High Dynamic Range Imaging for Automotive Stereo Vision Applied to Tunnel Scenarios

Alexander Gillert, Dominik Urbaniak, Gang Chen, Kai Huang

***Abstract*—In this paper we apply high dynamic range (HDR) imaging techniques in automotive stereo vision to solve the problem of difficult light conditions in tunnels, especially at entrances and exits. We present experimental results for two automatic multi-exposure control methods. Moreover, a modification of an image fusion algorithm that improves the performance of the stereo matching is proposed. We conclude that HDR imaging can be a vital tool for real world situations with difficult light conditions. The ghosting effect due to multi-exposure fusion is present but not very significant for the computation of a disparity map.**

I. Introduction

Stereo vision, being a vital part in autonomous automotive applications and advanced driver assistance systems (ADAS), requires high quality image data to deliver acceptable results. Real world scenarios like tunnel entrances, tunnel exits, camera facing low sun or reflections of the sun can produce over- or underexposed regions in the image data if acquired without care.

Figure 1 shows screenshots from a video published by Hernandez-Juarez et al. [4] and demonstrates the problem statement. In the first frame, the scene is evenly illuminated by the sun and the disparity map thus does not show significant artefacts. In the second frame, the camera is blinded by the sun and its reflections on the road. The disparity map algorithm cannot deal with this situation, creating artefacts in the overexposed regions. The third frame depicts a tunnel exit. Since the scene outside the tunnel is vastly brighter than the interior, it is completely overexposed in the image data, which erases the majority of the information. A useful disparity map cannot be produced. The exterior of the tunnel could be exposed correctly by reducing the exposure time, however this in turn would underexpose the remaining parts of the image, e.g. the road directly in front of the vehicle.

These problems are due to a rather low dynamic range (LDR) of the image sensor in the camera. In image acquisition, dynamic range refers to the ratio between the brightest and the darkest light intensities of an image [8]. Modern commercially available camera sensors have a dynamic range of around 65db [10,11]. However, in the real world scenarios mentioned above, the dynamic range can exceed 80db. During night, with headlights and reflecting road signs this may even exceed 100db [3].

To deal with this, several techniques to increase the dynamic range can be applied. In this paper we first give an overview over current HDR techniques and then present our experimental data with two of those techniques for use in automotive stereo vision.

II. Previous Work

Akhavan, Yoo & Gelautz [2] have explored possibilities for combining stereo matching algorithms with HDR image techniques so that the 3D-scene reconstruction benefits from the increased level of detail. Their work does not focus on a specific application area and only considers generated datasets, therefore does not deal with the challenges of real-world outdoor environments.

Görmer et al. [3] analyzed the problem of low dynamic range with automotive applications in mind, however focusing more on single camera tracking tasks, whereas our area of interest lies on dual camera stereo vision. Moreover, they used simultaneous multi-exposure image sensors, which are currently difficult to obtain.

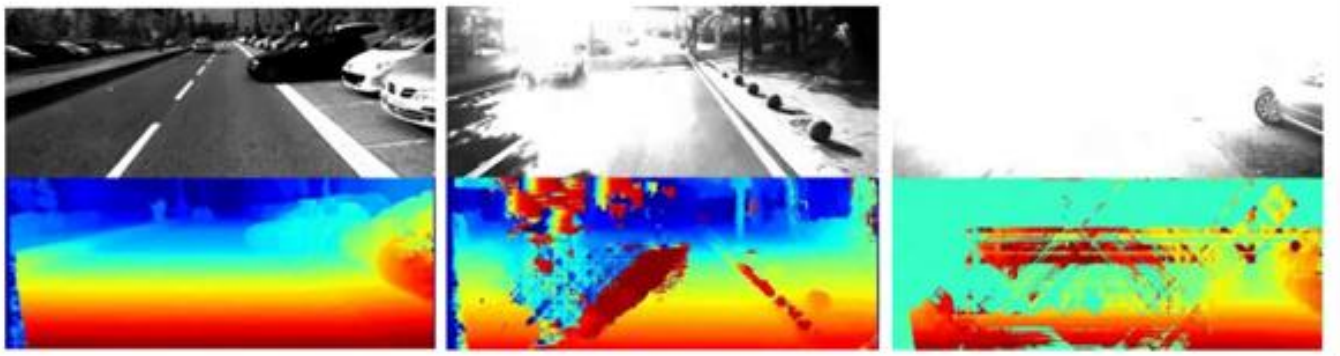

Figure 1: Three scenarios that are common in automotive vision applications and the corresponding disparity map below: evenly exposed scene (left), camera blinded by the sun and reflections (center) and a tunnel exit (right). Data published by [4].

Alexander Gillert and Kai Huang are with the School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China
(e-mail: alex.gillert@gmail.com, huangk36@mail.sysu.edu.cn).

Dominik Urbaniak is with the Munich School of Engineering, Technical University Munich, Germany. (e-mail: d20.urbaniak@gmail.com).
Gang Chen is with the TODO University
(e-mail: TODO@TODO)

Paul and Chung [14] applied HDR imaging to improve the performance of lane detection algorithms when driving in direct sunlight at dusk or at dawn. They compared several HDR multi-exposure fusion algorithms and concluded that the Mertens algorithm performs best.

III. HDR-Imaging Techniques

A. Sequential Multi-Exposure Imaging

A well-known technique in the field of photography to produce HDR images, is to shoot multiple images with different exposure times one after another to cover the whole dynamic range of the scene and subsequently combine them to a single picture preserving the details [8]. The result is an image that can cover the dynamic ranges of the source images.

Drawbacks, especially for real-time and automotive applications, include a reduced frame rate, since several images are required to create the final image and artefacts from non-static scenes, often referred to as the ghosting effect. Ghosting occurs when an object moves fast in-between the acquired frames, thus recorded by the camera as being in different positions. After combining the frames, this object appears twice in the HDR image, half-transparent resembling a ghost, hence the name.

Nevertheless, it is a viable option, since it is well researched and simple to implement. Therefore, we analyse it in this paper whether it is practical for automotive stereo vision

B. Simultaneous Multi-Exposure Imaging

Aggarwal & Ahuja [1] have proposed a custom hardware setup to create HDR images. The light beam between the camera lens and the image sensor is split into multiple parts and redirected onto different image sensors using an assembly of mirrors. The sensors are then programmed to acquire data with different exposure times and the resulting LDR images are combined in the same manner as in the technique described above to create an image with higher dynamic range.

Alternatively, sensors which can acquire two different exposure times at the same time have been developed, e.g. by [9], however those cannot be easily purchased as off-the-shelf commercial products at the moment.

One obvious advantage of this method is that the images are acquired simultaneously, which reduces the undesired ghosting effect to a minimum and does not decrease the frame rate. On the other hand, it requires additional non-standard hardware components which are not readily available on the market and have to be custom made, thus increasing the cost, size and complexity of the system. Due to this, this approach was not considered in this paper.

C. Non-linear Imaging

There are several variations of this method but in the end they all lead the same result, that the image sensor pixels have a response curve that resembles a logarithmic rather than a linear function. This means that, the more light a photodiode has already received, the more light is needed to increase its quantized pixel value. Many commercial image sensors (e.g. [11]) offer the feature of multi-slope or multi-reset integration, i.e. they allow to set one or more knee-points at which the photodiodes reduce their gain setting, yielding a piecewise linear response curve.

Although originally not designed to increase the dynamic range, similar results can be achieved with anti-blooming. In this case, special circuitry which allows a pixel to bleed off excess charge before it spills over to neighboring pixels and the image sensor response curve becomes logarithmic, similar to Figure 2 [12].

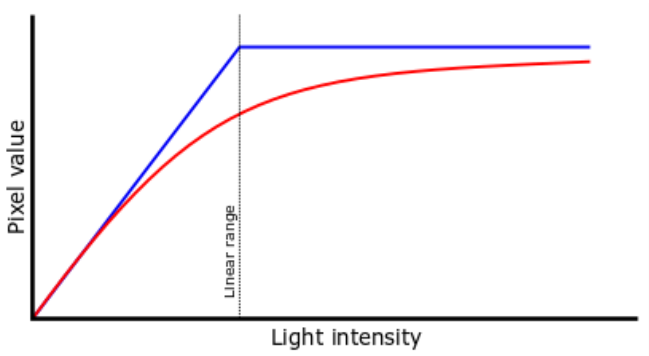

Figure 2: Schematic comparison of linear (blue) and logarithmic (red) image sensor response curves

Contrary to the multi-exposure techniques only one image has to be acquired and no further post-processing is needed to yield a HDR image. On the other hand, in general the bright regions of an image suffer from reduced contrast.

IV. Methods

A. HDR Image Fusion

There exist several well established algorithms for the creation of HDR images from multiple differently exposed images. However, all of those algorithms were designed for the field of photography to provide aesthetically pleasing results from a set of already quite well exposed images, rather than with computer vision in mind. Therefore, to adjust to the difficult light conditions of our application, we use a modified version of the algorithm that was published by Mertens et al. [6]. Inspired by the original paper we compute preliminary weights for each pixel *(i,j)* in image *k* as follows:

$$W_{ij,k} = \min \{ \omega_C C_{ij,k} + \omega_E E_{ij,k}, \quad 1 \}$$

where *C* is a contrast measure defined as the absolute value of the Laplacian filter response, applied to the image and *E* is a measure of how well exposed a pixel is, which we redefine as:

$$E_{ij,k} = \begin{cases} 1, & if\ I_{ij} > 0.5\ and\ k = 0 \\ 1, & if\ I_{ij} < 0.5\ and\ k = 1 \\ \exp\left(-\frac{(I_{ij} - 0.5)^2}{2\sigma^2}\right), & otherwise \end{cases}$$

where I_{ij} is the intensity at pixel *(i,j)*, *k=0* indicates the first, darker image, *k=1* indicates the second, brighter image and σ a constant which we set to a rather low value of *0.05* to minimize the influence of over- and underexposed pixels on

the final result. The case differentiation is possible because we are only using two images. We thus assume that a bright pixel in the dark image cannot be even better exposed in the brighter image. We combine the two measures by linear combination instead of a product, because for us it sufficient to have only one of them. The saturation measure *S* from the original paper is not used because we are only dealing with grayscale images.

After that, the weights have to be normalized to sum up to one. Partially due to the low value of σ, the normalization in the original paper is numerically unstable in regions which are both overexposed in the brighter image and underexposed in the darker image. Therefore, we opted for a more stable version, though changing proportions slightly:

$$\widehat{W}_{ij,k} = W_{ij,k} + \left(1 - \sum_{k'} W_{ij,k'}\right)/N$$

with *N* being the number of images, in our case *N=2*. To avoid sharp edges at weight transitions we apply a Gaussian blur filter to the weight maps. The differently exposed images are finally weighted and summed together to create the HDR result *R*:

$$R = \sum_{k'} \widehat{W}_{ij,k'} \times I_{ij,k'}$$

Overall, these modifications result in a higher contrast in extreme regions than the original algorithm. An example of the results is shown in Figure 3. Note that in this example the image area directly around the tunnel exit is either over- or underexposed in both of the two original images. The dynamic range of this scene exceeds even dual-exposure HDR. A third frame with an exposure time in-between the two others would be required to recover structure in this region.

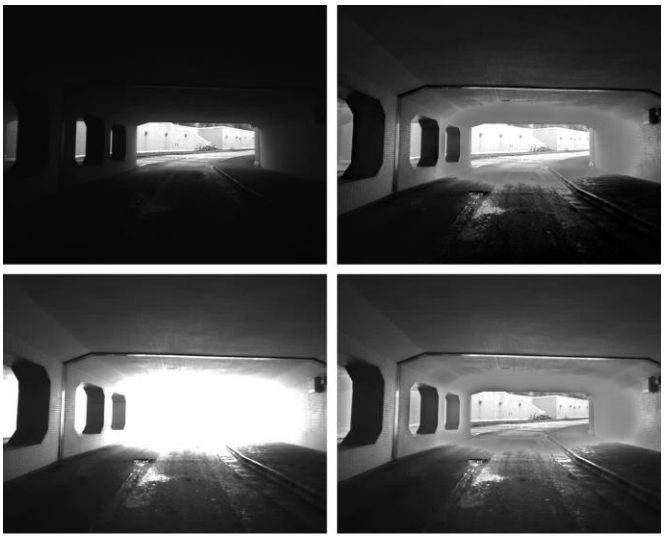

Figure 3: An example of HDR image fusion. Top left and bottom left: two images showing the same scene taken with short and long exposure times respectively. Top right: the two images on the left merged with the Mertens algorithm as implemented by the OpenCV library [16]. Bottom right: The same images merged with our modified algorithm.

B. Automatic Multi-Exposure Control

Most exposure control algorithms for single-exposure imaging try to keep the mean value of an image around a predefined target, usually the mid-tone *0.5*. For example, the calculation of the exposure value *EV* for the next image frame *n+1* can be expressed in an iterative way similar to [13]:

$$EV_{n+1} = EV_n + s(\log_2 B_n - \log_2 B_{target})$$

with *B* being the average brightness of the image and *s* a scaling constant which influences the adjustment speed. An additional small margin around the target value in which no changes are applied prevents oscillation of the exposure time. This leads to acceptable results as long as the dynamic range of the scene does not exceed the capabilities of the image sensor.

We employ a similar strategy for multi-exposure imaging, by setting the target values for the different exposures at different levels, for example at $\frac{1}{3}$ and $\frac{2}{3}$ of the pixel value range respectively. The two exposures are then controlled independently of each other. This means that the exposure time of the first frame in a sequence is not influenced by the second frame (and the other way around). We call this method *Average-Based* auto-exposure.

In an alternative approach we try to maintain the number of dark pixels for the first image, and respectively the number of bright pixels for the second image in a small target range, thus guaranteeing that the darkest as well as the brightest areas of a scene are well exposed in at least one of the two images. The calculation of the exposure for the next dark frame is very similar to the previous equation:

$$EV_{n+1} = EV_n + s(\log_2 N_n^t - \log_2 N_{target}^t)$$

where N^t stands for the number of pixels below threshold *t*. The exposure for the next bright frame is computed analogous by counting the number of pixels above a threshold.

We will refer to this method as *Histogram-Based* auto-exposure. This method has the advantage that it can recognize when HDR is required at all. If the amount of dark and bright pixels in a single image does not exceed the threshold, the dynamic range of the scene does not exceed single-exposure imaging and HDR can be turned off.

In both methods only a part of the image is used to calculate the new exposure value. We chose to consider only the region below the horizon line, to avoid that the auto exposure control gets influenced by the bright sky and the sun, which are of little interest for automotive vision, but instead only by the road surface and the objects on it.

As suggested by [3], we treat gain and exposure as one parameter consisting of two factors, since both have almost the same influence on the pixel value. Since gain reduces the signal-to-noise ratio, gain is only increased when the exposure reaches the maximum acceptable value and decreased until the minimum before exposure is decreased. Therefore, any mentioning of the term exposure above actually refers to the real exposure time multiplied by the gain.

V. RESULTS

A. Experimental Setup

We conducted experiments to test the viability of multi-exposure and non-linear imaging HDR for their use in stereo vision. The experiments consisted of two static scenarios, in front of a tunnel exit and a tunnel entrance and a dynamic scenario i.e. driving through a tunnel. A short 80m tunnel was used for this. The experiments were conducted during sunny weather with measured light differences of almost 80db inside the tunnel compared to the outside.

Two stereo cameras were employed. The first camera is based on a custom PCB which employs the two commercially available E2V EV76C570 CMOS 2MP image sensors [10] 270mm apart. According to the datasheet, the sensors have a dynamic range of 66db. KOWA LM8HC lenses with 8mm focal length were used and their aperture set to the minimum of F2.8 to let in as much light as possible. The second stereo camera is built from industrial grade JAI GO5000M cameras, placed 900mm apart.

To keep the ghosting at a minimum and to not reduce the frame rate too much, only two different exposure times were taken to be merged into one. An additional measure against ghosting is to ensure that the image with the shorter exposure time is taken first. This way, the second (longer exposed) image frame can start earlier during the data read-out of the first image. Both, left and right cameras used the same exposure time.

The EV76C570 sensor also supports control of the photodiode anti-blooming. For the anti-blooming experiments, the corresponding parameter was set to the highest possible value in the image sensor and turned off in the other cases.

A Velodyne VLP16 LIDAR, mounted in the center between the left and the right cameras, was used to provide ground truth data for the evaluation of the disparity maps. The error values shown in the rest of the paper are measured as the average absolute difference in pixel disparity between the disparity map computed by the stereo matching algorithm and the LIDAR data projected onto the disparity map.

The disparity maps were created with a modified version of the semiglobal matching (SGM) algorithm by Hirschmüller [5]. The modifications include using the census transform for the computation of the matching cost volume. For our case, the most important advantage of the census transform is illumination invariance because only relative intensities are compared [15]. A second important modification is the cross-based cost aggregation as described in [7]. This technique averages the costs of neighboring pixels with similar intensities, which we found to be very beneficial in high noise environments such as poorly illuminated tunnels. The parameters of the algorithm were kept the same across all test cases to ensure that the results are comparable.

However, note that our approach does not depend on this specific algorithm. Our main results should be reproducible with other stereo matching algorithms as well. Figure 4 provides an overview of our method.

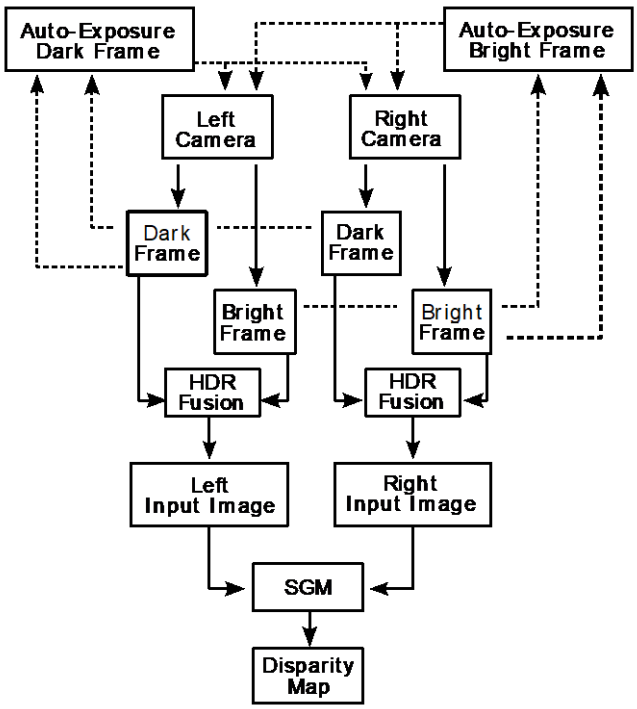

Figure 4: Overview of our dual-exposure HDR method

B. Discussion

Table 1 shows the error and standard deviations averaged over 10 frames recorded in front of the tunnel exit. As expected, all HDR methods perform much better than the single auto-exposure case. Among the HDR methods, multi-exposure proves to be more beneficial than nonlinear imaging and our modified image fusion algorithm more than the original version as implemented in the OpenCV library [16]. Both can be explained by higher contrast. We used the census transform as a cost function, for which at first glance contrast might be unimportant because only single bits are compared. However, the census transform is susceptible to noise and thus benefits from the higher contrast-to-noise ratio. The same should hold true for other cost functions like the often used sum of absolute differences (SAD).

Table 1: Average errors for the tunnel exit scenario in pixels

	Average Error	Standard Deviation
Single Exposure	7.69	1.981
OpenCV Mertens	2.06	0.130
Modified Mertens	1.92	0.082
Anti-Blooming	2.96	0.242

Similar results can be seen in table 2. In this scenario, images were acquired in front of the tunnel entrance. The same relationships between the methods as above are preserved. In the single exposure case a high deviation from frame to frame can be observed. The overall lower error than in the tunnel exit scenario can be explained by a comparatively lower dynamic range of the scene. Since light is flowing into the tunnel, the light intensity is only gradually decreasing and the underexposed image area is smaller, whereas in front of the

tunnel exit, light is hitting the camera all at once and overexposes an even larger image area.

Table 2: Average errors for the tunnel entrance scenario in pixels

	Average Error	Standard Deviation
Single Exposure	3.54	1.170
OpenCV Mertens	1.79	0.065
Modified Mertens	1.64	0.049
Anti-Blooming	2.11	0.129

Figure 5 provides a comparison of the auto-exposure strategies while driving through the tunnel. When using only single auto-exposure a high error peak in front of the tunnel can be observed. A useful disparity map cannot be computed until around 8 meters before the exit. That's when the auto-exposure control switches from a long to a short exposure time. With Average-Based dual-exposure we can increase this distance to around 18.5 meters. The Histogram-Based auto-exposure strategy avoids the large error peak at the end completely, however is rather noisy inside the tunnel due to an oscillation of the exposure time. The overall higher error inside the tunnel than outside in all strategies is due to low light intensity and higher signal-to-noise ratio.

Figures 6 and 7 show example disparity maps with and without HDR imaging. When facing the tunnel exit, the disparity map that was computed from a single exposure LDR image shows a wall, no information about the scene outside the tunnel can be inferred. Both HDR disparity maps show more information, albeit with artefacts in the anti-blooming case because the source image lacks contrast in the bright regions. Similarly, in the single exposed LDR image taken in front of a tunnel entrance, the inside of the tunnel is almost completely dark and the corresponding disparity map shows

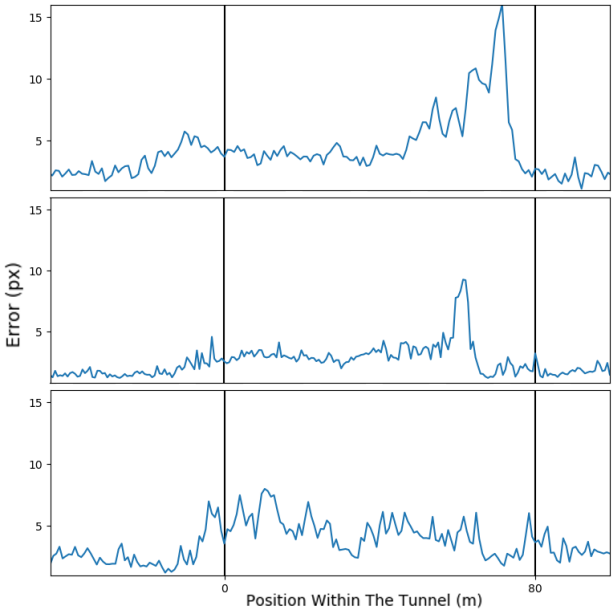

Figure 5: Error values measured during driving through an 80m long tunnel. The black vertical lines indicate the tunnel entrance (left) and exit (right). Top: Single Exposure LDR, Center: Average-Based Dual-Exposure HDR, Bottom: Histogram-Based Dual-Exposure HDR

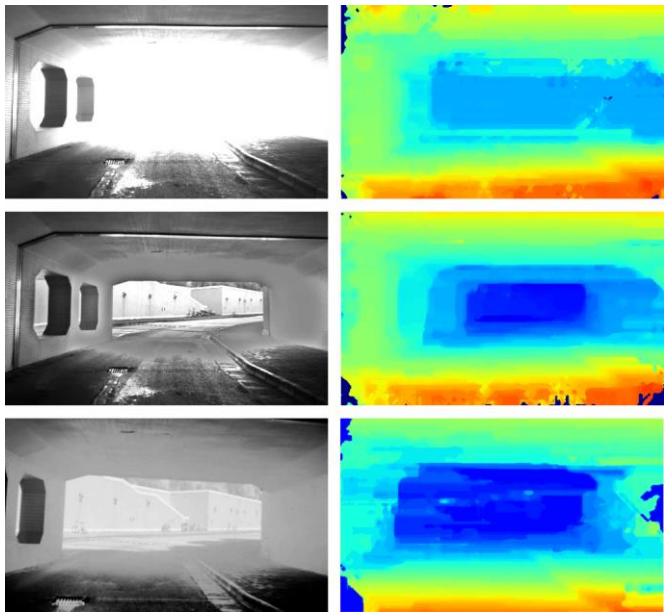

Figure 6: Comparison of LDR (top), HDR through dual-exposure fusion (center) and non-linear imaging HDR (bottom) with the corresponding disparity maps on the right when facing a tunnel exit

artefacts that vary from frame to frame. Those artefacts are not present in the HDR disparity maps.

C. *Possible Issues*

To examine the influence of the ghosting effect on the disparity map algorithm, scenes from a car, driving at a speed of 50km/h on a street were analyzed. As figure 4 shows, very little ghosting is present for cars moving in the same direction, since the speed difference relative to the viewer is comparatively small. The largest effect is visible on the road sign in the top left of the image: after merging it is displayed twice, slightly displaced. This might be a problem for sign detection and tracking algorithms, for disparity map

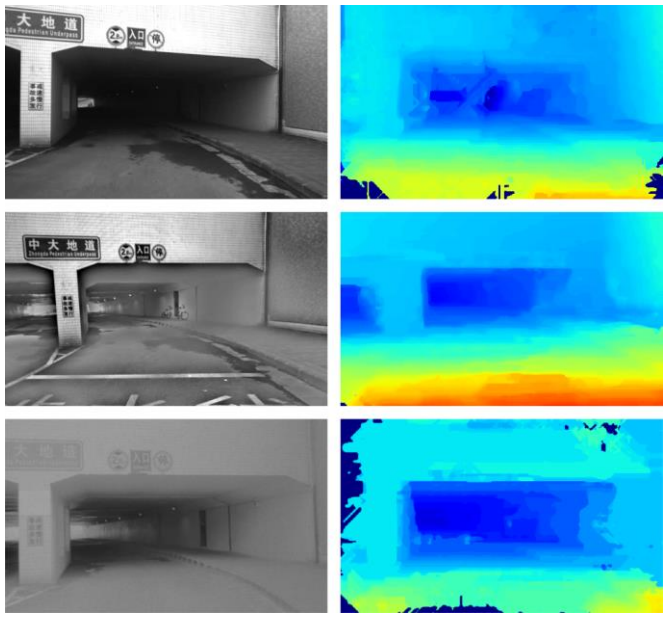

Figure 7: Comparison of LDR (top), HDR through dual-exposure fusion (center) and non-linear imaging HDR (bottom) with the corresponding disparity maps on the right when facing a tunnel entrance

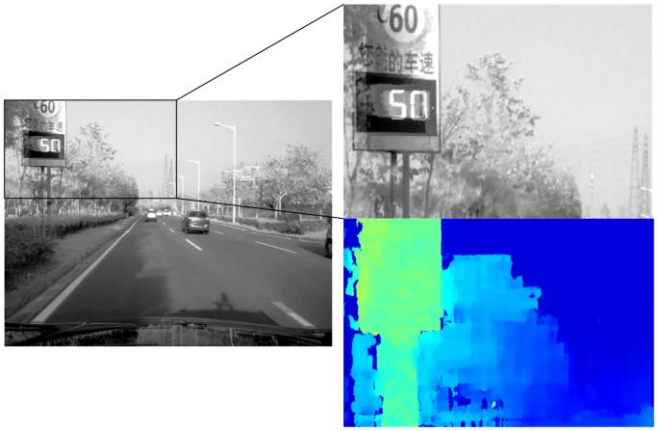

Figure 8: An example of the ghosting effect. Left: a multi-exposed HDR-image taken at a speed of 50km/h. Top right: an enlarged detail of the left image showing the ghosting effect. Bottom right: detail of the disparity map computed from the left image

computation however, this does not matter significantly. The result is simply, that the object on the disparity map becomes slightly wider than in reality, in the case of the road sign by 10 pixels or approximately 19%. Driving with speeds much larger than the examined 50km/h will certainly increase this effect.

Another issue is decreased distance measurement accuracy during driving. Since the two source pictures were taken sequentially and thus in two different locations, the merged HDR image is not (as one might naively think) in the center of those two locations, but in a meta-state of being in both locations at the same time. Thus, trying to measure to an object through disparity will have an uncertainty that is at least as large the distance between those two locations. Therefore, stereo vision with multi-exposure HDR might be best suited for obstacle detection.

VI. Conclusion

We presented experimental results for using high dynamic range imaging with stereo vision in the difficult low-light and high-noise conditions that are often present in tunnels. The largest benefit is present in scenarios when entering and exiting a tunnel, to be able to recognize the details on the inside or the outside, respectively. Overall, HDR through multi-exposure fusion performs slightly better than non-linear response curve HDR thanks to higher contrast, which is also the case for our modification of the Mertens HDR algorithm.

We tested two automatic dual-exposure strategies, both of which perform better than single exposure. To further improve the performance, it might be possible to combine both.

Although the ghosting effect is obviously present in multi-exposure fusion, for the disparity map computation it is not very significant when driving within city speed limits. Moreover, it could be reduced by increasing the frame rate. A larger problem might be the decreased distance measurement accuracy, which would disqualify this technique for tasks like map building. Therefore, we recommend to keep HDR off in normal light conditions and only activate it when needed.

Acknowledgment

We would like to thank Di Zhu, WenQuan Zhang, Fan Li and JunJie Yang for their driving assistance and patience during the driving trials.

References

[1] Aggarwal, M. and Ahuja, N., “Split Aperture Imaging for High Dynamic Range”, In: Int. Journal of Computer Vision, vol. 58, no 1, pp. 7-17, Hingham, MA, USA, 2004

[2] Akhavan T., Yoo H., Gelautz M.: “A framework for HDR stereo matching using multi-exposed images”, In HDRi2013 - First International Conference and SME Workshop on HDR imaging (2013), pp. 1–4. 1, 2

[3] Görmer, S & Hold, Stephanie & Kummert, Anton & Iurgel, Uri & Meuter, Mirko. “Multi-exposure image acquisition for automotive high dynamic range imaging”, IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. 1881 - 1886. 10.1109/ITSC.2010.5625024, (2010)

[4] D. Hernandez-Juarez, A. Chacón, A. Espinosa, D. Vázquez, J.C. Moure, A.M. López,, “Embedded Real-time Stereo Estimation via Semi-global Matching on the GPU”, Procedia Computer Science, Volume 80, 2016, Pages 143-153
(The mentioned video footage was retrieved on 2018/02/11 from https://youtube.com/watch?v=Olo0lPRyU5s)

[5] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30(2):328–341, 2008.

[6] T. Mertens, J. Kautz and F. V. Reeth, "Exposure Fusion," 15th Pacific Conference on Computer Graphics and Applications (PG'07), Maui, HI, 2007, pp. 382-390

[7] Zbontar, J., & LeCun, Y. (2015). Computing the stereo matching cost with a convolutional neural network. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1592-1599

[8] Bloch, C., “The HDRI Handbook: High Dynamic Range Imaging for Photographers and CG Artists”, 1st ed.: Rocky Nook Inc., Santa Barbara, CA, USA, 2007.

[9] Manabe, S. et. al., “High Dynamic Range Sensor with Blooming Drain”, U.S. Patent Pub. no. 2009/0002528A1, January 2009.

[10] EV76C570 2 Mpixels B&W and Color CMOS Image sensor, e2v semiconductors SAS, 29/03/2013

[11] CMV8000 8 Megapixel machine vision CMOS image sensor, CMOSIS, version1.2, 28/08/2015

[12] Theuwissen, A.J., “Solid-State Imaging with Charge-Coupled Devices”, 1st ed., Springer Netherlands, 1995

[13] JiaYi Liang, YaJie Qin and ZhiLiang Hong, "An Auto-exposure algorithm for detecting high contrast lighting conditions," 2007 7th International Conference on ASIC, Guilin, 2007, pp. 725-728.

[14] Nicholas Paul, ChanJin Chung, “Application of HDR algorithms to solve direct sunlight problems when autonomous vehicles using machine vision systems are driving into sun”, Computers in Industry,Volume 98,2018,Pages 192-196

[15] H. Hirschmüller and D. Scharstein, "Evaluation of Stereo Matching Costs on Images with Radiometric Differences," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 9, pp. 1582-1599, Sept. 2009.

[16] G. Bradski, The OpenCV Library, Dr. Dobb’s Journal of Software Tools (2000)