scene albedo and depth. As the underlying problem is inherently bilinear, we believe that the two variables, scene albedo and depth, should be estimated jointly and both solution spaces must be reduced with realistic constraints. The key observation we make is that both the scene albedo and scene depth carry over their latent structures to the observed foggy image. To facilitate this structural information of the scene, we formulate the image formation on a foggy (or hazy) day with a Factorial Markov Random Field (FMRF) (Kim and Zabih, 2002).

The FMRF model consists of a single observation, the foggy image, and two statistically independent latent layers that represent the scene albedo and the scene depth. This particular formulation with a probabilistic graphical model allows us to seamlessly integrate the structural constraints that we may extract from the input image as priors on the latent layers. In particular, we show that we can extract parameters of a heavy-tail distribution on the gradients of the scene albedo and impose it as a prior that is specific to the scene. We also show that, depending on the scene, we can impose different priors on the depth layer that better capture the scene structure.

For the actual factorization, we derive an Expectation Maximization algorithm tailored to the specific FMRF formulation using conventional graph-cuts algorithms in the expectation steps to obtain maximum a posteriori (MAP) estimates of the albedo and depth values. In this paper we will use the term defogging, but the method does not rely on the density of the suspended particles as long as single scattering dominates, and can be also applied to hazy images.

We experimentally evaluate the effectiveness of the method by showing the factorization results on a number of images. The results show that the method successfully arrives at physically plausible estimates, i.e., realistic recovery of clear-day images and scene structure even for images that prove challenging for past methods.

## II. Related Work

The visual effect induced by bad weather conditions has been of great interest in a number of fields–most prominently in meteorology–and their distinct characteristics, depth-dependent contrast loss and color shifts, can even be found in medieval paintings. Scientific studies of such weather-degraded scene appearances in images, especially of their physical and optical processes, however, only started in the mid 1900s (Middleton, 1952; McCartney, 1975). Early approaches to image restoration of weather-degraded images (a.k.a. defogging) treat the problem as yet another instance of image contrast enhancement. For instance, Grewe and Brooks (1998) learn an optimal set of parameter values, from pairs of images of the same scene with and without weather degradation, in order to fuse multiple images to reduce the atmospheric effects.

Image formation under static bad weather conditions such as fog and haze can be described with rigorous physically-based analytical models that provide a sound foundation for tackling its inversion. Based on earlier analysis of atmospheric light transport (Middleton, 1952; McCartney, 1975), Narasimhan and Nayar (2002) recently derived a dichromatic atmospheric scattering model that represents the light attenuation caused by single scattering of light with suspended atmospheric particles. By using this dichromatic model and the constraints it imposes in the color space, analogous to the dichromatic reflectance model (Shafer, 1985) and its use for reflectance component separation (Klinker et al, 1990) and illumination estimation (Tominaga and Wandell, 1989), they showed that a clear-day image and scene depth can be recovered from multiple (at least two) images of the same scene taken under different bad weather conditions, i.e., different fog densities. They further showed that essentially the same model can be used to impose

chromatic constraints to estimate scaled depth structure and then to restore the contrast of the scene from two images taken under different bad weather conditions. Schechner et al (2003) exploit the difference in the polarization characteristics of airlight and scene reflection to dehaze a scene from two images taken with polarizers of different angles.

Capturing multiple images, especially under different weather conditions, can be prohibitive in practical situations. Defogging becomes particularly challenging when we only have access to a single input image because of the inherent bilinearity of scene albedo and scene depth. A few past methods alleviate this ambiguity altogether by assuming known scene depth by manually specifying the rough 3D scene structure (Narasimhan and Nayar, 2003c), by assuming a planar terrain viewed with an airborne camera (Oakley and Satherley, 1998; Tan and Oakley, 2001), or by aligning geo-referenced digital city models (Kopf et al, 2008b).

More recent work on single image defogging formulate the problem as local contrast enhancement while accounting for the scene albedo and depth relation. These approaches essentially impose empirical constraints on the scene albedo and treat scene depth as a byproduct of the estimation process. Tan (2008b) imposes a locally constant constraint on the albedo values to increase the contrast in local block regions of the image. The results inevitably suffer from block artifacts. The method is not meant to restore the actual scene albedo but rather just enhance the visibility, and does not lead to physically meaningful estimates of the albedo or depth. Tarel and Hautière (2009) estimate the "atmospheric veil," an image of the scattered airlight, by using combinations of min, max, and median filters to enforce piecewise constant, and use the estimate to obtain a contrast enhanced image of the scene. Fattal (2008b) imposes locally constant constraints of albedo values together with decorrelation of the transmission in local areas, and then estimates the depth value using the albedo estimates and the original image. The method does not constrain the scene's depth structure, only the scene albedo. Similarly recent independent work by He et al (2009b) imposes constraints only on the depth structure induced by an empirical observation of clear-day intensity values in local region and refines the blocky depth estimates with an image matting algorithm. We will discuss the similarity of this constraint to the initial depth estimates in our method and show that it naturally arises as an upper bound that can be computed per pixel.

In sharp contrast to these past methods, we tackle defogging as factorization of an image into scene albedo and depth. Most important, we treat scene depth as an equally important

player in the image formation process and jointly estimate it with the scene albedo. We achieve this with a novel probabilistic graphical model formulation of the underlying image formation process that enables the integration of structural constraints on both the depth and albedo values as statistical priors. Past work on single image defogging (Fattal, 2008b; He et al, 2009b; Tan, 2008b) may be viewed as a special case of our framework–one with a single latent layer opposed to the two layers we consider and jointly estimate. Finally, our method exploits natural structural information extracted from the input image itself, resulting in scene-specific priors that impose stronger constraints for resolving the bilinear ambiguity.

## III. The Bilinear Ambiguity in a Foggy Image

Let us first review image formation of a foggy image. Readers interested in further details are referred to (Narasimhan and Nayar, 2002, 2003b). When an image is formed through static bad weather conditions–we consider only fog or haze in this paper–the scene appearance is altered by the minute atmospheric particles suspended in air. The light rays traveling through the air get scattered by these particles before they reach the camera. In this paper, we only consider single scattering as multiple scattering occurs only when the fog is very dense[1].

The light reflected by the scene that would have determined the scene appearance on a clear day becomes attenuated by this scattering. As the number of times a light ray hits an atmospheric particle is proportional to the distance it travels, this attenuation is clearly scene depth dependent. Let us denote the airlight that illuminates the scene with a 3-vector of RGB values $\mathbf{L}_\infty$. Allard's law, originally derived for night visual range of light sources, can be used to express the image irradiance $\mathbf{L}_s$ at each pixel $\mathbf{x} = (x, y)$ due to reflected scene radiance (Middleton, 1952; Narasimhan and Nayar, 2002):

$$\mathbf{L}_s(\mathbf{x}) = \frac{\mathbf{L}_\infty \rho(\mathbf{x}) e^{-\beta d(\mathbf{x})}}{d(\mathbf{x})^2},$$

where $\beta$ is the attenuation coefficient that we assume to be uniform across the entire scene, i.e., the fog is spatially homogeneous, and also is independent of the wavelength; and $\rho(\mathbf{x})$ and $d(\mathbf{x})$ are the scene albedo and depth, respectively, at the scene point corresponding to image coordinates $\mathbf{x}$. The scene albedo is a 3-vector of RGB color channels and the scene depth a

[1]Eliminating its visual effects can be achieved through deconvolution (Narasimhan and Nayar, 2003a).

scalar value shared among all color channels, so to be more precise $d(\mathbf{x})$ should be considered as $d(\mathbf{x})[1\ 1\ 1]^{\mathsf{T}}$. Also note that the clear-day image of the scene is

$$\mathbf{L}_c(\mathbf{x}) = \frac{\mathbf{L}_\infty \rho(\mathbf{x})}{d(\mathbf{x})^2}\,.$$

The scene appearance is also altered by the airlight itself. The atmospheric particles suspended along the line of sight intercepts the airlight and scatters some portion of it towards the camera, in effect making it act as an additional light source. The amount of airlight scattered towards the viewer is dependent on the unblocked distance along the view direction for each pixel: the scene depth at each pixel. This can be expressed using the airlight formula, also known as the Koschmieder's law (McCartney, 1975):

$$\mathbf{L}_a(\mathbf{x}) = \mathbf{L}_\infty(1 - e^{-\beta d(\mathbf{x})})\,.$$

The scene appearance in a foggy image is the linear combination of the attenuated, reflected scene light and the intercepted airlight:

$$\begin{aligned}\mathbf{I}(\mathbf{x}) &= \mathbf{L}_s + \mathbf{L}_a \\ &= \frac{\mathbf{L}_\infty \rho(\mathbf{x}) e^{-\beta d(\mathbf{x})}}{d(\mathbf{x})^2} + \mathbf{L}_\infty(1 - e^{-\beta d(\mathbf{x})})\,.\end{aligned}$$

Ideally, we would like to directly estimate the true scene albedo $\rho(\mathbf{x})$. This, however, is intractable due to the $d^2$-falloff term. Instead we resort to estimating the scene albedo in the image space by redefining it to include the foreshortening effect arising from the conversion from scene radiance to image irradiance:

$$\tilde{\rho}(\mathbf{x}) = \frac{\rho(\mathbf{x})}{d(\mathbf{x})^2}\,.$$

It is also clear that we can only recover the scaled scene depth: the optical depth $\tilde{d} = \beta d$. If we have prior knowledge of the fog/haze density, we can guess $\beta$ reasonably well and use that value to re-scale the scene depth estimates later.

Our goal then is to factorize the single foggy image $\mathbf{I}$ into the scene albedo $\tilde{\rho}(\mathbf{x})$ and scene depth $\tilde{d}(\mathbf{x})$ for each pixel across the entire image:

$$\mathbf{I}(\mathbf{x}) = \mathbf{L}_\infty \tilde{\rho}(\mathbf{x}) e^{-\tilde{d}(\mathbf{x})} + \mathbf{L}_\infty(1 - e^{-\tilde{d}(\mathbf{x})})\,. \tag{1}$$

The observed input image will inevitably scale the scene radiance with a value that accounts for the radiometric characteristics of the camera[2]. As such, the qualitative albedo and depth we estimate, $\tilde{\rho}$ and $\tilde{d}$, respectively, will encode this unknown global scaling. If necessary, and if we know some ground truth values in the scene, e.g., the reflectance and/or the absolute distance of a scene point a scene point, we may estimate the scaling factors for the estimated albedo and depth images to arrive at quantitatively accurate values.

We can assume that the airlight $\mathbf{L}_\infty$ is known, as it can be easily estimated (semi-)automatically. As pointed out by Narasimhan and Nayar (2002), the observed image color vector $\mathbf{I}$ reduces to the airlight $\mathbf{L}_\infty$ vector for image points corresponding to scene points at infinity $d(\mathbf{x}) = \infty$ or at scene points with no reflection $\tilde{\rho} = 0$ where we also assume $d(\mathbf{x}) = \infty$. This suggests that we can determine $\mathbf{L}_\infty$ by picking image points corresponding to direct observations of the sky or extremely dark scene points, e.g., inside a window as we show in latter experiments.

Even with knowledge of the airlight vector, factorizing the input image into scene albedo and depth is an ill-posed problem. A simple algebraic manipulation of Equation 1 reveals the inherent bilinearity:

$$\ln\left(1 - \frac{\mathbf{I}(\mathbf{x})}{\mathbf{L}_\infty}\right) = \ln(1 - \tilde{\rho}(\mathbf{x})) - \tilde{d}(\mathbf{x}) \quad (2)$$

$$\widetilde{\mathbf{I}}(\mathbf{x}) = \mathbf{C}(\mathbf{x}) + \mathbf{D}(\mathbf{x})\,, \quad (3)$$

where we have explicitly separated the contributions of scene albedo as $\mathbf{C}(\mathbf{x}) = \ln(1 - \tilde{\rho}(\mathbf{x}))$ and depth as $\mathbf{D}(\mathbf{x}) = -\tilde{d}(\mathbf{x})$. Note that we are mixing scalar values and 3-vectors to simplify the notation, but the scalar should be read as 3-vectors of the same element values. It is also worth mentioning that the airlight-normalized input image $\frac{\mathbf{I}}{\mathbf{L}_\infty}$ usually is a 3-vector whose elements are in $(0, 1)$ as the airlight color normally has the brightest intensity in the image. We discuss other cases when this assumption does not hold in Section V-C.

It is clear from Equation 3 that the scene albedo and scene depth are in a bilinear relationship– we can scale up one of them with an arbitrary scalar and scale down the other accordingly to arrive at the same output $\widetilde{\mathbf{I}}$. This clearly means that the factorization of a single foggy image

[2]We assume a linear camera. This assumption does not hold for general images taken with cameras that were not radiometrically calibrated. The nonlinear radiometric camera properties cumulated in its response function will simply be embedded in the estimated scene albedo and depth images.

into scene albedo and depth suffers from this inherent bilinear ambiguity and thus is ill-posed. Past methods have not paid full attention to this bilinear ambiguity and have passively treated it by constraining just one of the two terms, for instance, by forcing locally constant or locally maximized contrast on the scene albedo (or clear-day appearance) (Tan, 2008b; Fattal, 2008b) or by using simple smoothing constraints on the depth (He et al, 2009b). It is worth pointing out that local contrast constraints are not necessary physically meaningful, as there is no reason to believe that the scene albedo or clear-day scene appearance has locally constant or maximized contrast. In that regard, it is important to note that our goal is not contrast enhancement nor visibility restoration, but estimation of scene albedo and depth from a single foggy image.

## IV. A Foggy Image as a Factorial Markov Random Field

The inherent bilinearity indicates that both the scene albedo and depth play equally important roles in the determination of the scene appearance in a foggy image and thus should be treated equally. That is, we should jointly estimate these two variables and must properly constrain their solution spaces to arrive at meaningful estimates. For this, we turn to a probabilistic formulation based on a Factorial Markov Random Field (FMRF) (Kim and Zabih, 2002) to model the dependence between the two variables and the input image canonically and incorporate suitable constraints as statistical priors on each variable.

### *A. Factorial Markov Random Field Formulation*

As depicted in Figure 2, the FMRF formulation consists of a single observation, the foggy image, and two layers of latent variables. We associate one hidden layer with the albedo term $\mathbf{C}$ and the other with the depth $\mathbf{D}$. These two latent layers can be considered statistically independent, since surface color and its geometric structure are not statistically tied to each other. We can then factor the albedo and depth from the foggy image by maximizing the posterior probability

$$p(\mathbf{C}, \mathbf{D}|\widetilde{I}) \propto p(\widetilde{\mathbf{I}}|\mathbf{C}, \mathbf{D})p(\mathbf{C})p(\mathbf{D})\,,$$

where we have dropped the image coordinates as the probabilities are computed over the entire image; $p(\widetilde{\mathbf{I}}|\mathbf{C}, \mathbf{D})$ is the likelihood of the observation; and $p(\mathbf{C})$ and $p(\mathbf{D})$ are the priors. We assume uniform probability for the observation $p(\widetilde{\mathbf{I}})$. The structure of the FMRF permits

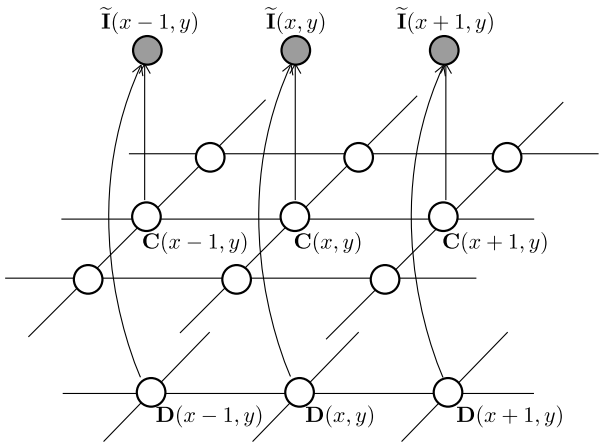



Fig. 2. The Factorial Markov Random Field formulation consists of two layers of latent variables associated with the depth factor $\mathbf{D}(\mathbf{x})$ and albedo factor $\mathbf{C}(\mathbf{x})$, and a single observation layer $\widetilde{\mathbf{I}}(\mathbf{x})$.

constraints to be imposed simultaneously and independently on the scene albedo and depth in the form of priors on each layer.

Although the structure of each latent layer is enforced by the priors imposed on them, they are interrelated through the observation layer. The simultaneous estimation of both factors relies on their interaction, i.e., the likelihood of the observation. We model the noise inherent in the observations with Gaussian distributions of the same variance $\sigma^2$. The likelihood then becomes

$$p(\widetilde{\mathbf{I}}|\mathbf{C},\mathbf{D}) = \prod_{\mathbf{x}} \mathcal{N}(\widetilde{\mathbf{I}}(\mathbf{x})|\mathbf{C}(\mathbf{x}) + \mathbf{D}(\mathbf{x}), \sigma^2)\,. \tag{4}$$

The significance of this factorial graphical model formulation of foggy image formation lies in the fact that we can now impose realistic constraints on the two latent variables, the scene albedo and depth, as statistical priors on the latent layers. This enables us to fully leverage realistic constraints on these variables and seek optimal factorization of the input image with a canonical probabilistic inference algorithm while jointly constraining and estimating the two variables. In this paper, we only consider conventional two-cliques and priors imposed on them as described in Section V-A, but we may also consider higher-order cliques and use appropriate energy minimization techniques (Komodakis and Paragios, 2009).

### *B. Estimating Scene Albedo and Depth*

Following the original inference algorithm for FMRF (Kim and Zabih, 2002), we derive a joint energy minimization algorithm based on the expectation maximization principal to factorize the input image into the scene albedo layer $\mathbf{C}$ and scene depth layer $\mathbf{D}$. The algorithm alternates between the expectation step that computes the posterior probabilities of the latent layers, i.e., estimating the scene albedo and depth, and the maximization step that maximizes the expected log likelihood.

In the expectation step, the posterior probabilities of the latent variables are computed and maximized assuming that the likelihood parameter, the noise variance $\sigma^2$ (Equation 4), is known– they are set to the current estimates from the maximization step. If we use standard exponential potential functions for the priors, maximizing the posterior (Equation 4)

$$\begin{aligned}\underset{\tilde{\rho},\tilde{d}}{\operatorname{argmax}}\, p(\mathbf{C},\mathbf{D}|\widetilde{I}) &= \underset{\tilde{\rho},\tilde{d}}{\operatorname{argmax}}\, p(\widetilde{\mathbf{I}}|\mathbf{C},\mathbf{D})p(\mathbf{C})p(\mathbf{D}) \\ &= \underset{\tilde{\rho},\tilde{d}}{\operatorname{argmax}} \exp\left[-\mathcal{Q}(\widetilde{\mathbf{I}},\mathbf{C},\mathbf{D})\right] \exp\left[-\mathcal{V}_{\mathbf{C}}(\mathbf{C})\right] \exp\left[-\mathcal{V}_{D}(\mathbf{D})\right] ,\end{aligned}$$

becomes equivalent to minimizing the negative log posterior that amounts to an energy minimization:

$$\begin{aligned}&\underset{\tilde{\rho},\tilde{d}}{\operatorname{argmin}} \left(-\ln p(\mathbf{C},\mathbf{D}|\widetilde{\mathbf{I}})\right) = \\ &\underset{\tilde{\rho},\tilde{d}}{\operatorname{argmin}} \sum_{\mathbf{x}} \mathcal{Q}(\widetilde{\mathbf{I}}(\mathbf{x}),\mathbf{D}(\mathbf{x}),\mathbf{C}(\mathbf{x})) + \sum_{\mathbf{x}}\sum_{\mathbf{y}\in\mathsf{N}_{\mathbf{x}}} \mathcal{V}_{\mathbf{C}}(\tilde{\rho}(\mathbf{x}),\tilde{\rho}(\mathbf{y})) + \sum_{\mathbf{x}}\sum_{\mathbf{y}\in\mathsf{N}_{\mathbf{x}}} \mathcal{V}_{\mathbf{D}}(\tilde{d}(\mathbf{x}),\tilde{d}(\mathbf{y})) , \quad (5)\end{aligned}$$

where $\mathsf{N}_{\mathbf{x}}$ is the set of pixels neighboring pixel $\mathbf{x}$, $\mathcal{Q}(\widetilde{\mathbf{I}}(\mathbf{x}),\mathbf{C}(\mathbf{x}),\mathbf{D}(\mathbf{x}))$ is the data energy reflecting the exponent of the Gaussians in the likelihood (Equation 4), and $\mathcal{V}_{\mathbf{C}}(\tilde{\rho}(\mathbf{x}),\tilde{\rho}(\mathbf{y}))$ and $\mathcal{V}_{\mathbf{D}}(\tilde{d}(\mathbf{x}),\tilde{d}(\mathbf{y}))$ are the potential energies imposed by the 2-clique priors for the albedo and depth layers, respectively.

Since the data energy term $\mathcal{Q}(\widetilde{\mathbf{I}}(\mathbf{x}),\mathbf{D}(\mathbf{x}),\mathbf{C}(\mathbf{x}))$ is dependent upon both factors $\mathbf{D}(\mathbf{x})$ and $\mathbf{C}(\mathbf{x})$ we cannot minimize the total energy directly. To efficiently minimize Equation 5, we use the pseudo observable introduced by Kim and Zabih (2002), where each layer is estimated in an alternating fashion by assuming that the other layer's values are known and by minimizing the corresponding partial energy. This amounts to alternating between solving two independent single-layer MRFs while fixing one of the latent variables to the current estimate and estimating

the other. Thus Equation 5 becomes two separate partial energies minimized in alternating orders

$$\sum_{\mathbf{x}} \frac{\widetilde{\mathbf{I}}(\mathbf{x}) - \bar{\mathbf{D}}(\mathbf{x}) - \mathbf{C}(\mathbf{x})}{\bar{\sigma}^2} + \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathsf{N}_{\mathbf{x}}} \mathcal{V}_{\mathbf{C}}(\tilde{\rho}(\mathbf{x}), \tilde{\rho}(\mathbf{y}))\,, \tag{6}$$

$$\sum_{\mathbf{x}} \frac{\widetilde{\mathbf{I}}(\mathbf{x}) - \bar{\mathbf{C}}(\mathbf{x}) - \mathbf{D}(\mathbf{x})}{\bar{\sigma}^2} + \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathsf{N}_{\mathbf{x}}} \mathcal{V}_{\mathbf{D}}(\tilde{d}(\mathbf{x}), \tilde{d}(\mathbf{y}))\,, \tag{7}$$

where $\bar{\mathbf{C}}(\mathbf{x})$ and $\bar{\mathbf{D}}(\mathbf{x})$ are the expected values–the current estimates–of $\mathbf{C}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$, respectively. The likelihood parameter $\sigma$ is also set to the current estimate $\bar{\sigma}$ given the maximization step, which is fixed throughout one expectation step. Since the pseudo-observables depend on the current estimate from the other layer, the expectation step is iterated until convergence. The maximization step is a simple parameter estimation of $\sigma$ given the latent layers with the current estimates of $\tilde{\rho}$ and $\tilde{d}$.

We solve each single-layer MRF in the expectation step, i.e., Equations 7 and 6, using graph-cuts algorithms (Boykov and Kolmogorov, 2004; Boykov et al, 2001; Kolmogorov and Zabih, 2004). The graph-cuts algorithm we use depends on the form of the prior. For instance, when we use a piecewise constant prior (Potts model expressed with a Kronecker delta prior) or a Laplace prior, we use the expansion-move algorithm (Boykov et al, 2001; Szeliski et al, 2008).

## V. Leveraging Latent Structures of the Scene

The factorial Markov random field formulation provides a sound foundation for jointly estimating the scene albedo and depth from a single foggy image. Most important, it allows us to leverage constraints that we may consider on the two latent variables as priors, i.e., clique potentials, to simultaneously tighten their solution spaces. It turns out that despite the veil of fog, the input foggy image encodes significant structural information of the scene albedo and depth that we may extract or identify and use as such statistical priors. Furthermore, we may extract good initial estimates of the scene depth directly from the input image. These scene-specific priors and initial estimates help navigate the optimization to arrive at meaningful estimates–physically-plausible local minima that embody the statistics of real-world scenes.

### *A. Scene-Specific Priors*

One of the widely used natural image statistics that captures the variability of scene color is the gradient distribution. Usually an image of a natural scene exhibits a heavy-tail distribution in its

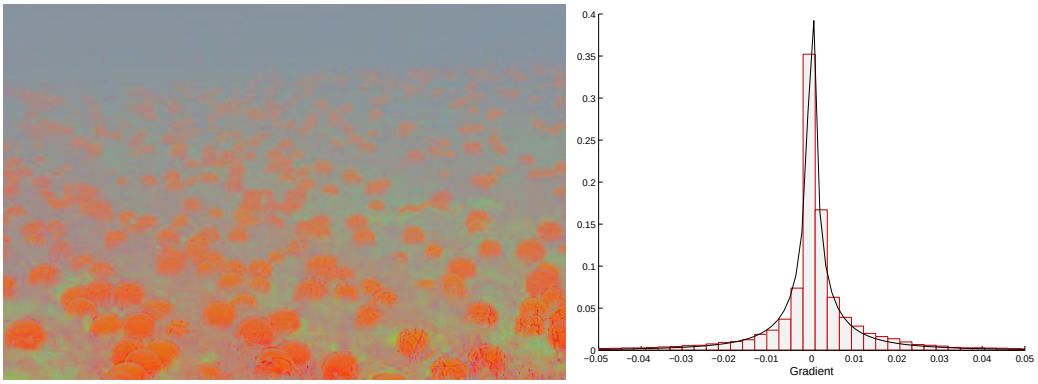
gradient distribution, whose statistical characteristics are often used as statistical priors for image processing tasks such as noise removal. This invaluable images statistics is often approximated with an analytic statistical distribution, as the direct use of non-parametric representations (histograms) incur a prohibitive computational burden. Although various parametric representations can be considered, such as that of the conditional distribution of the gradient components based on wavelet of Fourier bases, a simple analytical model such as an exponential power distributions (generalized Gaussians) is generally sufficient. Then the question is what parameter values should be used for this analytic model to impose a realistic constraint on the albedo layer. Past work that use natural image statistics often learn the parameters from a training image set or simply model them with a Laplace distribution. It is, however, important to notice that the kurtosis of the gradient distribution is usually scene-specific and directly encodes the statistical characteristics of the variability of scene colors. This kurtosis, i.e., the power of the exponential power distribution, is the key to a strong prior.

We notice that the input foggy image contains sufficient information to directly determine the albedo prior parameter values that approximates the clear-day image statistics of the very same scene. Imagine the gradients of a clear-day image of the scene, whose foggy image we have as an input. The fog alters the scene colors by scaling down the original scene albedo color and adding an airlight vector, whose magnitudes are scene depth dependent, $e^{-\bar{d}(\mathbf{x})}$ and $1 - e^{-\bar{d}(\mathbf{x})}$, respectively. Although this results in complex nonlinear changes in the scene color vectors and induces changes in their spatial gradients, it is well known that the orientation of those gradients are insensitive to such illumination changes (Chen et al, 2000). This suggests that the first-order statistics, most important the kurtosis, of the joint distribution of the spatial gradients are preserved as the intensity of the image is altered. To this end, we first calculate the chromaticity of the input image $\mathbf{I}(\mathbf{x}) = \{I_k | k \in \{R, G, B\}\}$

$$i_k(\mathbf{x}) = \frac{I_k(\mathbf{x})}{\sum_{k \in \{R,G,B\}} I_k(\mathbf{x})} ,$$

where $k$ denotes the color channel. We then calculate the gradient $\nabla i_k(\mathbf{x})$, and compute its histogram, which may be viewed as an approximation of the joint distribution of the gradients under a clear-day sky. Figure 3 shows the chromaticity image and its gradient distribution for the foggy pumpkin patch image in Figure 1.

We model the gradient distribution by fitting an exponential power distribution (generalized




(a) (b)

Fig. 3. (a) The chromaticity image of the foggy image of a pumpkin patch (Figure 1) and (b) the distribution of its gradients. A scene-specific albedo prior is estimated by approximating the chromaticity gradients (bar graph) with an exponential power distribution (solid curve) for each color channel.

normal distribution) to the histogram and use it as a prior on the albedo layer:

$$p(\mathbf{C}) = \prod_{\mathbf{x}} \prod_{\mathbf{y} \in \mathsf{N}_{\mathbf{x}}} \exp \frac{|\tilde{\rho}(\mathbf{x}) - \tilde{\rho}(\mathbf{y})|^{\gamma}}{\lambda}, \tag{8}$$

where $\lambda$ and $\gamma$ are the variance and power of the exponential power distribution, respectively. Figure 3(b) shows the exponential power distribution estimated for the corresponding gradient distribution. Note that, although dropped for brevity in Equation 8, the albedo prior is computed and imposed on each color channel independently and the exponential power distribution is computed and imposed as a prior for the mixed distribution of the horizontal and vertical gradients. Also we ignore the normalization factor of the exponential power distribution, as it only results in a constant in the total and partial energies. By estimating the exponential power distribution parameters for the gradient distributions of each channel of the chromaticity image, we may impose priors that are specifically tailored to the scene in the input image, which leads to strong constraints on the joint estimation that are faithful to the scene.

For the scene depth, we can consider a number of different statistical distributions as its prior. For instance, if it varies smoothly, we may impose a Gaussian distribution

$$p(\mathbf{D}) = \prod_{\mathbf{x}} \prod_{\mathbf{y} \in \mathsf{N}_{\mathbf{x}}} \exp \frac{\left(\tilde{d}(\mathbf{x}) - \tilde{d}(\mathbf{y})\right)^2}{\xi}.$$

This is often the case for natural scenes of landscapes such as terrains and fields. Conversely, if the scene consists of planar regions, as is often the case in urban scenes, a piecewise constant model would provide a more appropriate prior on the scene depth:

$$p(\mathbf{D}) = \prod_{\mathbf{x}} \prod_{\mathbf{y} \in \mathsf{N}_{\mathbf{x}}} \delta\left(\tilde{d}(\mathbf{x}) = \tilde{d}(\mathbf{y})\right) ,$$

where $\delta$ is a Kronecker delta. Finally, if the scene contains both characteristics, one may choose to use the Laplace distribution

$$p(\mathbf{D}) = \prod_{\mathbf{x}} \prod_{\mathbf{y} \in \mathsf{N}_{\mathbf{x}}} \exp \frac{\left|\tilde{d}(\mathbf{x}) - \tilde{d}(\mathbf{y})\right|}{\xi} .$$

Ideally, we would like to automatically identify the most suitable depth prior for the input foggy image and its parameter values, i.e., the variances for the Gaussian and Laplace priors. This can potentially be done by running the factorization multiple times with different depth priors with different parameter values and by seeking the combination that achieves the lowest energy. On several images, we indeed tried this and confirmed that this approach works albeit the extra computational cost. We, however, chose to simply manually specify the depth prior and its parameter values in the latter experimental examples. The ability to use depth priors that are most suitable for the scene, which is realized via the graphical model formulation, together with the similarly scene-specific albedo priors, enables reliable factorization.

### *B. Initial Estimates*

The EM-based FMRF estimation algorithm described in Section IV-B can be initialized with arbitrary values assigned to the latent layers of scene albedo and depth and one can start the alternating minimization in the expectation step with either layer. We, however, exploit the fact that the scene depth is allowed to vary less than the albedo to extract good initial depth estimates and start the expectation step with the albedo layer. As can be seen in Equation 1, the albedo $\tilde{\rho}(\mathbf{x})$ is a 3-vector, while the depth $\tilde{d}(\mathbf{x})$ is a single scalar value. In other words, each pixel contains three measurements (the three channels of the observed image, $\mathbf{I}(x)$) that may contribute to the depth estimation, while only one measure for each albedo channel. This simple observation suggests that we may obtain a good bound on the possible scene depth at each pixel from the corresponding observed RGB color values.

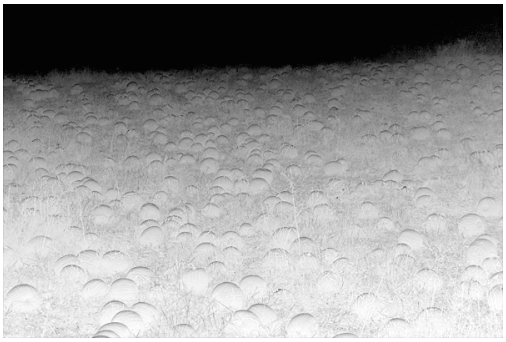


Fig. 4. The initial depth image computed from the pumpkin patch image in Figure 1. The initial estimates already capture the scene depth structure well, and its redundant details arising from scene albedo variations are refined through the joint estimation.

We estimate the initial depth values as the largest possible depth at each pixel. To do so, we note that the scene albedo layer value at each pixel $\mathbf{C}(\mathbf{x})$ is inversely related to its value in the depth layer $\mathbf{D}(\mathbf{x})$ as can be seen in Equation 3. Thus the farthest possible depth value occurs when $\mathbf{C}(\mathbf{x}) = 0$, which is equivalent to $\tilde{\rho}(\mathbf{x}) = 0$, and the corresponding depth estimate (or the negative optical depth $\tilde{d}(\mathbf{x})$) is

$$\mathbf{D}(\mathbf{x}) = \widetilde{\mathbf{I}}(\mathbf{x})\,.$$

The observed image contains three color channels, thus we may obtain a single depth estimate for each channel. Since the depth estimate is the largest possible depth the channel's observed intensity allows, i.e., the depth when the albedo is $0$, we set the initial depth estimate $\mathbf{D}_0(\mathbf{x})$ to the farthest possible depth over all three channels

$$\mathbf{D}_0(\mathbf{x}) = \max_{k \in \{R,G,B\}} \widetilde{\mathbf{I}}_k(\mathbf{x})\,.$$

In other words, we find the upper bound on the depth value at each pixel from the corresponding, observed RGB color values and use it as the initial estimate of the depth layer. Note that, since $-\infty \leq \mathbf{D}(\mathbf{x}) \leq 0$, the *maximum* value is in fact the *closest* depth to the viewer.

Figure 4 shows the initial depth estimates computed from the foggy, pumpkin patch image in Figure 1. As Figure 4 demonstrates, these initial estimates already capture the scene depth structure well and provide a good starting point for the FMRF-based factorization. We initialize the depth layer with $\mathbf{D}_0$ and use it as the pseudo-observable (current estimate) to start the expectation step by first solving the albedo MRF.

Recent independent work (He et al, 2009b) refer to a similar estimate of the depth as the dark channel prior. The dark channel prior, however, is computed in local regions resulting in blocky estimates, while our initial depth estimates are computed pixel-wise resulting in finer estimates. We have also shown that these initial estimates naturally fall out as upper bounds on the depth estimates, not mere empirical observations as discussed in (He et al, 2009b). It is also worth emphasizing that these are initial estimates of the depth, not priors that enforce the constraints in the estimation process. The dark channel "priors" in (He et al, 2009b) are also only used as initial estimates for the matting algorithm that imposes conventional smoothing on the depth estimation (but without any constraints on the albedo).

### *C. Scaling and Discretization*

Note that the left-hand side of Equation 3 is undefined when $\mathbf{I}(\mathbf{x}) \geq \mathbf{L}_\infty$ which can sometimes occur in real images. For instance, specularities can cause scene colors to exceed the intensity of airlight. To alleviate the numerical issues in such cases, we normalize the airlight to a unit vector, and scale $\mathbf{I}(\mathbf{x})$ by an arbitrary value $\eta$ to redefine $\widetilde{\mathbf{I}}(\mathbf{x})$, $\mathbf{C}(\mathbf{x})$, and $\mathbf{D}(\mathbf{x})$ in Equation 3:

$$\ln\left(1 - \frac{\mathbf{I}(\mathbf{x})|\mathbf{L}_\infty|}{\eta \mathbf{L}_\infty}\right) = \ln(1 - \tilde{\rho}(\mathbf{x})) - \tilde{d}(\mathbf{x})\,. \tag{9}$$

We select $\eta$ to minimize instances where $\widetilde{\mathbf{I}}(\mathbf{x})$ is not real. To do so, we select $\eta$ to be the largest intensity (color vector magnitude) in the input image $\mathbf{I}(\mathbf{x})$

$$\eta = \max_{\mathbf{x}} |\mathbf{I}(\mathbf{x})|\,.$$

Even with this scaling, however, there may still be instances where $\widetilde{\mathbf{I}}(\mathbf{x})$ is not real. This happens rarely, and we set the resulting pixel value to the airlight color.

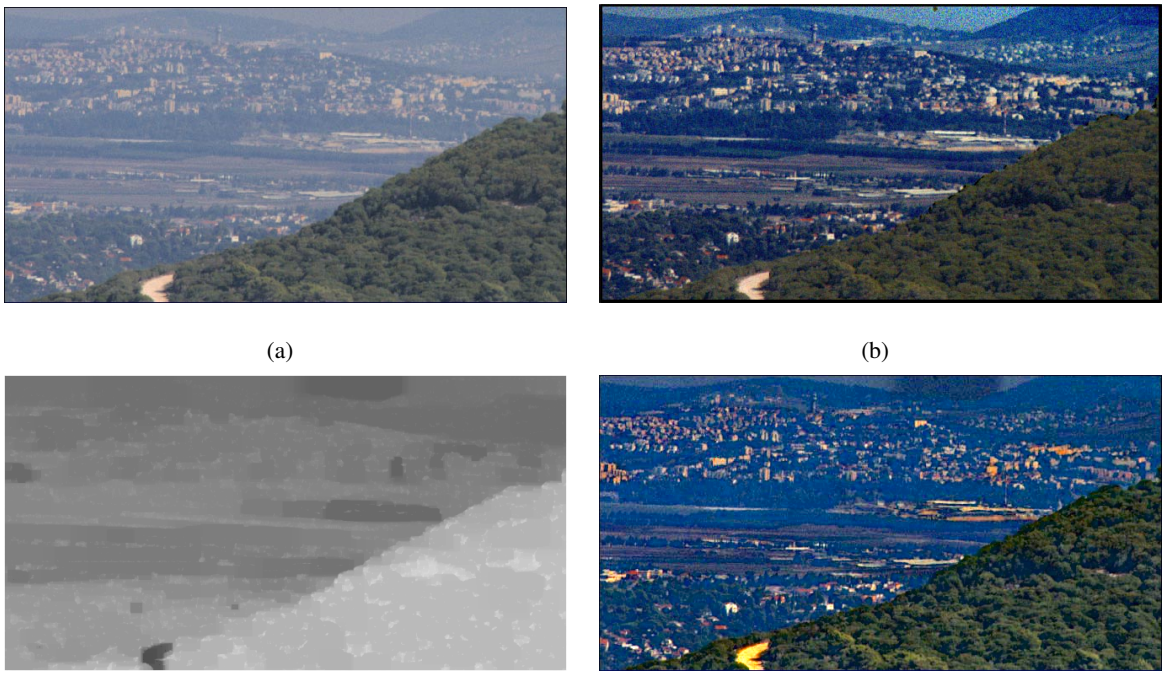
We assume that our source images are represented with 8-bits per channel, and discretize our albedo estimates to $|\mathbf{C}| = 256$ possible values. From Equation 9 we can see that $\mathbf{C}(\mathbf{x})$ is logarithmically related to $\rho(\mathbf{x})$. Thus the values for $\mathbf{D}(\mathbf{x})$ that minimize the error will also be in log scale. We discretize $\mathbf{D}(\mathbf{x})$ to $256$ possible labels. Specifically, given a label $D_L(\mathbf{x})$ between $0$ and $255$ the depth is

$$\mathbf{D}(\mathbf{x}) = \ln\left(\frac{D_L(\mathbf{x})}{255}\right)\,. \tag{10}$$

## VI. Experimental Results

We evaluated the effectiveness of our method on a number of real and synthetic foggy images. For each image, we estimated the exponential power distribution for each of the three chromaticity channels. For the depth prior, we used a Laplace, Gaussian, or piecewise constant distribution, which was manually picked depending on the input image. As the depth value of a scene point $\bar{d}(\mathbf{x})$ tends to infinity, i.e., for smaller $\mathbf{D}$ labels, its albedo values $\bar{\rho}(\mathbf{x})$ approaches the airlight vector and thus the computation of $\mathbf{C}$ becomes numerically unstable as it approaches $\ln 0$. To avoid this numerical issue, we simply limit the maximum depth value to the label of $D_L = 40$ in Equation 10 and, for those with smaller depth labels, recompute the albedo values using the depth value corresponding to $D_L = 40$. For all experimental results, we visualize the scene depth with a gray scale image in which the brighter the closer (the intensity is inverse proportional to $\bar{d}$) and the albedo values are shown with airlight multiplied to them that corresponds to the clear-day image (Equation 1) but without the $d^2$-falloff. The albedo image is also gamma corrected for better visualization–naturally the estimated albedo values lie in the lower end of the 8-bit intensity range of an image as a result of the factorization and are thus generally dark.

Figure 5 shows the results of our method compared with a polarization-based method that requires minimum two images (Schechner et al, 2003). The method by Schechner et al (2003) uses the polarization characteristics of scene reflection and airlight to physically remove their attenuation and transmittance, thus we may assume that their results represent the true clear-day scene. The results in Figure 5 show that our method computes similar albedo estimates (again, shown as an image with airlight vector multiplied to them) that demonstrate the accuracy of our method. The airlight color that was automatically chosen from the brightest point in the image was bluish, which in turn makes the airlight-normalized image yellowish, and thus the resulting albedo image (clear-day image) have a yellowish tint (Figure 5(d)). Note that areas that have the same color as the airlight will have an insolvable ambiguity as the albedo and airlight color vectors line up in the color space, and thus will be estimated as a scene point at infinity with airlight color albedo. This causes the erroneous depth estimates, such as the trail in the front hill (Figure 5(c)). Otherwise, the results have almost exactly the same visibility and scene color estimates as the physically-based method, which demonstrates the accuracy of the joint estimation.



(a) (b)

(c) (d)

Fig. 5. Factorization result of a real-world hazy skyline compared with that of a polarization-based method by Schechner et al (2003). (a) the input hazy image; (b) the dehazed clear-day image computed by Schechner et al (2003); (c) the depth image computed by our method; and (d) the albedo (clear-day) image computed by our method. Note that the polarization-based method requires two images. The polarization-based method computes a physically accurate clear-day image. The albedo image of our method achieves similar factorization that demonstrates its accuracy. Images (a) and (b) are reprinted from (Schechner, 2003).

Figure 6 shows the factorization results of a synthetically generated foggy image. The results of our method are compared with those computed with the method introduced by Fattal (2008b). Although at first glance the albedo and depth estimates by Fattal (2008b) (Figure 6(c)) may seem closer to the ground truth (Figure 6(b)) they suffer from inaccuracies that indicate remaining mixing of scene albedo and depth. These are visible in the somewhat shallow contrast of the albedo image, as if some fog still veils the image, and the depth estimates contaminated by the albedo (texture). Our results (Figure 6(d)), in contrast, show clear factorization of the scene albedo and depth, where the two are not contaminated by each other. Our method directly estimates the image albedo which are the true colors as if there were no airlight at all. This explains the increase in contrast–the more vivid albedo image–of our results. The clear albedo

(a) (b) (c) (d)

Fig. 6. Factorization results of a synthetic foggy image. (a) The input image; (b) the ground truth albedo (left) and depth images (right); (c) the estimated albedo (left) and depth images (right) computed with the algorithm by Fattal (2008b); (d) the estimated albedo (right) and depth images (left) computed with our method. Fattal's algorithm results in fog-veiled albedo image and texture-contaminated depth image (c) indicating remaining mixing of albedo and depth effects. In contrast, our algorithm successfully factorizes the input image into clear and vivid albedo and smooth and texture-free depth images (d).

and depth estimates in our result (Figure 6(d)) exemplifies the significance of achieving a joint estimation while canonically constraining both latent variables, unlike constraining and solving for just one of the variables (albedo in Fattal (2008b)).

Figure 7 shows a direct comparison of our approach to the method by Fattal (2008b) using real-world foggy images[3]. Our scene-specific albedo priors produce better consistency in the color of the recovered albedo, resulting in a more uniform appearance across the varying geometric structures of the scene. This, for example, can be seen in the green colors across the field and orange colors of different pumpkins in Figure 7(a), the wheat colors across the field in Figure 7(b), and the colors of the bricks and different greens in Figure 7(c). In all cases, the albedo estimates by Fattal (2008b) show color variations indicating remaining fog effects, e.g., the brighter green in the middle of the pumpkin patch, the darker wheat colors at far distance, and grayish tint to the bricks around the bird house on the house wall and the greens at different distances. In contrast, our albedo images are vivid and consistent in colors across the entire image. These characteristics agree well with the synthetic image results in Figure 6.

Similarly, and as an interrelated consequence, our approach estimates the depth variation of the scene more accurately. This can be observed in the globally smooth variation towards the top of the image for the examples in Figure 7(a) and 7(b), as well as the robustness against highlights observed on the pumpkins at grazing angles. Also, the estimation by Fattal (2008b)

[3]For the input image in Figure 7(c), we used the inside of the window to determine the airlight vector. Again, for regions with scene colors similar to airlight, e.g., the window frame, the depth estimates will be saturated.

lead to overly far depth values for further scene areas for the pumpkin patch and wheat field (Figure 7(a) and 7(b), respectively), and for the house exterior (Figure 7(c)), and lacks the details for these regions altogether. In sharp contrast, our method estimates depth at finer granularity even in these areas (for instance, observe the wheat bundles at the top region of our depth image in Figure 7(b) and the walls of the house in our depth image in Figure 7(c)). These errors in the depth estimates also explain the fog-like veil in the albedo images computed by Fattal (2008b) and the more accurate and consistent albedo estimates of our method. These results demonstrate the importance of constraining both latent variables with scene-specific priors, which is enabled with our Bayesian graphical model formulation.

Figure 8 shows a few comparisons of the factorization results by our method with those of He et al (2009b). The estimated albedo and depth images are more or less the same. As we discussed in Section V-B, He et al (2009b) independently introduced a depth initialization method, although they refer to it as a prior, that is similar to ours but computed block-wise unlike our pixel-wise initialization. In sharp contrast to their matting based method that refines the depth image, we achieve factorization via joint estimation with true statistical priors (constraints) imposed on both the albedo and depth values. The importance of this contribution in comparison to He et al (2009b) becomes particularly significant when defogging challenging scenes like that in Figure 8(c), where the scene colors nor the scene depth alone does not embody strong structures–there is not much variability individually and a mere smoothing constraint on one does not suffice (He et al, 2009b). Clearly, our method achieves more accurate removal of the fog (Figure 8(c) right), while fog persists in the albedo estimates by He et al (2009b) (Figure 8(c)). Similar observations can be made between our results on the images used by Fattal (2008b) (Figure 7) and those computed by He et al (2009b) shown in (He et al, 2009a). These results demonstrate the effectiveness of constraining both latent variables with scene-specific priors and jointly solving for them, which effectively compensates for each other's lack of discriminative information to successfully arrive at an accurate factorization.

Figure 9 shows comparison of our method with another recent defogging method (Kopf et al, 2008b). Except for the differences in the overall tint of colors, which are again due to the automatically chosen airlight vector in our method, the estimated albedo images are of similar quality–they exhibit similar visibility and local contrast–demonstrating the effectiveness of our method. The method by Kopf et al (2008b) require a semi-automatically aligned geo-referenced

3D terrain model to obtain the depth of the scene[4]. We, on the other hand, jointly estimate the scene depth as shown in the last column for each example in Figure 9. Except for the erroneous depth estimates for regions that happen to have albedo colors similar to the airlight, e.g., the clouds in Figure 9(b) and ground for the hikers in Figure 9(c), the depth image agree well with the scene structure.

In Figure 10, we compare our results with those by Tan (2008b). Since each pixel is modeled as an observation in the FMRF model, our approach can isolate fine edges between different depth objects with much finer detail, as visible in the lack of block artifacts the approach by Tan (2008b) suffers from. For instance, observe the tree trunk in Figure 10(b). Our method achieves much greater consistency in the colors of the recovered albedo layer when compared with the input images. Our scene-specific priors enforce a strict structure for each color layer ensuring that the resulting albedo value is consistent with the initial observation. For instance, the street sign in Figure 10(a) should be yellow as our method successfully estimates, while the local contrast maximization method by Tan (2008b) estimates it to be orange.

## VII. Conclusion

In this paper, we presented a novel probabilistic method for factorizing a single image of a foggy scene into its albedo and depth values. We formulated this problem as joint estimation of scene albedo and depth with energy minimization of a factorial Markov random field, enabling full exploitation of natural image and depth statistics in the form of scene-specific priors. The experimental results demonstrate superior accuracy over past methods, especially on challenging scenes where only constraining and solving for a single latent variable does not suffice. We believe the overall framework we have introduced for this particular problem of defogging would also be useful for tackling other computer vision problems that suffer from inherent bilinearity.

## Acknowledgements

This work was supported in part by National Science Foundation CAREER Award IIS-0746717.

[4]The focus of their method is rather on the different applications of such 3D-aligned images, and defogging is just one example.

## REFERENCES

Boykov Y, Kolmogorov V (2004) An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. IEEE Trans on Pattern Analysis and Machine Intelligence 26(9):1124–1137

Boykov Y, Veksler O, Zabih R (2001) Fast Approximate Energy Minimization via Graph Cuts. IEEE Trans on Pattern Analysis and Machine Intelligence 23(11):1222–1239

Chen H, Belhumeur P, Jacobs D (2000) In Search of Illumination Invariants. In: Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition, vol 1, pp 254–261

Fattal R (2008a) http://www.cs.huji.ac.il/ raananf/projects/defog/index.html

Fattal R (2008b) Single Image Dehazing. In: Proc. of ACM SIGGRAPH, vol 27, pp 1–9

Grewe L, Brooks R (1998) Atmospheric Attenuation Reduction Through Multi-Sensor Fusion. In: Proc. of SPIE Sensor Fusion: Architectures, Algorithms, and Applications II, vol 3376, pp 102–109

He K, Sun J, Tang X (2009a) http://personal.ie.cuhk.edu.hk/ hkm007/cvpr09/

He K, Sun J, Tang X (2009b) Single Image Haze Removal Using Dark Channel Prior. In: Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition, vol 1

Kim J, Zabih R (2002) Factorial Markov Random Fields. In: Proc. of European Conf on Comp. Vision, pp 321–334

Klinker G, Shafer S, Kanade T (1990) A Physical Approach to Color Image Understanding. International Journal of Computer Vision 4:7–38

Kolmogorov V, Zabih R (2004) What Energy Functions can be Minimized via Graph Cuts? IEEE Trans on Pattern Analysis and Machine Intelligence 26(2):147–159

Komodakis N, Paragios N (2009) Beyond Pairwise Energies: Efficient Optimization for Higher-Order MRFs. In: Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition, pp 2985–2992

Kopf J, Neubert B, Chen B, Cohen M, Cohen-Or D, Deussen O, Uyttendaele M, Lischinski D (2008a) http://johanneskopf.de/publications/deep_photo/dehazing/index.html

Kopf J, Neubert B, Chen B, Cohen M, Cohen-Or D, Deussen O, Uyttendaele M, Lischinski D (2008b) Deep Photo: Model-Based Photograph Enhancement and Viewing. ACM Transactions on Graphics 27(5):116:1–116:10

McCartney E (1975) Optics of the Atmosphere: Scattering by Molecules and Particles, John Wiley and Sons: New York

Middleton W (1952) Vision Through the Atmosphere, University of Toronto Press

Narasimhan S, Nayar S (2003a) Shedding Light on the Weather. In: Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition, pp 665–672

Narasimhan SG, Nayar SK (2002) Vision and The Atmosphere. Int'l Journal on Comp Vision 48(3):233–254

Narasimhan SG, Nayar SK (2003b) Contrast Restoration of Weather Degraded Images. IEEE Trans on Pattern Analysis and Machine Intelligence 25(1):713–724

Narasimhan SG, Nayar SK (2003c) Interactive (De)weathering of An Image Using Physical Models. In: ICCV Workshop on Color and Photometric Methods in Comp. Vision (CPMCV)

Oakley J, Satherley B (1998) Improving Image Quality in Poor Visibility Conditions Using a Physical Model for Contrast Degradation. IEEE Trans on Image Processing 7(2):167–179

Schechner Y (2003) http://webee.technion.ac.il/ yoav/research/instant-dehazing.html

Schechner Y, Narasimhan S, Nayar S (2003) Polarization-Based Vision Through Haze. Applied Optics 42(3):511–525

Shafer S (1985) Using Color to Separate Reflection Components. COLOR Research and Application 10(4):210–218

Szeliski R, Zabih R, Scharstein D, O Veksler VK, Agarwala A, Tappen M, Rother C (2008) A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors. IEEE Trans on Pattern Analysis and Machine Intelligence 30(6):1068–1080

Tan K, Oakley J (2001) Physics-based Approach to Color Image Enhancement in Poor Visibility Conditions. Journal of Optical Society America A 18(10):2460–2467

Tan R (2008a) http://people.cs.uu.nl/robby/fog/index.html

Tan R (2008b) Visibility in Bad Weather from a Single Image. In: Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition, pp 1–8

Tarel JP, Hautière N (2009) Fast Visibility Restoration from a Single Color or Gray Level Image. In: Proc. of IEEE Int'l Conf on Comp. Vision

Tominaga S, Wandell B (1989) Standard surface-reflectance model and illuminant estimation. JOpt Soc Am A 6(4):576–584

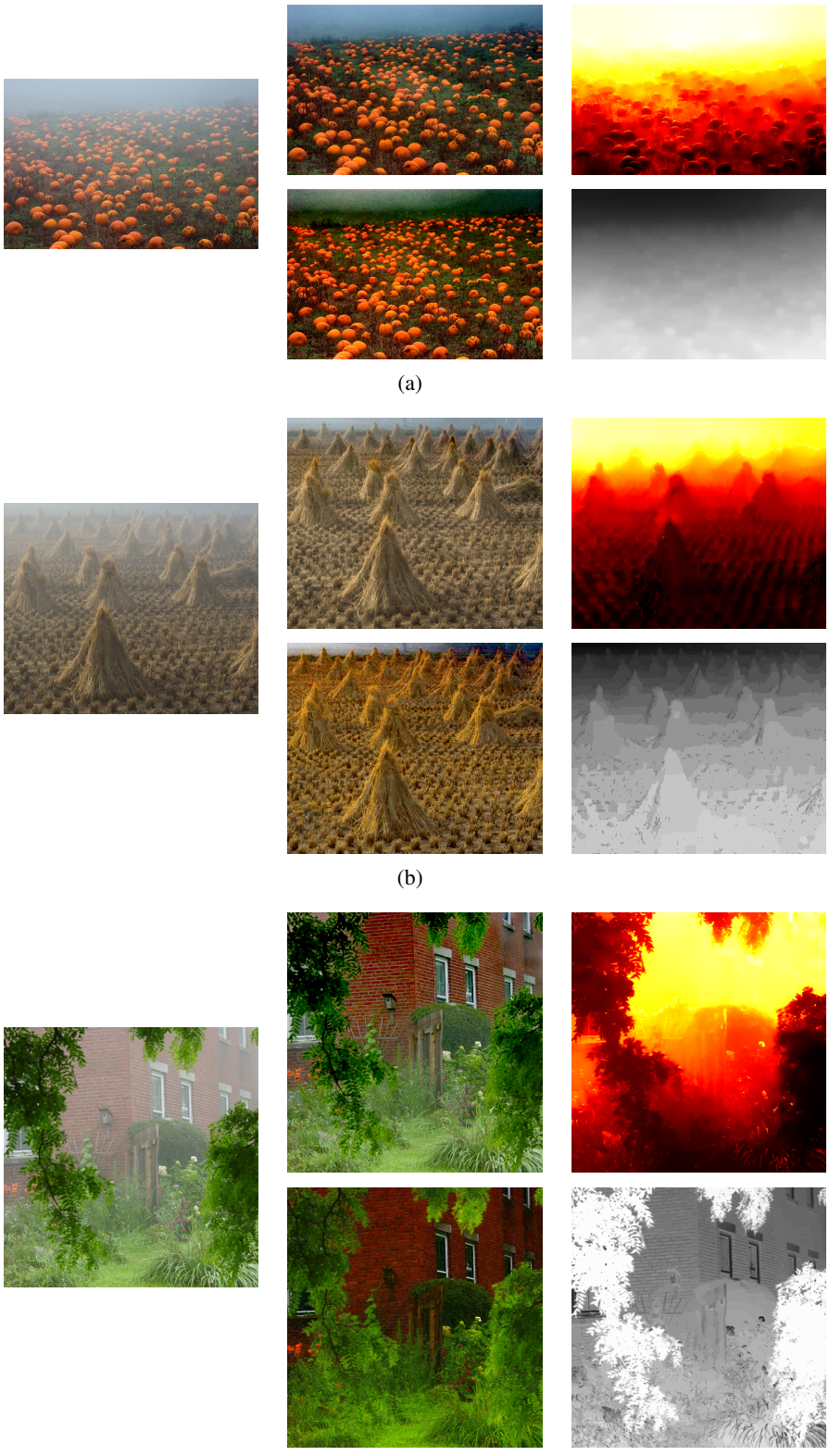


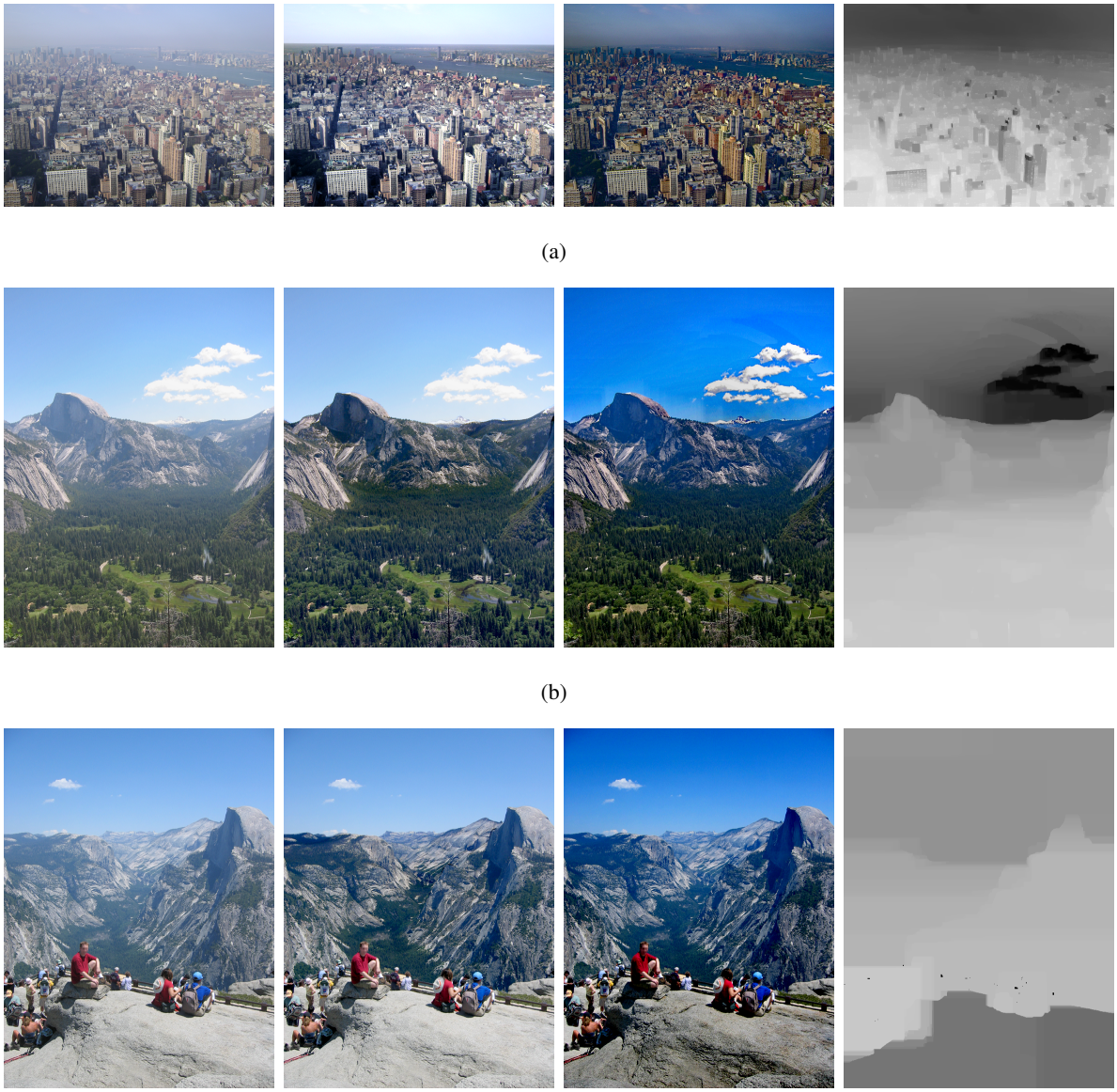
Fig. 7. Factorization results of real-world foggy images, (a) a pumpkin patch, (b) a wheat field, and (c) a house, compared with those of another single image defogging method by Fattal (2008b). For each image, in the columns, the input foggy image (left), the estimated albedo image (middle), and the estimated depth image (right) are shown. In the rows, the results by Fattal (2008b) (top), and by our method (bottom) are shown. Our results demonstrate more accurate factorization of scene albedo and depth, resulting in more vivid and fog-free albedo images and more detailed depth images. The input images and the corresponding results by Fattal (2008b) are reprinted from (Fattal, 2008a).

(a)

(b)

(c)

Fig. 8. Factorization results of real-world foggy images, of (a) an aerial city view, (b) an urban view, and (c) a cliff, compared with those of another method by He et al (2009b). For the example in (a), in the columns, the input foggy image (left), the estimated albedo image (middle), and the estimated depth image (right) are shown; and in the rows, the results by He et al (2009b) (top), and by our method (bottom) are shown. For the examples in (b) and (c), the input image, the albedo image by He et al (2009b), and the albedo image by our method are shown from left to right, respectively. Overall the factorization results are of similar quality, but our method achieves clearer defogging on particularly challenging scenes like that in (c). The input images and the corresponding results by He et al (2009b) are reprinted from (He et al, 2009a).



(a)

(b)

(c)

Fig. 9. Factorization results of real-world foggy images, of (a) a view of Manhattan, and (b) (c) two scenic views of the Half Dome, compared with those of another method by Kopf et al (2008b). For each example the input image, the albedo image by Kopf et al (2008b), the albedo image and the depth image computed by our method are shown from left to right, respectively. The color differences in the estimated albedo images are due to the particular selection of the airlight vector by our method. Otherwise, overall the factorization results are of similar quality. Most important, our method jointly estimates the scene albedo and depth, while the method by Kopf et al (2008b) requires a semi-automatically aligned 3D terrain model. The input images and corresponding results by Kopf et al (2008b) are reprinted from (Kopf et al, 2008a).

(a)

(b)

Fig. 10. Factorization results of real-world foggy images, of (a) a street and (b) a lake, compared with those of a contrast enhancement method by Tan (2008b). For each example, the input image, the albedo image by Tan (2008b), the albedo image and the depth image computed by our method are shown from left to right, respectively. Again, the color differences in the estimated albedo images are due to the particular selection of the airlight vector by our method. Our results do not suffer from the block artifacts observed in Tan's results, e.g., at the occluding boundary of tree trunk in (b), as the FMRF formulation and inference achieve pixel-wise factorization. The input images and corresponding results by Tan (2008b) are reprinted from (Tan, 2008a).