Automatic exposure control for video cameras towards HDR techniques

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Research Article

Automatic Level Control for Video Cameras towards HDR Techniques

Sascha Cvetkovic,[1] Helios Jellema,[1] and Peter H. N. de With[2, 3]

[1] *Bosch Security Systems, 5616 LW Eindhoven, The Netherlands*
[2] *Department of Electrical Engineering, University of Technology Eindhoven, 5600 MB Eindhoven, The Netherlands*
[3] *CycloMedia Technology, 4181 AE Waardenburg, The Netherlands*

Correspondence should be addressed to Sascha Cvetkovic, sacha.cvetkovic@nl.bosch.com

We give a comprehensive overview of the complete exposure processing chain for video cameras. For each step of the automatic exposure algorithm we discuss some classical solutions and propose their improvements or give new alternatives. We start by explaining exposure metering methods, describing types of signals that are used as the scene content descriptors as well as means to utilize these descriptors. We also discuss different exposure control types used for the control of lens, integration time of the sensor, and gain control, such as a PID control, precalculated control based on the camera response function, and propose a new recursive control type that matches the underlying image formation model. Then, a description of commonly used serial control strategy for lens, sensor exposure time, and gain is presented, followed by a proposal of a new parallel control solution that integrates well with tone mapping and enhancement part of the image pipeline. Parallel control strategy enables faster and smoother control and facilitates optimally filling the dynamic range of the sensor to improve the SNR and an image contrast, while avoiding signal clipping. This is archived by the proposed special control modes used for better display and correct exposure of both low-dynamic range and high-dynamic range images. To overcome the inherited problems of limited dynamic range of capturing devices we discuss a paradigm of multiple exposure techniques. Using these techniques we can enable a correct rendering of difficult class of high-dynamic range input scenes. However, multiple exposure techniques bring several challenges, especially in the presence of motion and artificial light sources such as fluorescent lights. In particular, false colors and light-flickering problems are described. After briefly discussing some known possible solutions for the motion problem, we focus on solving the fluorescence-light problem. Thereby, we propose an algorithm for the detection of fluorescent lights from the image itself and define a set of remedial actions, to minimize false color and light-flickering problems.

1. Introduction

A good video-level control is a fundamental requirement for any high-performance video camera. (By video-level control, we mean the control of the image luminance level, often referred to as exposure control. However, since we are also controlling exposure time of the sensor and value of the gain, instead of exposure, we will use the term video level.) The reason is that this function provides a basis for all the subsequent image processing algorithms and tasks, and as such it is a pre-requisite for a high image quality. With "high quality" we mean that we pursue a high-fidelity output image, where all relevant scene details have a good visibility and the image as a whole conveys sufficient scene context and information for good recognition. This paper gives an overview of the complete exposure processing chain and presents several improvements for that chain. Our improvements are applicable for both standard as well as high-dynamic range image processing pipelines.

In practice, high-performance imaging should give a good quality under difficult circumstances, that is, for both high- and low-dynamic range scenes. It will become clear that special signal processing techniques are necessary for correct rendering of such scenes. The required image processing functions involved with "standard concepts of exposure control" are, for example, iris control, sensor

integration time and gain control. These functions have to be combined with signal processing tasks such as tone mapping, image enhancement, and multiple exposure techniques. Summarizing, the integration involves therefore the marriage of both exposure techniques and advanced processing. This brings new challenges which will be addressed in this paper.

It is evident that a good image exposure control starts with a good exposure metering system, performing stable and correct control and improving image fidelity. It should also align well with tone mapping and enhancement control. The discussed techniques have received little attention in publications, while a good exposure control is at least as important as all the other stages of the image processing chain. First, the inherent complexity of the complete imaging system is large. This system includes camera, lens, peripheral components, software, signal transport and display equipment which were not optimized and matched with each other and are having large tolerances and deviations. Therefore, it becomes increasingly difficult to design a viable video-level control that guarantees a good "out-of-the-box" performance in all cases. Second, cameras have to operate well regardless of the variable and unknown scene conditions for many years.

The discussed themes are built up from the beginning. In Section 2, we start with an introductory part of the video-level system, where we describe the exposure metering methods. This gives ideas "where and what" to measure. We will only consider digital exposure measurement techniques that are performed on the image (video) signal self (so called trough-the lens) and do not use additional sensors. Section 3 discusses the types of signals that are used as the scene content descriptors as well as means to utilize these descriptors. From that discussion, we adopt signal types of which the typical examples are the average, median, and peak-white luminance levels within measured image areas. These measurements are used to control the iris, exposure time of the sensor, and gain of the camera, where each item is controlled in a specific way for obtaining a high quality. Then, in Section 4, we discuss different video-level control types used for the control of the lens, integration time of the sensor, and gain control, such as a PID control, precalculated control based on the camera response function and recursive control.

Afterwards, we develop control strategies to optimize the overall image delivery of the camera, for example, by optimizing the SNR, stability of operation under varying conditions, and avoiding switching in operational modes. The purpose of these discussions breaks down in several aspects. The main question that is addressed is the design and operation of image level control algorithms and a suitable overall control strategy, to achieve stable, accurate, and smooth level control, avoiding switching in operational modes and enabling subsequent perceptual image improvement. The output image should have as good SNR as possible and signal clipping should be avoided, or only introduced in a controllable fashion. The level control strategy should provide a good solution for all types of images/video signals, including low-, medium-, and high-dynamic range images. One of the problems in the control system is the lens, as it

has unknown transfer characteristics, the lens opening is not known, and the involved mechanical control is unpredictable in accuracy and response time. As already mentioned, many other parameters need to be controlled as well, so that a potentially attractive proposal would be to control those parameters all in parallel and enforce an overall control stability, accuracy, and speed. The design of such a parallel control system, combined with a good integration with the tone mapping and enhancement part of the image pipeline, is one of the contributions of this paper, which will be presented in Section 5. The presentation of the novel design is preceded by a standard overall control strategy for lens, exposure time, and gain.

Section 6 is devoted to exploiting the full dynamic range of the signal under most circumstances. For this reason we develop specific means to further optimize the visibility of important scene objects, the amount of signal clipping, and the dynamic range. We have found that these specific means are more effective with a parallel control system. We present three subsections on those specific means of which two contain new contributions from our work. The first subsection of Section 6 is containing an overview of level control for standard cases and does not contain significant new work. It starts with an overview of existing typical solutions and strategies used for determining the optimal level control of HDR images in standard video processing pipelines and cameras. These proposals overexpose the complete image to enable visualization of important dark foreground objects. The key performance indicator in these scenarios is how good can we distinguish the important foreground objects from unimportant background regions. However, these approaches come with high complexity, and even though they can improve visibility of important objects for many HDR scene conditions, there are always real-life scenes where they fail. Another disadvantage is that clipping occurs in the majority of the bright parts of the displayed image. However, for standard dynamic range video cameras, this is the only available strategy.

The second subsection of Section 6 presents the saturation control strategy to optimize the overall image delivery of the camera, with the emphasis on an improved SNR and global image contrast. The third subsection of Section 6 discusses the control of the amount of signal clipping. After presenting the initial clipping solution, thanks to the saturation control, we propose a better solution for signal clipping control. It can be intuitively understood that when the saturation control is operating well, the clipping of the peak signal values can be more refined, making less annoying artifacts. The principle is based on balancing between the highest-dynamic range with the limited amount of clipping. These special modes in combination with multiple-exposure techniques will prepare the camera signal for the succeeding steps in the processing on tone mapping and enhancement functionalities, which are discussed in the remainder of this paper.

The last part of this paper is devoted to high-dynamic range imaging. We have previously described handling of the high-dynamic range scenes for standard dynamic range image pipelines. The primary disadvantage of these

procedures is that clipping of the signal is introduced due to overexposing of the bright background for the visualization of dark foreground (or vice versa). By employing HDR techniques for extending the sensor dynamic range, we can achieve better results without introducing additional signal clipping. In particular, we can optimize the image delivery by using the video-level control to reduce or completely remove any signal clipping. Although very dark, because of exposure bracketing, the resulting image will have sufficient SNR for further tone mapping and visualization of all image details.

Section 7 first briefly introduces several techniques used for obtaining HDR images and describes some of their drawbacks. In particular, we are concerned by image fidelity and color distortions introduced by nonlinear methods of HDR creation. This is why we focus on the exposure bracketing, since this is currently the only visible HDR solution for the real-time camera processing in terms of costperformance. However, this technique also has certain drawbacks and challenges, such as motion in the scene and the influence of light coming from non-constant light sources. In Section 8 we focus on the problems originating from artificial light sources such as fluorescent lights and propose two solutions for their handling. By presenting some experimental results, we show the robustness of our solution and demonstrate that this is a very difficult problem. Finally, we give some hints and conclude this paper in Section 9.

2. Metering Areas

Each exposure control algorithm starts with exposure metering. We will discuss three metering systems which are used depending on the application or camera type. In some cases, they can even be used simultaneously, or as a fall-back strategy if one metering system provides unreliable results.

2.1. Zone Metering Systems. The image is divided in a number of zones (sometimes several hundred) where the intensity of the video signal is measured individually. Each image zone has its own weight and the contributions of them are mostly combined into one output average measurement. Higher weights are usually assigned to the central zones (center-weighted average metering [1, 2]) or zones in the lower half of the screen, following an assumption that interesting objects are typically located in that area. Simultaneously, we avoid measuring in the sky area, which mostly occurs in the upper part of the image. The zone weights can also be set based on an image database containing a large number of pictures with optimal setting of the exposure [3]. Here, the authors describe a system where images are divided in 25 equal zones and all weights are calculated based on the optimization procedure, having values as in Figure 1(a). In some cases, the user is given the freedom to set the weights and positions of several zones of interest. This is particularly important in the so-called back-lit scenes, where the object of interest is surrounded by very bright areas in scenarios like tunnel exits, persons entering the building on a bright sunny day while the camera is inside of the building, or in a video-phone application where a bright sky behind

the person dominates the scene. These solutions are often used for low- to medium-dynamic range sensors which cannot capture the dynamics of the High-Dynamic Range (HDR) scenes without losing some information. Generally, these problems were typically solved by overexposing the image so that details in the shadows have a good visibility. However, all the details in the bright parts of the image are then clipped and lost. In case when no object of interest is present, the exposure of the camera is reduced to correctly display the background of the image. This explains why it is important to correctly set the metering zones to give a higher weight to important foreground that is often darker than a bright background. Otherwise, the object of interest will be underexposed and will vanish in shadows. This scheme is called back-light compensation and is discussed further in Section 6.

2.2. Matrix (Multizone) Metering. This metering mode is also called honeycomb or electroselective pattern metering, as the camera measures the light intensity in several points of the image and then combines the results to find the settings for the best exposure. The actual number of zones can range from a few up to a thousand, and various layouts are used (see [1] and Figures 1(b)–1(d)). A number of factors are considered to determine the exposure: the autofocus point, areas in focus and out of focus, colors in the image, dynamic range, and back-light in the image, and so forth. A database of features of interest taken from many images (often more than 10,000) is prestored in the camera and algorithms are used to determine what is being captured and accordingly determine the optimal exposure settings. Matrix metering is mainly used in high-end digital still cameras whereas this technology is not very suitable for video cameras due to its complexity and stability for dynamic scenes. This is why other types of metering systems are needed to solve the problem of optimal exposure for video.

2.3. Content-Based Metering Systems. The basic problem of the classical zone metering system is that large background areas of high brightness are spoiling the measurement, resulting in an underexposed foreground. To avoid this situation, intelligent processing in the camera can consider only important scene parts, based on statistical measures of "contrast" and "focus", face and skin-tones, object-based detection and tracking, and so forth. For example, it can be assumed that well-focused/high-contrast/face/object regions are more relevant compared to the others and will be given a higher weight accordingly. Content-based metering systems are described in more detail in Section 6.

3. Measurement Types Used for the Exposure Control

In this section we discuss various measurement types used for the exposure controller. Starting from the standard average measurement, we will introduce other types of measurements which are used in some specific applications, for instance, HDR scenes. We will not discuss focus, contrast,

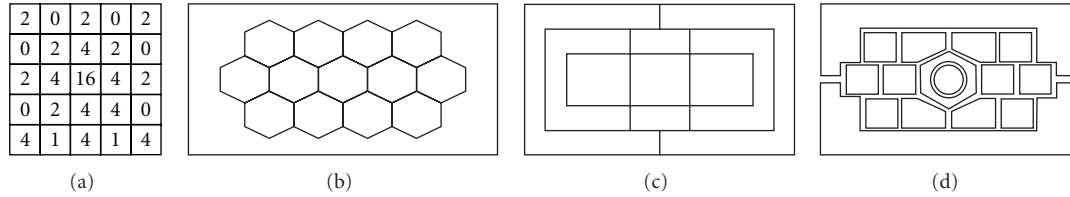| 2	0	2	0	2
0	2	4	2	0
2	4	16	4	2
0	2	4	4	0
4	1	4	1	4

(a) (b) (c) (d)

FIGURE 1: Multizone metering mode used by several camera manufacturers, adapted from [1, 3]. (a) The weighting matrix in a 25-zone system, (b) 14-zone honeycombs, (c) 16-zone rectangulars, and (d) 16-zone flexible layout.

skin-tone, or other types of measurement that are not directly based on the image intensity [1, 3].

3.1. Average Luminance Measurement (AVG).

The average luminance measurement Y_{AVG} is used in most exposure applications. It is defined as an average value of pixel luminance in the area of interest and is measured by accumulating pixel luminance values within the measurement window. Depending on the application, different weights can be used throughout the image, by dividing the measurement window to subareas. In cases when video-level controller uses only AVG measurement, it tunes camera parameters to make the measured average luminance value equal to the desired average luminance value.

3.2. Median and Mode Measurement.

Using a median intensity measurement within an area of interest has certain advantages over the average intensity measurement. Namely, exposure problems with the HDR scenes result from the fact that the average luminance measurement Y_{AVG} of such a image is high due to the very bright background image, so that an interesting foreground image remains dark. On the other hand, the median value Y_{MED} of such an image is much lower due to a bulk of dark pixels belonging to the foreground, as in Figure 2(a) [4]. Consequently, the actual brightness of the pixels in the background is irrelevant, since the median of the image is not taking them into account if there are enough dark foreground pixels. This is in most cases satisfied, particularly for the HDR images. The *mode* of the histogram distribution can also be used in a similar manner, as in [5], where a camera exposure system is presented that finds the mode of the histogram and controls the exposure such that the mode drifts towards a target position bin in the histogram. In case of a simple video-level control with only one measurement input, the median is a better choice than the average measurement. However, in more complex video-level control algorithms which include saturation control from Section 6, an average level control suffices.

Unfortunately, the output of the median calculation can show large variations. Let CF be a scaled cumulative distribution function of an input image, normalized to a unity interval. The median is calculated as an luminance value Y_{MED} which is defined by $\mathrm{CF}(Y_{\mathrm{MED}}) = 0.5$. In other words, $Y_{\mathrm{MED}} = \mathrm{CF}^{-1}(0.5)$. For instance, in cases when the input image histogram is bimodal with a similar amount of

dark and bright pixels, as in Figure 2(b), a small change in the input image can move the median value from the dark side of the image to the bright side. This is illustrated in Figure 2(c), where we present a cumulative histogram function of image from Figure 2(b). It becomes obvious that if histogram $H(i)$ changes from a starting shape a to a shape b, its CF changes from CF_a to CF_b, which can considerably change the position of the median (the median changes from $\mathrm{CF}_a^{-1}(0.5)$ to $\mathrm{CF}_b^{-1}(0.5)$). This change of control measurement would introduce potential instabilities and large changes in the response of the system. To mitigate this effect, we propose to calculate the median as $Y_{\mathrm{MED}} = [\mathrm{CF}^{-1}(0.5 + \delta) + \mathrm{CF}^{-1}(0.5 - \delta)]/2$, where δ is a small number (e.g., 0.05). In this way, we prevent large changes of the median, even if the standard-definition median would change considerably, thereby improving the stability of the exposure control.

3.3. Peak White Measurement (PW).

In some cases, especially in the HDR scenes, where high-intensity parts of the image are clipped, a Peak White measurement Y_{PW} is used in addition to the average measurement to fine-tune the exposure level of the camera and decrease a number of the clipped pixels. Thereby, user can see potential details of the image that were lost (clipped at bright intensities). There is no unique definition for the computation of a PW measurement. However, its result in terms of control should be that the overall intensity level is lowered globally, for the sake of visualization of important bright details. Let us first give some introductory comments about the use of PW measurement, after which we briefly discuss several definitions.

Firstly, using only PW measurement in the exposure control of the camera is not desired, since it can lead to control stability problems when bright objects or light sources enter (appear) or leave the scene. In these cases, large variations in the measured signal lead to large average intensity variations as a response to the exposure controller.

Secondly, if very bright light sources like lamps and sun or large areas of specularly reflected pixels are directly visible in the scene, it is difficult to decide whether they should be included in the PW measurement. Lowering the average intensity value of the image to better visualize clipped bright areas is then not effective, due to a very high intensity of these areas which can be several times higher that the available dynamic range of the imaging sensor. We now discuss three possible PW measurements.
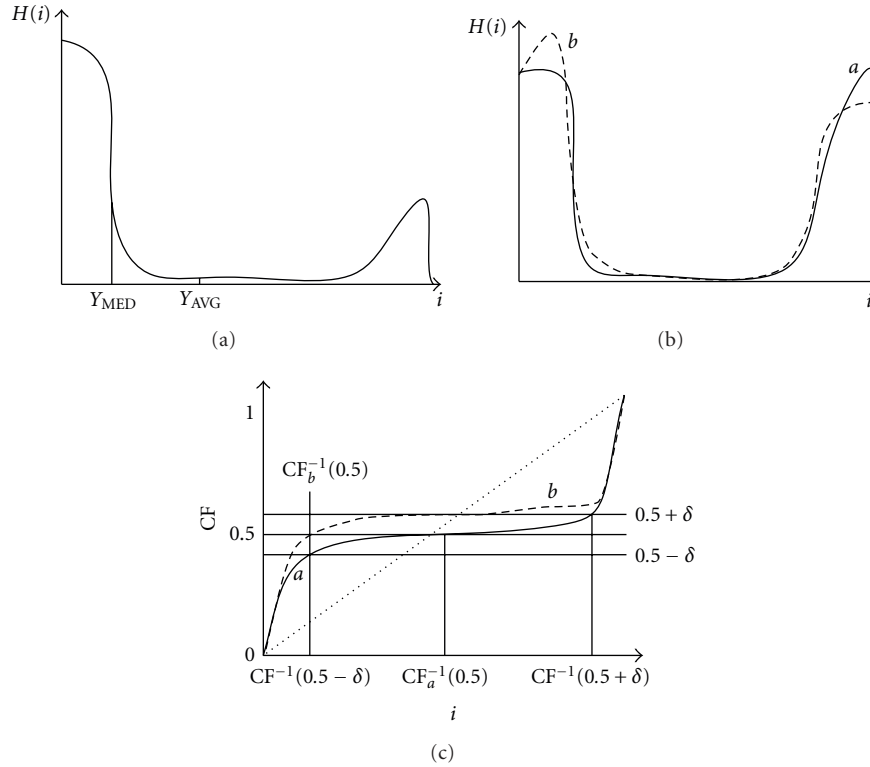
FIGURE 2: (a) Median of the input image is less sensitive than the average to very bright pixels in the background. (b) Histogram of two input signals, a and b. (c) Cumulative functions based on input signals a and b. Although not very different, they yield very different medians, which can lead to instability. We propose to modify calculation of the median by looking at its 2δ neighborhood.

3.3.1. Max of Min Measurement. The PW measurement can be naively defined as the brightest luminance pixel in the image, but to avoid noisy pixels and lonely bright pixels, it can be better defined as a global maximum value of the local minimum of pixel luminance Y in a small window of size $(2a_k + 1)(2b_k + 1)$. By finding the local minimum value $\min_l$ around each pixel (at a position (m, n)), we can exclude outliers from the subsequent calculation of a global maximum value $\max_g$ in the image:

$$Y_{\text{PW}} = \max_g\left(\min_l\left(Y(m + i, n + j)\right)\right),$$
$$i = -a_k, \ldots, a_k, \quad j = -b_k, \ldots, b_k. \tag{1}$$

By adjusting the size of the local window, we can skip small specular reflectance pixels which do not carry any useful information. Still, with this approach, we cannot control the amount of pixels in the image that determine the peak information. This is why we would like to include number of pixels in the PW calculation, which will be described next.

3.3.2. Threshold-Based Measurement. The PW measurement can also be defined in terms of the number of pixels above a certain high threshold: if more pixels are above that threshold, a larger reaction is needed from the controller. However, this kind of measurement does not reveal the distribution of pixels and can lead to instabilities and challenges for smooth control. Particularly, if pixels are close

to the measurement threshold, they can easily switch their position from one side of the threshold to the other. In one case, we would measure a significant number of bright pixels and in the other case much less or even none. From the previous discussion, it is clear that a better solution is required to solve such difficult cases. This solution is a histogram-based measurement.

3.3.3. Histogram-Based Measurement. A histogram measurement provides a very good description of the image, since it carries more information than just the average intensity or the brightest pixels in the image. We can define a better definition of the PW measurement which is the intensity level of the top $n\%$ of pixels (usually n is in the range 0.5%–3%). Likewise, we combine information of the number of pixels with their corresponding intensity to ensure that a significant number of the brightest pixels are considered and that all the outliers are skipped. If a large number of specularly reflected pixels exist in the image, we can consider applying a prefiltering operation given by (1) to skip them.

4. Control Types in Video Cameras

Video cameras contain three basic mechanisms for the control of the output image intensity: a controllable lens (a closed-loop servo system as, e.g., DC or AC iris lens), variable integration time of the sensor, and the applied gain (analog or digital) to the image. Each of these controls has its

own peculiarities, different behavior, and effect on the image. The task of the video-level control algorithm is to maintain the correct average luminance value of the displayed image, regardless of the intensity of the input scene and its changes. For example, when certain object moves into the scene or if scene changes its intensity due to a light switched on or off, video-level controller reacts to maintain a correct visibility of image details, which would otherwise be either lost in shadows or oversaturated. If the scene becomes darker, level control is achieved by opening the lens or using the lager sensor integration time or larger value of gain, and vice versa. The level-control process should result in a similar output image impression regardless of the intensity level in the scene and should be fast, smooth, and without oscillations and overshoots. The video-level control input is often an average input exposure value Y_{AVG} or some other derived feature of interest, as described in Section 3. We briefly address the above control mechanisms and then present specific control algorithms for each of them.

Adjustable iris lenses can be manual or automatic. For the manual lenses, user selects a fixed setting, while the automatic ones feature a dynamical adjustment following a measurement. If this measurement and the aperture control occur in the lens unit using the actual video signal as input, it is said to be a video (AC) iris lens. Alternatively, when the measurement occurs outside the lens unit, it is called a DC iris and an external signal is used to drive the lens. The iris is an adjustable opening (aperture), that controls the amount of light coming through the lens (i.e., the "exposure"). The more the iris is opened, the more light it lets in and the brighter the image will be. A correct iris control is crucial to obtain the optimum image quality, including a balanced contrast and resolution and minimum noise.

To control its opening, the AC iris lens has a small integrated amplifier, which responds to the amount of scene light. The amplifier will open or close the iris automatically to maintain the same amount of light coming to the image sensor. By adding positive or negative offsets and multiplying this video signal, we explicitly guide the controller in the lens, to open or close the iris. To obtain a stable operation of AC iris lenses, they are constructed to have very slow response to dynamic changes. There are cases where the response is fully absent or follows special characteristics. First, such lenses often have large so-called dead-areas in which they do not respond to the driving signal. Second, the reaction to an intensity change can be nonlinear and nonsymmetrical. Third, a stable output value can have static offset errors.

The DC iris lens has the same construction but is less expensive since there is no amplifier integrated in the lens. Instead, the amplifier is in the camera which drives the lens iris through a cable plugged into the camera. For the DC iris lens, the signal that controls the iris opening and closing should have a stable value if the input signal is constant and should increase/decrease when the input signal decreases/increases. This control is most of the times achieved by a PID controller [6]. The use of a custom PID type of video level control allows an enhanced performance compared to AC iris lens type. For high-end video applications, the DC iris lens is adopted and discussed

further below. However, since it is not known in advance which DC iris lens will be attached to the camera, a PID loop should be able to accommodate all DC iris lenses. Hence, such a control is designed to be relatively slow and stability and other problems as for the AC iris lens often occur due to the large variations in characteristics of the various lenses.

The sensor exposure time and applied gain can also be used for video-level control. The control associated with these parameters is stable and fast (change is effective next video frame already) and offers good linearity and known response. In addition, any possible motion blur reduces only with the shorter exposure time and not with closing of the lens. (Motion is even more critical for rolling-shutter CMOS sensors, which introduce geometrical distortions. In these cases, sensor exposure time must be kept low, and lens control should be used to achieve a desired average video level.) Therefore, when observing motion scenes like traffic or sport events, the sensor integration time is set deliberately low (depending on the speed of objects in the scene) to prevent the motion blur. For traffic scenes, integration time can be as low as 1 millisecond for license-plate recognition applications.

The above discussion may lead to the desire of using the exposure time for the video-level control. However, lens control is often preferred to integration time or a gain control, even though it is less stable and more complex. While the operating range of the integration time is from $1/50$ s (or $1/60$ s) to $1/50,000$ s (a factor of 1000), this range is much larger for lenses with iris control. (If camera employs small-pixel size sensors, to avoid a diffraction-limit problem and a loss of sharpness, opening of the lens can be kept to more than F11, which then limits the lens operating range and imposes a different control strategy. However, this discussion is beyond the scope of this paper.) Furthermore, lenses are better suited to implement light control, as they form the first element of the processing chain. For example, when the amount of light is large, we can reduce the exposure time of the sensor, but still the same light reaches the color dies on the sensor and can cause their deterioration and burn-in effect. Besides this, closing the lens also improves the field of depth and generally sharpens the image (except for very small sensor pixel sizes which suffer from diffraction-limit problems).

4.1. PID Control for DC Iris Lens. The working principle of a DC iris lens consists of moving a blocking part, called an iris blade, in the pathway of the incoming light (Figure 3). Iris is a plant/process part of the control system. To prevent the iris blade from distorting the information content of the light beam, the iris blade must be positioned before the final converging lens. Ideally, the iris blade should be circularly shaped, blocking the incoming light beam equally over a concentric area; however, circular shape is seldom used for practical reasons. A voltage delivered to a coil controls the position of a permanent magnet and hence the opening of the lens via a fixed rod. Two forces occur in this configuration: F_{el}, resulting electrical force exerted on the magnet as a result of a voltage on the coil, and F_{mech}, mechanical force exerted on the magnet as a result
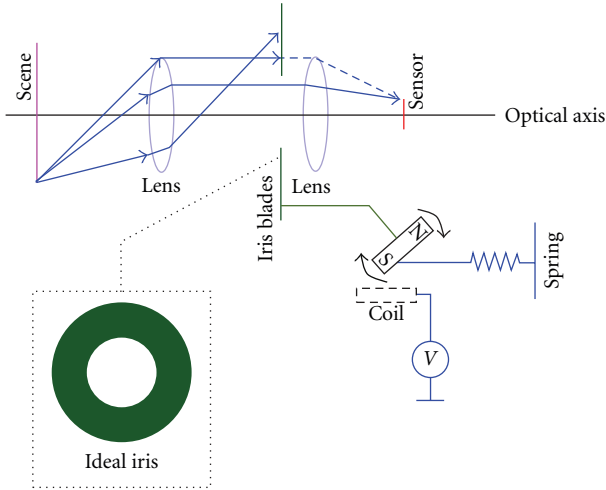
FIGURE 3: Adjustable iris control.

of the rigidity of the spring. When $F_{el} = F_{mech}$, the current position of the iris does not change (the equilibrium, Lens Set Point (LSP)). For $F_{el} < F_{mech}$, the mechanical force is larger than the electrical force, and the iris closes until it reaches the minimum position. Finally, for $F_{el} > F_{mech}$, the iris opens until it reaches the maximum opening position. The control system is realized by software, controlling an output voltage for driving the iris. The driving voltage in combination with the driving coil and the permanent magnet results in the electromagnetic force. These represent the actuator of the system.

The core problem for DC iris control is the unknown characteristics of the forces and the attached DC iris lens as a system. Each DC iris lens possesses a specific transfer function due to a large deviation of the LSP in addition to the differences in friction, mass, driving force, equilibrium force, iris shape, and so forth. Using a single control algorithm for all lenses results in a large deviation of control parameters. To cope with this variable and unknown characteristics, we have designed an adaptive feed-back control. Here, the basic theory valid for the linear time invariant systems is not applicable, but it is used as a starting point and aid for the design. As such, to analyze the system stability, we cannot employ the frequency analysis and a root-locus method [7], but have to use a time-series analysis based on a step and sinus responses.

Due to the unknown nonlinear lens components, it is not possible to make a linear control model by feedback linearization. Instead, a small-signal linearization approach around the working point (LSP) is used [8]. Furthermore, DC iris lenses have a large spread in LSPs: for example, temperature and age influence the LSP in a dynamic way (e.g., mechanical wear changes the behavior of the DC iris lens and with that the LSP). An initial and dynamic measurement of the lens' LSP is required. The initial LSP is fixed, based on an averaged optimum value for a wide range of lenses, and the dynamic LSP value is obtained by observing a long-term "lowpass" behavior of the lens. In addition, the variable friction and mechanical play result in

a momentous dead area around the LSP, which we also have to estimate.

The simplest way to control a DC iris is with a progressive control system. However, a major disadvantage of such a controller is the static error, which is enlarged by the presence of the dead-area. An integrating control is added to the control system software to reduce the static error to acceptable levels. Software integrators have the added advantage that they are pure integrators and can theoretically cancel the static error completely. Finally, derivative action anticipates where the process is heading, by looking at the rate of change of the control variable (output voltage). Let us now further discuss the PID control concept for such a lens.

We will mark the Wanted luminance Level of the output image with Y_{WL} and measured average luminance level with Y_{AVG}. An error signal $\Delta Y = Y_{WL} - Y_{AVG}$ is input to the exposure controller, which has to be minimized and kept at zero if possible. However, this error signal is nonzero during the transition periods, for instance, during scene changes or changes of the WL set by the user. The mathematical representation of the PID controller is given by [6]

$$V(t) = \text{LSP} + k_p \cdot \Delta Y(t) + \frac{1}{T_i} \cdot \sum \Delta Y(t) + T_d \cdot \frac{d(\Delta Y(t))}{dt}. \tag{2}$$

Here, $V(t)$ represents the driving voltage of the DC iris lens, LSP is a Lens Set Point, and terms $(1/T_i) \cdot \sum \Delta Y(t)$ and $(T_d) \cdot d(\Delta Y(t))/dt$ relate to the integral and the differential action of the controller, respectively. The DC iris lens is a nonlinear device, and it can be linearized only in a small area around the LSP. To achieve the effective control of the lens, we have to deviate from the standard design of the PID control and modify the controller. This discussion goes beyond the scope of this paper; so we will only mention several primary modifications.

First of all, LSP and dead area are not fixed values but are lens dependent and change in time. This is why an initial and dynamic measurement of the lens' LSP is required. Secondly, proportional gain k_p is made proportional to the error signal. Likewise, we will effectively have a quadratic response to the error signal, by which the reaction time for DC iris lenses with a large dead area is decreased. The response is given by a look-up table, interpolating intermediate values, such as depicted in Figure 4(a). Thirdly, the integrator speed has been made dependent of the signal change, in order to decrease the response time for slow lenses and reduce the phase relation between the progressive and the integrating part. The larger the control error is, the faster the integrator will react. A representation of the integrator parameter is shown in Figure 4(b). In addition, if the error is large and points at a different direction than the integrator value, a reset of the integrator is performed to speed up the reaction time. Once stability occurs, the necessity for the integrator disappears. The remaining integrator value keeps the driving voltage at one of the edges of equilibrium, which a small additional force can easily disturb. The strategy is to slowly reset the integrator value to zero which also helps in the event of a sudden change of the LSP value, as the slow reset of
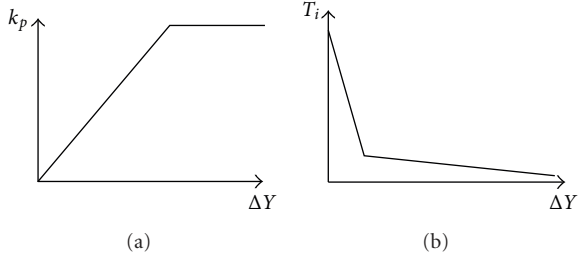
FIGURE 4: (a) Proportional gain factor k_p as a function of a given error, (b) integrator parameter T_i as a function of a given error.

the integrator value disturbs the equilibrium and adds a new chance for determining the correct LSP.

4.2. LUT-Based Control. A simulated Camera Response Function (CRF) gives an estimate of how light falling on the sensor converts into final pixel value. For many camera applications, the CRF can be expressed as $f(q) = 255/(1 + \exp(-Aq))^C$, where q represents the light quantity given in base-2 logarithmic units (called stops) and A and C are parameters used to control the shape of the curve [1]. These parameters are estimated for a specific video camera, assuming that the CRF does not change. However, this assumption is not valid for many advanced applications that perform global tone mapping and contrast enhancement. If the CRF is constant, or if we can estimate parameters A and C in real-time, then the control error prior to the CRF is equal to $\Delta Y = f^{-1}(Y_{WL}) - f^{-1}(Y_{AVG})$. The luminance of each pixel in the image is modified in a consecutive order, giving an output luminance $Y' = f(f^{-1}(Y) + \Delta Y)$. The implementation of this image transformation function is typically based on a Look-Up Table (LUT).

An alternative realization of the exposure control system also uses an LUT but does not try to compensate for the CRF. It originates from the fact that the measured average value of the image signal Y_{AVG} is made as a product of brightness L of the input image, Exposure (integration) Time t_{ET} of the sensor, gain G of the image processing pipeline, and a constant K, see [9], and computed with $Y_{AVG} = K \cdot L \cdot G \cdot t_{ET}$. The authors derive a set of LUTs that connect exposure time t_{ET} and gain G with the brightness L of the object. Since the brightness changes over more than four orders of magnitude, the authors apply a logarithm to the previous equation and set up a set of LUTs in the logarithmic domain, where each following entry of L is coupled with the previous value with the multiplicative factor. Likewise, they set up a relationship LUT structure between the logarithmic luminance of the object and t_{ET} and G, giving priority to the exposure time to achieve a better SNR.

Since the previous two methods are based on an LUT implementation, they are very fast; however, they are more suitable for the digital still cameras. Namely, the quantization errors in the LUTs can give rise to a visible intensity fluctuation in the output video signal. Also, they do not offer the flexibility needed for more complex controls such as a saturation control. In addition, the size of the LUT and

correct estimation of parameters A, C, and K limits these solutions.

4.3. Recursive Control. As an alternative to a PID control, we propose a new control type that is based on recursive control. This control type is very suitable and native for the control of the exposure time of the sensor (shutter control) and gain (gain control). The advantage of the recursive control is its simplicity and ease of use. Namely, for a PID type of control, three parameters have to be determined and optimized. Although some guidelines exist for tuning the control loop, numerous experiments have to be performed. However, for each particular system to be controlled, different strategies are applicable, depending on the underlying physical properties. This discussion is beyond the scope of this paper; we recommend [6, 10] for more information.

4.3.1. Exposure Control. Image sensors (CCD and CMOS) are approximately linear devices with respect to the input light level and charge output. A linear model is then a good approximation of the sensor output video level $Y = C \cdot t_{ET}$, where Y is the output luminance, t_{ET} is the Exposure Time of the sensor, and C denotes a transformation coefficient (which also includes the input illumination function). If a change of the exposure time occurs, the output average luminance change can be modeled as $\Delta Y = C \cdot \Delta t_{ET}$, yielding a proportional relation between the output video level and the exposure time. Let us specify this more formally. A new output video level Y'_{AVG} is obtained as

$$Y'_{AVG} = Y_{AVG} + \Delta Y = C \cdot t'_{ET} = C \cdot (t_{ET} + \Delta t_{ET}), \quad (3)$$

by change of the exposure time with

$$t_{ET}(n+1) = t_{ET}(n) + \Delta t_{ET} = \frac{Y_{AVG}}{C} + \frac{\Delta Y}{C}, \quad (4)$$

which results in

$$t_{ET}(n+1) = \frac{Y_{AVG}}{C}\left(1 + \frac{\Delta Y}{Y_{AVG}}\right) = t_{ET}(n)\left(1 + \frac{\Delta Y}{Y_{AVG}}\right). \quad (5)$$

Hence, the relative change of the video level is $\Delta Y / Y_{AVG} = \Delta t_{ET}/t_{ET}$. The parameter n is a time variable which represents discrete moments nT, where T is the length of the video frame (in broadcasting sometimes interlaced fields). Such a control presumes that we will compensate the exposure time in one frame for a change of $\Delta Y = Y_{WL} - Y_{AVG}$. For smooth control, it is better to introduce time filtering with factor k, which determines the speed of control, so that the exposure time becomes

$$t_{ET}(n+1) = t_{ET}(n)\left(1 + k \cdot \frac{\Delta Y}{Y_{AVG}}\right), \quad (6)$$

where $0 \le k \le 1$. A small value of parameter k implies a slow control and vice versa (typically $k < 0.2$). This equation presents our proposed recursive control, which we will use to control the exposure time of the sensor and the gain value.

4.3.2. Gain Control. The output video level (if clipping of the signal is not introduced) after applying the gain G equals to $Y_{\text{out}} = GY$; so the same proportional relation holds between the output video level and the gain (assuming that the exposure time is not controlled), being $\Delta Y / Y_{\text{AVG}} = \Delta G / G$, leading to a controlled gain:

$$G(n + 1) = G(n)\left(1 + k \cdot \frac{\Delta Y}{Y_{\text{AVG}}}\right). \tag{7}$$

In this computation, parameters t_{ET} and G are interchangeable and their mathematical influence is equivalent. The difference is mainly visible in their effect on the noise in the image. Namely, increasing the exposure time increases the SNR, while increasing the gain generally does not change the SNR (if the signal is not clipped), but it increases the amplitude (and hence visibility) of the noise. This is why we prefer to control the exposure time, and only if the output intensity level is not sufficient, the controller additionally starts using gain control. As mentioned, for scenes including fast motion, the exposure time should be set to a low value, and instead, the gain (and iris) control should be used.

5. Video-Level Control Strategies

In this section we will discuss the strategy employed for overall video-level control of the camera, which includes lens control, exposure control of the sensor, and gain control of the image processing chain. We will apply the concept of a recursive control proposed in previous section, intended for the control of sensor integration time and the gain, whereas the lens is controlled by a PID control. First we will discuss a state-of-the-art sequential concept for overall video level control. In most cases, to achieve the best SNR, sensor exposure control is first performed and only when the sensor exposure time (or the lens opening) reaches its maximum, digital gain control will be used supplementary. (The maximum sensor exposure time is inversely proportional to the camera capturing frame frequency, which is often 1/50 s or 1/60 s. Only in cases when fast moving objects are observed with the camera, to reduce the motion blur, the maximum integration time is set to a lower value depending on the object speed. This value is, e.g., 1/1000 s when observing cars passing by with a speed of 100 km/h.) However, in cases when the video camera system contains a controllable lens, the system performance is degraded due to the unknown lens transfer characteristics and the imposed control delay. To obtain a fast response time, we will propose a parallel control strategy to solve these delay drawbacks.

5.1. Sequential Control. In case of a fixed iris lens, or if the lens is completely open, we can perform video-level control by means of changing the exposure time t_{ET} and digital gain G. A global control model is proposed where, instead of performing these two controls individually, we have one control variable, called integration time (t_{IT}), which can be changed proportionally to the relative change of the video signal, and from which the new t_{ET} and G values can

be calculated. This global integration time is based on the proposed recursive control strategy explained in the previous section and is given by

$$t_{\text{IT}}(n + 1) = t_{\text{IT}}(n)\left(1 + k \cdot \frac{\Delta Y(n)}{Y_{\text{AVG}}(n)}\right). \tag{8}$$

In this equation, $Y_{\text{AVG}}(n)$ represents the measured average luminance level at discrete time moment n, $\Delta Y(n)$ is the exposure error sequence from the desired average luminance value (wanted level Y_{WL}), and $k < 1$ is a control speed parameter. Preferably, we perform the video-level control by employing the sensor exposure time as a dominant factor and a refinement is found by controlling the gain. The refinement factor, the gain G, is used in two cases: (1) when t_{ET} contains the noninteger parts of the line time for CCD sensors and some CMOS sensors, and (2) when we cannot reach the wanted level Y_{WL} set by the camera user using t_{ET}, as we already reached its maximum ($t_{\text{ET}} = T$, full frame integration). Figure 5 portrays the sequential control strategy. We have to consider that one frame delay (T) always exists between changing the control variables t_{ET} and G and their effective influence on the signal. Also, the control loop responds faster or slower to changes in the scene, depending on the filtering factor k. The operation of the sequential control is divided into several luminance intervals of control, which will be described. An overview of these intervals and their associated control strategy is depicted in Figure 6.

5.1.1. Lens Control Region. When sufficient amount of light is present in the scene and we have a DC or AC iris lens mounted on the camera, we use the iris lens to perform video-level control. The DC iris lens is controlled by a PID control type, whereas the AC iris lens has a build-in controller that measures the incoming video signal and controls the lens to achieve an adequate lens opening. When this lens control is in operation, other controls (exposure and gain control) are not used. Only when the lens is fully open and the wanted video level is still not achieved, we have to start using exposure and gain controls. A problem with this concept is that we do not have any feedback from the lens about its opening status; so we have to detect a fully open condition. A straightforward approach for this detection is to observe the error signal ΔY. If the error remains large and does not decrease for a certain time t_{check} during active lens operation, we assume that the lens is fully open and we proceed to a second control mode (*Exposure control*, see at the top of Figure 6). This lens opening detection (in sequential control) always introduces delays, especially since time t_{check} is not known in advance and has to be assumed quite large to ensure lens reaction, even for the slowest lenses with large dead areas. Coming from the other direction (*Exposure control* or *Gain control* towards the *Lens control*) is much easier, since we know exactly the values of the t_{ET} and G, and whether they have reached their nominal (or minimal) values. In all cases, hysteresis has to be included in this mode transition to prevent fast mode switching.
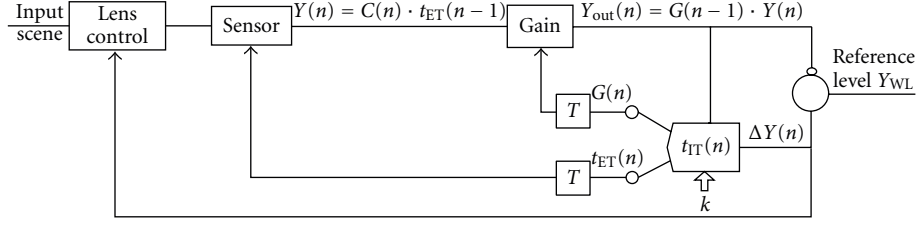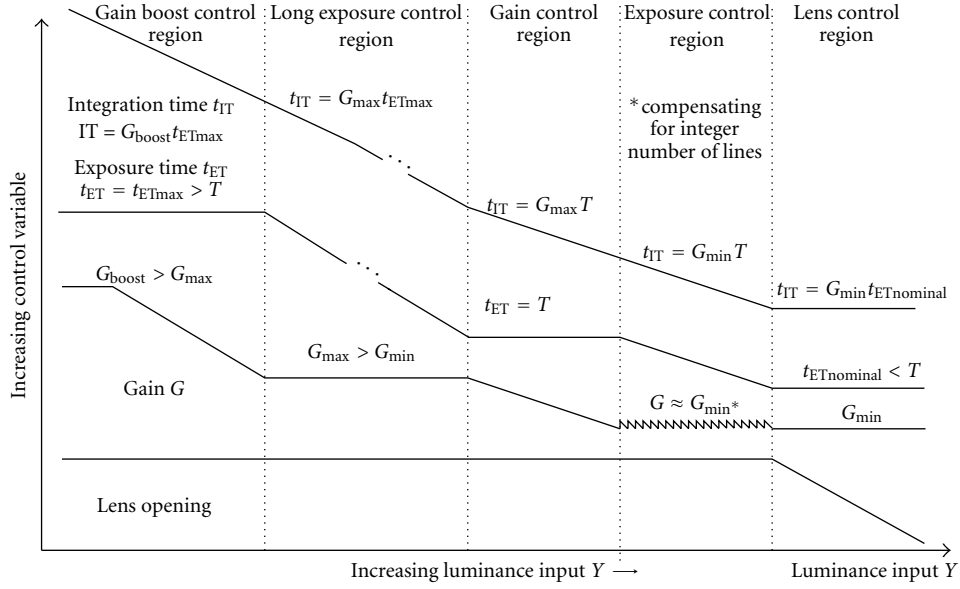
FIGURE 5: Model of the sequential control loop for video-level control.

FIGURE 6: Video-level control regions, where the Lens opening, Gain G, Exposure time t_{ET}, and Integration time t_{IT} are shown as related to the Luminance input Y (more light in the right direction).

5.1.2. Exposure Control Region ($G = G_{min}$). Assuming that we can deploy the exposure time only for an integer number of lines, we have

$$t_{ET}(n+1) = T_L \cdot \left\lfloor \frac{t_{IT}(n+1)}{T_L} \right\rfloor, \qquad (9)$$

where T_L is the time span of one video line and $\Delta t_{ET} = t_{IT} - t_{ET}$ represents the part of the t_{IT} that we cannot represent with t_{ET}. Therefore, instead of achieving $Y_{AVG} = Y_{WL} = C \cdot t_{IT} \cdot G_{min}$, we reach $Y_{AVG} = C \cdot t_{ET} \cdot G_{min}$. Hence, we have to increase the gain with ΔG in order to compensate for the lacking difference, and achieve $Y_{AVG} = Y_{WL}$ by

$$C \cdot t_{ET} \cdot (G_{min} + \Delta G) = C \cdot t_{IT} \cdot G_{min}. \qquad (10)$$

This implies the application of an additional gain:

$$\Delta G = G_{min} \frac{\Delta t_{ET}}{t_{ET}}, \qquad (11)$$

so that the new gain becomes

$$G = G_{min}\left(1 + \frac{\Delta t_{ET}}{t_{ET}}\right) = G_{min}\left(1 + \frac{t_{IT} - t_{ET}}{t_{ET}}\right) = G_{min}\frac{t_{IT}}{t_{ET}}. \qquad (12)$$

5.1.3. Gain Control Region. In this region, the exposure time is $t_{ET} = t_{ETmax} = T$ (frame time), so that the compensation of ΔY is performed by gain. We reuse the form of (8), where the gain is equal to

$$G(n+1) = G(n)\left(1 + k \cdot \frac{\Delta Y}{Y_{AVG}}\right) = G(n)\frac{t_{IT}(n+1)}{t_{IT}(n)}$$
$$= G_{min}\frac{t_{IT}(n+1)}{t_{ET}(n+1)}. \qquad (13)$$

The last expressions are mathematically equal because we are compensating for insufficient exposure time, so that

$$Y_{AVG}(n) = G_{min} \cdot C \cdot t_{IT}(n) = G(n+1) \cdot C \cdot t_{ETmax}$$
$$\implies G(n+1) = G_{min}\frac{t_{IT}(n)}{t_{ETmax}}. \qquad (14)$$

The control strategy when using the gain is to compensate as much as possible the level error using the exposure time of the sensor and compensate the remainder of the level error with gain. This implies that we do not separate exposure and gain regions but rather consider it as one region, where the exposure time is limited to the maximum integration of one field/frame. We can also impose the maximum gain G_{max},

after which we switch to the *long exposure control* region, where $t_{ET} > T$. The reason for this approach is that a too high gain would deteriorate the image quality by perceptually annoying noise.

5.1.4. Long Exposure Control Region. A similar control strategy is adopted for the *long exposure control* region: if the parameter setting $t_{ET} = T$ and $G = G_{max}$ is insufficient for achieving $Y_{AVG} = Y_{WL}$, we have to increase the exposure time, while keeping $G = G_{max}$. In this case, we only have to find a new exposure time (which is larger than T), but now compensating on top of t_{IT}/G_{max}. Effectively, the sensor will integrate the image signal over several field/frame periods. We can also limit the maximum exposure time $t_{ETmax}(>T)$ to prevent serious motion degradation.

5.1.5. Gain Boost Control Region. If the integration time of t_{ETmax} is insufficient, the system moves the operation to the gain-boost region, where the remainder of the gain is used. Now we keep $t_{ET} = t_{ETmax}$ and just calculate a new gain to compensate from t_{ETmax} to the desired integration time t_{IT}. Typical values are $G_{max} = 4, t_{ETmax} = 4T \cdots 8T$, and $G_{boost} = 16$. The integration time t_{IT} is now confined to the range: $0 < t_{IT} \leq G_{boost} \cdot t_{ETmax} = 64 \cdot T$.

Example of Control Realization. If the digital gain can be adjusted in 128 steps, the digital value of the gain is computed by

$$G(n + 1) = \frac{1}{128} \left\lfloor \frac{t_{IT}(n)}{t_{ET}(n)} \cdot 128 \right\rfloor. \tag{15}$$

In the *Exposure control* and *long exposure control* region, the gain is fixed to G_{min} and G_{max}, respectively, (except in the *Exposure control* region for the compensation between the achieved exposure time by integrating over an integer number of lines and wanted exposure time). The exposure time t_{ET} accordingly becomes

$$t_{ET}(n) = \min\left(T, \left\lfloor \frac{t_{IT}(n)}{T_L} \right\rfloor \cdot T_L\right) \tag{16}$$

for the *exposure control* region. For the *long exposure control* region, t_{ET} is specified by

$$t_{ET}(n) = \min\left(t_{ETmax}, \left\lfloor \frac{t_{IT}(n)/G_{max}}{T_L} \right\rfloor \cdot T_L\right), \tag{17}$$

whereas $t_{ET} = t_{ETmax}$ in the *gain boost control* region.

The value of the theoretical specification of the past paragraphs is covered in several aspects. First, the overview of Figure 6 is rarely or not discussed in literature, and it provides ways for a large range of control of the luminance and with defined intervals. Second, the equations form a framework for performing control functions. Third, the equations quantify the conversion of exposure time to gain control and finally video level.

5.2. Parallel Control. Despite the clarity of the previously discussed sequential control strategy and the presented

theoretical model, the sequential control has considerable disadvantages: the reaction speed and delays of the total control loop. As mentioned, the lens control operates according to the "best effort" principle, but due to versatility of lenses with different and unknown characteristics, it is difficult to ensure a predetermined reaction time and the absence of a nonconstant static error. To obtain a much faster control response and flexibly manipulate control modes, we propose a parallel control concept, in which we control the lens in parallel with the gain. Additionally, we can fully compensate the static error of the lens.

Figure 7 portrays the diagram of our new parallel control system. The diagram reflects also our design philosophy. In the first part, the lens/sensor and the digital gain algorithms ensure that the desired video level is obtained at Point B instead of at the end of the camera (Point D). This has the benefit that all enhancement functions, of which some are nonlinear, will operate on a user-defined setting and will not disturb the video level control itself. If these nonlinear processing steps would be inside the control loop, the control algorithm would be complicated and less stable. Hence, we separate the dynamic tone mapping effects which take place in the camera from the global video level setting. Due to dynamic tone mapping, the transfer function of the total camera changes depending on the input signal distribution and the user preferences. We isolate these functions in the Enhancement (contrast) control block of (Figure 7).

The video-level control is now operating prior to the enhancement control and its objective is to make the average digital signal level at Point B equal to the Wanted Level Y_{WL} set by the user. Afterwards, the enhancement control will further improve the signal but also lead to a change at the output level that is different from the controlled level at Point B. However, the assumption is that this change is for the benefit of creating a better image at the output. Finally, the digital gain control and post-gain control will stabilize the output video level and act as a refinement if necessary.

Let us now discuss the diagram of Figure 7 in more detail. The video-level control is performed by (1) front-end control involving the control of sensor Exposure Time (t_{ET}) and lens control (voltage V)), and (2) Digital Gain (DG) Control, which manipulates the gain parameter G. (Instead of digital gain, an analog gain control in the sensor can also be used. However, for the sake of simplicity, we will discuss the digital gain case only.) The DG control and ET control are performed as recursive (multiplicative) controls in the same way as in the sequential control strategy and as proposed in Section 4. This approach is chosen since they follow (mimic) the nature of integrating light, which has a multiplicative characteristic. The DC and AC iris lens controls are realized as a PID control system, because their response is not multiplicative by nature.

In a typical case, the front-end and DG control loops share the same control reference value (Wanted video Level, $Y_{WL_A} = Y_{WL_B} = Y_{WL}$). Let us further detail why we have chosen to close the DG loop at Point B and the lens/sensor control at Point A in Figure 7. Generally, and as already mentioned, this choice separates the video-level control
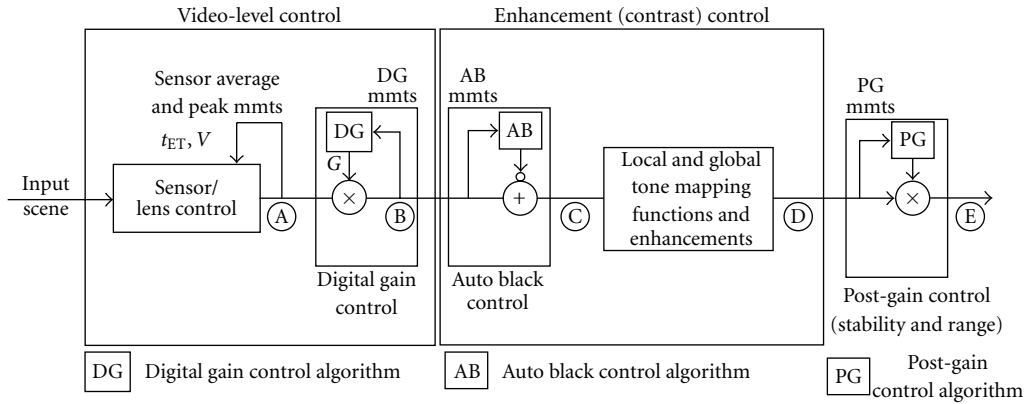
FIGURE 7: Overview of the proposed parallel video-level controller and enhancement processing chain.

loops from enhancement control loops (like Auto Black and Tone-mapping loops) and avoids introducing nonlinear elements (local and global tone mapping, Gamma function) within the video-level control loop. The enhancement control contains an Auto Black (AB) control loop, which sets the minimum value of the input signal to a predefined black level. This effectively lowers the video level setting after the wanted video level was already set by the user. This problem is typically solved by closing the lens/sensor control at Point C, hence, creating effectively a feed-back control to the sensor/lens control block at the start. Unfortunately, this leads to a control loop that includes other control loops like DG and AB.

This is exactly what we want to avoid. Therefore, we implement a saturation control which effectively increases the level at Point A, to optimize the SNR. As a consequence, AB now becomes a feed-forward loop which is much more stable and easier to control. An additional benefit of having the AB control loop separated from the ET control loop is that no additional clipping of the signal is introduced due to the corresponding level rectification (compensation of lowered video level as a result of the AB control) by means of the gain control (or perhaps increased lens opening or longer exposure time). When saturation control is performed (as explained in Section 6), the lens opening will be close to optimal (without introducing additional clipping), and so compensation for the intensity level drop due to the AB control becomes obsolete.

Let us now briefly describe the control strategy for the parallel control system. By making the wanted levels at Points A and B equal, hence $Y_{WL_A} = Y_{WL_B}$, we perform parallel level control. This action improves general camera performance and speeds up the video-level control. If the wanted video level after the sensor at Point A from Figure 7 cannot be reached due to exceeding the control range (maximum integration time or maximum lens opening), the remaining video level gap is compensated the same way as explained in the sequential control. This process is also dynamic, as the gain control loop is usually much faster than the Lens control, so that the wanted level at Point B Y_{WL_B} will become equal to the final Y_{WL}, while Y_{WL_A} will be converging slower

to Y_{WL}. As Y_{WL_A} gets closer to the Y_{WL}, the gain G returns to its nominal value, since more of the output level is achieved by the correct position of the lens. The above discussion on the dynamics and the parallel control strategy holds for the general case. However, there are cases which are very specific and where this strategy will not work sufficiently well. This leads to some special control modes which will be addressed in the next section.

6. Defining Optimal Average Luminance Level for Video Cameras: Special Control Modes

The previously described overall control strategies aim at achieving the average image luminance level to become equal to the user-desired average level. However, there are various cases when this scenario is overruled, for the sake of better visualization of important scene details. In particular, the desired average image luminance can be set higher than the user-desired average value. These cases occur when (1) HDR images are processed with standard dynamic range cameras, or (2) in case of low-dynamic range input scenes. Contrary to this, if we wish to control/limit the amount of signal clipping, the desired average image luminance can be set lower than the user set value. Both sets of cases require a more complex dynamic control due to the constant scene changes. This section describes special control modes for serving those purposes.

6.1. Processing HDR Images with Standard Dynamic Range Cameras. In general, there is a class of HDR scenes where the imaging sensor has a lower dynamic range than the scene of interest. These low- to medium-dynamic range sensors cannot capture the full dynamics of the scene without losing information. In such back-lighted or excessive front-lighted scene conditions, considerable luminance differences exist between the object(s) of interest and the background. As a typical result, the average luminance is dominated by the luminance of the background. Typical scenarios where this situation occurs are tunnel exits, persons entering the building on a bright sunny day while the camera is inside of the building, or in a video-phone application where a bright

sky behind the person at the foreground dominates the scene. In these cases, exposure problems are typically solved by overexposing the image so that details in the shadows have a good visibility. However, all the details in the bright parts of the image are then clipped and lost. In case when no object of interest is present, the exposure of the camera is reduced to correctly display the background of the image. This processing is called back-light compensation.

It becomes obvious that it is difficult to obtain correct exposure of the foreground objects if the average level of the overall image is used. This is why areas of interest are chosen in the image where measurements are made. The average image intensity is then measured as $Y_{\mathrm{AVG}} = \sum_i w_i Y_{\mathrm{AVG}_i}$. Two basic ideas can be employed. First, we can use selective weights w_i that depend on the classification of the corresponding measured areas. To correctly choose the weights, intelligent processing in the camera can consider only important image parts, which are identified as regions containing more information, based on features such as intensity, focus, contrast, and detected foreground objects. Second, we can detect the degree of back-lighting/front-lighting, as commonly exploited in fuzzy logic systems. In this section, we will describe these ideas including several possible modifications. The content of this subsection is known from literature but it is added for completeness and providing an overview. Our contribution will be discussed in the remaining subsections.

6.1.1. Selective Weighting. To cope with the HDR scene conditions in case of a stationary video camera, the user is often given the freedom to set the area weights and positions of several zones of interest. The idea is to set higher weights at areas where the interesting foreground objects are likely to appear, for instance, at moving glass doors of the building entrance. In cases when the darker foreground object is present in the zone of interest, it will dominate the measurement as the bright background will be mostly ignored and hence the image display will be optimized for the foreground. This explains why it is important to correctly set the metering zones, or otherwise, the object of interest will be underexposed and will vanish in shadows. We will now describe two general cases of selective weighting: (1) static weighting, when weights (and metering areas) are once selected and set by the user, and (2) dynamic weighting, when weights depend on the content of the metering areas. It will be shown that dynamic weighting, although more complex, provides better results than the static weighting.

Static Weighting. The user can assign higher weights to various areas of interest such that the desired amount of back-light compensation is achieved and good perception of objects of interest is ensured. Hence, if a certain object enters the area of interest, this is detected and the video-level control overexposes the image so that object details become visible. However, there are two principal disadvantages of this approach.

First, methods for back-light compensation detection and operation, that are based on the (difference of) measured signals in various areas of the image, have intrinsic problems if the object of interest is miss-positioned, or if it leaves the area of interest. The consequence is the severe underexposure of the important foreground object. To detect the change of object position, areas of interest are often set several times larger than the size of the object. However, the average intensity level of the whole metering window can be so high and the size of the object of interest can be very small, that insufficient back-light compensation occurs and the object details still remain invisible. Second, the changed object position can also give problems to the video-level controller, due to a considerable change of the measured signal because of the large differences in weights of the metering zones. These problems can be solved by dynamic weighting schemes.

Dynamic Weighting. A first solution is to split the areas of interest in several subareas and to apply a dynamic weighting scheme that gives a high gain to sub-areas that contain dark details and low gains to bright sub-areas. Likewise, we can ignore unimportant bright sub-areas which can spoil the measurement. To achieve temporal measurement consistency, sub-areas are usually overlapping, so that when the relevant object is moving within the area of interest, one sub-area can gradually take over the high weight from the other one where that object is just leaving. To additionally stabilize the video-level controller, asymmetric control behavior is imposed, so that when a low video level is measured (dark object entered the area of interest), the controller responds rapidly and the image intensity increases to enable a better visualization of the object of interest. However, if the object exits the area of interest, a slow control response is preferred, and the video level decreases gradually. Hence, if the considered object reenters the area of interest, the intensity variation stays limited. It is also possible to give priority to moving objects and nonstatic parts of the possibly changing image background. For example, when an object enters the scene and remains static for a certain time, we stop assigning it a high weight, so that the bright background is correctly displayed (the video level is lowered).

A second solution employs histogram-based measurements which do not use various areas to measure the signal. Therefore, they are not influenced by the position of the object. Based on the histogram shape or the position and volume of histogram peaks, unimportant background is given less weight [11, 12] and hence the video-level control is primarily based on the foreground objects.

A third solution is to adapt area weights based on the detected mode of operation. An example is presented in [13], where the luminance difference between the main object and the background is detected and represents the degree of back-lighting, defined by

$$D_b = \left[Y_{\mathrm{AVG}_0} + \max \left(Y_{\mathrm{AVG}_2}, Y_{\mathrm{AVG}_3} \right) \right] - \left(Y_{\mathrm{AVG}_1} + Y_{\mathrm{AVG}_4} \right). \tag{18}$$

Here, the Y_{AVG_i} measurements represent the average luminance values of various metering areas ($i = 0, \ldots, 5$) from Figure 8(a). If the degree D_b is large, higher weights w_i are assigned to the presumed main object areas 1 and 4, than
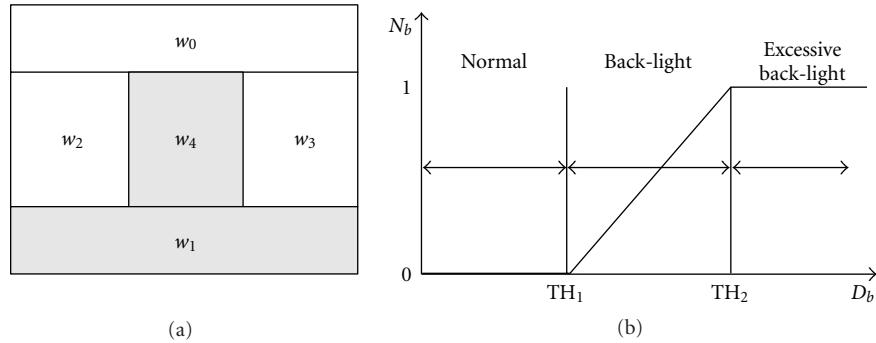
FIGURE 8: (a) Back-light compensation windows with corresponding weights that are varying based on the degree of back-lighting. (b) Relative weights are changed based on the detected state.

to the background areas 0, 2, and 3. This is achieved by a transfer function presented in Figure 8(b), which shows the additional weight of Regions 1 and 4, based on the degree of back-lighting D_b.

The dynamic weighting schemes (sometimes also the static) can provide a good exposure setting in many cases, but can also fail simply because they determine the importance of a certain (sub)area only by its average intensity value, which proves to be insufficient in many real-life situations. There is an extension to these approaches that offers an improved performance at the cost of additional system complexity. This extension involves a detection of important image regions that is based not only on the intensity but also on other features such as focus, contrast, and detected foreground objects, as with the content-based metering systems. Still, in this case, higher measuring weights are given to detected important objects. A second possibility is to use a rule-based fuzzy-logic exposure system, that incorporates various measurement types. These measurements include the experience of a camera designer, to define a set of distinctive operating modes. In turn, these modes optimize the camera parameters, based on extensive expert preference models. These possibilities are discussed in the following subsections.

6.1.2. Content-Based Metering Systems. The second class of systems that is aiming at the correct display of HDR scenes in standard dynamic-range image processing pipelines is content-based metering. In this approach, the objective is to distinguish relevant and/or meaningful metering parts in the image. The basic problem of the conventional metering systems is that large background areas of high luminance are spoiling the average luminance measurement, resulting in an underexposed foreground. The dynamic-weighting metering schemes can partially improve this drawback. However, a possible and more powerful approach would be to apply intelligent processing in the camera to better distinguish the important image parts.

In one of the approaches that is able to identify image regions containing semantically meaningful information, the luminance plane is subdivided in blocks of equal dimensions. For each block, statistical measures of contrast and focus are computed [1, 14]. It is assumed that well-focused

or high-contrast blocks are more relevant compared to the others and will be given a higher weight accordingly. In certain applications, features like face and skin-tones can also be used for the weight selection [1, 3, 14]. In cases where skin tones are absent in the image, classical average luminance metering is performed. This approach is often used in video applications for mobile phones, or in general, when humans occupy large parts of an HDR image. However, this rarely occurs for standard cameras. Especially in surveillance applications, the complete person's body is of interest, which is much larger than his face. This is why object-based detection and tracking is of high importance. Such background estimation and adaptation system discriminates interesting foreground objects from the uninteresting background by building the background model of the image [15, 16]. The model stores locations of foreground objects in a separate foreground memory that is used to discard background of the image from the luminance measurements. In cases when no objects of interest are detected, again classical average metering is performed [17]. These object detection models are much better than a simple frame-differencing method, since frame differencing can only distinguish parts of moving objects, and when moving objects suddenly become static, the detection completely fails. On the other hand, a background-modeling metering scheme enables much better results than the conventional approaches, since it is insensitive to the position of an object in the image and it maintains a correct exposure of that object of interest.

Let us elaborate further on object detection to provide better metering. The object-based detection is already challenging on its own, especially with respect to correct and consistent object detection and its correct detection behavior when scene and light changes occur. These changes happen by default when video-level control reacts to changes in the scene. For example, if a person enters the HDR scene that had correctly exposed background, the person will be displayed in dark color(s). After finding that the object of interest is underexposed, the video-level controller increases the average video level rapidly to enable object visibility. This action changes a complete image, which is a significant challenge for the subsequent operation of object detection during this transition period. To avoid erroneous operation

when an image change is detected, the background detection module should skip such transition periods and maintain the control as if it is measuring the image just prior to the reaction of the video-level controller. When the exposure level and scene changes are stabilized, regular operation of the system is resumed. During scene and exposure transition periods, the object detection system updates the background model with a new image background and continues to operate from new operation conditions. A similar operation mode occurs when the object of interest leaves the scene. These scene-change transition problems can be avoided by building the background subtraction models that do not depend on the intensity component of the image [18], which unfortunately is still in the experimental phase.

6.1.3. Fuzzy Logic Systems. Fuzzy logic can also be employed to achieve a higher flexibility, stability, and smoothness of control. Fuzzy logic systems classify an image scene to a scene type based on a set of features and perform control according to the classification. In this framework, a set of rules is designed which cover a space of all possible light situations and apply smooth interpolation between them. Fuzzy logic systems can incorporate many different types of measurements which can be taken over various spatial positions, in an attempt to achieve an optimal and smooth control strategy. Besides obvious measurements like peak white, average, median, and maximum intensities, less obvious examples of features that are used by fuzzy logic systems are the degree of back- and front-lighting (contrast) in different measurement areas [19, 20], colors of the objects and histogram shape [21], luminance distribution in the image histogram [22], and cumulative histogram of the image [12]. Various areas of operation are established, based on these measurements, and the system selects the appropriate control strategy, based on, for example, open/close lens, set gain, use of adaptive global tone mapping to visualize details in shadows [20, 22], and so forth.

Content-based metering systems and especially fuzzy logic systems can offer a very good and versatile solution for the difficult problem of obtaining an optimal image exposure, especially for standard dynamic-range image processing pipelines. However, inherently, both exposure systems have rather high complexity and they completely determine the design of the camera. Also, they are difficult to change, maintain, and combine with other camera subsystems. An unresolved problem of both conventional and content-based metering systems is the overexposure of the image to enable visualization of objects of interest. This drawback can be avoided by using the sensor dynamic-range extension techniques such as exposure bracketing, when capturing the scene and subsequent tone mapping for its correct visualization [23–25]. However, prior to explaining these solutions, we will describe how to employ the video-level control system in order to exploit the full benefit of these approaches.

6.2. Saturation Control. At the beginning of this section, we have explained that in particular cases the luminance is set

to a higher value than normal, which overrules the user setting. One possibility to do that is by saturation control. In this subsection we provide insight into a saturation control which increases the exposure of the sensor above the level needed to achieve a desired average output luminance value. We also describe two approaches for the compensation of this increased luminance level. Essentially, in addition to the regular video-level control, to achieve a better SNR, we propose to open the lens more than needed to achieve the wanted level Y_{WL} (required by the user), as long as signal clipping is avoided. If the lens cannot be controlled dynamically, we can employ a longer sensor exposure time. This action increases the overall dynamic range and is analogous to a white point correction [26] or white stretch function [27]. The idea is to control an image exposure to achieve a Peak White (PW) image value equal to $Y_{PW_{TH}}$, which is a value close to the signal clipping range. This approach is particularly interesting for Low-Dynamic Range (LDR) scenes, such as objects in a foggy scene (gray, low contrast). We call these actions saturation control and we can perform them only if the original PW value is below the desired PW value, hence if $Y_{PW} < Y_{PW_{TH}}$. The desired PW level $Y_{PW_{TH}}$ should not be set too high to avoid distortion of the video signal due to the excessive saturation of the sensor.

Our contribution is based on the previous statement that we aim at higher SNR, created by a larger lens opening, without introducing clipping. The approach is that we introduce a control loop with a dynamic reference signal, where the reference is adaptive to the level of a frame-based PW measurement. To explain the algorithm concept, we will reuse a part of Figure 7, up to Point C.

Algorithm Description. The purpose of our algorithm is as follows. The saturation control is effectively performed in such a way that it increases the wanted average video level Y_{WL_A} (from Figure 7) to make the PW of the signal equal to a predetermined reference level $Y_{PW_{TH}}$. This is achieved by setting the desired average value after the sensor (Point A) to a new value that we will call Wanted Level saturation (Y_{WL_s}). The key to our algorithm is that we compute this wanted level Y_{WL_s} with a following specification:

$$Y_{WL_s} = Y_{WL_A} = \max\left(Y_{AVG} * \frac{Y_{PW_{TH}}}{Y_{PW}}, Y_{WL}\right). \qquad (19)$$

The max function is used to enforce one-side control, which allows only output average values higher than the one set by the user. The meaning of (19) is that this dynamic-reference control loop modifies the desired average value of the loop at Point A with frame-based iterations and effectively controls the camera video level to an operational point such that the following holds: the measured PW of the image signal Y_{PW} becomes equal to the predefined PW value; hence $Y_{PW} = Y_{PW_{TH}}$. Hence, the system control acts as a convergence process. As a refinement of the algorithm, we set a limit for the level increase, that is, a maximum saturation level, which is equal to $r \cdot Y_{WL}$, where Y_{WL} is a wanted average video level as set by the camera user. Parameter r is a real number.

The rationale behind this control formula is that in the space of average and Peak White measurements, the current state is represented as a point (Y_{AVG}, Y_{PW}), of which the involved parameters have the mutual relation $Y_{PW} = g_{PA} * Y_{AVG}$, where $g_{PA} > 1$. If only reflective objects are present in the scene (no visible clipped light sources or specular reflectance areas), g_{PA} is nearly constant. Hence, we are effectively evolving the starting point (Y_{AVG}, Y_{PW}) to an end point $(Y_{WL_s}, Y_{PW_{TH}})$, where it also holds that $Y_{PW_{TH}} = g_{PA} * Y_{WL_s}$. The dynamic-reference control loop from (19) changes the desired average video level and converges to the average luminance level which corresponds to the peak white of $Y_{PW_{TH}}$ (in both operation points, g_{PA} is identical). Only when PW is clipped to its maximum value, the relation between the measured Y_{AVG} and Y_{PW} is distorted compared to their actual relation in the scene. In such cases, the PW measurement does not correspond to the real PW value in the scene, but this only reduces the control speed and is hence good for the overall stability.

Consequence of the Saturation-Control Algorithm. The increase of the video level after saturation control has to be compensated to ensure that the image signal does not entirely pass through the compression part of the gamma function (wrong working point). Two approaches for compensating the increased video level are proposed: (1) using a gain value smaller than unity in the DG control loop from Figure 7, or (2) using Auto Black control (in the AB control loop).

In the first case, when digital gain is used for the compensation, the maximum saturation level of $r \cdot Y_{WL}$ is coupled to the minimum negative gain that is equal to $1/r$. (Technical camera experts call the situation with the gain smaller than unity, a *negative gain*, as the gain is often expressed in dB units.)

The second option for the compensation of the increased level is the use of the Auto Black control that sets the darkest parts of the signal to the proper black level. This processing approach increases the amount of subtracted black level, as compared to situations when only negative gain is used. A benefit of this approach is the increased signal contrast and the corresponding improved image fidelity. However, the increased video level is not compensated (in contrast with the negative gain concept) and the output video level is not constant and is higher than the input level (e.g., $Y_{WL_C} > Y_{WL}$ in Figure 7). However, we claim that the improved image contrast is more important than the constant video level being equal to the reference level Y_{WL}, since the video level setting is anyhow subjectively set. Let us now explain these two concepts for the compensation of the increased level in more detail.

To better explain the effect of saturation control (compensation), we present Figure 9. Function 1 depicts the histogram of the original signal after the video-level control and Auto Black control, but without the saturation control. After the saturation control (Auto Black control is not applied), we obtain the image histogram, as depicted by Function 2, where the PW of the signal is placed at the saturation level $Y_{PW_{TH}}$ (chosen as 90% of the maximum signal level). The
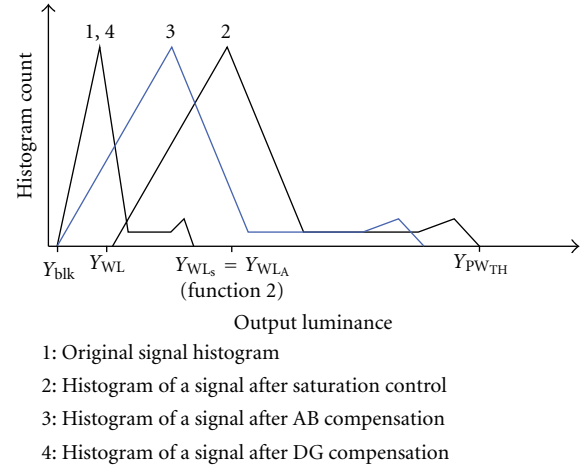

FIGURE 9: Histograms of the original image and of the resulting image after saturation control, and two compensation strategies (negative gain and Auto Black control).

1: Original signal histogram
2: Histogram of a signal after saturation control
3: Histogram of a signal after AB compensation
4: Histogram of a signal after DG compensation

saturation control expands the dynamic range of the whole signal in the analog domain, leading to new digital values in the signal (opposite to the case when the image signal is multiplied with a digital gain). We achieve a better SNR, since we expose the signal longer than needed to achieve the wanted video level of the user. The improved SNR is needed for enhancement and tone-mapping steps afterwards (enhancement control in Figure 7). Let us reconsider the above two options for saturation compensation, but now in the framework of Figure 9.

The first option to compensate for the level increase is basically equal to multiplying the image signal with a gain smaller than unity. Consequently, we use the Auto Black function afterwards, to compensate for the remaining image offset and put the minimum image luminance to the desired black level. As a result, the output image histogram is virtually identical to the starting image histogram (depicted with Function 4). However, although the output image has the same content, the SNR is increased because of the longer exposure time. Multiplication with a negative gain occurs automatically by the DG control loop, since the DG loop reference level is set to the user selected level ($Y_{WL_B} = Y_{WL}$ in Figure 7).

The second option to compensate for the level increase is to shift the video signal downwards in amplitude by means of the Auto Black (AB) control, instead of compensating for the increased video level. Hence, we achieve the correct black level (depicted with Y_{blk} at the bottom left in Figure 9), resulting in the output histogram Function 3. This compensation strategy is enforced by setting $Y_{WL_B} = Y_{WL_A} = Y_{WL_s}$. It can be noticed that the histogram of Function 3 has a larger dynamic range, and thus better contrast than Histogram 4.

A disadvantage of the second option based on AB control is that it gives undesirable effects in certain cases. Those cases occur in two situations: (1) large AB values are subtracted in case of very foggy scenes and (2) color faders are used for video signals close to the saturation level. Let us now address both cases.

(i) For example, if very large AB values are subtracted, this leads to increased noise visibility. Photon shot noise that is dominant for higher signal values is proportional to the square root of the signal amplitude and when the whole image signal is shifted down by the AB control, parts of the signal with higher noise values are shifted to the lower luminance values where lower noise amplitudes, are expected. (This is not the case with saturation compensation using negative gain, since the noise is scaled back to its original amplitude before saturation control.) This effect is further amplified by global and local tone mapping functions, creating the impression that the noise amplitude in the signal is quite large, giving a lower SNR impression. This can be partially alleviated by reducing the strength of the image enhancement and hence decreasing the perception of the noise.

(ii) Nonuniform saturation effects always occur in (near) clipped parts of the signal. In these cases, for CMYG sensors one line is clipped and the other is not, which creates nonlinear effects and an artificial "contrast" between those lines. In some cases, both lines are not clipped, but then the color fader, typically used in cameras, operates differently for subsequent lines (fading more color than in the other line). Some color distortion effect can also be observed with a Bayer type of sensor (a sensor with alternating RGRG and GBGB pixel lines), where the same effect can be observed for the saturation of individual color pixels. When the AB subtraction is used, the increased contrast between lines (pixels) is not reduced and becomes quite visible, but now at low intensity levels. This visibility does not occur when negative gain is used for the level compensation, since the "contrast" between lines (pixels) is reduced.

To cope with these potential problems, an intermediate solution can be used where the AB compensation is used completely if the compensation gap is small, and when large, a negative gain is gradually introduced. This intermediate solution is not further elaborated here.

6.3. Peak-Average-Based Control

6.3.1. Standard Peak-Average Control. The conventional video-level controller tunes the camera system such that an average luminance level of the measured area (Y_{AVG}) becomes equal to a predefined Wanted Level (Y_{WL}), that can be set by a user. One of the pillars of our "optimal exposure strategy" is to additionally use a Peak White measurement Y_{PW} and achieve an average video level intensity that leads to less or even no clipping of the video signal. This is especially beneficial for HDR cameras which create a video signal having a sufficient SNR for subsequent local and global tone mapping operations. We call this operation a Peak Average (PA) control. The PA mechanism should lower the average (and PW) video level of the image to mostly avoid clipping and only allow it in a small fraction which is acceptable for the user.
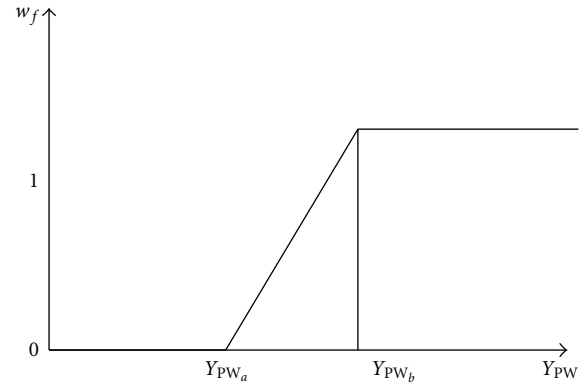

FIGURE 10: The peak white weight factor w_f is used to disable the use of PW measurement when it is small.

Let us first discuss a common approach for PA control. To achieve lowering of the video level, one possibility is to mix an average measurement Y_{AVG} with an often much higher Peak White measurement Y_{PW}, which results in the Peak Average measurement Y_{PA}, where

$$Y_{PA} = (1 - w) \cdot Y_{AVG} + w \cdot Y_{PW}, \qquad (20)$$

with $w \ll 1$. This method substitutes the average measurement in the controller with the effectively increased PA measurement, that now becomes the total level measurement, where $Y_{PA} > Y_{AVG}$. Increasing the relative weight factor w leads to an increased importance of bright pixels, which effectively results in an increase of the PA measurement. When detecting the increase of the intensity measurement, the video-level controller lowers the average intensity of the image, enabling visualization of important bright pixels and resulting in the *fewer clipped pixels*. The parameter w can be seen as a user-based setting, which tunes towards the user preferences for a particular scene.

As a refinement, to ensure that the average video level will be lowered only when clipped pixels exist in the image, we make weight w dependent on the PW measurement, as in Figure 10. It is important not to lower the video level when very bright (or even clipped) pixels are absent from the image. Hence, we introduce the weight factor w_f such that $w = w \cdot w_f$, where w_f disables the usage of the PW measurement for values where $Y_{PW} < Y_{PW_a}$ and allows its full use if $Y_{PW} > Y_{PW_b}$.

However, the previous common approach shows disadvantages when employing the PW signal in such a way. This is particular critical as the PW value can change much faster than the average value. Hence, the idea of mixing the potentially fast-changing PW measurement with the average exposure measurement lowers the stability of control. This effect is giving significant problems to the lens control due to a nonlinear nature of the lens transfer characteristics. As such, a better solution of incorporating the PW information to minimize the signal clipping is required and it can be obtained by employing the previously discussed saturation control.

6.3.2. New Proposal for a Peak-Average Control. Our contribution is based on the previous requirement that we aim at creating video signals with less or even no clipping. The approach is that we operate the saturation control in parallel with the peak-average control. As a consequence, we can modify the standard PA control to make it simpler and more stable.

Algorithm Description. The purpose of our algorithm is as follows. When saturation control is used in parallel with the peak-average control, the overall control has two regions: *PW control* region which is active when $Y_{\mathrm{PW}} \geq Y_{\mathrm{PW_{TH}}}$, and the *saturation control* region valid for $Y_{\mathrm{PW}} < Y_{\mathrm{PW_{TH}}}$. As the objective of the PA control is to lower an output average level to reduce signal clipping, we now allow an output average values lower than the one set by the user. Hence the maximum function is now not used, compared to the saturation control only, as given in (19). Hence,

$$Y_{\mathrm{WL}_s} = Y_{\mathrm{AVG}} * \frac{Y_{\mathrm{PW_{TH}}}}{Y_{\mathrm{PW}}}. \qquad (21)$$

Instead of mixing the PW measurement with the average measurement, the PA control is achieved by reducing the desired average video level Y_{WL} to a value of Wanted Level peak (Y_{WL_p}). The reduction is implemented with a scaling factor p, so that

$$Y_{\mathrm{WL}_p} = \frac{Y_{\mathrm{WL}}}{p}, \qquad (22)$$

with $p > 1$. For example, for a maximum signal clipping reduction effect, we can set $p = 4$, no clipping reduction will be $p = 1$, whereas the intermediate values are interpolated. As a result, if the overall control is in the PW *control* region ($Y_{\mathrm{PW}} \geq Y_{\mathrm{PW_{TH}}}$), the camera video level contributing factors (lens opening, exposure time, gain) will be lowered and the average (and PW) level of the image will decrease, reducing the amount of signal clipping. However, if the PW level drops below the PW saturation level as in $Y_{\mathrm{PW}} < Y_{\mathrm{PW_{TH}}}$, the overall control will enter the *saturation control* region, which will again increase the average video level to make the PW of the signal equal to $Y_{\mathrm{PW_{TH}}}$. Likewise, lowering of the PW level of the signal will be stopped and Y_{PW} will be set to the saturation level. This control behavior can be imposed if we set the desired average video level at Point A from Figure 7, to a value

$$Y_{\mathrm{WL}_A} = \max\left(Y_{\mathrm{WL}_s}, Y_{\mathrm{WL}_p}\right). \qquad (23)$$

The original proposal of a mixing-based PA control has a control stability problems, since unstable PW information directly influenced the measurement signal that was used in the control. With the new proposal, the control stability is improved as we are modifying the desired average video level Y_{WL} instead. We can now impose better restrictions on the speed of change of this desired value, as influenced by the value of the PW measurement.

7. Extending the Sensor Dynamic Range

The dynamic range of an image signal is defined as the ratio between the saturation value of the sensor and the value of the noise level [23]. A good linear imaging sensor in CCD or CMOS technology can capture scenes with the dynamic range of 74 dB which is sufficient for most applications. However, for HDR scenes, for example, such as outdoor scenes with bright sunlight, a larger dynamic range should be captured by the sensor in order to obtain images with a satisfactory quality. For example, the contrast ratio in a sunny outdoor scene can be as high as 1000 (60 dB). For the lowest level in that image, the SNR needs to be 40 dB in order to achieve an acceptable quality. Therefore, the total sensor dynamic range should be about 100 dB. For a given CCD/CMOS sensor, the saturation voltage (corresponding to maximum image brightness) is fixed, leaving us only with the possibility to reduce the noise level in order to increase the dynamic range. Creating such HDR images reduces the need of back-light compensation strategies described in the previous chapter, since this image has sufficient SNR for consequent tone mapping enabling good visualization of details in dark image parts. The exposure control strategy with these images is to use peak white control to prevent (excessive) clipping of the signal. Allowing some clipping can accommodate for very bright light sources visible in the image.

There are several often used techniques for extending the dynamic range of the sensor [28]. First, there is a group of a nonlinear response (OECF) sensors, such as *Logarithmic response sensor*, *Multiple-slope sensor* which approaches a logarithmic response by a piece-wise linear curve having usually 3 segments, and a *Linlog* sensor that behaves linearly for low light intensity and logarithmically for higher intensities.

Second group is made of a linear-response sensors, such as a *Dual-pixel sensor* and a *Linear sensor using exposure bracketing*. *Dual-pixel sensor* is made of two interlaced arrays of pixels with different responsiveness (high and low). It produces two images acquired at the same time, which are then combined in a higher-dynamic-range image. In some cases a single sensitive element has two (or more) storage nodes to store the multiple images. *Linear pixel and exposure bracketing* is a standard approach in which two (or more) images with different exposure (integration time) of the sensor are taken after each other and afterwards merged. In video applications, there are two general possibilities for this action. If we can sacrifice the frame rate and halve it, then we can consecutively take long-exposure image in odd frames and short-exposure image during the even frames (or the other way around). Otherwise, to keep the frame rate, we have to take two images after each other during the same frame. To prevent disturbances, long-exposure image has to be obtained during the active video period, and short-exposure image should be recorded during the vertical blanking period. (Some new CMOS sensor architectures allow taking the short-exposure image during the active video period, which can reduce image blur.) This immediately poses a restriction on the duration of the short
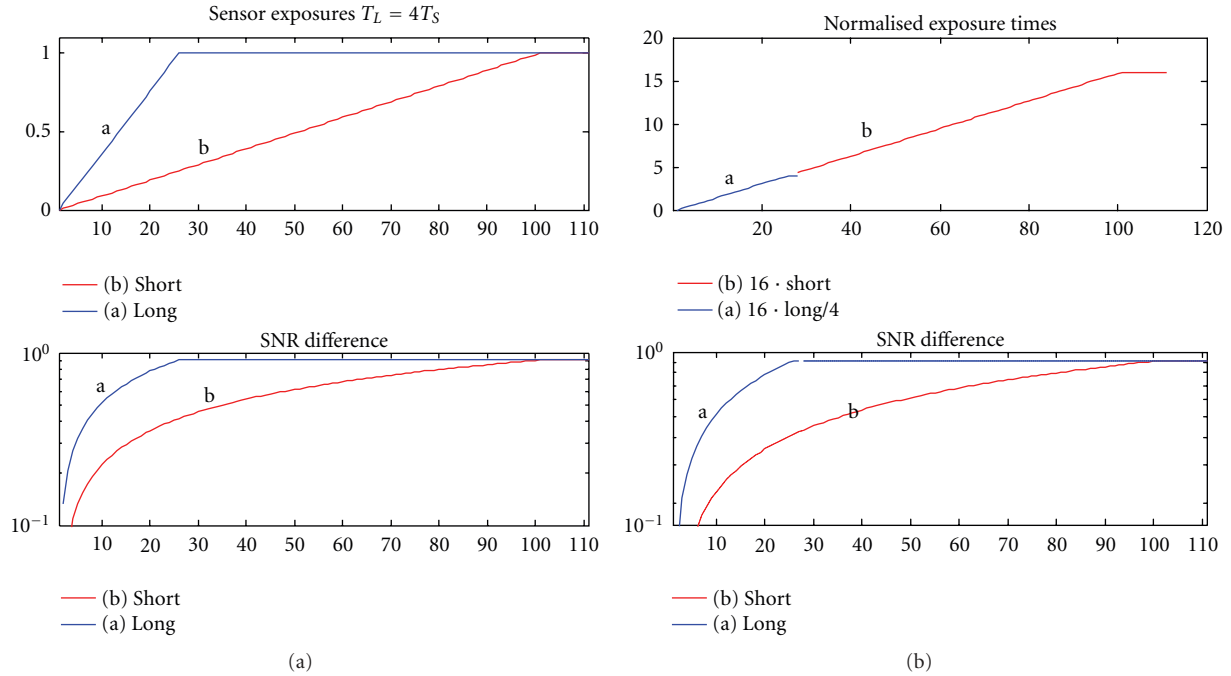
FIGURE 11: (a) Example of a dual exposure process of an input pixel intensity, where the long exposure time T_L equals four times the short exposure time T_S. The top subfigure portrays the sensor output values for long- and short-exposure images, while the bottom subfigure depicts the corresponding SNR curves. (b) Images originating from two exposure times after weighted normalization, when using switching between two exposures. The top subfigure portrays the sensor output values for a combined image, while the bottom subfigure depicts the resulting SNR curve.

exposure image, which has to be obtained before the end of the frame.

One of the main criteria for choosing the adequate sensor type is its sensitivity and the flexibility. For example, using nonlinear response sensors implicitly "builds-in" a certain output-input characteristics (tone mapping) of the original image, which is not desired for high-fidelity imaging, and has to be removed. Our desire is to have the freedom to chose the transfer characteristics based on the image content, to achieve the best possible output quality and visibility of details. Furthermore, in terms of sensitivity, a dual-pixel sensor is not acceptable since it often has lower sensitivity due to a fact that high- and low-sensitivity pixels have to share the area of the pixel element.

In addition to the complexity, sensitivity, and flexibility, the color performance is still very important when choosing the method for extending the dynamic range. In case of a nonlinear pixel response (e.g., logarithmic, multiple-slope or linlog sensor), ratios between the color pixels are nonlinearly changed and mutual relation between different colors is distorted. Furthermore, if intensity of the pixels is changed, color values are also changed nonlinearly. This generally implies use of a linear-response sensors or exact inversion of the OECF of the nonlinear sensors.

For these reasons, we choose to further employ exposure bracketing as a method to extend the dynamic range of the image since we can produce linear sensor output with limited amount of color distortions. In the following section we will focus on the exposure bracketing technique.

7.1. Exposure Bracketing and Image Merging. In this subsection, we discuss exposure bracketing and the creation of double-exposed images to reduce the sensor noise level. A popular concept known from the work of Alston et al. [29] is a double-exposure system, where two images are captured after each other. Images are taken with a short and a long exposure time, where the ratio between the exposure times varies from 4 to 32. For example, this is possible by means of a special sensor that physically stores images captured with two exposure times by one sensor. The combination of these two images results in a good SNR in the dark parts of the image, due to the long exposure time of one of the captured images. Furthermore, there is almost no clipping in the bright parts of the image, since the other image is captured with a short exposure time.

An example of this process is given in Figure 11(a), where we can observe a graphical representation of an image taken with a long exposure time, which has a good SNR but it is clipped in bright parts of the image already at low input levels. We can also notice a short-exposure image, which is a standard image with a lower SNR, that is under-exposed in dark parts. The long- and the short-exposed images are combined into a single image, and the simplest way to combine them is to assign an individual weight to them, to retain the luminance relations occurring in the real scene (see the continuing intensity curve in Figure 11(b)-top). For example, if the long exposure time equals four times the short exposure time, then we would give the short-exposed image four times more gain than the long-exposed
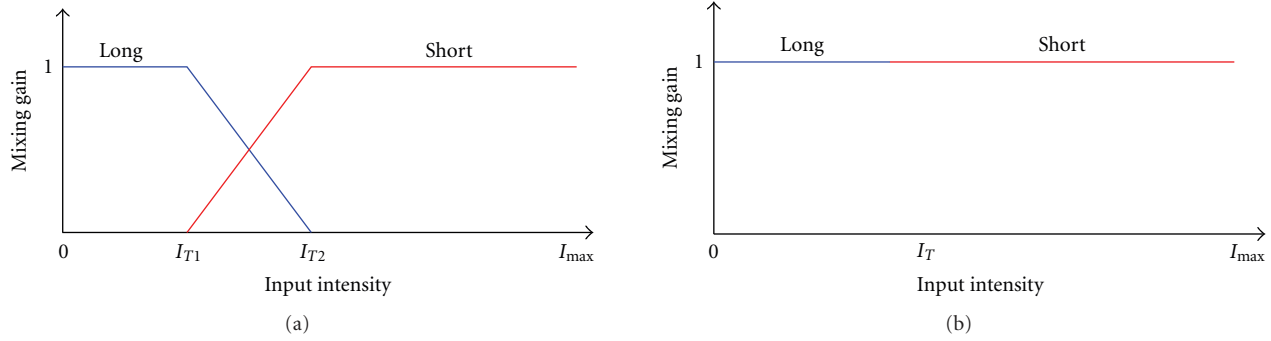
FIGURE 12: Merging long- and short-exposure images into one image: (a) mixing and (b) hard switching.

image, to retain the luminance relation. As a result, after combining these two images into one image, the first quarter of the input intensity range is derived from the long-exposure image and the other three quarters are derived from the short-exposure image (Figure 11(b)-top). Consequently, a difference in SNR between short- and long-exposed image parts occurs (Figure 11(b)-bottom)).

An additional important consideration is the detailed mixing or combining short- and long-exposure images into one image. There are several possibilities by which multiexposed images can be merged. For a complete overview, we recommend the reading of [25] and [3]. We will describe two basic methods: mixing of images and hard switching between images. Figure 12(a) depicts a soft switch between long- and short-exposure images, where two images are mixed in a transition region with weights proportional to their local intensity values. Figure 12(b) presents a hard switch between two images: if the input level is lower than a threshold I_T, a pixel from the long-exposed image is used, and vice versa. According to the example from Figure 11 in which the exposure ratio of four was used, the setting of threshold parameter is $I_T = (I_{T1} + I_{T2})/2 = I_{max}/4$.

However, the exposure bracketing technique has the following drawbacks and challenges and we have to deal with three problems: (1) nonlinearity of the sensor output, (2) motion in the scene, and (3) the influence of light coming from nonconstant light sources. We already discussed how to solve the problem of sensor nonlinearity in previous publication [23]. Due to its importance, we will briefly discuss a problem of motion in the scene and present some typical solutions. Our contribution will be given for the third problem, which will be presented in the following section.

7.2. Motion Problems and Misregistration. One of the problems when combining various exposure images is motion in the scene, since the intensity of pixels changes over time due to the motion, leading to differences between long- and short-exposed image pixels. Consequently, a misregistration appears and the linear relationship between two differently exposed images is no longer valid. In such a case, the mixing scheme performs more smoothly, unlike the switching scheme where misregistration effects may become visible. When motion is absent from the scene, it can be more advantageous to use a hard-switch threshold (Figure 12(b)),

since then the corruption of the SNR in the transition area does not occur. An example of how motion can be handled in the image fusion process is presented in [25, 26, 30, 31]. The easiest way to partially solve the motion problem is to discard the long-exposed image part with motion and use only the short-exposed image for those problematic pixels. Since the short-exposed image is integrated over less time than the long-exposed image, it exhibits much less motion problems, but it has a worse SNR. To improve the consistency of this approach and use a single, short exposure for the complete moving object, an image-differencing algorithm can be used, followed by a region-growing technique [26, 31]. All the proposed methods can improve the final exposure bracketing result. However, some motion errors always remain and can be observed and appear as colored regions at the edges of moving objects.

Furthermore, most of these approaches work well for static digital images but cannot work well with digital video cameras, where large object or camera motion is involved and no video delays are allowed. To solve for camera motion, Lasang et al. [31] proposed to use a feature-based image alignment technique [31]. Unfortunately, its complexity prohibits the real-time application, so that it remains a nonsolved problem in real-time imaging.

In addition, the choice of using the short-exposed image when local motion is detected can lead to much worse results in the presence of artificial light sources, such as fluorescent lights. We will discuss this problem in the following section.

8. HDR Imaging Problems with Motion and Fluorescent Light Sources

8.1. Problem Description. In this section, we develop a performance improvement for a double exposure camera in the presence of fluorescent light sources, which may give intensity flickering and color-error effects. Let us describe how this problem evolves.

The mixing of the short- and the long-exposure image as discussed in the previous section presents problems in case of specific lighting conditions and motion. Problems occur in the presence of artificial light sources, particularly fluorescents, where light intensity and color are strongly modulated at twice the local mains frequency. If the integration (exposure) time of the sensor is not a multiple
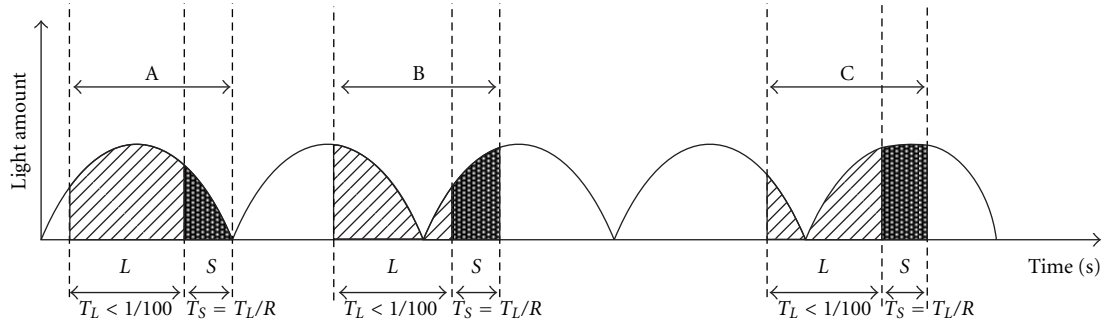
FIGURE 13: Due to a slow drift in the mains frequency or a variable exposure time, the amount of gathered light varies in time. This results in a light flicker and variable coloration due to the various positions of both long and short exposure times with respect to the oscillation period of fluorescent light.
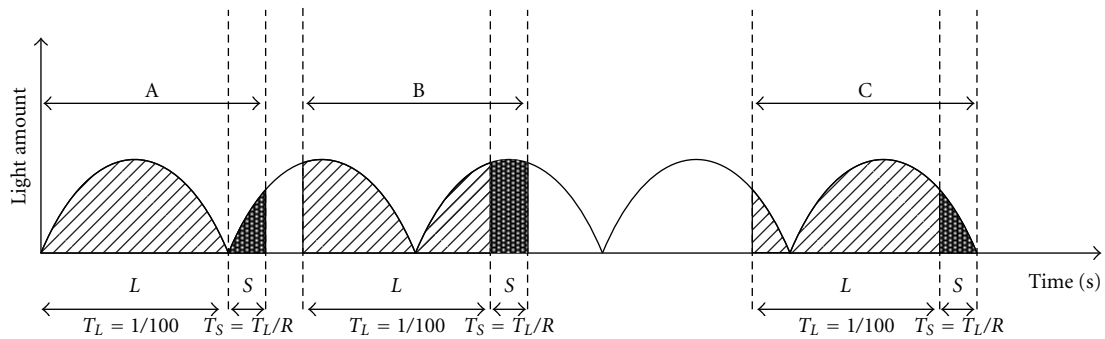

FIGURE 14: The long exposure time is set to a multiple of the fluorescence light period (e.g., to 1/100 s or 1/120 s, depending on the mains frequency). The short exposure time has a variable level output and color content, depending on the sampling moment.

of the period of the fluorescent light source, the amount of integrated light varies per field (frame), which results in temporal intensity flickering and changing colors. The frequency of light flickering is either 100 Hz or 120 Hz, according to national mains standards, and can vary up to 2% of the mains frequency. To cope with this problem, the sensor integration is set manually or a flicker detection mechanism is activated, so that the integration time becomes an integer number of the fluorescence period ($n/100$ s or $n/120$ s consecutively, depending on the national mains frequency, $n = 1, 2, 3, \ldots$). This special operation is a valid solution in single-exposure time sensors, but not in multiexposure time sensors. For example, if the longer exposure time is $T_L = 1/100$ s, the shorter exposure time will be several times shorter (the exact relation depends on their ratio R) and will not be adequate for the operation in fluorescence light condition. In this section, we propose two solutions to improve the performance of a double exposure camera in the presence of fluorescent light sources.

In Figure 13, we can observe the influence of the fluorescent light source on the amount of light in the image. In case of 50 Hz mains frequency, the output light oscillates with 100 Hz frequency. If the long exposure time is not a multiple of the fluorescence period, the amount of integrated light can vary per field due to a slow drift in the mains

frequency. Here, T_L and T_S represent long (L) and short (S) exposure time periods, interlinked with the ratio R, as in $T_L = R \cdot T_S$. This is why a long exposure time has to be set to a multiple of the fluorescence period, for instance, 1/100 s in a 50 Hz mains area (as in Figure 14) and 1/120 s in a 60 Hz mains area. For both integration cases, frequency/phase drift of the mains does not influence the amount of gathered light during the long exposure period. However, although this provides a good solution for the long exposure period, the light gathered within the short exposure period is inevitably sampled at various positions of the oscillation period of the fluorescent light. Let us now detail on three problematic aspects when fluorescent light sources appear in the scene.

First, due to a slow frequency/phase drift, the amount of light gathered within a short exposure time period is variable and can be observed as the low-frequency flicker in brighter parts of the image. Furthermore, the inevitable intensity differences between the long- and the short-exposed images in this condition are detected and considered scene motion. For pixel intensities below the threshold I_T, the intensity image would normally be derived from the long-exposure image. However, the output image in these "motion" regions is constructed from the short-exposure image to reduce motion blur, which introduces intensity flickering, even in the darker image parts.

Second, the output of the fluorescent light tube is also not constant in color but has different colors within the period. Depending on the type of fluorescent light, for example, when switching on, the fluorescent light is more red and yellow (Period A in Figure 13), while at the peak of its periodic interval it is white (Period B), and at the end (switching off) it turns blue (Period C). This property effectively creates various colors in image parts that are normally colorless.

Third, the outputs of the used light sources do not comply with a $\sin^2$ characteristic but often exhibit various distortions at moments of switching on and off, as the measured curves in Figure 15. Consequently, our algorithm has to be very robust and should not be influenced by all possible distortions and interferences.

We briefly outline a solution here, which is based on two stages, where the second stage consists of two options. If fluorescent light is detected in the image, solving the problem of low-frequency intensity flicker and variable coloration occurring in the short exposure periods, the first indispensable step is to make the long exposure time equal to a multiple of the fluorescence light period (e.g., to 1/100 s or 1/120 s, depending on the mains frequency; see Figure 14). This involves the detection of fluorescent light to determine the long exposure time. Afterwards, in the second stage, we propose two following basic options.

8.1.1. Shifting the Short-Exposure Image Out of the Display Range.
Problematic image parts which are constructed from the short-exposure image are removed from the display range as much as possible, by modifying the gain control. Besides this, the image color saturation in bright parts is reduced.

8.1.2. Fluorescence Locking.
This is performed such that the time interval where the short-exposure image is captured is always positioned at the optimal moment within the fluorescent light period, namely, at the peak (maximum) of the fluorescent light output (see the dark interval at the top in Period B in Figure 14). Hence, we ensure that light integrated during the short exposure time is nearly constant over time and has a correct color (not influenced by the fluorescence light source).

In the next subsection we will discuss the first solution consisting of the first stage and followed by the first option of shifting the short-exposure image. Afterwards, in Section 8.3, we will present the florescence locking proposal. However, the latter concept is very recent immature work, of which only the concept is proposed and explained. Some further details are found in publication [32].

8.2. Algorithm 1: Detection of Fluorescent Light and Shifting the Short-Exposure Image.
Figure 16 presents the concept of the proposed algorithm for the detection of fluorescent light in the scene and then applying then shifting the short-exposure image out of the display range. Although it is possible to manually trigger a fluorescent mode of processing, we omit this option and directly pursue the design of an automatic fluorescence detector.

First, measurements of intensity errors and color errors present in the short-exposure image are performed. These errors will show sinusoidal behavior in the presence of fluorescent light. However, motion in the scene and other light sources can significantly affect the intensity- and color-error measurements. For this reason, we have to perform filtering of these measurements and ensure more accurate and reliable results. After the filtering stage, we detect the frequency, amplitude, and temporal consistency of the error signals. The algorithm for the fluorescent light detection uses these measurements and makes a decision about the existence of the fluorescent light in the scene. When fluorescent light is detected, as a second step, we shift the corrupted short-exposure image out of the display range and apply color faders to remove any remaining color errors.

In the remainder of this subsection, we will describe the steps of the complete algorithm in more detail.

8.2.1. Intensity- and Color-Error Measurements.
We have already described that in the presence of fluorescent light, the long-exposure time should be set to a multiple of the fluorescence period. The intensity of the long-exposed image will then be constant because the 1/100 s (or 1/120 s) integration time equals the duration of a 100 Hz (120 Hz) cycle of the fluorescent light source. The short-exposed image may contain large intensity and color errors, as it integrates only a small part of the squared sine wave (e.g., 10 ms/R for the 100 Hz fluorescent cycle; $R = T_L/T_S$). If the camera is not locked to the mains frequency, the errors will sinusoidally change over time. To be able to detect fluorescent light conditions and to adapt the dual exposure processing, we propose two measurement types, which are performed each field/frame.

(i) *Intensity-Error Measurements.* calculate the amplitude differences between the long- and the short-exposed pixels in several intensity regions.

(ii) *Color-Error Measurements.* They measure the color error by accumulating the differences in color between the long- and the short-exposure pixels within a certain intensity range.

The differences in intensity and color between the long- and the short-exposed pixels depend on the exposure times and phase relation between the exposure moments and the mains frequency. We can perform these measurements only in the intensity areas where both long- and short-exposed pixels are not saturated (for input intensity smaller than I_T in Figure 12). Both intensity- and color-error measurements will display similar sinusoidal behavior under fluorescent lighting condition. A more detailed discussion and the actual implementation of the error measurements are presented in the appendix. Here we briefly discuss these two measurement types. To increase robustness, we will use the outcome of both measurement types simultaneously for the fluorescent detection algorithm.

(a)   (b)

FIGURE 15: We depict the intensity output of various light sources. They do not follow a $\sin^2$ characteristic. Additionally, some light sources exhibit various distortions at the moments of on and off switching. In (a), three fluorescent lamps are presented, and in (b) the output of a Sox lamp is shown.
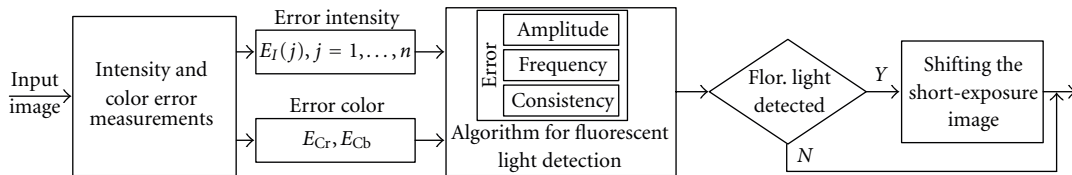

FIGURE 16: The proposed algorithm for the detection of fluorescent light and shifting the short-exposure image.

Intensity Error Measurements. The average values of the corrected long-I_{Lc} and the short-exposure pixels I_S in n intensity regions are accumulated. We also count the number of pixels that are accumulated. The differences between these measurements at n intensity levels can be plotted as a waveform, which will show a periodic (sinusoidal) behavior in the presence of artificial light in the considered intensity range. We will call these intensity differences Error Intensity signals E_{I_j} for $j = 1, \dots, n$, which are defined as

$$E_{I_j} = \sum_j (I_{Lc}(j) - I_S(j)), \quad j = 1, \dots, n. \tag{24}$$

The value of this measurement will be used as the input signal for the fluorescent detection algorithm.

Color-Error Measurements. The color-error measurement involves the differences in color between the short- and long-exposed pixels. For example, while color originating from the long-exposure image is white, it changes from red to blue in the short-exposure image. We will call these color-error measurements Error Color signals E_C. Color difference signals are created as the differences between two subsequent neighboring pixels For example, for a Bayer type of image sensors, color differences between R and G channels ($R - G$) can be compared (subtracted) for the long- and short-exposed pixels, and then produce the red error E_{Cr} signal. (Similar reasoning holds for the complementary

mosaic (Cyan Magenta Yellow Green, CMYG) sensor.) The same holds for the difference of B and G color channels ($B - G$) between the long- and short-exposed pixels, yielding a blue error signal E_{Cb}. Each field/frame, these differences are measured separately for red and blue lines and also the number of pixels that are accumulated is counted. This leads to the following specification:

$$E_{Cr} = \sum_{\forall \text{pix window}} \left((R - G)_{Long} - (R - G)_{Short} \right),$$

$$E_{Cb} = \sum_{\forall \text{pix window}} \left((B - G)_{Long} - (B - G)_{Short} \right). \tag{25}$$

As mentioned, if color-error measurements show periodic (sinusoidal) behavior, this indicates the presence of fluorescent light. Example of color-error measurements are shown in Figures 17–19. The horizontal axis represents the time scale at frame resolution.

In Figure 17, we present color difference errors of the signals Cr and Cb in a typical scene with a fluorescent light source. In Figure 18, we show the influence of motion on the same color errors: a noticeable disturbance of the measurement can be observed. Finally, in Figure 19, we depict the influence of other light sources on the color-error signals. The errors shown in this figure are recorded under essentially the same conditions as in Figure 17, where the only difference in the set-up is an active LCD screen visible
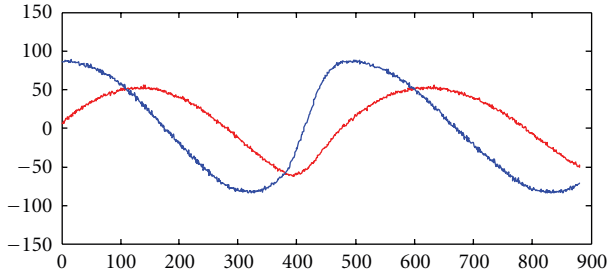
FIGURE 17: The time variation of the average color differences measured in a small region, for Cr and Cb in a typical scene with fluorescent light source.

FIGURE 19: The color-error measurements can also be disturbed by other light sources. Errors in this figure are recorded under the same conditions as in Figure 17. The only difference is an active LCD screen visible in a part of the scene, causing high-harmonic noise.
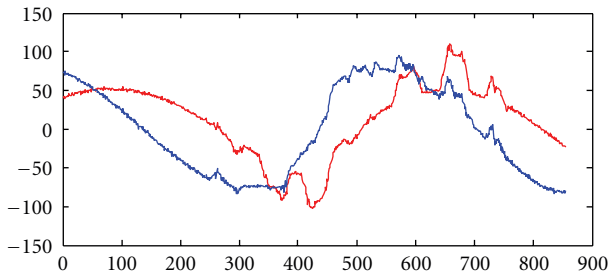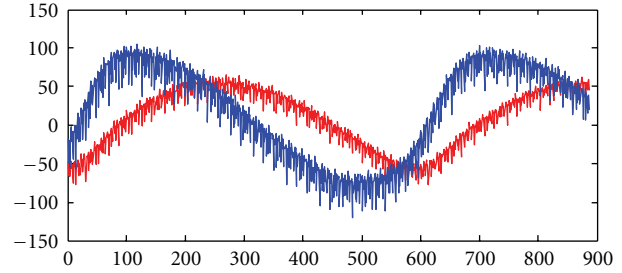

FIGURE 18: In the presence of motion in the scene, the above color-error measurements of the previous figure are disturbed.

in a part of the scene, leading to the high-harmonic noise distortion superimposed on the sinusoidal waveform.

The described color and intensity measurements subsystems and a part of the filtering were implemented as a subsystem in an Application Specific Integrated Circuit (ASIC). The corresponding block diagram of the logic is shown in Figure 23. The blocks in the diagram with inequalities are determining the selected pixel windows, where the filtering takes place. Further implementation details are presented in the appendix.

Removing Motion, Noise, and Other Artifacts from the Measurement Signals. One has the following.

(i) *Spectral Filtering.* Once measured, error signals should be filtered to remove all the spectral components that do not belong to the waveform model of fluorescent light. If we assume that the deviation of the mains frequency from its nominal value is ±1%, the filters have to remove all spectral components that do not belong to this range, such as the superimposed noise depicted in Figure 19.

(ii) *Motion-Effect and Light-Change Filtering.* Moreover, the remaining disturbances originating from motion or other light changes in the scene are filtered subsequently. For example, when light(s) in the scene are switched on and off, and/or when large-scale motion is present, the color-error measurement signals are considerably disturbed and are not reliable. If any of these conditions is detected, measurements are disregarded until the image stabilizes. In some cases, measurements are reset to accommodate scene

changes. All these consistency and reliability measures are implemented for both the intensity as well as for the color-error measurements.

8.2.2. The Fluorescent-Light Detection Algorithm. The detection algorithm presented in the following paragraphs employs the output of the intensity- and color-error measurements discussed previously. Let us now focus on the detection algorithm.

When intensity and color-error measurements are showing periodic (sinusoidal) behavior and as they are difference signals, the dominant mains frequency waveform is removed by the substraction. Then, the resulting waveform shows the deviations, that is, a frequency of 1% of the mains frequency. The intensity and color-error signals should also have a significant amplitude, to indicate the presence of florescent light sources in the scene. The detection strategy is thus to remove all the disturbances and perform the following three measurements: (1) *amplitude* of the error signals, (2) *frequency* of those signals, and (3) *temporal detection consistency* of the detected fluorescent light.

The amplitude of the error signals is measured by a robust envelope detector, whereas the frequency is obtained by calculating a period between two zero-crossing moments. Actually, error signals do have a DC value that depends on the mains frequency and on the scene contents. We have to estimate this DC value. Hence, points of the crossing through this DC level are used for sinusoidal period determination. Finally, we check whether the calculated amplitude is significant and the frequency of the color error is within 1% of the mains frequency for a certain amount of time. This duration measurement improves the consistency of the detection. Despite the structure of the measurements, our experimental results will reveal that sometimes complicated situations occur, which lead to special signal patterns. This will be discussed in the experimental results of the fluorescent light detector in Section 8.2.4.

In the special case that the fluorescent locking is used in combination with the detector discussed here (described in Section 8.3), the error measurement signals are likely to be constant over time and have nonzero values, which is then an indication of the presence of florescent light sources in the scene.
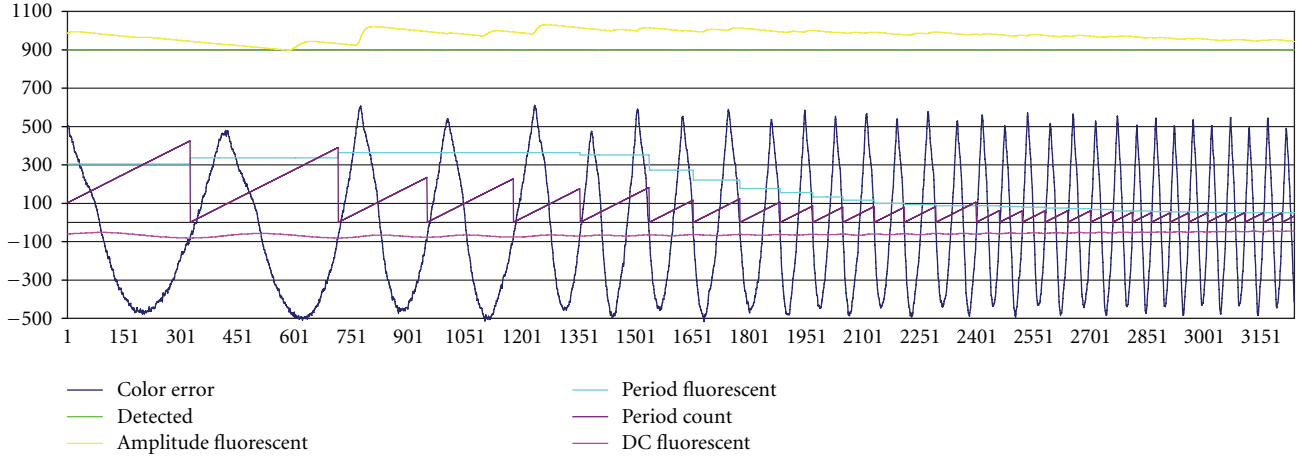
FIGURE 20: Color-error signal and derived signals used for detection of fluorescent light. This figure shows measurements in case of fluorescent light with increasing frequency over time.

Resuming with the normal case without fluorescent locking, if the above detection conditions are satisfied, our fluorescent light detector provides a positive detection and we can perform the procedure of shifting the short-exposed image parts from the output intensity range. By doing so, we will lose the benefit of having an improved SNR as achieved by the exposure bracketing. However, the intensity and color errors caused by the fluorescent light can be quite severe and the choice for shifting results in a more stable image quality. If we still would like to keep the benefit of exposure bracketing, we would have to ensure that the output of the short-exposure image has constant intensity and color. This can be achieved by means of a "fluorescence locking" procedure, which is proposed as a second option for handling fluorescent light in the image. The locking procedure is discussed in Section 8.3.

8.2.3. Shifting the Short-Exposure Image and Reducing the Color Saturation. To avoid fluorescent-light problems, we now detail the solution of shifting the short-exposure image out of the display range, so that only the long-exposure image remains effective. Once the fluorescent-light detection is performed, the following operations have to take place. Similar to the single exposure camera in the presence of fluorescent light, the long-exposure time T_L is made equal to a multiple of the fluorescence light period. The long- and short-exposed images are afterwards combined to a single output image I_O. Gain G is applied to the combined signal, so that the image parts constituted of short exposure are removed from the output range. Hence, image parts with input intensity smaller than I_T (see Figure 12(b)) are shifted to a clipping range, which effectively results in $I_O = RI_{Lc}$. This implies the use of a gain setting of, for example, $G = R = 4$ in Figure 12, since the ratio of exposure times shown there is $R = 4$. Consequently, the long-exposed image parts (having the integration time of e.g., 1/100 s or 1/120 s) will constitute the majority of the output signal. If some parts of the short-exposed image are left in the image, color reduction (fading) will be applied to them to remove false colors. At the same

time, the lens is closed in such a way that the same average light output is achieved prior to shifting the short-exposure image.

8.2.4. Experimental Results of the Fluorescent Light Detector. We present three examples of the performance of our fluorescence detection algorithm (Figures 20–22). The figures show color- (intensity-) error measurements and several derived variables which are used in the algorithm. The signal *Amplitude Fluorescent* is an estimate of the amplitude of the *Color-Error* signal. We derive the signal amplitude by detecting a robust positive and negative envelope of the color-error signal. To detect frequency shifts of the fluorescent light with respect to the mains frequency, we measure the time between subsequent crossings of the *Color-Error* signal through a DC value of the color-error signal *DC Fluorescent*. The signal *DC Fluorescent* is calculated as a robust, long-term average value of the color-error signal. The value *Period Count* is a counter that measures oscillation time (period) of the color-error signal and is used to derive a period of the fluorescent light *Period Fluorescent*. Using the previous measurements, we can decide whether fluorescent light is present in the scene, indicated by the signal *Detected*.

In Figure 20, we show the detector response to a scene containing a fluorescent light source with increasing frequency over time. In Figure 21, we present the detector response to florescent light that is switched off and on again. In Figure 22, we show a more complex scene which includes both motion, changing frequency of fluorescent light and switching other light sources on and off. In all these cases, the detection is correctly performed.

These experiments reveal in a clear way that the primary design challenge for a fluorescent light detector is to provide a sufficiently high detection robustness. This is a difficult task for which we had to build various mechanisms to stabilize each of the discussed signals and measurements, in order to preserve a correct detection. The combination of large-scale motion, sudden light changes, and multiple light
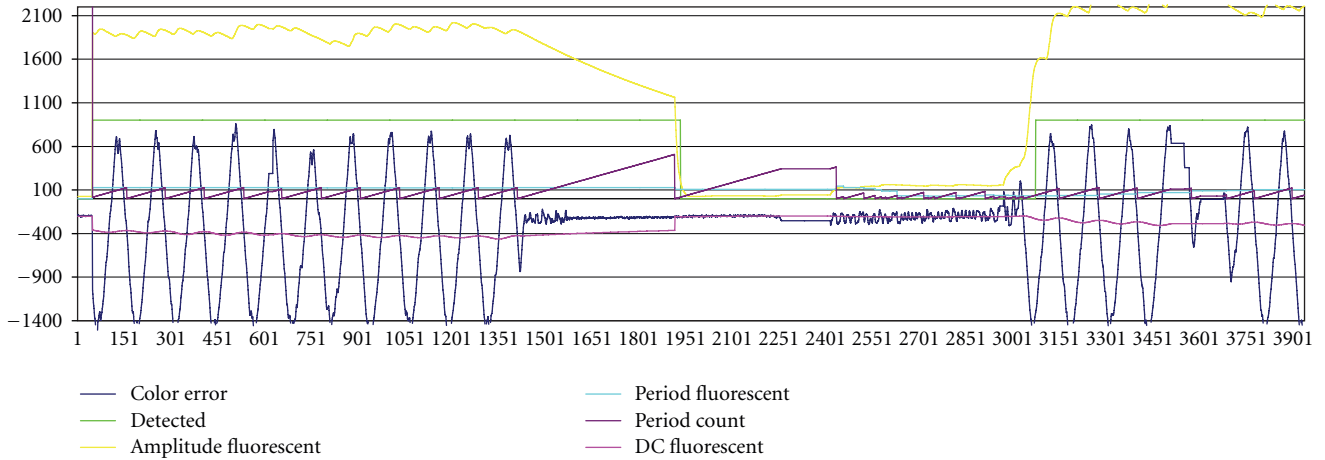
FIGURE 21: Color-error signal and derived signals used for detection of fluorescent light. This figure shows measurements in case of switching fluorescent light on and off.
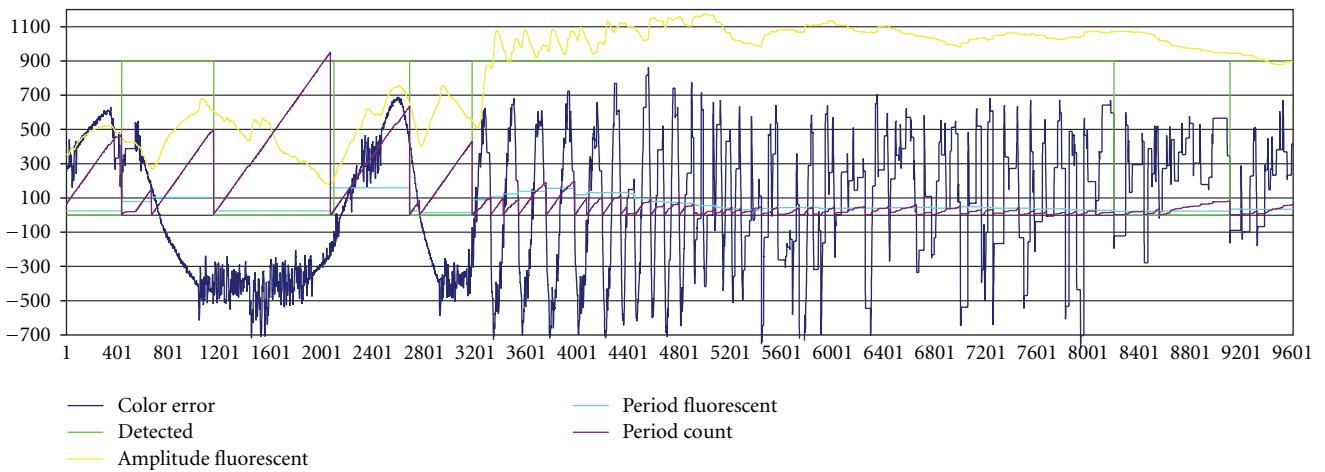

FIGURE 22: Color-error signal and derived signals used for detection of fluorescent light. This figure shows measurements in case when motion occurs, the fluorescent light has changing frequency, and other light sources are switched on and off, at the same time.

sources with different behavior and phases poses significant challenges for the detection algorithm.

The experimental results presented here show that a significant progress has been made in an area that is not reported in scientific literature. This makes our contribution already highly valuable, although we recognize that further research needs to be performed to clarify the interdependencies between the parameters and establish additional robustness improvements.

8.3. Algorithm 2: Fluorescence Locking. Fluorescence locking is a procedure that has the objective to synchronize the exposure measurement with the mains frequency, such that the moment at which the short-exposure integration is performed is positioned at the optimal moment within the fluorescent light period. This optimal moment occurs at the peak (maximum) of the light output (Period B in **Figure 14**). Therefore, we ensure that light integrated during the short-exposure time is constant over time and has a correct color,

which is then not influenced by the variable fluorescence light output and on/off switching effects of the fluorescent tubes. To achieve this, the intensity and color errors between the long- and short-exposure images are observed and used as a control signal to drive a Phase-Locked Loop (PLL), such that it assures the correct phase (read-out moment) of the short-exposure time with respect to the fluorescent lighting. In case that the correct read-out moment is selected for the short-exposure time, color errors either are constant or do not exist, so that an oscillatory (periodic) behavior is absent.

The input for the PLL can be, for example, one or more of the intensity- and color-error signals, or some combination of them. This proposal using a PLL compensates not only for the phase difference between the optimal read moment and the current read moment of the *Short Exposure Image* but also for the frequency difference of the actual and ideal mains lock frequencies. Namely, due to a frequency drift of the mains signal (usually up to 1% of a nominal value), integrated light in the short-exposure image changes the color temperature
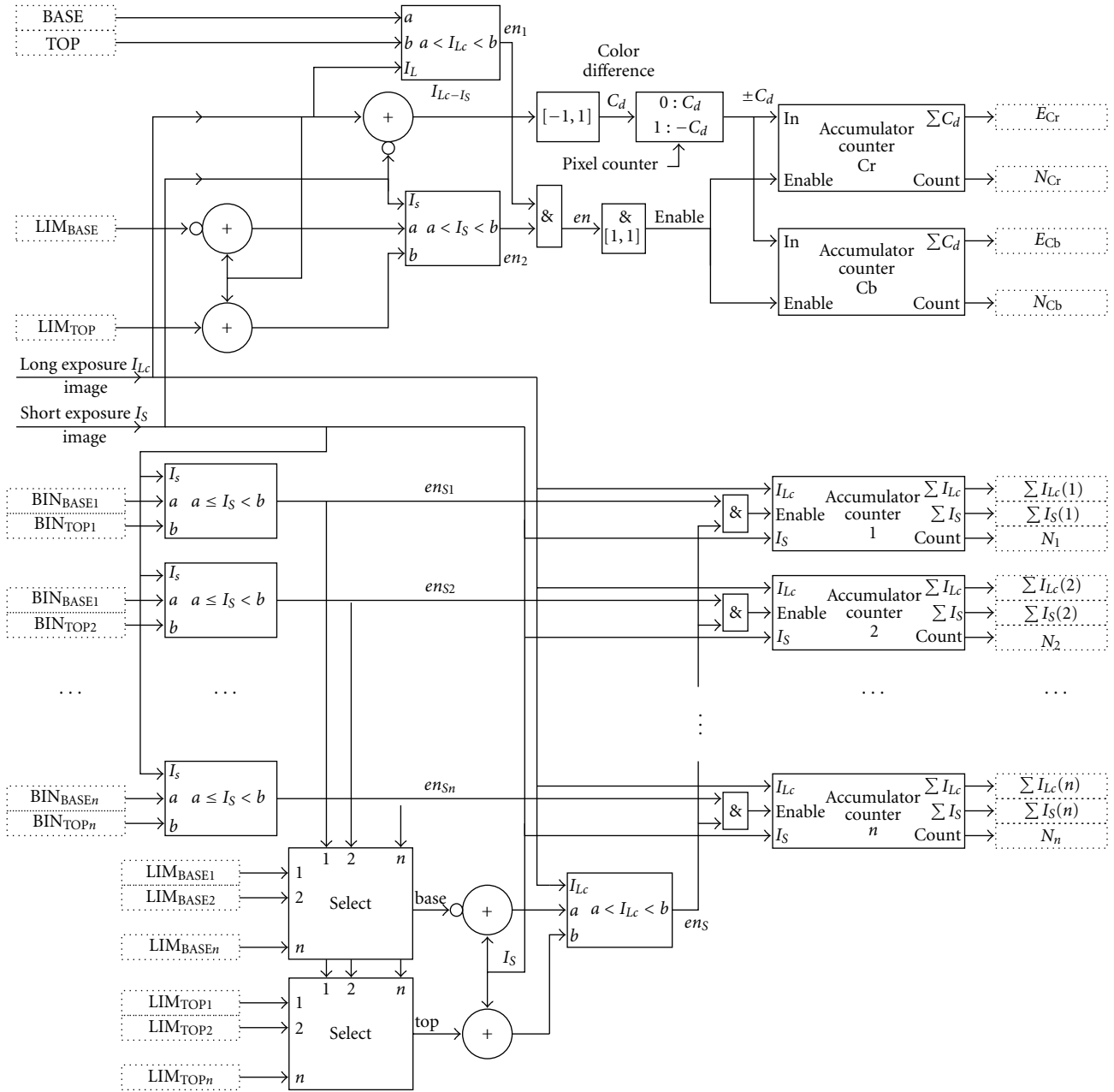
Figure 23: The fluorescent light detection block.

and dominant color content over time, which is also prevented in the above proposal. The fluorescent locking control is achieved by realizing two important aspects:

(i) changing the camera picture operating frequency, so that it runs on the same current mains frequency, which is used for driving the fluorescent light sources;

(ii) adjusting the camera phase, such that the short-exposure time is positioned at the peak (maximum) of the fluorescent light output (Period B in Figure 14).

When multiphase fluorescent light is present in the scene, the camera will lock to the phase that gives the largest output signal. Usually, all mains signal phases are running synchronously with each other, and if the camera locks to one of them, their mutual synchronization will be maintained. This means that light sources having a phase other than the one the camera is locked to will have a constant phase relationship and will be giving a constant light output and associated color.

Preliminary results of the phase locking procedure are promising and we can achieve a constant intensity and color output. However, this is recent, ongoing work of which the results are still emerging. For example, tuning of the PLL loop and its correct operation under all circumstances is a complicated matter and still under investigation. The biggest

challenge lies in removing all the interferences to the measurement errors, as, for instance, presented in Figures 18 and 19. A considerable advantage of phase locking is that we still enjoy the benefits of an increased SNR due to the exposure bracketing technique, which enables us to perform optimal tone mapping and image enhancement techniques. An alternative technique to image-based fluorescent light detection and locking can be based on a light-metering diode, which has an added value of being much less sensitive to motion objects in the scene. In such a case, the same procedure of detection and locking can be performed as previously described. However, the disadvantage of this approach is that it requires an additional sensor, the measuring photo diode.

9. Conclusions and an Outlook

In this paper, we have presented fundamental functionality for video cameras in the form of various forms of exposure (level) control. Camera level control is very important since it provides a basis for all the subsequent image processing algorithms and is a prerequisite for a good image quality. Moreover, we claim that a good level control is at least as important as all the other subsequent stages of the image processing chain. We have given a comprehensive overview of the complete level-control processing chain for video cameras. The overview involves metering techniques, types of measurements used for exposure control, control methods, and strategies, and we wrappedup the level control with special control operation modes.

With respect to control strategy, we prefer parallel control because it offers improved speed in adapting to signal changes and stability of operation. We have described special signal processing techniques that are necessary for correct rendering of both low- and high-dynamic range scenes. The special control modes are preferably based on both peak-average-based control and saturation control. Besides this, back-light compensation strategies are required to produce good visibility of foreground objects for low-dynamic range processing pipelines. These techniques are especially needed in low-to-medium dynamic range pipelines and sensors to ensure good visibility of dark image parts. However, in these cases, we have to sacrifice image fidelity in bright image parts.

Furthermore, we discussed various sensor dynamic-range extension techniques and have chosen the exposure bracketing for this task. In the remainder of the paper, we have discussed some solutions for the problem of moving scene objects and nonconstant light sources such as fluorescent light, that introduce false colors and light flickering. In particular, two methods were proposed for fluorescent-light handling: automatic fluorescent light detection and fluorescence locking. Our experiments showed that it is possible to design such a detector as well as control mechanisms based on PLL principles. However, the robustness of such a system is difficult to achieve when various interferences occur simultaneously.

Due to its inherent complexity and various problems, exposure bracketing is a very challenging candidate for creating HDR images, but on the other hand, it offers a significant extension of the sensor dynamic range, which becomes about 100 dB. High-dynamic range sensors whose output-to-input conversion is based on logarithmic functions are also not a good alternative for high-fidelity imaging, since they introduce color distortions and color shifts. A good alternative that offers an output dynamic range of about 90 dB can be a technique where two types of sensor pixels are used: one with high and the other with low sensitivity. High-sensitivity pixel output mimics long-integration time output and low-sensitivity output corresponds to short-exposure time. This method has an advantage that integration time of both types of pixels is the same and can be set equal to a fluorescent light period to solve the fluorescent light problem. However, this approach has lower sensitivity due to smaller pixel sizes. In a specific sensor implementations, instead of having two sets of pixels, only one set can be used with two different conversion settings, which maintains the image resolution and sensitivity [33]. This approach works well also in the presence of moderate motion, and only when fast motion occurs in the scene, we may have to lower the integration time and lose the benefit of an increased SNR. An alternative is the use of a so-called "flutter shutter" camera, where the shutter of the camera lens is opened and closed during the field/frame time, with a binary pseudorandom sequence [34]. This method enables the recovery of high-frequency spatial details for the constant speed objects. However, this approach would often imply employment of a special, more expensive lens, which might not be acceptable for the user due to its size and cost.

Finally, the presented level control algorithms have to cooperate with the subsequent tone mapping and image enhancement processing. This gives a new paradigm and opens possibilities for further exploration to achieve better camera performance. This new class of high imagefidelity algorithms integrates "conventional concepts of exposure control" that account for the functions of iris control, sensor integration time, and gain control with signal processing tasks such as tone mapping, image enhancement, and object detection.

Appendix

Intensity and Color-Error Detector for Artificial Light Sources

This appendix gives a detailed description of the intensity and color-error measurements used for the detection of artificial light sources. Both measurement types are used in parallel, as we want to avoid any annoying artifact and increase the correctness of the detection.

Color-Error Measurement. Figure 23 presents a measurement block performing the previously described fluorescent light detection. The differences in intensity and color between the long- and the short-exposed pixels depend on the exposure times and phase relation between the exposure moments and the mains frequency. Accumulators/counters Cr and Cb calculate the differences in color between the short- and the long-exposure pixels (the accumulator Cr

measures error of Cr color and the accumulator Cb measures the error of Cb color). To calculate the color differences C_d, we implement a differentiator (FIR filter with coefficients 1 and -1), whereas a pixel-alternating-sign multiplier is used to always take the same sign of the difference. For example, for CMYG type of sensor, the color difference C_d is always equal to "Cyan-Yellow" in a Cr line and "Green-Magenta" in a Cb line, whereas in the RGB Bayer sensor, color difference C_d is always equal to "Red-Green" in a Cr line and "Green-Blue" in a Cb line. For spatial consistency, it is required that neighboring pixel also satisfies detection conditions, which is why logical and operation is performed on two neighboring pixels. The color differences are measured in a range that can be set with BASE and TOP registers; hence, if the *Long-Exposure Image* pixel I_{Lc} is between BASE and TOP values, signal en_1 is active. To exclude large color differences that potentially come from moving objects, we only allow reasonably small differences, which are between $-\text{LIM}_{\text{BASE}}$ and $+\text{LIM}_{\text{TOP}}$, checked by a signal en_2. If both signals en_1 and en_2 are active, then both pixels are accumulated and the counter is incremented. The accumulator/counter values are copied to the registers at the end of the field/frame.

Intensity Error Measurement. The lower part of the detector block measures accumulated intensity of the normalized *long-Exposure Image* I_{Lc} and *Short-Exposure Image* I_S in n programmable bin ranges, from which differences between I_{Lc} and I_S can be calculated. One such range is selected by $\text{BIN}_{\text{TOP}j}$ and $\text{BIN}_{\text{BASE}j}$ registers ($j = 1,\dots,n$). We also use $\text{LIM}_{\text{BASE}j}$ and $\text{LIM}_{\text{TOP}j}$ registers to remove extremes that could spoil the measurement. Such extremes can, for instance, occur in the presence of motion and/or light changes in the scene. Using $\text{LIM}_{\text{BASE}j}$ and $\text{LIM}_{\text{TOP}j}$ as well as LIM_{BASE} and LIM_{TOP} registers, we will exclude the majority of these disturbances from the image: if long- and short-exposure signals are very different from each other, we assume that these differences originate from disturbances and not from fluorescent light. Likewise, we will disable the measurement of these image parts, by setting the signals en_s, en_1 and en_2 to zero. Looking at Figure 23, the enable signal en_{sj} is equal to unity for the jth signal $j = 1,\dots,n$ when $\text{BIN}_{\text{BASE}j} < I_S < BIN_{\text{TOP}j}$. If the signal is $en_{sj} = 1$, then the selector switches to base $= \text{LIM}_{\text{BASE}j}$ and top $= \text{LIM}_{\text{TOP}j}$. When a *Short-Exposure Image* pixel value falls into a bin, a second test is done where the normalized *Long-Exposure Image* pixel value should fall in a range about the *Short-Exposure Image* pixel value. If both tests are $en_s = 1$ and $en_{sj} = 1$, then accumulator j is enabled and it accumulates long and short intensity signals and counts the number of pixel occurrences within the intensity range j.

References

[1] S. Battiato, G. Messina, and A. Castorina, "Exposure correction for imaging devices: an overview," in *Single-Sensor Imaging: Methods and Applications for Digital Cameras*, chapter 1, pp. 323–349, CRC Press/Taylor & Francis, Boca Raton, Fla, USA, 2008.

[2] M. Reichmann, "The luminous landscape," 2010, http://www.luminous-landscape.com/.

[3] W. C. Kao, C. C. Hsu, C. C. Kao, and S. H. Chen, "Adaptive exposure control and real-time image fusion for surveillance systems," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '06)*, pp. 935–938, Island of Kos, Greece, May 2006.

[4] M. Skow and H. Tran, "Automatic exposure control system for a digital camera," Unated States Patent Application 7173663, 2002.

[5] S. Johnson, S.-C. Chao, N. R. Itani et al., "Image processor circuits, systems and methods," US patent application 20020176009, 2002.

[6] K. J. Astrom and T. Hagglund, *PID Controllers: Theory, Design, and Tuning*, International Society for Measurement and Control, Seattle, Wash, USA, 1995.

[7] W. R. Evans, "Control systems synthesis by root locus method," *Transactions of the American Institute of Electrical Engineers*, vol. 69, pp. 66–69, 1950.

[8] J.-J. E. Slotine and W. Li, *Applied Nonlinear Control*, Prentice Hall, Upper Saddle River, NJ, USA, 1991.

[9] T. Kuno, H. Sugiura, and N. Matoba, "A new automatic exposure system for digital still cameras," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 1, pp. 192–199, 1998.

[10] O. H. Bosgra, H. Kwakernaak, and G. Meinsma, "Design methods for control systems," in *Notes for the Course of the Dutch Institute for System and Control*, pp. 397–405, University of Twente, Enschede, The Netherlands, 2000.

[11] Y. Haitao, C. Yilin, and W. Jing, "A new automatic exposure algorithm for video cameras using luminance histogram," *Frontiers of Optoelectronics in China*, vol. 1, no. 3-4, pp. 285–291, 2008.

[12] S. Shimizu, T. Kondo, T. Kohashi, M. Tsuruta, and T. Komuro, "A new algorithm for exposure control based on fuzzy logic for video cameras," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 3, pp. 617–623, 1992.

[13] J. S. Lee, Y. Y. Jung, B. S. Kim, and S. J. Ko, "An advanced video camera system with robust AF, AE, and AWB control," *IEEE Transactions on Consumer Electronics*, vol. 47, no. 3, pp. 694–699, 2001.

[14] S. Battiato, A. Bosco, A. Castorina, and G. Messina, "Automatic image enhancement by content dependent exposure correction," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 12, pp. 1849–1860, 2004.

[15] S. Cvetkovic, P. Bakker, J. Schirris, and P. H. N. de With, "Background estimation and adaptation model with light-change removal for heavily down-sampled video surveillance signals," in *IEEE International Conference on Image Processing*, pp. 1829–1832, Atlanta, Ga, USA, 2006.

[16] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, pp. 246–252, June 1999.

[17] M. Koyanagi and T. Tomitaka, "Video camera system with exposure control," European patent application EP0975152, 1995.

[18] G. Finlayson and S. Hordley, "Color signal processing," US patent application 7227586, 2007.

[19] T. Haruki and K. Kikuchi, "Video camera system using fuzzy logic," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 3, pp. 624–634, 1992.

[20] A. Morimura, K. Uomori, Y. Kitamura et al., "A digital video camera system," *IEEE Transactions on Consumer Electronics*, vol. 36, no. 4, pp. 866–876, 1990.

[21] M. Murakami and N. Honda, "Exposure control system of video cameras based on fuzzy logic using color information,"

in *Proceedings of the 5th IEEE International Conference on Fuzzy Systems*, vol. 3, pp. 2181–2187, New Orleans, La, USA, 1996.

[22] S. Sakaue, A. Tamura, M. Nakayama, and S. Maruno, "Adaptive gamma processing of the video cameras for the expansion of the dynamic range," *IEEE Transactions on Consumer Electronics*, vol. 41, no. 3, pp. 555–562, 1995.

[23] S. Cvetković, J. Klijn, and P. H. N. de With, "Tone-mapping functions and multiple-exposure techniques for high dynamic-range images," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 904–911, 2008.

[24] S. Cvetković, J. Schirris, and P. H. N. de With, "Non-linear locally-adaptive video contrast enhancement algorithm without artifacts," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 1, pp. 1–9, 2008.

[25] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*, Morgan Kaufmann Publishers, San Francisco, Calif, USA, 2005.

[26] L. Meylan, *Tone mapping for high dynamic range images*, Ph.D. thesis, EPFL, 2006.

[27] G. de Haan, *Video Processing for Multimedia Systems*, University Press Facilities, Endhoven, The Netherlands, 2000.

[28] A. Darmont, "Methods to extend the dynamic range of snapshot active pixels sensors," in *Sensors, Cameras, and Systems for Industrial/Scientific Applications IX*, Proceedings of SPIE, San Jose, Calif, USA, January 2008.

[29] L. E. Alston, D. S. Levinstone, and W. T. Plummer, "Exposure control system for an electronic imaging camera having increased dynamic range," US patent 4647975, 1987.

[30] W. C. Kao, "High dynamic range imaging by fusing multiple raw images and tone reproduction," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 1, pp. 10–15, 2008.

[31] P. Lasang, C. P. Ong, and S. M. Shen, "Cfa-based motion blur removal using long/short exposure pairs," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 332–338, 2010.

[32] S. Cvetkovic, P. Sturm, J. Schirris, and J. Klijn, "Fluorescent artifact reduction for double exposure cameras," European patent application registration no. 2009/5805, r. 330995, 2009.

[33] B. Fowler, C. Liu, S. Mims et al., "A 5.5Mpixel 100 frames/sec wide dynamic range low noise CMOS image sensor for scientific applications," in *Sensors, Cameras, and Systems for Industrial/Scientific Applications XI*, vol. 7536 of *Proceedings of SPIE*, San Jose, Calif, USA, January 2010.

[34] R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," in *Proceedings of the 33rd International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH '04)*, pp. 795–804, Boston, Mass, USA, August 2006.

Photograph © Turisme de Barcelona / J. Trullàs

EUSIPCO 2011

19th European Signal Processing Conference
August 29- September 2, 2011, Barcelona (Spain)

Preliminary call for papers

The 2011 European Signal Processing Conference (EUSIPCO-2011) is the nineteenth in a series of conferences promoted by the European Association for Signal Processing (EURASIP, www.eurasip.org). This year edition will take place in Barcelona, capital city of Catalonia (Spain), and will be jointly organized by the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) and the Universitat Politècnica de Catalunya (UPC).

EUSIPCO-2011 will focus on key aspects of signal processing theory and applications as listed below. Acceptance of submissions will be based on quality, relevance and originality. Accepted papers will be published in the EUSIPCO proceedings and presented during the conference. Paper submissions, proposals for tutorials and proposals for special sessions are invited in, but not limited to, the following areas of interest.

Areas of Interest

- Audio and electro-acoustics.
- Design, implementation, and applications of signal processing systems.
- Multimedia signal processing and coding.
- Image and multidimensional signal processing.
- Signal detection and estimation.
- Sensor array and multi-channel signal processing.
- Sensor fusion in networked systems.
- Signal processing for communications.
- Medical imaging and image analysis.
- Non-stationary, non-linear and non-Gaussian signal processing.

Submissions

Procedures to submit a paper and proposals for special sessions and tutorials will be detailed at www.eusipco2011.org. Submitted papers must be camera-ready, no more than 5 pages long, and conforming to the standard specified on the EUSIPCO 2011 web site. First authors who are registered students can participate in the best student paper competition.

Important Deadlines:

Proposals for special sessions	15 Dec 2010
Proposals for tutorials	18 Feb 2011
Electronic submission of full papers	**21 Feb 2011**
Notification of acceptance	23 May 2011
Submission of camera-ready papers	6 Jun 2011

Webpage: www.eusipco2011.org

Organizing Committee

Honorary Chair
Miguel A. Lagunas (CTTC)

General Chair
Ana I. Pérez-Neira (UPC)

General Vice-Chair
Carles Antón-Haro (CTTC)

Technical Program Chair
Xavier Mestre (CTTC)

Technical Program Co-Chairs
Javier Hernando (UPC)
Montserrat Pardàs (UPC)

Plenary Talks
Ferran Marqués (UPC)
Yonina Eldar (Technion)

Special Sessions
Ignacio Santamaría (Unversidad de Cantabria)
Mats Bengtsson (KTH)

Finances
Montserrat Nájar (UPC)

Tutorials
Daniel P. Palomar
(Hong Kong UST)
Beatrice Pesquet-Popescu (ENST)

Publicity
Stephan Pfletschinger (CTTC)
Mònica Navarro (CTTC)

Publications
Antonio Pascual (UPC)
Carles Fernández (CTTC)

Industrial Liaison & Exhibits
Angeliki Alexiou
(University of Piraeus)
Albert Sitjà (CTTC)

International Liaison
Ju Liu (Shandong University-China)
Jinhong Yuan (UNSW-Australia)
Tamas Sziranyi (SZTAKI -Hungary)
Rich Stern (CMU-USA)
Ricardo L. de Queiroz (UNB-Brazil)