An Algorithm for Extended Dynamic Range Video in Embedded Systems

Alberto Ferrante[1], Massimo Chelodi[1], Francesco Bruschi[1] and Valeria Mozzetti[2]
[1]*ALaRI, Faculty of Informatics, Università della Svizzera italiana, Lugano, Switzerland*
[2]*ViSSee AG, Zurich, Switzerland*

Keywords: Video Camera, Extended Dynamic Range, Dynamic Range, Video, Visual Sensor.

Abstract: Video cameras are gaining popularity in embedded systems, either used directly or as sensing devices. Being embedded systems often portable, they can be employed in different environments with various lightning conditions; mutating lightning conditions may pose some problems to conventional video-cameras. Yet, high-dynamic range video-cameras can be too expensive for certain applications. In this paper we propose an algorithm that extends the dynamic range of common video-cameras by using limited computational resources, yet exploiting the original frame rate of the camera. Furthermore, we discuss some results obtained by using our algorithm coupled with a commercial webcam.

1 INTRODUCTION

Video-cameras are used for different applications, ranging from the consumer market to video surveillance and other industrial applications. Among all of these applications, video-cameras can be also used as input devices of different kinds of sensors. In this case, the final user is not interested in the camera output; his interest is in the final output of the sensor: image quality as perceived by the human eye is not important in this case. It is fundamental, though, that the output of the image sensor exhibits the characteristics required by the successive elaborations performed in the sensor. For example, ViSSee AG (vis, 2009) produces speed sensors that are based on visual inputs. When this sensor is used, the output is a speed measure and the images from the video-camera are normally not visible to the final application or user.

One of the fundamental characteristics of image sensors is dynamic range. Dynamic range represents the maximum difference in luminosity among different zones of the image that the sensor can register. Unfortunately, many of the typical scenes contain ranges of luminosities that cannot be represented by the camera sensor (i.e., the dynamic range of the scene is greater than the one allowed by the sensor); this causes some zones of the images to be overexposed (i.e., completely white, with no details) or underexposed (i.e., completely black, with no details). Overexposed or underexposed portions of the images are unusable in most of the cases as they contain little or no detail. Furthermore, due to the reduced dynamic range of image sensors, most of the cameras require their exposure to be adjusted depending on the lightning conditions. Most of the automatic exposure adjustment algorithms, though, require some time to change the exposure appropriately. During this time most of the images may result to be completely overexposed or underexposed.

A well known photographic technique for increasing dynamic range relies on merging multiple exposures (typically 2 or 3) of the same scene taken with different exposure times. This technique is sometimes also used in videos.

In this paper we discuss an algorithm aimed at cameras used in embedded devices; video-cameras, in this case, are not required to provide a good visual output. Our algorithm increases dynamic range by relying on the merge of different exposures, yet maintaining the original frame rate of the camera.

The remainder of this paper is organized as follows: in Section 2 we discuss the related work; in Section 3 we describe the algorithms that we developed; in Section 4 we discuss an implementation of the algorithms and we show the results that we have obtained.

2 RELATED WORK

Extending dynamic range of image sensors has been the subject of a number of research and industrial works. In many currently available digital cameras

Ferrante A., Chelodi M., Bruschi F. and Mozzetti V..
An Algorithm for Extended Dynamic Range Video in Embedded Systems.
DOI: 10.5220/0004207001810186
In *Proceedings of the 2nd International Conference on Sensor Networks* (SENSORNETS-2013), pages 181-186
ISBN: 978-989-8565-45-7

there is the possibility of creating extended dynamic range images by merging different frames (e.g., the Canon EOS 5D Mark III, can automatically produce an extended dynamic range image from up to 3 photos (EOS, 2012)). This functionality, though, is not available in current consumer video cameras.

Hardware solutions are based on extended dynamic range image sensors. One of these sensors, for example, is the IcyCAM (Arm et al., 2008). This image sensor, though, is only available as as an IP core that should be implemented on chip.

One software method that requires low level access to the sensor is discussed in (Bandoh et al., 2010). Other methods, such as the ones discussed in (Schulz et al., 2007; Jinno and Okuda, 2012; Guarnieri et al., 2011; Xinqiao and Gamal, 2003; Mertens et al., 2007; Lu et al., 2009; Castro et al., 2011; Wang et al., 2011), are software only, even though they may require the knowledge of different low level parameters of the sensors. All the software techniques are based on the combination of multiple frames to obtain a final image with extended dynamic range.

In the present paper we propose an algorithm for extended dynamic range videos that relies on general purpose video-cameras; our algorithm is inspired to the one presented in (Schulz et al., 2007). Our method differs in the fact that it requires no knowledge of the physical sensor parameters and that it produces videos with the same frame rate of the camera. Our method is also low on computational resources demand and can be easily used in embedded systems.

3 ALGORITHMS

We have designed two different variants of our algorithm, both of them allowing the system to produce final images with the same frame rate of the video-camera employed as image source. The first variant merges two frames as explained in (Schulz et al., 2007); the sequence of images to be merged, though, is different than the original one. The second variant allows considering three different exposure values and quickly switching among them to optimize dynamic range. We have tuned the parameters of the algorithm without any specific knowledge on the image sensors and by using, as a reference, only the images obtained from the camera.

In this section we introduce the two variants of the algorithm that we have designed and we discuss a methodology to determine the correct exposure values for the frames that should be considered during merge. We have designed our algorithm with black and white images in mind, as our applications do not require color images; the algorithm, though, can be used without any modification.

3.1 Two-frame Merge

In this variant of the algorithm two frames are considered, one with shorter exposure time, named *dark*, and another one with longer exposure time, named *light*. The dark frame is merged to the light one with the goal of extracting the maximum number of details from both images: the loss of details in the black areas of the dark frame are compensated with the details that are in the same zones of the light frame. Details lost in the white areas of the light frame are compensated by using the details taken from the dark frame. The merge is performed by applying the following formula for each pixel j of the image:

$$m_j = \begin{cases} l_j, & l_j < L_1 \\ \left(1 - \frac{l_j - L_1}{L_2 - L_1}\right) l_j + \frac{l_j - L_1}{L_2 - L_1} d_j & L_1 \leq l_j \leq L_2 \\ d_j, & l_j > L_2 \end{cases} \tag{1}$$

where l_j and d_j represent the pixels of the light and the dark images, respectively; L_1 is a threshold on the luminosity that defines underexposed pixels; similarly, L_2 is a threshold that defines overexposed pixels.

Let us name each frame generated by the camera F_i. When i is odd, the shortest exposure time is set and, therefore, F_i corresponds to a dark image; when i is even, the longest exposure time is set and, therefore, F_i corresponds to a light image. In our algorithm, differently than the one explained in (Schulz et al., 2007), we adopted a way to merge images that allows obtaining a resulting image at each i (i.e., at each image acquired by the camera), instead of only at even values of i:

$$R_i = \begin{cases} merge(A_i, A_{i-1}) & \text{i odd}, \quad i > 1 \\ merge(A_{i-1}, A_i) & \text{i even} \end{cases} \tag{2}$$

The merge parameters need to be determined by considering the characteristics of the camera and the lightning conditions foreseen for the usage of the device. In particular, the two exposure values need to be determined in such a way that the dynamic range of the final image is maximized and that also the midtones of the image are present. In fact, if the exposure values are too far away one another, details in the middle gray levels are lost. An example of these parameters is shown in Section 4.

The merge function needs to be applied an all pixels of the image and, supposing that the values of the function f are tabbed instead of computed every time, it uses two integer sums and two integer multiplications for each pixel. Two frames at a time (if the oldest frame is overwritten with the results from the merge) need to be kept in memory. Therefore, the resources used by the algorithms are limited and compatible with most of the embedded systems where the algorithm can be adopted.

3.2 Two-frame Adaptive Merge

The dynamic range of a frame obtained by merging two images is higher than the one of a single frame obtained from the camera, but it is still limited. In some situations it may be desirable to increase the dynamic range even further. For this purpose, more than two frames can be merged (e.g., three). The solution that we propose here, though, uses only two frames for merging, even though it relies on three fixed exposure values. The algorithm that we propose, in fact, is able to choose dynamically which couple of exposure values should be considered at a given time. As mentioned earlier, three exposure values are considered instead of three, one for the darkest frame, one for lightest frame, and one that sits in the middle. The distance between the exposure time of dark and the one of the medium frame, as well as the distance between the exposure time of the medium and the one of the light frame can be considered similar to the one adopted between the dark and the light frames in the non adaptive version of the two-frame merge. Considering the three exposure values, there are two different kinds of sequences of exposure times that can be used: one corresponds to dark and medium frames; the other one to medium and light frames. Whether to use the light or the dark images for the merge is decided by comparing the previously obtained merged image with the one exposed for mid-tones: if the image obtained through the merge has similar average luminosity of the one exposed for mid-tones, it means that the merge operation has revealed limited details (i.e., both the merge and the image exposed for mid-tones are possibly overexposed or underexposed). The average luminosity can be computed by summing up the luminosity of each pixel. No division by the total number of pixels is necessary as it is a constant that in the comparison can be omitted. This evaluation is only performed for even values of *i*. Let us call L_i^M the average luminosity of the frame obtained from the merge and L_i^m the average luminosity of the frame with mid-level exposure. Let us also call S the threshold that we use to define similarity between L_i^M and L_i^m. S is expressed as a percentage.

The camera will start acquiring sequences of frames in which the dark frames are considered for odd values of *i* and the medium frames are considered for even values of *i*. For even values of *i* L_i^M and L_i^m are evaluated with the following possibilities:

1. $\left|L_i^M - L_i^m\right| \geq S * L_i^m$: no adaptation required;
2. $\left|L_i^M - L_i^m\right| < S * L_i^m$: the new sequence will be composed by medium frames for even *i*s and and light frames for odd *i*s; the resulting images will be obtained as follows:

$$R_i = \begin{cases} merge(A_i, A_{i-1}) & \text{i odd}, \quad i > 1 \\ merge(A_{i-1}, A_i) & \text{i even} \end{cases} \tag{3}$$

A change from medium and light frames to dark and medium ones is performed in a similar way when $\left|L_i^M - L_i^m\right| < S$,

As in the two-frame merge, the merge function needs to be applied an all pixels of the image and it uses two integer sums, two integer multiplications for each pixel. Two additional sums are required when even frames are considered to compute the average luminosity values of the medium frame and of the frame resulting from the merge operation. Two frames at a time need to be kept in memory. Therefore, the resources used by the algorithms are limited and compatible with most of the embedded systems where the algorithm can be adopted.

3.3 Exposure Values

By using our algorithm, two frames are merged to obtain a single image with extended dynamic range; therefore, exposure parameters of the two frames must be chosen carefully to obtain the best possible results in the largest possible range of lightning conditions. Exposure is determined by the couple exposure time-gain; the latter is also called ISO in photography (ISO, 2012). Our main goal is to preserve as many details as possible in the light areas, in the dark areas, and in the mid-tones. The main constraint in computing the exposure values resides in the longest exposure time that must be shorter than $1/frame-rate$; this constraint may be problematic when lightning is scarce. Gain can be increased to obtain shorter exposure times at the price of increasing image noise.

The exposure parameters are determined by means of experiments performed in the different lightning conditions considered. Different images of a reference scene are taken by considering these lightning conditions and different couple of exposure parameters. The reference scene if represented by a Kodak/Tiffen Q-13 gray scale. This printed image a quality control device is chart consisting of 20 zones, labeled 0-19, which have optical densities from 0.0 (white) to 1.90 (practical printing black) in steps of 1/3rd of EV/f-stop. The total range of values represented by the target is of 6 and 2/3rd f-stops. The obtained images are evaluated by considering the number number of gray patches retained by them, both in the highlights and in the shadows. In particular, we consider the most underexposed image for the dark scene and the most overexposed one for the light scene in which all gray patches can be distinguished one another. If the exposure parameters of the underexposed image of the dark scene are compatible with the ones of the overexposed image of the light scene, the correct value for the exposure parameters is chosen among the ones belonging to the intersection of the results obtained in the two extreme lightning conditions. Otherwise it is not possible to cover the considered EV range.

4 IMPLEMENTATION AND RESULTS

In this section we discuss the results obtained by with our algorithm.

A Logitech Webcam Pro9000 has been used for testing. The algorithm discussed in Section 3 have been implemented in C under Linux and the *Video4Linux2* (v4l2)(Schimek et al., 2008) API have been used to control the webcam and to acquire the video frames. The ImageMagic suite of programs have been used for image manipulation and conversion. The values chosen for L_1 and L_2 of Equation 1 are 330 and $165'000$, respectively. S of Section 3.2 have been set to 15%

4.1 Dynamic Range Measurement

The dynamic range measurement methodology that we have used relies on the aforementioned Kodak/Tiffen Q-13 gray scale. The target need to be front-lighted evenly; in our case we used a diffused light lamp. By considering a constant light level, a series of pictures, of the target are shot with different exposure values; we used exposure values in steps of of 1/3 f-stop. Dynamic range is obtained by counting the number of gray patches that are visible in the images. Since the dynamic range of the sensor is, in most of the cases, larger than 6 and 2/3rd f-stops, multiple images need to be used for this purpose. Using multiple images obtained with different exposure levels is approximately equivalent to having a longer grayscale. Multiple images are used in the following way:

1. Thee set A of the images in which all the 20 gray levels are distinguishable is identified.
2. The images with lowest (a_l) and highest (a_h) exposure values are selected in set A.

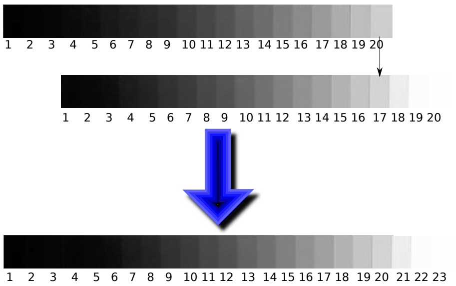


Figure 1: Overlapping of two different photos of the target to compute the dynamic range. Please notice that some gray levels may not be distinguishable one another on screen or paper, but they are by measuring the luminosity by using appropriate pieces of software.

3. The smallest set (I) of images with intermediate exposure values are selected in set A.
4. A single image (g) is formed by joining a_l, a_h, and the images in the set I such as the values of the gray levels overlap as shown in Figure 1.
5. The number of gray levels that are present in g are counted. As 3 gray patches correspond to EV1, the dynamic range in dB is obtained by multiplying the number of gray levels of g by 2.0667.

If no image shows the full range of gray levels (i.e., the dynamic range of the camera is lower than 6 and 2/3 f-stop), the dynamic range can be simply obtained by selecting one of the images where the highest number of gray patches can be distinguished and counting their number. To validate the procedure we performed a test by using a camera with a known dynamic range.

The procedure to measure the dynamic range of images resulting from merge is different as the notion of exposure time gets lost in the final images. Also in this case a set of images, obtained through merge, are considered. Of all the obtained pictures, the brighter valid merge (all the 20 gray levels can be seen) and the darker valid merge are selected. The darker original (i.e., before merge) fame of the darker image obtained from merge and the lighter original frame of the lighter image obtained from merge need to be considered. A gap in the gray scale will be present between the two images; this gap needs to be filled, with the same procedure described above for measuring the dynamic range of non-merged images, by using other images (non-merged) taken with exposure settings in between the two of the darker and of the lighter images. The total number of gray patches is obtained as 40 plus the number of gray patches used to fill the distance between the two original images.

Figure 2: Ghosting effect resulting from the merge of two frames; the camera was moved horizontally. The ghosting effect is visible on the borders of the lamp.

4.2 Ghosting Evaluation

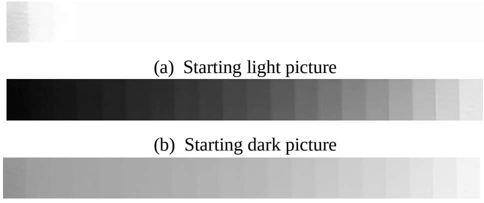
Ghosting is due to motion of the camera and/or of the subject between successive frames that are later merged; due to this motion, the border of some zones of the image are repeated multiple times in the resulting image, thus creating a ghost image, as shown in Figure 2. Ghosting depends on different parameters such as the speed of movement, the frames per second of the camera, the merge algorithm, the subject considered, and the lightning conditions. The higher the speed and the lower the frames per second, the higher the ghosting may potentially be. Ghosting is concentrated around the borders between zones with high luminosity and zones with with low luminosity.

Some of the software methods for extending the dynamic range of cameras discussed in Section 2 include techniques to prevent or mitigate the problem of ghosting; hardware solutions do not usually have ghosting problems. (Khan et al., 2006) describes an approach to removing ghosting artifacts from high dynamic range images, without the need for explicit object detection and motion estimation.

The presence of ghosting in the images is very difficult to quantify without considering a proper reference scene. We considered artificially created scenes, composed by gray patches. This allows leaving as variables only the ones connected to the merge algorithm considered. In particular, we considered two different merges obtained from: a dark frame with half of the patches underexposed and a light frame with half of the patches overexposed; a dark frame with all the bins underexposed and a light frame with all bins correctly exposed. For simulating motion we shifted the second frame on the right by a certain number of pixels.

4.3 Results

We measured the dynamic range and we evaluated ghosting for each variants of the algorithm introduced in Section 3. The dynamic range measured for the Logitech Webcam Pro9000 that we used in our experiments is 47.5dB, as shown in Figure Figure 1.



(a) Starting light picture

(b) Starting dark picture

(c) Merge

Figure 3: Original pictures and the merge result obtained with EV 15 (full sunlight at noon).

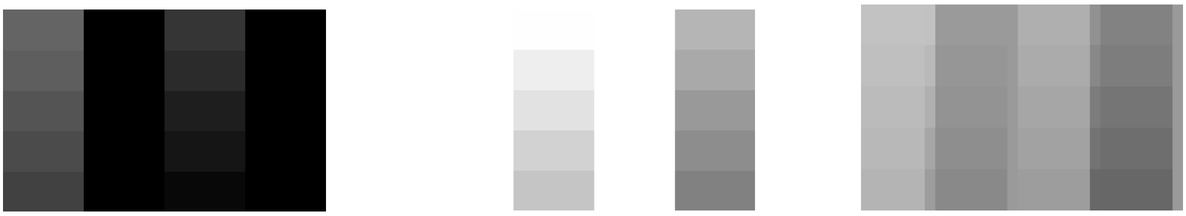
When considering the two-frame merge, we have chosen exposure times that can be used in most of the normal natural lightning conditions that can occur outdoor during the day in Summer (from dark, EV 1 to full light, direct sunlight at noon, EV 15). Considering that the frame rate of our webcam camera is of 20 frames per second, the longer exposure time must be shorter than 1/20s. The webcam does not provide accurate results for exposure values below EV 7 (typical light 10 minutes after sunset). This greatly limits the possibility of obtaining exposure times that can be used in darks environments. The exposure values chosen for these experiments are 1/1000s and 1/30s (5 f-stops distance).

The dynamic range measured when applying the merge algorithm is 93dB, that is almost double the one of the considered camera (47.5dB).

From the ghosting stand point the worst case happens when half of the bins of the dark image are underexposed and half bins of the light image are overexposed. In this case, as shown in Figure 4, halos are present in all vertical borders of the image (the movement is simulated to be horizontal in these experiments). Due to the sequence of dark and light images considered during merge, ghosting is not constant in one direction: halos will be in the same direction of the motion in one frame and in the opposite direction in the successive frame.

The dynamic range of the IcyCAM is 132dB (Rüedi and Gray, 2008). The IcyCAM is also free from ghosting thus exhibiting better performances than our system. The IcyCAM, though, is much more expensive than a standard video-camera sensor.

Two-frame adaptive merge provides the same results, in term of dynamic range and ghosting, as the two-frame merge, when single final frames are considered. The main difference, in this case, is that there are two different possible sets of exposure values that can be considered and this virtually doubles the number of lightning conditions supported. This algorithm



(a) Half bins under-exposed (b) Half bins over-exposed (c) Merge all bins with all bins

Figure 4: Halos when half bins of the dark image are underexposed and half bin of the light image are overexposed.

does not provide the same dynamic range, in a single image, as a merge obtained from three images. Merging three frames, though, increases significantly the problems with ghosting.

5 CONCLUSIONS AND FUTURE WORK

In this paper we discussed two variants of an algorithm aimed at extending the dynamic range of videocameras. The main characteristics of this algorithm is that it allows using a normal video camera (i.e., with no special sensors and with no specific knowledge about it) and it allows maintaining the same frame rate of which the camera is capable. Furthermore, the algorithm is quite simple and, therefore, it uses limited computational resources. In the paper we show, through experimental results, the effectiveness of the algorithm in enlarging the dynamic range of the camera adopted.

Future work will concentrate on methods for removing ghosting in the final frames. The method of choice must be compatible with low-cost and low-resources embedded systems: it must not be computationally too expensive.

REFERENCES

(2009). ViSSee AG. http://www.vissee.ch.

(2012). *Canon EOS 5D Mark III Instruction Manual*, pages 173–176.

(2012). ISO 5800:1987 - Colour negative films for still photography – Determination of ISO speed. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm? csnumber=11948.

Arm, C., Gyger, S., Heim, P., Kaess, F., Nagel, J. L., Rüedi, P. F., and Todeschini, S. (2008). IcyCAM - a high dynamic range vision system. Technical report, CSEM - Centre Suisse d'électronique et de microelectronique.

Bandoh, Y., Qiu, G., Okuda, M., Daly, S., Aach, T., and Au, O. C. (2010). Recent advance in high dynamic range imaging technology. *Sensors Peterborough NH*, pages 3125–3128.

Castro, T., Chapiro, A., Cicconet, M., and Velho, L. (2011). Towards Mobile HDR Video. Poster: IEEE International Conference on Computational Photography.

Guarnieri, G., Marsi, S., and Ramponi, G. (2011). High Dynamic Range Image Display With Halo and Clipping Prevention . *Image Processing, IEEE Transactions* , 20(5):1351–1362.

Jinno, T. and Okuda, M. (2012). Multiple Exposure Fusion for High Dynamic Range Image Acquisition. *Image Processing, IEEE Transactions* , 21(1):358–365.

Khan, E., Akyuz, A., and Reinhard, E. (2006). *Ghost Removal in High Dynamic Range Images*, pages 2005–2008. IEEE.

Lu, P.-y., Wu, M.-s., Cheng, Y.-t., and Chuang, Y.-y. (2009). High dynamic range image reconstruction from hand-held cameras. *IEEE Conference on Computer Vision and Pattern Recognition (2009)*, 54(1):509–516.

Mertens, T., Kautz, J., and Van Reeth, F. (2007). Exposure fusion. *15th Pacific Conference on Computer Graphics and Applications PG07*, pages 382–390.

Rüedi, P.-F. and Gray, S. (2008). IcyCAM - high dynamic range system-on-chip for vision systems combining a 132db qvga pixel array and a 32-bit dsp/mcu processor. Technical report, CSEM - Centre Suisse d'électronique et de microelectronique.

Schimek, M. H., Dirks, B., Verkuil, H., and Rubli, M. (2008). Video for linux two API specification.

Schulz, S., Grimm, M., and rainer Grigat, R. (2007). Using brightness histogram to perform optimum auto exposure. In *WSEAS Transactions On Systems and Control*, volume 2, pages 93–100. World Scientific and Engineering Academy and Society.

Wang, H., Cao, J., Tang, L., and Tang, Y. (2011). Hdr image synthesis based on multi-exposure color images. In Tan, H., editor, *Informatics in Control, Automation and Robotics*, volume 132 of *Lecture Notes in Electrical Engineering*, pages 117–123. Springer Berlin Heidelberg.

Xinqiao, L. and Gamal, A. E. (2003). Synthesis of High Dynamic Range Motion Blur Free Image From Multiple Captures . *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions* , 50(4):530–539.