IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 4, 2021, accepted March 7, 2021, date of publication March 12, 2021, date of current version March 25, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3065817

Highlights Analysis System (HAnS) for Low Dynamic Range to High Dynamic Range Conversion of Cinematic Low Dynamic Range Content

ANA STOJKOVIC[1], JAN AELTERMAN[1], HIEP LUONG[1], HANS VAN PARYS[2,3], AND WILFRIED PHILIPS[1]

[1]IPI-imec Research Group, Department of Telecommunications and Information Processing (TELIN), Faculty of Engineering and Architecture, Ghent University, 9000 Ghent, Belgium
[2]Televic Conference, 8870 Izegem, Belgium
[3]TP Vision, 9052 Ghent, Belgium

Corresponding author: Ana Stojkovic (ana.stojkovic@ugent.be)

The work of Jan Aelterman was supported by the Ghent University Postdoctoral Fellowship under Grant BOF15/PDO/003.

ABSTRACT We propose a novel and efficient algorithm for detection of specular reflections and light sources (highlights) in cinematic content. The detection of highlights is important for reconstructing them properly in the conversion of the low dynamic range (LDR) to high dynamic range (HDR) content. Highlights are often difficult to be distinguished from bright diffuse surfaces, due to their brightness being reduced in the conventional LDR content production. Moreover, the cinematic LDR content is subject to the artistic use of effects that change the apparent brightness of certain image regions (e.g. limiting depth of field, grading, complex multi-lighting setup, etc.). To ensure the robustness of highlights detection to these effects, the proposed algorithm goes beyond considering only absolute brightness and considers five different features. These features are: the size of the highlight relative to the size of the surrounding image structures, the relative contrast in the surrounding of the highlight, its absolute brightness expressed through the luminance (luma feature), through the saturation in the color space (maxRGB feature) and through the saturation in white (minRGB feature). We evaluate the algorithm on two different image data-sets. The first one is a publicly available LDR image data-set without cinematic content, which allows comparison to the broader State of the art. Additionally, for the evaluation on cinematic content, we create an image data-set consisted of manually annotated cinematic frames and real-world images. For the purpose of demonstrating the proposed highlights detection algorithm in a complete LDR-to-HDR conversion pipeline, we additionally propose a simple inverse-tone-mapping algorithm. The experimental analysis shows that the proposed approach outperforms conventional highlights detection algorithms on both image data-sets, achieves high quality reconstruction of the HDR content and is suited for use in LDR-to-HDR conversion.

INDEX TERMS Cinematic content, high dynamic range, highlights detection, inverse tone mapping, low dynamic to high dynamic range (LDR-to-HDR) conversion, specular reflections.

I. INTRODUCTION

Nowadays, the huge demand of high quality multimedia content production, influenced by the rapid rise of the high dynamic range display technology, creates a big competitiveness on the market for devices capable for capturing and displaying high quality multimedia content (in high resolution and in the high dynamic range (HDR)). Displaying such content on the newest generations of TV screens is not considered a problem. However, there still exist a huge amount of legacy video content, that is stored in low dynamic range (LDR). In order to display this legacy content on HDR screens, it should be properly converted. This conversion implies increasing the contrast, by boosting the brightest parts

The associate editor coordinating the review of this manuscript and approving it for publication was Ananya Sen Gupta.

of the scene that are commonly considered as highlights. For that reason, in order to achieve realistic effect (HDR effect) of displaying the scene in its high dynamic range, the highlights should be correctly detected in the low dynamic range (LDR) content and then appropriately boosted. Therefore, the standard LDR-to-HDR conversion pipeline combines the two types of techniques: highlights detection (HLD) and inverse tone mapping (ITM), as it is presented in [1].

The early works [2]–[5] in the field of LDR-to-HDR conversion of image and video content, employ classical approaches for detecting the brightest regions, expanding the luminance with a pixel-wise ITM function and increasing the contrast. In the same time, approaches for performing a psycho-visual evaluation analysis [6] were proposed. References [7]–[9] present a good overview of the field of LDR-to-HDR. The recent works [10]–[18] are more oriented towards making the inverse tone mapping process more automatic, robust, implemented in real time, applicable to wide range of light and contrast conditions, while aiming to preserve the color appearance and details visibility. Some of these approaches use more advanced classical techniques [10], [15], [16], [18], while some are deep learning based approaches [12]–[14] and rely on the training data.

In our approach we focus on obtaining the HDR effect, by correctly detecting and selectively, but gradually/continuously boosting the highlights in low dynamic range cinematic content. The highlights detection for the purposes of LDR-to-HDR conversion is commonly performed by threshold-based techniques [4], [11], [19]. The reason for this is the assumption that highlights are the brightest parts in the scene. Another class of techniques (see [20]–[29]), relies on using the Shafer's dichromatic model introduced in [20], [21]. The latter techniques are known as specular highlights removal techniques that separate the specular reflections from the diffuse reflections. A thorough survey on these techniques is given in [30]. However, both types of techniques fail in terms of correctly detecting highlights in cinematic LDR content. This is because, in the cinematic content (which is subject of artistic arrangements), the specular reflections and the light sources, which we will refer as ''highlights'', are not necessarily the brightest parts of the scene. We use the term ''cinematic content'' to refer to content that is subject to artistic arrangements and artistic use of effects, such as use of complex multi-lighting setup with lights of diverse colors and intensities, use of limited depth of field during the acquisition process and performing artistic grading (post-processing on the recorded content, which is done by professionals in the movie industry, in order to meet the requirements of movie directors and producers).

In [1], we addressed the difficulty of detecting highlights in cinematic LDR content, we introduced two features as inherent features of the highlights, which improve the highlights detection and overcome the disadvantages of the aforementioned HLD algorithms.

A. STRUCTURE

In Section II, by addressing the problem of highlights detection in cinematic LDR content, we first focus on defining highlights in the LDR content. In Section III, we first give insight into the goal of the devised highlights detection algorithm (which we refer as ''Highlights Analysis System'', abbreviated as HAnS) and then we proceed with describing it in more details. Later, in Section IV, we propose an approach for inverse-tone mapping, in which HAnS is integrated. In Section V, we present the experimental analysis through which we evaluate HAnS and the proposed ITM and discuss the results. In the final Section VI, we summarize the results from the experimental analysis into a conclusion about the advanteougness and the efficiency of the proposed algorithm for detecting highlights in LDR content in general (not only in cinematic LDR content), about its suitability for use in inverse tone-mapping, with an emphasis on the correct detection and boosting of the highlights.

II. HIGHLIGHTS IN CINEMATIC CONTENT

The image content consists of highlight and non-highlight areas. In the group of highlight areas we classify areas belonging to either light sources or specular reflections, while in the group of non-highlights areas we classify areas belonging to light reflections from diffuse surfaces.

In Section II-A, we define highlights as they exist in the real world and in the context of cinematic LDR content production. In Section II-B, we proceed with addressing the problems for highlights detection in cinematic LDR content, while in Section II-C we present the features that we use for highlights detection in HAnS.

A. HIGHLIGHTS DEFINITION

In the real world, lights sources represent origins of electromagnetic radiation to which the human visual system is sensitive. Depending on the spectral distribution of the emitted light, light sources can produce light in different colors. In regards of cinematic LDR content production, when light sources are part of the cinematic scene, most often they belong to small parts of the scene with a purpose of producing artistic effects and in the same time to avoid big interfering with the main context of the scene.

From a physical aspect of defining specular reflections, they represent direct reflections of light from shiny surfaces and unlike the light reflections from diffuse surfaces (i.e. surfaces with Lambertian reflectance properties), they appear as most pronounced when being observed from one certain direction. Specular reflections occur as small, bright spots on illuminated shiny objects. The color of the specular reflections is dependent on the color of the light that is illuminating the surface and the color of the surface from which the light is reflected. In other words, the color of the reflected light highly depends on the spectral distribution of the incident light and the reflectance properties of the surface that is being illuminated. Specular reflections cover surfaces relatively small to the surfaces of the illuminated objects on

which they occur. Exceptions of this definition are objects with flat mirror-like surfaces. The shape of the specular reflections varies depending: on the shape of the illuminated surface, on the reflectance properties of the surface, on the texture of the surface, on the shape and the position of the light sources.

In Fig.1 we present an illustration for occurrence of specular reflections on a glossy surface illuminated by single light source and from one direction. The light reflected from the glossy surface consists of specular and diffuse parts. The both parts (diffuse and specular) on Fig.1 are presented with different cross section lines: the green line and the red line. In the LDR content production, due to the limitations of the camera sensitivity to the high dynamic range of the light, the intensity of the specular reflections is being either clipped or compressed.

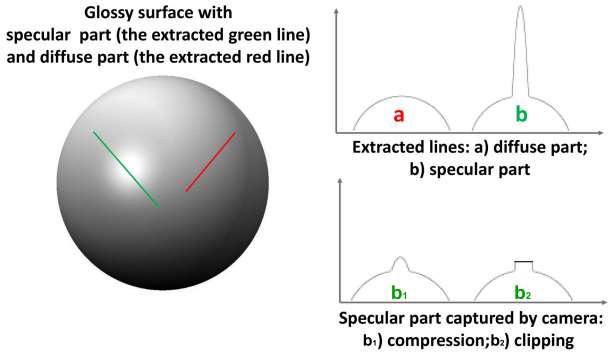


FIGURE 1. Illustration of a glossy surface with specular and diffuse reflections in two cases: 1) in the scene before it is captured with the camera (graph on the top right) and 2) in digital form after the scene is captured with the camera (graph on the bottom right) in LDR. In the LDR content, highlights are either compressed or clipped. The illustration is a modified version of the illustration presented in [4].

In Fig.2 we present a frame extracted from the cinematic LDR content, which is property of the National Public-Service Broadcaster for the Flemish Region and Community of Belgium, VRT [51], with marked areas for the specular reflections and the light sources.

B. DIFFICULTIES FOR HIGHLIGHTS DETECTION IN CINEMATIC CONTENT

In the dynamic range of the original scene, highlights (specular reflections and light sources) occur several degrees of intensity brighter than the light reflections from diffuse (Lambertian) surfaces, (see [31]). There are multiple reasons why highlights are difficult to be detected in the cinematic LDR content. In what follows we present a list of the most important ones, which are described in details in [1].

- The dynamic range of the highlights in the LDR representation of the scene, is either clipped or compressed. In that way, a color preserving and non-color preserving LDR content is produced, respectively as illustrated in Fig.3 (to which, in more details, we refer in Section II-C).
- Highlights that are out of focus, due to the camera settings, appear blurred and low in brightness in the LDR content.
- Complex lighting setup which results in multiple lighting sources, diverse in color and intensity.
- The cinematic content is subject to artistic arrangements, through which the brightness ratios in different parts of the scene are differently adjusted in order to achieve an artistic effect.

Therefore, by only applying threshold-based HLD algorithms, there is a high risk that light reflections from bright diffuse surfaces will be also detected. Consequently, when inverse tone mapping is applied, the incorrectly detected areas will be boosted and in the reconstructed HDR content they will appear as light-emitting areas (making the reconstructed HDR content unpleasant and unnatural to watch [1]). Considering these issues, highlights detection, especially in cinematic LDR content, is considered to be non-trivial and challenging. Therefore, we propose an algorithm that combines the features presented in Section II-C.

C. FEATURES FOR HIGHLIGHTS DETECTION

The goal of HAnS is the detection of highlights in cinematic LDR content, that are bright, that cover areas relatively small to the surrounding image structures and are distinguished from their vicinity.

In [1], we introduced two inherent features of specular reflections that improve the detection of specular reflections. For highlights detection in HAnS, we use the same features, assuming that content-wise the light sources are subject to same processes of cinematic LDR content production as the specular reflections. We additionally use three features related to the brightness, the color and the saturation of the highlights. These three features we refer to as ''pixel-wise features''. In what follows, we present a list of all features that are used in HAnS.

LIST OF FEATURES

- The local contrast feature [1]: defined as the relative contrast in the surrounding of the highlight (reasoning that the area of the actual highlight is brighter than its surrounding which belongs to the light reflections from diffuse surfaces).
- The relative size feature [1]: the size of the area belonging to the highlight relative to the size of the area of the surrounding image structures (reasoning that specular reflections and light sources in cinematic LDR content commonly cover areas that are smaller than the areas belonging to light reflections from diffuse surfaces).
- The pixel-wise features:
  - The minRGB feature: since the lighting conditions in most scenes are achromatic, the highlights are more likely to be less color saturated. In the recent literature for specular reflection separation [19], [28], many algorithms use the dark-channel prior, which we will refer to as ''minRGB'' feature.

FIGURE 2. Frame extracted from cinematic LDR content, which is property of the National Public-Service Broadcaster for the Flemish Region and Community of Belgium VRT [51]. The scene consists of: light sources (marked with blue), specular reflections (marked with green) and light reflections from diffuse surfaces (left unmarked). We consider the halo around the lamp as highlight because physically it originates from small specular reflections on dust particles and humid in the air.

According to the description of the dark channel, presented in [28], a diffuse pixel of natural images is expected to have a very low intensity in at least one color channel. In other words, pixels belonging to the white region of a clipped highlight in the LDR content, will have high values in all color channels. Therefore the dark channel is used as a feature that gives the level of presence of white in the tristimulus representation of the scene and hence, it is suitable for detecting achromatic specular reflections and localizing highlights that are saturated in the full dynamic range of the light due to clipping (see Fig.3).

- The maxRGB feature: in the real world, there are exceptional cases in which the highlights are color saturated. Examples for these cases are colored man-made light sources and specular reflections from surfaces illuminated with these light sources. The feature that is related with the high saturation in color will easily expose the colored highlights, [34] and [35]. This feature is known as the maximum RGB feature, which we will refer to as "maxRGB" feature. It considers a high intensity in at least one color channel. According to the max-RGB hypothesis presented in [33], based on earlier work of Van de Weijer ([34] and [35]) where the maxRGB feature was originally introduced,

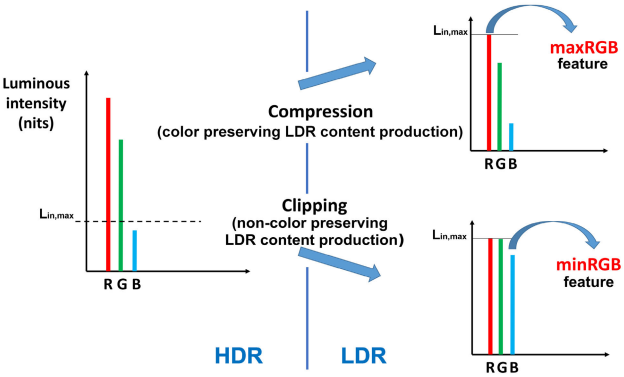


FIGURE 3. Illustration of the LDR content production from HDR scene, with compression and clipping (to the maximal luminous intensity of the LDR screen, to which later we will refer as $L_{in,max}$). The LDR content production in the both cases (clipping and compression), is associated with two pixel-wise features that we use in our highlights detection algorithm and are listed in Section II-C. The color-preserving LDR content production (compression of the highlights) is associated with the maxRGB feature, while the non-color preserving LDR content production (clipping of the highlights) is associated with the minRGB feature.

a surface with a perfect reflectance property will reflect the full range of light colors it captures, when being illuminated by single light source. Hence, the reflected color from the surface is the actual color of the light source. We use the maxRGB image feature for localizing specular reflections and

light sources that are color saturated due to compression (see Fig.3).
- The luma feature: in order to account for the luminance of the highlights in the LDR content, we use the luma as an additional feature. The luma feature, as it will be noticed from the following example presented in Fig.4, is a trade-off between the minRGB and the maxRGB features and it is related with the actual luminance of the highlights.

In Fig.4, we present a night scene abounding with artificial lights and specular reflections. Here, we additionally present the three pixel-wise features (minRGB, luma and maxRGB) extracted from the RGB LDR image. By observing the images in Fig.4, we can notice that the colored light sources are partially or not localized with the minRGB feature, while they are well localized with the maxRGB feature. On the other hand, the achromatic highlights are better localized with the minRGB feature. Since the two features (minRGB and maxRGB) noticeably differ in the way of exposing the highlights, the luma feature balances between them. Considering that the highlights are differently exposed with each of these three pixel-wise features, by using each of them separately (rather than using them in a combination or using only one), most of the highlight candidates will be well localized and hence, easier for detection.

FIGURE 4. Night scene with artificial lights and specular reflections: 1) RGB LDR image, 2) minRGB features; 3) luma features, 4) maxRGB features. The maxRGB feature is advantageous in localizing highlights that are color saturated, while the minRGB feature is advantageous in localizing highlights that are saturated in the full dynamic range of the light. The luma feature accounts for the actual luminance of the highlights and balances between the maxRGB feature and the minRGB feature.

III. PROPOSED ALGORITHM FOR HIGHLIGHTS DETECTION (HANS)

In Section III-A, we first focus on stating the objective of HAnS. Later, in Section III-B we describe each component of HAnS in detail and discuss the objective for highlights detection. In Section III-C, through visual analysis of the outputs of the essential HAnS procedure, we discuss the performance of HAnS in terms of overcoming the difficulties for highlights detection in cinematic LDR content.

A. OBJECTIVE

Starting from the assumption that in the cinematic LDR content, areas belonging to highlights are brighter and smaller than their surrounding areas that belong to non-highlights (light reflections from diffuse surfaces), we define the highlights to be dependent:
- on the size of the area of the highlight relative to the size of the surrounding non-highlight areas;
- on the relative contrast between the highlight area and the surrounding non-highlight areas;
- on the absolute per-pixel intensities of the highlight areas.

If we express the previous statement mathematically, the objective of the proposed highlights detection algorithm is obtaining a per-pixel output value for the detected highlight, which would be proportional to $\frac{(k-w)^+}{k} \cdot (b-a)^+ \cdot b$, where k is the size of the surrounding non-highlight area, w is the size of the highlight area, b is the absolute intensity of the highlight area and a is the absolute intensity of the surrounding non-highlight area. The operator $(.)^+$, only accounts for the positive values of the considered difference, while the negative values are set to 0. In other words, in the way it is used, it ensures that the conditions $w < k$ and $a < b$, are met. The term $\frac{(k-w)^+}{k}$ is associated with the relative size feature. The term $(b-a)^+$ is associated with the local contrast feature. With these two terms, structures surrounding the highlight which are brighter and bigger in size than the highlight, will be suppressed. These structures commonly belong to non-highlight (light reflections from diffuse surfaces) areas. With the term b the absolute intensity of the highlight is taken into consideration. Following this formulation, we distinguish three general cases for the highlight candidate:
- the highlight candidate is very small in size, $w \ll k$, so the output becomes approximately $(b-a)^+ \cdot b$;
- the highlight candidate is smaller in size than the surrounding structures, $w < k$, so the output becomes $\frac{(k-w)^+}{k} \cdot (b-a)^+ \cdot b$;

- the highlight candidate is indistinguishable from the surrounding or it can't be considered as highlight according to the definition for highlights in cinematic LDR content. If at least one of the conditions: $w \geq k$ and $a \geq b$, is satisfied, the output is 0.

We design the algorithm in a way that it extracts the relative size feature and the local contrast feature in subsequent procedures applied identically on each pixel-wise feature (minRGB, maxRGB and luma). For the output of HAnS, we refer as to ''Highlights Map''. The different pixel intensity values in the Highlights Map depend proportionally on the strength (expressed through the relative contrast, the relative size and the absolute intensity) of the detected highlights areas.

B. THE IMPLEMENTATION OF HANS

In order to enable implementation of HAnS on a fast executing system, such as the Field-Programmable Gate Array (FPGA) integrated circuit and to enable easy integration within simple inverse tone-mapping approach (described in Section IV), in the design of HAnS we are limited to use simple image processing techniques. These techniques consider streaming low-pass filters (i.e. moving average filter), mathematical morphology operations (dilation and erosion) and look-up tables (LUTs) for the more complex per-pixel mathematical operations (exponential and square functions). The block diagram for HAnS is shown in Fig.5.

The algorithm is devised in three parts (see Fig.5 (1)):

- pixel-wise feature extraction (explained in Section III-B1);
- main procedure (explained in Section III-B2) for highlights detection (applied on each extracted pixel-wise feature separately) to which we refer as ''main HLD procedure'';
- fusion (explained in Section III-B3) of the separate outputs (to which we refer as ''Feature Highlights Maps'' and denote each of them with $\mathbf{M}_f$, where $f \in \{\text{minRGB, luma, maxRGB}\}$) of the main HLD procedure.

In what follows we focus on describing each procedure of HAnS in details.

1) PIXEL-WISE FEATURE EXTRACTION

The pixel-wise features (minRGB, luma and maxRGB) are calculated by (1), (2) and (3). They are extracted in a pixel-wise manner from the LDR RGB input image, with the color component $\mathbf{I}_c(\mathbf{x})$, where $c \in \{r, g, b\}$ is the color component label and $\mathbf{x} = [x_1, x_2]$ is the vector of the image coordinates.

$$\mathbf{I}_{\text{minRGB}}(\mathbf{x}) = \min\left\{\mathbf{I}_r(\mathbf{x}), \mathbf{I}_g(\mathbf{x}), \mathbf{I}_b(\mathbf{x})\right\} \tag{1}$$

$$\mathbf{I}_{\text{luma}}(\mathbf{x}) = 0.2989 \cdot \mathbf{I}_r(\mathbf{x}) + 0.5870 \cdot \mathbf{I}_g(\mathbf{x}) + 0.1140 \cdot \mathbf{I}_b(\mathbf{x}) \tag{2}$$

$$\mathbf{I}_{\text{maxRGB}}(\mathbf{x}) = \max\left\{\mathbf{I}_r(\mathbf{x}), \mathbf{I}_g(\mathbf{x}), \mathbf{I}_b(\mathbf{x})\right\} \tag{3}$$

Hereafter, we use $\mathbf{I}_f$ to denote the extracted pixel-wise feature, where $f \in \{\text{minRGB, luma, maxRGB}\}$ refers to the label of the extracted pixel-wise feature.

2) MAIN HLD PROCEDURE

The main HLD procedure together with the inputs is presented in Fig.5 (2). The main procedure is applied on each extracted pixel-wise feature, $\mathbf{I}_f$, in the same way.

The main HLD procedure is divided into three parts:

- extraction of the relative size feature (step B and step C)
- extraction of the local contrast feature (step D and step E)
- adaptation with the absolute intensities from the input pixel-wise feature (step F).

a: INPUTS OF THE MAIN HLD PROCEDURE (STEP A)

The inputs of the main HLD procedure (see Fig.5 (2).A) are the extracted pixel-wise feature $\mathbf{I}_f(\mathbf{x})$ and the size parameter k. As it is described in [1], the size parameter imposes a soft upper limit on the size feature to exclude bright areas that would be too large and likely to belong to light reflections from diffuse surfaces. In HAnS, the size parameter represents the size of a 2D moving average filter with boundaries that are supposed to cover the full areas of the desired highlights. The moving average filter of size k x k, is denoted with $\mathbf{H}$.

b: EXTRACTION OF THE RELATIVE SIZE FEATURE (STEP B AND STEP C)

In the first part of the main HLD procedure the extracted pixel-wise feature is first processed with a low-pass moving average filter (see Fig.5 (2).B). The output of this step, $\mathbf{I}_{f,lp}(\mathbf{x})$, is calculated with (4).

$$\mathbf{I}_{f,lp}(\mathbf{x}) = (\mathbf{H} * \mathbf{I}_f)(\mathbf{x}) \tag{4}$$

Then, on the output from the step B we apply mathematical dilation (see Fig.5 (2).C), with a structural element of the same size as the moving average filter. The output from step C, $\mathbf{I}_{f,lpd}(\mathbf{x})$, is calculated with (5),

$$\mathbf{I}_{f,lpd}(\mathbf{x}) = \max_{\mathbf{y} \in \alpha(\mathbf{x})} \left(\mathbf{I}_{f,lp}(\mathbf{y})\right) \tag{5}$$

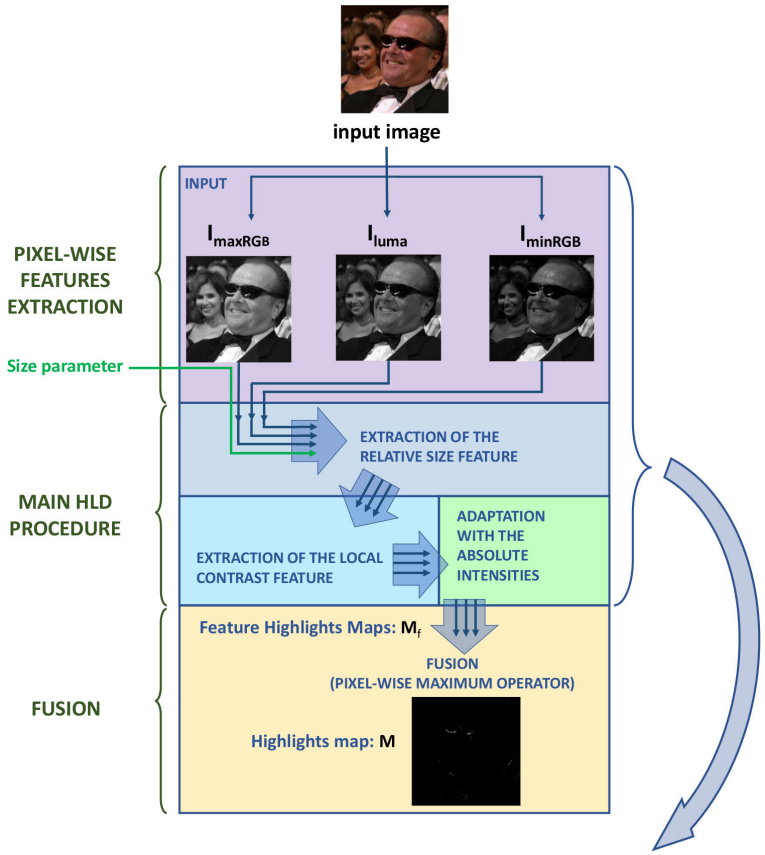
where $\alpha(\mathbf{x})$ is area (of size k defined with the structural element) in the vicinity of $\mathbf{x}$. The reason for using mathematical dilation, is to enable suppression of relatively large in size bright areas that commonly belong to non-highlights.

c: LOCAL CONTRAST FEATURE EXTRACTION (STEP D AND STEP E)

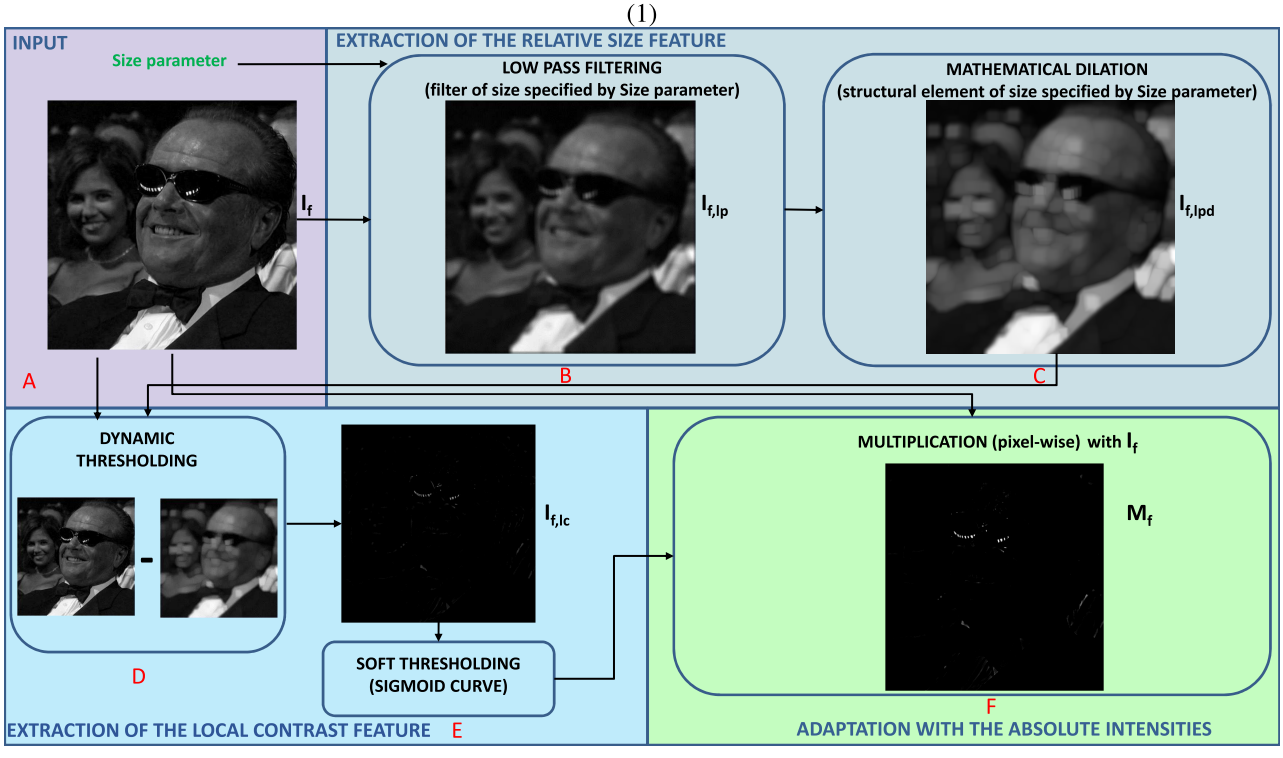
Here, we describe the extraction of the local contrast feature $\mathbf{I}_{f,lc}(\mathbf{x})$. For that purpose we use dynamic thresholding (step D) followed by soft thresholding (step E) for normalisation in range [0-1].

The applied dynamic thresholding is mathematically expressed by (6).

$$\mathbf{I}_{f,lc}(\mathbf{x}) = \left(\mathbf{I}_f(\mathbf{x}) - \mathbf{I}_{f,lpd}(\mathbf{x})\right)^+ \tag{6}$$



(1)



(2)

FIGURE 5. (1) Block diagram for HAnS, splitted into three parts: pixel-wise features extraction; highlights detection (main HLD procedure) applied on each extracted pixel-wise feature (minRGB, maxRGB, luma) separately; fusion of the outputs (Highlights Maps: M_f) obtained with the main HLD procedure; (2) Main HLD procedure with the inputs, here presented for one pixel-wise feature.

With (6), we consider the difference between $\mathbf{I}_{\mathrm{f}}(\mathbf{x})$ and $\mathbf{I}_{\mathrm{f,lpd}}(\mathbf{x})$, from which only the positive values are taken into account.

The simple implementation of the dynamic thresholding is closely related to certain algorithms for morphological gray-scale reconstruction [41]. Compared to these techniques, the proposed dynamic thresholding has the advantage of not being iterative and hence, being time efficient.

In order to distinguish strong highlights from weak highlights (where the strength is expressed through the relative size feature and the local contrast feature) and to normalize the output obtained in step D to always be in range $[0-1]$, in step E we perform soft thresholding with a sigmoid curve. The output of the soft-thresholding step, $g\left(\mathbf{I}_{\mathrm{f,lc}}(\mathbf{x})\right)$, is calculated by (7).

$$g\left(\mathbf{I}_{\mathrm{f,lc}}(\mathbf{x})\right) = \frac{1}{1+e^{-\mathrm{p}\cdot(\mathbf{I}_{\mathrm{f,lc}}(\mathbf{x})-\mathrm{t})}} \tag{7}$$

In (7), p is a parameter that controls the steepness of the sigmoid function, while t is the soft threshold.

We empirically set the value of the parameter p to be equal to 20 (i.e. $\mathrm{p}=20$), so that the steepness of the sigmoid curve is high enough and in the same time the continuous rise is preserved (which is different from hard-thresholding that is applied with a step function). The soft threshold is in range $[0-1]$ when the input RGB LDR image is normalized in the same range and it is in range $[0-255]$ when the input RGB LDR image is not normalized. With the soft threshold t, associated with the "local contrast threshold" described in [1], we control the allowed relative difference between the pronounced areas and their local surrounding. In [1] it was shown that closer this value to zero is, the more sensitive the detector becomes to the local contrast feature. In most of the experiments we performed (with the input RGB LDR image normalized in range $[0-1]$), the value of the soft threshold was set empirically to either 0.2 or 0.3. However, this value is not fixed and we consider it as one degree of freedom of HAnS. The second parameter that we consider as a degree of freedom of HAnS is the size parameter k.

d: ADAPTATION WITH THE ABSOLUTE INTENSITIES (STEP F)

In the final step F, we adapt the output from step E, with the absolute intensities from the pixel-wise feature (see Fig.5 (2).F).

The Feature Highlights Map, $\mathbf{M}_{\mathrm{f}}(\mathbf{x})$, as output of the main HLD procedure, is calculated by (8), where $\mathrm{f}\in\{\mathrm{minRGB}, \mathrm{luma}, \mathrm{maxRGB}\}$.

$$\mathbf{M}_{\mathrm{f}}(\mathbf{x}) = \mathbf{I}_{\mathrm{f}}(\mathbf{x}) \cdot g\left(\mathbf{I}_{\mathrm{f,lc}}(\mathbf{x})\right) \tag{8}$$

In (8), with $g\left(\mathbf{I}_{\mathrm{f,lc}}(\mathbf{x})\right)$, we denote the output of the soft thresholding step (step E, see also Fig.5 (2).E), obtained with a sigmoid function and presented with (7).

e: ACCOMPLISHING THE OBJECTIVE FOR HIGHLIGHTS DETECTION

Here, we discuss how the objective for highlights detection presented in Section III-A is accomplished with the main HLD procedure from HAnS. To simplify the explanation, we stick to the one-dimensional case.

We associate the input pixel-wise feature $\mathbf{I}_{\mathrm{f}}(\mathbf{x})$ (with $\mathrm{f}\in\{\mathrm{minRGB}, \mathrm{luma}, \mathrm{maxRGB}\}$) from the main HLD procedure with one-dimensional square-shaped signal, where by following the description of the objective presented in Section III-A, to the square impulse we associate the highlight area and to the surrounding flat area of the square impulse we associate the non-highlight area. Referring to step A of the main HLD procedure, we associate the size parameter (i.e. the size of the moving average filter) to the value of k described in Section III-A, which is related to the size of the non-highlight areas. The size of the actual highlight area (i.e. width of the square impulse in the one-dimensional signal) remains to be w.

Referring to step B, with the performed moving-average filtering, a maximum weight for the highlight area, equal to $\frac{(\mathrm{k}-\mathrm{w})\cdot\mathrm{a}+\mathrm{w}\cdot\mathrm{b}}{\mathrm{k}}$ is extracted, where b is the intensity of the square impulse and a is the intensity of the flat area surrounding the square impulse.

With the mathematical dilation performed in step C, the area of the filtered square impulse is dilated to area of size $\mathrm{w}+2\cdot\mathrm{k}$. In that way, it is ensured that by performing step C (the mathematical dilation) the maximal value $\frac{(\mathrm{k}-\mathrm{w})\cdot\mathrm{a}+\mathrm{w}\cdot\mathrm{b}}{\mathrm{k}}$ of the filtered signal (the output of step B) will stay constant within the full area of the square impulse. In other words, with the mathematical dilation the existence of the term $(\mathrm{k}-\mathrm{w})^{+}$ instead of $(\mathrm{k}-\mathrm{w})$, is ensured.

With the dynamic thresholding performed in the step D, the filtered output of step C is subtracted from the input one-dimensional square-shaped signal and only the positive values are taken into account. After the performed subtraction, the output of the step D is one-dimensional signal of constant intensity $\frac{(\mathrm{k}-\mathrm{w})^{+}}{\mathrm{k}}\cdot(\mathrm{b}-\mathrm{a})^{+}$ in the area where the original square impulse exists (i.e. in the area of size w) and 0 in its vicinity. Beside for normalisation we use the soft thresholding for giving higher priority to smaller and at the same time pronounced structures.

With the final step F of the main HLD procedure, multiplication with b (i.e. the intensity of the square impulse) is ensured.

In Fig.6, we visually present the outputs from the B-D steps of the main HLD procedure, when the input is one-dimensional square-shaped signal (with the square impulse associated to the highlight (case 1 and case 2) and when it is not a highlight, i.e. it is considered to belong to light reflections from bright surfaces (case 3 and case 4). Note that in every of the represented cases, by extracting the relative size feature and the local contrast feature, we succeed to

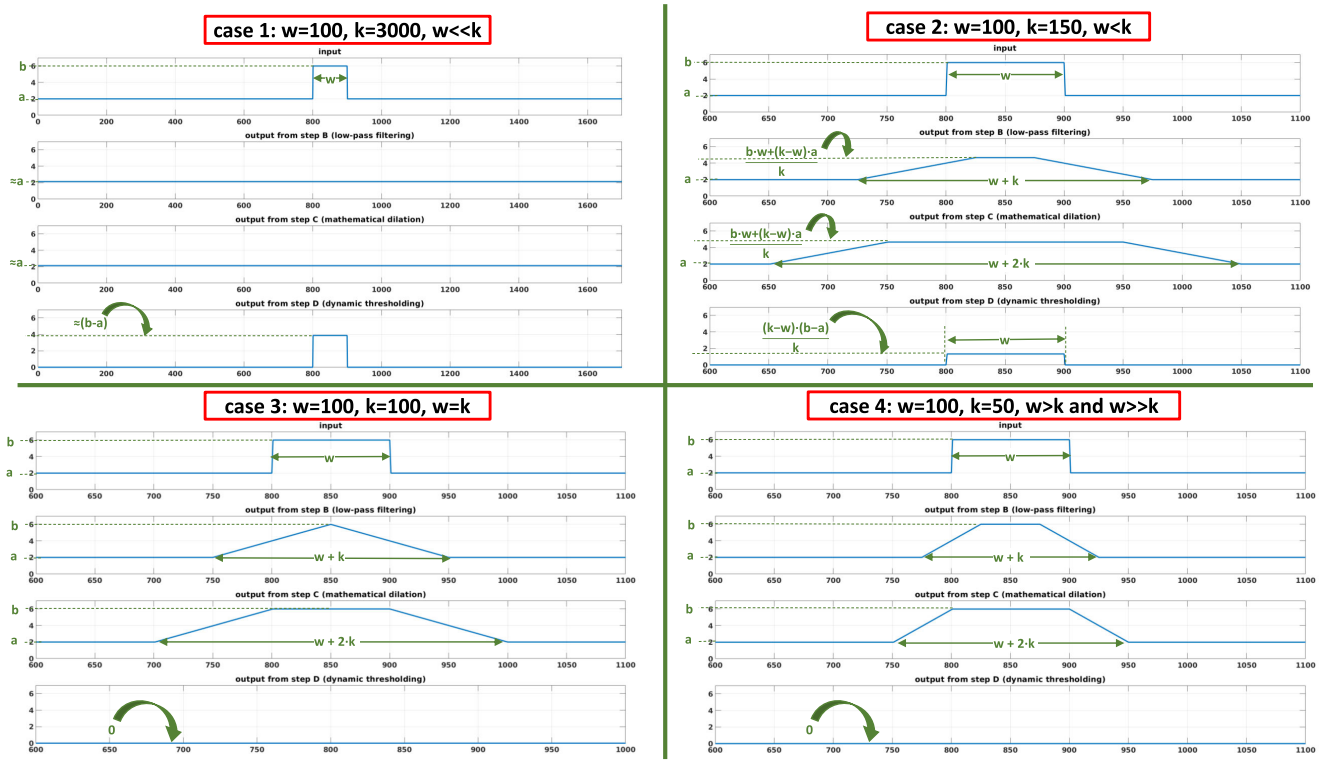


FIGURE 6. Accomplishing the goal for highlights detection in cinematic LDR content. The input signal is represented as one-dimensional square-shaped signal. We analyze 4 cases: $w \ll k$, $w < k$, $w = k$ and $w > k$ for which its subcase ($w \gg k$) results in the same 0 output. For each of the analyzed cases, we show the outputs from B-D steps of the main HLD procedure. For better visibility, we direct the reader to the electronic version of the paper.

accomplish the initial objective for highlights detection (see Section III-A) given with the term $\frac{(k-w)^+}{k} \cdot (b-a)^+ \cdot b$.

With the relative size feature extraction, the local contrast feature extraction, and the adaptation with the absolute intensities, with HAnS we succeed to extract image structures that are small (relatively to the surrounding area), distinguished from their vicinity and bright.

3) FUSION

In the final part of HAnS, (see Fig.5 (1)) we combine the three Feature Highlights Maps obtained from each pixel-wise feature separately.

$$\mathbf{M}(\mathbf{x}) = \max\{\mathbf{M}_{\text{minRGB}}(\mathbf{x}), \mathbf{M}_{\text{luma}}(\mathbf{x}), \mathbf{M}_{\text{maxRGB}}(\mathbf{x})\} \quad (9)$$

For obtaining the Highlights Map, $\mathbf{M}(\mathbf{x})$, we use the pixel-wise maximum operator, as it is presented with (9). We use the pixel-wise maximum operator in order to consider all detections obtained for each of the three pixel-wise features (minRGB, maxRGB and luma), without making biased decisions towards some of them,

The obtained Highlights Map, can be further used in combination with any type of pixel-wise inverse-tone mapping function for the purpose of LDR-to-HDR conversion. Therefore, in Section IV we present a simple inverse-tone mapping approach in which HAnS is integrated.

C. OVERCOMING THE DIFFICULTIES FOR HLD IN CINEMATIC LDR CONTENT

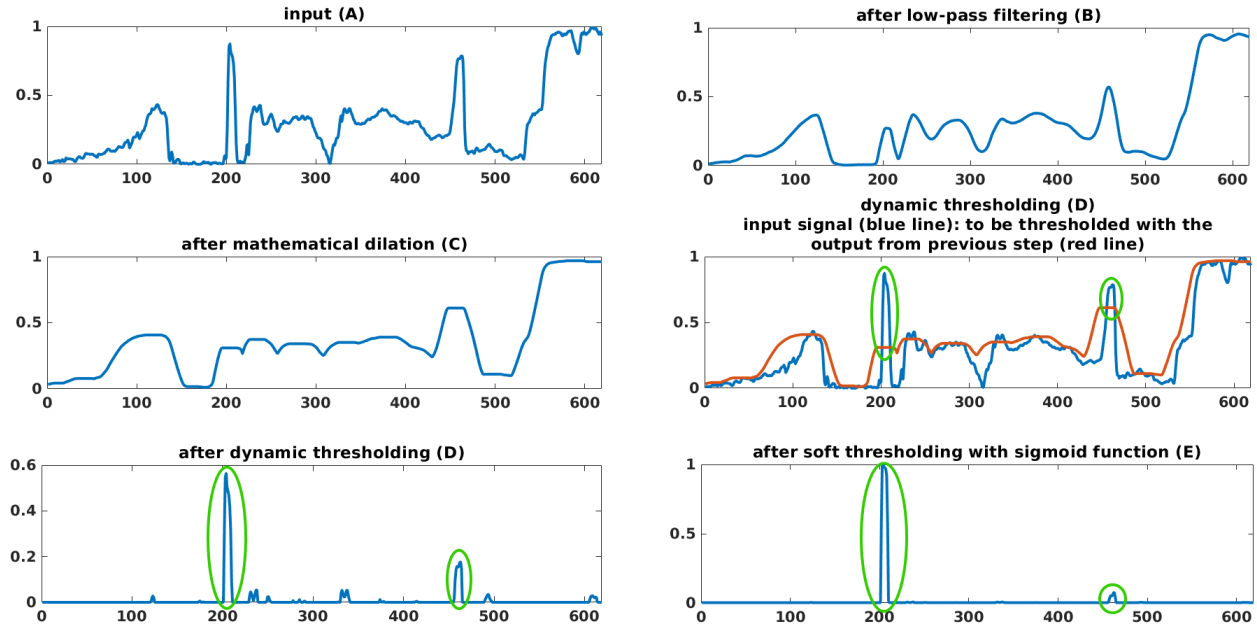
Here, we show that HAnS overcomes the difficulties for detecting highlights in cinematic LDR content (see Section II-B).

For that purpose, we use a representative example (see Fig.7) of cinematic LDR content for which we display the output from each of the A-E steps in the main HLD procedure. We only show the outputs with the minRGB feature is processed.

The input LDR image presented in Fig. 7, considers specular reflections and light reflections from bright diffuse surfaces. Content-wise, for the areas which belong to specular reflections from the sunglasses in the LDR image, we refer to as "Sunglasses" and for the light reflections from the shirt we refer to as "Shirt". In Fig.7 (2), we present the extracted cross-section lines from the outputs (see Fig.7 (1)) for each step of the main HLD procedure.

It can be seen that after dynamic thresholding (see the fifth image in the row from Fig.7 (1) and the fifth graph from Fig.7 (2)), the only extracted areas as highlights are those for which the extracted pixel-wise feature has higher absolute intensities than the output of step C. On the graphs, (see Fig.7 (2), step D) these parts are marked with green ellipses. On the fifth image in the row from Fig.7 (1), the two areas that are more pronounced than other detected areas, are marked in a same way. The first detected high peak belongs

(1)



(2)

FIGURE 7. Performance of HAnS, presented through the image outputs and cross-section lines extracted from them, when each of the A-E steps, is applied. (1) Images (from left to right): original RGB LDR image, extracted minRGB feature (step A), output (step B), output (step C), output (step D), output (step E). The red lines represent the extracted cross section lines and the green ellipses represent the areas that are detected as highlights; (2) Graphs for the extracted cross-section lines from the image outputs presented in (1).

(see the last graphic in Fig.7 (2)) to the area associated with the specular reflections from the Sunglasses. The second high peak belongs to a small area, which is associated with light reflections from bright diffuse surface (i.e. Shirt) and hence, it is incorrect detection. However, following the definition of the highlights (see Section II) in cinematic LDR content, this part is correctly detected by HAnS. Although it is detected as a small bright area, the maximum output intensity of this peak (after the soft-thresholding step E from the main HLD procedure) is quite lower than the maximum output intensity of the first detected peak.

Relying on the conclusions from the experimental analysis (about the advanteougness of using the local contrast feature and the relative size feature for highlights detection) presented in [1] and additionally with the results from the presented example, we show that HAnS is capable of distinguishing highlights from light reflections from diffuse surfaces and at the same time it is suitable for detecting blurred (out of focus) specular reflections. Moreover, HAnS has the advantage of having two degrees of freedom, which are given with the two controlling parameters: the size parameter and the local contrast threshold. By adjusting these parameters, we are able to decide which highlights (relatively to their surrounding: small and pronounced highlights, small and less pronounced highlights, wide and pronounced highlights, wide and less pronounced highlights) we want to prioritize when performing the inverse-tone mapping approach. In other words, by having two degrees of freedom, we accomplish to preserve the freedom of using HAnS in cases of making different artistic choices on the reconstructed HDR content.

IV. HANS AS PART OF INVERSE TONE MAPPING

The main goal of the algorithms for LDR-to-HDR conversion is to increase the dynamic range of the LDR content, by increasing the contrast between the bright and the dark regions without loosing details from the LDR content. Relying on the model for the reflected scene radiance in real world, in the work of Wolff and Lawrence [31], it is stated that the radiance of the highlights (specular reflections and light sources) is always of higher magnitude than the radiance of the reflections from diffuse surfaces. Therefore, it is expected that in the original scene the highlights will have the highest brightness levels. Since the initial content of the original scene is lost in the LDR content production, there is no unique solution and the LDR-to-HDR conversion is subject to human perception. Each solution is interpreted differently and depends on the human perception of the HDR content.

In what follows, in Section IV-A we propose an inverse tone mapping (ITM) approach, that makes a distinction between reproducing the highlights and reproducing light reflections from diffuse surfaces. In the proposed ITM approach we combine local and global expansion of the dynamic range of the input LDR content. The usage of local boosting has the advantage of providing the ability to prioritize highlights, over light reflections from diffuse surfaces. The usage of global boosting has the advantage of prioritizing bright areas over dark and normally exposed areas.

In Section IV-B we discuss and analyze the output of the proposed ITM approach in function of the input variables.

A. PROPOSED ITM

The proposed ITM approach is only applied on the luminance representation of the input tristimulus image. This is commonly done in ITM to avoid affecting the colors. Therefore, an implementation in CIELUV or CIELAB color spaces is most suitable. In the proposed ITM approach we use simple pixel-wise ITM function for the global expansion (for which we refer to as ''global boosting'') of the dynamic range. We ensure sufficient space for additional boosting of the detected highlights areas (for which we refer to as ''local boosting'').

In Fig.8, we present an example of the pixel-wise ITM function for global boosting. This function is plotted in green, while the red dotted line is an example for the linear pixel-wise ITM function. The slope of the linear pixel-wise ITM function is denoted with s_1.

For the case when the linear pixel-wise ITM function is considered, the space for local highlights boosting is constrained to be within the region defined with the term: $c_2 = \mathbf{L}_{\text{out,max}} - s_1 \cdot \mathbf{L}_{\text{in}}(\mathbf{x})$, where $\mathbf{L}_{\text{in}}(\mathbf{x})$ is the per-pixel luminance of the input LDR content (for which we refer to as ''input luminance'') when it is displayed on LDR screen, and $\mathbf{L}_{\text{out,max}}$ is the maximal luminance intensity that the HDR screen is able to produce.

When non-linear pixel-wise boosting ITM function is considered, the space for local highlights boosting is constrained to be within the region defined with the term:

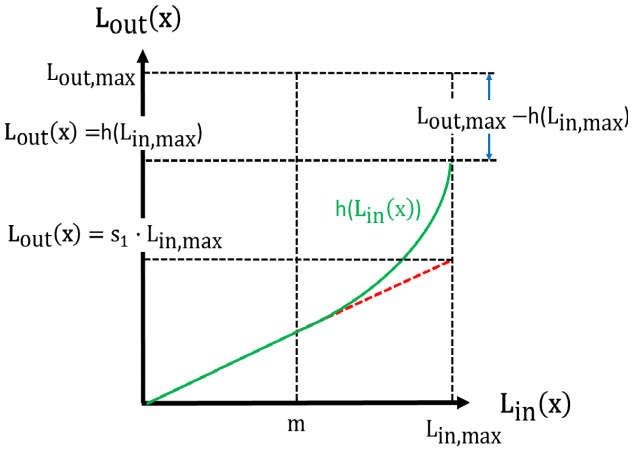


FIGURE 8. Example of a simple pixel-wise inverse-tone mapping (ITM) function. The bright regions from the input LDR content (high pixel intensity of the displayed luminance of the input LDR image) are boosted with a non-linear (exponentially rising) function. For the darker part of the LDR content, there is a linear dependence between the output luminance (luminance of the displayed HDR content on a HDR screen) $\mathbf{L}_{\text{out}}(\mathbf{x})$, and the input luminance (luminance of the displayed LDR content on an LDR screen), $\mathbf{L}_{\text{in}}(\mathbf{x})$. The space within the constraint: $\mathbf{L}_{\text{out,max}} - h(\mathbf{L}_{\text{in}}(\mathbf{x}))$, is reserved for local boosting of the highlights detected by HAnS.

$c_1 = \mathbf{L}_{\text{out,max}} - h(\mathbf{L}_{\text{in}}(\mathbf{x}))$, where the global boosting ITM function denoted with $h(\mathbf{L}_{\text{in}}(\mathbf{x}))$ is required to satisfy the term $0 \leq h(\mathbf{L}_{\text{in}}(\mathbf{x})) \leq \mathrm{L}_{\text{out,max}}$.

The proposed ITM approach considers two input variables: the input luminance $\mathbf{L}_{\text{in}}$ and the Highlights Map $\mathbf{M}(\mathbf{x})$. Therefore we introduce two terms for obtaining the luminance $\mathbf{L}_{\text{out}}(\mathrm{x})$ of the HDR content reproduced to be displayed on HDR screen (for which we refer to as ''output luminance''). The first term denoted by $\mathbf{L}_{\text{out,gb}}(\mathrm{x})$, represents the part of the output luminance that is obtained with global pixel-wise boosting, while the second term denoted by $\mathbf{L}_{\text{out,lb}}(\mathrm{x})$, represents the part of the output luminance obtained with local pixel-wise boosting. The mathematical formulation is given by (10), (11), (12) and (13).

$$\mathbf{L}_{\text{out}}(\mathbf{x}) = \mathbf{L}_{out,gb}(\mathbf{x}) + \mathbf{L}_{out,lb}(\mathbf{x}) \tag{10}$$

$$\mathbf{L}_{\text{out,gb}}(\mathbf{x}) = h(\mathbf{L}_{\text{in}}(\mathbf{x})) \tag{11}$$

$$\mathbf{L}_{\text{out,lb}}(\mathbf{x}) = c_1 \cdot q(\mathbf{L}_{\text{in}}(\mathbf{x})) \cdot \mathbf{M}(\mathbf{x})^{\gamma} \tag{12}$$

$$q(\mathbf{L}_{\text{in}}(\mathbf{x})) = \begin{cases} \mathrm{L}_{\text{in,max}} & \mathbf{L}_{\text{in}}(\mathbf{x}) \geq \dfrac{\mathrm{L}_{\text{in,max}}}{2} \\ \dfrac{\mathbf{L}_{\text{in}}(\mathbf{x})}{\mathrm{L}_{\text{in,max}} - \mathbf{L}_{\text{in}}(\mathbf{x})} & \mathbf{L}_{\text{in}}(\mathbf{x}) < \dfrac{\mathrm{L}_{\text{in,max}}}{2} \end{cases} \tag{13}$$

In (12) and (13), $q(\mathbf{L}_{\text{in}}(\mathbf{x}))$ is a gain function that controls the amplification slope for boosting the highlights according to the intensities in the Highlights Map, $\mathbf{M}$. The gain function depends on the input luminance. In order to avoid amplification bigger than 1 (i.e. avoid clipping), the presented gain function is constrained with the maximum intensity of the input luminance, i.e. $\mathrm{L}_{\text{in,max}}$, in the case when the input luminance reaches value that is equal or higher than $\frac{\mathrm{L}_{\text{in,max}}}{2}$. The γ parameter, with the constraint $\gamma > 0$, controls how

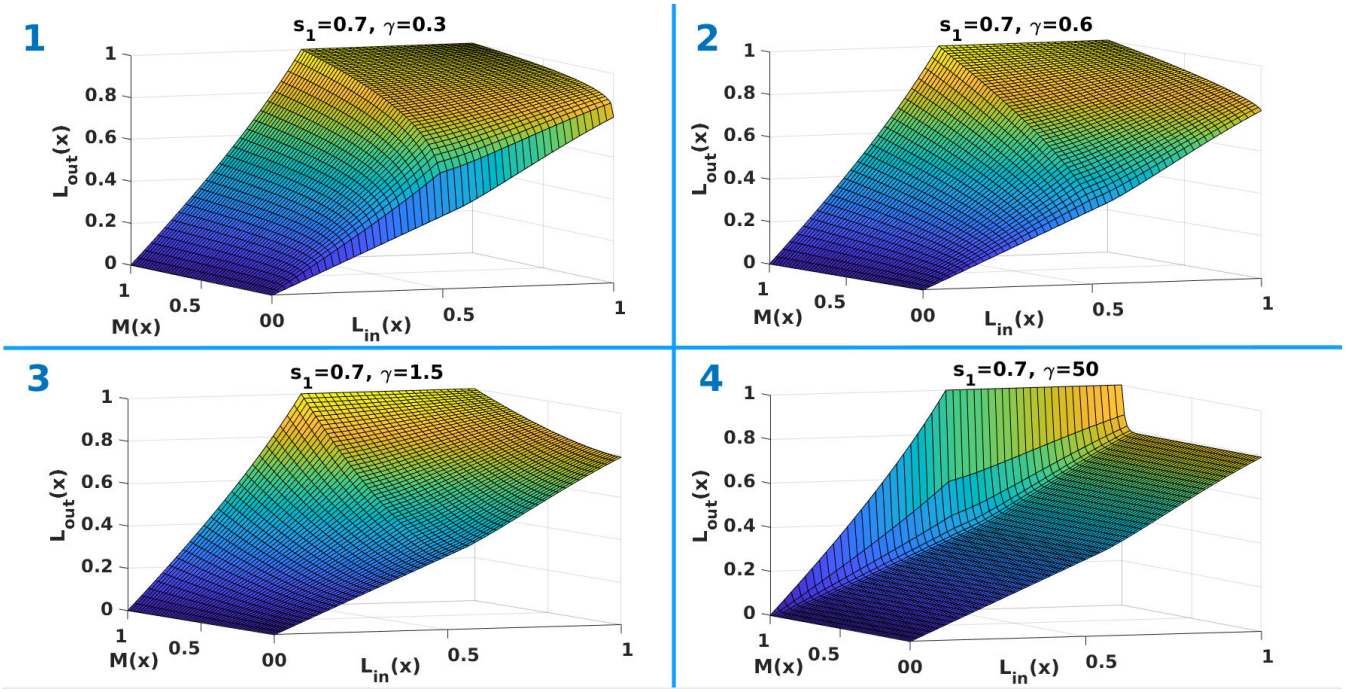


FIGURE 9. Behaviour of the proposed ITM approach, considering the two inputs: the input luminance, $\mathbf{L_{in}(x)}$, in range [0-1] and the Highlights Map, M, in range [0-1]. Four cases are presented where for the non-linear pixel-wise ITM function $h(\mathbf{L_{in}(x)})$ (given by (14)), the input parameters settings are: $s_1 = 0.7$, and $\gamma \in \{0.3, 0.6, 1.5, 50\}$.

much all of the detected highlights in the Highlights Map, **M**, have influence on the reconstructed HDR image. Closer the value to 0 is, more pronounced in the HDR content the detected weak highlights become. As the value of the γ parameter increases, higher priority is given to the detected strong highlights.

In order to analyze the ITM approach, we use a simple pixel-wise ITM function, given by (14). This function is associated with the green curve presented in Fig.8 and it is a combination of a linear and an exponential part. The slope for the linear part is denoted by s_1. For input luminance higher than m, the function starts to rise exponentially.

$$h(\mathbf{L}_{\text{in}}(\mathbf{x})) = s_1 \cdot \mathbf{L}_{\text{in}}(\mathbf{x}) + c_2 \cdot u(\mathbf{L}_{\text{in}}(\mathbf{x})) \cdot (\mathbf{L}_{\text{in}}(\mathbf{x}) - m)^{1.5} \tag{14}$$

$$u(\mathbf{L}_{\text{in}}(\mathbf{x})) = \left\{ \begin{array}{ll} 1 & \mathbf{L}_{\text{in}}(\mathbf{x}) \geq m \\ 0 & \mathbf{L}_{\text{in}}(\mathbf{x}) < m \end{array} \right\} \tag{15}$$

$$c_2 = L_{\text{out,max}} - s_1 \cdot \mathbf{L}_{\text{in}}(\mathbf{x}) \tag{16}$$

The constraint factor, by which clipping of $\mathbf{L}_{\text{out,gb}}(\mathbf{x})$ to $\mathbf{L}_{\text{out,max}}$ is avoided, is denoted with c_2. The constraint factor for the $\mathbf{L}_{\text{out,lb}}(\mathbf{x})$ term, has the same form, which was presented earlier in this Section (given with $c_1 = \mathbf{L}_{\text{out,max}} - h(\mathbf{L}_{\text{in}}(\mathbf{x}))$).

B. ON THE PERFORMANCE OF THE PROPOSED ITM

Here, we discuss the behavior of the proposed ITM approach depending on the two input variables: the input luminance and the Highlights Map obtained with HAnS. The analysis is presented for four different settings of the input parameters: γ and s_1. In Fig.9 we present a three-dimensional plot of the output, for 4 cases. Here, the input luminance and the output luminance are normalized to be in range [0-1]. It can be noticed that clipping is avoided in all cases. It can also be noticed that, the higher the value of the exponent γ (used in (12)) is, the less dependent to the local boosting term $\mathbf{L}_{\text{out,lb}}(\mathbf{x})$ the output luminance $\mathbf{L}_{\text{out}}(\mathbf{x})$ becomes.

V. EXPERIMENTAL ANALYSIS

To assess the performance of HAnS we perform three experiments. With the first experiment we perform an objective evaluation of the highlights detection. With the second experiment we perform an objective evaluation of the highlights detection when used in the boosting, as part of a complete ITM (LDR-to-HDR conversion) pipeline. With the third experiment we perform a subjective evaluation of the proposed highlights detection and boosting. The experimental framework and the results for the both objective experiments are presented in Section V-A and in Section V-B, while for the subjective experiment they are presented in Section V-C. With the objective experiment for highlights detection, we evaluate the accuracy of HAnS and compare it with four different highlights detection algorithms. With the objective experiment for highlights boosting, we evaluate HAnS with the proposed ITM, to properly boost the regions that are important in obtaining the HDR effect. The aim of the subjective experiment is to show that by using HAnS as part of an ITM process (which combines global and local boosting of the highlights), we succeed to produce results that are more pleasant (according to the human perception) than results produced with a global ITM approach only.

A. OBJECTIVE EVALUATION OF HIGHLIGHTS DETECTION

In the first objective experiment, we compare HAnS with four different methods for highlights detection, which are listed in Subsection V-A1. We use two image data-sets, out of which the first one is publicly available and does not consider cinematic content and the second one is a data-set we create for performing the analysis on the difficult cases from cinematic content and real-world images. In Subsection V-A2, we describe the two image-data sets in details. In order to perform thorough objective analysis we use nine metrics, out of which six are commonly used for binary classification evaluation, and the other three are used for object level segmentation evaluation. Additionally, we perform a time performance analysis for HAnS. The used metrics are briefly described in Subsection V-A3. For further explanation on the object-level evaluation metrics, we direct the reader to Appendix A. In Subsection V-A4 and Subsection V-A5, we present the quantitative and the qualitative results from the objective evaluation.

1) HLD METHODS USED FOR COMPARISON

Here, we give a brief description of the algorithms for highlights detection used for comparison. Most of them, for distinguishing the specular reflections from light reflections from diffuse surfaces, are threshold-based techniques. We classify these algorithms in two categories:

- highlights detection algorithms used for removal of specular reflections;
- highlights detection algorithms used for the purposes of LDR-to-HDR conversion.

a: HLD METHODS USED FOR REMOVAL OF SPECULAR REFLECTIONS

These techniques are based on the dichromatic model introduced by Shafer and Steven [20]. Although this model correctly describes the physical process of light reflection from surfaces that have the both reflectance properties (specular and/or diffuse), it is constrained to light conditions with the existence of single light source. Moreover, this model is not applicable for highlights detection in images abundant with mirror-like surfaces or of low quality. Also, this model is not applicable for cases where the highlights are clipped or compressed, as it is the case in the cinematic content.

Representative algorithms from this category with available implementations are the algorithms of Shen *et al.* (see [24] and [25]), and Huo *et al.* [11]. The latter algorithm is threshold based and the valley-emphasis method [40] is used for determining the threshold values.

b: HLD METHODS USED FOR THE PURPOSE OF LDR-TO-HDR CONVERSION

In this category we classify the algorithms of Huo *et al.* [11] and Meylan *et al.* [4].

For calculating the threshold values in the algorithm of Huo *et al.* [11], the authors used the Otsu's automatic thresholding method [39], which is a histogram based method, similar to the valley-emphasis method [40].

In the algorithm of Meylan *et al.* [4], which is also threshold based, the global (hard) thresholds are calculated from the luma representation of the input RGB image, after it is low-pass filtered with a two-dimensional filter of a predefined size that is dependent of the image size. In this case, the filtering approach is used for obtaining the thresholds in order to distinguish areas that belong to specular reflections from the areas that belong to light reflections from bright diffuse surfaces. The disadvantage of the algorithm of Meylan *et al.* is that the highlights detection is dependent on the whole image content, and not on the near surrounding of the highlights. This is due to the hard thresholding performed globally on the whole image content.

2) DATA-SETS

The experimental evaluation is performed on two image data-sets:

- the publicly available SPEC-DB [37], [38], which consists of LDR images that do not consider cinematic content and is annotated for specular reflections;
- a novel image data-set, which is manually annotated for highlights (specular reflections and light sources) and consists of cinematic content and real-world photographs.

In what follows, we describe the publicly available data-set SPEC-DB in more details. Then we proceed with describing the experimental setup for creating the proposed image data-set and the content of the proposed image data-set itself.

a: SPEC-DB IMAGE DATA-SET

As it is presented in [36], the SPEC-DB image data-set consists of 300 images from 100 objects, diverse in color and specularity properties. Each object was acquired under three different illumination conditions: diffuse, ambient, and directed. The images with ambient and directed light were annotated for specular reflections by one observer. All objects in this data-set show specularity properties when being illuminated. Therefore, for the objective evaluation, we only use the images taken of the objects illuminated by ambient (the SPEC-DB 11 subset) and directed light (the SPEC-DB 22 subset).

The images from the SPEC-DB data-set:

- are of resolution lower than the resolution of the commonly produced cinematic content;
- consider scenes that consist of one or multiple objects with same specularity properties;
- consider cases with only specular reflections;
- do not consider cases in which all the problems of LDR cinematic content production (see Section II-B) are addressed;

We use this image data-set in order to have a fair comparison on general LDR content with the highlights detection algorithms presented in Subsection V-A1.

b: PROPOSED IMAGE DATA-SET CONSISTING OF CINEMATIC CONTENT AND REAL-WORLD PHOTOGRAPHS

In order to perform evaluation analysis with a data-set that considers cinematic content in which all challenges for highlights detection (see Section II-B) are addressed, we propose a novel image data-set, which is manually annotated for highlights (specular reflections and light sources).

For the design of this image data-set we used real-world photographs that are freely available on the Internet and content from movies that are property of the National Public-Service Broadcaster for the Flemish Region and Community of Belgium VRT [51].

In what follows, we give a brief description for the content of this image data-set and the annotation procedure.

- Content:
  - 23 movie frames (LDR frames) and selected 15 additional real-world photographs, as representative examples for highly diverse in content scenes.
  - Scenes:
    - ∗ vary from very dark to very bright;
    - ∗ are indoor and outdoor;
    - ∗ are abundant with objects with different surface reflectance properties;
    - ∗ are abundant with diverse in color and size, light sources;
    - ∗ are abundant with natural content;
    - ∗ include people and visible faces.
  - The illumination conditions for all captured scenes from the selected images and movie frames are not known and are not subject of estimation. In that way, we give the opportunity to annotators to use their own perception and subjective opinion, when annotating a region of interest as highlight.
  - In order to examine cases for which common HLD algorithms have incorrect detections, areas that only belong light reflections of bright objects with diffuse surface properties are also considered in the creation of the proposed image data-set.
- Annotation procedure:
  - Three annotators, experts in the field of HDR imaging, were independently to select areas they consider as highlights (specular reflections or light sources);
  - For simplification, regions of interest (patches extracted from each image) were shown to annotators (one patch at a time) in the context of the whole image content and separately (zoomed version) where they are able to select the area of the actual highlight more precisely;
  - The annotations are per pixel and binary;
  - 960 patches are annotated by each annotator:
    The procedure is splitted into 9 annotation sessions (in order to keep the annotator's attention on a desired level), so that in each session approximately 107 patches were annotated in maximum time of 1h.

The annotators are non-biased experts in the field and therefore we do not proceed in examining the inter and intra annotator's agreement, which is commonly done when a single ground-truth data set is designed and proposed for usage in different applications. For that purpose, the annotations made by each annotator in the objective evaluation are considered to be separate reference (ground-truth) image data-sets.

3) METRICS

We treat the highlights detection as a binary classification/segmentation problem, i.e. certain area is either highlight or light reflection from diffuse surface. For that purpose, we use six metrics for binary classification evaluation and three metrics for object level segmentation evaluation. In the analysis we compare binary reference images with binary segmented images, obtained with different parameter settings when HAnS and the HLD algorithms from the comparison are applied. In what follows we present the metrics we use for objective evaluation, in two separate lists: metrics for binary classification evaluation and metrics for object level segmentation evaluation, with a short description.

a: METRICS FOR BINARY CLASSIFICATION EVALUATION

- Accuracy (ACC) per annotated image or annotated patch, in order to account for: large amount of images diverse in resolution and size and challenging content for highlights detection.
- True positive rate (TPR) per annotated image or annotated patch, in order to support the ACC.
- False positive rate (FPR) per annotated image or annotated patch, in order to support the ACC.
- True negative rate (TNR) per annotated image or annotated patch, in order to support the ACC.
- False negative rate (FNR) per annotated image or annotated patch, in order to support the ACC.
- Sørensen-Dice index (DSC) [43], which is commonly used for comparing the similarity and the diversity of sample sets. Similar to the Sørensen-Dice index is the Jaccard index [44], [45]. Because one can be derived from the other, we only use the Sørensen-Dice index.
- Symmetric partition distance ($D_{sym}{}'$) [50], which counts the minimal number of pixels that must be removed from the both binary images that are being compared, so that the remaining pixels in each of them will be identical. The authors of [46] use the complement of the metric proposed by Cardoso and Corte Real in [50]. Consequently, the modified metric is described as the maximum number of pixels that remain such that the reference image equals the segmented image. Since the two metrics are equivalent, by following the positive logic analogy, we use the metric proposed by the authors of [46], which is derived from the metric proposed in [50]. For this metric, we use the same notation, $D_{sym}{}'$, as in [46]. This metric has the same interpretation as accuracy per image (or image patch).

b: METRICS FOR OBJECT LEVEL SEGMENTATION EVALUATION

Image segmentation is considered as a process of pixel labeling and its performance is evaluated by metrics at object level. Many of these metrics, as explained in [48], consider the number of mis-segmented pixels, the position of the mis-segmented pixels, the number of the objects in the image and their geometric features. In most of them, the image segmentation is considered as a process of pixel labeling and for that reason they are known as metrics for object-level segmentation evaluation.

- Martin's error metric (in both versions, Local Consistency Error (LCE) and Global Consistency Error (GCE)) [47], tolerates over- and under-segmentation, in terms of the number of segmented objects and the object sizes. Therefore, it is suitable for being used in cases when the annotations (reference images) are manually performed by different people (annotators) with allowed deviations in the region boundaries of the highlights candidates. On the other hand, it has a disadvantage that (as explained in [48]) it does not discriminate among segments that are scaled version of each other. Therefore, in case of segments existing in both the reference image and the segmented image, the LCE and the GCE, in most of the cases will result in values equal or very close to 0, although there is a big difference between the segments in the reference and the segmented image.
- Object level consistency error (OCE) [48], overcomes the disadvantage of Martin's error metric of not being sensitive to the size and the shape of the segments in both the reference and the segmented image.
- Quality rate (QR with the both versions, QR_{SR} and QR_{RS}) [49], beside being region-based it also considers the direction of matching between the reference and the segmented image.

The used metrics for object-level segmentation evaluation, are not defined for cases where no detections and/or no reference segmentations are available. Therefore, we adapt Martin's error metric, the OCE and the QR metrics to have the highest penalty for two cases:

- when there are existing segments in the output of the detection algorithm and there are no positive annotations in the reference binary images (penalizing for false positive segments)
- when there are no existing segments in the output of the detection algorithm and there are positive annotated regions in the binary reference image (penalizing for false negative segments).

For better description of these metrics and the way how they are adapted for false positive and false negative segments, we direct the reader to Appendix A of this article.

4) RESULTS FROM THE QUANTITATIVE ANALYSIS

Here, we present the quantitative results from the objective evaluation analysis. We additionally present the results from the running time performance analysis.

For the evaluation of the algorithms of Meylan *et al.* [4], Shen *et al.* [24] and [25], Zou *et al.* [19] and HAnS, we sweep over a wide range for each parameter and select the best results (for each evaluation metric) per image. In case of the proposed algorithm, the free parameters are the size parameter (sizes of the applied filter: 1×1, 2×2, 4×4, 8×8, 16×16, 32×32, 64×64, 128×128, 256×256, 512×512, 1024×1024) and the local contrast threshold (changed with a step factor of 0.05 in a range [0,1]). HAnS and the algorithm of Shen *et al.* [24] and [25], produce output maps with per-pixel real numbers in the range [0,1]. In order to have a proper comparison between the reference binary images (from both image data-sets) and the outputs of the highlights detection algorithms, the binary maps for these algorithms are created by thresholding the outputs with 0. When the algorithm of Huo *et al.* [11] is evaluated, no parameters are adjusted because it automatically detects highlights in an image. We select the best results (per metric) for each HLD algorithm and calculate the average and the standard-deviation over the separate image data-sets, SPEC-DB 11 and 22 and the proposed image data-set consisted of LDR cinematic content and real-world photographs.

a: RESULTS ON SPEC-DB

From the results presented on Table 1, in the objective evaluation analysis, HAnS achieves best performance. Its overall performance is most similar to the performance of the algorithm of Meylan *et al.* [4]. Since both algorithms implicitly rely on using the relative size to detect highlights, the similarity in performance is expected. With HAnS we succeed to correctly detect highlights (see DSC, QR and TPR) and avoid false detections (see results for Martin's error metric, OCE and FPR). HAnS outperforms the algorithms of Zou *et al.* [19], Huo *et al.* [11] and Shen *et al.* [24], [25]. Therefore, we will focus on the differences in results between HAnS and the algorithm of Meylan *et al.* [4]. The algorithm of Meylan *et al.* shows slightly better (yet, very similar) results than HAnS, only in terms of having less false detections (results from Martin's error metric and FPR). This is due to the fact that the SPEC-DB data-set is consisted of images that only consider achromatic specular reflections. It is expected that by using only the minRGB image feature in HAnS, many of the false detections will be avoided. From the results for the Martin's error metric, in case when the algorithm of Meylan *et al.* and HAnS are applied, it can be noticed that the calculated error values are very close to 0. This is due to the insensitivity of the metric to over- and under-segmentation (discussed in Subsection V-A3), the advantage of the two algorithms to introduce small amount of false positive detections (which is not the case with the other algorithms from the comparison (see Table 1)) and the statistics of the analyzed image data-set (all images consider specular highlights and were analyzed at a full content level, so there are no examples where only non-highlight areas are present). Regarding the OCE and QR metrics, from the presented results, it can be noticed that HAnS outperforms the algorithms from the

TABLE 1. Evaluation results on SPEC-DB 11 and SPEC-DB 22 (averaged over all images in each subset). Evaluation metrics (in range [0,1]) for which higher value means better performance: Dice coefficient (DSC), Quality Rate (QR), Complement of the symmetric partition distance (D_{sym}') which is equal to accuracy (ACC), True Positive Rate (TPR) and True Negative Rate (TNR). Evaluation metrics (in range [0,1]) for which lower value means better performance: Martin's error metric, Object level consistency error (OCE), False Positive Rate (FPR) and False Negative Rate (FNR). The results in the fields marked with green are associated with the best performing algorithms for each different metric. The results from the algorithms that are second best performing are presented with blue color. Since the results for TPR, TNR, FPR and FNR support ACC (which is marked for the best performing algorithms), they are left unmarked.

Evaluation on SPEC-DB						
SPEC-DB 11						
Evaluation metric (↑) average ± standard deviation in range [0-1] over the whole data-set		Highlights detection algorithms				
		Meylan et al.	Zou et al.	Huo et al.	Shen et al.	Proposed HAnS
DSC		0.767 ± 0.155	**0.442** ± 0.24	**0.566** ± 0.233	**0.151** ± 0.118	**0.812** ± 0.144
QR	QR_{SR}	0.422 ± 0.263	**0.197** ± 0.185	**0.28** ± 0.242	**0.051** ± 0.077	**0.46** ± 0.259
	QR_{RS}	0.427 ± 0.252	**0.187** ± 0.174	**0.26** ± 0.222	**0.055** ± 0.072	**0.59** ± 0.229
Dsym'=ACC		0.97 ± 0.027	**0.776** ± 0.18	**0.896** ± 0.077	**0.186** ± 0.101	**0.973** ± 0.032
TPR		**0.71** ± 0.269	**0.981** ± 0.102	**0.959** ± 0.126	**0.899** ± 0.091	**0.769** ± 0.232
TNR		**0.985** ± 0.026	**0.755** ± 0.199	**0.892** ± 0.081	**0.121** ± 0.095	**0.987** ± 0.022
Evaluation metric (↓) average ± standard deviation in range [0-1] over the whole data-set		Highlights detection algorithms				
		Meylan et al.	Zou et al.	Huo et al.	Shen et al.	Proposed HAnS
Martin's error metric	LCE	**0.00001** ± 0.00015	**0.011** ± 0.063	**0.019** ± 0.067	**0.007** ± 0.021	0.00059 ± 0.0044
	GCE	**0.00004** ± 0.00028	**0.016** ± 0.078	**0.03** ± 0.089	**0.007** ± 0.021	0.00081 ± 0.0053
OCE error		0.35 ± 0.192	**0.732** ± 0.217	**0.603** ± 0.241	**0.757** ± 0.224	**0.254** ± 0.19
FPR		**0.016** ± 0.026	**0.245** ± 0.199	**0.108** ± 0.081	**0.879** ± 0.095	**0.013** ± 0.022
FNR		**0.291** ± 0.269	**0.02** ± 0.102	**0.041** ± 0.126	**0.101** ± 0.091	**0.232** ± 0.232
SPEC-DB 22						
Evaluation metric (↑) average ± standard deviation in range [0-1] over the whole data-set		Highlights detection algorithms				
		Meylan et al.	Zou et al.	Huo et al.	Shen et al.	Proposed HAnS
DSC		0.805 ± 0.157	**0.583** ± 0.233	**0.685** ± 0.206	**0.294** ± 0.199	**0.82** ± 0.133
QR	QR_{SR}	0.504 ± 0.242	**0.319** ± 0.251	**0.374** ± 0.258	**0.128** ± 0.162	**0.513** ± 0.252
	QR_{RS}	0.495 ± 0.243	**0.312** ± 0.242	**0.361** ± 0.255	**0.144** ± 0.163	**0.652** ± 0.234
Dsym'=ACC		**0.952** ± 0.035	**0.786** ± 0.151	**0.885** ± 0.082	**0.262** ± 0.145	0.945 ± 0.057
TPR		**0.787** ± 0.231	**0.982** ± 0.05	**0.934** ± 0.112	**0.895** ± 0.099	**0.784** ± 0.227
TNR		**0.969** ± 0.041	**0.742** ± 0.186	**0.887** ± 0.091	**0.111** ± 0.08	**0.959** ± 0.079
Evaluation metric (↓) average ± standard deviation in range [0-1] over the whole data-set		Highlights detection algorithms				
		Meylan et al.	Zou et al.	Huo et al.	Shen et al.	Proposed HAnS
Martin's error metric	LCE	**0.00005** ± 0.00043	**0.009** ± 0.033	**0.227** ± 0.055	**0.0064** ± 0.014	0.00091 ± 0.0057
	GCE	**0.00005** ± 0.00043	**0.013** ± 0.048	**0.036** ± 0.074	**0.0064** ± 0.014	0.00097 ± 0.0058
OCE error		0.295 ± 0.161	**0.607** ± 0.262	**0.502** ± 0.245	**0.702** ± 0.228	**0.233** ± 0.181
FPR		**0.031** ± 0.041	**0.258** ± 0.186	**0.114** ± 0.091	**0.889** ± 0.08	**0.041** ± 0.079
FNR		**0.213** ± 0.231	**0.018** ± 0.05	**0.066** ± 0.112	**0.105** ± 0.099	**0.216** ± 0.227

comparison, and it achieves most similar performance to the algorithm proposed by Meylan *et al.* [4]. The OCE metric and the QR metric, beside penalizing the false detections severely, they also reward the cases when the size of the corresponding overlapping areas is similar (see the explanation in [48] and in the Appendix A). Therefore, considering the results from these two metrics and the results from the metrics that were previously discussed, we can conclude that HAnS achieves best performance among the algorithms from the comparison and it is most similar in performance with the algorithm proposed by Meylan *et al.* [4].

If we proceed with the analysis towards making a comparison between the results from the two parts of the SPEC-DB image data-set, we can notice the shift in values. This shift is due to the different statistics between the two subsets, referring to the different illumination conditions under which the images were taken. It is obvious that for SPEC-DB 22, the rate for correct detections is increased with the rate for false detections, which leads to slight drop in the accuracy (ACC) results for most of the HLD algorithms. The reason for this is that the directed illumination conditions create higher intensity of the reflected light, which is then clipped due to the camera capture. Consequently, because the images in SPEC-DB 22 contain larger, saturated (flat) areas for the specular reflections, it is harder to distinguish them from bright diffuse surfaces.

b: RESULTS ON THE PROPOSED IMAGE DATA-SET

We now discuss the results on the proposed image data-set consisted of cinematic LDR content and real world photographs, annotated for highlights. For the procedure of the performed experimental evaluation on this image data-set, we direct the reader to Appendix B.

From the results presented in Table 2, it can be seen that most importantly for cinematic and real-world LDR content, HAnS achieves best performance among all highlights detection algorithms from the comparison.

TABLE 2. Evaluation results (averaged over the whole data-set when the binary images obtained with each of the highlights detection algorithms, are compared with the binary images from each separate ground-truth and then averaged over the three ground-truth data-sets) on the annotated data-set consisted of cinematic content and real-world photographs. Evaluation metrics (in range [0,1]) for which higher value means better performance: Dice coefficient (DSC), Quality Rate (QR), Complement of the symmetric partition distance (D_{sym}') which is equal to accuracy (ACC), True Positive Rate (TPR) and True Negative Rate (TNR). Evaluation metrics (in range [0,1]) for which lower value means better performance: Martin's error metric, Object level consistency error (OCE), False Positive Rate (FPR) and False Negative Rate (FNR). The results in the fields marked with green are associated with the best performing algorithms for each different metric. The results from the algorithms that are second best performing are presented with blue color. Since the results for TPR, TNR, FPR and FNR support ACC (which is marked for the best performing algorithms), they are left unmarked.

Evaluation on the proposed data-set						
		Highlights detection algorithms				
Evaluation metric (↑) average ± standard deviation in range [0-1] over the whole data-set		Meylan et al.	Zou et al.	Huo et al.	Shen et al.	Proposed HAnS
DSC		0.424 ± 0.361	**0.31** ± 0.34	**0.301** ± 0.346	**0.254** ± 0.246	0.535 ± 0.357
QR	QR_{SR}	0.463 ± 0.415	**0.356** ± 0.394	**0.38** ± 0.41	**0.151** ± 0.192	0.542 ± 0.392
	QR_{RS}	0.47 ± 0.415	**0.36** ± 0.394	**0.386** ± 0.409	**0.152** ± 0.188	0.547 ± 0.395
Dsym'=ACC		0.905 ± 0.122	**0.764** ± 0.296	**0.824** ± 0.241	**0.374** ± 0.299	0.925 ± 0.098
TPR		**0.250** ± 0.374	**0.551** ± 0.438	**0.454** ± 0.434	**0.938** ± 0.128	**0.498** ± 0.395
TNR		**0.975** ± 0.103	**0.777** ± 0.356	**0.863** ± 0.284	**0.283** ± 0.322	**0.976** ± 0.069
Evaluation metric (↓) average ± standard deviation in range [0-1] over the whole data-set		Meylan et al.	Zou et al.	Huo et al.	Shen et al.	Proposed HAnS
Martin's error metric	LCE	0.234 ± 0.418	**0.348** ± 0.47	**0.365** ± 0.474	**0.344** ± 0.46	0.163 ± 0.362
	GCE	0.235 ± 0.418	**0.35** ± 0.469	**0.366** ± 0.473	**0.344** ± 0.46	0.164 ± 0.361
OCE error		0.5 ± 0.415	**0.616** ± 0.4	**0.589** ± 0.412	**0.773** ± 0.265	0.406 ± 0.380
FPR		**0.026** ± 0.103	**0.223** ± 0.356	**0.137** ± 0.284	**0.717** ± 0.322	**0.024** ± 0.069
FNR		**0.75** ± 0.374	**0.449** ± 0.438	**0.546** ± 0.434	**0.062** ± 0.128	**0.502** ± 0.395

From the results for ACC, TPR, QR, and DSC, we conclude that HAnS is better in detecting artificial light sources and specular reflections that are color saturated, which cannot be detected by using only the minRGB feature, as it is in the algorithms of Meylan *et al.* [4], Zou *et al.* [19] and Huo *et al.* [11]).

Based on the Martin's error metric, OCE and FPR results, it can also be concluded that when the other algorithms are applied (see the FPR results when the algorithm of Shen *et al.* [24], [25] is applied) on the whole image content, the risk for false detections is higher. This is due to the fact that most of the algorithms from the comparison are threshold based, dependent on the whole image content and do not distinguish highlights from light reflections from bright diffuse surfaces.

From the presented results, we also arrive with a conclusion that the proposed new image data-set annotated for highlights in cinematic LDR content and real-world photographs is justified for usage in experimental analysis for future research related to highlights detection, in terms that it abounds with challenging examples (areas belonging to light reflections from bright diffuse surfaces and areas belonging to blurred and not pronounced specular reflections), for which the classical highlights detection algorithms fail.

c: RUNNING TIME ANALYSIS

The results for the running time performance analysis are presented on Table 3.

The running time for each of the algorithms from the comparison, is calculated in Matlab R2015b. We analyze the running-time performance of HAnS, with optimized integer friendly implementation, in which only simple linear operations and LUTs are considered. This version can easily be adapted to any program language and for any system (such as FPGA) with limited memory or processing power. Because most of the algorithms from the comparison rely on using only one pixel-wise feature, which is minRGB, for HAnS we calculated the time when only one image feature is used and when all pixel-wise features are used. From the results presented on Table 3, it can be seen that the optimized version of HAnS is most efficient among all algorithms, in the both cases: when only one pixel-wise feature is processed and when all pixel-wise features (minRGB, luma and maxRGB) are taken into account.

TABLE 3. Results for the performance of each highlights detection algorithm, in terms of running time (measured using the program package MATLAB R2015b) for one movie frame of size 1920 × 1080. When HAnS is applied the running time is measured in two cases: 1) when only one pixel-wise feature (minRGB, maxRGB or luma) is considered for highlights detection (since that is a case for the rest of the detection algorithms from the comparison), and 2) when all pixel-wise features are processed and the Highlights Map is obtained. The results in the fields marked with green are associated with the best performing algorithm. The result from the algorithm that is second best performing algorithm is presented with blue color.

Highlights detection algorithms	Meylan et al.	Zou et al.	Huo et al.	Shen et al.	Proposed HAnS	
running time	≈0.46 s	≈0.09 s	≈0.19 s	≈9.13 s	per image feature	total
					≈0.08 s	≈0.38 s

5) RESULTS FROM THE QUALITATIVE ANALYSIS

Here, we visually present the results from HAnS and the HLD algorithms from the comparison, for two image examples.

The input image from the first example is the image presented on Fig.4. In this image some of the highlights are color saturated. Therefore, this image serves as a good example to show that by additionally (beside the minRGB image feature) using the maxRGB and the luma ppixel-wise features, we succeed not only to detect highlights that are achromatic, but also highlights that are saturated in color.

The input image from the second example is a frame extracted from a video of a broadcast TV show. The image abounds with bright specular reflections and bright light reflections from diffuse surfaces. We use this image as an example to visually show the difference in the detection quality (specifically for the areas that belong to bright diffuse surfaces) between HAnS and the HLD algorithms.

The detection results for the first example (Example 1) are presented on Fig.10 and the detection results for the second example (Example 2) are presented on Fig.11.

From Example 1, we notice that the color saturated highlights are best detected with HAnS. This is not the case for the HLD algorithms from the comparison because they rely on using only one image feature, the luma image feature (Meylan *et al.* [4]) or the minRGB image feature (Shen *et al.* [24], Huo *et al.* [11] and Zou *et al.* [19]).

From the results of Example 1, it can also be noticed that when HAnS is applied, non-binary detections with clear distinction between the highlights with different intensity are obtained. The advantage of applying different weights on the detected highlights, depending on their strength, makes HAnS easily adjustable for usage in any LDR-to-HDR conversion pipeline.

In the results of Example 2, least amount of incorrect detections occurs in the case of HAnS. The other algorithms fail in distinguishing highlights bright diffuse surfaces, because they are threshold based and the calculation of the global threshold is dependent on the whole image content (it is not locally performed as in the case of HAnS). The algorithm of Shen *et al.* [24] fails for the case of making a distinction between areas belonging to highlights and areas belonging to light reflections from bright diffuse surfaces, due to not being devised for cases where clipping and compression occur (which is a very common case in the LDR content production).

B. OBJECTIVE EVALUATION OF HIGHLIGHTS BOOSTING AS A PART OF A COMPLETE ITM (LDR-TO-HDR CONVERSION) PIPELINE

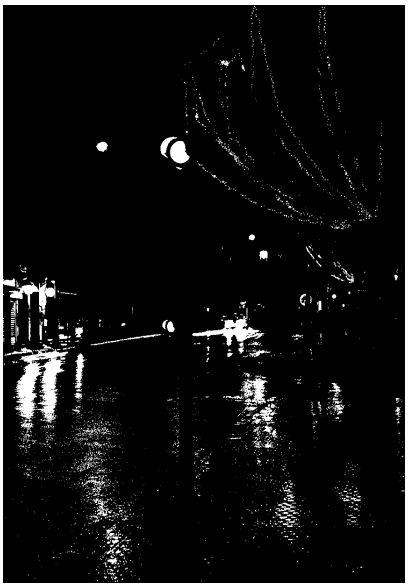
In the second objective experiment, we evaluate the performance of HAnS as part of an inverse tone mapping (LDR-to-HDR conversion) pipeline in two cases: when only global boosting is applied and when both global and local boosting are applied. The aim is to show that with proper detection and boosting of the highlights (the actual regions of interest), HAnS achieves more accurate reconstructions of HDR scenes by objective quality metrics. For that purpose we use three different state of the art algorithms for LDR-to-HDR conversion, listed in Subsection V-B1. For the purpose of evaluating HAnS with the proposed ITM on cinematic content, we use the Stuttgart data-set [53], to which we refer in Subsection V-B2. The metrics used in the quantitative evaluation are listed in Subsection V-B3. The results from the quantitative and the qualitative evaluation as part of the second objective experiment are presented in Subsection V-B4 and Subsection V-B5.

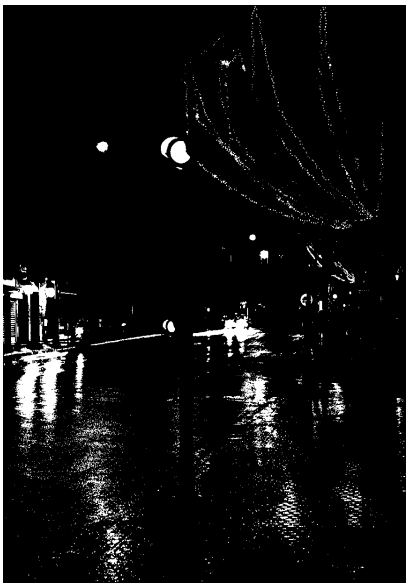
1) ITM METHODS

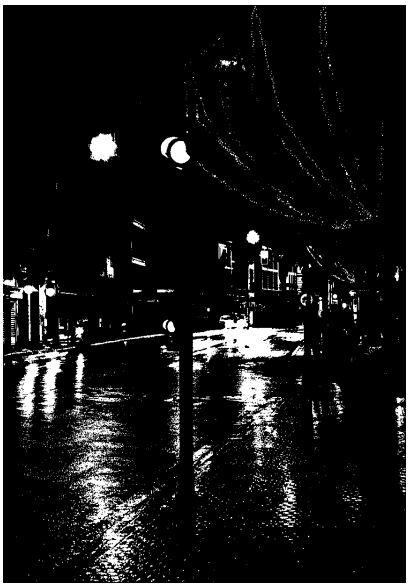
Here, we list the three state of the art algorithms for LDR-to-HDR conversion, used as comparison to the proposed ITM in combination with HAnS. As a representative algorithm from the group of deep learning based approaches, we use ExpandNet [14], while as representative state of the art algorithms from the group of classical approaches we use the ITM algorithm proposed by Meylan *et al.* [4] and the ITM algorithm proposed by Kovaleski *et al.* [10].

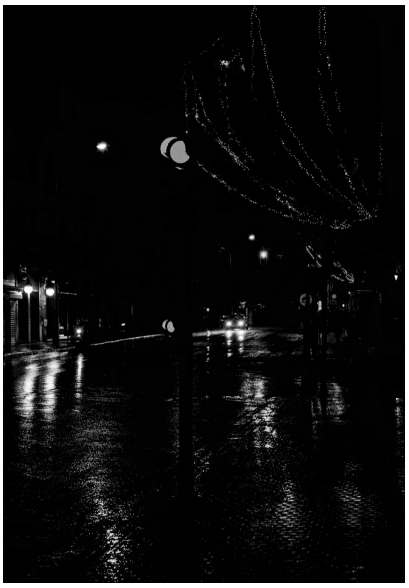
The ExpandNet algorithm [14] is a fully automatic, data driven, parameter-free algorithm for inverse tone mapping (LDR-to-HDR conversion), based on a multi-scale CNN architecture which avoids the use of upsampling layers to obtain high quality LDR-to-HDR conversion. The ExpandNet architecture consists of three branches: global branch (for preserving the global context of the image through the higher level image-wide features), local branch and dilation branch (for handling the local details and the medium level details). The advantage of ExpandNet is that it is fast, parameter free and a fully automatic algorithm, easily adaptable to any LDR-HDR training data-set. Its disadvantage, since it is a learning based approach, is in the fact that it is dependant on the diversity and the statistics of the training data-set. Additionally, when compared to the classical ITM approaches,


input LDR image


Highlights detection result: Huo et al. [11]


Highlights detection result: Zou et al. [19]


Highlights detection result: Meylan et al. [4]


Highlights detection result: Shen et al. [24]

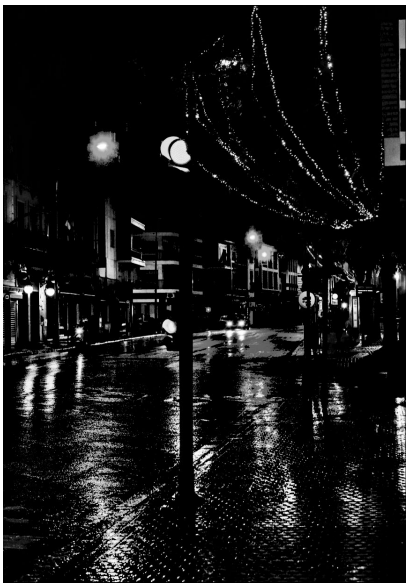
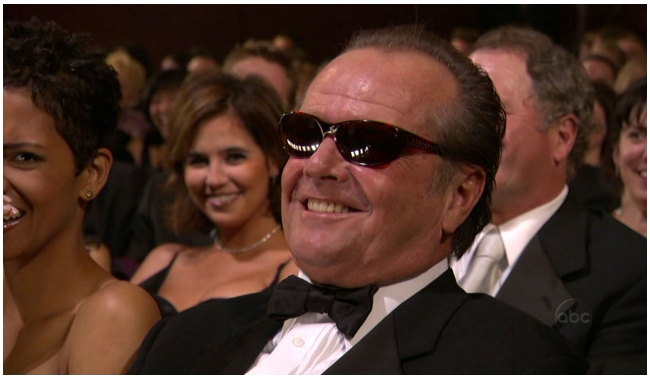

Highlights detection result: HAnS

FIGURE 10. **Results from highlights detection algorithms on Example 1: Note that the color saturated highlights are best detected with HAnS. The reason for the good detection of color saturated highlights as well as highlights saturated in the full dynamic range of the light is the usage of the three image features: the minRGB image feature, the luma image feature and the maxRGB image feature. Note that the halo surrounding the light source is detected only with the algorithm of Meylan *et al.* [4] and HAnS. Moreover, there is a clear distinction between the strong and the weak highlights, in the result obtained by HAnS, which is then very suitable to any LDR-to-HDR conversion pipeline.**

it requires extra memory resources. The software for ExpandNet is available online at [52].

The ITM algorithm of Meylan *et al.* [4], as previously explained, is a segmentation-based algorithm. After the regions are segmented into specular highlights and diffuse parts, a pixel-wise two-slope linear function is applied on the image. The first slope scales (by linearly decreasing the luminance) the diffuse parts and the second slope scales (boosts by linearly increasing the luminance) the specular parts. Since the complete approach (for LDR-to-HDR conversion) proposed by Meylan *et al.* is conceptually similar to HAnS (the both algorithms perform highlights segmentation/detection and apply a pixel-wise ITM function), we include it in the objective experimental analysis and comparison.

The algorithm of Kovaleski *et al.* [10], is another algorithm for inverse tone mapping (LDR-to-HDR conversion) that belongs in the group of classical approaches. The algorithm computes per pixel brightness enhancement function (BEF), for which the authors refer to as "expansion map". The calculation of the BEF is based on cross-bilateral filtering. The obtained BEF is then multiplied (pixel-wise) with the scaled version (to cover the dynamic range of the available display)



input LDR image

Highlights detection result: Huo et al. [11]

Highlights detection result: Zou et al. [19]

Highlights detection result: Meylan et al. [4]

Highlights detection result: Shen et al. [24]

Highlights detection result: HAnS

FIGURE 11. Results from highlights detection algorithms on Example 2. Note that HAnS produces the least amount of incorrect detections (areas that belong to light reflections from bright diffuse surfaces).

of the input image in order to obtain the resulting HDR image. Since this algorithm has shown high performance among the classical approaches in the analyses from recent works in the field of LDR-to-HDR conversion [16]–[18], we include it in our analysis together with the algorithm of Meylan *et al.* [4] and ExpandNet [14]. The software for the algorithm of Meylan *et al.* [4] and the algorithm of Kovaleski *et al.* [10] is licensed and provided with [8].

2) DATA-SET

In our experiment, we use images from the Stuttgart data-set [14]. We select this data-set because it consists of sequences of cinematic content with diverse scene content and wide in range contrast and lighting conditions. This data-set (both LDR and HDR versions) was subject to prior professional grading done by experts from VRT [51]. The details about the grading process are presented in the work of Luzardo *et al.* [18]. From the 15 different sequences of this data-set, we extracted 33 representative frames, each belonging to a different scene and/or sequence, on which in the performed objective experiment we evaluate the performance of HAnS with the proposed ITM as part of a complete LDR-to-HDR conversion pipeline.

3) METRICS

To evaluate the performance of HAnS with the proposed ITM and the ITM approaches from the comparison (see Subsection V-B1), we use three metrics:

TABLE 4. Evaluation results on 33 representative scenes from the Stuttgart data-set. Evaluation metrics for which higher value means better performance: HDR-VDP-3.0.6 (Q metric) in range [0-10], Structural similarity index (SSIM) in range [0-1], peak-signal-to-noise-ratio (PSNR) in range [0-∞]. The results in the fields marked with green are associated with the best performing algorithms for each different metric. The results from the algorithms that are second best performing are presented with blue color.

Evaluation metric (↑) average ± standard deviation	Evaluation on the Stuttgart data-set				
	ITM algorithms (LDR to HDR conversion)				
	ExpandNet	Kovaleski and Oliveira's ITM	Meylan et al. ITM	Proposed ITM (global highlights boosting)	Proposed ITM (global and local highlights boosting)
Q metric (HDR-VDP-3) in range [0-10]	7.37 ± 1.13	8.47 ± 0.64	8.47 ± 0.69	8.55 ± 0.66	8.57 ± 0.66
SSIM in range [0-1]	0.90 ± 0.13	0.97 ± 0.04	0.98 ± 0.03	0.97 ± 0.04	0.98 ± 0.04
PSNR in range [0-∞]	31.79 ± 6.84	36.83 ± 7.24	39.25 ± 6.70	38.82 ± 6.76	38.77 ± 7.04

HDR-VDP-3.0.6 metric [54], the structural similarity index (SSIM) [56] and peak-signal-to-noise ratio (PSNR). PSNR is a simple metric for objective quality assessment. SSIM tries to account for subjective human perception. HDR-VDP-3.0.6 is a specialised reference-based metric. It is based on a psychovisual model for various luminance conditions. It is designed to predict visibility and quality as perceived by human observers (in terms of mean opinion score). We use the quality (Q) metric from HDR-VDP-3.0.6 to asses the performance of the considered ITM approaches. The software is available online at [55].

4) RESULTS FROM THE QUANTITATIVE ANALYSIS

We perform this analysis in order to show that with correct highlights detection with HAnS and proper highlights boosting with the proposed ITM, we achieve high quality performance as it is the case with the other well established ITM algorithms from the comparison. For this purpose, on the selected 33 representative image samples from the Stuttgart data-set, we apply every ITM algorithm considered in the comparison. We do this, by sweeping over the wide range of values for the algorithm's specific parameters. For every version of the reconstructed image, we calculate the values of HDR-VDP-3.0.6 (the Q metric), PSNR and SSIM. For every algorithm and every single image sample (from the selected 33 representative images), we select the best results across the used metrics. The results are then averaged and presented on Table 4.

For calculating the HDR-VDP-3.0.6 Q metric, we use the following settings: resolution of 3840 × 2160 pixels, viewing distance of 0.5 meters, diagonal display size of 30 inches, ''rgb-native'' color encoding for HDR images and dynamic range of 700 nits.

The presented results show that the proposed ITM achieves high performance comparable to the state of the art algorithms. This means that with proper detection and boosting of the highlights, with HAnS and the proposed ITM (the both versions: only global highlights boosting and global with local highlights boosting) we successfully obtain the HDR effect.

5) RESULTS FROM THE QUALITATIVE ANALYSIS

Here, we visually present few of the reconstructed images, selected to have high and comparable performance between the ITM algorithms considered in the comparison. We present four interesting examples, for which we show the reconstructed HDR images and the corresponding maps for the top 6 % brightest pixels in the reconstructed images. The visual results are presented on Figures 12, 13, 14 and 15.

From the presented visual examples it can be noticed that the proposed ITM (especially when global and local highlights boosting is applied) succeeds to best pronounce and distinguish most of the highlights (light sources and specular reflections), while still decently boosting the bright areas as it is the case with the other ITM algorithms.

C. SUBJECTIVE EVALUATION

In order to subjectively evaluate the performance of HAnS as part of the ITM (LDR-to-HDR conversion) pipeline, we perform psycho-visual experiment. The intention of this analysis is to show that the local highlights boosting is important for perceiving an HDR effect and that the combination of local and global highlights boosting is perceived as better than only using global highlights boosting. We present the experimental framework for the psycho-visual experiment in Section V-C1, the results from the quantitative analysis in Section V-C2 and the results from the qualitative analysis in Section V-C3.

1) PSYCHO-VISUAL EXPERIMENT

For the psycho-visual experiment 30 diverse in content LDR images were processed in three different ways:

- by applying linear pixel-wise ITM function;
- by applying non-linear pixel-wise boosting ITM function (global highlights boosting);
- by applying non-linear pixel-wise boosting ITM function in combination with the Highlights Maps obtained with HAnS (global and local highlights boosting, see Section IV).

In this experiment 15 volunteers participated. The three differently processed images for every LDR scene, were shown

ITM result: ExpandNet

Top 6 % brightest pixels: ExpandNet

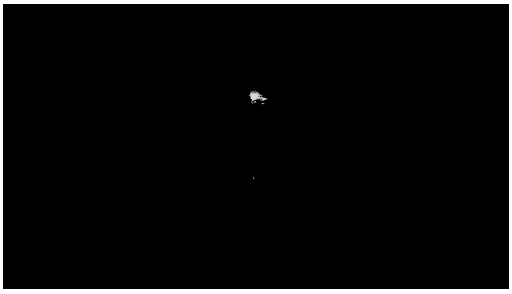
ITM result: Kovaleski and Oliveira

Top 6 % brightest pixels: Kovaleski and Oliveira

ITM result: Meylan et al.

Top 6 % brightest pixels: Meylan et al.

ITM result: proposed ITM (global boosting)



Top 6 % brightest pixels: proposed ITM (global boosting)

ITM result: proposed ITM (global + local boosting)

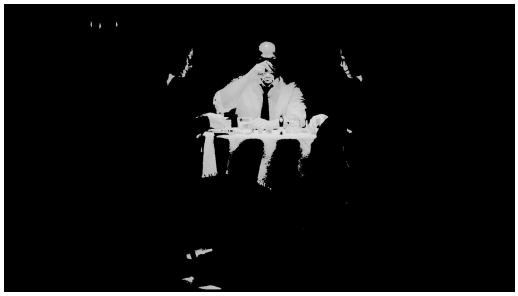
Top 6 % brightest pixels: proposed ITM (global + local boosting)

FIGURE 12. Example 1. Results obtained with different ITM algorithms and maps of the top 6 % brightest pixels in the reconstructed images. Note that the sunlight and its reflection in the water is boosted well with most of the ITMs. Using the proposed ITM with global and local highlights boosting is advantageous over using it with only global highlights boosting.

ITM result: ExpandNet

Top 6 % brightest pixels: ExpandNet

ITM result: Kovaleski and Oliveira

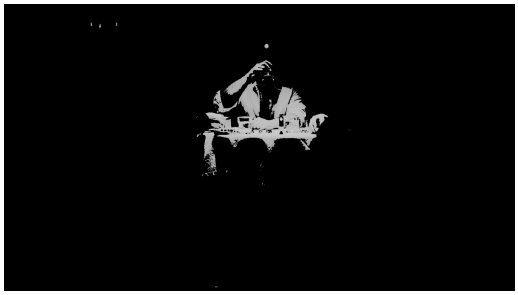


Top 6 % brightest pixels: Kovaleski and Oliveira

ITM result: Meylan et al.

Top 6 % brightest pixels: Meylan et al.

ITM result: proposed ITM (global boosting)



Top 6 % brightest pixels: proposed ITM (global boosting)

ITM result: proposed ITM (global + local boosting)

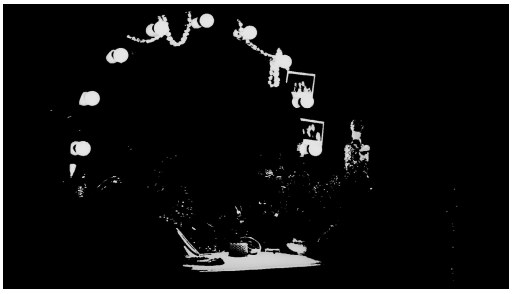
Top 6 % brightest pixels: proposed ITM (global + local boosting)

FIGURE 13. **Example 2. Results obtained with different ITM algorithms and maps of the top 6 % brightest pixels in the reconstructed images. Note that the candle flames are best pronounced with the proposed ITM (global + local boosting). The bright diffuse parts are pronounced with every ITM. With the proposed ITM they are pronounced by the global boosting term.**

ITM result: ExpandNet

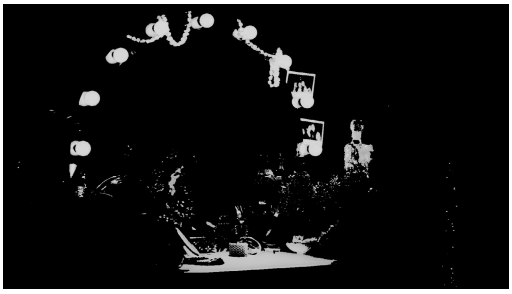
Top 6 % brightest pixels: ExpandNet

ITM result: Kovaleski and Oliveira



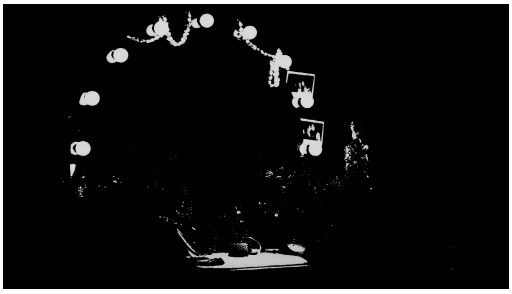
Top 6 % brightest pixels: Kovaleski and Oliveira

ITM result: Meylan et al.



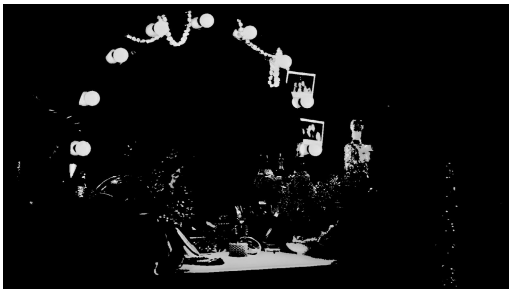
Top 6 % brightest pixels: Meylan et al.

ITM result: proposed ITM (global boosting)



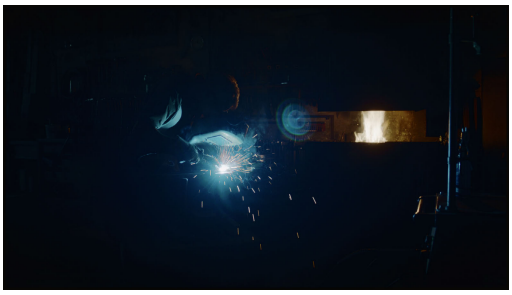
Top 6 % brightest pixels: proposed ITM (global boosting)

ITM result: proposed ITM (global + local boosting)

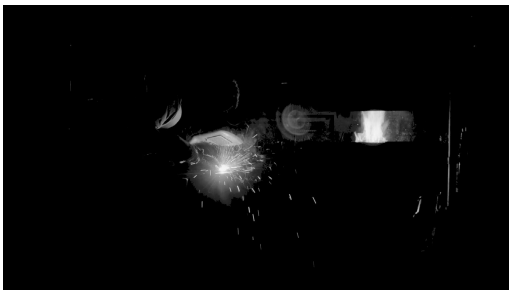


Top 6 % brightest pixels: proposed ITM (global + local boosting)

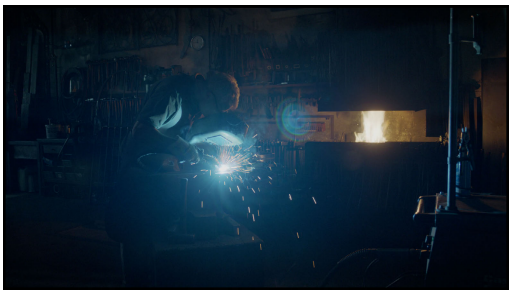
FIGURE 14. Example 3. Results obtained with different ITM algorithms and maps of the top 6 % brightest pixels in the reconstructed images. Note that the light sources and most of the specular reflections are best pronounced with the proposed ITM (global + local boosting), while with the other ITMs they are pronounced with less intensity.



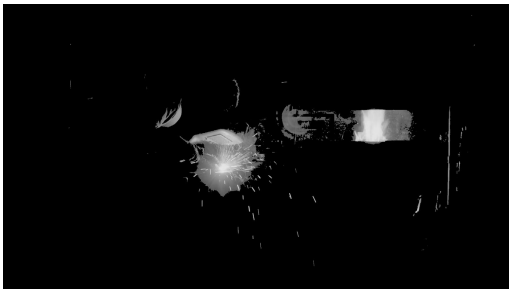
ITM result: ExpandNet



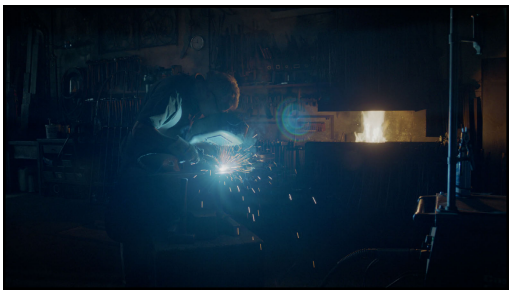
Top 6 % brightest pixels: ExpandNet



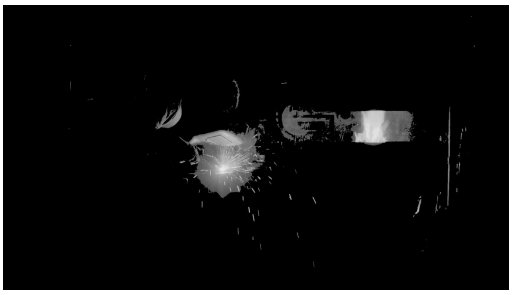
ITM result: Kovaleski and Oliveira



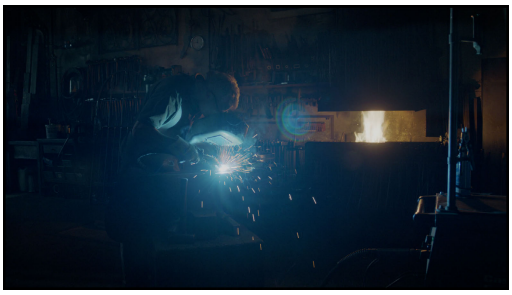
Top 6 % brightest pixels: Kovaleski and Oliveira



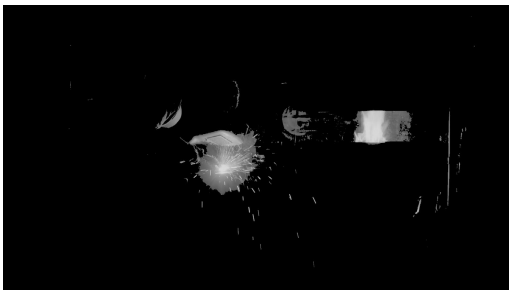
ITM result: Meylan et al.



Top 6 % brightest pixels: Meylan et al.

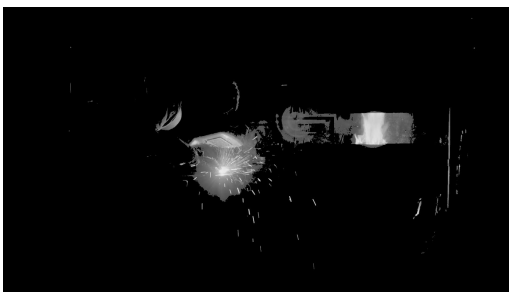


ITM result: proposed ITM (global boosting)



Top 6 % brightest pixels: proposed ITM (global boosting)

ITM result: proposed ITM (global + local boosting)



Top 6 % brightest pixels: proposed ITM (global + local boosting)

FIGURE 15. **Example 4. Results obtained with different ITM algorithms and maps of the top 6 % brightest pixels in the reconstructed images. Note that the welding sparks are well pronounced with ExpandNet and the proposed ITM (global + local boosting). Most welding sparks and specular reflections are pronounced with the proposed ITM (global + local boosting).**

TABLE 5. Quantitative results from the psycho-visual experiment: for global linear pixel-wise ITM function; for global non-linear pixel-wise boosting ITM function and for global non-linear pixel-wise boosting ITM function in combination with the Highlights Maps obtained by HAnS.

Results from the psychovisual experiment: fraction of participants (%) preferring the ITM type in a blind test		
1. Linear pixel-wise ITM	2. Non-linear pixel-wise ITM (global highlights boosting)	3. Proposed ITM (global and local highlights boosting)
10.88	19.56	69.56

sequentially in random order (for the purpose of performing a blind test), to each participant in the psycho-visual experiment. For every scene, each participant was asked to make a choice of subjective preference HDR version, without having a reference image, and only relying on his/her subjective opinion for most pleasant content. The three HDR versions, were repeatedly shown three times in the same order, in order to facilitate the decision making and to allow the participant to remember the presented content.

The independent variables (IVs) in the performed experiment are the participants and the order of the presented ITM reconstructed versions per image sample. We considered 15 participants out of which: 4 experts and 11 non-experts in HDR imaging; 9 males and 6 females, 7 coming from Belgium and 8 coming from countries in Europe, South America and Asia. The participants are between 25 and 40 years old. The order for the three presented ITM reconstructed versions for every image sample and participant was random (uniform distribution).

In the performed experiment we measured (dependent variable - DV) votes per ITM reconstructed version. We counted the single votes from every participant for every image sample, per reconstructed ITM version. The votes are then summarized over the image samples and participants and normalized with the product of the number of participants and the number of image samples. Given that the test is blind (the participants don't know which reconstructed HDR version is presented), the contribution of each participant (15 participants) per image sample (30 image samples) is equal to 0.22 % with a probability of 33.33 % to select one of the 3 presented reconstructed versions as best. Hypothetically, if one of the 15 participants, always selects the same version (as best) for all 30 image samples, his contribution in the obtained results would be 6.67 % for the preferred version and that would be the extreme case scenario. We consider this fraction sufficient for the accuracy of this experiment.

The control variables (CVs) are the selected image samples (diverse in content: scenes vary from dark to bright, scenes abundant with natural content, indoor and outdoor scenes, scenes extracted from movies, scenes from TV shows with artificial lights, scenes with diverse in color and size light sources, scenes with people and visible faces and skin), the number of image samples (30 image samples), the time allocated for one session per participant (30 minutes), observation distance (0.5 meters), type of screen and resolution (LDR screen of HD resolution which is sufficient for the use-case of our experiment), contrast and brightness settings of the screen (set to default), ITM parameters (previously tuned for every image sample so that they remain unchanged during the experiment), room lighting (uniform ambient light: the light is the same at the view-point of the participant and where the screen is positioned).

2) RESULTS FROM THE QUANTITATIVE ANALYSIS

Here, we present the results (see Table 5) from the performed psycho-visual experiment. Most of the participants ($\approx 70\,\%$) preferred the results that are obtained with the proposed ITM approach presented in Section IV (global and local highlights boosting), over the results obtained with global boosting.

In Fig.17 we visually present some of the results shown to the participants in the psycho-visual experiment.

3) RESULTS FROM THE QUALITATIVE ANALYSIS

Considering the results that are obtained with linear pixel-wise ITM function as not interesting (see results shown on Table 5), here we visually present only the results when the ITM is applied in two different ways: with a non-linear ITM function (global highlights boosting, see Section IV) and with non-linear ITM function combined with the output Highlights Maps from HAnS (global and local highlights boosting, see Section IV).

First in Fig.16, we show the input LDR images (as observed on an LDR screen) and the obtained Highlights Maps. Afterwards, with the results from inverse tone-mapping that are presented on Fig.17 (simulated as being observed on an HDR screen) we make visual comparison between the two cases of applied ITM.

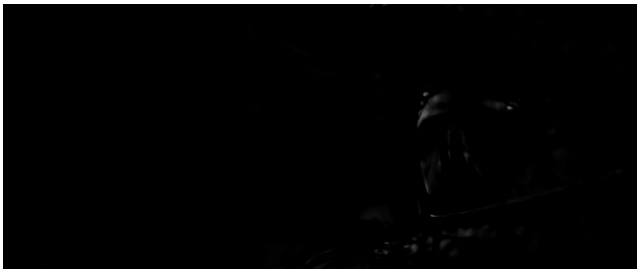
From the Highlights Maps presented in Fig.16, we notice that the specular reflections and the light sources (among which there are color saturated lights and color saturated specular reflections) are successfully detected by HAnS, while at the same time most of the areas belonging to bright diffuse surfaces are excluded.

From the visual results presented in Fig. 17, it can be noticed that there is a clear distinction between the results obtained with only global highlights boosting and the results obtained with global and local highlights boosting and moreover, the results obtained with global and local highlights boosting are more pleasant to the observer.

Input LDR image (example 1)

Highlights Map (example1)

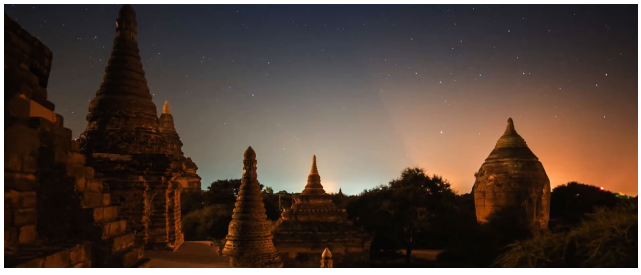
Input LDR image (example 2)



Highlights Map (example2)

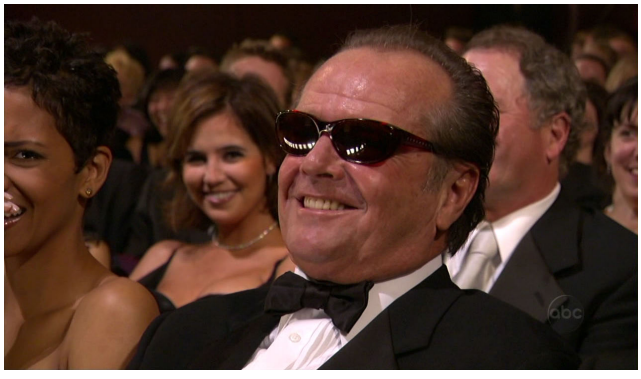
Input LDR image (example 3)

Highlights Map (example3)

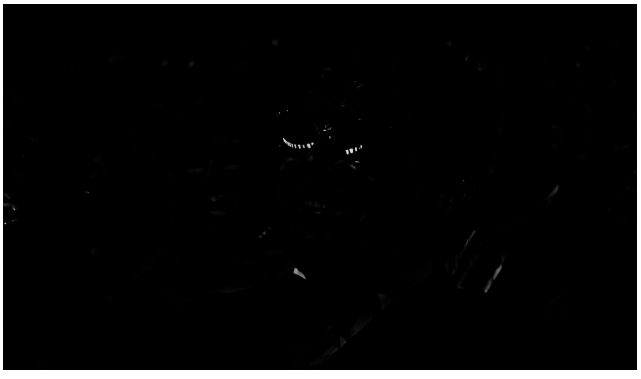


Input LDR image (example 4)

Highlights Map (example4)



Input LDR image (example 5)



Highlights Map (example5)

FIGURE 16. **Input LDR images (as observed on an LDR screen) and the corresponding Highlights Maps generated with HAnS.**

ITM result for example 1: global boosting

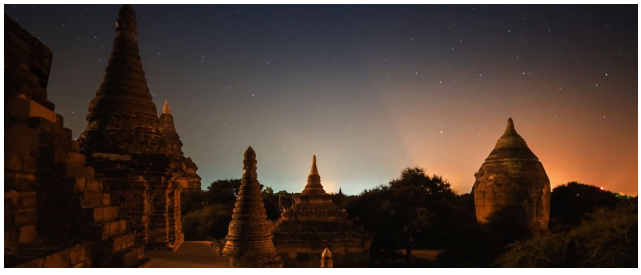
ITM result for example 1: global and local boosting

ITM result for example 2: global boosting

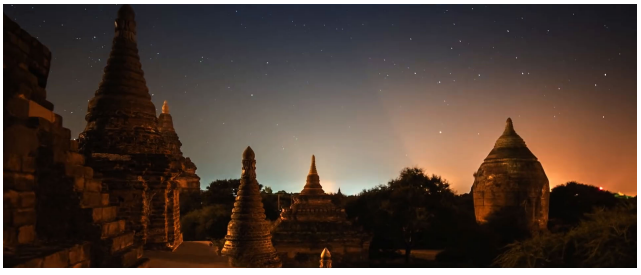
ITM result for example 2: global and local boosting

ITM result for example 3: global boosting

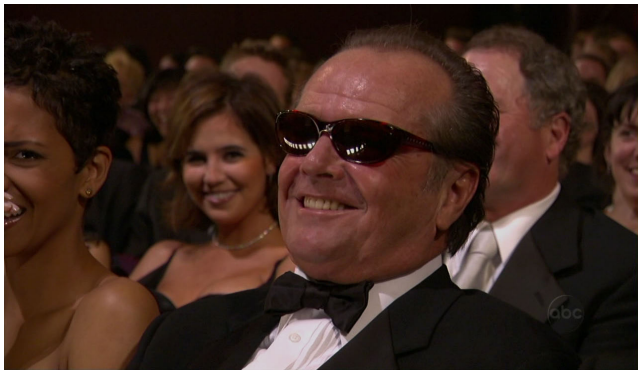
ITM result for example 3: global and local boosting



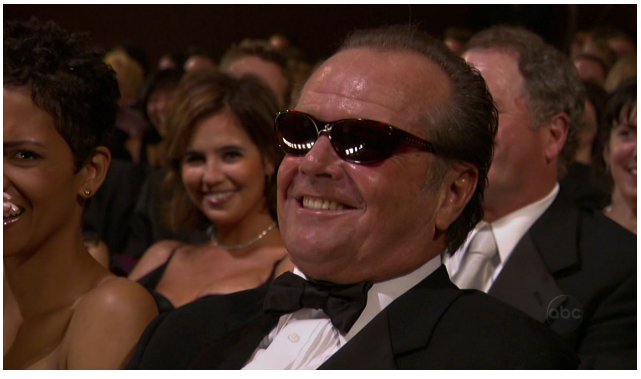
ITM result for example 4: global boosting



ITM result for example 4: global and local boosting



ITM result for example 5: global boosting



ITM result for example 5: global and local boosting

FIGURE 17. **Results (simulated as being observed on a HDR screen) obtained with ITM: when only global boosting is applied; when global and local boosting (with the Highlights Maps presented in Fig.16) is applied.**

From both quantitative and qualitative results from the subjective evaluation analysis (see Table 5 and Fig.17), we conclude that global and local highlights boosting is undoubtedly more advantageous than only global highlights boosting in creating the HDR effect. Hereby, we justify the advanteougness of HAnS in LDR-to-HDR conversion of cinematic LDR content.

VI. CONCLUSION

Considering low hardware requirements and simple and efficient implementation, we propose a new algorithm for detecting specular reflections and light sources in cinematic LDR content. The proposed highlights detection algorithm (HAnS) is devised as part of an inverse-tone mapping approach for the purposes of the LDR-to-HDR conversion.

The proposed algorithm for highlights detection is evaluated objectively, on two image data-sets consisted of LDR content. Additionally, HAnS with the proposed ITM, is objectively and subjectively evaluated as part of a complete LDR-to-HDR pipeline.

The results from the objective evaluation analysis on highlights detection show that HAnS outperforms the classic highlights detection algorithms, both quantitatively and qualitatively. These results also show that HAnS is successful in avoiding false positive detections, while at the same time (with the appropriate selection of the controlling parameters), is able to correctly detect specular reflections and light sources not only in cinematic LDR content, but also in LDR content in general (SPEC-DB image data-set).

The results from the objective evaluation on highlights boosting as part of a complete LDR-to-HDR conversion pipeline, show that HAnS with the proposed ITM achieves high quality performance, already established with the ITM algorithms included in the comparison. The presented results (quantitative and qualitative) justify use of HAnS in a complete LDR-to-HDR conversion pipeline.

The results from the subjective evaluation analysis, both quantitatively and qualitatively, show that the usage of HAnS as part of simple ITM approach for global and local highlights boosting, produces more pleasant content than the conventional global pixel-wise boosting ITM approaches.

The implementation of HAnS with the proposed ITM approach, was adapted for FPGA and it is already validated, approved and released as a complete product. This fact is a testament to the efficacy of using the proposed approach for highlights detection and boosting in an LDR-to-HDR conversion pipeline.

APPENDIX A
OBJECT-LEVEL SEGMENTATION METRICS USED IN THE OBJECTIVE EVALUATION

Here, we describe the object-level segmentation metrics that we use for objective evaluation of the highlights detection algorithms.

A. MARTIN'S ERROR METRIC

The Martin's error metric is used for measuring the qualitative similarity between man-made segmentations, which are expected to be different. In order to present the details for calculating this error metric we will refer to [48] and use a similar set notation. Let $I_R = \{R_1, R_2, R_3, \ldots, R_M\}$ represent the binary reference image (ground-truth image obtained from one annotator) with R_j being the j-th fully-connected segment in I_R. Similarly, $I_S = \{S_1, S_2, S_3, \ldots, S_N\}$ is the binary image obtained by the highlights detection algorithms with S_i being the i-th fully connected segment in I_S. As described in [48], the errors P_{ji} and Q_{ji} between segments R_j and S_i are presented with (17) and (18), where $\setminus$, $\cap$ and $|.|$, are the set difference operator, the intersection operator and the number of elements in a set operator, respectively.

$$P_{ji} = \frac{|R_j \setminus S_i|}{|R_j|} \times |R_j \cap S_i| \tag{17}$$

$$Q_{ji} = \frac{|S_i \setminus R_j|}{|S_i|} \times |R_j \cap S_i| \tag{18}$$

The number of pixels in the complete intersection region between the two binary images is calculated by $n = \sum_{j=1}^{M} \sum_{i=1}^{N} |R_j \cap S_i|$. The Martin's error metric is defined with two calculated errors, global consistency error (GCE) and local consistency error (LCE). These are calculated using (19) and (20), appropriately.

$$GCE(I_R, I_S) = \frac{1}{n} \min \left\{ \sum_{j=1}^{M} \sum_{i=1}^{N} P_{ji}, \sum_{j=1}^{M} \sum_{i=1}^{N} Q_{ji} \right\} \tag{19}$$

$$LCE(I_R, I_S) = \frac{1}{n} \sum_{j=1}^{M} \sum_{i=1}^{N} \min \left(P_{ji}, Q_{ji}\right) \tag{20}$$

In cases where no region is annotated and yet there are detected segments in the result from the detection algorithms (cases with false positives: $I_R = \emptyset$ and $I_S \neq \emptyset$), and vice versa (cases with false negatives: $I_R \neq \emptyset$ and $I_S = \emptyset$), we assign the highest value of 1 to these errors, i.e. $GCE(I_R, I_S) = 1$ and $LCE(I_R, I_S) = 1$. In order to preserve the matching direction symmetry the false negative cases are treated equally as the false positive cases.

B. OBJECT CONSISTENCY ERROR

Similarly to Martin's error metric, the object consistency error metric (OCE), as proposed in [48], performs object by object comparison of the segmented image and the reference image. It takes into account the size, the shape and the position of each fully connected segment. This error is normalized within a range [0,1], where 0 means that there is no error (i.e. the compared images are equal) and 1 when there is no match between the reference and the segmented image. This metric, also has properties of being symmetric and scale invariant. The authors of [48], define the partial error measure

$E_{R,S}(I_R, I_S)$, calculated using (21), (22) and (23),

$$E_{R,S}(I_R, I_S) = \sum_{j=1}^{M} \left[1 - \sum_{i=1}^{N} \frac{|R_j \cap S_i|}{|R_j \cup S_i|} \times W_{ji} \right] W_j \quad (21)$$

$$W_{ji} = \frac{\bar{\delta}\left(|R_j \cap S_i|\right)|S_i|}{\sum_{k=1}^{N} \bar{\delta}\left(|R_j \cap S_k|\right)|S_k|} \quad (22)$$

$$W_j = \frac{|R_j|}{\sum_{l=1}^{M} |R_l|} \quad (23)$$

where W_{ji} weighs each S_i that intersects with R_j, according to the size of S_i relative to all segments in I_S that intersect with R_j. Additionally, W_j weighs the importance of R_j, relative to all the segments in the reference image. In (22), $\bar{\delta}(x)$ is a function that equals 0 for input being 0, and 1 otherwise. The object consistency error is then calculated using (24).

$$OCE(I_R, I_S) = \min\left(E_{R,S}, E_{S,R}\right) \quad (24)$$

In the same way as it was done for the Martin's error metric, we adapted the OCE metric to have the highest penalty (=1), for false positive cases, $I_R = \emptyset$ and $I_S \neq \emptyset$, and for false negative cases, $I_R \neq \emptyset$ and $I_S = \emptyset$.

C. QUALITY RATE

Another object by object comparison metric is the quality rate (QR) metric, proposed by Weidner [49]. We use this metric, as it is proposed in [46], where the authors consider the matching direction. Therefore, two versions of this metric are introduced, QR_{SR} given with (25), for matching the segments to the reference objects and QR_{RS} given with (26) for matching the reference objects to the segments.

$$QR_{SR}(I_R, I_S) = \sum_{j=1}^{M} \frac{|R_j \cap S_{max}| \times |R_j|}{|R_j \cup S_{max}| \times |R|} \quad (25)$$

$$QR_{RS}(I_R, I_S) = \sum_{i=1}^{N} \frac{|S_i \cap R_{max}| \times |S_i|}{|S_i \cup R_{max}| \times |S|} \quad (26)$$

Its values are within the range [0,1], with 1 for the best match between the segments and the reference objects, and 0 when there is no match. We extend this metric to have highest value (=1), when $I_R = \emptyset$ and $I_S = \emptyset$.

APPENDIX B
EVALUATING THE PERFORMANCE OF THE HLD ALGORITHMS ON THE PROPOSED IMAGE DATA-SET

Here, we describe how the evaluation of the HLD algorithms was conducted on the proposed image data-set consisted of cinematic LDR content and real world photographs.

In this data-set only extracted patches from the images are annotated for highlights, and not the whole content. This is because the process of annotating the complete content of the images would be too difficult and inefficient for the annotators.

First the images are processed with the HLD algorithms (with different settings of the controlling parameters) in their full size. Afterwards, for evaluation, in order to make the comparison with the annotated reference patches, we extract the corresponding patches of the results obtained with the HLD algorithm and calculate the value of each metric per extracted patch. Then, we calculate the average value for every metric, over the patches considered in one single image that is processed (by the HLD algorithm) for every fixed setting of the controlling parameters. The best setting of controlling parameters for every HLD algorithm and every image, is then selected and the corresponding results for every patch and every metric stored. In some cases the patch is small and consists of small highlights and/or non-highlights areas and in other cases the patch is big and content-wise it covers areas of the whole object with various specularity properties. The obtained evaluation results for every metric and every patch, are then averaged over the whole data-set of 960 patches, for every subset of ground-truth annotations. The metrics, with the average and the standard deviation are presented in Table 2.

ACKNOWLEDGMENT

The authors would like to thank to all participants in the psycho-visual experiment performed for the purpose of the subjective evaluation, Gonzalo Raimundo Luzardo Morocho for assisting in the creation of the novel image data-set, annotated for highlights in cinematic LDR content, and TP Vision for their participation and open collaboration in this research. This work was part of a joint project for dynamic range expansion, between Ghent University and TP Vision.

REFERENCES

[1] A. Stojkovic, J. Aelterman, H. Luong, H. Van Parys, L. Jovanov, and W. Philips, "How size and relative contrast can improve specular highlight detection," in *Proc. 14th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2018, pp. 258–265.

[2] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Inverse tone mapping," in *Proc. 4th Int. Conf. Comput. Graph. Interact. Techn. Australasia Southeast Asia*, 2006, pp. 349–356.

[3] L. Meylan, "Tone mapping for high dynamic range images," Ph.D. dissertation, École Polytechnique Fédérale Lausanne, Lausanne, Switzerland, 2006.

[4] L. Meylan, S. Daly, and S. Süsstrunk, "Tone mapping for high dynamic range displays," *Proc. SPIE*, vol. 6492, Feb. 2007, Art. no. 649210.

[5] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, "Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs," *ACM Trans. Graph.*, vol. 26, no. 3, p. 39, 2007.

[6] F. Banterle, P. Ledda, K. Debattista, M. Bloj, A. Artusi, and A. Chalmers, "A psychophysical evaluation of inverse tone mapping techniques," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 13–25, Mar. 2009.

[7] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2010.

[8] F. Banterle, *Advanced High Dynamic Range Imaging: Theory and Practice*. Natick, MA, USA: AK Peters, 2011.

[9] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging*. Boca Raton, FL, USA: CRC Press, 2017.

[10] R. P. Kovaleski and M. M. Oliveira, "High-quality reverse tone mapping for a wide range of exposures," in *Proc. 27th SIBGRAPI Conf. Graph., Patterns Images*, Aug. 2014, pp. 49–56.

[11] Y. Huo and F. Yang, "High-dynamic range image generation from single low-dynamic range image," *IET Image Process.*, vol. 10, no. 3, pp. 198–205, Mar. 2016.

[12] Y. Endo, Y. Kanamori, and J. Mitani, "Deep reverse tone mapping," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–177, 2017.
[13] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep CNNs," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–15, Nov. 2017.
[14] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content," *Comput. Graph. Forum*, vol. 37, no. 2, pp. 37–49, May 2018.
[15] G. Luzardo, J. Aelterman, H. Luong, W. Philips, and D. Ochoa, "Real-time false-contours removal for inverse tone mapped HDR content," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 1472–1479.
[16] G. Luzardo, J. Aelterman, H. Luong, W. Philips, D. Ochoa, and S. Rousseaux, "Fully-automatic inverse tone mapping preserving the content Creator's artistic intentions," in *Proc. Picture Coding Symp. (PCS)*, Jun. 2018, pp. 199–203.
[17] G. Luzardo, T. Vyvey, J. Aelterman, T. Paridaens, G. Van Wallendael, P. Lambert, S. Rousseaux, H. Luong, W. Durnez, J. Van Looy, W. Philips, and D. Ochoa, "An experimental study on the perceived quality of natively graded versus inverse tone mapped high dynamic range video content on television," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 5559–5576, Feb. 2021.
[18] G. Luzardo, J. Aelterman, H. Luong, S. Rousseaux, D. Ochoa, and W. Philips, "Fully-automatic inverse tone mapping algorithm based on dynamic mid-level tone mapping," *APSIPA Trans. Signal Inf. Process.*, vol. 9, no. E7, Feb. 2020, doi: 10.1017/ATSIP.2020.5.
[19] D. Makkar and M. Malhotra, "Single image haze removal using dark channel prior," *Int. J. Eng. Comput. Sci.*, vol. 29, no. 5, pp. 835–849, Jan. 2016.
[20] S. A. Shafer, "Using color to separate reflection components," *Color Res. Appl.*, vol. 10, no. 4, pp. 210–218, 1985.
[21] G. J. Klinker, S. A. Shafer, and T. Kanade, "The measurement of highlights in color images," *Int. J. Comput. Vis.*, vol. 2, no. 1, pp. 7–32, Jun. 1988.
[22] R. T. Tan and K. Ikeuchi, "Separating reflection components of textured surfaces using a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 178–193, Feb. 2005.
[23] K.-J. Yoon, Y. Choi, and I. S. Kweon, "Fast separation of reflection components using a specularity-invariant image representation," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 973–976.
[24] H.-L. Shen, H.-G. Zhang, S.-J. Shao, and J. H. Xin, "Chromaticity-based separation of reflection components in a single image," *Pattern Recognit.*, vol. 41, no. 8, pp. 2461–2469, Aug. 2008.
[25] H.-L. Shen and Q.-Y. Cai, "Simple and efficient method for specularity removal in an image," *Appl. Opt.*, vol. 48, no. 14, pp. 2711–2719, 2009.
[26] H.-L. Shen and Z.-H. Zheng, "Real-time highlight removal using intensity ratio," *Appl. Opt.*, vol. 52, no. 19, pp. 4483–4493, 2013.
[27] Q. Yang, S. Wang, and N. Ahuja, "Real-time specular highlight removal using bilateral filtering," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 87–100.
[28] H. Kim, H. Jin, S. Hadap, and I. Kweon, "Specular reflection separation using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1460–1467.
[29] Y. Akashi and T. Okatani, "Separation of reflection components by sparse non-negative matrix factorization," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 611–625.
[30] A. Artusi, F. Banterle, and D. Chetverikov, "A survey of specularity removal methods," *Comput. Graph. Forum*, vol. 30, no. 8, pp. 2208–2230, Dec. 2011.
[31] L. B. Wolff, "Relative brightness of specular and diffuse reflection," *Opt. Eng.*, vol. 33, no. 1, pp. 285–294, 1994.
[32] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
[33] S. N. Youseph and R. R. Cherian, "Pixel and edge based illuminant color estimation for image forgery detection," *Procedia Comput. Sci.*, vol. 46, pp. 1635–1642, Jan. 2015.
[34] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.
[35] J. van de Weijer and T. Gevers, "Color constancy based on the grey-edge hypothesis," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2005, p. II-722.
[36] J. B. Park and A. C. Kak, "A truncated least squares approach to the detection of specular highlights in color images," in *Proc. IEEE Int. Conf. Robot. Automat.*, vol. 1, Sep. 2003, pp. 1397–1403.
[37] J. B. Park and A. Kak, "A new color representation for non-white illumination conditions," School Elect. Eng., Purdue Univ., West Lafayette, IN, USA, Tech. Rep. TR-ECE-05-06, Jun. 2005.
[38] *Purdue RVL SPEC-DB Image Data-Set*. Accessed: Jan. 2021. [Online]. Available: https://engineering.purdue.edu/RVL/Database/specularity_database/gallery.htm
[39] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
[40] H.-F. Ng, "Automatic thresholding for defect detection," *Pattern Recognit. Lett.*, vol. 27, no. 14, pp. 1644–1649, Oct. 2006.
[41] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993.
[42] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons," *Biologiske Skrifter*, vol. 5, pp. 1–34, Jun. 1948.
[43] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.
[44] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et du Jura," *Bull. Soc. Vaudoise Sci. Nat.*, vol. 37, pp. 547–579, Jan. 1901.
[45] P. Jaccard, "The distribution of the flora in the alpine zone.1," *New Phytologist*, vol. 11, no. 2, pp. 37–50, Feb. 1912.
[46] X. Zhang, X. Feng, P. Xiao, G. He, and L. Zhu, "Segmentation quality evaluation using region-based precision and recall measures for remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 73–84, Apr. 2015.
[47] D. R. Martin, J. Malik, and D. Patterson, "An empirical approach to grouping and segmentation," Dept. Comput. Sci. Division, Univ. California, Berkeley, Berkeley, CA, USA, Tech. Rep. UCB/CSD-3-1268, Aug. 2003.
[48] M. Polak, H. Zhang, and M. Pi, "An evaluation metric for image segmentation of multiple objects," *Image Vis. Comput.*, vol. 27, no. 8, pp. 1223–1227, Jul. 2009.
[49] U. Weidner, "Contribution to the assessment of segmentation quality for remote sensing applications," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 37, no. B7, pp. 479–484, 2008.
[50] J. S. Cardoso and L. Corte-Real, "Toward a generic evaluation of image segmentation," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1773–1782, Nov. 2005.
[51] *National Public-Service Broadcaster for the Flemish Region and Community of Belgium VRT*. Accessed: Jan. 2021. [Online]. Available: https://www.vrt.be
[52] *GitHub Repository for ExpandNet*. Accessed: Jan. 2021. [Online]. Available: https://github.com/dmarnerides/hdr-expandnet
[53] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, "Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays," *Proc. SPIE*, vol. 9023, Mar. 2014, Art. no. 90230X.
[54] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–14, 2011.
[55] *HDR Visual Difference Predictor*. Accessed: Jan. 2021. [Online]. Available: https://sourceforge.net/projects/hdrvdp/files/hdrvdp/3.0.6/
[56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

ANA STOJKOVIC was born in Skopje, Macedonia, in 1989. She received the B.S. and M.S. degrees in electrical engineering and information technologies from Ss. Cyril and Methodius University, Skopje, in 2012 and 2016, respectively. She is currently pursuing the Ph.D. degree in computer science engineering with Ghent University, Ghent, Belgium. She is currently working as a Doctoral Researcher with the imec Research Group Image Processing and Interpretation (IPI), Department of TELIN, Faculty of Engineering, Ghent University. Her research interests include signal processing, image and video processing, image and video analysis, image and video reconstruction, HDR imaging, computer vision, machine learning, and speech processing.

JAN AELTERMAN received the Ph.D. degree in engineering from Ghent University, in 2014. He has been working as a Postdoctoral Researcher with the Image Processing and Interpretation (IPI) Group, an imec Research Group, Ghent University, where he is currently a Professor. His research interests include image and video restoration, reconstruction, and estimation problems in application fields, such as (HDR) consumer video, MRI, CT, (electron) microscopy, photography, and multi-view processing.

HIEP LUONG received the Ph.D. degree in computer science engineering from Ghent University, in 2009. He has been working as a Postdoctoral Researcher and the Project Manager of the Image Processing and Interpretation (IPI), an imec Research Group, Ghent University. He is currently a Professor at IPI and leads the UAV Research Centre of UGhent. His research interests include image and real-time video processing for various fields, such as HDR imaging, (bio) medical imaging, depth and multi-view processing, and multi-sensor fusion for UAV and AR applications.

HANS VAN PARYS received the M.S. degree in electrical and electronics engineering from Ghent University, in 1991. He is currently an Electrical Engineer Specialist in video architectures, video, and image processing, and experienced in hardware and software development. He has worked at Barco, Kortrijk, and Philips, Brugge/TP Vision, Ghent. He is also a Senior Research and Development Engineer at Televic Conference. His research interests include digital signal processing, mathematical computing, data science, and machine learning.

WILFRIED PHILIPS was born in Aalst, Belgium, in October 1966. He received the Diploma and Ph.D. degrees in electrical engineering from the University of Ghent, Belgium, in 1989 and 1993, respectively. Since October 1989, he has been working with the Department of Electronics and Information Systems, Ghent University, as a Research Assistant for the Flemish Fund for Scientific Research (FWO). He is also a Professor with Ghent University and leads the IPI-imec Research Group at TELIN, UGhent. His main research interests include image and video restoration, image analysis, and lossless and lossy data compression of images and video, and processing of multimedia data.