SERVER APPARATUS, METHOD, AND SYSTEM FOR PREDICTING OPTIMAL EXPOSURE TIME OF IMAGES

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This patent application claims priority to and the benefit of Singapore Patent Application No. 10202400568Y, filed on 29 February 2024 in the Intellectual Property Office of Singapore, the disclosure of which is incorporated by reference herein in its entirety.

TECHNICAL FIELD

[0002] Various aspects of this disclosure relate to a method, server apparatus, and system for predicting optimal exposure time of images.

BACKGROUND

[0003] The following discussion of the background art is intended to facilitate an understanding of the present disclosure only. It should be appreciated that the discussion is not an acknowledgement or admission that any of the material referred to was published, known or is part of the common general knowledge of the person skilled in the art in any jurisdiction as of the priority date of the disclosure.

[0004] Dynamic range is defined as the ratio of a maximum value to a minimum value that a certain quantity can assume. The dynamic range of a real-world high dynamic range (HDR) image can reach 100,000,000:1 whereas the dynamic range of an 8-bit low dynamic range (LDR) image is 256:1. Our human eyes can detect dynamic ranges varying from 10,000:1 to 1,000,000:1. HDR imaging is thus highly demanded by many real-world applications. Merging differently exposed LDR images to expand the dynamic range is an effective way for the HDR imaging. The differently exposed images are usually captured sequentially. Unfortunately, ghosting artifacts caused by moving objects are an issue for such a capturing method.

[0005] Recent HDR image capturing devices are emerging to capture HDR images simultaneously so as to avoid ghosting artifacts from appearing in a synthesized image of an HDR scene. Among them, an adaptive row-wise autoexposure capturing method was proposed in the article titled "Coded rolling shutter photography: flexible space-time sampling" by Jinwei Gu, Yasunobu Hitomi, Tomoo Mitsunaga, and Shree Nayar. The exposure is changed for every row by using a heuristic method. A temporary image is first captured for an HDR scene. An empirical equation between the scene radiance and the temporary image is then built

up. An optimal exposure is finally derived for every row on top of the empirical equation to minimize the number of over- and under-exposed pixels while keep most pixels well-exposed in the row. However, the empirical equation may not adequately represent the nonlinear relationship between the captured image and the irradiance.

[0006] Therefore, there exists a need for improved image processing for obtaining HDR images.

SUMMARY

[0007] The present disclosure relates to an artificial intelligence (AI) based high dynamic range (HDR) image processing, and more specifically, a neural spatially-varying auto-exposure method, for generating optimal exposure times for non-overlapping blocks of an input image. The exposure time of each non-overlapping block may then be adjusted according to the generated optimal exposure time to produce a resultant HDR image that captures information in brightest and darkest regions of the first image.

[0008] In some embodiments, artificial intelligence algorithms, such as, but not limited to, neural networks, which may further include convolutional neural networks (CNN), may be used to represent nonlinear functions. It is thus highly demanded to develop a neural network based learnable spatially-varying auto-exposure algorithm for HDR imaging.

[0009] According to an aspect of the present disclosure there is provided a method of predicting optimal exposure times for all non-overlapping blocks of an image, comprising: training an artificial intelligence module on a dataset comprising temporary images of a first scene and corresponding sets of ground-truth optimal exposure times for all pixels of each temporary image of the first scene; obtaining a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene; inputting the temporary image of the second scene and its initial exposure time into the trained artificial intelligence module; generating, by the artificial intelligence module, a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.

[0010] In some embodiments, the artificial intelligence module comprises a convolutional neural network (CNN).

[0011] In some embodiments, the method further comprises defining a loss function based on a comparison of a predicted exposure correction with a ground truth exposure correction at different spatial locations.

[0012] In some embodiments, the method further comprises generating an exposure adjustment map based on the set of predicted optimal exposure times, wherein the exposure adjustment map represents per-pixel exposure adjustments of the temporary image of the second scene.

[0013] In some embodiments, the exposure adjustment map comprises a greyscale heatmap.

[0014] In some embodiments, the method further comprises reconstructing a spatially-adapted HDR image based on the exposure adjustment map applied on the temporary image of the second scene.

[0015] In some embodiments, the CNN has a U-Net architecture, and a shallow residual U-Net configuration.

[0016] In some embodiments, the U-Net architecture comprises an encoder-decoder structure to process exposure variations at different scales.

[0017] In some embodiments, the encoder of the U-Net architecture comprises a first 3x3 convolution kernel, a parametric ReLU (PReLU) activation module, and a second 3x3 convolution kernel.

[0018] In some embodiments, the method further comprises defining a 16 by 16 non-overlapping block or a 32 by 32 non-overlapping block.

[0019] According to another aspect of the present disclosure there is provided a non-transitory computer-readable storage medium storing instructions that, when executed by a processor, cause the processor to perform the aforementioned method.

[0020] According to another aspect of the present disclosure there is provided a server apparatus of predicting optimal exposure times for non-overlapping blocks of an image, the server apparatus comprises a processor, the processor configured to: train an artificial intelligence module on a dataset comprising temporary images of a first scene and corresponding sets of optimal exposure times for all pixels of each temporary image of the first scene; obtain a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene; input the temporary image of the second scene and its initial exposure time into the trained artificial intelligence module; and generate, by the artificial intelligence module, a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.

[0021] In some embodiments, the artificial intelligence module comprises a convolutional neural network (CNN).

[0022] In some embodiments, the processor is configured to define a loss function, wherein the loss function based on a comparison of a predicted exposure correction with a ground truth exposure correction at different spatial locations.

[0023] In some embodiments, the processor is configured to generate an exposure adjustment map based on the set of predicted optimal exposure times, and wherein the exposure adjustment map represents per-pixel exposure adjustments of the temporary image of the second scene.

[0024] In some embodiments, the exposure adjustment map comprises a greyscale heatmap.

[0025] In some embodiments, the processor is configured to reconstruct a spatially-adapted HDR image based on the exposure adjustment map.

[0026] In some embodiments, the CNN has a U-Net architecture, and a shallow residual U-Net configuration.

[0027] In some embodiments, the U-Net architecture comprises an encoder-decoder structure to process exposure variations at different scales.

[0028] In some embodiments, the encoder of the U-Net architecture comprises a first 3x3 convolution kernel, a parametric ReLU (PReLU) activation module, and a second 3x3 convolution kernel.

[0029] In some embodiments, the server apparatus is configured to define a 16 by 16 non-overlapping block or a 32 by 32 non-overlapping block.

[0030] According to another aspect of the present disclosure there is provided a CMOS image sensor configured to capture a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene; input the temporary image of the second scene and its initial exposure time into the trained artificial intelligence module of the aforementioned server apparatus; and generate, by the artificial intelligence module, a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] The disclosure will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

- FIG. 1A is a schematic block diagram of a server apparatus for predicting optimal exposure times for one or more non-overlapping blocks of an image.

- FIG. 1B is a schematic block diagram comprising the server apparatus of FIG. 1A, comprising various modules.
- FIG. 2 is a flowchart depicting a method of predicting optimal exposure times for one or more non-overlapping blocks of an image.
- FIG. 3 is a proposed framework of a convolutional neural network (CNN) based learnable block-wise auto-exposure for predicting optimal exposure times for one or more non-overlapping blocks of an image. The framework is especially suited for implementation with CMOS sensors.
- FIG. 4 is an encoder-decoder arrangement for facilitating the method of predicting optimal exposure times for one or more non-overlapping blocks of an image.

DETAILED DESCRIPTION

[0032] The following detailed description refers to the accompanying drawings that show, by way of illustration, specific details, and embodiments in which the disclosure may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the disclosure. Other embodiments may be utilized, and structural and logical changes may be made without departing from the scope of the disclosure. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

[0033] Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments. Features that are described in the context of an embodiment may correspondingly be applicable to the other embodiments, even if not explicitly described in these other embodiments. Furthermore, additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0034] In the context of various embodiments, the articles "a", "an" and "the" as used with regard to a feature or element include a reference to one or more of the features or elements. As used herein, the term "and/or" includes any and all combinations of one or more of the associated listed items.

[0035] While such terms as "first", "second", etc., may be used to describe various elements, such elements must not be limited to the above terms. The above terms are used only to distinguish one element from another, and do not define corresponding elements, for

example, an order and/or significance of the elements. Without departing from the scope of rights of the specification, a first element may be referred to as a second element, and similarly, the second element may be referred to as the first element.

[0036] As used herein, the term "data" may be understood to include information in any suitable analog or digital form, for example, provided as a file, a portion of a file, a set of files, a signal or stream, a portion of a signal or stream, a set of signals or streams, and the like. The term data, however, is not limited to the aforementioned examples and may take various forms and represent any information as understood in the art.

[0037] As used herein, the term "processor" refers to a circuit, including analog circuits, digital circuits, or hybrid circuits, or their constituent components. Any other kind of implementation of the respective functions which will be described in more detail below may also be understood as a "circuit" in accordance with an alternative embodiment. A digital circuit may be understood as any kind of a logic implementing entity, which may be special purpose circuitry or a processor executing software stored in a memory, or a firmware.

[0038] As used herein, the term "module" refers to, forms part of, or includes an application specific integrated circuit (ASIC); an electronic circuit; a combinational logic circuit; a field programmable gate array (FPGA); a processor (shared, dedicated, or group) that executes code; other suitable hardware components that provide the described functionality; or a combination of some or all of the above, such as in a system-on-chip. The term module may include memory (shared, dedicated, or group) that stores code executed by the processor. A single module or a combination of modules may be regarded as a device. A processor may include one or more modules. For example, multiple modules described in this disclosure may form a processor.

[0039] As used herein, the term "associate", "associated", and "associating" indicate a defined relationship (or cross-reference) between two items.

[0040] As used herein, "memory" may be understood as a non-transitory computer-readable medium in which data or information can be stored for retrieval. References to "memory" included herein may thus be understood as referring to volatile or non-volatile memory, including random access memory ("RAM"), read-only memory ("ROM"), flash memory, solid-state storage, magnetic tape, hard disk drive, optical drive, etc., or any combination thereof. Furthermore, it is appreciated that registers, shift registers, processor registers, data buffers, etc., are also embraced herein by the term memory. It is appreciated that a single component referred to as "memory" or "a memory" may be composed of more than one different type of memory, and thus may refer to a collective component including one or more

types of memory. It is readily understood that any single memory component may be separated into multiple collectively equivalent memory components, and vice versa. Furthermore, while memory may be depicted as separate from one or more other components (such as in the drawings), it is understood that memory may be integrated within another component, such as on a common integrated chip.

[0041] As used herein, the term "configured to" broadly refers to the design, arrangement, or adaptation of a system, device, component, or module to perform a specific function or achieve a particular outcome. The term includes both hardware and software implementations wherein in a hardware implementation, the physical components are arranged, programmed, or structured to carry out the intended function(s), and in the context of programming and software, a device is operable under executable instructions (e.g., software, firmware) to perform the specified function(s) when executed by one or more processors. The resultant configuration allows the system or component to perform the stated function, either inherently or after suitable programming or activation, without requiring substantial modifications to its structure or operational logic.

[0042] According to various embodiments, a circuit may include analog circuits or components, digital circuits or components, or hybrid circuits or components. Any other kind of implementation of the respective functions which will be described in more detail below may also be understood as a "circuit" in accordance with an alternative embodiment. A digital circuit may be understood as any kind of a logic implementing entity, which may be special purpose circuitry or a processor executing software stored in a memory, firmware, or any combination thereof. Thus, in various embodiments, a "circuit" may be a digital circuit, e.g. a hard-wired logic circuit or a programmable logic circuit such as a programmable processor, e.g. a microprocessor (e.g. a Complex Instruction Set Computer (CISC) processor or a Reduced Instruction Set Computer (RISC) processor). A "circuit" may also include a processor executing software, e.g. any kind of computer program, e.g. a computer program using a virtual machine code such as e.g. Java.

[0043] As used herein, the term "dynamic range" refers to the ratio between the maximum and minimum tonal values that can be captured in a single image. In image capturing, the dynamic range may be measured in exposure value (EV) differences or stops. It represents the range of luminosity that can be accurately represented in an image. High dynamic range imaging refers to an imaging technique that captures and combines multiple exposures of the same scene to create an image with a greater dynamic range than what is possible with a single

exposure. HDR imaging aims to produce images that more closely resemble what the human eye perceives, preserving details in both bright and dark areas of a scene.

[0044] As used herein, the term "non-overlapping blocks" refers to distinct, contiguous regions of an image that do not share any common pixels. When dividing an image into non-overlapping blocks, each pixel of the image belongs to exactly one block.

[0045] As used herein, the term "convolutional neural networks or CNN" may refer to a class of artificial neural networks commonly used in image processing and computer vision tasks. These networks apply convolutional filters to input data, enabling them to learn and extract hierarchical features from images. In some embodiments, the CNN may include a U-net, the U-net comprising two main paths: a contracting (encoder) path and an expanding (decoder) path. The encoder path may be configured as a feature extractor, and the decoder path may be configured to precisely localize the features learned by the encoder.

[0046] According to an aspect of the present disclosure there is provided a server apparatus for predicting optimal exposure times for one or more non-overlapping blocks of an image. The server apparatus may be, or may form part of a processor or circuit of an image capturing device, or an image sensor, such as a complementary metal oxide semiconductor (CMOS) sensor. In some embodiments, the server apparatus may be arranged in data or signal communication with the CMOS sensor to receive image signals or data captured by the CMOS sensor for further processing. The server apparatus may be remote with respect to the CMOS sensor or may be integrated with the CMOS sensor in an integrated circuit chip.

[0047] In some embodiments, the predicted optimal exposure times may be used to generate an enhanced image with high dynamic range.

[0048] In some embodiments, the server apparatus may be part of a distributed system, the server apparatus arranged or operable to receive input image data, which may include a temporary image requiring exposure correction, to generate a corrected image, in the form of a HDR image.

[0049] The server apparatus may comprise a processor and a memory, the processor is capable of being configured to execute instructions stored in the memory to receive image data. In the embodiment illustrated in FIG. 1A, the server apparatus may be a communications server apparatus. The communications server apparatus may be in the form of a server computer 100, the server computer 100 may be a single server as illustrated schematically in FIG. 1A, or have the functionality performed distributed across multiple server components.

[0050] In some embodiments, the server computer 100 includes a communication interface 102. The communication interface 102 may be configured to send and receive data, which may include image data, the image data may include a temporary image, denoted as Z_N, to be processed, and its exposure time, denoted as Δt_N. The output of the server computer 100 may be an estimated optimal exposure time of all the pixels $\Delta t^*(p)$ in the image Z_N.

[0051] The communication interface 102 may include a transmitter module and/or a receiver module allowing the server apparatus to communicate over a communications network. The communication interface 102 may include one or more user-interfaces configured to provide users for user control and may include, for example, one or more computing peripheral devices such as display monitors, computer keyboards and the like.

[0052] The server computer 100 may further include a processor in the form of processing unit 104 and a memory 106. The memory 106 may be used by the processing unit 104 to store, for example, the image data, and the exposure time associated with the image data, and may optionally include other associated metadata, identifiers, etc.

[0053] In some embodiments, the server apparatus as shown in FIG. 1B may comprise an input module 110, the input module 110 configured to obtain the input image data from, for example, a CMOS sensor, and a predictor module 120 configured to receive pre-process or converted image data 111, process the input image data 111 and generate predicted optimal exposure time 112 associated with each pixel of the input image. An output module 130 may be configured to further process or display an enhanced image 113, based on the predicted optimal exposure time, on an interface, such as a user interface. In some embodiments, the CMOS sensor may provide RAW images which may be converted to RGB images by the input module 110 before the prediction by the predictor module 120.

[0054] Referring to the flowchart of FIG. 2, there is provided a method 200 of predicting optimal exposure times for one or more non-overlapping blocks of an image. The image may be a RAW image captured by a CMOS sensor, and subsequently converted to an RGB image.

[0055] In step S201, the method comprises training a convolutional neural network (CNN) on a dataset comprising temporary images of a first scene and corresponding sets of ground-truth optimal exposure times for all pixels of each temporary image.

[0056] In step S202, the method comprises obtaining a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene.

[0057] In step S203, the method comprises inputting the temporary image of the second scene and its initial exposure time into the trained CNN.

[0058] In step S204, the method comprises generating, by the CNN, a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.

[0059] In some embodiments, the method further comprises generating an exposure adjustment map based on the set of predicted optimal exposure times, wherein the exposure adjustment map represents per-pixel exposure adjustments of the temporary image of the second scene.

[0060] In some embodiments, the exposure adjustment map comprises a greyscale heatmap.

[0061] In some embodiments, the method further comprises reconstructing a spatially-adapted HDR image based on the exposure adjustment map applied on the temporary image of the second scene.

[0062] In some embodiments, the CNN has a U-Net architecture, and a shallow residual U-Net configuration.

[0063] In some embodiments, the U-Net architecture comprises an encoder-decoder structure to process exposure variations at different scales.

[0064] In some embodiments, the encoder of the U-Net architecture comprises a first 3x3 convolution kernel, a parametric ReLU (PReLU) activation module, and a second 3x3 convolution kernel.

[0065] In some embodiments, the method further comprises defining a 16 by 16 non-overlapping block or a 32 by 32 non-overlapping block.

[0066] In some embodiments, the training of the CNN in step S201 may include constructing a training dataset and testing dataset to facilitate training, validation, and testing. The training dataset may comprise a population of samples, for example 100 samples or 1000 samples, to facilitate supervised learning of the CNN. In some embodiments, each sample in the training dataset may comprise a temporary image of one or more first scenes, which may be one or more HDR scenes, and its exposure time, as well as the optimal exposure times of all the pixels for the first scene. The optimal exposure times of all pixels associated with the first scene(s) may be obtained from historical records or generated based on feedback from experts, including human experts. It is worth noting that the optimal exposure times of two pixels in the same 16x16 block are usually different. The optimal exposure times of all the pixels will serve as the ground-truth for the training, validation, and testing of the CNN algorithm. It is appreciable that the training dataset comprise variations of the first scene(s) so as to enhance the versatility and/or robustness of the trained CNN.

[0067] In some embodiments, the loss function for facilitating the supervised training may be based on a comparison of a predicted exposure correction with a ground truth exposure correction at different spatial locations. In other words, the loss function may be defined by using the optimal exposure times (ground-truth) rather than the captured image. This represents a uniqueness of the proposed training of the CNN.

[0068] In some embodiments, a temporary image of a second scene is captured by using the learnt block-wise exposure times by the proposed trained CNN. Rather than minimizing the number of over- and under-exposed pixels while keeping most pixels well-exposed, the exposure time of each non-overlapping block is made to approach the optimal exposure times of all the pixels in the block.

[0069] In some embodiments, a CNN having a U-Net architecture may be used for the training and learning of the optimal exposure times. It may be appreciable that a simple CNN may be chosen as the learnable architecture, as CNNs may be well established architectures for feature encoding in the image domain, and are thus a natural choice for learning the mapping.

[0070] As indicated, each sample in the training dataset comprises a temporary image $\boldsymbol{Z_N}$ and its exposure time $\boldsymbol{\Delta t_N}$ as well as the optimal exposure times of all the pixels $\boldsymbol{\Delta t^*(p)}$'s which are supposed to be approached by the proposed CNN structure, which comprises a block-wise auto-exposure algorithm during the training of the proposed algorithm. For two pixels $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ in the same 16x16 block, $\boldsymbol{\Delta t^*(p_1)}$ and $\boldsymbol{\Delta t^*(p_2)}$ are usually different. The $\boldsymbol{\Delta t^*(p)}$ is small if the pixel p is in a bright region, and large if the pixel p is in a dark region.

[0071] The proposed framework on learnable block-wise auto-exposure for CMOS sensors is illustrated in FIG. 3. The inputs are the temporary image $\boldsymbol{Z_N}$ and its exposure time $\boldsymbol{\Delta t_N}$, and the outputs are the optimal exposure times of all the non-overlapping 16 x16 blocks $\boldsymbol{\Delta t_l}$'s. It is contemplated that the non-overlapping blocks may have other sizes, not limited to 16 x 16.

[0072] FIG. 3 represents a deep learning-based exposure correction network that adaptively adjusts exposure at different spatial locations.

[0073] An input image Z_N at the top may represent an input scene with varying exposures. The dimensions of W x H x 3 indicates that the input image Z_N may be an RGB image.

[0074] In some embodiments, the input image Z_N may be a low-dynamic-range (LDR) capture that needs exposure correction. It may be appreciable that spatial variations in exposure may exist, requiring local adjustments rather than a global exposure shift.

[0075] In the feature extraction layers 310 or initial convolution layers, spatial features may be extracted from the input image, including underexposed and overexposed regions. Multi-scale context are captured to facilitate exposure correction adaptation to local structures rather than applying a uniform exposure shift.

[0076] The initial convolution layers may perform shallow feature extraction, wherein the input image Z_N (size W × H × 3) passes through convolutional layers (Conv 3×3), the PReLU (Parametric ReLU) activation introduces non-linearity, ensuring exposure features are efficiently captured. The initial convolution layers results in a feature map of size W × H × 16, where 16 represents the number of channels.

[0077] The CNN comprises encoding layers 320. The encoding layers may comprise convolutional layers (Conv 3×3) and PReLU activations to extract exposure-relevant features. The feature extraction & encoding layers 310 may be configured to (i.) extract spatial features from the input image Z_N, including underexposed and overexposed regions, (ii.) capture multi-scale context to ensure exposure correction adapts to local structures rather than applying a uniform exposure shift.

[0078] The encoder may be configured to capture multi-scale spatial features, including regions that are underexposed or overexposed.

[0079] The encoder layers 320 may feature hierarchical feature learning, wherein the encoder progressively reduces spatial resolution while increasing feature channels:

[0080] Encoder 1: W/2 × H/2 ×32 (Down-sampling, increasing depth)

[0081] Encoder 2: W/4 × H/4 ×64

[0082] Encoder 3: W/8 × H/8 × 128

[0083] In some embodiments, the down-sampling may be based on strided convolution or max pooling.

[0084] In other words, the encoder acts like a spatial pyramid, capturing local, mid-range, and global exposure variations.

[0085] The decoding and concatenation layers 330 are configured to reconstruct the full-resolution spatially-varying exposure correction map, and recover fine details lost in encoding, hence facilitating exposure adjustments are not overly smoothed.

[0086] The decoding and concatenation layers 330 is configured to up-sample feature maps while reducing depth, and may comprise

[0087] Decoder 1: W/4 × H/4 × 64

[0088] Decoder 2: W/2 × H/2 × 32

[0089] Decoder 3: W × H × 16

[0090] Each decoder block may comprise up-sampling (which may include bilinear interpolation or transposed convolution)

[0091] Convolutional layers (Conv 3×3)

[0092] Activation (PReLU)

[0093] At each decoder stage, feature maps from the corresponding encoder stage may be concatenated as follows:

[0094] Decoder 1+Encoder 2

[0095] Decoder 2+Encoder 1

[0096] Such an arrangement may help to restore spatial details, preventing exposure artifacts, for example, halo effects.

[0097] The concatenation facilitates the preservation of fine-grained local details from early layers while benefiting from deeper context-aware features.

[0098] After the last decoding layer, a final convolutional layer (Conv 3×3) + PReLU 340 generates the exposure correction map 350, which may be a spatially-varying exposure correction map 350. This is a grayscale heatmap (W × H × 1), wherein:

[0099] Brighter areas indicates positive exposure compensation (brightening underexposed regions).

[00100] Darker areas indicates negative exposure compensation (darkening overexposed regions).

[00101] This means each pixel is corrected individually based on its local exposure, rather than applying a uniform exposure correction to the entire image.

[00102] The exposure adjustment map 350 may be applied to reconstruct a spatially-adaptive HDR image with balanced exposure 360.

[00103] It may be appreciable that the structure depicted in FIG. 3 shows a CNN utilizing a shallow residual U-Net architecture with three levels, 3x3 convolution kernels, 1x1 convolution kernels, PReLU nonlinearities, downsampling with 2x 2 average pooling, upsampling by 3x 3 deconvolution layers (transpose convolution).

[00104] The number of channels in the first U-Net level (encoder 1, encoder 2, encoder 3) increases by a factor of 2. The channels are doubled in each of the three levels of the U-Net.

[00105] In some embodiments, in addition to learning the optimal exposure times Δt_i, the ratios of the optimal time of Δt_i's of all non-overlapping blocks to all input exposure times,

i.e. $\Delta t_l/\Delta t_N$'s are learnt by using the U-Net for all the non-overlapping 16x16 blocks. The objective is to increase the robustness of the proposed algorithm. It may be appreciable that other forms of CNN, not limited to U-Net, may be contemplated.

[00106] FIG. 4 shows a general block diagram of an encoder-decoder arrangement of a U-Net architecture used in some embodiments of present disclosure.

[00107] In summary, the encoder module 410 is configured to extract spatially-varying exposure features from one or more input image(s), and the decoder module 420 is configured to reconstruct an exposure-modulated image while maintaining spatial details.

[00108] The encoder 410 comprises a first Conv 3×3 layers 412, the first Conv 3×3 layers 412 configured to capture local textures and variations in brightness, PReLU (Parametric ReLU) module 414, the PReLU module 414 introduces non-linearity while allowing small negative values, helping in preserving fine exposure details, a second Conv 3×3 layers 416, the second Conv 3×3 layers 416 further refines feature extraction, emphasizing spatially varying exposure characteristics.

[00109] The decoder 420 comprises a first DeConv 3×3 (transposed convolution) module 422, the DeConv 3x3 module 422 configured to upsample the encoded feature maps to restore spatial resolution.

[00110] The PReLU module 424 ensures non-linearity in reconstruction, helping adapt to different exposure levels.

[00111] The final Conv 1×1 module 426 maps features back to an appropriate output format, such as an exposure-modulated HDR image or a tone-mapped image suitable for display.

[00112] The U-Net based CNN network is configured to learn spatially adaptive exposure adjustments, ensuring that both dark and bright regions are properly exposed in the final HDR image.

[00113] By using U-Net's skip connections (see FIG. 3), fine-grained spatial details are preserved, preventing over-smoothing.

[00114] In some embodiments, the loss function, denoted L, is mathematically expressed in Equation (1) as follows.

$$L = \sum_l \sum_{p \in \Omega_l} |\Delta t_l - \Delta t^*(p)| \quad (1)$$

[00115] where Δt_l's are the outputs of the proposed network in FIG. 3 and FIG. 4. Ω_l denotes the *l*th non-overlapping 16x16 block.

[00116] The equation facilitates the training of the CNN by comparing predicted exposure shifts $\Delta t^*(p)$ with the actual candidate solution Δt_l at different spatial locations.

[00117] This ensures per-pixel exposure accuracy.

[00118] The Absolute Difference $|\Delta t_l - \Delta t^*(p)|$ measures the absolute error between:

[00119] Δt_l, which could be the predicted or computed exposure value (candidate solution) at level *l*, and $\Delta t^*(p)$, which is the ground truth or reference exposure value at pixel *p*. The absolute value ensures non-negative differences.

[00120] It is appreciable that as an alternative to the loss function of Equation (1), which is also referred to as a L1 loss function, a L2 loss function based on a squared difference instead of the absolute difference may be contemplated. The L2 loss function may penalize larger errors more heavily than smaller ones due to squaring, but may produce a smoother and more continuous gradient, making optimization more stable.

[00121] It is contemplated that the method 200 may be performed using a non-transitory computer-readable storage medium storing instructions that, when executed by a processor, cause the processor to perform the method.

[00122] It is contemplated that the steps of the various disclosed methods are not restricted to the sequence set forth in the claims unless explicitly required. Steps may be performed in a different order, concurrently, or omitted entirely, depending on the embodiment.

[00123] Although the present disclosure is described in the context of the training and usage of convolutional neural network (CNN), it is contemplated that other suitable artificial intelligence or machine-learning based tools, such as vision transformers, capsule networks, graph neural networks (GNNs), fourier neural operators, diffusion models, self-attention mechanisms, and/or hybrid/combinations of any the aforementioned algorithms.

[00124] The following examples may be provided solely for illustrative purposes and may be not intended to limit the scope of the method, system, and/or server apparatus for the prediction of optimal exposure time of images, such as HDR images. The described examples illustrate specific implementations but do not encompass all possible variations. Features and aspects from different examples may be combined in various ways to suit particular applications, and modifications may be made without departing from the spirit and scope of the method, system, and/or server apparatus as defined by the claims.

[00125] Example 1 may be a method of predicting optimal exposure times for all non-overlapping blocks of an image, the method including: training an artificial intelligence module on a dataset comprising temporary images of a first scene and corresponding sets of ground-truth optimal exposure times for all pixels of each temporary image; obtaining a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene; inputting the temporary image of the second scene and its initial exposure time into the trained artificial intelligence module; generating, by the artificial intelligence module ,a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.

[00126] In Example 2, the artificial intelligence module of the subject matter of Example 1 can optionally include a neural network based algorithm, such as, but not limited to, a convolutional neural network (CNN).

[00127] In Example 3, the subject matter of any one of Examples 1 or 2 can optionally include defining a loss function based on a comparison of a predicted exposure correction with a ground truth exposure correction at different spatial locations.

[00128] In Example 4, the subject matter of any one of Examples 1 to 3 can optionally include generating an exposure adjustment map based on the set of predicted optimal exposure times, wherein the exposure adjustment map represents per-pixel exposure adjustments of the temporary image of the second scene.

[00129] In Example 5, the exposure adjustment map of Example 4 can optionally include a greyscale heatmap.

[00130] In Example 6, the subject matter of any one of Examples 3 or 4 can optionally include reconstructing a spatially-adapted HDR image based on the exposure adjustment map applied on the temporary image of the second scene.

[00131] In Example 7, the subject matter of Example 2 can optionally include a U-Net architecture, and a shallow residual U-Net configuration.

[00132] In Example 8, the subject matter of Example 7 can optionally be configured such that the U-Net architecture comprises an encoder-decoder structure to process exposure variations at different scales.

[00133] In Example 9, the subject matter of Example 8 can optionally be configured such that the encoder of the U-Net architecture comprises a first 3x3 convolution kernel, a parametric ReLU (PReLU) activation module, and a second 3x3 convolution kernel.

[00134] In Example 10, the subject matter of any one of Examples 1 to 9 can optionally include defining a 16 by 16 non-overlapping block or a 32 by 32 non-overlapping block.

[00135] Example 11 may be a non-transitory computer-readable storage medium storing instructions that, when executed by a processor, cause the processor to perform any of the subject matter of any one of Examples 1 to 10.

[00136] Example 12 may be a server apparatus for predicting optimal exposure times for non-overlapping blocks of an image, the server apparatus comprises a processor or processing unit, the processor or processing unit configured to: train an artificial intelligence module on a dataset comprising temporary images of a first scene and corresponding sets of optimal exposure times for all pixels of each temporary image; obtain a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene; input the temporary image of the second scene and its initial exposure time into the trained artificial intelligence module; and generate, by the artificial intelligence module, a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.

[00137] In Example 13, the artificial intelligence module of the subject matter of Example 12 can optionally include a neural network based algorithm, such as, but not limited to, a convolutional neural network (CNN).

[00138] In Example 14, the processor or processing unit of the subject matter of Example 13 can optionally be configured to define a loss function, wherein the loss function based on a comparison of a predicted exposure correction with a ground truth exposure correction at different spatial locations.

[00139] In Example 15, the processor or processing unit of the subject matter of any one of the subject matter of Examples 12 to 14 can optionally be configured to generate an exposure adjustment map based on the set of predicted optimal exposure times, and wherein the exposure adjustment map represents per-pixel exposure adjustments of the temporary image of the second scene.

[00140] In Example 16, the exposure adjustment map of the subject matter of Example 15 can optionally comprise a greyscale heatmap.

[00141] In Example 17, the processor or processing unit of the subject matter of claim 16can optionally be configured to reconstruct a spatially-adapted HDR image based on the exposure adjustment map.

[00142] In Example 18, the CNN of the subject matter of Example 13 can optionally include a U-Net architecture, and a shallow residual U-Net configuration.

[00143] In Example 19, the U-Net architecture of the subject matter of Example 18 can optionally include an encoder-decoder structure to process exposure variations at different scales.

[00144] In Example 20, the encoder of the U-Net architecture of Example 19 can optionally include a first 3x3 convolution kernel, a parametric ReLU (PReLU) activation module, and a second 3x3 convolution kernel.

[00145] In Example 21, the processor or processing unit of the subject matter of any one of Examples 12 to 20 can be configured to define a 16 by 16 non-overlapping block or a 32 by 32 non-overlapping block.

[00146] Example 22 may be a CMOS image sensor configured to capture a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene; input the temporary image of the second scene and its initial exposure time into the trained artificial intelligence module of the server apparatus of any one of Examples 12 to 20; and generate, by the artificial intelligence module, a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.

[00147] In Example 23, the subject matter of Example 22 can optionally integrate the server apparatus of any one of Examples 12 to 20, or can be arranged in data or signal communication with the server apparatus of any one of Examples 12 to 20.

[00148] While the disclosure has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the disclosure as defined by the appended claims. The scope of the disclosure is thus indicated by the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

CLAIMS

1. A method of predicting optimal exposure times for all non-overlapping blocks of an image, comprising:

training an artificial intelligence module on a dataset comprising temporary images of a first scene and corresponding sets of ground-truth optimal exposure times for all pixels of each temporary image;

obtaining a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene;

inputting the temporary image of the second scene and its initial exposure time into the trained artificial intelligence module ;

generating, by the artificial intelligence module , a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.

2. The method of claim 1, wherein the artificial intelligence module comprises a convolutional neural network (CNN).

3. The method of claim 1 or 2, further comprises defining a loss function based on a comparison of a predicted exposure correction with a ground truth exposure correction at different spatial locations.

4. The method of any one of claims 1 to 3, further comprises generating an exposure adjustment map based on the set of predicted optimal exposure times, wherein the exposure adjustment map represents per-pixel exposure adjustments of the temporary image of the second scene.

5. The method of claim 4, wherein the exposure adjustment map comprises a greyscale heatmap.

6. The method of claim 3 or 4, further comprising reconstructing a spatially-adapted HDR image based on the exposure adjustment map applied on the temporary image of the second scene.

7. The method of claim 2, wherein the CNN has a U-Net architecture, and a shallow residual U-Net configuration.

8. The method of claim 7, wherein the U-Net architecture comprises an encoder-decoder structure to process exposure variations at different scales.

9. The method of claim 8, wherein the encoder of the U-Net architecture comprises a first 3x3 convolution kernel, a parametric ReLU (PReLU) activation module, and a second 3x3 convolution kernel.

10. The method of any one of the preceding claims, further comprises defining a 16 by 16 non-overlapping block or a 32 by 32 non-overlapping block.

11. A non-transitory computer-readable storage medium storing instructions that, when executed by a processor, cause the processor to perform a method of any one of the preceding claims.

12. A server apparatus of predicting optimal exposure times for non-overlapping blocks of an image, the server apparatus comprises a processor, the processor configured to:

train an artificial intelligence module on a dataset comprising temporary images of a first scene and corresponding sets of optimal exposure times for all pixels of each temporary image;

obtain a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene;

input the temporary image of the second scene and its initial exposure time into the trained artificial intelligence module; and

generate, by the artificial intelligence module, a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.

13. The server apparatus of claim 12, wherein the artificial intelligence module comprises a convolutional neural network (CNN).

14. The server apparatus of claim 13, wherein the processor is configured to define a loss function, wherein the loss function based on a comparison of a predicted exposure correction with a ground truth exposure correction at different spatial locations.

15. The server apparatus of any one of claims 12 to 14, wherein the processor is configured to generate an exposure adjustment map based on the set of predicted optimal exposure times, and wherein the exposure adjustment map represents per-pixel exposure adjustments of the temporary image of the second scene.

16. The server apparatus of claim 15, wherein the exposure adjustment map comprises a greyscale heatmap.

17. The server apparatus of claim 16, wherein the processor is configured to reconstruct a spatially-adapted HDR image based on the exposure adjustment map.

18. The server apparatus of claim 13, wherein the CNN has a U-Net architecture, and a shallow residual U-Net configuration.

19. The server apparatus of claim 18, wherein the U-Net architecture comprises an encoder-decoder structure to process exposure variations at different scales.

20. The server apparatus of claim 19, wherein the encoder of the U-Net architecture comprises a first 3x3 convolution kernel, a parametric ReLU (PReLU) activation module, and a second 3x3 convolution kernel.

21. The server apparatus of any one of claims 12 to 20, further comprises defining a 16 by 16 non-overlapping block or a 32 by 32 non-overlapping block.

22. A CMOS image sensor configured to capture a temporary image of a second scene and determining an initial exposure time of the temporary image of the second scene;

input the temporary image of the second scene and its initial exposure time into the trained artificial intelligence module of the server apparatus of any one of claims 12 to 20; and

generate, by the artificial intelligence module, a set of predicted optimal exposure times for the non-overlapping blocks of the temporary image of the second scene.