Single Image Dehazing

Raanan Fattal*
Hebrew University of Jerusalem, Israel

Figure 1: Dehazing based on a single input image and the corresponding depth estimate.

Abstract

In this paper we present a new method for estimating the optical transmission in hazy scenes given a single input image. Based on this estimation, the scattered light is eliminated to increase scene visibility and recover haze-free scene contrasts. In this new approach we formulate a refined image formation model that accounts for surface shading in addition to the transmission function. This allows us to resolve ambiguities in the data by searching for a solution in which the resulting shading and transmission functions are locally statistically uncorrelated. A similar principle is used to estimate the color of the haze. Results demonstrate the new method abilities to remove the haze layer as well as provide a reliable transmission estimate which can be used for additional applications such as image refocusing and novel view synthesis.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms; I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Radiometry

Keywords: image dehazing/defogging, computational photography, image restoration, image enhancement, Markov random field image modeling

1 Introduction

In almost every practical scenario the light reflected from a surface is scattered in the atmosphere before it reaches the camera. This

*e-mail: raananf@cs.huji.ac.il

is due to the presence of aerosols such as dust, mist, and fumes which deflect light from its original course of propagation. In long distance photography or foggy scenes, this process has a substantial effect on the image in which contrasts are reduced and surface colors become faint. Such degraded photographs often lack visual vividness and appeal, and moreover, they offer a poor visibility of the scene contents. This effect may be an annoyance to amateur, commercial, and artistic photographers as well as undermine the quality of underwater and aerial photography. This may also be the case for satellite imaging which is used for many purposes including cartography and web mapping, land-use planning, archeology, and environmental studies.

As we shall describe shortly in more detail, in this process light, which should have propagated in straight lines, is scattered and replaced by previously scattered light, called the *airlight* [Koschmieder 1924]. This results in a multiplicative loss of image contrasts as well as an *additive* term due to this uniform light. In Section 3 we describe the model that is commonly used to formalize the image formation in the presence of haze. In this model, the degraded image is factored into a sum of two components: the airlight contribution and the unknown surface radiance. Algebraically these two, three-channel color vectors, are convexly combined by the *transmission* coefficient which is a scalar specifying the visibility at each pixel. Recovering a haze-free image requires us to determine the three surface color values as well as the transmission value at every pixel. Since the input image provides us three equations per pixel, the system is ambiguous and we cannot determine the transmission values. In Section 3 we give a formal description of this ambiguity, but intuitively it follows from our inability to answer the following question based on a single image: are we looking at a deep red surface through a thick white medium, or is it a faint red surface seen at a close range or through a clear medium. In the general case this ambiguity, which we refer to as the *airlight-albedo* ambiguity, holds for every pixel and can not be resolved independently at each pixel given a *single* input image.

In this paper we present a new method for recovering a haze-free image given a single photograph as an input. We achieve this by interpreting the image through a model that accounts for surface shading in addition to the scene transmission. Based on this refined image formation model, the image is broken into regions of a constant albedo and the airlight-albedo ambiguity is resolved by de-

riving an additional constraint that requires the surface shading and medium transmission functions to be locally statistically uncorrelated. This requires the shading component to vary significantly compared to the noise present in the image. We use a graphical model to propagate the solution to pixels in which the signal-to-noise ratio falls below an admissible level that we derive analytically in the Appendix. The airlight color is also estimated using this uncorrelation principle. This new method is *passive*; it does not require multiple images of the scene, any light-blocking based polarization, any form of scene depth information, or any specialized sensors or hardware. The new method has the minimal requirement of a *single* image acquired by an ordinary consumer camera. Also it does not assume the haze layer to be smooth in space, i.e., discontinuities in the scene depth or medium thickness are permitted. As shown in Figure 1, despite the challenges this problem poses, this new method achieves a significant reduction of the airlight and restores the contrasts of complex scenes. Based on the recovered transmission values we can estimate scene depths and use them for other applications that we describe in the Section 8.

This papers is organized as follows. We begin by reviewing existing works on image restoration and haze removal. In Section 3 we present the image degradation model due to the presence of haze in the scene, and in Section 4 we present the core idea behind our new approach for the restricted case of images consisting of a single albedo. We then extend our solution to images with multi-albedo surfaces in Section 6, and report the results in Section 8 as well as compare it with alternative methods. In Section 9 we summarize our approach and discuss its limitations.

2 Previous Work

In the context of computational photography there is an increasing focus on developing methods that restore images as well as extracting other quantities at minimal requirements in terms of input data, user intervention, and sophistication of the acquisition hardware. Examples are recovery of an all-focus image and depth map using a simple modification to the camera's aperture in [Levin et al. 2007]. A similar modification is used in [Veeraraghavan et al. 2007] to reconstruct the 4D light field of a scene from a 2D camera. Given two images, one noisy and the other blurry, a deblurring method with a reduced amount of ringing artifacts is described in [Yuan et al. 2007]. Resolution enhancement with native-resolution edge sharpness based on a single input image is described in [Fattal 2007]. In [Liu et al. 2006] intensity-dependent noise levels are estimated from a single image using Bayesian inference.

Image dehazing is a very challenging problem and most of the papers addressing it assume some form of additional data on top of the degraded photograph itself. In [Tan and Oakley 2000] assuming the scene depth is given, atmospheric effects are removed from terrain images taken by a forward-looking airborne camera. In [Schechner et al. 2001] polarized haze effects are removed given two photographs. The camera must be identically positioned in the scene and an attached polarization filter is set to a different angle for each photograph. This gives images that differ only in the magnitude of the polarized haze light component. Using some estimate for the *degree of polarization*, a parameter describing this difference in magnitudes, the polarized haze light is removed. In [Shwartz et al. 2006] this parameter is estimated automatically by assuming that the higher spatial-bands of the direct transmission, the surface radiance reaching the camera, and the polarized haze contribution are uncorrelated. We use a similar but more refined principle to separate the image into different components. These methods remove the polarized component of the haze light and provide impressive results. However in situations of fog or dense haze the polarized light is not the major source of the degradation and may also be too weak as to undermine the stability of these methods. In [Schechner and Averbuch 2007] the authors describe a regularization mecha-

nism, based on the transmission, for suppressing the noise amplification involved with dehazing. A user interactive tool for removing weather effects is described in [Narasimhan and Nayar 2003]. This method requires the user to indicate regions that are heavily affected by weather and ones that are not, or to provide some coarse depth information. In [Nayar and Narasimhan 1999] the scene structure is estimated from multiple images of the scene with and without haze effects under the assumption that the surface radiance is unchanged.

In [Oakley and Bu 2007] the airlight is assumed to be constant over the entire image and is estimated given a single image. This is done based on the observation that in natural images the local sample mean of pixel intensities is proportional to the standard deviation. In a very recent work [Tan 2008] image contrasts are restored from a single input image by maximizing the contrasts of the direct transmission while assuming a smooth layer of airlight. This method generates compelling results with enhanced scene contrasts, yet may produce some halos near depth discontinuities in scene.

Atmospheric haze effects also appear in environmental photography based on remote sensing systems. A multi-spectral imaging sensor called the Thematic Mapper is installed on the Landsats satellites and captures six bands of Earth's reflected light. The resulting images are often contaminated by the presence of semi-transparent clouds and layers of aerosol that degrade the quality of these readings. Several image-based strategies are proposed to remove these effects. The *dark-object subtraction* [Chavez 1988] method subtracts a constant value, corresponding the darkest object in the scene, from each band. These values are determined according to the offsets in the intensity histograms and are picked manually. This method also assumes a uniform haze layer across the image. In [Zhang et al. 2002] this process is automated and refined by calculating a *haze-optimized transform* based on two of the bands that are particularly sensitive to the presence of haze. In [Du et al. 2002] haze effects are assumed to reside in the lower part of the spatial spectrum and are eliminated by replacing the data in this part of the spectrum with one taken from a reference haze-free image.

General contrast enhancement can be obtained by tonemapping techniques. One family of such operators depends only on pixels values and ignores the spatial relations. This includes linear mapping, histogram stretching and equalization, and gamma correction, which are all commonly found in standard commercial image editing software. A more sophisticated tone reproduction operator is described in [Larson et al. 1997] in the context of rendering high-dynamic range images. In general scenes, the optical thickness of haze varies across the image and affects the values differently at each pixel. Since these methods perform the same operation across the entire image, they are limited in their ability to remove the the haze effect. Contrast enhancement that amplifies local variations in intensity can be found in different techniques such as the Laplacian pyramid [Rahman et al. 1996], wavelet decomposition [Lu and Jr. 1994], single scale unsharp-mask filter [Wikipedia 2007], and the multi-scale bilateral filter [Fattal et al. 2007]. As mentioned earlier and discusses below, the haze effect is both multiplicative as well as additive since the pixels are averaged together with a constant, the airlight. This additive offset is not properly canceled by these procedures which amplify high-band image components in a multiplicative manner.

Photographic filters are optical accessories inserted into the optical path of the camera. They can be used to reduce haze effects as they block the polarized sunlight reflected by air molecules and other small dust particles. In case of moderately thick media the electric field is re-randomized due to multiple scattering of the light limiting the effect of these filters [Schechner et al. 2001].

3 Image Degradation Model

Light passing through a scattering medium is attenuated along its original course and is distributed to other directions. This process is commonly modeled mathematically by assuming that along short distances there is a linear relation between the fraction of light deflected and the distance traveled. More formally, along infinitesimally short distances $d\mathbf{r}$ the fraction of light absorbed is given by $\beta d\mathbf{r}$ where β is the medium extinction coefficient due to light scattering. Integrating this process along a ray emerging from the viewer, in the case of a spatially varying β, gives

$$t = \exp\left(-\int_0^d \beta(\mathbf{r}(s))ds\right), \qquad (1)$$

where $\mathbf{r}$ is an arc-length parametrization of the ray. The fraction t is called the *transmission* and expresses the relative portion of light that managed to survive the entire path between the observer and a surface point in the scene, at $\mathbf{r}(d)$, without being scattered. In the absence of black-body radiation the process of light scattering conserves energy, meaning that the fraction of light scattered from any particular direction is replaced by the same fraction of light scattered from all other directions. The equation that expresses this conservation law is known as the *Radiative Transport Equation* [Rossum and Nieuwenhuizen 1999]. Assuming that this added light is dominated by light that underwent multiple scattering events, allows us to approximate it as being both isotropic and uniform in space. This constant light, known as the *airlight* [Koschmieder 1924] or also as the *veiling light*, can be used to approximate the true in-scattering term in the full radiative transport equation to achieve the following simpler image formation model

$$I(\mathbf{x}) = t(\mathbf{x})J(\mathbf{x}) + (1-t(\mathbf{x}))A, \qquad (2)$$

where this equation is defined on the three RGB color channels. I stands for the observed image, A is the airlight color vector, J is the surface radiance vector at the intersection point of the scene and the real-world ray corresponding to the pixel $\mathbf{x} = (x,y)$, and $t(\mathbf{x})$ is the transmission along that ray. This degradation model is commonly used to describe the image formation in the presence of haze [Chavez 1988; Nayar and Narasimhan 1999; Narasimhan and Nayar 2000; Schechner et al. 2001; Narasimhan and Nayar 2003; Shwartz et al. 2006]. Similar to the goal of these work, we are interested here in recovering J which is an image showing the scene through a clear haze-free medium. By that we do not eliminate other effects, the haze may have on the scene, such as a change in overall illumination which in turn affects the radiant emittance. Also, we assume that the input image I is given in the true scene radiance values. These radiance maps can be recovered by extracting the camera raw data or inverting the overall acquisition response curve, as described in [Debevec and Malik 1997].

This model (2) explains the loss of contrasts due to haze as the result of averaging the image with a constant color A. If we measure the contrasts in the image as the magnitude of its gradient field, a scene J seen through a uniform medium with $t(\mathbf{x}) = t < 1$ gives us

$$\|\nabla I\| = \|t\nabla J(\mathbf{x}) + (1-t)\nabla A\| = t\|\nabla J(\mathbf{x})\| < \|\nabla J(\mathbf{x})\|, \qquad (3)$$

4 Constant Albedo Images

The airlight-albedo ambiguity exists in each pixel independently and gives rise to a large number of undetermined degrees of freedom. To reduce the amount of this indeterminateness, we simplify the image locally by relating nearby pixels together. This is done in two steps which we describe next. In the first step we model the unknown image J as a pixelwise product of surface albedo coefficients and a shading factor, Rl, where R is a three-channel RGB vector of surface reflectance coefficients and l is a scalar describing

the light reflected from the surface. We use this refined model to simplify the image by assuming that $R(\mathbf{x})$ is piecewise constant. In this section we consider one of these localized sets of pixels $\mathbf{x} \in \Omega$ that share the same surface albedo, i.e., pixels in which $R(\mathbf{x}) = R$ for some constant vector R. At these pixels, the standard image formation model (2) becomes

$$I(\mathbf{x}) = t(\mathbf{x})l(\mathbf{x})R + (1-t(\mathbf{x}))A. \qquad (4)$$

Instead of having three unknowns per pixel in $J(\mathbf{x})$, we now have only one unknown $l(\mathbf{x})$ and $t(\mathbf{x})$ per pixel plus another three constants in R. We proceed by breaking R into a sum of two components, one parallel to the airlight A and a residual vector R' that lies in the linear sub-space which is orthogonal to the airlight, $R' \in A^\perp$. In terms of these normalized components, the equation above becomes

$$I(\mathbf{x}) = t(\mathbf{x})l'(\mathbf{x})(R'/\|R'\| + \eta A/\|A\|) + (1-t(\mathbf{x}))A, \qquad (5)$$

where $l' = \|R'\|l$ and $\eta = \langle R,A\rangle/(\|R'\|\|A\|)$ measuring the component that is mutual to the surface albedo and the airlight. By $\langle\cdot,\cdot\rangle$ we denote the standard three-dimensional dot-product in the RGB space. In order to obtain independent equations, we project the input image along the airlight vector, which results in a scalar given by

$$I_A(\mathbf{x}) = \langle I(\mathbf{x}),A\rangle/\|A\| = t(\mathbf{x})l'(\mathbf{x})\eta + (1-t(\mathbf{x}))\|A\|, \qquad (6)$$

and project along R', which equals to the norm of the residual that lies within $A^\perp$, i.e.,

$$I_{R'}(\mathbf{x}) = \sqrt{\|I(\mathbf{x})\|^2 - I_A(\mathbf{x})^2} = t(\mathbf{x})l'(\mathbf{x}). \qquad (7)$$

The transmission t can then be written in terms of these two quantities as

$$t(\mathbf{x}) = 1 - (I_A(\mathbf{x}) - \eta I_{R'}(\mathbf{x}))/\|A\|. \qquad (8)$$

In this equation the airlight-albedo ambiguity is made explicit; assuming for the moment that we have some estimate for the airlight vector A, the image components $I_A(\mathbf{x})$ and $I_{R'}(\mathbf{x})$ give us two constraints per pixel $\mathbf{x} \in \Omega$ while we have two unknowns per pixel $l(\mathbf{x})$ and $t(\mathbf{x})$ *plus* an additional unknown constant η. A third equation cannot be obtained from I since, according to our model, any vector in $A^\perp$ yields the same equation as $I_{R'}$. Thus, this model (5) reduces the pointwise ambiguity into a problem of determining a single scalar η for all $\mathbf{x} \in \Omega$. Finding this number, that expresses the amount of airlight in R', allows us to recover the transmission from (8) and ultimately the output image J according to (2).

Now we are ready to present the key idea which we use to resolve the airlight-albedo ambiguity in this reduced form. The transmission t depends on the scene depth and the density of the haze, $\beta(\mathbf{x})$, while the shading l depends on the illumination in the scene, surface reflectance properties, and scene geometry. Therefore it is reasonable to expect that the object shading function l and the scene transmission t do not exhibit a simple local relation. This leads us to the assumption that the two functions are statistically uncorrelated over Ω, meaning that $C_\Omega(l,t) = 0$, where the sample covariance C_Ω is estimated in the usual way

$$C_\Omega(f,g) = |\Omega|^{-1}\sum_{\mathbf{x}\in\Omega}(f(\mathbf{x})-E_\Omega(f))(g(\mathbf{x})-E_\Omega(g)), \qquad (9)$$

and the mean by

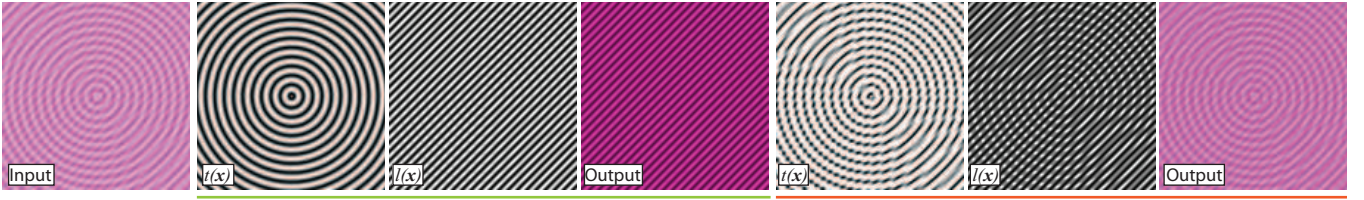$$E_\Omega(f) = |\Omega|^{-1}\sum_{\mathbf{x}\in\Omega}f(\mathbf{x}). \qquad (10)$$

Figure 2: Synthetic example. Left to right are: the input image, the recovered transmission t and shading l, and the recovered J. The last three image show the results obtained using a *wrong* η, where the correlation between t and l is apparent.

A simpler formulation is obtained by expressing this lack of correlation between t and $1/l'$ instead[1], where the latter is given by

$$\left(l'(\mathbf{x})\right)^{-1} = \left(1 - \left(I_A(\mathbf{x}) - \eta I_{R'}(\mathbf{x})\right)/\|A\|\right)/I_{R'}(\mathbf{x})$$
$$= \frac{1 - I_A(\mathbf{x})/\|A\|}{I_{R'}(\mathbf{x})} + \eta/\|A\|. \quad (11)$$

If we define a new variable

$$h(\mathbf{x}) = (\|A\| - I_A)/I_{R'}, \quad (12)$$

substitute (8) and (11) into the covariance operator, and extract η, we finally get that

$$\eta = \frac{C_\Omega(I_A, h)}{C_\Omega(I_{R'}, h)}. \quad (13)$$

This function is easily computed given the input image and, as explained earlier, η allows us to compute the transmission from (8) and recover the unknown image J according to (2). The procedure described here is, in essence, an *independent component analysis* [Hyvrinen and Oja 2000]. Unlike certain models, equation (13) fully determines η and decorrelates l and t. This is true as long as the denominator $C_\Omega(I_{R'}, h)$ is non-zero, which we describe how this is ensured in the next section. As we mentioned in the beginning, independent component analysis is used in [Shwartz et al. 2006] to determine a certain parameter, the degree of polarization, that is required for polarization-based multi-image dehazing.

In Figure 2 we show a synthetic example where the shading l consists of vertical ripples and the transmission t is made of radial ones. When estimating η from (13) these functions are accurately recovered, whereas the ones computed based on a false η appear to be mixed together, i.e., correlated.

5 Noise Estimation

Unfortunately no measurement comes free of noise, including the input image I. This error or lack of data in the input plays a critical role in this inverse problem of recovering a faint signal masked by haze. This noise introduces an uncertainty into the estimation of η by (13) and therefore into t as given by (8). In order to use the estimated transmission later, with the right amount of confidence, we must be able to assess the amount of error it contains. This is done by modeling the error present in I as an additive white Gaussian noise. More specifically, we assume that $I = \tilde{I} + \xi$, where ξ is a three dimensional vector of random Gaussian variables, which is independent both componentwise and pixelwise, with zero mean and variance σ^2. In the Appendix we follow the propagation of ξ, starting from I through the calculation of η up to t, and obtain an expression that is *linear* in ξ. This linearity means that at the end we are left again with a Gaussian noise whose variance σ_t^2 can be computed. In the next section we describe how this information is factored into the estimation of the final t.

---

[1] If X and Y are two independent random variables then so are $f(X)$ and $g(Y)$ for every pair of continuous functions f and g.

Along the derivation of the error estimate, given in detail in the Appendix, several conditions which allow us to truncate Taylor expansions are assumed to hold. Here we list these conditions and briefly explain why we believe that they are truly unavoidable and not merely artificial byproducts of the error formulation or analysis. The first condition requires the noise variance σ^2 to be less than the three components of $I/10$ and also less than $I_{R'}/6$ which is the haze-free component of the input. These are basic signal-to-noise requirements stating that data must be available and that pixels which are too opaque, compared to σ, contain little or no information about J. The second condition requires the variance of the noise term $\xi_{C_{R'}}$ to be less than $(C_\Omega(I_{R'}, h)/5)^2$, where $\xi_{C_{R'}}$ is the noise present in $C_\Omega(I_{R'}, h)$ which is defined in the Appendix. This condition breaks down when $|\Omega|$ is too small and there is not enough averaging to reduce the variance in $\xi_{C_{R'}}$ or when there is not enough variation in l which appears both in $I_{R'}$ and h. Intuitively, these conditions ensure the stability of the statistics E_Ω and C_Ω over Ω and the availability of enough multiplicative variation in $I_{R'}$ which is needed to resolve airlight-albedo ambiguity. These conditions are essential to the way we make our predictions and are explicitly quantified by this noise model.

By the above we do not mean to justify or assume that any of these conditions actually hold. On the contrary, every pixel that does not meet these error estimation criteria is omitted. And as the discussion above suggests, these pixels are discarded for not containing a reliable information on top of violating conditions needed to evaluate the error they contain. We denote the set of these discarded pixels by $\mathcal{B}$. As we show in the Appendix, the estimate of t at the rest of the pixels $\mathbf{x} \notin \mathcal{B}$ contains zero-mean Gaussian noise with a variance σ_t^2, defined in the Appendix. This error estimate is used in the next section to extrapolate t for *all* $\mathbf{x}$, needed for computing the output image J, when we extend our approach to handle images with multiple albedo values.

Given an input image we must supply σ that best approximates the variance of the noise present in the input image. We leave this to be set by the user and used $\sigma = 1/100 - 1/200$ in our experiments (for images with pixel values ranging between zero and one). Pixels with low signal-to-noise ratios are either discarded or receive a high σ_t^2, meaning that this parameter need not be precise. Recently [Liu et al. 2006] proposed an automated method to estimate the noise level based on a single image.

6 Multi-Albedo Images

In the previous section we considered the restricted case in which we assume that all the pixels correspond to the same surface, i.e., a constant albedo value $R(\mathbf{x}) = R$, and hence result in a constant airlight-albedo mutual component η. In order to handle general images, containing surfaces with different reflectance values, we perform these evaluations on a pixel basis using robust estimators defined on neighboring pixels. For that we relate pixels based on their similarity in albedo using the following shading and airlight invariant measure. We compute a naive haze-free image by $I' = I - AI_A/\|A\|$ (which may contain negative values) and project
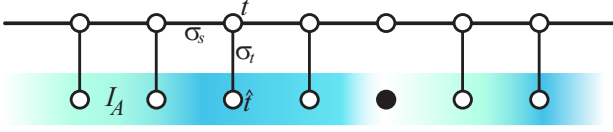
Figure 4: Random graph model illustrated in 1D. This model links the unknown transmission variables t to the estimated values $\hat{t}$ as well as to neighboring ones. Unreliable pixels for which we cannot assess the error are not linked. The airlight image I_A, from which the spatial dependencies are derived, is depicted by the lower blue strip.

it onto the two chroma channels U and V of the YUV color space. Since we are not interested in differentiating pixels based on their shading, we extract a single chroma channel based on the ratio of these projections

$$\theta(\mathbf{x}) = \tan^{-1}\left(\langle I', U\rangle / \langle I', V\rangle\right). \qquad (14)$$

This maps pixel color into angles where we measure distances by the length of the shortest arc, i.e.,

$$d(\theta_1, \theta_2) = \min\{|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|\}. \qquad (15)$$

We use this similarity measure to define the weights of the following w-estimators [Andrews et al. 1972]

$$C_\mathbf{x}(f, g) = \frac{1}{W_\mathbf{x}} \sum_{\mathbf{y} \in \Omega_\mathbf{x}} \left(f(\mathbf{y}) - E_\mathbf{x}(f)\right)\left(g(\mathbf{y}) - E_\mathbf{x}(g)\right)w(\mathbf{x}, \mathbf{y}),$$
$$\text{and} \quad E_\mathbf{x}(f) = \frac{1}{W_\mathbf{x}} \sum_{\mathbf{y} \in \Omega_\mathbf{x}} f(\mathbf{y})w(\mathbf{x}, \mathbf{y}), \qquad (16)$$

where the weighting function $w(\mathbf{x}, \mathbf{y})$ is given by $\exp(-d(\theta(\mathbf{x}), \theta(\mathbf{y}))^2/\sigma_\theta^2)$, the normalizing weight $W_\mathbf{x} = \sum_{\mathbf{y} \in \Omega_\mathbf{x}} w(\mathbf{x}, \mathbf{y})$, and $\Omega_\mathbf{x}$ is the window of pixels centered around $\mathbf{x}$ excluding pixels from $\mathscr{B}$. In our testings we set $\sigma_\theta = 1/8$ and use windows of 24-by-24 pixels which can be enlarged in case of noisy images. We replace E_Ω and C_Ω in (13) with these robust estimators and compute $\eta(\mathbf{x})$ at every pixel $\mathbf{x} \notin \mathscr{B}$. We compute the transmission at these pixels according to (8) and denote this primary estimate by $\hat{t}$. As noted earlier, in the Appendix we compute the estimated noise variance σ_t^2 present in these estimated transmission values.

We proceed with a statistical smoothing to account for the noise present in $\hat{t}$ and in order to fill in the transmission values at pixels $\mathbf{x} \in \mathscr{B}$. The airlight contribution is fully contained in I_A and can therefore be used as a prior to derive the spatial regularity of t. We assume independe between these dependencies and formalize them using a *Gauss-Markov random field* model [Pérez 1998], defined by

$$P(t) \propto \prod_{\mathbf{x} \notin \mathscr{B}} e^{-(t(\mathbf{x}) - \hat{t}(\mathbf{x}))^2/\sigma_t^2(\mathbf{x})} \prod_{\forall \mathbf{x}, \mathbf{y} \in N_\mathbf{x}} e^{-(t(\mathbf{x}) - t(\mathbf{y}))^2/(I_A(\mathbf{x}) - I_A(\mathbf{y})))^2/\sigma_s^2}, \qquad (17)$$

where $N_\mathbf{x}$ is the set of pixel $\mathbf{x}$'s four nearest-neighbors in the lattice and use $\sigma_s = 1/5$. The links defined by this model are illustrated in Figure 4. We maximize this model by solving the linear system resulting from $d \log P/dt = 0$ and take this optimum to be our final t which we use to recover the final haze-free output image J from (2).

7 Estimating the Airlight Color

We can apply the principle of uncorrelation, described in Section 4, once again in order to estimate the airlight color vector A. One possibility is to restrict to small windows of constant albedo and

search for a vector A that yields the least correlated t and l. This is based on a requirement for consistency between the equations derived and the input pixels; the airlight vector used in equations (8), (11), and (12) must be identical to the airlight present in the input pixels so that the t and l resulting from our computations will indeed be uncorrelated. If this is not the case the η, computed from (13) and meant to achieve zero correlation, is meaningless. Thus given an initial guess for A, we minimize $C(J, t)^2$ by updating A's components using the steepest descent method. This process is done only within small windows of 24-by-24 pixels and takes a few seconds to perform. Similarly to [Narasimhan and Nayar 2003], the most haze-opaque pixel can be used as an initial guess.

This three-dimensional minimization can be reduced to a one-dimensional search if two or more regions that correspond to different but uniform albedos are selected. As indicated by (4), in such regions all the pixels lay within a two-dimensional linear sub-space spanned by the vectors R and A. A principal component analysis will not extract the directions of these vectors since their coefficients tl and $-t$ are not independent. Nevertheless, we use this analysis in order to find the sub-spaces themselves by omitting the least-active component of the three. This is applied in each of the given regions, indexed by i, and we denote the two spanning components by $\mathbf{v}_1^i$ and $\mathbf{v}_2^i$. Since we expect A to be contained in each of these sub-spaces, its direction can be recovered by intersecting these planes. We do this by searching for a vector with the highest projection onto all these sub-spaces, i.e.,

$$\max_A \sum_i \langle A, \mathbf{v}_1^i \rangle^2 + \langle A, \mathbf{v}_2^i \rangle^2 \text{ such that } \|A\|^2 = 1. \qquad (18)$$

The solution to this problem, according to the Lagrange-multipliers rule, is given by the eigenvector that corresponds to the highest eigenvalue of the 3-by-3 matrix given by $\sum_i \mathbf{v}_1^i (\mathbf{v}_1^i)^\top + \mathbf{v}_2^i (\mathbf{v}_2^i)^\top$. The search described above, where we minimize $C(J, t)^2$, can now be used to find the magnitude of A. This is a one-dimensional optimization problem which we solve using a naive search and takes less than one second.

8 Results

In Figure 3 we show an input image and the output produced by our approach. We also show the scene depth values computed by $d = -\log t/\beta$ which follows from (1) when assuming a spatially-uniform extinction coefficient $\beta(\mathbf{r}) = \beta$. These depth values are defined up to an unknown multiplicative constant $1/\beta$. In Figure 1 we show the result for a more complicated image with surfaces of distinct albedos. As mentioned earlier, aerial and underwater photography are also prone to scattering artifacts, and in Figure 5 we show the results for these types of images.

In Figure 7 we compare our method with existing techniques. The result obtained by our method given the best-polarized image from [Schechner et al. 2001] is comparable to the ones they obtain using two registered and differently polarized images. The results obtained by the dark-object subtraction method [Chavez 1988] shown in this figure correspond to two different choices of dark object values. The outcome depends greatly on this choice and demonstrates the limited ability of this method to cope with images that contain multiple layers of haze depth. Tonemapping methods such as histogram equalization and gamma correction achieve some enhancement of contrast, yet most of the additive airlight component remains unchanged. The same is true for filtering-based enhancement such as unsharp masking and the multi-scale bilateral filter [Fattal et al. 2007].

In Figure 8 we show the results on a real scene where the ground-truth solution is known. This example is synthesized from an image and its corresponding disparity map which we used to create the scene transmission. This image is taken from [Hirschmller and

Figure 3: Wheat cones. The estimated airlight is shown at the top-left corner of the input image. Source image courtesy of Jason Hill.
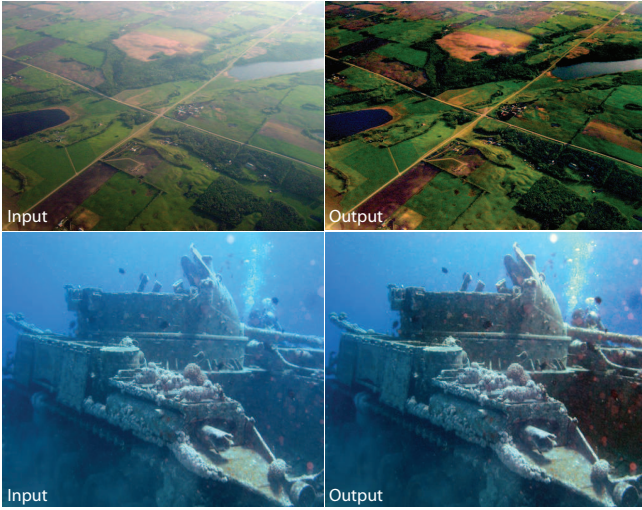

Figure 5: Aerial and underwater photography.

Figure 6: In the left image we simulate a virtual focus plane. The right image is rendered with a recolored haze layer multiplied by a random noise.

Scharstein 2007]. The result is overall quite satisfactory yet shows some inaccuracies. The mean absolute error in transmission and haze-free image J are both less than 7%. In another test, shown in Figure 9, we show two images: left image has its middle square masked by a thicker haze, and in the right image the square is slightly brighter exactly in the direction of the airlight. It is hard to tell this difference visually, yet the method is able to disambiguate the two scenarios and provide near perfect decompositions.

Additional Applications. We use the predicted transmission function $t(\mathbf{x})$ for other applications on top of haze removal. In Figure 6 we show an image simulating a denser medium by multiplying the extinction coefficient β by a factor of $\lambda = 2$. According to (1) this is achieved by applying the following simple power law transformation of the transmission values

$$e^{-\int \lambda \beta(s)ds} = e^{-\lambda \int \beta(s)ds} = \left(e^{-\int \beta(s)ds}\right)^{\lambda} = t^{\lambda}. \qquad (19)$$

In this figure we also show a scene rendered with an altered haze layer, where the fog layer is modulated by random noise at different scales and recolored.

Estimated depth maps are shown in figures 1 and 3. Based on these maps we change the viewpoint and arbitrarily refocus the image, as shown in Figure 6. Animated examples of these effects are shown in the supplementary video, which contains more examples and offers a better viewing experience.

We implemented our method in C++ and it takes it about 35 seconds to process a 512-by-512 pixel image on a 1.6GHz Intel Pentium Dual-Core Processor.
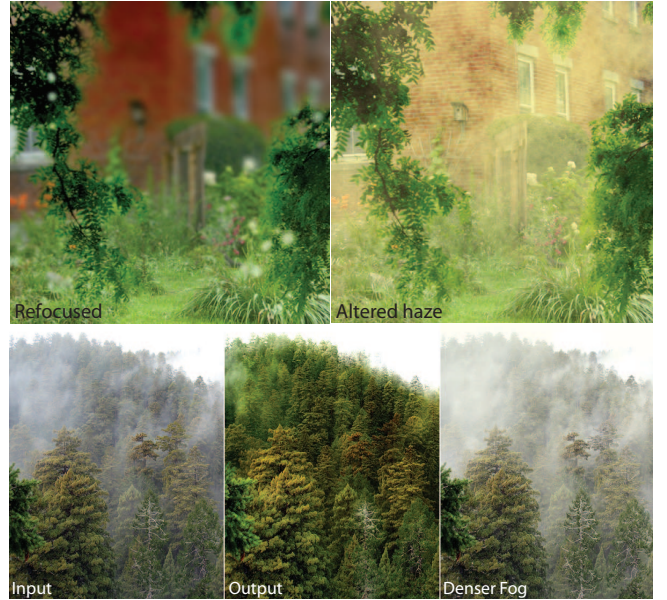
9   Conclusions

We presented a new method for estimating the transmission in hazy scenes based on minimal input, a single image. This new approach is physically sound and relies only on the assumption that the transmission and surface shading are locally uncorrelated. The use of robust statistics allows us to cope with complicated scenes containing different surface albedos and the use of an implicit graphical model makes it possible to extrapolate the solution to pixels where no reliable estimate is available. Despite the challenges in resolving the ambiguities involved in this problem, the images produced by this method offer a lucid view of the scene and regain contrast that is typical to haze-free scenes.

In its essence this method solves a non-linear inverse problem and therefore its performance greatly depends on the quality of the input data. We derive an estimation for the noise present in the input and use statistical extrapolation to cope with large errors. Nevertheless, insufficient signal-to-noise ratio or the absence of multiplicative variation in significant portions of the image will cause our method to fail. These scenarios are easy to find and in Figure 10 we
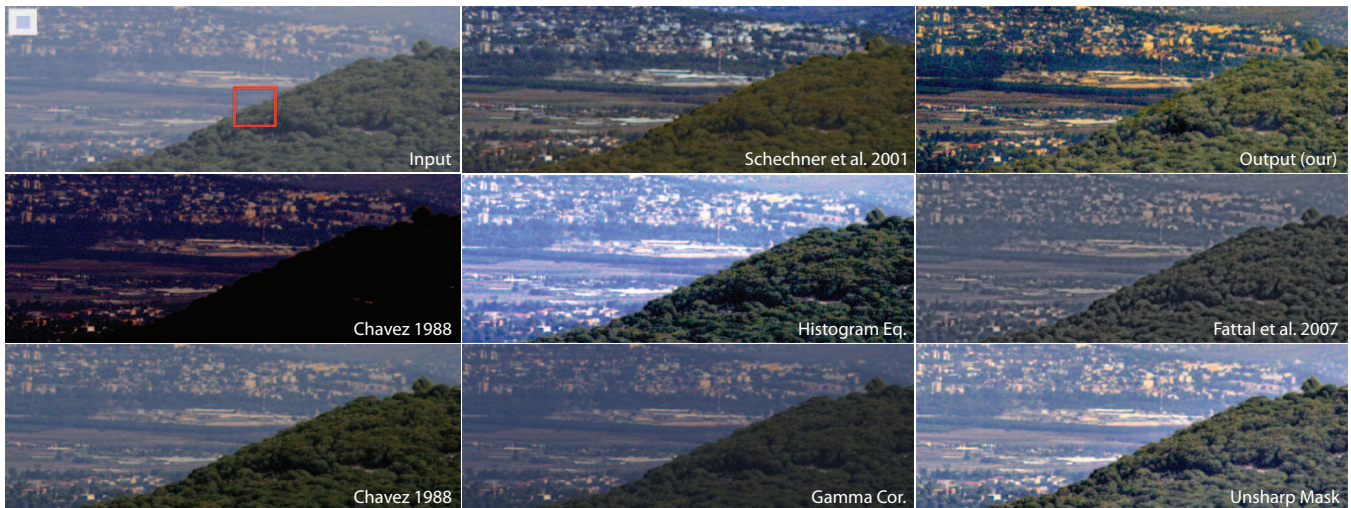
Figure 7: Comparison with other approaches. Image taken from Schechner et al. 2001.

show one such example. Also, when too few pixels are available the accuracy of the statistical estimation is undermined such as in the case of the white flowers in Figure 1.

Haze removal from images provides a clearer view of the scene, but we do not regard it as a default preference. On the contrary, the presence of haze can be desired for many reasons, for example it is known to be an important monocular depth cue [Coon 2005]. Painters use a technique called *atmospheric perspective* where they wash out regions of the painting using white paint to create an illusion of depth.

Recent theories of perceptual transparency [Singh and Anderson 2002] suggest that we, the human observers, perceive the medium transmission based on the contrast ratio of a pattern seen through a semi-transparent filter and through a clear medium. Although we do not assume here that any haze-free data is available to us, we also rely on variations in surface shading in order to disambiguate and require a significant $C(I_{R'}, h)$.

Atmospheric scattering causes some amount of blurriness in the captured image [Grewe and Brooks 1998]. As a future work we intend to investigate the possibility of incorporating a debluring mechanism that is regulated by our estimated transmission. This idea can be pushed even further where the blurriness itself is used as an additional source of information when estimating the transmission function.

Acknowledgements

I would like to thank the Siggraph reviewers for their comments which helped improving this manuscript as well as thank Robert Carroll, Raz Kupferman, Dani Lischinski and Yehuda Arav for their feedback and helpful discussion.

References

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H., AND W.TUKEY, J. 1972. *Robust Estimates of Location: Survey and Advances*. Princeton University Press; London, Oxford University Press.

CHAVEZ, P. S. 1988. An improved dark-object subtraction technique for atmonspheric scattering correction of multispectral data. *Remote Sensing of Environment 24*, 450–479.

COON, D. 2005. *Psychology: A Modular Approach To Mind And Behavior*. Wadsworth Pub Co, July.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 1997*, 369–378.

DU, Y., GUINDON, B., AND CIHLAR, J. 2002. Haze detection and removal in high resolution satellite image with wavelet analysis. *IEEE Transactions on Geoscience and Remote Sensing 40*, 1, 210–217.

FATTAL, R., AGRAWALA, M., AND RUSINKIEWICZ, S. 2007. Multiscale shape and detail enhancement from multi-light image collections. In *ACM SIGGRAPH*, 51.

FATTAL, R. 2007. Image upsampling via imposed edge statistics. *ACM SIGGRAPH 26*, 3, 95.

GREWE, L., AND BROOKS, R. R. 1998. Atmospheric attenuation reduction through multi-sensor fusion in sensor fusion: Architectures, algorithms, and applications. 102–109.

HIRSCHMLLER, H., AND SCHARSTEIN, D. 2007. Evaluation of cost functions for stereo matching.

HYVRINEN, A., AND OJA, E. 2000. Independent component analysis: Algorithms and applications. *Neural Networks 13*, 411–430.

KOSCHMIEDER, H. 1924. Theorie der horizontalen sichtweite. In *Beitr. zur Phys. d. freien Atm.*, 171–181.

LARSON, G. W., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics 3*, 4, 291–306.

LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and depth from a conventional camera with a coded aperture. *ACM Transaction on Graphics 26*, 3, 70.

LIU, C., FREEMAN, W. T., SZELISKI, R., AND KANG, S. B. 2006. Noise estimation from a single image. In *Proceedings of IEEE CVPR*, 901–908.

LU, J., AND JR., D. M. H. 1994. Contrast enhancement via multi-scale gradient transformation. In *IEEE International Conference on Image Processing*, 482–486.

NARASIMHAN, S. G., AND NAYAR, S. K. 2000. Chromatic framework for vision in bad weather. In *Proceedings of IEEE CVPR*, 598–605.
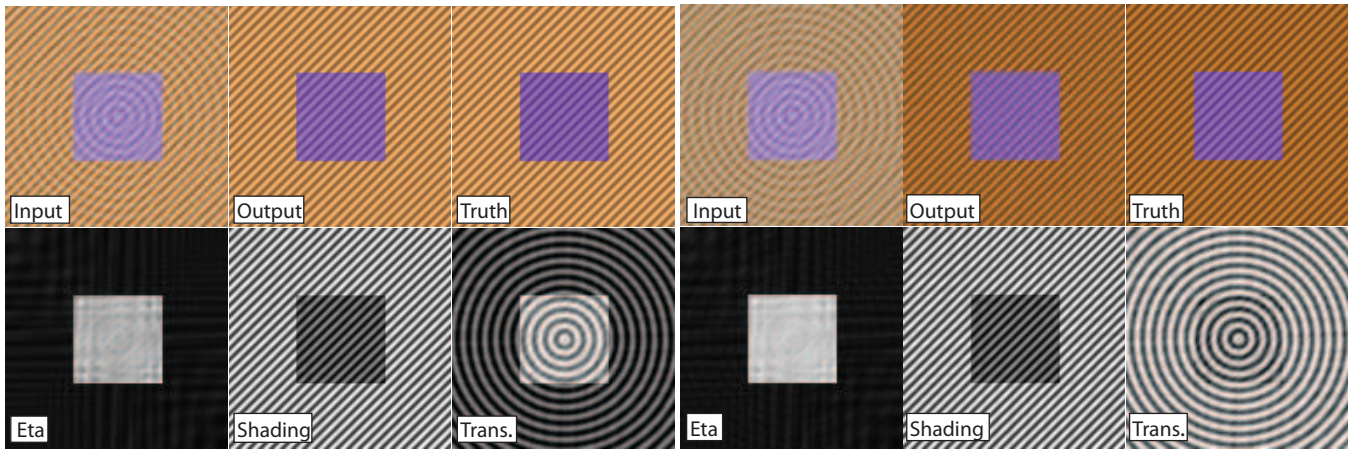
Figure 9: Disambiguation test. Transmission mean absolute error is 56×10^{-4} and 13×10^{-3} in the left and right tests respectively.
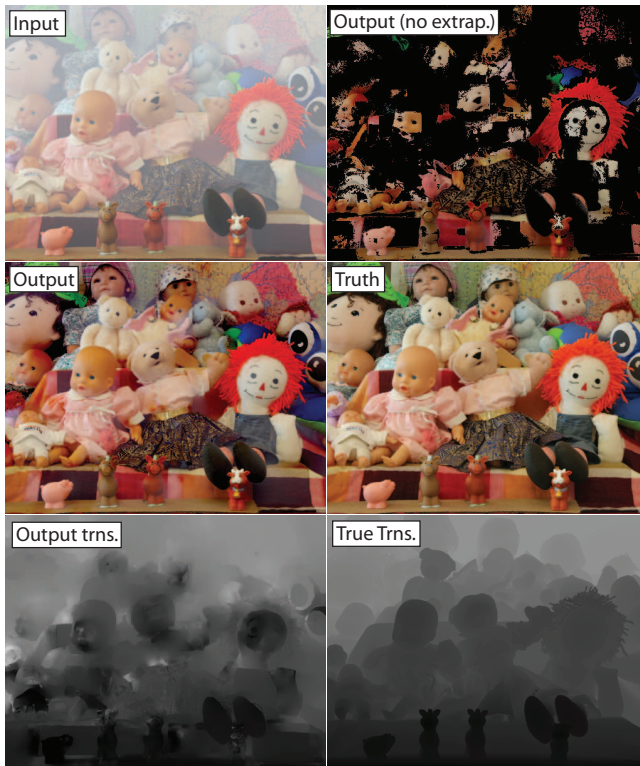

Figure 8: Real scene test. Mean absolute error in estimated transmission is 43×10^{-3} and 68×10^{-3} in J.

Figure 10: At the roof $I_{R'}$ is near zero and causes an overshoot in the estimated transmission.

NARASIMHAN, S. G., AND NAYAR, S. K. 2003. Interactive (De)weathering of an Image using Physical Models. In *ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*.

NAYAR, S. K., AND NARASIMHAN, S. G. 1999. Vision in bad weather. In *Proceedings of IEEE CVPR*, 820.

OAKLEY, J. P., AND BU, H. 2007. Correction of simple contrast loss in color images. *IEEE Transactions on Image Processing 16*, 2, 511–522.

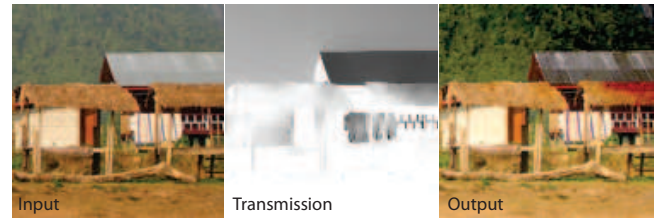PÉREZ, P. 1998. Markov random fields and images. In *CWI Quar-terly*, vol. 11, 413–437.

RAHMAN, Z., JOBSON, D., AND WOODELL, G. 1996. Multiscale retinex for color image enhancement.

ROSSUM, M. V., AND NIEUWENHUIZEN, T. 1999. Multiple scattering of classical waves: microscopy, mesoscopy and diffusion. vol. 71, 313–371.

SCHECHNER, Y. Y., AND AVERBUCH, Y. 2007. Regularized image recovery in scattering media. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*, 9, 1655–1660.

SCHECHNER, Y. Y., NARASIMHAN, S. G., AND NAYAR, S. K. 2001. Instant dehazing of images using polarization. 325–332.

SHWARTZ, S., NAMER, E., Y., Y., AND SCHECHNER. 2006. Blind haze separation. In *Proceedings of IEEE CVPR*, 1984–1991.

SINGH, M., AND ANDERSON, B. 2002. Toward a perceptual theory of transparency. No. 109, 492–519.

TAN, K., AND OAKLEY, J. P. 2000. Enhancement of color images in poor visibility conditions. *Proceedings of International Conference on Image Processing 2*, 788–791.

TAN, R. T. 2008. Visibility in bad weather from a single image. *Proceedings of IEEE CVPR*.

VEERARAGHAVAN, A., RASKAR, R., AGRAWAL, A., MOHAN, A., AND TUMBLIN, J. 2007. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *ACM SIGGRAPH*, 69.

WIKIPEDIA, 2007. Unsharp masking — wikipedia, the free encyclopedia.

YUAN, L., SUN, J., QUAN, L., AND SHUM, H.-Y. 2007. Image deblurring with blurred/noisy image pairs. *ACM Transactions on Graphics 26*, 3, 1.

ZHANG, Y., GUINDON, B., AND CIHLAR, J. 2002. An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images. *Remote Sensing of Environment 82* (October), 173–187.

Appendix: Derivation of the Noise Estimates

Here we derive our estimation for the noise present in the transmission t, given by (8). As described in Section 5, this analysis is based on modeling the noise present in the input data I as an additive white Gaussian noise with zero mean and variance σ^2, given by

$$I = \tilde{I} + \xi, \tag{20}$$

where $\tilde{I}$ is the true image and the noise ξ is a three-dimensional vector (corresponding to the RGB color channels) with each component at each pixel being an independent Gaussian variable $\xi \sim N(0, \sigma^2)$. This breaks I_A into signal and noise terms,

$$I_A = \langle \tilde{I} + \xi, A \rangle / \|A\| = \tilde{I}_A + \xi_A, \tag{21}$$

with $\tilde{I}_A = \langle \tilde{I}, A \rangle / \|A\|$ and $\xi_A = \langle \xi, A \rangle / \|A\|$. Similarly, for $I_{R'}$ we get

$$I_{R'}^2 = \|I\|^2 - I_A^2 = \|\tilde{I}\|^2 + 2\langle \tilde{I}, \xi \rangle + \|\xi\|^2 - \tilde{I}_A^2 - 2\tilde{I}_A \xi_A - \xi_A^2$$
$$= \tilde{I}_{R'}^2 + \langle 2\tilde{I} - \xi, \xi - \xi_A A \rangle, \tag{22}$$

where

$$\tilde{I}_{R'} = \sqrt{\|\tilde{I}\|^2 - \langle \tilde{I}, A \rangle^2},$$

and by assuming that the noise ξ is less than $\tilde{I}$ we neglect the quadratic noise terms. This leaves us

$$I_{R'} = \sqrt{\tilde{I}_{R'}^2 + 2\langle \xi, \tilde{I} - \tilde{I}_A A \rangle}, \tag{23}$$

and since $\tilde{I}_{R'}^2 = \|\tilde{I} - \tilde{I}_A A\|^2$ we need ξ to be less than $\tilde{I}_{R'}$ (and not than $\tilde{I}_{R'}^2$) in order to use a first-order Taylor approximation for $\sqrt{x}$ centered around $\tilde{I}_{R'}$. This gives us

$$I_{R'} \approx \tilde{I}_{R'} + \left(\langle \tilde{I}, \xi \rangle - \tilde{I}_A \xi_A \right) / \tilde{I}_{R'} = \tilde{I}_{R'} + \xi_{R'}, \tag{24}$$

where $\xi_{R'}$ can be written more explicitly as

$$\xi_{R'} = \langle \xi, \tilde{I} - A\tilde{I}_A / \|A\| \rangle / \tilde{I}_{R'}. \tag{25}$$

Based on the same assumption, $\xi < \tilde{I}_{R'}$, we use a first-order Taylor approximation for $1/x$ around $\tilde{I}_{R'}$ in order to remove the noise term from the denominator of h

$$h \approx (\|A\| - I_A)/\tilde{I}_{R'} - \xi_{R'}(\|A\| - I_A)/\tilde{I}_{R'}^2 = (\|A\| - \tilde{I}_A)/\tilde{I}_{R'}$$
$$\quad - \xi_A/\tilde{I}_{R'} - \xi_{R'}(\|A\| - \tilde{I}_A)/\tilde{I}_{R'}^2 + \xi_{R'}\xi_A/\tilde{I}_{R'}^2 \tag{26}$$
$$= \tilde{h} + \xi_h + O(\|\xi\|^2),$$

where $\tilde{h} = (\|A\| - \tilde{I}_A)/\tilde{I}_{R'}$, and

$$\xi_h = -(\xi_A + \xi_{R'}\tilde{h})/\tilde{I}_{R'} = \langle \xi, A(\tilde{h}\tilde{I}_A/\tilde{I}_{R'} - 1)/\|A\| - \tilde{I}\tilde{h}/\tilde{I}_{R'} \rangle / \tilde{I}_{R'}. \tag{27}$$

The covariance terms gives

$$C_\Omega(I_{R'}, h) = C_\Omega(\tilde{I}_{R'} + \xi_{R'}, \tilde{h} + \xi_h)$$
$$= C_\Omega(\tilde{I}_{R'}, \tilde{h}) + C_\Omega(\xi_{R'}, \tilde{h}) + C_\Omega(\tilde{I}_{R'}, \xi_h) + O(\|\xi\|^2)$$
$$\approx C_\Omega(\tilde{I}_{R'}, \tilde{h}) + E_\Omega(\xi_{R'}\tilde{h} + \xi_h\tilde{I}_{R'}) - E_\Omega(\xi_{R'})E_\Omega(\tilde{h}) - E_\Omega(\xi_h)E_\Omega(\tilde{I}_{R'})$$
$$= C_\Omega(\tilde{I}_{R'}, \tilde{h}) + \xi_{C_{R'}}, \tag{28}$$

where by taking into account that $\xi_{R'}\tilde{h} + \xi_h\tilde{I}_{R'} = -\xi_A$ the noise term becomes

$$\xi_{C_{R'}} = -E_\Omega(\xi_A) - E_\Omega(\xi_{R'})E_\Omega(\tilde{h}) - E_\Omega(\xi_h)E_\Omega(\tilde{I}_{R'}). \tag{29}$$

Similarly, we get

$$C_\Omega(I_A, h) \approx C_\Omega(\tilde{I}_A, \tilde{h}) + \xi_{C_A}, \tag{30}$$

where

$$\xi_{C_A} = E_\Omega(\xi_A\tilde{h} + \xi_h\tilde{I}_A) - E_\Omega(\xi_A)E_\Omega(\tilde{h}) - E_\Omega(\xi_h)E_\Omega(\tilde{I}_A). \tag{31}$$

Finally, under the assumption that $C_\Omega(\tilde{I}_{R'}, \tilde{h})$ is larger than $\xi_{C_{R'}}$, we use again a first-order Taylor expansion of $1/x$ around $C_\Omega(\tilde{I}_{R'}, \tilde{h})$ and get

$$\eta = \frac{C_\Omega(\tilde{I}_A, \tilde{h}) + \xi_{C_A}}{C_\Omega(\tilde{I}_{R'}, \tilde{h}) + \xi_{C_{R'}}} \approx \tilde{\eta} + \xi_\eta, \tag{32}$$

where

$$\xi_\eta = \frac{\xi_{C_A} C_\Omega(\tilde{I}_{R'}, \tilde{h}) - \xi_{C_{R'}} C_\Omega(\tilde{I}_A, \tilde{h})}{\left(C_\Omega(\tilde{I}_{R'}, \tilde{h})\right)^2}, \tag{33}$$

and $\tilde{\eta}$ is the the noiseless $C_\Omega(\tilde{I}_A, \tilde{h})/C_\Omega(\tilde{I}_{R'}, \tilde{h})$ that we are estimating. Note that E_Ω and C_Ω are the mean and variance of the pixels sample in Ω and have nothing to do with the moments of ξ. Finally, from (8) we get

$$t = \tilde{t} + (\xi_A - \tilde{\eta}\xi_{R'} - \tilde{I}_{R'}\xi_\eta - \xi_\eta\xi_{R'})/\|A\| = \tilde{t} + \xi_t, \tag{34}$$

where $\tilde{t} = 1 - (\tilde{I}_A - \tilde{\eta}\tilde{I}_{R'})/\|A\|$ and $\xi_t = (\xi_A - \tilde{\eta}\xi_{R'} - \tilde{I}_{R'}\xi_\eta)/\|A\|$. Here we drop the quadratic noise term, under the assumption that $\xi_{R'} < \tilde{I}_{R'}$, i.e., ξ is small.

Along this derivation, we made several assumptions which we quantify now. In order for $a\xi + \xi^2$ to still behave like a Gaussian, i.e., have a distribution with a similar shape and a variance of $\approx a^2\sigma^2$, we need $\sigma < a/10$. In our derivation (23) this condition requires that $\sigma < \tilde{I}/10 \approx I/10$ for each of the RGB components and at each pixel. This is also the case in (34) where σ must be less than $1/10$. Similarly, in order to remove the noise from $\sqrt{a + \xi}$ and still maintain an approximately Gaussian distribution, we need that $\sigma < a/3$. In our derivation this condition arises in (24) and requires $\sigma < \tilde{I}_{R'}/6 \approx I_{R'}/6$. The ratio between a and ξ that allows us to move the noise from the denominator in $1/(a + \xi)$ and still end up with approximately a Gaussian variable is $\sigma < a/5$. This applies to the two cases that we encountered above, the first in (26) which gives a less restrictive condition $\sigma < (\tilde{I} - \tilde{I}_A A)/5$ than what we already have and the second case in (32) that requires

$$\mathrm{Var}_\xi(\xi_{C_{R'}}) < \left(C_\Omega(\tilde{I}_{R'}, \tilde{h})/5\right)^2 \approx \left(C_\Omega(I_{R'}, h)/5\right)^2, \tag{35}$$

where Var_ξ is the variance with respect to the noise randomness. In Section 5 we summarize these conditions and in Section 6 we explain how this uncertainly is accounted for in the estimation of t.

The estimated transmission noise term ξ_t is now a linear combination of ξ. Therefore its variance is computed according to the following arithmetic rules for Gaussian variables: (i) $a\xi(\mathbf{x}) + b\xi(\mathbf{x}) \sim N(0, (a + b)^2\sigma^2)$, and (ii) $a\xi(\mathbf{x}) + b\xi(\mathbf{y}) \sim N(0, (a^2 + b^2)\sigma^2)$ for any scalars a, b. The variance of $\xi_A, \xi_{R'}$ and ξ_h are computed according to the first rule as no different pixels get mixed together. Mixing occurs in $\xi_{C_{R'}}$ and ξ_{C_h} where the second rule is applied. We use the first rule also to calculate the variance of the intermediate noise terms $\xi_A\tilde{h} + \xi_h\tilde{I}_A$ in (31) and for $\xi_A - \tilde{\eta}\xi_{R'}$, found in ξ_t.