Image Based Visibility Estimation During Day and Night

Sami Varjo and Jari Hannuksela

The Center for Machine Vision Research, Department of Computer Science and Engineering, P.O. Box 4500, FI-90014 University of Oulu, Finland
sami.varjo@ee.oulu.fi, jari.hannuksela@ee.oulu.fi

Abstract. The meteorological visibility estimation is an important task, for example, in road traffic control and aviation safety, but its reliable automation is difficult. The conventional light scattering measurements are limited into a small space and the extrapolated values are often erroneous. The current meteorological visibility estimates relying on a single camera work only with data captured in day light. We propose a new method based on feature vectors that are projections of the scene images with lighting normalization. The proposed method was combined with the high dynamic range imaging to improve night time image quality. Visibility classification accuracy (F1) of 85.5% was achieved for data containing both day and night images. The results show that the approach can compete with commercial visibility measurement devices.

1 Introduction

Meteorological visibility is an important measure in many fields such as road traffic safety, flight control and aviation safety, as well as coastal and marine activities. Visibility is usually reduced by different sized particles or aerosols, such as fog, rain, or snow [1]. In addition to precipitation, visibility can be reduced by air pollution which levels can be monitored using visibility estimation[2]. Traditionally, visibility is estimated by a human observer who estimates the visibility range from known distant objects like forest line, stars or other light sources. Manual observations are not only costly and time consuming but also biased by the individuals.

The visibility degradation can be modeled using physical models that are based on light absorption and scattering in the air. One commonly applied model is based on the Beer-Lambert law that states that the light intensity drops exponentially as a function of the traveled distance and the medium absorption coefficient. The non-imaging based measurements with lasers are typically carried out in very small volumes and the results are extrapolated to cover distances up to 20 kilometers away from the actual measurement location. Local weather phenomenon at some distance cannot be captured by these methods.

Cameras enable monitoring affordably panoramic scenes at all directions. Tan [3] has applied random Markov fields on single images to improve the image

quality in low visibility images, but no metric visibility measurements were attempted. The Beer-Lambert law has been also applied with images, which allows learning of simple regression models for visibility estimation [4, 5].

The observations are usually required to be carried out continuously. So far the large illumination changes have not been addressed and the image based estimation methods are usually shown to work only at day light or during night but not both. Du et alia[6] have used two cameras at different distances to capture images of the same target during night and transmission based estimation was obtained. The transmission measurements are limited on a line of sight and multiple cameras are required. Also distant light sources and scattering models have been used to estimate the visibility during night[7]. That approach has limited usability during the day time as the sky luminance is often dominant over the man-made light sources. We suggest an approach that does not rely on external light sources and use a single camera to estimate the visibility 24 hours a day.

We show that commonly applied gradient and light attenuation based regression models fail in continuous monitoring. We propose a new method where Retinex filtering is used to create light intensity invariant images that are used for projection based features. The visibility estimation is based on machine learning with support vector machines. Further, HDR techniques have not been previously utilized in metric visibility estimation and we show that it is beneficial when combined with the proposed feature classification. The proposed method is targeted for scenes with a visible horizon, like at air ports or city skylines.

2 Related Work

Human observations are still widely used in visibility estimation as no extra equipment is required. The estimation is based on known distant objects whose contrast against the background is assessed and, if the visibility vary in different directions, usually only the smallest visibility is reported. Manual observations are relatively expensive and may be biased by the observer.

The visibility sensors do not directly measure how far one can see but instead the clarity of the air is measured. Locality of the non-imaging sensors is a major problem that would require multiple measuring sites, which usually is not cost efficient. Cameras can be utilized for estimating visibility, but large lighting changes limit the current methods to day time imaging. We discuss these issues in the following chapters in more detail.

2.1 Non-Imaging Based Estimation

Non-imaging based automatic measuring systems rely either on light scattering or transmission. Light scattering measurements are carried out in a small volume, for example, a cubic decimeter and the estimates are extrapolated to cover ranges up to 20 kilometers. Transmission meters measure the intensity drop in a light beam on a given path. While not as point-like as the scattering methods,

these are also considered local measurement devices. Both measuring methods are considered too expensive to be utilized in numbers for more comprehensive coverage.

The extrapolated visibility measurements do not always correspond to the true situation at some distance. For example, by visual inspection the Vaisala FD12P forward scattering device used in this study, produced estimates that were wrong in about 23% of the 6650 captured images. Figure 1 shows few examples of images and visibility estimates from the device where the measurements are clearly misleading. Too high visibility estimates announced by aviation traffic control are problematic at airports where reliable information is required 24 hours a day. This is one of the main motivations behind this work.

Fig. 1. Examples of cases where the light scattering equipment failed: (left,day) visibility measure 8180 meters and (right,night) 13 kilometers. The proper visibilities are less than 1000 meters and less than 5 km respectively. The clearly false measurements cause problems for example in aviation during day and especially at night.

2.2 Image Based Models

The existing visibility measures are typically based on contrast or gradients computed from the captured images. Multivariate linear regression, for example, can be used to learn visibility models from edges obtained using Sobel-filters [4]. The model coefficient of determination values (R^2) ranged between 0.780 and 0.845 for day time images.

Babari et al. also used gradients to estimate a physical model described by the Beer-Lambert law [5]. They reported the average errors of 30% (R^2 between 0.89 and 0.95) for images taken during day with luminance between 10 and 8000 cd/m^2. Their approach uses Sobel filtering to extract gradient magnitudes from an image I and the response is weighted by an estimate of the Lambertianess $W_{x,y}$ of the pixels in the scene for a contrast measure:

$$G = \sqrt{\left(\begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * I\right)^2 + \left(\begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I\right)^2} \cdot W_{x,y} \tag{1}$$

$$W_{x,y} = corr(I^{set}_{x,y}, I^{set}_{sky}),$$

where $[*]$ denotes convolution, and $[\cdot]$ point wise multiplication at pixels x, y. The $W_{x,y}$ is calculated as the correlation of pixel intensities respect to the sky illumination over time. The final measure $\hat{E}$ is formed by integrating the response G over the image. The obtained responses are used to form a regression model with parameters A and B that relates the measure to the Beer-Lambert law:

$$\hat{E} = A + B \log\left(V_{met}\right). \tag{2}$$

The inverse of the model can be used to convert the measured response to metric visibility V_{met}. Figure 2 shows a fit to data containing both day and night images from Matilda database[5].

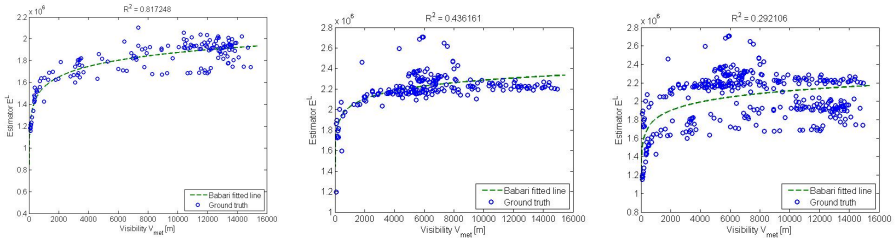

Fig. 2. Model (2) fitted on day and night data from the Matilda database and combined data. Only the day data fits the model with a reasonable coefficient of determination, and weighting with a surface Lambertianess estimate improves the fit as shown in [5]. However, data is separated so that no static weighting can remove all the lighting covariance from the data and the day and night data is clearly separated to sub groups.

The data clearly separates into two groups and a single model is clearly not enough. The changing sky luminance cannot be handled with the regression based models. The ambient luminance can change over a day radically at the northern and southern hemisphere making the reliable estimation of the Lambertian weights $W_{x,y}$ almost impossible in these scenarios, at least the estimation should be adaptive instead of static. Further, as the weighting is based on the illumination of the sky this cannot be done during night.

Other approach is to use machine learning to categorize the samples into visibility classes. Yin et al. [8] used support vector regression to learn visibility

from regions of interest where the features were formed from the mean local contrast of 4x4 sub blocks resulting in 16-dimensional features for each image. A mean success ratio calculated from the reported results show 81% performance for day time images. In this case poor visibility was most often misinterpreted, which is the most important class to be classified correctly.

3 Proposed Method

The main problem with the existing image based visibility estimation is the covariance of the extracted response with the ambient illumination. The estimation of the ambient illumination based on the sky is not reliable as clouds change and reflect light and at night there is only light from moon and ground reflections.

We suggest forming a new, lighting invariant, descriptor for scene visibility classification. The features are formed by projecting Retinex filtered images on the horizontal direction corresponding to different distances in the view as shown in Fig. 3. While we propose to use HDR imaging to capture the images, the approach is shown to work also with conventional low dynamic range images. The features are further processed for robustness by taking the absolute values of their gradients which are normalized. The visibility estimation is then finally based on support vector machine based classification. In the following, the steps are described in more detail.

3.1 HDR Imaging for Visibility Estimation

Luminance of the night sky is typically in range of millicandelas per square meter while street lamps are capable of producing some tens of kilocandelas per square meter. Therefore night scenes with urban surroundings have a dynamic range of six to seven decades, which is on the lower limit of an adapted human eye dynamic contrast ratio. Digital cameras commonly use analog to digital converters with 8 to 16 bit precision enabling at most contrast ratios of 1:65536, which is not enough for representing the low light scenes as can be seen from Fig. 4. Short exposure times enable sufficient capture of the street lights vicinity but unilluminated vegetation is poorly shown. Long exposure times capture even clouds and the stars from the night sky but the lit streets are severely over exposed.

High dynamic range (HDR) imaging in combination with tone mapping techniques is typically used to produce visually pleasing images to overcome the limited camera sensor and display dynamic ranges [9]. It is also possible to enhance an image quality by producing HDR images without using physical models [10, 11].

HDR images are commonly constructed by solving the camera response function (CRF) that describes the relation of recorded intensity and the exposure time to the scene radiance. Inverse CRF is then used to linearize the collected data which is combined for a radiance map [12]. Further the radiance map is used with tone mapping to represent the color information correctly for the human

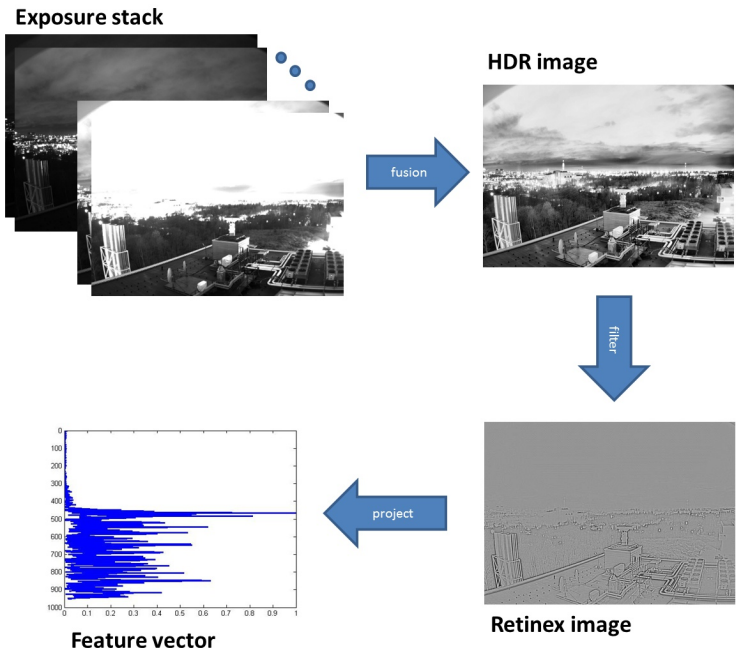

Fig. 3. An exposure stack of images is captured for a high dynamic range image, which is Retinex filtered and the rows of the image are projected on the y-axis for the feature vector.

Fig. 4. Six subsequent images from an exposure time stack captured during a night. The exposure times are 72.9ms, 218.7ms, 656.1ms, 1.968s, 5.905s, and 11.810s.

viewer using low dynamic range displays or print outs[13]. It has been shown that a dense set of images produces an improved signal to noise ratio for HDR radiance map compared with a minimal number of images [14]. The dense image sampling strategy is selected to suite the needs of this application.

There are no real-time requirements for the system and it is enough to have an updated estimate every ten minutes. To capture stacks of images, exposure times between 25 microseconds and 23.6 seconds were used yielding maximum total exposure time of 35.4 seconds. This enables the reconstruction of HDR images under a large range of lighting conditions. In this work, we are more interested in the normalization of the lighting changes and the visibility than the esthetics of the images.

An adaptive method was utilized to stack the HDR images. As the sensor behaves here linearly (see Section Data, Fig. 7, the fusion of the images I_{HDR} in an exposure stack is an integration of the intensity values I_i over the pixels weighted with the exposure times t_i:

$$I_{HDR} = \sum_{i}^{N} I_i \cdot t_i. \tag{3}$$

The number of used images N, was selected by estimating the contrast in result image I_{HDR} after each summing step and selecting the result image with the highest obtained contrast at the end. Fig. 5 shows an example of the combined images for a scene during a night and a day. The night image corresponds to the inputs shown also in the Fig. 4. The HDR image clearly contains more information than the single input images.

Fig. 5. HDR images obtained in a night and during a day. The images are tone mapped for viewing using contrast limited adaptive histogram equalization (CLAHE)[15].

3.2 Retinex Filtering

Multiscale retinex (MSR) filtering was originally developed to improve the visual quality of both 8-bit images and wide dynamic range images [10]. MSR aims for

creating images that match lightness and color perception of human vision. While the filtering is aimed for color images, it handles each color channel separately and thus can be used also on gray scale images. The response R_{MSR} for MSR is a weighted sum of single scale retinex filters R_i where the Gaussian filter F_i scale c_i is varied:

$$R_{MSR} = \sum_{n=1}^{N} w_n R_n, \tag{4}$$

$$R_i(x,y) = logI(x,y) - log\left[F_i(x,y) * I(x,y)\right], \tag{5}$$

$$F_i(x,y) = Ke^{-r^2/c_i^2}. \tag{6}$$

The parameters suggested in [10] were applied here; $N = 3$, $c_i = 15, 80, 250$ and $w_i = 1/3$. For the $r = sqrt(x^2 + y^2)$ 9x9 local window was applied. The applied filter does not produce here visually stunning results, but the differences between day and night time images are very small in visual inspection (see Fig. 6).

3.3 Feature Extraction

The retinex filtering can be characterized as high pass filtering in logarithmic space, which makes the result much less susceptible towards the lighting changes. The scene distance from the camera increases towards to upper part of the image. According to (2) the contrast or here the high frequencies are reduced with the increased distance in case of visibility degradation.

The projection of the frequency response in the image on y-axis can be considered a good descriptor for the visibility when the horizon is captured levelly in the images. In this work, no image rectification was used but the camera was only positioned carefully to capture horizon images. In practice, the feature can be formed by summing up the retinex filter response row wise over the image. The robustness is further increased by taking a gradient of the y-projection. Finally, the magnitude vector of the obtained gradient is normalized in range [0,1]. The features were orthogonalized using principal component analysis (PCA). The best results were obtained containing all the components of the PCA transformation indicating that all the feature vector components contain relevant visibility information that is useful in classification.

The Fig. 6 exhibits examples of retinex filtered images and the corresponding feature vectors. The difference between night and day images is visually compared small. The corresponding good visibility and poor visibility feature vectors appear very similar.

3.4 SVM Training

The feature vectors extracted from the images were used to learn the required visibility classes. Here support vector machines with RBF kernels were utilized

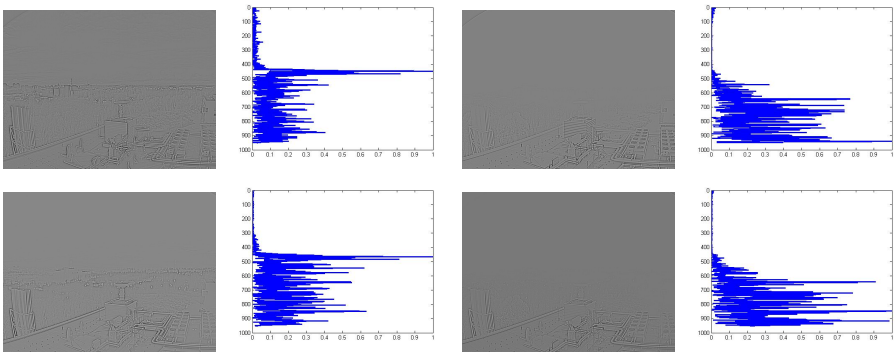

Fig. 6. Retinex filtered images in good (left) and poor visibility (right) during day (up) and night (bottom) and the corresponding feature vectors.

to learn the five classes named in Tab. 1. The applied learning strategy for the given class was one-against-all. In the multiclass classification the distance from the decision boundary of each class was used as the classification score and the largest distance was selected as a winner.

The learning parameters for soft margin C and the kernel Gaussian variance σ were optimized using a grid search and the performance was tested using leave one out cross validation for Matilda data and 5-fold cross validation for HDR data. F1-measure, that is the harmonic mean of precision and recall, was used as an accuracy measure.

4 Data

Images from two sources were used and all images top part of the sky was cropped out. With the Matilda database 31 pixel rows, containing time stamps, were cropped out leaving 100 pixel rows sky and 348 rows ground. With the HDRvisMe data 40 pixel rows of the sky were left leaving 511 pixel rows of ground. The Matilda database was utilized for the experiments with 8-bit images [5]. The data consist of CCTV camera images and meteorological optical range and sky luminance values measured with meteorological instruments. The similar division into visibility classes as in [5] was utilized, with the distinction that also images having visibility above 15 km were included. Tab. 1 contains the number of samples in each of the class. Images with luminance below 10 cd/m^2 were treated as night images. It should be noted that in [5] only day images and four first classes were utilized.

For HDR data set (HDRvisMe)[1] a 14 bit gray scale camera (AVT, Prosilica GT1290) was used for capturing the images. It was housed in a weatherproof casing with a heater element to prevent water condensation in the casing and on the case window. The setup was installed on a roof about 30 meters from

[1] The data and Matlab implementation are available upon a request from the authors.

Table 1. Visibility classes and number of samples.

Database		<400m	400-1000m	1-5km	5-15km	>15km
Matilda	day	17	6	26	109	49
	night	9	3	44	148	26
	total	26	9	70	257	75
HDRvisMe	day	48	58	208	725	1765
	night	15	68	393	955	1339
	dawn and dusk	10	20	104	317	625
	total	73	146	705	1997	3729

the ground. The camera was secured so that image registration due to camera movement was not required. For comparison also here the data was divided to day, night, and dusk&dawn sets based on the capture time. Data was manually labeled to the visibility classes in Tab. 1.

Vaisala FD12P light scattering measurement device was used to collect the reference visibility data. The manufacturer of FD12P gives the error tolerance of ±10% for the range 10-10000 m and ±20% for range 10-50 m and instrument consistency of 4%. It should be noted that the device has the sampling volume of 0.1 liters.

Lack of the color filters in the camera removes the need for separate camera response calibration for color channels. The color filters would also reduce slightly the incident light hitting the sensor. The images were captured using manual settings so that exact parameters, such as gain and exposure time, were used. Also all the in-build data manipulations, including gamma correction, were disabled. This enabled the near linear behavior of the camera sensor, which was ensured by inspecting the camera response curve. This was constructed by selecting uniform intensity patches from sets of images with different exposure times, which were further reduced to median values for each of the exposure time settings. These values were converted to relative intensity values using exposure times as presented in Fig. 7. In addition the method from Mitsunaga and Nayar [16] was used to fit polynomials with different degrees. The best fit was a straight line, proving the linear response of the imaging setup.

5 Results and Discussion

The proposed method reached 86.5% F1-accuracy for the combined day and night images in the Matilda database with leave one out cross validation (soft margin $C = 3.5$ and $\sigma_{RBF} = 1.05$) . With separation to day and night, the accuracies were 85.2% and 82.6% respectively. One would expect that the classification would work better with separated data sets, but apparently the data contains a limited number of samples, especially the second class for visibility range 400-1000 meters. This can also be seen in the confusion matrix in Fig. 8 as large variation with the second class.

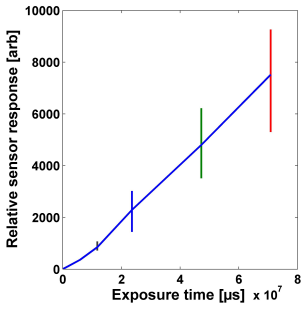

Fig. 7. The camera response function for the utilized camera.

A more comprehensive data set was captured for HDR testing with over 10 times more images available than in Matilda data set. Here for the combined day and night data, the 5-fold cross validation yielded 85.5% F1-accuracy. With the HDR data, the confusion matrix is less scattered around the true classes and the result is more reliable than with the low dynamic range image set. Data separated to day and night sets resulted in 86.2% and 85.0% classification accuracy respectively. The utilized training parameters here for SVMs were for soft margin $C = 5.3593$ and for the RBF-kernel shape $\sigma_{RBF} = 1.5625$ with all classes.

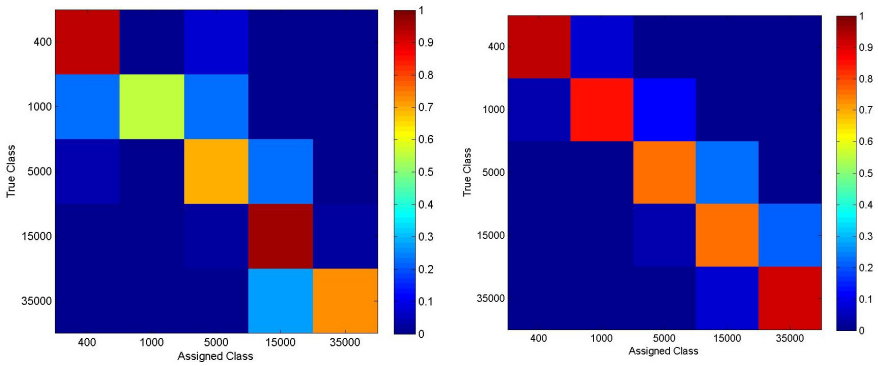

Fig. 8. Confusion matrices for all Matilda data (left) and for the HDR data (right).

Classifiers were tested also by using only the day data for training and night data for testing and vice versa. Here the mean accuracy was 71%, showing that some lighting invariance has been obtained for the descriptor. The slight drop indicates that the invariance is not perfect, but with samples of all possible lighting conditions the changing illumination can be handled with the descriptor and SVMs.

It should be noted that the HDR data was manually annotated to the visibility categories as the visibility sensor was realized to produce correct estimates for about 77% of input images. However, it is difficult also for human to estimate visibilities around 10 km in the wild and here annotation was done based on images only. Therefor, some errors may exist in annotations. The confusion matrix with HDR data supports this assumption as the first two classes are classified more robustly than the problematic classes 1-5 km and 5-15 km, while the clear weather was easier to classify for both the human and the proposed system.

6 Conclusions

We show the first time that also night time images can be used for estimating the visibility using a single camera. High dynamic range imaging can be used to capture images under varying lighting conditions and combined with Retinex filtering the illumination changes become almost negligible. The projection of filtered images on the horizontal direction captures the visibility degradation as function of distance which can be used for visibility classification.

The proposed features were used to train support vector machines for five visibility classes using data from two different sources. The image sets resulted in 85% accuracy measured using F1-score. The confusion matrices show the HDR data set to be more reliable than the CCTV-based image set. While the sample size differences in the data sets leave room for speculation, HDR imaging can be considered a good way to capture images in varying lighting conditions.

It was shown that the method reached 85% classification accuracy, which can be safely stated to be at least on the same level or better than the reference commercial light scattering instrument, that produced visually correct results in 77% of the cases. The confusion matrices show that the proposed classification approach tends to misclassify some of the samples to the neighboring classes, which can be the indication of a data annotation problem, while the commercial instrument failed some times grossly. Also the classifiers were trained using the same training parameters for all the classes and it can be expected that some improvement can be obtained using more advanced machine learning approach.

The main criticism for the method, and in fact for most of the image based approaches, can be considered to be the need for training data. It can be very time consuming to gather the required training data that contains all the possible weather conditions. For future work, one might consider if computational methods can be used to generate simulated training data, so that only a few good visibility images in varying lighting conditions would be enough for the training.

Acknowledgment

The Finnish meteorological institute is acknowledged for both the image data and the reference data used in the HDR experiments. Nicolas Hautière and Éric

Dumont are thanked for the Matilda database. Eliska Nyrönen is acknowledged for the implementation of the regression model based method described in [5].

References

1. Nayar, S., Narasimhan, S.: Vision in bad weather. In: The Proceedings of the Seventh IEEE International Conference on Computer Vision. Volume 2. (1999) 820–827
2. Hyslop, N.P.: Impaired visibility: the air pollution people see. Atmospheric Environment (2009) 182–195
3. Tan, R.: Visibility in bad weather from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition. (2008) 1–8
4. Graves, N., Newsam, S.: Visibility cameras: Where and how to look. In: Proceedings of the 1st ACM International Workshop on Multimedia Analysis for Ecological Data. (2012) 7–12
5. Babari, R., Hautière, N., Dumont, È., Paparoditis, N., Misener, J.: Visibility monitoring using conventional roadside cameras emerging applications. Transportation Research Part C: Emerging Technologies **22** (2012) 17 – 28
6. Du, K., Wang, K., Shi, P., Wang, Y.: Quantification of atmospheric visibility with dual digital cameras during daytime and nighttime. Atmospheric Measurement Techniques (2013) 2121–2130
7. Narasimhan, S., Nayar, S.: Shedding light on the weather. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. (2003) I–665–I–672 vol.1
8. Yin, X.C., He, T.T., Hao, H.W., Xu, X., Cao, X.Z., Li, Q.: Learning based visibility measuring with images. In Lu, B.L., Zhang, L., Kwok, J., eds.: Neural Information Processing. Volume 7064 of Lecture Notes in Computer Science., Springer Berlin Heidelberg (2011) 711–718
9. Cadík, M., Wimmer, M., Neumann, L., Artusi, A.: Evaluation of HDR tone mapping methods using essential perceptual attributes. Computers & Graphics **32** (2008) 330 – 349
10. Daniel J. Jobson, Z.u.R., Woodell, G.A.: A multiscale retinex for building the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing **6** (1997) 965–976
11. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion: A simple and practical alternative to high dynamic range photography. Computer Graphics Forum **28** (2009) 161–171
12. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: Proc. of the 24th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '97 (1997) 369–378
13. Kirk, A.G., O'Brien, J.F.: Perceptually based tone mapping for low-light conditions. In: ACM Transactions on Graphics. Volume 30(4) of SIGGRAPH '11. (2011) 42:1–10
14. Barakat, N., Darcie, T., Hone, A.: The tradeoff between snr and exposure-set size in hdr imaging. In: 15th IEEE International Conference on Image Processing 2008. (2008) 1848–1851
15. Zuiderveld, K.: Contrast limited adaptive histograph equalization. In: Graphics Gems IV, Academic Press (1994) 474–485
16. Mitsunaga, T., Nayar, S.: Radiometric self calibration. In: IEEE Conference on Computer Vision and Pattern Recognition. Volume 1. (1999) 374–380