Factorizing Scene Albedo and Depth from a Single Foggy Image

Louis Kratz   Ko Nishino

Department of Computer Science, Drexel University

Philadelphia, PA

{lak24, kon}@drexel.edu

Abstract

Atmospheric conditions induced by suspended particles, such as fog and haze, severely degrade image quality. Restoring the true scene colors (clear day image) from a single image of a weather-degraded scene remains a challenging task due to the inherent ambiguity between scene albedo and depth. In this paper, we introduce a novel probabilistic method that fully leverages natural statistics of both the albedo and depth of the scene to resolve this ambiguity. 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. We show that we may exploit natural image and depth statistics as priors on these hidden layers and factorize a single foggy image via a canonical Expectation Maximization algorithm with alternating minimization. Experimental results show that the proposed method achieves more accurate restoration compared to state-of-the-art methods that focus on only recovering scene albedo or depth individually.

1. Introduction

Poor weather conditions caused by suspended particles in the air, such as fog and haze, may significantly reduce the contrast and distort the colors of the scene, resulting in a severely degraded image. Discerning content in such images poses significant challenges even for human observers and can cause computer vision algorithms for scene analysis to fail miserably. Fortunately, the physical interaction of light rays with such particles, i.e. scattering, is well understood: the image intensities of a foggy image may be written as a function of the corresponding original scene color (albedo) and depth [10][1].

Restoring the true scene color from a single weather-degraded image is, however, an inherently ill-posed problem that cannot be analytically solved unless one of the two constituents, either the scene depth or colors, is known.

---

[1]In this paper we assume that single scattering dominates the weather effect on the image.

This is due to the inherent bilinearlity between scene albedo and depth as we make explicit in this paper. To this end, as in any other ill-posed computer vision problem we encounter, past work has focused on imposing additional constraints to resolve the ambiguity. In particular, recent work focuses on estimating the scene albedo by significantly constraining their values to maximize contrast [15] or to be decorrelated from depth-induced effects (transmission) within constant albedo regions [4]. In these work, the scene depth is rather a byproduct that can be computed once the scene albedo is estimated using a physically-based scattering image model. Due to their formulation these methods result in overly contrast-stretched images [15] and inaccurate restoration of the color or depth [4]. A recent independent work [5] leverages an empirical observation, that within local regions the scene albedo values are very small in at least one color channel, to constrain the depth variation. The method still suffers from the ambiguity between color and depth leading to inaccurate restoration.

In this paper, we derive a novel method for "defogging" a single image by factorizing the image into scene albedo and depth. The key insight underlying our method is that both the scene albedo and scene depth convey valuable structural information to the resulting foggy image and thus these two factors should be jointly estimated. Due to the inherent ambiguity between the scene albedo and depth, such joint estimation necessitates a canonical method that can seamlessly incorporate additional constraints on each of the factors. To this end, we derive a canonical probabilistic formulation based on a Factorial Markov Random Field (FMRF) [6]. The image generation is modeled as an FMRF with a single observation, the foggy image, and two statistically independent latent layers that represent the scene albedo (clear day image) and the scene depth. With this formulation, we may impose natural statistics priors, such as a heavy-tail prior on the gradients of the scene albedo and piecewise constant prior on the depth. We show that we may obtain a maximum a posteriori (MAP) estimate through an Expectation Maximization algorithm tailored to the specific FMRF formulation. We also show that we may directly extract useful

albedo priors, i.e., scene-specific priors, from the given single foggy image and incorporate them into the estimation.

We experimentally evaluate the effectiveness of the method by showing the factorization results on a number of images. The results show the method achieves more physically plausible estimates of both the scene albedo and depth, resulting in more accurate decomposition of the image.

2. Related Work

Early approaches to defogging weather-degraded images have focused on achieving the image decomposition with additional observations or scene information. For instance, methods in [3, 11] require multiple exposures of the same scene to estimate parameter values related to the atmospheric particles or the scene depth. Active methods require the scene to be photographed using different imaging conditions such as the degree of polarization [13] or require manually specified depth values [12]. A more recent work leverages scene depth directly provided by georeferenced digital terrain or city models [8]. Although additional scene information provides sound constraints to resolve the ambiguity between albedo and depth or makes one of them (depth) known, such information requires additional equipment or conditions, e.g., imaging in different fog densities, that may not be available.

Recent work that operate on single images impose constraints solely upon either the albedo or depth, but not directly on both. Tan [15] imposes a locally constant constraint on the albedo channel to increase the contrast of the image. The method is not meant to restore the actual scene albedo but rather just enhance the visibility, and thus does not lead to physically meaningful estimates of the albedo or depth. Fattal [4] imposes locally constant constraints of albedo values together with decorrelation of the transmission in local areas, and then estimates the depth value from the result. The method does not constrain the scene's depth structure, but only the scene albedo. Similarly recent independent work by He et al. [5] imposes constraints only on the depth structure induced by an empirical observation made on the possible values of scene albedo within a local region. We will show the similarity of this constraint to the initial estimation of (not the prior on) depth values in our method.

By estimating the depth and albedo simultaneously through a canonical probabilistic formulation, our approach can impose structural constraints on both the depth and albedo values. Past work on single image defogging [4, 5, 15] 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 work exploits natural structural information extracted from the input image itself, i.e. creating scene-specific priors on the scene albedo, which has not been leveraged in the past.

3. Factorial Formulation

Suspended atmospheric particles cause scattering of direct light from the air and its reflection from the scene surfaces. The amount of attenuation caused by this scattering is proportional to the distance from the observer. As a result, the observation, i.e., a foggy image, becomes a function of the true scene color and depth.

Under realistic conditions, in particular, where Allard's law that describes the distance dependent exponential decay of reflected light transmission and the light sources modeled as aggregated *airlight* that exponentially increases along the path length it travels (see [10] for a detailed derivation with a complete review of the literature), the imaged colors in a single foggy image can be modeled as a function of the corresponding scene albedo $\rho(\mathbf{x})$ (a vector of 3 color channels) and depth $d(\mathbf{x})$

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

In this equation, β is the attenuation coefficient that we assume to be uniform across the entire scene, that is we assume spatially homogeneous fog, $\mathbf{L}_\infty$ is the airlight vector of 3 color channels, and $\mathbf{x}$ is the 2D image coordinates. To be more precise, the depth is shared among all color channels and thus $d(\mathbf{x})$ should be considered as $d(\mathbf{x})[1\ 1\ 1]^\mathsf{T}$, making it a 3-vector. Note that when $d(\mathbf{x}) = \infty$, Equation 1 reduces to $\mathbf{I}(\mathbf{x})|_{d(\mathbf{x})=\infty} = \mathbf{L}_\infty$ indicating that we directly observe the airlight $\mathbf{L}_\infty$. In practice, this means that we may obtain a good estimate of $\mathbf{L}_\infty$ by picking an image color at locations where no objects occlude the sky. We may also assume the attenuation coefficient β to be known as it is simply a multiplicative constant on the depth [3].

Estimating the scene albedo $\rho(\mathbf{x})$ and depth $d(\mathbf{x})$ for each scene point corresponding to individual image coordinates from a single foggy image is an inherently ill-posed problem. We may see this in a more explicit form through simple algebraic rearrangements of Equation 1:

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

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

where we have explicitly separated the contributions of scene albedo as $\mathbf{C}(\mathbf{x}) = \ln(\rho(\mathbf{x}) - 1)$ and depth as $\mathbf{D}(\mathbf{x}) = -\beta d(\mathbf{x})^2$. The left hand side $\widetilde{\mathbf{I}}(\mathbf{x})$ denotes the log of *airlight-normalized* input image offset by 1. As made explicit in Equation 3, estimating the scene albedo and depth by estimating $\mathbf{C}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$ at each image coordinate is a completely ill-posed problem that suffers from the bilinearity of those two factors. Reaching meaningful estimates of these two factors from the single observation requires additional constraints on the solution space. We achieve this by

---

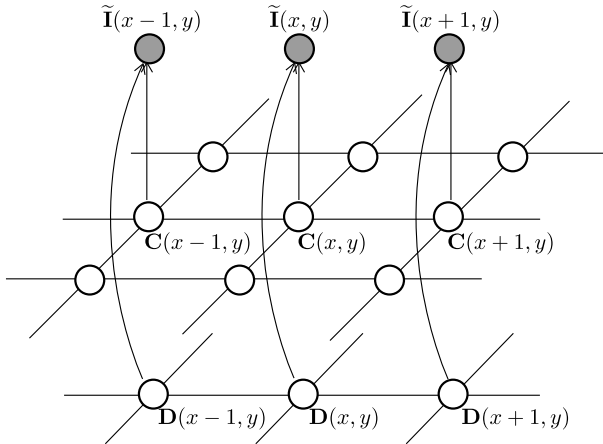[2] Again, this is a 3-vector with shared depth values.

Figure 1. 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 $\mathbf{C}(\mathbf{x})$.

imposing realistic constraints on both the scene albedo and depth (and thus on $\mathbf{C}$ and $\mathbf{D}$) and decompose the image into these two factors by jointly estimating them.

In order to canonically incorporate constraints on each factor, we construct a Factorial Markov Random Field [6] as depicted in Fig. 1. 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

$$p(\mathbf{C}, \mathbf{D}|\widetilde{I}) \propto p(\widetilde{\mathbf{I}}|\mathbf{C}, \mathbf{D})p(\mathbf{C})p(\mathbf{D}) , \qquad (4)$$

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

Note that, though the structure of each underlying 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. In order to retain the structure imposed on the factors through the priors, and to model the noise inherent in the observations, we model the likelihood as a mixture of Gaussian distributions with spatially uniform variance and weights

$$p(\widetilde{\mathbf{I}}|\mathbf{C}, \mathbf{D}) = \prod_{\mathbf{x}} \mathcal{N}(\widetilde{\mathbf{I}}(\mathbf{x})|\mathbf{C}(\mathbf{x}) + \mathbf{D}(\mathbf{x}), \sigma^2) , \qquad (5)$$

where σ^2 is the variance of the noise.

Note that in this factorial framework, we may impose any type of priors on the scene albedo and depth independently, which enables us to fully leverage any realistic constraints on these factors. The prior on the albedo layer can reflect the natural image statistics, in particular the heavy-tail distribution of its gradients. As such we may model it as a Laplace distribution. The prior on the depth layer can reflect the geometric structure of the scene, which in real-world scenes often exhibits sparse gradients similar to natural image statistics [9] or piecewise constant layering, each of which can be modeled with a Laplace distribution or a delta function (Potts model), respectively. These priors will accurately capture the geometric structures of urban settings where the scene is composed of planar surfaces. For continuously varying scene geometry, we may impose a Gaussian prior on the depth layer. The significance of the factorial formulation lies in the fact we may incorporate any of such priors in a canonical manner.

4. Factorial MRF Estimation

The estimation of the albedo $\rho(x)$ and depth $d(x)$ corresponds to maximizing the posterior probability given in Equation 4. Direct maximization of this equation is intractable due to the interdependence of the hidden variables. In order to estimate the latent layers, we adopt and tailor the Expectation Maximization (EM) approach proposed by Kim and Zabih [6]. This approach corresponds to alternatively minimizing two Markov random fields that are tied together by their current estimates and the generative model of the observations.

The general EM approach alternates between computing the posterior probabilities of the latent variables and maximizing the expected log likelihood of the distribution parameters. Thus in the expectation step we estimate the albedo and depth values assuming a known variance σ^2. This is identical to minimizing the log energy of

$$\sum_{\mathbf{x}} \mathcal{Q}(\widetilde{\mathbf{I}}(\mathbf{x}), \mathbf{D}(\mathbf{x}), \mathbf{C}(\mathbf{x})) + \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathsf{N}_{\mathbf{x}}} \mathcal{V}_{\mathbf{D}}(d(\mathbf{x}), d(\mathbf{y}))$$
$$+ \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathsf{N}_{\mathbf{x}}} \mathcal{V}_{\mathbf{C}}(\rho(\mathbf{x}), \rho(\mathbf{y})) , \quad (6)$$

where $\mathsf{N}_{\mathbf{x}}$ is the set of pixels neighboring pixel $\mathbf{x}$, $\mathcal{Q}(\widetilde{\mathbf{I}}(\mathbf{x}), \mathbf{D}(\mathbf{x}), \mathbf{C}(\mathbf{x}))$ is the data energy reflecting the likelihood, $\mathcal{V}_{\mathbf{D}}(d(\mathbf{x}), d(\mathbf{y}))$ is the potential energy for the depth imposed by the prior on it, and $\mathcal{V}_{\mathbf{C}}(\rho(\mathbf{x}), \rho(\mathbf{y}))$ is the potential energy for the albedo imposed by its prior.

Since the data energy 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 6, we use the pseudo observable [6], where each layer will be estimated in an alternating fashion by minimizing the corresponding
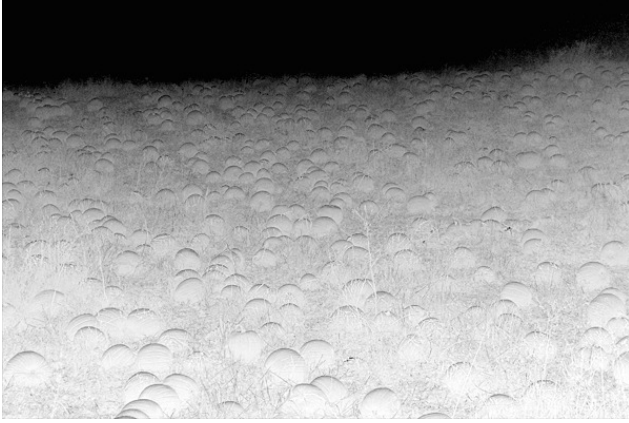
1703

Figure 2. An initial depth estimate for a pumpkin field image. Redundant structure due to albedo intensity variations, such as the specular highlights on the pumpkins, is removed during the FMRF minimization.
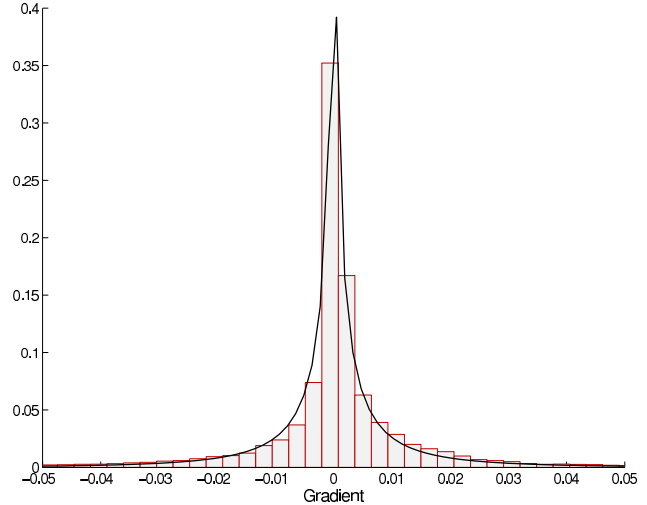

Figure 3. A scene-specific albedo prior is estimated by approximating the chromaticity gradients with an exponential power distribution for each color channel.

partial energies assuming that the other layer's values are known. Thus Equation 6 becomes two separate partial energies

$$\sum_{\mathbf{x}} \mathcal{Q}(\widetilde{\mathbf{I}}(\mathbf{x}), \mathbf{D}(\mathbf{x}), \bar{\mathbf{C}}(\mathbf{x})) + \sum_{\mathbf{x}} \sum_{\mathbf{y} \in N_{\mathbf{x}}} \mathcal{V}_{\mathbf{D}}(d(\mathbf{x}), d(\mathbf{y})) \quad (7)$$

$$\sum_{\mathbf{x}} \mathcal{Q}(\widetilde{\mathbf{I}}(\mathbf{x}), \bar{\mathbf{D}}(\mathbf{x}), \mathbf{C}(\mathbf{x})) + \sum_{\mathbf{x}} \sum_{\mathbf{y} \in N_{\mathbf{x}}} \mathcal{V}_{\mathbf{C}}(\rho(\mathbf{x}), \rho(\mathbf{y})) , \quad (8)$$

where $\bar{\mathbf{C}}(\mathbf{x})$ and $\bar{\mathbf{D}}(\mathbf{x})$ are the expected values of $\mathbf{C}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$, respectively. We estimate the expected values as the maximum a posterior values of the hidden variables, that is, their current estimate. Although more flexible estimations are possible using the priors [6], we found that there is little improvement for the increased computational cost.

The pseudo-observables for the depth layer are simply $\widetilde{\mathbf{I}}(\mathbf{x}) - \bar{\mathbf{C}}(\mathbf{x})$, and similarly for the albedo layer they are $\widetilde{\mathbf{I}}(\mathbf{x}) - \bar{\mathbf{D}}(\mathbf{x})$. We solve Equations 7 and 8 using well-established energy-minimization techniques [14]. In particular, we use derivatives of graph-cuts [1, 2, 7], depending on the functional forms of the priors. Since the pseudo-observables depend on the current estimate from the other layer, the expectation step is iterated until convergence. The maximization step consists of estimating σ^2 using the maximum likelihood estimator.

We treat each channel of the image as statistically independent and estimate their albedos separately. The depth values, however, are shared across all of the channels and we use all three channels of the image when estimating the depth to exploit as much information as possible.

Energy minimization techniques generally assume that the values of the hidden variables are discrete, as continuous values vastly increase the computational cost. We assume that our source images are represented by 8-bits per channel, and thus we discretize our albedo estimate to

$|\mathbf{C}| = 256$ possible values. The number of depth values $|\mathbf{D}|$ depends on the scene. We found that a sufficient representation rarely requires more than 100 possible values.

The convergence time of the EM algorithm depends on the initial guesses of the albedo and depth values. While random initializations result in the same minimization [6], we exploit a general property of our formulation to initialize the depth values. Note that, in Equation 1, the maximum depth value $d(\mathbf{x}) = \infty$ occurs when $\mathbf{I}(\mathbf{x}) = \mathbf{L}_\infty$. When $\mathbf{I}(\mathbf{x}) \neq \mathbf{L}_\infty$, the maximum possible distance of point $\mathbf{x}$ occurs when $\rho(\mathbf{x})$ is farthest from the airlight, or when $\rho(\mathbf{x}) = 0$. We choose the initial depth $\widehat{d}(\mathbf{x})$ to be

$$\widehat{d}(\mathbf{x}) = \min_k \frac{\widetilde{\mathbf{I}}_k(\mathbf{x})}{-\beta} , \quad (9)$$

where k is the red, green, and blue color channels. In other words, we initialize $d(\mathbf{x})$ to the farthest possible value that the observation allows. This proves to be an effective estimate, as illustrated in Fig. 2. Recent independent work [5] refer to a similar estimate of the (optical) depth as the dark channel prior. The dark channel prior is computed in local regions resulting in blocky estimates, while our initial depth estimates are computed pixelwise. Note that these are simply initially estimates which are then refined by the EM algorithm to accurately factorize albedo and depth layers.

The EM algorithm using the pseudo-observables provides an efficient and effective process for jointly estimating the foggy image factors. The ill-posed nature of Equation 3 suggests multiple local minima, and thus we rely on the priors $p(\mathbf{D})$ and $p(\mathbf{C})$ to steer the minimization to physically-plausible local minimum. Our use of the FMRF model allows for maximum flexibility in incorporating these priors, as the depth and albedo statistics vary depending on the

captured scene. In fact, the only restriction on the form of the priors is imposed by the energy-minimization technique employed [7].

5. Scene-Specific Priors

The FMRF model provides a general method for factorizing a single foggy image into its scene albedo and depth by leveraging natural constraints imposed on the latent layers. For instance, again, depth variation of real-world scenes can be naturally modeled with piecewise constant or smooth priors using delta or Gaussian functions, respectively. Although surface color variation will obey the natural image statistics, well-captured in the heavy-tail distribution of its gradients often modeled with Laplace distributions, the actual gradient distribution can vary fairly significantly depending on the scene in the image. As such, determining the exact parameter values to incorporate these constraints on the gradient distribution as a prior can become a tedious task.

We observe that the statistics of the image gradients for a clear day image (the albedo image) can be approximated directly from the input image. By converting the foggy original input image into a chromaticity image, we may minimize the effect of fog on the gradients, given the fact that gradient orientations are fairly invariant to illumination and the chromaticity normalizes the magnitudes of the color vectors in the image. To this end, we first calculate the chromaticity of the image for each channel:

$$i_k(\mathbf{x}) = \frac{I_k(\mathbf{x})}{I_r(\mathbf{x}) + I_g(\mathbf{x}) + I_b(\mathbf{x})} \ , \qquad (10)$$

where k indicates the channel of interest and r, g, b indicate the red, green, and blue channels, respectively. 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. An example distribution is shown in Fig. 3.

To accurately capture the gradient distribution, we model it with an exponential power distribution, since its shape may not align with a simple Laplace distribution depending on the scene, and use it as the prior on the albedo layer:

$$p(\mathbf{C}) = \prod_{\mathbf{x}} \prod_{\mathbf{y} \in \mathsf{N}_{\mathbf{x}}} \exp \frac{|\rho(\mathbf{x}) - \rho(\mathbf{y})|^\gamma}{\lambda} \ , \qquad (11)$$

where λ is the variance parameter and γ relates to the kurtosis of the distribution. Note that we ignore the scaling factor of the exponential power distribution, which normalizes the distribution, since it only results in a constant in the total and partial energies. By estimating these parameters and fitting Equation 11 to 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.

6. Results

We evaluated our method on six real-world images degraded by fog. For each image, we estimated the exponential power distribution for each of the three chromaticity channels and empirically set the attenuation coefficient to an appropriate value[3]. For the depth prior, we used a Gaussian distribution for smoothly varying scene structure and a delta function for planar scenes. Although we set these empirically, we may simply run the method and pick the instance that produces better quantitative results by examining the residual errors after the minimization. For all of the images, we used 100 depth labels and 256 possible albedo labels.

Fig. 4 shows the results of our method compared with a polarization-based method that requires minimum two images [10]. An attenuation coefficient of 1.75 and a delta depth prior are used. The polarization result reflects the removal of actual physical attenuation and thus we consider it to be the ground truth clear day image. Our method directly estimates the image albedo which is not exactly the clear day image but rather the true colors as if there were no airlight at all. This discrepancy explains the increase in contrast of our results. The overall consistency between our estimates and the polarization-based method demonstrates the accuracy of our method. Readers may also compare this result with Fig. 7 in [4] to confirm the higher accuracy of our method against other methods.

A direct comparison of our approach to that of Fattal [4] is shown in Fig. 5. We used an attenuation coefficient of 2 together with a Gaussian depth prior. Our scene-specific albedo priors produce superior consistency in the color of the recovered albedo, resulting in a more uniform appearance across the varying geometric structures of the scene (compare the green colors across the field and orange colors of different pumpkins). As shown in Fig. 5(d), our results do not suffer from artifacts seen in the results of [4]. Our approach estimates the depth variation of the scene more accurately, as can be seen in its globally smooth variation towards the top of the image as well as its robustness against highlights observed on the pumpkin surface at grazing angles[4].

Fig. 6 shows the results on a scene with shallow depth and sharp geometric objects. For this image, we used a delta depth prior to identify the piecewise constant structures in the scene. As highlighted in Fig. 6(d), our approach recovers finer details of objects closer to the viewer, such as the leaves, and farther away, such as the facade of the house, that are not recovered with the method in [4]. Due

---

[3]Note that this only scales the depth labels with a constant value and we may instead set it to one and estimate the optical depth.

[4]This results in overly far estimates of depth towards the top parts of the pumpkins in [4], whereas our method correctly estimates the depth of a single pumpkin to be more or less uniform compared to the global variation of the depth of the field.
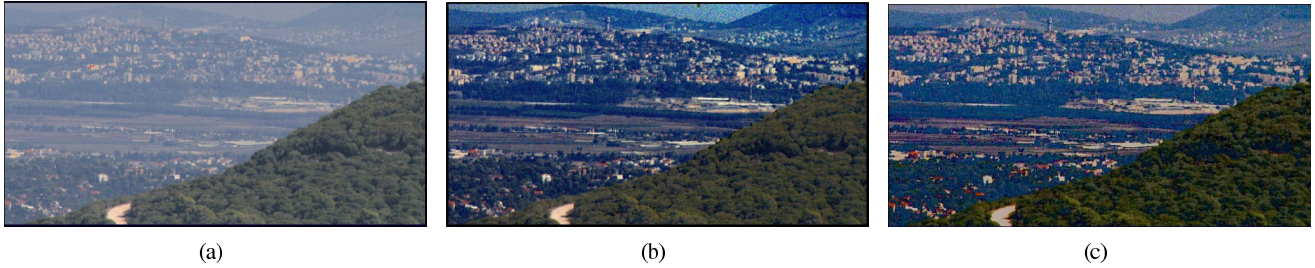
(a)   (b)   (c)

Figure 4. A real-world hazy skyline (a), the result using polarization-based [10] removal (b), and our estimated albedo (c). Our method estimates the scene albedo, resulting in higher saturation when compared to the result of the polarization-based method.

(a)   (b)   (c)
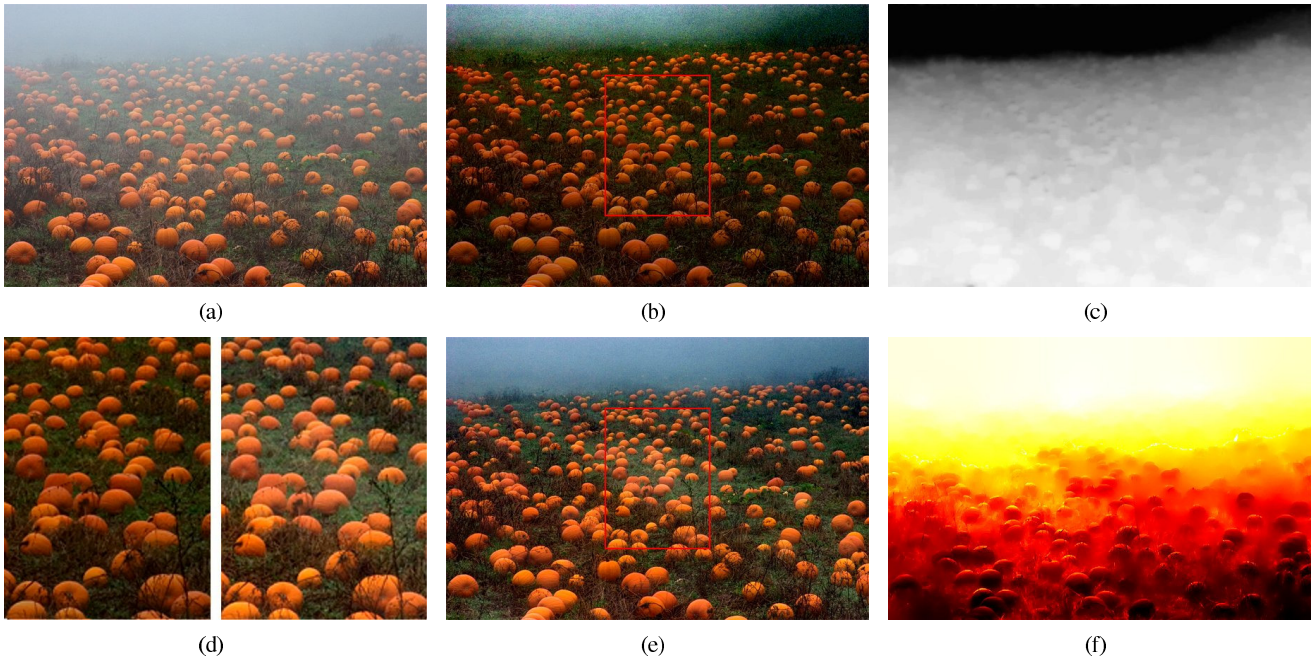
(d)   (e)   (f)

Figure 5. The results of factorizing a foggy image of a pumpkin patch (a) into the albedo (b) and depth images (c) compared with Fattal [4] in (e) and (f). Our approach produces more accurate decomposition, completely removing the atmospheric interference. Observe in the inset (d) the saturation in the pumpkin and green patch colors in the albedo estimates of [4] (right) indicating remaining fog. Our method estimates consistent colors for the same region indicating higher accuracy (left).

to the inherent ambiguity in regions with colors similar to the airlight, some portions of this scene have incorrect depth estimates, such as the window frames. These regions will have saturated depth estimates and can be easily detected for post-processing, if necessary [5].

A comparison of our approach with that of He et al. [5] is shown in Fig. 7. Our approach recovers significantly more detail, especially in dense fog areas such as the right rock face and distant trees. The deeper colors in the rocks and trees reveal a more accurate restoration of scene albedo. The recovered image appears dark due to the density of the fog, an effect removed from the results of [5] by re-exposure.

In Fig. 8 and Fig. 9, we compare our results with the results of Tan [15]. We used an attenuation coefficient of 1.2 and a Gaussian depth prior. 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 in [15] suffers from (for instance, observe the tree branches in Fig. 8). Our method achieves much greater consistency in the colors of the recovered albedo layer when compared with the original image as shown in Fig. 9. Our scene-specific priors enforce a strict structure for each color layer ensuring that the resulting albedo value is consistent with the initial observation (the street sign should be yellow as our method successfully estimates, while the contrast maximization method estimates it to be orange).

---

[5]For instance, we may discard the depth estimates in these regions and simply fill in from surrounding estimates to arrive at better (but hallucinated) albedo and depth estimates.
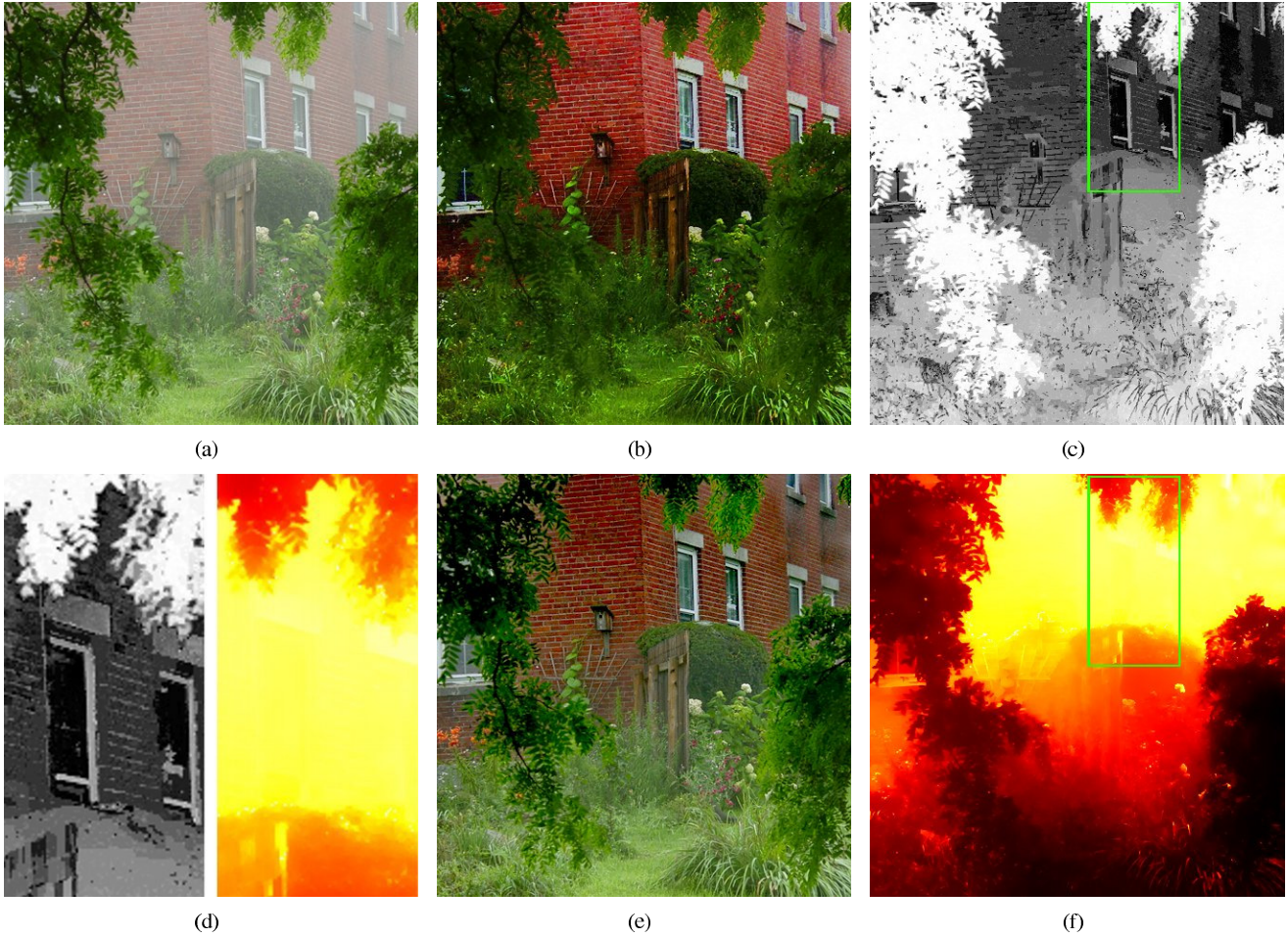
Figure 6. The results of factorizing a foggy image of a yard (a) into an albedo image (b) and a depth image (c) compared with the results from Fattal [4] (e) and (f). Our approach identifies a greater level of detail in the depth (d). Observe that we recover the geometric structure of the leaves and the facade of the house, both of which the results in [4] miss.
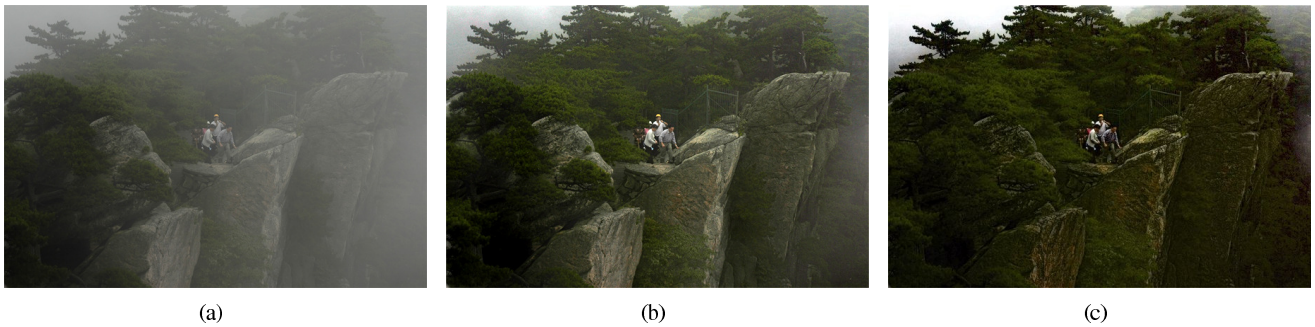

Figure 7. A foggy mountain scene (a) processed by He et al. [5] (b) and our approach (c). Our approach recovers significantly more detail around dense fog areas such as the rock face on the right.

7. 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 an energy minimization of a factorial Markov random field, enabling full exploitation of natural image and depth statis-tics in the form of scene-specific priors. The experimen-tal results demonstrate superior accuracy to state-of-the-art methods for single image defogging, resulting in improved depth reconstruction and consistency in the recovered col-ors. Currently, we are investigating the possibility of con-structing scene-specific depth priors to further improve the

(a)	(b)	(c)

Figure 8. A foggy lake scene (a) with results from contrast enhancement [15] (b) and our factorization method (c). Observe that due to our pixel-level method, our results do not suffer from blocky artifacts as in [15] caused by enforcing piecewise contrast maximization.

(a)	(b)	(c)

Figure 9. A foggy street scene (a) processed with contrast enhancement [15] (b) and our factorization method (c). Our results retain a greater consistency in the recovered colors when compared to the original image.

decomposition.

Acknowledgement

References

[1] Y. Boykov and V. Kolmogorov. 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, Sep. 2004.

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov. 2001.

[3] F. Cozman and E. Krotkov. Depth from scattering. In *Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition*, pages 801–806, 1997.

[4] R. Fattal. Single image dehazing. In *Proc. of ACM SIGGRAPH*, volume 27, pages 1–9, 2008.

[5] X. T. K. He, J. Sun. Single image haze removal using dark channel prior. In *Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition*, volume 1, 2009.

[6] J. Kim and R. Zabih. Factorial markov random fields. In *Proc. of European Conf on Comp. Vision*, pages 321–334, 2002.

[7] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(2):147–159, Feb. 2004.

[8] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski. Deep photo: Model-based photograph enhancement and viewing. *ACM Transactions on Graphics*, 27(5):116:1–116:10, 2008.

[9] J. H. Lee and D. Mumford. Statistics of range images. In *Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition*, volume 1, pages 324–331, 2000.

[10] S. G. Narasimhan and S. K. Nayar. Vision and the atmosphere. *Int'l Journal on Comp. Vision*, 48(3):233–254, Jul. 2002.

[11] S. G. Narasimhan and S. K. Nayar. Contrast restoration of weather degraded images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(1):713–724, June 2003.

[12] S. G. Narasimhan and S. K. Nayar. Interactive (de)weathering of an image using physical models. In *ICCV Workshop on Color and Photometric Methods in Comp. Vision (CPMCV)*, 2003.

[13] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar. Instant dehazing of images using polarization. In *Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition*, volume 1, pages 325–332, June 2001.

[14] R. Szeliski, R. Zabih, D. Scharstein, V. K. O. Veksler, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields. In *Proc. of European Conf on Comp. Vision*, volume 2, pages 16–29, 2006.

[15] R. T. Tan. Visibility in bad weather from a single image. In *Proc. of IEEE Int'l Conf on Comp. Vision and Pattern Recognition*, pages 1–8, June 2008.