Combining Stereo and Atmospheric Veil Depth Cues for 3D Reconstruction

Laurent Caraffa, Jean Philippe Tarel

IPSJ Transactions on Computer Vision and Applications Vol.6 1–11 (Feb. 2014)

[DOI: 10.2197/ipsjtcva.6.1]

Research Paper

Combining Stereo and Atmospheric Veil Depth Cues for 3D Reconstruction

Laurent Caraffa[1,2,a)] Jean-Philippe Tarel[1,2,b)]

Received: May 31, 2013, Accepted: November 14, 2013, Released: February 17, 2014

Abstract: Stereo reconstruction serves many outdoor applications, and thus sometimes faces foggy weather. The quality of the reconstruction by state of the art algorithms is then degraded as contrast is reduced with the distance because of scattering. However, as shown by defogging algorithms from a single image, fog provides an extra depth cue in the gray level of far away objects. Our idea is thus to take advantages of both stereo and atmospheric veil depth cues to achieve better stereo reconstructions in foggy weather. To our knowledge, this subject has never been investigated earlier by the computer vision community. We thus propose a Markov Random Field model of the stereo reconstruction and defogging problem which can be optimized iteratively using the α-expansion algorithm. Outputs are a dense disparity map and an image where contrast is restored. The proposed model is evaluated on synthetic images. This evaluation shows that the proposed method achieves very good results on both stereo reconstruction and defogging compared to standard stereo reconstruction and single image defogging.

Keywords: 3D Reconstruction, Stereo Reconstruction, Contrast Restoration, Defogging, Dehazing, MRF model.

1. Introduction

The first dense stereo reconstruction algorithms were proposed forty years ago. There is now more than one hundred algorithms listed on the Middlebury evaluation site. Nevertheless, several new algorithms or improvements are proposed each year. The reason for this constant interest is the high usefulness of the 3D reconstruction which serves in many applications such as: driver assistance, automatic driving, environment simulators, augmented reality, data compression, 3D TV. While the Middlebury database contains only indoor scenes of good quality, outdoor applications are confronted with more difficult weather conditions such as fog, rain and snow. These weather conditions reduce the quality of the stereo pairs and introduce artifacts. Reconstruction results are thus usually degraded.

The principle of stereo reconstruction is to find, for every pixel in the left image, the pixel in the right image which minimizes a matching cost along the epipolar line. Depending on the scene, the matching cost can be ambiguous or wrongly minimal. A prior on the disparity map is thus added, for instance to enforce that close pixels have similar disparity. As a consequence, the stereo reconstruction is set as the minimization of an energy which derives from a Markov Random Field (MRF) model, see for instance [1], [16]. Thanks to recent advances in numerical analysis, the optimization of this energy can be performed quickly without being trapped by most of the local minima.

We observed that stereo reconstructions are degraded in the presence of fog. As an illustration, in Fig. 1, we show disparity maps obtained on a foggy stereo image by four stereo reconstruction algorithms: α-expansion on MRF [1], Elas [3], correlation windows and dynamic programing on each column. Results are not satisfactory; in the best case, they are correct only up to a critical distance. Indeed, in a foggy scene, the more distant an object, the whiter its color. As a consequence, contrast is a decreasing function of distance, which makes matching all the more difficult to perform.

However, image processing in foggy image condition has been studied for a while, especially image defogging. The goal of image defogging is to find the original intensity of a foggy scene. For this purpose, several algorithms has been proposed for single image defogging.

One common point of many single image defogging algorithms is the use of an intermediate depth map to estimate the restored image. Indeed, the farthest an object is, the stronger the contrast should be improved. The first method for single image defogging is in [10], where an approximate depth-map of the scene geometry is built interactively depending of the scene. This method, due to the depth map dependency, cannot be used when the 3D model of the scene is unknown. To tackle this problem, several methods has been proposed to estimate the depth from the foggy image. In [7], [12], [15], three defogging methods are introduced which are able to process a gray-level or as well as a color image. These three methods rely on a single principle: the use of a local spatial regularization. The single image defogging problem is ill-posed. Indeed, there is an ambiguity between the original color of an object, and the color of the fog added with the distance. Consequently, the real distance cannot be computed with a single image, and leads to compute an approximated depth map

[1] Paris-Est University, France.
[2] IFSTTAR, LEPSiS, 14-20 Boulevard Newton, Cité Descartes, Champs-sur-Marne, F-77420, France.
[a)] laurent.caraffa@ifsttar.fr
[b)] jean-philippe.tarel@ifsttar.fr

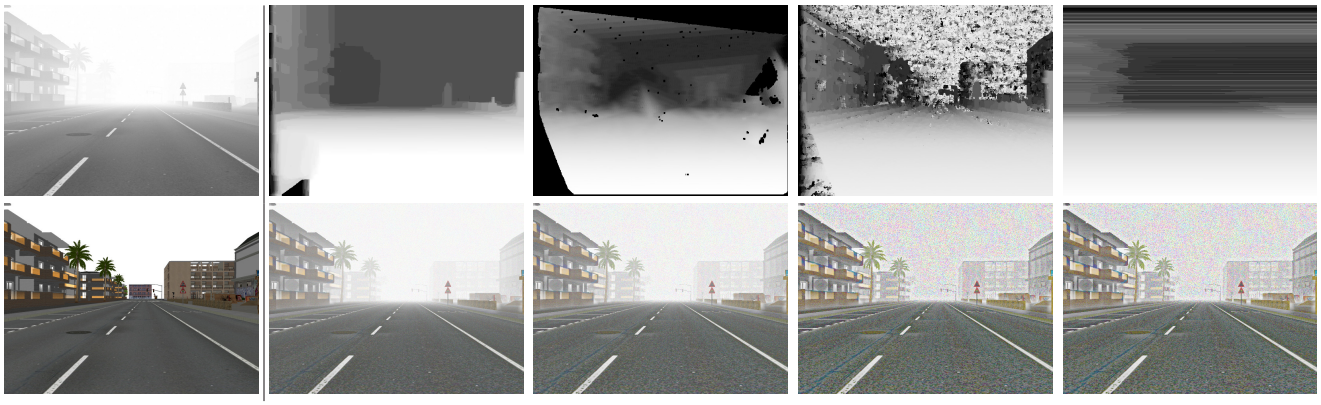

Fig. 1 Examples of stereo reconstruction in foggy weather and of single image defogging. First line, from left to right: original left image of the stereo pair, disparity maps obtained using: α-expansion on MRF [1], Elas [3], correlation windows and dynamic programing on each line. Second line, the left image without fog, and defogging results using [15] with four increasing degrees of enhancement.

and biased restored image. Fig. 1 also shows results obtained by [15] with different values of the restoration strength. For close objects, only a small amount of enhancement is necessary. On the contrary, remote objects needs more contrast restoration. But this stronger restoration usually leads to over-restoration in the close area. Various camera-based Advanced Driver Assistance Systems (ADAS) can be improved if more efficient algorithms are designed for visibility enhancement of road environment images. For this kind of images, a large area of the image is covered by the road, which is gray and thus subject to the white color ambiguity, i.e, no depth information can be obtained in a reliable way from a single image. To remedy this difficulty, a strong geometric constraint about the 3D position of the road is added in [14]. It is the so-called planar assumption (PA). With the PA, a good restoration of the road area is obtained, but the problem remains in gray areas which are outside the road plane.

If stereo disparity is important for 3D reconstruction, in foggy scenes, the gray-level of distant objects is also a depth cue. This depth cue is used in contrast restoration algorithms but had not been used in 3D reconstruction yet. The defogging problem, like the stereo reconstruction, can be set as a MRF problem, see [2], [11]. The atmospheric veil depth cue is particularly interesting since it is complementary to the stereo depth cue: the former is reliable only for remote objects, while the latter is reliable only for near by objects. Our idea is thus to combine MRF models of both stereo reconstruction and defogging problems into a unified MRF model to take advantages of both depth cues. As far as we know, there is no algorithm dedicated to dense stereo reconstruction in foggy weather conditions.

The article is structured as follows. In Sec. 2, we state the problem, and describe how fog affects the scene image. The classical dense stereo reconstruction and the image defogging problems are derived from a general formulation. In Sec. 3, our model of the stereo reconstruction and defogging problem is proposed. At last, Sec. 4 is dedicated to an evaluation on synthetic images and to tests on camera images.

2. Problem statement

The inputs are the left and right images of a stereo pair $\{I_L, I_R\}$. These images are observed after perturbation by atmospheric scattering and camera optics. The images without all these perturbations are denoted I_{0L} and I_{0R}, respectively, and are of course unknown. Also unknown is the depth map represented by its disparity map D. Our goal being to fuse depth cues from the stereo and from the atmospheric veil to achieve better reconstruction, it seems natural to search for a Bayesian formulation of the problem so that prior knowledge can be included to remove possible ambiguities. The two unknowns that we want to estimate are the disparity map D and the clean left image I_{0L}. The right one I_{0R} is not an unknown since, not considering occluded objects, it is a function of D and I_{0L}.

The maximum a posteriori principle tells us to maximize the following posterior probability, which can be rewritten using Bayes' rule as:

$$p(D, I_{0L}|I_L, I_R) \propto p(I_L, I_R|D, I_{0L})\ P(D, I_{0L}) \qquad (1)$$

where $p(I_L, I_R|D, I_{0L})$ is the data likelihood and $P(D, I_{0L})$ is the prior on the unknowns (D, I_{0L}). Instead of posterior probability maximization, in practice, it is its $-\log$ which is minimized, leading to the following formulation in terms of energy, or log-likelihood:

$$E(D, I_{0L}|I_L, I_R) = \underbrace{E(I_L, I_R|D, I_{0L})}_{E_{data}} + \underbrace{E(D, I_{0L})}_{E_{prior}} \qquad (2)$$

The term E_{data} is also known as the data cost or fidelity term, and E_{prior} as the prior or regularization term.

2.1 Dense stereo reconstruction without fog

Without fog, I_L and I_R are only affected by the noise of the sensor.

2.1.1 Data term

I_L and I_R are subject to a centered noise with same probability density function (pdf). The disparity D links the random variable I_R in the right image, with I_{0L}. The two random variables being independent, the data cost is obtained as a sum over $-\log$ the pdfs:

$$E_{data} = \sum_{(i,j)\in X} \rho_S\left(\frac{|I_{0L}(i,j) - I_L(i,j)|}{\sigma_S}\right) + \rho_S\left(\frac{|I_{0L}(i,j) - I_R(i - D(i,j), j)|}{\sigma_S}\right) \qquad (3)$$

where X is the set of image pixels, ρ_S is a function related to the distribution of the intensity noise with scale σ_S. This intensity noise takes into account the camera noise, but also the occlusion, and ρ_S can be one of the functions used in robust estimation to remove outliers.

In clear day scene, when the sensor noise is quite low, the intensity of the left image I_L tends to the intensity of the scene I_{0L}. Following [16], the Bayesian formulation of the dense stereo reconstruction is approximated by assuming that I_L is without noise. In (2), the unknown variable I_{0L} can be thus substituted by I_L leading to the approximate but simpler energy minimization:

$$E(D|I_L, I_R) = \underbrace{E(I_R|D, I_L)}_{E_{data\_stereo}} + \underbrace{E(D|I_L)}_{E_{prior\_stereo}} \tag{4}$$

E_{data_stereo} is the error in intensity between a pixel in the left image and a pixel in the right image given the disparity. By substitution of I_{0L} by I_L, it is obtained from (3) as:

$$E_{data\_stereo} = \sum_{(i,j)\in X} \rho_S\left(\frac{|I_L(i,j) - I_R(i - D(i,j), j)|}{\sigma_S}\right) \tag{5}$$

2.1.2 Prior term

This term enforces the smoothness of the disparity map. Because of constant intensity objects, the data term can be rather ambiguous. It is thus necessary to introduce a prior on the disparity map to interpolate the ambiguous areas correctly. The smoothness prior tells that two close pixels have a greater chance to be the projection of a same object with the same depth than remote pixels. This assumption is not always true due to gaps in depth for example. As a consequence, a robust function ρ_D should be used in this term. The classical prior term is:

$$E_{prior\_stereo} = \lambda_D \sum_{(i,j)\in X} \sum_{(k,l)\in N} W_D(\nabla I_L(i,j))\, \rho_D(|D(i,j) - D(i+k, j+l)|) \tag{6}$$

where λ_D is a factor weighting the strength of the prior on D, N is the set of relative positions of pixel neighbors (i, j) (4, 8 connectivity or other), and W_D is a monotonically decreasing function of image intensity gradients. The weight W_D is introduced to smooth low-gradient ambiguous areas more than gradient edges. Usually W_D is chosen as a decreasing exponential function of the image gradient: $W_D(\nabla I) = e^{-\frac{|\nabla I|}{\sigma_g}}$, where σ_g is a scale parameter. It is even better to use a function of the image Laplacian in order to avoid sensitivity to linear intensity variations: $W_D(\nabla I) = e^{-\frac{|\Delta I|}{\sigma_g}}$. When a segmentation of the image in objects is available, the image I can be substituted by this segmentation, with advantages, in the weight W_D.

2.2 Effects of fog

With a linear response camera, assuming an object of original intensity I_0, the apparent intensity I in presence of a fog with extinction coefficient β is modeled by Koschmieder law:

$$I = I_0 e^{-\beta p} + \underbrace{I_s(1 - e^{-\beta p})}_{V} \tag{7}$$

where p is the object depth, and I_s is the intensity of the sky. From (7), it can be seen that fog has two effects: first an exponential decay $e^{-\beta p}$ of the original intensity I_0, and second the addition of the atmospheric veil V which is an increasing function of the object distance p. The depth p can be rewritten as a function of the disparity $p = \frac{\delta}{D}$ where δ is related to the stereo calibration parameters.

After substitution, (7) becomes:

$$I = I_0 e^{-\frac{\beta\delta}{D}} + I_s(1 - e^{-\frac{\beta\delta}{D}}) \tag{8}$$

It is important, for the following, to notice that there is one situation where D can be obtained from a single image using V: when I_0 is close to zero, i.e when the object is black. It is also important to notice that when the disparity D is zero, the intensity I_0 cannot be obtained. Moreover, I_0 being positive, the photometric constraint $V < I$ is deduced from (7).

The parameter β is assumed known in the following, as well as I_s. I_s can be obtained simply as the maximum over the image or using a more dedicated algorithm [12]. Depending of the application field, the value of β can be measured from a visibility-meter, or estimated from the images, see [6] for road images.

2.3 Single image defogging knowing the depth

Before we describe our model for fused stereo reconstruction and defogging, we focus on the simpler problem of defogging from a single image I given the disparity map D. Using the previous notations, only the left image is used in this section. We thus drop L in the indexes. The unknown I_0 is the image without fog and noise. Both I and D are assumed known. The defogging problem knowing the disparity D can be set as a particular case of (1), i.e the maximization of the posterior probability:

$$p(I_0|D, I) \propto p(I|D, I_0)P(I_0|D)\; P(D) \tag{9}$$

or equivalently as the minimization of the energy:

$$E(I_0|D, I) = \underbrace{E(I|D, I_0)}_{E_{data\_fog}} + \underbrace{E(I_0|D)}_{E_{prior\_fog}} \tag{10}$$

2.3.1 Data term

The data term is the log-likelihood of the noise probability on the intensity, taking into account that I_0 is observed through the fog, see (7):

$$E_{data\_fog} = \sum_{(i,j)\in X} \rho_P\left(\frac{|I_0(i,j)e^{-\frac{\beta\delta}{D(i,j)}} + I_s(1 - e^{-\frac{\beta\delta}{D(i,j)}}) - I(i,j)|}{\sigma_P}\right) \tag{11}$$

where ρ_P is a function related to the intensity noise due to the camera and σ_P is the scale of this noise. ρ_P and σ_P are thus directly related to the probability density function (pdf) of the camera noise and can be estimated off-line when calibrating the camera. It can be noticed for D close to zero that the data term does not constrain the distribution of I_0 which tends to the uniform pdf.

2.3.2 Prior term

This term enforces the smoothness of the restored image. This term is necessary to handle the image intensity noise. The smoothness prior tells that two close pixels have a greater chance to have similar intensity when at similar disparities, compared to pixels with large disparity changes. This assumption is valid for objects of uniform color. As a consequence, a robust function ρ_{I_0} should be used in this term. We found that the following prior

Fig. 2 Comparison of defogging results knowing the same depth map. From left to right: the input foggy image, the restoration result without regularization and without noise, the restoration result without regularization and with noise on the input image, previous method with smoothing of the input image, our restoration result with noise on the image first without the extra factor, and with the factor added on the regularization term.

term produces nice restoration results:

$$E_{prior\_fog} = \lambda_{I_0} \sum_{(i,j)\in X} \sum_{(k,l)\in N} e^{-\frac{\beta\delta}{D(i,j)}} W_{I_0}(\nabla D(i,j))\, \rho_{I_0}(|I_0(i,j) - I_0(i+k, j+l)|) \tag{12}$$

where λ_{I_0} is a factor weighting the strength of the prior on image I_0. Function W_{I_0} plays the same role than W_D in the stereo, only now it is applied on the disparity map gradient rather than on image gradient. We use $W_{I_0}(\nabla D) = e^{-\frac{|\Delta D|}{\sigma'_g}}$, where σ'_g is a scale parameter. An extra weight $e^{-\frac{\beta\delta}{D(i,j)}}$ is introduced, and it is a key point, to take into account that in presence of fog, there is an exponential decay of contrast with respect to (w.r.t.) depth. This has the effect of giving less and less importance to the prior as depth increases. This is necessary to be consistent with the fact that the distribution of I_0 is less and less constrained by the data term for large distances. Without this extra factor, the intensity of close objects may wrongly diffuse on remote objects. Fig. 2 shows the impact of the extra weight.

2.4 Optimization

While MRF formulations are successful to model image processing and computer vision problems, it is also necessary to have reliable optimization algorithms to minimize the derived energies. A large class of useful MRF energies is of the form:

$$f(L) = \sum_{x\in X} \Phi_x(L_x) + \sum_{x\in X, x'\in X} \Phi'_{x,x'}(L_x, L_{x'}) \tag{13}$$

where L_x is a label located in x. The previously introduced energy is in this class since regularization term is of first order. When the function Φ' is sub-modular, it has been shown that the global minimum of the previous problem can be obtained in polynomial time.

2.4.1 α-expansion

A particularly fast algorithm which can be used to find a local minima of (13) is the fusion move introduced in [9]. From two label sets, the result of the fusion move is obtained as a combination of these labels which minimize (13). Let L^t a label set which is our current solution of (13), and let L^p a label set which is a proposal solution we would like to test. The result of the fusion move is the label set L^b described by the following linear combination:

$$L^b(B) = (1 - B)L^t + BL^p \tag{14}$$

where B is binary. B is selected in such a way that the energy (13) is decreased, thus:

$$\min_B f(L^b(B)) \tag{15}$$

When Φ' is sub-modular, the resulting boolean optimization problem is still sub-modular and the optimal B can be found in polynomial time. The fusion move is thus guaranteed to reduce locally the energy (13). Consequently, the fusion move can be iterated using different proposal L^p, and the energy will decrease at each iteration.

The well-known α-expansion algorithm is a special case of fusion move where each proposal L^p is made of a unique label. It is guaranteed to find a local minimum of the energy (13).

3. Stereo reconstruction and defogging

The model we now propose for fused stereo reconstruction and defogging shares similarities with the single image defogging model presented in [11]. Indeed in [11], the model is set as a MRF model and both depth p and restored image I_{0L} are estimated successively. The main difference is that stereo images are used as input in our approach, while the approach in [11] is monocular. In particular, this last approach cannot work with gray-level images, contrary to our stereo approach. Another difference is that, in [11], the Koschmieder's law (7) is rewritten using new variables, like the log of the intensity. This rewriting implies that a uniform additive noise is transformed non-linearly differently depending of the intensity value. We thus preferred not to proceed in such a way.

3.1 MRF model

3.1.1 Data term

In stereo with fog, the data term (11) applies on the left image. On the right, a similar term taking into account the disparity D is also introduced. This leads to the following log-likelihood of the stereo data in fog:

$$E_{data\_fog\_stereo} = \sum_{(i,j)\in X} \rho_P\left(\frac{|I_{0L}(i,j)e^{\frac{-\beta\delta}{D(i,j)}} + I_s(1 - e^{\frac{-\beta\delta}{D(i,j)}}) - I_L(i,j)|}{\sigma_P}\right) + \rho_P\left(\frac{|I_{0L}(i,j)e^{\frac{-\beta\delta}{D(i,j)}} + I_s(1 - e^{\frac{-\beta\delta}{D(i,j)}}) - I_R(i - D(i,j), j)|}{\sigma_P}\right) \tag{16}$$

Notice that when $\beta = 0$, i.e without fog, the first term in (16) enforces $I_{0L} = I_L$, and the second term is the stereo log-likelihood E_{data_stereo}. This shows that D can be estimated from both log-likelihoods. We thus propose to add the following log-likelihoods with a weight α:

$$E_{data^*} = E_{data\_fog\_stereo} + \alpha E_{data\_stereo} \tag{17}$$

During the estimation of both I_{0L} and D, the value of I_{0L} can be temporarily far from the true value. The advantage of introducing E_{data_stereo} in the data term is that the minimization of E_{data_stereo}

provides correct estimates of D at short distances even if I_{0L} is badly estimated.

3.1.2 Photometric constraint and assumption on white pixels

As introduced in Sec. 2.2, the photometric constraint on the atmospheric veil V must be verified both on the left and right images of the stereo pair. Due to noise, the photometric constraint is not very strict but it helps to reduce the search space of I_{0L}.

Due to fog, the contrast of remote objects is very low and stereo does not work. As remote objects are nearly white, we add a zero disparity assumption on those pixels with an intensity equal to I_s. This assumption is of course wrong for white objects. Taking into account the photometric constraint and the assumption on white pixels, the data term is:

$$E_{data} = \begin{cases} E_{data^*} & \text{if } V(i,j) \leq I_L(i,j) + 3\sigma_P \\ & \text{and } V(i,j) \leq I_R(i - D(i,j), j) + 3\sigma_P \\ & \text{and } I_L(i,j) \neq I_s \\ 0 & \text{if } I_L(i,j) = I_s \text{ and } D(i,j) = 0 \\ +\infty & \text{else.} \end{cases} \tag{18}$$

3.1.3 Prior term

In (1), the prior probability $P(D, I_{0L})$ is related to two variables: the disparity D and the intensity I_{0L}. Unfortunately, it is not clear which probability function is a good choice for this mixed prior term, knowing that the two variables D and I_{0L} cannot be assumed independent of one another. Our idea is thus to take advantages of the previously introduced prior terms: the stereo prior term (6) and the defogging prior term (12). We thus propose to write the mixed prior probability as $P(D, I_{0L}) = P(D|\ddot{I}_{0L})P(I_{0L}|\ddot{D})$, where $\ddot{D}$ and $\ddot{I}_{0L}$ are fixed approximations of D and I_{0L}, given as priors. The log-probability of D given $\ddot{I}_{0L}$ is assumed as in (6) and the log-probability of I_{0L} given $\ddot{D}$ is assumed as in (12). We thus propose the following prior term for the stereo reconstruction and defogging problem:

$$\begin{aligned} E_{prior} = & \sum_{(i,j)\in X} \sum_{(k,l)\in N} \lambda_D \; W_D(\nabla \ddot{I}_{0L}(i,j)) \; \rho_D(|D(i,j) - D(i+k, j+l)|) \\ & + \lambda_{I_0} \; e^{-\frac{\beta\delta}{D(i,j)}} \; W_{I_0}(\nabla \ddot{D}(i,j)) \; \rho_{I_0}(|I_{0L}(i,j) - I_{0L}(i+k, j+l)|) \end{aligned} \tag{19}$$

Similarly to section 2.1, the weight W_D is set to $W_D(\nabla \ddot{I}) = e^{-\frac{|\Delta \ddot{I}|}{\sigma_g}}$. Similarly to section 2.2, the weight W_{I_0} is set to $W_{I_0}(\nabla \ddot{D}) = e^{-\frac{|\Delta \ddot{D}|}{\sigma_g}}$. The weights W_D and $W_{I_{0L}}$ in E_{prior} are introduced to attenuate the regularization through object edges. Due to the exponential decreasing function in these two weights, a not too large error on $\ddot{D}$ or $\ddot{I}_{0L}$ usually leads to a small variation of the associated weight value. This justify why the previous approximation is not too bad.

3.1.4 Initial $\ddot{D}$ and $\ddot{I}_{0L}$

We propose to rely on an approximate atmospheric veil $\ddot{V}$ to find the approximates $\ddot{D}$ and $\ddot{I}_{0L}$. The atmospheric veil can be approximately estimated on the left image using a single image defogging algorithm, see for instance [2], [13], [15]. Here, it is approximated by minimizing the following w.r.t. $\ddot{V}$:

$$\sum_{(i,j)\in X} |I_L(i,j) - \ddot{V}(i,j)| + \lambda \sum_{(k,l)\in N} |\ddot{V}(i,j) - \ddot{V}(i+k, j+l)| \tag{20}$$

using α-expansion. The small features in the image I_L are lost in $\ddot{V}$, but thanks to the L_1 robust terms in (20), large objects with low contrast are kept. This atmospheric veil $\ddot{V}$ has the important property: it contains object edges.

By definition from (7), $V = 1 - e^{-\frac{\beta\delta}{D}}$. As a consequence, $\ddot{D}$ can be obtained from $\ddot{V}$. This implies that the factor $e^{-\frac{\beta\delta}{D}}$ in E_{prior} can be substituted by $1 - \ddot{V}$. Another consequence is that $\Delta\ddot{D}$ in W_{I_0} can be approximated by $\Delta\ddot{V}$. Moreover, rather than to rely on a close approximation of $\ddot{I}_{0L}$, we can substitute $\Delta\ddot{I}_{0L}$ by $\Delta\ddot{V}$ in the weight W_D, following the remark in the end of section 2.1.2. These approximations allows to rewrite the prior E_{prior} as a function of $\ddot{V}$ rather than a function of $\ddot{D}$ and $\ddot{I}_{0L}$.

3.1.5 Complete model

In summary, the stereo reconstruction and defogging problem is set at the following minimization:

$$\min_{D, I_{0L}} E_{data} + E_{prior} \tag{21}$$

In our experiments, the functions ρ_D and ρ_{I_0} are chosen as the identity function. The noise on the image being assumed Gaussian, ρ_P is selected as the square function. Therefore, for those pixels which verify the photometric constraint and which are not white, the energy which is minimized is, after rewriting using $\ddot{V}$:

$$\begin{aligned} E(D, I_{0L}, \sigma_p) = & \sum_{(i,j)\in X} \left\{ \frac{1}{\sigma_P^2} \left(|I_{0L}(i,j) e^{\frac{-\beta\delta}{D(i,j)}} + I_s(1 - e^{\frac{-\beta\delta}{D(i,j)}}) - I_L(i,j)|^2 \right.\right. \\ & \left. + |I_{0L}(i,j) e^{\frac{-\beta\delta}{D(i,j)}} + I_s(1 - e^{\frac{-\beta\delta}{D(i,j)}}) - I_R(i - D(i,j), j)|^2 \right) \\ & + \alpha \; \rho_S\left(\frac{|I_L(i,j) - I_R(i - D(i,j), j)|}{\sigma_S}\right) \\ & + \sum_{(k,l)\in N} \left\{ \lambda_{I_0}(1 - \ddot{V}(i,j)) e^{-\frac{|\Delta\ddot{V}(i,j)|}{\sigma_g}} |I_{0L}(i,j) - I_{0L}(i+k, j+l)| \right. \\ & \left. \left. + \lambda_D \; e^{-\frac{|\Delta\ddot{V}(i,j)|}{\sigma_g}} |D(i,j) - D(i+k, j+l)| \right\} \right. \\ & + 4log(\sigma_p) \Big\} \end{aligned} \tag{22}$$

As this energy is known up to a scale factor, (22) can be arbitrarily divided by λ_{I_0}. This is used in the next section to estimate σ_P from image residuals. In order to take into account color images, it is enough to sum the cost for each color channel. Moreover, the photometric constraint is applied on each color channel.

3.2 Optimization

In (22), D and I_{0L} appear in non-linear unary functions and independently in binary functions. It is thus possible to optimize (22) by means of a two-step alternate minimization: one step consists in minimizing w.r.t. I_{0L} and the other in minimizing w.r.t. D. The first step is defogging and the second step is stereo reconstruction. The energies in both steps being sub-modular, α-expansion is used for the minimization. With the alternate minimization, convergence towards a local minima is guaranteed. Before the first step, the disparity is initialized by stereo reconstruction assuming no fog.

As pointed in [11], the gradient distribution of a hazy image can be very different from that of a foggy image. This implies that after division of (22) by λ_{I_0}, the factor $\sigma_P \sqrt{\lambda_{I_0}}$ must be set differently from one image to another. When this factor is not correctly set, the chance to converge towards an interesting local minimum decreases. Hopefully, the first term of (22) being quadratic, the factor $\sigma_P \sqrt{\lambda_{I_0}}$ can be easily estimated by estimating the standard deviation of the intensity residuals of the data term (16).

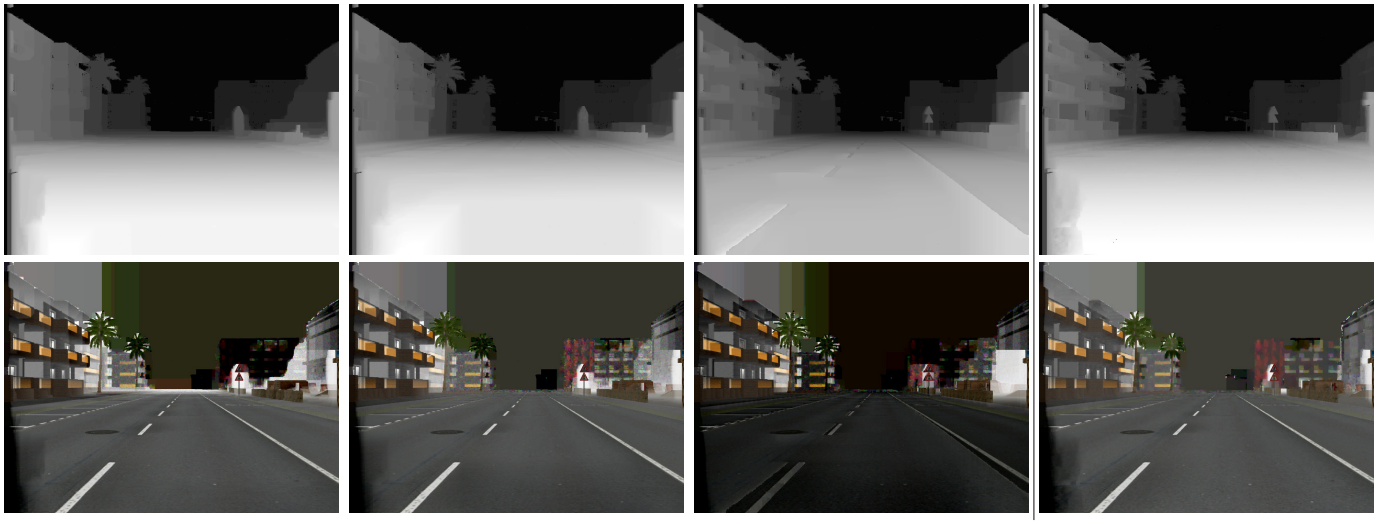

Fig. 3 Effect of the choice of α and $\sigma_P\sqrt{\lambda_{I_0}}$ on obtained results. First column: disparity map and restoration obtained with $\alpha = 0$ and $\sigma_P\sqrt{\lambda_{I_0}} = 2$. Second column: for $\alpha = 0$ and $\sigma_P\sqrt{\lambda_{I_0}} = 5$. Third column: for $\alpha = 0$ and $\sigma_P\sqrt{\lambda_{I_0}} = 10$. Last column: $\alpha = 4$ and $\sigma_P\sqrt{\lambda_{I_0}} = 20$.

In summary, the optimization scheme is:

- Compute $\ddot{V}$ by minimizing (20), using α-expansion.
- Initialize D by minimizing $\alpha E_{data\_stereo} + E_{prior}$ w.r.t D, using α-expansion.
- Until convergence, iterate:
  - Until convergence, iterate:
    - Minimization of (22) w.r.t I_{0L}, using α-expansion.
    - Minimization of (22) w.r.t D, using α-expansion.
  - Update σ_P by computing the standard deviation of the intensity residuals.
- $\sigma_P\sqrt{\lambda_{I_0}}$ is enforced to value 1 and a last optimization w.r.t. I_{0L} is performed to better emphasize the detailed texture.

In this algorithm, the choice of the value of α is important. Fig. 3 shows the obtained results after algorithm convergence with different values of α. When $\alpha = 0$ results are not correct at close distances. This is due to the first order prior which enforces a fronto-parallel solution when the scale $\sigma_P\sqrt{\lambda_{I_0}}$ is large. When $\alpha = 4$, thanks to the use of E_{data_stereo}, the obtained solution is better at close distances.

At each step, the energy (22) is decreased as illustrated in Fig. 4. This figure shows several iterations during the optimization with $\sigma_P\sqrt{\lambda_{I_0}} = 20$. One can notice that, after one iteration, the large scale of $\sigma_P\sqrt{\lambda_{I_0}}$ allows a better reconstruction and restoration around the closest vehicle. When the number of iteration increases, the scale $\sigma_P\sqrt{\lambda_{I_0}}$ becomes smaller, and the restoration is improved step by step for remote objects. Thus, the two far away vehicles appear.

3.2.1 Refinement using second order prior

In the previous model, thanks to the sub-modular properties of the energy, each MRF layer can be efficiently optimized using α-expansion. However, first order prior gives advantages to fronto-parallel solutions and thus it is not fully adequate for planar objects which are not fronto-parallel. When the scene is made of multiple planes in various directions, a refinement using a prior based on the the second order derivative of the disparity is interesting to apply. During the refinement, only the regularization

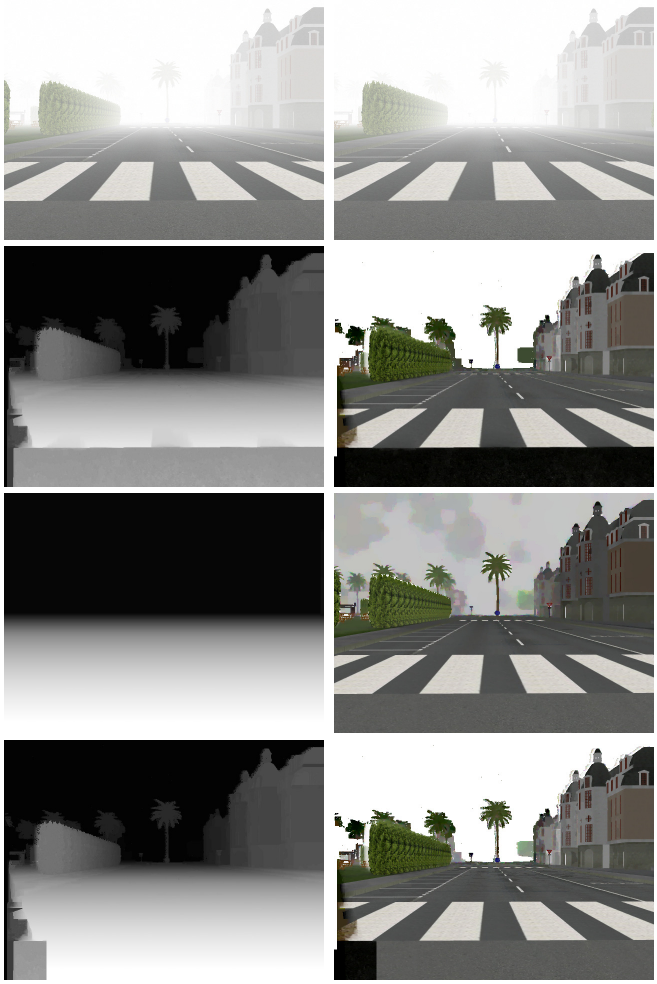

Fig. 5 First line, the left and right foggy stereo pair. Second line, the disparity map and the restored image obtained with the first order prior. Third line, the road plane and single image defogging obtained by [2]. Fourth line, the disparity map and restoration obtained after fusion move of the images in two previous lines.

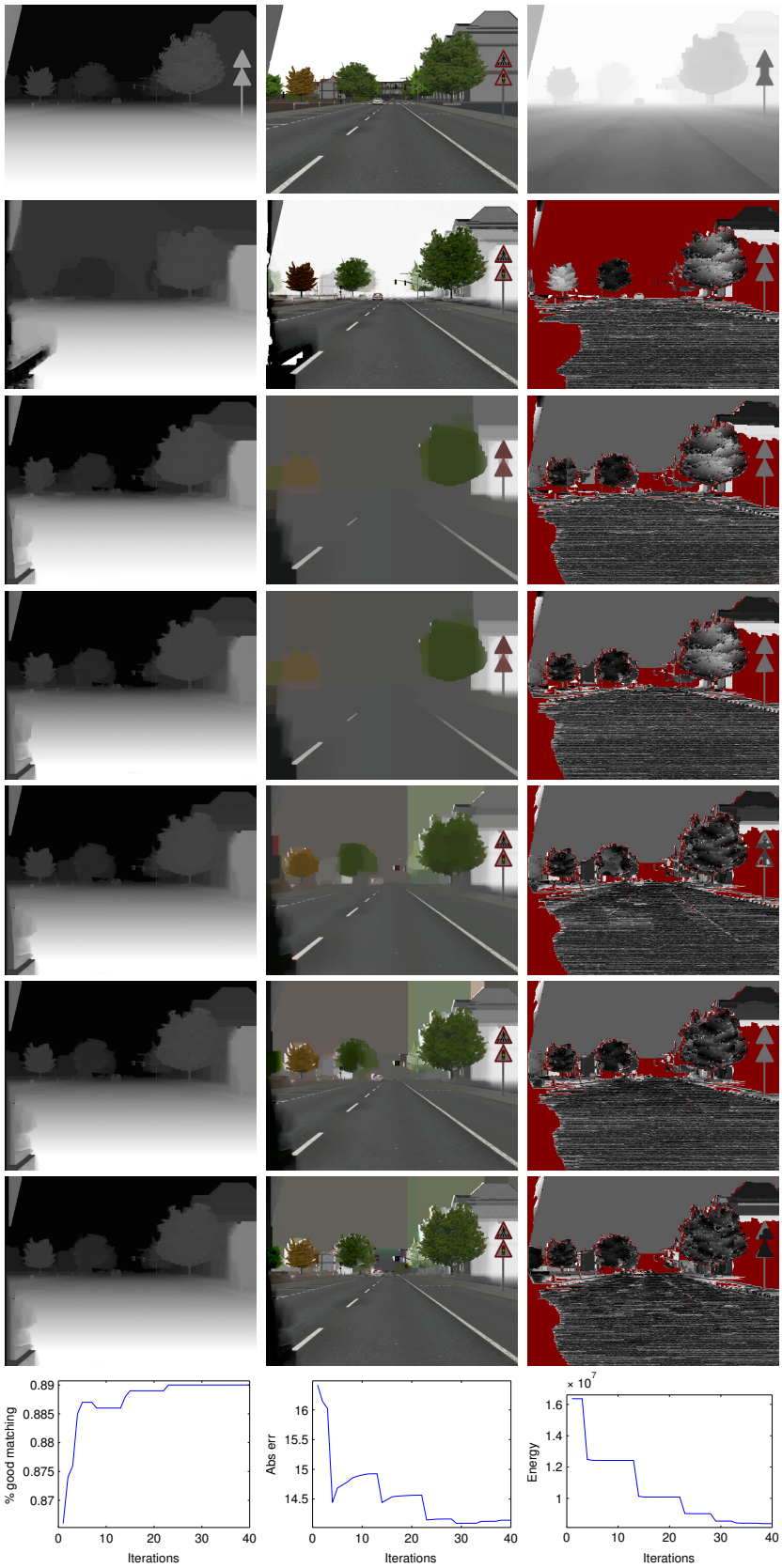

Fig. 4 First line, from left to right: the ground truth disparity map, the image without fog and the estimated veil $\breve{V}$. Second line: the disparity map obtained with stereo reconstruction without considering fog, the restored image with $\sigma_P \sqrt{\lambda_{I_0}} = 1$ and the associated disparity error map. Third line: solution during initialization when $\sigma_P \sqrt{\lambda_{I_0}} = 20$. Fourth line: solution after one iteration. Fifth line: after 10 iterations. Sixth line: after 20 iterations. Seventh line: after convergence. Last line: evolution along iterations of the number of good matches, of the absolute error between the restored image and the image without fog, and of the energy.

term on the disparity is changed, and it involves triple cliques which are more complex to handle than binary cliques involved with a first order term.

In [16], a complete scheme is proposed for stereo reconstruction with a second order prior. The main idea is to use fusion move with, at each iterations, a new proposal, until converging. The difficulty is to choice the good proposal which leads to good convergence properties. Indeed, if the proposal is not well chosen, the energy is subject to fall into a bad local optimum. In order to explore a large set of solutions, the following proposal are proposed in [16]:

- a disparity plane obtained by fitting segmented areas,
- a smooth version of the current solution,
- random solution obtained from a uniform distribution.

When a second order prior is selected, the fusion move algorithm can be also used with advantages to minimize the proposed model (22). The idea consists to estimate the solution with the first order prior as described before with eventually a reduced number of iterations. Then, proposal solutions are combined iteratively with the first order solution using fusion moves. Two kinds of proposal can be used: proposal on the disparity map and proposal on the restored left image.

For road images, most of the time, the road plane is known in the neighborhood of the vehicle. As pointed in the introduction, the road area is also rather uniform in color, thus is an ambiguous area for stereo reconstruction. To illustrate these difficulties, Fig. 5 shows the result of the previously proposed algorithm on a stereo pair, and one can see in the bottom of the obtained disparity map that the disparity is wrongly constant. In this ambiguous area, the first order prior forces the solution to be fronto-parallel. The planar road model introduced in [14] can be used as a proposal D^p for the disparity map D.

The restored image obtained using a single image defogging algorithm can be used as a proposal I_0^p for I_0. Fig. 5 shows an example of defogging obtained using a single image defogging algorithm described in [2].

The question is then how fusion move can be applied in this situation where a pair of label sets I_0^p and D^p is proposed? This can be solve easily by considering the fusion of the variable (D, I_0) with the variable (D^p, I_0^p).

Fig. 5 illustrates a fusion move step between the result of the first order scheme and a proposal composed by the pair: the road plane for the disparity, and the result of a single image defogging algorithm with the planar assumption for the restored intensity. The obtained result shows nice improvements, particularly in the bottom area of the disparity map. Notice also that the buildings and others gray objects like the sidewalk close to the road, are not modified since their are too dark in the proposal of restored image.

4. Evaluation

4.1 Parameters setting

The proposed MRF model is mainly parametrized by α which is the weight between the photometric log-likelihood $E_{photo\_fog\_stereo}$ of left and right images and the log-likelihood E_{photo_stereo} of the stereo. When α is close to zero, the obtained disparity map is smooth in homogeneous areas, but the disparity of close objects may be biased as well as the intensity I_{0L}. When α is large, the disparity obtained from the stereo log-likelihood is usually correct for close objects but the quality of the reconstruction decreases with the contrast and thus with the depth. Therefore, we usually set α to the value 4.

Another important parameter is the initial value of $\sigma_P \sqrt{\lambda_{I_0}}$. The choice of this value can have an effect on the local minima selected at convergence. The bigger $\sigma_P \sqrt{\lambda_{I_0}}$ at the initialization, the smoother is the depth map after convergence.

4.2 Synthetic images

To evaluate the stereo reconstruction and image defogging in foggy weather, we rely on synthetic images due to the difficulty to have the 3D model of a scene and images of this scene with and without fog. We generate synthetic stereo images using SiVIC software which allows to build physically-based road environments. Uniform fog is added knowing the depth map, see Fig. 6. To make the image more realistic and evaluate the ability of the algorithm to manage the noise, we also added a Gaussian noise on every pixels of left and right images, with standard deviation 1. This database is named FRIDA3 and is available online for comparative studies on stereo reconstruction[*1]. For image defogging evaluation, we use the database FRIDA2, which is composed by the left image of FRIDA3. The database FRIDA2 was used in [2], [13] to compare single image defogging algorithms.

4.3 Stereo reconstruction evaluation

For the stereo reconstruction evaluation, we compared the result of the proposed algorithm with three other algorithms: the standard Markov random field algorithm (MRF_Stereo, see Sec. 2.1), the semi-global matching algorithm proposed in [8] (SGBM) and the Elas algorithm [3]. Each method is applied on the original foggy stereo pair and on the stereo pair after restoration. Two different restoration methods are tested: the No-Black-Pixel-Constrain (NBPC) [15] and the method based on a MRF modeling (MRF_Defog) described in [2]. For the proposed algorithm, the evaluation is decomposed in three steps: the initialization using the photometric constraint, after convergence (with initial $\sigma_P \sqrt{\lambda_{I_0}} = 20$ and $\alpha = 4$), and finally, after refinement using fusion with the planar assumption.

Results shown in Tab. 1 are obtained in average on 66 stereo pairs. These percentages take into account only the pixels seen in both images with a disparity larger than one, i.e not considering the sky. When applied on the foggy images, the MRF_Stereo algorithm achieves 81.2% of correct disparities, for a maximum disparity error of 1 pixels, SGMB achieves 73.9 and Elas achieves 82.2%. When the reconstruction is performed on restored images, the scores is not better, and in some case, is drastically reduced. This is mainly due to artifacts introduced by the restoration: details removed when MRF_Defog is used, or emphasis of the noise when NBPC is used. The proposed algorithm always achieve better results, even at the first step, when the photometric constraint is introduced (Initialization), with a percentage of correct dispar-

[*1] http://perso.lcpc.fr/tarel.jean-philippe/visibility/fogstereo.zip

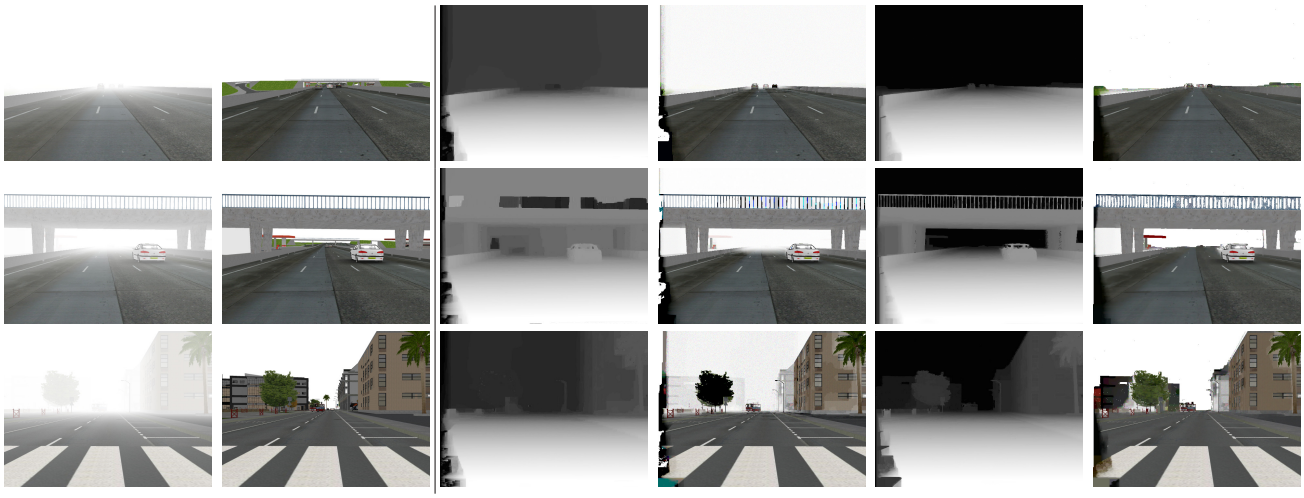

Fig. 6 Results on three images of the synthetic FRIDA3 stereo image database. First column: foggy left images. Second column: same scene without fog. Third and fourth columns: disparity maps obtained using standard stereo reconstruction and restored images using these disparity maps. Fifth and sixth columns: disparity maps with the proposed method and associated restored images.

Table 1 Comparison of the percentage of correct disparities in average on 66 synthetic stereo pairs using different algorithms. First: the standard MRF approach without fog (MRF_Stereo, see Sec. 2.1), second: a semi-global matching algorithm proposed in [8] (SGBM) and the Elas algorithm [3] applied on the original foggy stereo images. Then, both algorithms are applied on restored stereo pairs using the methods proposed in [2] (MRF_Defog) and in [15] (NBPC). Results on three steps of the proposed algorithm are displayed: first iteration (Initialization), after converging (Final) and after refinement with fusion move (Refined). Percentages are given for different values of the maximum error *err* on the disparity (in pixel).

Algorithm	$err < 1$	$err < 0.66$	$err < 0.33$
Input: original foggy images			
MRF_Stereo	81.2	75.9	53.8
SGBM	73.9	62.5	40.6
Elas	82.2	72.9	46.7
Input: restored images with MRF_Defog			
MRF_Stereo	82.5	77.0	53.2
SGBM	73.9	62.5	40.6
Elas	74.4	61.6	37.5
Input: restored images with NBPC			
MRF_Stereo	74.4	61.6	37.5
SGBM	72.7	62.4	41.3
Elas	77.6	68.6	45.2
Proposed algorithm on original foggy images			
Initialization	85.8	81.4	58.8
Final	86.9	82.6	61.8
Refined	93.4	89.5	75.7

ities of 85.8%. This step corresponds to the initialization of the proposed algorithm. After convergence (Final), this percentage is increased to 86.9%. Finally, the refinement (Refined) with the fusion move using the the planar road assumption gives even a better result at 93.4%. In percentage, the improvement due to iterations may seem reduced on the whole image, but these iterations are important to improve correct disparities at long distances. This fact is illustrated in Fig. 6 which displays obtained disparity maps and restored left images on three stereo pairs of the FRIDA3 database.

4.4 Restoration evaluation

Table 2 Average absolute error between restored image and target image without fog, for 9 defogging methods on 66 synthetic images with uniform noise and with standard deviation 1.

Algorithm	mean error (in gray levels)
Nothing	81.6 ± 12.3
DCP [7]	46.3 ± 15.6
FSS [5]	34.9 ± 15.1
NBPC [15]	50.8 ± 11.5
NBPC+PA [14]	31.1 ± 10.2
CM+PA [4]	19.1 ± 6.7
MRF+PA [2]	16.9 ± 5.1
Proposed	22.9 ± 18.8
Proposed+refined	18.9 ± 13.0

For image defogging, we compared the proposed method after convergence to 6 methods also evaluated in [2], [13]. Results are shown in Tab. 2. In this comparison, pixels with horizontal coordinate lower than the maximum disparity are not considered. Indeed, pixels in this area can be incorrectly reconstructed due to image borders. The proposed result, with refinement using the planar assumption, is second and fourth without refinement.

4.5 Camera images

We compared the proposed method to the stereo reconstruction without fog described in Sec. 2.1 and image defogging described in [15]. β is manually selected. Results show that both the reconstruction and restoration are of better quality. In Fig. 7, results

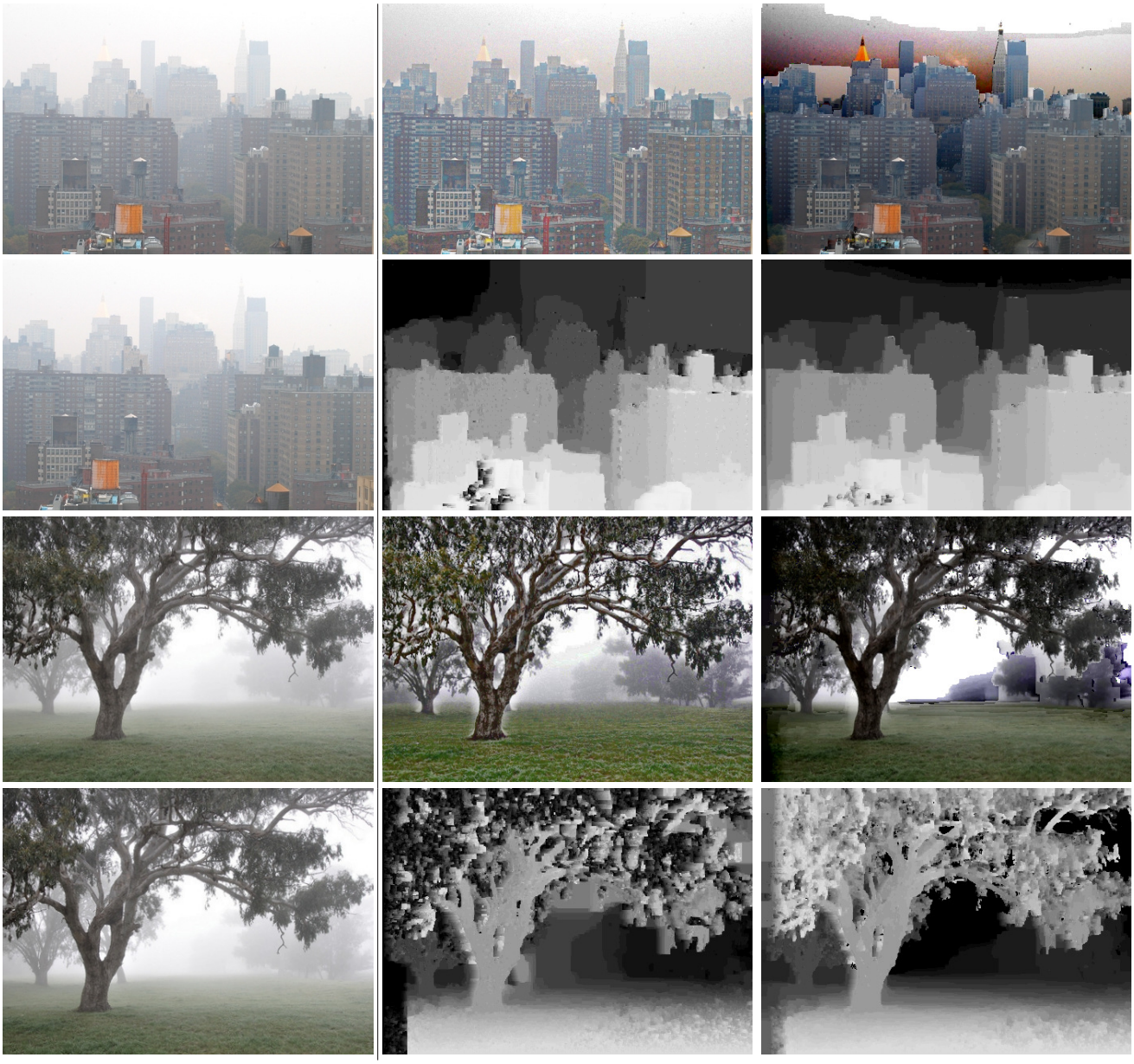

Fig. 7 First column: foggy stereo pair. Second column: Single image defogging with [15] and disparity map obtained by standard stereo reconstruction. Third column: restored left image and disparity map obtained using the proposed method.

are compared on urban and country side stereo pairs. One may note that the obtained stereo reconstruction are dense at both short and long distances, contrary to stereo reconstruction without taking into account the fog. The stereo restoration obtained by the proposed method is of good quality compared to single image defogging results. At close distances, outliers are avoided thanks to the photometric constraint and the true intensity of objects is kept. At far distances, the contrast is greatly enhanced without amplifying the noise to much.

5. Conclusion

We proposed a MRF model to solve the stereo reconstruction and image defogging in daytime fog. It is an extension of two sub-models: the classical stereo reconstruction without fog and newly introduced image defogging when the depth is known. The proposed model includes the photometric constraint and priors on white pixels. It leads to the optimization of an energy which can be solved by an alternate scheme based on the application of successive α-expansion optimizations. The convergence towards a local minimum is thus guaranteed. Tests on both synthetic stereo pairs and camera stereo pairs show the relevance of the model. Thanks to the stereo depth clue, the disparity is correct at short distances, and thanks to the atmospheric veil depth cue, the disparity is drastically improved at long distances. The restoration results on close objects are better than the ones obtained without stereo thanks to the simultaneous estimation with the disparity map. Perspectives for future research are to take into account non constant sky, non Gaussian noise to improve scale estimation, to explicitly take into account occlusions in the formulation, to speed up the algorithm for real time applications and to extend the previous model to heterogeneous fog.

References

[1] Boykov, Y., Veksler, O. and Zabih, R.: Fast Approximate Energy Minimization via Graph Cuts, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 23, pp. 1222–1239 (2001).

[2] Caraffa, L. and Tarel, J.-P.: Markov Random Field Model for Single Image Defogging, *Proceedings of IEEE Intelligent Vehicle Symposium (IV'2013)*, Gold Coast, Australia, pp. 994–999 (online), available from ⟨http://perso.lcpc.fr/tarel.jean-philippe/publis/iv13.html⟩ (2013).

[3] Geiger, A., Roser, M. and Urtasun, R.: Efficient large-scale stereo matching, *Proceedings of the 10th Asian conference on Computer vision*, ACCV'10, Vol. Part I, pp. 25–38 (2011).

[4] Halmaoui, H., Cord, A. and Hautière, N.: Contrast restoration of road images taken in foggy weather, *Computational Methods for the Innovative Design of Electrical Devices*, pp. 2057–2063 (2011).

[5] Hautière, N., Tarel, J.-P. and Aubert, D.: Mitigation of Visibility Loss for Advanced Camera based Driver Assistances, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 11, No. 2, pp. 474–484 (online), available from ⟨http://perso.lcpc.fr/tarel.jean-philippe/publis/its10.html⟩ (2010).

[6] Hautière, N., Tarel, J.-P., Lavenant, J. and Aubert, D.: Automatic fog detection and estimation of visibility distance through use of an onboard camera, *Machine Vision and Applications*, Vol. 17, No. 1, pp. 8–20 (online), available from ⟨http://perso.lcpc.fr/tarel.jean-philippe/publis/mva06.html⟩ (2006).

[7] He, K., Sun, J. and Tang, X.: Single Image Haze Removal using Dark Channel Prior, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 12, pp. 2341–2353 (2010).

[8] Hirschmuller, H.: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information, *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02*, CVPR '05, Washington, DC, USA, IEEE Computer Society, pp. 807–814 (online), DOI: 10.1109/CVPR.2005.56 (2005).

[9] Lempitsky, V., Rother, C., Roth, S. and Blake, A.: Fusion Moves for Markov Random Field Optimization, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 32, No. 8, pp. 1392–1405 (online), DOI: 10.1109/TPAMI.2009.143 (2010).

[10] Narashiman, S. G. and Nayar, S. K.: Interactive Deweathering of an Image using Physical Model, *IEEE Workshop on Color and Photometric Methods in Computer Vision*, Nice, France (2003).

[11] Nishino, K., Kratz, L. and Lombardi, S.: Bayesian Defogging, *International Journal of Computer Vision*, Vol. 98, pp. 263–278 (2012).

[12] Tan, R.: Visibility in bad weather from a single image, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*, Anchorage, Alaska, pp. 1–8 (2008).

[13] Tarel, J.-P., Hautière, N., Caraffa, L., Cord, A., Halmaoui, H. and Gruyer, D.: Vision Enhancement in Homogeneous and Heterogeneous Fog, *IEEE Intelligent Transportation Systems Magazine*, Vol. 4, No. 2, pp. 6–20 (online), available from ⟨http://perso.lcpc.fr/tarel.jean-philippe/publis/itsm12.html⟩ (2012).

[14] Tarel, J.-P., Hautière, N., Cord, A., Gruyer, D. and Halmaoui, H.: Improved Visibility of Road Scene Images under Heterogeneous Fog, *Proceedings of IEEE Intelligent Vehicle Symposium (IV'2010)*, San Diego, California, USA, pp. 478–485 (online), available from ⟨http://perso.lcpc.fr/tarel.jean-philippe/publis/iv10b.html⟩ (2010).

[15] Tarel, J.-P. and Hautière, N.: Fast Visibility Restoration from a Single Color or Gray Level Image, *Proceedings of IEEE International Conference on Computer Vision (ICCV'09)*, Kyoto, Japan, pp. 2201–2208 (online), available from ⟨http://perso.lcpc.fr/tarel.jean-philippe/publis/iccv09.html⟩ (2009).

[16] Woodford, O., Torr, P., Reid, I. and Fitzgibbon, A.: Global Stereo Reconstruction under Second-Order Smoothness Priors, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 31, No. 12, pp. 2115–2128 (2009).

Laurent Caraffa received a M.S. degree in Computer Vision and Image Processing from the University of Nice Sophia-Antipolis in 2010. He received his PhD degree in Computer Science Paris VI-P. and M. Curie University in 2013 on stereo 3D reconstruction taking into account bad weather conditions. From 2011, he is with the french institute of science and technology for transport, development and networks (IFSTTAR).

Jean-Philippe Tarel graduated from the Ecole Nationale des Ponts et Chaussées (ENPC), Paris, France (1991). He received his PhD degree in Applied Mathematics from Paris IX-Dauphine University in 1996. He was with the Institut National de Recherche en Informatique et Automatique (INRIA) from 1991 to 1996 and from 2001 to 2003. From 1997 to 1998, he worked as a research associate at Brown University, USA. From 1999, he is a researcher in the french institute of science and technology for transport, development and networks (IFSTTAR and formerly LCPC), Paris, France. His research interests include 3D reconstruction, pattern recognition and detection.