AN OPTIMAL VIDEO-SURVEILLANCE APPROACH FOR HDR VIDEOS TONE MAPPING

Alberto Boschetti ⋆ *Nicola Adami* ⋆ *Riccardo Leonardi* ⋆ *Masahiro Okuda* †

⋆ Department of Information Engineering, University of Brescia, Italy
† Department of Information and Media Sciences, The University of Kitakyushu, Fukuoka, Japan

ABSTRACT

Recently, a new type of camera, which allows the recording of almost the entire range of luminosity of the acquired scene, has been introduced to the market. The produced High Dynamic Range images (HDRi), can simplify some video-surveillance tasks, such as object detection, recognition and tracking, since they can simultaneously capture the visual scene content of both dark and bright areas. Applications where lighting conditions cannot be controlled can greatly benefit from the adoption of HDR cameras.

For example, the achievement of a good and constant visual quality of images is one of the key aspects in video-surveillance applications, because it allows to robustly detect salient events in the captured scenes. However, the use of HDR images in traditional video-surveillance systems requires the reduction of their dynamic range appropriately, by applying a tone mapping operator.

In this paper a fast method for tone mapping HDR videos, which combines the benefits of both local and global operators, is presented. It is applied in the context of object detection and tracking. It also enhances the visual quality of the image in all light conditions, which facilitated surveillance tasks for both human and automatic operators.

1. INTRODUCTION

The quality of the acquired images is one of the most relevant issues in video-surveillance because it highly impacts the performance of automatic tasks, such as automatic object detection, recognition and tracking.

Cameras for video monitoring are selected according to the requirements of their specific application. However, the final image quality is greatly affected by the illumination conditions of the scene, which sometimes cannot be completely controlled, for example in outdoor situations. As a matter of fact, typical professional cameras, equipped with a CCD or a CMOS sensor, cannot accurately describe the real world luminance in every type of light condition, because of the limited dynamic range of their sensors. A few years ago a new model of camera, which provides High Dynamic Range (HDR) images, has been introduced to the market. Thanks to its new acquisition sensor technology, it can capture much higher contrast than classical cameras. In fact, HDR sensors are able to capture up to 12 orders of magnitude of luminance. Due to their improved dynamic resolution, the captured images contain a greater amount of information, which can significantly improve automatic surveillance tasks as well al the viewing conditions for human operators.

Since HDR images contain more information than classical ones, in order to use an HDR video in conventional surveillance or monitoring system, it is required to adapt them. The adaptation is accomplished by companding the dynamic range of the images to the one supported by the processing and visualization device. In fact, classical monitors are only able to display only 8 *bcc* images (hence a 2-magnitude order of luminance), therefore a dynamic reduction process should be applied to the HDR image. This process, called Tone Mapping (TM), aims to optimally map the information from the HDR range ($[0, +\infty)$, real) to the displayable range (typically $[0, 255]$, integer).

According to [1], four classes of tone mapping operators (TMO) have been proposed in literature. They can be subdivided into: *local operators*, *global operators*, *frequency domains operators* and *gradient domain operators*. A simple and fast global TM operator is provided by the following expression:

$$LDR(x) = \frac{HDR(x)}{HDR(x) + 1} \tag{1}$$

where $HDR(x)$ is the absolute intensity of luminance of the pixel in position $\mathbf{x}$ and $LDR(x)$ represents the tone mapped (or *Low Dynamic Range*) value in the same position $\mathbf{x}$. The Eq. 1 remaps original values from the range $[0, +\infty)$ to the limited range $[0, 1)$, in a fast and simple way. This is a global operator because the tone mapped value of a single pixel depends only on its original luminance value and not on those of its neighbors.

Frequency domain, gradient domain and local domain operators work locally and the provided tone mapped value of a generic pixel is a function of the absolute luminance of the pixel itself and of its neighbours. For this reason, in the following parts, all these operators will be referred to as "local".

Due to their adaptability, local operators typically provide good image quality with an optimal local contrast, at the expense of high computational cost. On the other hand, global operators are faster but they provide less accurate results. A comparison between the average tone mapping time required by some local and global operators is presented in Figure 1, as a function of the number of the pixels in the image, using [2].

Table 1: Pros and cons of local and global tone mapping algorithms

Feature	Local TMO	Global TMO
Time	↑	↓
Complexity	↑	↓
Visual result	↑	↓
Surveillance quality	↑	↓
Edge quality quality	↑	↓
Contrast quality	↑	↓
Illumination independence	↑	↓

When image quality is a concern, local operators are the best choice ([3]), because they can locally adapt the remap function to the image's content. More precisely, they enhance visibility in dark areas even when very bright areas are present in the same image ([4], [5]), thus providing an optimal contrast for object detection. Conversely, global operators do not perform well in the above conditions, but they provide a fast result.

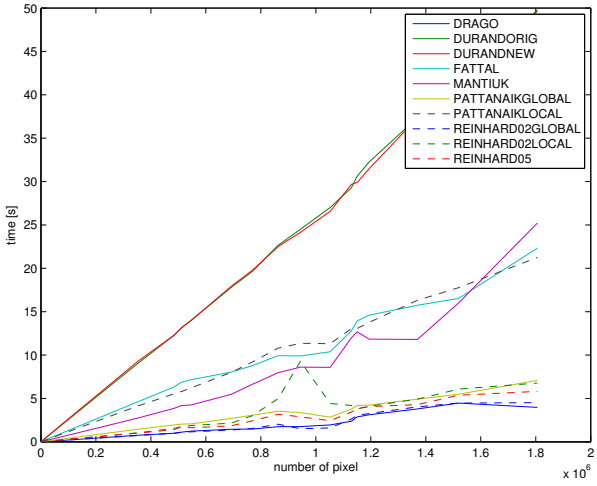

Figure 1: Average time required by different TM operators with respect to the frame size (in pixels), using [2]

According to the previous graph, in the context of video-surveillance, neither global nor local tone mapping functions can be efficiently and effectively used. The global operators provide low quality images, while the local ones take too much processing time. The field of video-surveillance typically requires both fast (possibly real-time) and accurate results ([6], [7] and [8]). Therefore, a mediation among local and global operators should be used. Fortunately, in some video-surveillance applications, like object detection, tracking and pattern recognition and identification, only some portions of the given image are of particular interest.

The main idea of the proposed algorithm is quite straightforward. In order to reduce the computational time and to achieve a good visual quality and high object detection accuracy, the video sequence is temporally sub-sampled by a factor GL (gop length). Video-surveillance algorithms, such as object detection, are then applied only to the sub-sampled sequence tone mapped with a local TMO. The object detection algorithms provide the Regions of Interest (RoI) containing the target objects in these frames. In the remaining images of the original video sequence, the tone map strategy is adaptively selected between local and global. Inside the RoIs, a local tone map operator is applied, while a global one is used in the remaining part. Then, in order to smooth possible discontinuity between RoIs and the other areas, visual data is interpolated along the RoI's boundary.

The paper is organized as follows: section 2 presents the proposed algorithm, section 3 describes and discusses the parameters of the system, some results and performance of the proposed approach. Finally, section 4 concludes the manuscript and provides some insight on future evolution of this work.

2. THE PROPOSED ALGORITHM

The goal of the proposed algorithm is to provide an efficient and effective tone map algorithm, built by exploiting the strength of both global and local tone map algorithms. A global view of our approach is shown in Figure 2.

The main idea is to apply a local tone map operator only inside the areas of interest for the considered surveillance application, while the remaining part of the picture is tone mapped using a global algorithm. In this way, it is possible to achieve several benefits: the computational time is reduced, the image contrast is enhanced in all the relevant areas, making these areas easily detectable by a human

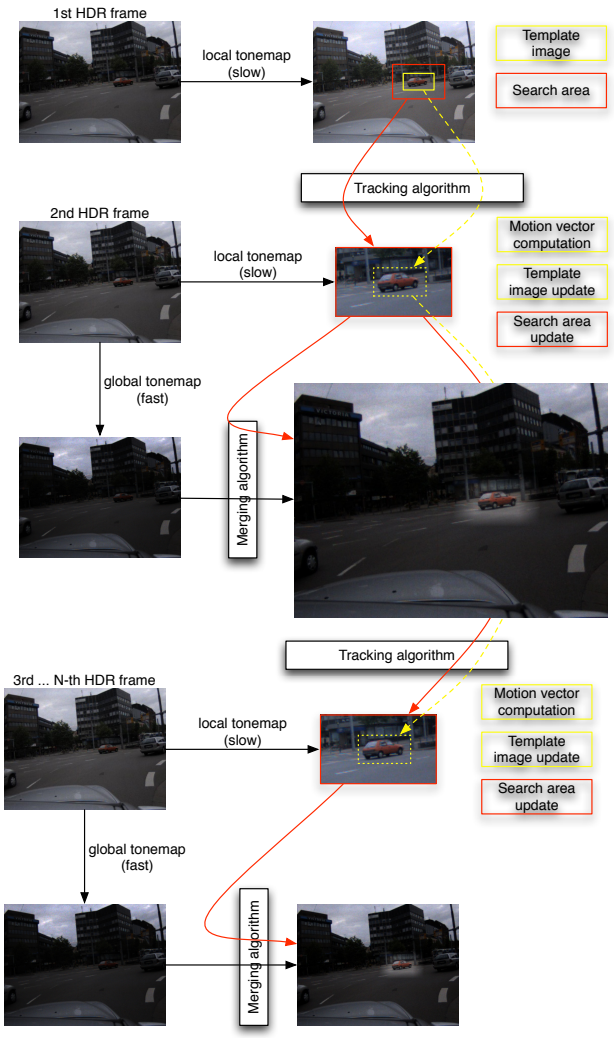

Figure 2: An intuitive block schema of the proposed algorithm

observer (see Fig. 2).

In the proposed method, similarly to what happens in video coding, the considered HDR video is initially partitioned into GOPs, (*Group Of Pictures*), where the first frame represents a "Key Picture". All of the Key Pictures, which are used to refresh the object detection and tracking processes, are tone mapped by applying a local algorithm, while all the other frames in a GOP are adaptively tone mapped.

The process starts with an initialization step, where a number of Regions of Interest (ROI) are selected from the key picture, according to the considered surveillance application. A local tone mapping is then performed on the previously selected ROIs, appropriately enlarged according to the search window defined in the object tracking algorithm of the second frame. Then, the video-surveillance algorithm (typically object tracking) is applied to each of these areas, in order to track temporal movements of the object in the image plane (e.g. the red car in Fig. 2). After this step, ROI parameters (displacement and size) are updated, according to the output of the object tracking algorithm. The updated parameters are then applied to the successive frame, the third in the given example. Finally the process is repeated for all the other frames inside the GOP. The transition between globally and locally tone mapped areas has

been smoothed by using a bilinear interpolation, described in Section 2.2.

2.1 Length of the GOP

The GOP length GL, i.e. the number of frames that compose a GOP, should be appropriately set in order to achieve the best trade off between speed and tracking robustness. The time required for tone mapping an entire movie, composed of N frames, is given by the following equation:

$$T_{TM}[s] = \frac{N}{GL}\overline{T_{TMglobal}} + \left(N - \frac{N}{GL}\right)\overline{T_{TMupdate}} \quad (2)$$

where $\overline{T_{TMglobal}}$ is the average time needed for the global tone mapping algorithm and $\overline{T_{TMupdate}}$ is the average time needed for the local tone mapping of ROIs and the bilinear interpolation.

If GL is increased, the time needed for tone mapping the whole movie T_{TM} decreases, because $\overline{T_{TMglobal}} << \overline{T_{TMupdate}}$. In this case, the visual quality of the video-surveillance is also decreased, because all the new objects (detected only in the key frames) are inspected with a low frequency.

Oppositely, if GL is decreased, the tracking robustness increases, as well as the tone mapping time T_{TM}, because the local tone map algorithm is applied with a higher rate.

In conclusion, an optimal GL needs to be computed and set during the start-up phase. It should be:

1. high enough to discover all new video-surveillance events in the key frames and
2. not excessive, to obtain the best tone map time (see Eq. 2)

This problem is very similar to the issue of computing the optimal GOP size in the field of video encoding.

2.2 Local tone mapping of ROIs and bilinear interpolation

Let assume that the video-surveillance algorithm, applied on the key picture determines k interesting ROIs, and each of them is a (different) rectangle with a size (w, h) and centred in the pixel coordinate (x, y). Since the movement of the interesting area is relatively slow (for hypothesis), it is possible to define a search area for each ROI. To avoid computing an exhaustive search, the following assumption is needed (note that this assumption can be easily modified with more/less permissive parameters): an object movement is never larger than half of the rectangle size. This hypothesis is also used in the motion estimation step of video encoding. Using this hypothesis, the rectangular search area associated with the generic ROI is centred at (x, y) and has a size of $(2w, 2h)$. According to this presumption, each search area is tone mapped with a local TMO and the tracking algorithm is applied to the area. Finally, in accordance with the algorithm output, the ROIs' parameters are updated.

In order to correctly visualize this image, a bilinear interpolation is then used. The goal of the blending is to merge the local-tone mapped ROIs with the global-tone mapped image. The interpolation filter has the shape of a truncated pyramid, as shown in Figure 3. The upper base of the pyramid is a generic ROI and the lower base is its correspondent search area tone mapped with the global algorithm. Both areas are relative to the current frame and the maximum height of the pyramid is unitary.

The bilinear interpolation process can be split into nine parts, according to Figure 4. In the following formulas $I_G(x, y, c)$ is the value of the component c ($c \in \{R, G, B\}$) of the pixel (x, y) in the global tone mapped image, and $I_L x, y, c$ is the same component of the same pixel in the local tone

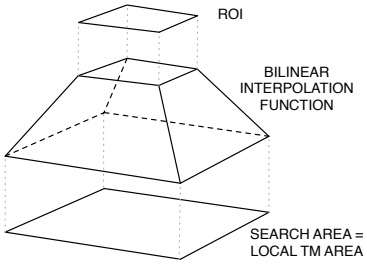

Figure 3: Truncated pyramid interpolation filter

mapped search area. dx and dy are the normalized distance between the pixel (x, y) and the center of the relative zone respectively (vertical and horizontal).

1. Corner zones (pieces A, C, I, K). The interpolated value of (x, y, c) is:[1]

$$I_C(x, y, c) = \min(dx, dy) \cdot I_G(x, y, c) + (1 - \min(dx, dy)) \cdot I_L(x, y, c) \quad (3)$$

2. ROI adjacent zones (pieces B, D,F,J). Interpolated value is:

$$I_C(x, y, c) = (1 - dx) \cdot I_G(x, y, c) + dx \cdot I_L(x, y, c) \text{ or}$$
$$I_C(x, y, c) = (1 - dy) \cdot I_G(x, y, c) + dy \cdot I_L(x, y, c) \quad (4)$$

3. ROI zone (piece E). Interpolated value is:

$$I_C(x, y, c) = I_L(x, y, c) \quad (5)$$

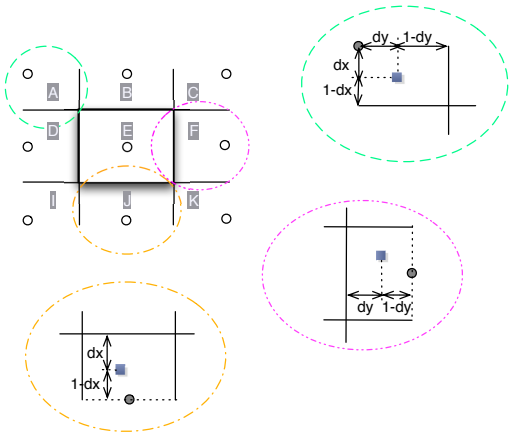

Figure 4: Bilinear interpolation in the four zones around the ROI

3. DISCUSSION: AVERAGE TIME NEEDED FOR TONE MAPPING A FRAME

The proposed algorithm fits the requirements of the video-surveillance industry, because it is as fast as possible, it does not need electronic-specific devices and it provides good visual quality images suitable for automatic analysis and for human viewing. The average computational time needed to tone map a generic frame in a HDR movie is provided by Eq.

[1]It can also be used the formula:
$I_C(x, y, c) = (1 - dx \cdot dy) \cdot I_G(x, y, c) + (dx \cdot dy) \cdot I_L(x, y, c)$

6. It is a sum of three factors: the time needed to globally tone map the whole frame plus two terms for each ROI inside it. These terms represent the time needed for the local tone mapping algorithm and the final interpolation step.

$$T_{TMO}[s] = T_{TMglobal}(W,L) + \sum_{i=0}^{K} T_{TMlocal}(W_i, L_i) + \sum_{i=0}^{K} T_{interp}(\Delta W_i, \Delta L_i) \quad (6)$$

It should be noted that, if the number of ROIs is high, it is less expensive to tone map the entire frame with a local tone map algorithm, than to apply the proposed algorithm. Therefore, the computational time formula becomes:

$$T_{optimal}[s] = \min\{T \text{ of Eq. (6)}, \quad T_{TMlocal}(W,L)\} \quad (7)$$

In Figure 5, the time needed for one real-world simulation is shown. The results are relative to the *sb-tunnel-exr*[2] movie, and the aim of the video-surveillance algorithm is to track the red car. The GOP size is set to 16, and the algorithm used for tracking is a Back-Project-Patch histogram template search. In the graph, along the Y axes, the percentage of real tone mapping time of the proposed algorithm is showed with respect to the tone mapping time needed for local tone mapping of entire frame. In correspondence to the key-frames of each GOP, the tone map time reaches the value 1, because these frames are completely tone mapped with a local algorithm. In the other frames, the computational time needed for the tone map operation is always lower (see Eq. 7).

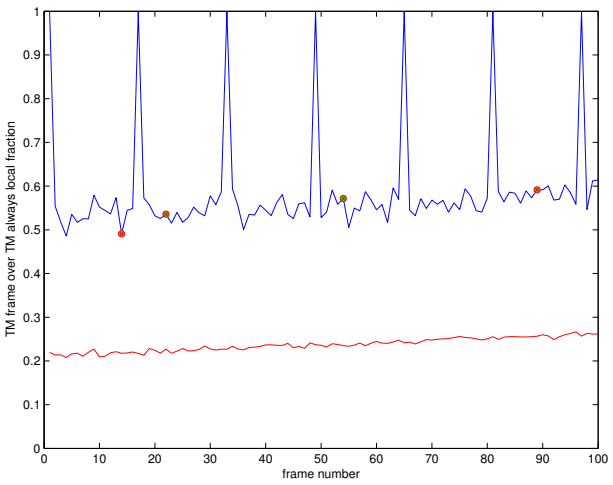

Figure 5: Percentage of time used for the tone mapping step with respect to always local tone mapping (GOP size = 16 frames). The red line represents the percentage of time used when only the global TMO is applied.

4. CONCLUSION AND FUTURE WORKS

We have proposed a new tone mapping strategy, which can be effectively used for HDR video-surveillance purposes. The method combines the benefits provided by both local and global TMOs, thus providing a fast and good quality LDR image. Future work will concentrate on increasing the speed of the TMO algorithm, using the motion vector field of the image to reduce the search areas of the possible ROIs and to place them in the best possible space position. By reducing the ROI size, it will be possible to further decrease the local tone map time, and therefore speed up the process. Nevertheless a future evolution of this work will try to merge the tone mapping process with the detection/tracking task in order to take advantage of the increased information carried by HDR images.

[2] "sb-tunnel-exr" by Grzegorz Krawczyk ©2004

REFERENCES

[1] Erik Reinhard, Greg Ward, Sumanta Pattanaik, and Paul Debevec, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics)*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.

[2] Rafal Mantiuk and Grzegorz Krawczyk, "Pfstools, http://www.mpi-sb.mpg.de/resources/pfstools/," 2005.

[3] Patrick Ledda, Alan Chalmers, Tom Troscianko, and Helge Seetzen, "Evaluation of tone mapping operators using a high dynamic range display," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 640–648, 2005.

[4] Grzegorz Krawczyk, Karol Myszkowski, and Hans-Peter Seidel, "Perceptual effects in real-time tone mapping," in *SCCG '05: Proceedings of the 21st spring conference on Computer graphics*, New York, NY, USA, 2005, pp. 195–202, ACM.

[5] Patrick Ledda, Luis Paulo Santos, and Alan Chalmers, "A local model of eye adaptation for high dynamic range images," in *AFRIGRAPH '04: Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, New York, NY, USA, 2004, pp. 151–160, ACM.

[6] F. Hassan and J. E. Carletta, "A real-time fpga-based architecture for a reinhard-like tone mapping operator," in *GH '07: Proceedings of the 22nd ACM SIGGRAPH/EUROGRAPHICS symposium on Graphics hardware*, Aire-la-Ville, Switzerland, Switzerland, 2007, pp. 65–71, Eurographics Association.

[7] Firas Hassan and Joan Carletta, "A high throughput encoder for high dynamic range images," in *Proceedings of the International Conference on Image Processing, ICIP 2007, September 16-19, San Antonio, Texas, USA*. 2007, pp. 213–216, IEEE.

[8] Tsun-Hsien Wang, Wei-Su Wong, Fang-Chu Chen, and Ching-Te Chiu, "Design and implementation of a real-time global tone mapping processor for high dynamic range video," in *Proceedings of the International Conference on Image Processing, ICIP 2007, September 16-19, San Antonio, Texas, USA*. 2007, pp. 209–212, IEEE.