

Enhancement of Low Dynamic Range Videos using High Dynamic Range Backgrounds

Francesco Banterle, Matteo Dellepiane, and Roberto Scopigno

Visual Computing Laboratory, ISTI-CNR, Pisa, Italy

Abstract

In this paper, we present a practical system for enhancing the quality of Low Dynamic Range (LDR) videos using High Dynamic Range (HDR) background images. Our technique relies on the assumption that the HDR information is static in the video footage. This assumption can be valid in many scenarios where moving subjects are the main focus of the footage and do not have to interact with moving light sources or highly reflective objects. Another valid scenario is teleconferencing via webcams, where the background is typically over-exposed, not allowing the users to perceive correctly the environment where the communication is happening.

Categories and Subject Descriptors (according to ACM CCS): I.4.1 [Image Processing and Computer Vision]: Enhancement—Filtering I.3.3 [Computer Graphics]: Picture/Image Generation—Bitmap and framebuffer operations

1. Introduction

HDR images capturing is now a well known and solved problem in many conditions, see Reinhard et al. [RWP*10] for an overview. However, HDR video capturing still remains a challenge.

HDR video cameras are currently under release, but their cost make them affordable only for specific high-standard applications, like cinematography. On the other side, HDR images can be obtained with an SLR camera following a very simple procedure. In addition, some SLR and compact cameras are now featuring HDR capturing for still images using bracketing techniques followed by tone mapping.

Our work addresses the problem of enhancing LDR videos using the HDR information in the scene (i.e. the background). This can be especially important when the environment contains light sources or reflective objects: if the exposure of the camera is chosen to focus on the important parts of the scene (i.e. a moving subject), part of the detail in the background can be lost.

In the assumption that the subjects are not interacting with light sources or reflective objects, we propose to acquire the HDR information of the static scene in a preliminary stage. Then, the LDR video can be acquired in the usual way, by setting the best exposure level for the moving subjects. In a completely automatic post-processing phase, the HDR

information is used to recover detail in all the over-exposed and under-exposed portions of each frame. In addition to the increase in visual quality of the video, also the video shooting phase is greatly helped. Even in the case of complex environments, it is possible to focus on the main elements of the scene, because the rest of the background will be enhanced in a subsequent phase.

The proposed approach can be useful in very different contexts: from the acquisition of medium-to-high quality videos, to the typical case of webcams, where a satisfying trade-off in exposure between the subject in foreground and the usually bright background is usually hard (if not impossible) to find.

The background enhancement technique is easy to obtain, automatic and fast to process. It represents a way to add HDR information to LDR videos, until the cost and usability of HDR video-cameras will make them usable for wide public.

2. Related Work

We organized the related work in three categories: HDR video acquisition using computational photography techniques, native HDR video-cameras, and techniques for enhancing videos from photographs.

2.1. Computational Photography

In computational photography, different approaches have been proposed to tweak different features of LDR video-cameras.

Aggarwal and Ahuja [AA01] proposed and implemented a video-camera design, where the aperture is split into multiple parts and the beam exiting from each is directed in a different direction using mirrors. This solution has recently become a popular technique for creating HDR videos [Sov10]. However, it requires two video-cameras and accurate alignment has to take place, to avoid misalignments artifacts. Moreover, practical solutions do not allow more than 2-3 cameras to be used. This limits the dynamic range that can be captured. Kang et al. [KUWS03] achieved to capture HDR videos by varying the shutter speed of the video-camera at subsequent frames. They use warping to register different exposure frames. Then, the aligned exposure frames are merged into a radiance map. Their method is effective, but a very high-speed camera needs to be employed, otherwise alignment artifacts can appear. Narasimhan and Nayar [NN02] proposed a Bayer-like pattern where exposure varies spatially. This method has the drawback that a more complex de-mosaicing algorithm needs to be applied and a trade-off in spatial resolution has to be found. Nayar and Branzoi [NB03] proposed a controllable liquid crystal light modulator in front of the camera, in order to have an adaptive dynamic range camera. Their modulator adapts the exposure of each pixel on the image detector, this allows to acquire scenes with a very large dynamic range. In this case, a perfect alignment between the sensor and the modulator is difficult to achieve, and filtering needs to take place to avoid artifacts.

2.2. HDR Video-cameras

In last few years, HDR CCD/CMOS sensors have been introduced by several companies, such as the HDRc Sensors line by IMS Chips [IC10]. These sensors allow to record into 10/12-bit channels in the logarithmic domain, but they typically capture at low resolutions, and can be very noisy. Therefore, they are mainly used in security and industrial manufacturing.

More recently, various HDR cameras have been presented such as the Gustavson's camera [UG07], HDRv by SpheronVR GmbH developed in collaboration with Warwick University [CBB*09], and RED's HDRx and easyHDR technologies [RED10]. All these solutions are extremely expensive. Moreover, they present a limited transportability as in the case of the HDRv due to the need of a 24 TB array in order to acquire videos.

2.3. LDR Videos Enhancement using Photographs

HDR video-cameras are extremely expensive and resource-demanding: for this reason, some research has been made

with the aim of enhancing LDR videos using images. One of the first methods for enhancing LDR videos was presented by Sand and Teller [ST04]. Their work matches different videos allowing to create HDR videos from two or more LDR videos. The system is prone to differences in the videos to match, but it cannot cope with large-scale differences between the images, such as an actor that appears in one frame but not the other.

Bennett and McMillian [BM05] introduced a system for enhancing low exposure LDR videos exploiting temporal coherence, bilateral filtering, and tone mapping. The main goal of the paper is improving the visual quality of a LDR video in terms of noise and dynamic range. However, the final dynamic range cannot be considered as HDR.

Ancuti [AHMB08] proposed a simple technique to transfer detail from an high resolution image to the video. Gupta [GBD*09] extended the concept by designing a framework that could be implemented in video-cameras which can acquire both images and videos. No HDR information is taken into account.

In a similar fashion to our method, Bhat et al. [BZS*07] proposed a method to enhance videos using photographs of the same scene. In their system, images can be taken from different positions, and the geometry of the scene needs to be estimated, possibly restraining the field of application. Moreover, this system works only for static scenes, while our approach deals with moving objects, from a fixed point of view.

Wang et al. [WWZ*07] proposed a method for increasing the dynamic range and transferring the detail from a source LDR image into a target LDR image. However, their method can not be applied to videos because it needs user interaction for increasing the dynamic range and transferring detail from a part of an image to the other one.

In our work, we automatically enhance LDR videos using HDR images in a set-up, where these are taken from the same position and device. The camera is fixed, but the scene is dynamic in contrast to Bhat et al.'s work [BZS*07]. This may be seen as the opposite of their methodology, where the camera is dynamic but the scene is static. Compared to Wang et al.'s work [WWZ*07] our method works with a real HDR background source, it uses straightforward techniques for transferring detail and dynamic range, and it is fully automatic. The main contributions of this paper are an acquisition technique for enhancing the dynamic range of LDR videos based on static camera, and a straightforward blending method for enhancing the input LDR videos which does not need image warping and/or complex techniques. To our knowledge, this kind of methodology, even if it seems to be quite straightforward, is not present in the literature.

3. The Acquisition Method

Our capturing method consists of three straightforward steps. Firstly, the acting scene is acquired in HDR captur-

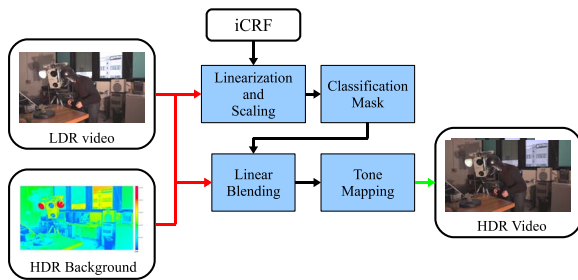


Figure 1: The proposed pipeline for augmenting LDR videos.

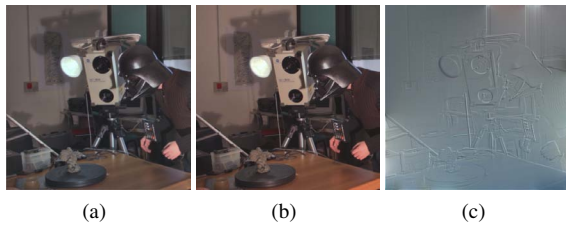


Figure 2: An example of different blending algorithm: a) Blending in the gradient domain. b) Blending in the spatial domain. c) The difference between a) and b). Note that gradients are slightly more enhanced in a) and colors are slightly shifted towards a bluish tint.

ing all the dynamic range of it, from dark to bright areas. Then, all different exposures are built into a radiance map using Debevec and Malik's method [DM97]. Secondly, an LDR video footage, where actors play in the scene, is acquired. The exposure of this footage is manually set in order to have actors well exposed. Finally, the background HDR image and the LDR image are processed by a straightforward blending algorithm.

3.1. The Blending Algorithm

In our approach, we blend the HDR image and LDR video in a straightforward way. The full pipeline is shown in Figure 1. Firstly, the LDR video footage is linearized applying the inverse Camera Response Function (iCRF) of the LDR video-camera. Assuming that the device that captured the HDR image and the LDR video were the same, the iCRF that is calculated during the creation of the radiance map can be reused. After linearization, the LDR video is scaled by the capturing shutter speed, obtaining a normalized LDR video with absolute values. Linearization and scaling are important in order to match intensities and colors with the reference. Moreover, this allows to use less computationally expensive techniques for the blending.

Secondly, the HDR image and the normalized and scaled LDR frame are linearly blended in the logarithm domain, using a selection mask M which classifies background and actors. The blending is applied in the logarithmic domain to



(a)



(b)



(c)

Figure 3: An example of the classification mask: a) A frame from an LDR video. b) The application of thresholding to a). c) The final mask after the application of the bilateral filter.

avoid seams at mask's boundaries. Then, the blended image is exponentiated to obtain the final radiance map. We found out from our experiments that this straightforward blend is enough to obtain no seams or other kind of artifacts. We tested blending using Laplacian Pyramids [BA87] and Gradient Domain techniques [BZCC10, PGB03, FLW02], but they produced similar results of the linear blending. Moreover, in some cases the colors are slightly shifted, see Figure 2 for a comparison. Therefore, we opted for the computationally cheaper solution of the linear blending.

3.2. The Classification Mask

The classification mask M is computed using thresholding on over-exposed and under-exposed pixels values on the luminance channel for each frame of the video. In our experiments, we found out that 0.95 and 0.05 are respectively good threshold for over-exposed and under-exposed thresholds (using normalized RGB values). Thresholding can produce groups of single pixels in the image which are typically to be considered as noise. Therefore, we apply morphological operators such as erosion followed by dilation. In our case, we empirically found that 3-5 iterations are typically enough to obtain good results on full HD content (1920×1080). Finally, the mask is cross bilateral filtered with the original frame LDR luminance using the bilateral grid [CPD07] ($\sigma_s = 16$ and $\sigma_r = 0.1$ for full HD content) in order to smoothly extend the classification to strong edges. Figure 3 shows an example of the different steps for calculating the mask.

4. Results

One of the goals of the proposed method was to create a method that could be used with a wide range of video-camera types. The acquisition step is very simple, and the processing is automatic. Hence, we decided to test the system using both a low-end and a medium/high-end camera.

In our experiments, we tested our capturing technique with a Logitech QuickCam Pro 9000 and a Canon 550D (or Rebel XTi) DSLR camera. The images depicting the reference HDR were tone mapped using Reinhard's operator implemented in [BADC10].

The Canon 550D is able to acquire videos at 1920×1080 full HD resolution, and it can be adapted to difficult lighting conditions. In the example shown in Figure 4, a tone mapped version of the HDR background is shown on the top. The middle image shows a frame of the original LDR video, where most of the background and the sky are over-exposed. Using the HDR background image, the enhanced video (bottom image) recovers the appearance of the sky and of several over-exposed parts.

A second example (Figure 5) shows an indoor environment, where the light and part of the outdoor scene are lost. The enhanced video recovers this information. The colors in the enhanced frame are different from the ones in the LDR frame because the linearization process matches color curves between HDR background and LDR frame.

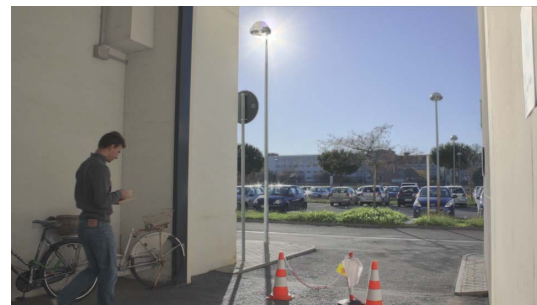
The second type of tested device (a medium-end webcam with a 1280×720 HD Ready resolution) covers for a different type of application: the enhancement of videos which are not able to adapt to difficult conditions. Webcams are usually not able to find a trade-off exposure between the foreground subject and the background. The example in Figure 6 allows to select the best exposure for the main



(a)



(b)



(c)

Figure 4: A frame of a video taken with Canon 550D (1920×1080 resolution): a) A tone mapped version the HDR reference of the background. b) A frame from the original LDR video. c) The same frame in b) after enhancement and tone mapping.

subject, while background information is recovered during the blending. The top image shows that a convincing HDR can be obtained also with this device, although part of the range is missing due its limitations.

Regarding timing, we ran our fully automatic enhancement algorithm, described in the previous Section, on an Intel Core 2 Duo at 2.33 Ghz equipped with 3Gb of memory and Windows 7. The algorithm was implemented in Matlab, and each frame took on average less than 9 seconds for a 1920×1080 frame. Note that the algorithm was run on a single core using unoptimized Matlab code expect the Mat-



(a)



(b)

Figure 5: An example of indoor scene taken with a Canon 550D (1920×1080): a) A frame from the original LDR video. c) The same frame in a) after enhancement and tone mapping.

lab Bilateral Grid implementation by Chen et al. [CPD07] which took around 4 seconds on average. All the operations of the algorithm are straightforward image processing operations such as thresholding, morphological operations (i.e. dilatation and erosion) which can be easily implemented on graphics hardware. Moreover, the Bilateral Grid can be implemented efficiently in real-time on GPUs [CPD07]. Therefore, a real-time implementation of the whole algorithm would be possible on graphics hardware, allowing to preview the final results.

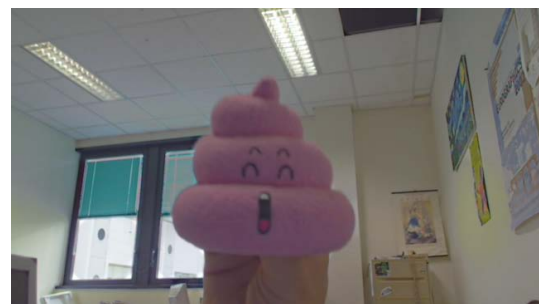
The main limitation of our algorithm is that the camera needs to be static. However, actors can play inside the scene, allowing to enhance the experience during teleconferencing or improve visual quality in movies. Another limitation is that our straightforward classifier based on threshold can have false positives. This means that an over-exposed moving object can be classified as background. However, this typically happens for very few frames, because the main exposure was set for an overall good exposure for actors and moving objects. Note that these artifacts cannot be noticed by the human eye, because they stand for only few frames, and they can be perceived as a reflection. A better classifier, based on motion estimation or tracking, would solve this problem.



(a)



(b)



(c)

Figure 6: A frame of a short sequence taken with a Logitech Quick-Cam Pro 9000 (1280×720 resolution): a) The tone mapped version of the HDR background. b) A frame from an LDR video of the scene. c) The same frame in a) after enhancement and tone mapping.

5. Conclusion and Future Work

In this paper, we presented a straightforward technique for increasing the dynamic range of LDR videos using HDR background images taken from the same camera and position. We showed that this technique is able to produce convincing videos at low cost. The method can be exploited for video teleconferencing when challenging backgrounds (i.e. with windows, and bright light sources) are present in the environment in order to enhance the teleconferencing experience. Moreover, the method can be employed for shooting movies in a similar manner as chroma key or blue-screen are used nowadays for enhancing the movie productions with vi-

sual effects.

In future work, we would like to improve the classification algorithm in order to remove the minor problems in case of over-exposed moving actors and/or objects. This classification could be coupled with reverse tone mapping operators [MAF*09, DMHS08, RTS*07] in order to improve the over-exposed areas of the actors and/or moving objects.

In addition, code re-factoring and GPU implementation could decrease the processing time to the possibility of having quasi-realtime enhancement: one of the possible applications of this could be video-conferencing.

Acknowledgements. We thank Gianpaolo Palma and Daniele Bernabei for their help in acquiring the video sequences. Marco Di Benedetto, Stefano Marras, and Daniele Bernabei played in these videos, and their contribution is gratefully acknowledged. We also thank the anonymous reviewers, whose suggestions improved the paper. The work presented in this paper was founded by the EC IST IP project “3D-COFORM” (IST-2008-231809).

References

- [AA01] AGGARWAL M., AHUJA N.: Split aperture imaging for high dynamic range. *Computer Vision, IEEE International Conference on 2* (2001), 10. 2
- [AHMB08] ANCUTI C., HABER T., MERTENS T., BEKAERT P.: Video enhancement using reference photographs. *The Visual Computer 24*, 7-9 (2008), 709–717. 2
- [BA87] BURT P. J., ADELSON E. H.: Readings in computer vision: issues, problems, principles, and paradigms. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987, ch. The Laplacian pyramid as a compact image code, pp. 671–679. 3
- [BAD10] BANTERLE F., ARTUSI A., DEBATTISTA K., CHALMERS A.: *Advanced High Dynamic Range Imaging: Theory and Practice*, first edition ed. AK Peters, Ltd, 2010. 4
- [BM05] BENNETT E. P., McMILLAN L.: Video enhancement using per-pixel virtual exposures. *ACM Trans. Graph. 24* (July 2005), 845–852. 2
- [BZCC10] BHAT P., ZITNICK C. L., COHEN M., CURLESS B.: Gradientshop: A gradient-domain optimization framework for image and video filtering. *ACM Trans. Graph. 29* (April 2010), 10:1–10:14. 3
- [BZS*07] BHAT P., ZITNICK C. L., SNAVELY N., AGARWALA A., AGRAWALA M., CURLESS B., COHEN M., KANG S. B.: Using photographs to enhance videos of a static scene. In *Rendering Techniques 2007 (Proceedings Eurographics Symposium on Rendering)* (jun 2007), Kautz J., Pattanaik S., (Eds.), Eurographics, pp. 327–338. 2
- [CBB*09] CHALMERS A., BONNET G., BANTERLE F., DUBLA P., DEBATTISTA K., ARTUSI A., MOIR C.: High-dynamic-range video solution. In *ACM SIGGRAPH ASIA 2009 Art Gallery & Emerging Technologies: Adaptation* (New York, NY, USA, 2009), SIGGRAPH ASIA '09, ACM, pp. 71–71. 2
- [CPD07] CHEN J., PARIS S., DURAND F.: Real-time edge-aware image processing with the bilateral grid. *ACM Trans. Graph. 26*, 3 (2007), 103. 4
- [DM97] DEBEVEC P., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1997), ACM Press/Addison-Wesley Publishing Co., pp. 369–378. 2
- [DMHS08] DIDYK P., MANTIUK R., HEIN M., SEIDEL H.-P.: Enhancement of bright video features for HDR displays. In *Proceeding of Eurographics Symposium on Rendering 2008* (2008), Computer Graphics Forum, Eurographics, Blackwell Ltd. 5
- [FLW02] FATTAL R., LISCHINSKI D., WERMAN M.: Gradient domain high dynamic range compression. *ACM Trans. Graph. 21*, 3 (2002), 249–256. 3
- [GBD*09] GUPTA A., BHAT P., DONTCHEVA M., CURLESS B., DEUSSEN O., COHEN M.: Enhancing and experiencing space-time resolution with videos and stills. In *International Conference on Computational Photography* (2009), IEEE. 2
- [IC10] IMS-CHIPS: HDRc Sensors. <http://www.ims-chips.de/home.php?id=a3b4c2en> (December 2010). 2
- [KUWS03] KANG S. B., UYTENDAELE M., WINDER S., SZELISKI R.: High dynamic range video. *ACM Trans. Graph. 22* (July 2003), 319–325. 2
- [MAF*09] MASIA B., AGUSTIN S., FLEMING R. W., SORKINE O., GUTIERREZ D.: Evaluation of reverse tone mapping through varying exposure conditions. *ACM Trans. Graph. 28*, 5 (2009), 1–8. 5
- [NB03] NAYAR S., BRANZOI V.: Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures over Space and Time. In *IEEE International Conference on Computer Vision (ICCV)* (Oct 2003), vol. 2, pp. 1168–1175. 2
- [NN02] NAYAR S., NARASIMHAN S.: Assorted Pixels: Multi-Sampled Imaging With Structural Models. In *European Conference on Computer Vision (ECCV)* (May 2002), vol. IV, pp. 636–652. 2
- [PGB03] PÉREZ P., GANGNET M., BLAKE A.: Poisson image editing. *ACM Trans. Graph. 22* (July 2003), 313–318. 3
- [RED10] RED: Hdrx. In <http://www.red.com/> (2010). 2
- [RTS*07] REMPEL A. G., TRENTACOSTE M., SEETZEN H., YOUNG H. D., HEIDRICH W., WHITEHEAD L., WARD G.: Ldr2hdr: on-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph. 26*, 3 (2007), 39. 5
- [RWP*10] REINHARD E., WARD G., PATTANAİK S., DEBEVEC P., HEIDRICH W., MYSZKOWSKI K.: *High Dynamic Range Imaging, Second Edition: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2010. 1
- [Sov10] SOVIETMONTAGEPRODUCTIONS: HDR Video Demonstration Using Two Canon 5D mark II's. In <http://www.sovietmontage.com/> (San Francisco, CA, USA, 2010). 2
- [ST04] SAND P., TELLER S.: Video matching. *ACM Trans. Graph. 23* (August 2004), 592–599. 2
- [UG07] UNGER J., GUSTAVSON S.: High dynamic range video for photometric measurement of illumination. In *In Proceedings of Sensors, Cameras and Systems for Scientific/Industrial Applications X, IS&T/SPIE 19th International Symposium on Electronic Imaging* (2007), vol. 6501. 2
- [WWZ*07] WANG L., WEI L.-Y., ZHOU K., GUO B., SHUM H.-Y.: High dynamic range image hallucination. In *Proceedings of Eurographics Symposium on Rendering* (Jun 2007). 2