SERVER APPARATUS, METHOD, AND SYSTEM FOR IMAGE PROCESSING

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This patent application claims priority to and the benefit of Singapore Patent Application No. 10202400526R, filed on 27 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, apparatus, and system for image processing. In some embodiments, the image processing may be based on multi-scale image fusion techniques to dynamically combine multiple exposure images into an output image.

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] Fusing multiple differently exposed low dynamic range (LDR) images of a high dynamic range (HDR) scene into an information enriched LDR image is an efficient way to overcome the limited dynamic ranges of cameras. Due to possible camera movement and moving objects, the LDR images are first aligned and all the moving objects are synchronized according to a pre-defined reference image.

[0005] Existing exposure fusion algorithms range from conventional methods to deep learning based ones. However, conventional methods may not be detailed due to their generic nature.

[0006] Therefore, there exists a need for improved image processing of differently exposed images for efficient image fusion with adaptive weight maps.

SUMMARY

[0007] The present disclosure was conceptualized to provide a technical solution for processing differently exposed images to produce an output composite image. The technical solution is based on a multi-scale exposure fusion (MEF) for high dynamic range (HDR) imaging. In some embodiments, the inputs to the MEF are a set of differently exposed LDR RGB images, and the bit-depth of the input image(s) could be from 8 to 14. The output is an HDR-like RGB image with the bit depth same as or higher than that of the input images. Hybrid pyramids and loss functions are proposed for neural augmentation based MEF. The loss functions and the set of inputting RGB images with different exposures may be decoupled based on two observations: 1) each set of inputting images may not be able to cover the whole dynamic range of the corresponding HDR scene; and 2) a set of sufficient RGB images with different exposures from an HDR scene is able to cover the whole dynamic range of the HDR scene even though each RGB image is with low dynamic range (LDR). In some embodiments, all the inputting differently exposed RGB images and additional RGB images with different exposures from the same HDR scene are used to define the loss functions. As a result, the fused image approaches the HDR scene rather than the set of inputting images.

[0008] According to an aspect of the present disclosure there is provided a server apparatus for image processing, the server apparatus comprising an input module, the input module configured to obtain input image data, the input image data comprising a plurality of differently exposed images; an image processing module configured to process the plurality of differently exposed images to generate a composite image; wherein the image processing module comprises a neural augmented multi-scale exposure fusion function, the neural augmented multi-scale exposure fusion function comprises: a plurality of levels; one or more weight maps of the plurality of differently exposed images, each weight in the one or more weight maps associated with an image of the plurality of differently exposed images; a plurality of first pyramids corresponding to the plurality of levels, the plurality of first pyramids generated based on the one or more weight maps; and a plurality of hybrid pyramids, each of the plurality of hybrid pyramids constructed based on at least one edge-preserving smoothing function using the plurality of first pyramids as input.

[0009] In some embodiments, the neural augmented multi-scale exposure fusion function further comprises a plurality of second pyramids are generated based on the differently exposed images.

[0010] In some embodiments, the plurality of first pyramids are Gaussian pyramids and the plurality of second pyramids are Laplacian pyramids.

[0011] In some embodiments, the Laplacian pyramids are combined with the hybrid pyramids to form a fused Laplacian pyramid associated with the composite image.

[0012] In some embodiments, the fused Laplacian pyramid is collapsed to produce the composite image.

[0013] In some embodiments, the edge-preserving smoothing function comprises edge-preserving smoothing (EPS) pyramids of the one or more weight maps, and content adaptive edge-preserving smoothing (CAS) pyramids of the one or more weight maps.

[0014] In some embodiments, the one or more weight maps is trained using an unsupervised learning method.

[0015] In some embodiments, the unsupervised learning method comprises a Multi-Exposure Fusion Structural Similarity Index (MEF-SSIM) based loss function, the MEF-SSIM based loss function independent of one or more ground-truth images used to train the one or more weight maps.

[0016] In some embodiments, the loss function further comprises a weighted mean absolute error loss function for smoothing the one or more weight maps.

[0017] In some embodiments, the loss function and the plurality of differently exposed images to be fused are decoupled in a manner such that the loss function is defined by another plurality of differently exposed images from the same scene which covers a higher dynamic range than the plurality of differently exposed images to be fused.

[0018] In some embodiments, the bit-depth of the images is from 8 to 14.

[0019] In some embodiments, the plurality of levels is eight levels.

[0020] According to another aspect of the present disclosure there is provided a system for image processing comprising any of the aforementioned server apparatus.

[0021] According to another aspect of the present disclosure there is provided a method of configuring or constructing a neural multi-scale exposure fusion (MEF) algorithm for use in processing a plurality of differently exposed images to produce a composite image, the method comprising: predefining a plurality of levels of pyramids of the neural augmented multi-scale exposure fusion function; generating a plurality of first pyramids; learning one or more weight maps of the plurality of first pyramids via an unsupervised method; constructing or computing a plurality of hybrid pyramids based on on at least one edge-preserving smoothing function

using the plurality of first pyramids as input; and obtaining the composite image by collapsing the Laplacian pyramid.

[0022] In some embodiments, the method further comprises generating a plurality of second pyramids based on the differently exposed images.

[0023] In some embodiments, the plurality of first pyramids are Gaussian pyramids and the plurality of second pyramids are Laplacian pyramids.

[0024] In some embodiments, the method further comprises combining the Laplacian pyramids with the hybrid pyramids to form a fused Laplacian pyramid associated with the composite image.

[0025] In some embodiments, the fused Laplacian pyramid is collapsed to produce the composite image.

[0026] In some embodiments, the edge-preserving smoothing function comprises edge-preserving smoothing (EPS) pyramids of the one or more weight maps, and content adaptive edge-preserving smoothing (CAS) pyramids of the one or more weight maps.

[0027] In some embodiments, the unsupervised learning method comprises a Multi-Exposure Fusion Structural Similarity Index (MEF-SSIM) based loss function, the MEF-SSIM based loss function independent of one or more ground-truth images used to train the one or more weight maps.

[0028] According to another aspect of the present disclosure there is provided a computer program element comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform any one of the aforementioned method.

[0029] According to another aspect of the present disclosure there is provided a non-transitory computer-readable medium comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform any one of the aforementioned method.

BRIEF DESCRIPTION OF THE DRAWINGS

[0030] 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 image processing, in particular processing a plurality of differently exposed images to obtain a output fused or composite image.
- FIG. 1B is a schematic block diagram comprising the server apparatus of FIG. 1A, comprising various modules for image processing.
- FIG. 2 is a flowchart depicting a method of configuring or constructing a multi-scale exposure fusion (MEF) algorithm.
- FIG. 3 is a flowchart depicting a generalised method for processing images according to some embodiments.

DETAILED DESCRIPTION

[0031] 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.

[0032] 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.

[0033] 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.

[0034] 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.

[0035] 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.

[0036] 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.

[0037] 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.

[0038] As used herein, the term "associate", "associated", and "associating" indicate a defined relationship (or cross-reference) between two items.

[0039] 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.

[0040] 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.

[0041] 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.

[0042] As used herein, the term "differently exposed image/images" broadly refers to one or more digital images capturing the same scene but with varying exposure levels. In some embodiments, the differently exposed image/images may be obtained by adjusting camera settings such as shutter speed, aperture, and/or ISO sensitivity. Such images may collectively represent a wider dynamic range of the scene than can be captured in a single exposure.

[0043] As used herein, the term "neural augmented multi-scale exposure fusion" may refer to any computational method that combines multiple differently exposed images into a single

high-quality image by utilizing both traditional multi-scale decomposition techniques and neural network-based refinement. Such fusion may involve decomposing input images into multiple scales, fusing them using predetermined rules, and then refining the result using a neural network trained to enhance image quality and preserve details from various exposure levels.

[0044] As used herein, the term "pyramid" refers to hierarchical image representations used in multi-scale image processing. Such image representations may include Gaussian pyramids, which comprises a series of images created by repeatedly smoothing and down-sampling an original image, resulting in a stack of increasingly lower-resolution versions of the image, and Laplacian pyramid, which comprises a series of difference images obtained by subtracting each level of the Gaussian pyramid from the up-sampled version of a next coarser level, preserving edge and detail information at different scales.

[0045] As used herein, the term "edge-preserving smoothing" refers to a function and/or an image processing algorithm designed to reduce noise and texture in an image while maintaining sharp edges and important structural features. In some embodiments, an edge-preserving smoothing function may employ adaptive filtering techniques that adjust smoothing strength based on local image characteristics, such as gradient magnitude or statistical properties of neighboring pixels.

[0046] As used herein, the term "level(s)" broadly refer to the different scales or resolutions at which image processing and fusion operations may be performed. In particular, pyramid levels refers to the different resolutions in image pyramids, such as Gaussian or Laplacian pyramids, used in multi-scale decomposition. Each level represents the image at a different scale, with higher levels containing coarser details and lower levels containing finer details. Feature representation levels refers to different layers of a neural network architecture, each extracting features at varying levels of abstraction. Lower levels may capture low-level features like edges, while higher levels capture more complex, abstract features. Fusion levels may refer to the stages at which fusion operations are performed, combining information from multiple exposure images at different scales or network depths.

[0047] According to an aspect of the present disclosure there is provided a server apparatus for image processing. The server apparatus may form part of a processor or circuit of an image capturing device.

[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 differently exposed images, to generate a composite image, in the form of a fused 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 the image data, the image data may include a plurality of differently exposed images having varying brightness levels, different exposure values, diverse dynamic range representation, varying detail preservation, variations in contrast, differences in color saturation, and/or metadata.

[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 speech data packets, historical quality data associated with similar speech data packets, and/or quality-based parameters such as scores, to be processed.

[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 input image data, the input image data comprising a plurality of differently exposed images, an image processing module 120 configured to obtain (which may include receive) the plurality of differently exposed images and generate a composite image 112. An output module 130 may be configured to further process or display the composite image on an interface.

[0054] In some embodiments, the image processing module 120 comprises a neural augmented multi-scale exposure fusion function, the neural augmented multi-scale exposure fusion function comprises: a plurality of levels; a weight map of the plurality of differently

exposed images, each weight in the weight map associated with an image of the plurality of differently exposed images; a plurality of first pyramids corresponding to the plurality of levels, the plurality of first pyramids generated based on the weight map; and a plurality of hybrid pyramids, each of the plurality of hybrid pyramids constructed based on at least one edge-preserving smoothing function using the plurality of first pyramids as input.

[0055] As shown in FIG. 1B, the image processing module 120 may be configured to receive, from the input module 110, input image data 111 for processing. The output may be a composite or fused image 112.

[0056] In some embodiments, the server computer 100 and one or more terminal devices 110 may be connected via a network, which may be an Internet or Intranet network.

[0057] FIG. 2 shows a flow-chart of a method 200 of configuring or constructing a multi-scale exposure fusion (MEF) algorithm according to some embodiments of the present disclosure. The MEF algorithm may be trained using unsupervised learning, the unsupervised learning may be incorporated as one of the steps. The neural augmented multi-scale exposure fusion function may be configured to receive N differently exposed images, denoted as Z_i's ($1 \leq i \leq N$). The output of the neural augmented multi-scale exposure fusion function may be an information enrich LDR image Z_f. $Z_i (1 \leq i \leq N)$ is a set of differently exposed LDR images with N denoted as a number of the LDR images and such a set is denoted as Ω_f. The bit depth of each Z_i may be from 8 to 14, such as, but not limited to, bit depth of 8, 10, 12, or 14. In some embodiments, the exposure time of Z_i may be denoted as Δt_i, and where the exposure time of the N images are in a relationship mathematically expressed in Equation (1) as follows.

$$\Delta t_1 < \Delta t_2 < \cdots < \Delta t_N \quad (1)$$

[0058] In some embodiments, the set of images Ω_f may be a typical geometric exposure variation, and each available exposure setting is a constant factor larger than its subsequent lower setting.

[0059] The various notation are follows. Z_f denotes the fused image. Y_i denotes the luminance component of the image Z_i, and Y_f denotes the luminance component of the image Z_f. For simplicity, five different pyramids are defined in Table 1 as follows.

Table 1

Notation	Pyramid
$G\{Y_i\}$	Gaussian pyramid of Y_i
$L\{Y_i\}$	Laplacian pyramid of Y_i
$E\{Y_i\}$	EPS pyramid of Y_i
$C\{Y_i\}$	CAS pyramid of Y_i
$H\{Y_i\}$	Hybrid pyramid of Y_i

[0060] In step S201: predefining a plurality of levels of pyramids of neural augmented multi-scale exposure fusion function. In some embodiments, the levels may be denoted as κ, and where $\kappa > 4$.

[0061] In step S202: Generating a plurality of second pyramids, e.g. Laplacian pyramids $L\{Z_i\}$'s and $L\{Y_i\}$'s as well as a plurality of first pyramids, e.g. Gaussian pyramids $G\{Z_i\}$'s and $G\{Y_i\}$'s.

[0062] In step S203: Learning the weight maps $G\{W_i\}^{(4)}$'s from the Gaussian pyramid $G\{Y_i\}^{(4)}$)'s via an unsupervised method. The superscript (4) denotes level 4 of the Gaussian pyramid. It is contemplated that the middle layer, in this case 4, may be selected.

[0063] In some embodiments, the weight maps may be formed using three quality measures Ci(p), Si(p), and Ei(p), which measure contrast, color saturation, and well-exposedness of pixel $Z_i(p)$ of the *i*th image, respectively. Ci(p) may be computed by applying a Laplacian filter to the gray-scale version of the *i*th image. Si(p) may be obtained as the standard deviation among the three-color channels of the pixel Zi(p). Ei(p) may be obtained/calculated by applying a Gauss curve to each channel separately and multiplying the results. In some embodiments, the weight associated with a pixel may be based on the product of Ci(p), Si(p), and Ei(p), i.e. let the product of Ci(p), Si(p), and Ei(p), be denoted as $\widetilde{W}_i(p)$. The weight map of a particular image Z_i may be computed according to Equation (2) as follows.

$$W_i = \frac{\widetilde{W}_i}{\sum_{j=1}^{N} \widetilde{W}_j}, \qquad (2)$$

[0064] In some embodiments, the total number of levels for all the pyramids is denoted as $\kappa(\kappa > 4)$. The recommended value of κ is 8. In some embodiments, all the weight maps Wi's ($1 \leq i \leq N$) may be first computed via the Equation (2) by using the full-size images Z_i. They may then be used to generate the Gaussian pyramids $G\{W_i\}$'s ($1 \leq i \leq N$).

[0065] In step S204: Constructing or computing the hybrid pyramids $H\{W_i\}$'s. Based on the computed weight maps Wi's, the pyramids of the weight maps may generated as follows:

[0066] The EPS and CAS pyramids may be constructed from the Gaussian pyramids $G\{W_i\}$'s by using the WGIF or iWGIF.

[0067] In some embodiments, let $G\{Y_i\}^{(l)}$ denote the lth level of the Gaussian pyramid $G\{Y_i\}$. To reduce the complexity of the proposed algorithm, the weight maps $G\{W_i\}^{(4)}$'s may be learnt from $G\{Y_i\}^{(4)}$'s via an unsupervised learning method.

[0068] The Gaussian pyramid $G\{W_i\}^{(l)}(4 \le l \le \kappa)$ is then constructed from the $G\{W_i\}^{(4)}$ by applying the Gaussian pyramid to the weight map $G\{W_i\}^{(4)}$. Subsequently, the EPS pyramid $E\{W_i\}^{(l)}(1 \le l \le \kappa)$ may be constructed by using the GFU, and the CAS pyramid $C\{W_i\}^{(l)}(4 \le l \le \kappa)$ is constructed by using the $E\{W_i\}^{(l)}(4 \le l \le \kappa)$ and $G\{W_i\}^{(l)}(4 \le l \le \kappa)$. The proposed hybrid pyramids $H\{W_i\}^{(l)}(1 \le l \le \kappa)$ are finally constructed, and denoted in Equation (3)

$$H\{W_i\}^{(l)} = \begin{cases} E\{W_i\}^{(l)}, & l < 4 \\ C\{W_i\}^{(l)}, & l \ge 4 \end{cases}. \quad (3)$$

[0069] Rather than computing the weight maps for all the differently exposed images by using the full size differently exposed images, the weight maps of all the differently exposed images are initially learnt from the differently exposed images at a middle level of the Laplacian pyramids. The learnt weight maps are then expanded by using the CAS pyramids from the middle level to all the down-sized levels and the EPS pyramids at the other levels. The resultant pyramids are called hybrid pyramids. The proposed hybrid pyramids fully utilize the availability of the weight maps from the middle level to all the down-sized levels. It is worth noting that the expanded CAS pyramids can further reduce possible artifacts caused by the learnt weight maps. This is similar to the effect of the Gaussian, EPS, and CAS pyramids to reduce the halo artifacts caused by the weight maps.

[0070] The hybrid pyramids of the weight maps and the Laplacian pyramids of the differently exposed images are then blended to have the Laplacian pyramid of the fused image which is finally collapsed to produce the fused image. Since the proposed algorithm is multi-scale, it can preserve the depth of scene better than the single-scale exposure fusion algorithms as the multi-scale can add scene depth to the fused image. This is very important for the fused image because it looks natural and does not look like a flat "cartoony" rendition.

[0071] In some embodiments, the weight maps are smoothed using Weighted Guided Image Filter (WGIF) or improved Weighted Guided Image Filter (iWGIF).

[0072] In step S205: Constructing or computing the Laplacian pyramid for the fused image Z_f via the equation (4). In this step, once the hybrid pyramids $H\{W_i\}$'s are available from Equation (3), all the images Z_i's (1≤ i ≤_N) at the different pyramid levels are blended or combined according to Equation (4) as follows.

$$L\{Z_f\}^{(l)} = \sum_{i=1}^{N} L\{Z_i\}^{(l)} H\{W_i\}^{(l)}, \quad (4)$$

[0073] In step S206: Producing or obtaining the fused image Z_f by collapsing the Laplacian pyramid $L\{Z_f\}$. In this step, the Laplacian pyramid $L\{Z_f\}$ is finally collapsed to produce the fused image Z_f.

[0074] In some embodiments, one or more loss functions may be utilized for the unsupervised training of the weight map(s) in step S204. An example loss function may be a Multi-Exposure Fusion Structural Similarity (MEF-SSIM). The MEF-SSIM may be adopted to train the data-driven component (e.g. weight map) because it may be independent of the ground-truth image. The MEF-SSIM loss function $L_S(\Omega_f, Z_f)$ is defined by using the set of images to be fused and the fused image. Besides the loss function in the equation $L_S(\Omega_f, Z_f)$, one more loss function may be derived from the MEF algorithms.

[0075] In some embodiments, a loss function may be a weight mean absolute error function, aimed to reduce sharp weight map transitions. The normalized weight maps W_i in the Equation (2) may be first smoothed by using the WGIF or the iWGIF functions with a guidance (reference) image. In some embodiments, the guidance image may include the luminance component of each input image. A loss function $L_W(\Omega_f, Z_f)$ may then be defined and mathematically expressed in Equation (5) as follows.

$$L_W(\Omega_f, Z_f) = \sum_{i=1}^{N} \sum_p W_i(p) \left| Z_i(p) - Z_f(p) \right| . \quad (5)$$

[0076] The function of Equation (5) represents a weighted sum of absolute differences between input images $Z_i(p)$ and the final fused image $Z_f(p)$, where $W_i(p)$ is a weighting function for each pixel p.

[0077] In general, MEF-SSIM-based loss functions typically aim to preserve perceptual quality by ensuring structural similarity between the fused image and the input images.

[0078] The loss function encourages the fused image to be a weighted combination of the input images while minimizing per-pixel absolute differences. The weights likely depend on local contrast, brightness, or exposure measures, aligning with MEF-SSIM principles.

[0079] It may be appreciated that the loss function $L_W(\Omega_f, Z_f)$ may also be independent of the ground-truth image, and the loss functions $L_S(\Omega_f, Z_f)$ and $L_W(\Omega_f, Z_f)$ may be defined by using the fused image Z_f and the set Ω_f. This may also be true for the deep learning based MEF algorithms. In other words, the loss functions and the set Ω_f are tightly coupled in the existing deep learning based MEF algorithms. In some cases, the relative brightness order may not be well preserved in the fused image when two large-exposure-ratio (LER) images are fused. This issue arises because high LER values can cause significant disparities in brightness across different regions, making it difficult to maintain a consistent intensity relationship when merging images. In some embodiments, details in the brightest and darkest regions are not preserved well when two or three normal-exposure-ratio (NER) images are fused. This issue arises because the two or three NER images are not able to cover the whole dynamic range of an HDR scene. As a result, undesired artifacts or distortions may emerge in the fused output, impacting the overall quality and realism of the image. This problem can be mitigated by introducing an additional set Ω_m to replace the set Ω_f. The key advantage of this approach is that it enables a decoupling between the loss functions and the set Ω_f, thereby offering greater flexibility in optimizing the fusion process. Instead of being constrained by Ω_f, which may impose direct dependencies on brightness ordering and image structure, the new formulation using $\Omega_m$allows for independent control over fusion constraints and loss calculations. Such decoupling ensures that loss functions can be designed to prioritize perceptual quality and structural coherence without being overly influenced by the limitations imposed by Ω_f, leading to more robust and visually consistent HDR image fusion results. The set Ω_f may be a subset of the set Ω_m. The set Ω_m may cover a higher dynamic range than the set Ω_f.

[0080] Two cases of the set Ω_f and set Ω_m may be elaborated as follows. Case 1 Exposure interpolation: The set Ω_f is composed of two images Z_1 and Z_4 while the set Ω_m consists of Z_1, Z_2, Z_3, and Z_4. Case 2 Exposure extrapolation: The set Ω_f is composed of two images Z_2 and Z_3 while the set Ω_m consists of Z_1, Z_2, Z_3, and Z_4. Since the ground-truth fused images are not available, the data-driven component in the proposed algorithm can be trained in an unsupervised way mathematically expressed in Equation (6) as follows.

$$\min_{\theta}\{L_S(\Omega_m, Z_f) + \gamma L_W(\Omega_m, Z_f)\}, \quad (6)$$

where θ are the parameters of the deep learning component, and γ is a constant hyper-parameter. It is contemplated that the loss functions required by object detection, depth estimation, semantic segmentation, and so on may also be considered if the fused images serve as the inputs to these processes.

[0081] According to another aspect of the present disclosure, there is provided a method of an input module, the input module configured to obtain input image data, the input image data comprising a plurality of differently exposed images, to generate an output composite image. The composite image may be a fused image.

[0082] FIG. 3 shows a flowchart of a method 300 for processing images. The method may be implemented as executable software codes stored in one or more non-transitory computer-readable medium of the processing unit 104 and/or server computer 100. The processing unit 104 may part of the data processing unit of an image capturing device, such as a camera or a video recorder/ camcorder, capable of capturing images with differently exposed images. In some embodiments, the server apparatus as described may form part of the image capturing device, or be arranged in data or signal communication with the image capturing device.

[0083] In step S301: obtaining a plurality of differently exposed images;

[0084] In step S302: processing each of the plurality of differently exposed images by a neural augmented multi-scale exposure fusion function, the neural augmented multi-scale exposure fusion function comprises: a plurality of levels; a weight map of the plurality of differently exposed images, each weight in the weight map associated with an image of the plurality of differently exposed images; a plurality of first pyramids corresponding to the plurality of levels, the plurality of first pyramids generated based on the weight map; and a plurality of hybrid pyramids, each of the plurality of hybrid pyramids constructed based on at least one edge-preserving smoothing function using the plurality of first pyramids as input.

[0085] According to another aspect of the present disclosure, the server apparatus and method may be used in applications to low-light imaging and/or digital HDR imaging. In some embodiments, the proposed loss functions for training the weight map(s) may be applied to study lowlight imaging.

[0086] In one non-limiting example, the input image data may comprises a 12 to 14-bit raw image denoted as I_1 rather than an 8-bit sRGB image Z_1 that is captured by using a very small exposure time Δt_1. A data-driven method may be designed to generate a high quality Standard

Red Green Blue (sRGB) composite or fused image Z_f. In some embodiments, the composite or fused image Z_f may be required to approach a desired sRGB image Z_2, which may be captured on the same scene with a much larger exposure time Δt_2. A possible issue with such a method is that information in the brightest regions could be saturated.

[0087] With the proposed loss functions, the composite image Z_f obtained may approach the image $Z_2$and one more image from the same scene with an exposure time smaller than Δt_2 but larger than Δt_1. As the present disclosure uses unsupervised learning instead of supervised learning, the desired sRGB image (for supervised training) may not available to be utilized. Thus, the proposed algorithm belongs to unsupervised learning to obtain the Z_f image which is enhanced to reveal details in low-light or near-dark conditions that would normally be invisible to the human eye. Another difference between the proposed algorithm and existing algorithms is that a raw image I_1 may be normalized by the ratio $1/\Delta t_1$ rather than being amplified by the ratio $\Delta t_2/\Delta t_1$ before serving as the input of the subsequent data-driven algorithm. The information in the brightest regions will be preserved better by the method 200.

[0088] In some embodiments, the proposed loss functions can also be applied to study digital HDR imaging. In one non-limiting example, one raw image $I_2$is captured with the exposure time as Δt_2 for an HDR scene. The image I_2 may be the input of digital HDR imaging and the output of the digital HDR imaging is an 8-bit sRGB image Z_f which looks like being fused from a set of differently exposed sRGB images. The raw image I_2 may be first normalized by dividing $1/\Delta t_2$ and then serves as the input of a subsequent unsupervised learning-based algorithm. The loss functions as described may be defined by multiple differently exposed sRGB images from the same HDR scene. The output will be the desired image Z_f.

[0089] In some embodiments, the server computer 100 and the method 200, 300 may be deployed in an image signal processor (ISP). The ISP may be utilized for image signal processing for HDR imaging and low-light imaging that disrupts the traditional image processing workflow.

[0090] It is appreciable that the weight maps of differently exposed images may be important in the exposure fusion, and techniques such as Gaussian and Laplacian pyramids are applied to reduce halo artifacts. The Gaussian pyramids of the weight maps and the Laplacian pyramids of the different exposed images may be blended to form the Laplacian pyramid of the fused image which is collapsed to produce the fused image. In the present disclosure, the

Gaussian pyramids of the weights may be replaced by edge-preserving smoothing (EPS) pyramids of the weight maps and content adaptive edge-preserving smoothing (CAS) pyramids of the weight maps. The multi-scale exposure fusion (MEF) algorithms can preserve information in the brightest and darkest regions of HDR scenes better and produce higher MEF-SSIM values than the MEF algorithm The CAS pyramids outperform the EPS pyramids while the CAS pyramids require the Gaussian pyramids of the weight maps at all the levels.

[0091] It is contemplated that guided filtering for up-sampling (GFU) may be adopted to simplify the MEF algorithm. One feature of the GFU is that the coefficients of weighted guided image filter (WGIF) are only computed at two levels of the pyramids and they are up-sampled to obtain the coefficients of the WGIF at other levels. The other is that the weight maps can be computed from the luminance components and the coefficients of the WGIF at all the other levels. The unsupervised learning-based exposure fusion algorithms are also based on the GFU. In the present disclosure, instead of computing the weight maps by using the full-size images, the weight maps may be first learnt from down-sampled images and then up-sampled to the full-size by using the GFU. All the differently exposed images may be blended together by using the up-sampled weight maps. In some cases, the exposure fusion algorithms may be single-scale and the depth of scene might not be preserved well by them. Besides the MEF, the GFU may also be adopted to develop a real time single image dehazing algorithm on mobile phones. It is appreciable that the coefficients of the weight maps cannot be up-sampled too many times. Otherwise, the coefficients may be too smooth and the structures of the guidance images cannot be transferred to the weighted maps well, and the quality of the final image may drop significantly. Therefore, in some embodiments, the coefficients of the WGIF are computed twice at two different levels.

[0092] In some embodiments, the algorithms may be used or implemented as software codes in smartphones (operating in, for example, a night mode), automotive night vision systems, and surveillance cameras operating in dim environments.

[0093] 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 processing a plurality of differently exposed images to obtain a composite or fused image. 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.

[0094] Example 1 may be a server apparatus for image processing, the server apparatus may include a processor, the processor configured to: obtain input image data, the input image data comprising a plurality of differently exposed images; process the plurality of differently exposed images to generate a composite image; wherein the processor is configured to execute a neural augmented multi-scale exposure fusion function, the neural augmented multi-scale exposure fusion function includes: a plurality of levels; one or more weight maps of the plurality of differently exposed images, each weight in the one or more weight maps associated with an image of the plurality of differently exposed images; a plurality of first pyramids corresponding to the plurality of levels, the plurality of first pyramids generated based on the one or more weight maps; and a plurality of hybrid pyramids, each of the plurality of hybrid pyramids constructed based on at least one edge-preserving smoothing function using the plurality of first pyramids as input.

[0095] In Example 1A, the processor of the subject matter of Example 1 can include an input module configured to obtain the input image data, and/or pre-process the input image data, and an image processing module configured to process the input image data.

[0096] In Example 2, the subject matter of Example 1 or Example 1A can optionally include the neural augmented multi-scale exposure fusion function further includes a plurality of second pyramids are generated based on the differently exposed images.

[0097] In Example 3, the subject matter of Example 2 can optionally include a plurality of first pyramids are Gaussian pyramids and the plurality of second pyramids are Laplacian pyramids.

[0098] In Example 4, the subject matter of Example 3 can optionally combine the Laplacian pyramids with the hybrid pyramids to form a fused Laplacian pyramid associated with the composite image.

[0099] In Example 5, the fused Laplacian pyramid of the subject matter of Example 4 can optionally be collapsed to produce the composite image.

[00100] In Example 6, the edge-preserving smoothing function of the subject matter of any of Examples 1 to 5 can optionally include edge-preserving smoothing (EPS) pyramids of the one or more weight maps, and content adaptive edge-preserving smoothing (CAS) pyramids of the one or more weight maps.

[00101] In Example 7, the weight maps of the subject matter of any of Examples 1 to 6 can be trained using an unsupervised learning method.

[00102] In Example 8, the unsupervised learning method of the subject matter of Example 7 can optionally include a Multi-Exposure Fusion Structural Similarity Index (MEF-SSIM) based loss function, the MEF-SSIM based loss function independent of one or more ground-truth images used to train the one or more weight maps.

[00103] In Example 9, the loss function of the subject matter of Example 8 can optionally include a weighted mean absolute error loss function for smoothing the one or more weight maps.

[00104] In Example 10, the loss function and the plurality of differently exposed images to be fused of Example 9 can optionally be decoupled in a manner such that the loss function is defined by another plurality of differently exposed images from the same scene which covers a higher dynamic range than the plurality of differently exposed images to be fused.

[00105] In Example 11, the bit-depth of the plurality of differently exposed images can optionally be in a range from Examples 8 to 14.

[00106] In Example 12, the plurality of levels of the subject matter of any of Examples 1 to 11 are eight levels.

[00107] Example 13 may be a system for image processing that optionally includes the server apparatus of any one of Examples 1 to 12.

[00108] Example 14 may be a method for configuring or constructing a neural multi-scale exposure fusion (MEF) algorithm for use in processing a plurality of differently exposed images to produce a composite image, the method including: predefining a plurality of levels of pyramids of the neural augmented multi-scale exposure fusion function; generating a plurality of first pyramids; learning one or more weight maps of the plurality of first pyramids via an unsupervised method; constructing or computing a plurality of hybrid pyramids based on on at least one edge-preserving smoothing function using the plurality of first pyramids as input; and obtaining the composite image by collapsing the Laplacian pyramid.

[00109] In Example 15, the subject matter of Example 14 further includes generating a plurality of second pyramids based on the differently exposed images.

[00110] In Example 16, the plurality of first pyramids and second pyramids of the subject matter of Example 15 can optionally be Gaussian pyramids and Laplacian pyramids respectively.

[00111] In Example 17, the subject matter of Example 16 can optionally include combining the Laplacian pyramids with the hybrid pyramids to form a fused Laplacian pyramid associated with the composite image.

[00112] In Example 18, the fused Laplacian pyramid of the subject matter of Example 17 can optionally be collapsed to produce the composite image.

[00113] In Example 19, the edge-preserving smoothing function of the subject matter of any of Examples 14 to 18 can optionally include edge-preserving smoothing (EPS) pyramids of the one or more weight maps, and content adaptive edge-preserving smoothing (CAS) pyramids of the one or more weight maps.

[00114] In Example 20, the unsupervised learning method of the subject matter of any of Examples 14 to 19 can optionally include a Multi-Exposure Fusion Structural Similarity Index (MEF-SSIM) based loss function, the MEF-SSIM based loss function independent of one or more ground-truth images used to train the one or more weight maps.

[00115] Example 21 may be a computer program element including program instructions, which, when executed by one or more processors, cause the one or more processors to perform any one of Examples 14 to 20.

[00116] Example 22 may be a non-transitory computer-readable medium including program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of Examples 14 to 20.

[00117] 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 server apparatus for image processing, the server apparatus comprising

an input module, the input module configured to obtain input image data, the input image data comprising a plurality of differently exposed images;

an image processing module configured to process the plurality of differently exposed images to generate a composite image; wherein the image processing module comprises a neural augmented multi-scale exposure fusion function, the neural augmented multi-scale exposure fusion function comprises:

a plurality of levels;

one or more weight maps of the plurality of differently exposed images, each weight in the one or more weight maps associated with an image of the plurality of differently exposed images;

a plurality of first pyramids corresponding to the plurality of levels, the plurality of first pyramids generated based on the one or more weight maps; and

a plurality of hybrid pyramids, each of the plurality of hybrid pyramids constructed based on at least one edge-preserving smoothing function using the plurality of first pyramids as input.

2. The server apparatus of claim 1, wherein the neural augmented multi-scale exposure fusion function further comprises a plurality of second pyramids are generated based on the differently exposed images.

3. The server apparatus of claim 2, wherein the plurality of first pyramids are Gaussian pyramids and the plurality of second pyramids are Laplacian pyramids.

4. The server apparatus of claim 3, wherein the Laplacian pyramids are combined with the hybrid pyramids to form a fused Laplacian pyramid associated with the composite image.

5. The server apparatus of claim 4, wherein the fused Laplacian pyramid is collapsed to produce the composite image.

6. The server apparatus of any one of the preceding claims, wherein the edge-preserving smoothing function comprises edge-preserving smoothing (EPS) pyramids of the one or more weight maps, and content adaptive edge-preserving smoothing (CAS) pyramids of the one or more weight maps.

7. The server apparatus of any one of the preceding claims, wherein the one or more weight maps is trained using an unsupervised learning method.

8. The server apparatus of claim 7, wherein the unsupervised learning method comprises a Multi-Exposure Fusion Structural Similarity Index (MEF-SSIM) based loss function, the MEF-SSIM based loss function independent of one or more ground-truth images used to train the one or more weight maps.

9. The server apparatus of claim 8, wherein the loss function further comprises a weighted mean absolute error loss function for smoothing the one or more weight maps.

10. The server apparatus of claim 9, wherein the loss function and the plurality of differently exposed images to be fused are decoupled in a manner such that the loss function is defined by another plurality of differently exposed images from the same scene which covers a higher dynamic range than the plurality of differently exposed images to be fused.

11. The server apparatus of claim 10, wherein the bit-depth of the plurality of differently exposed images is in a range is from 8 to 14.

12. The server apparatus of any one of the preceding claims, wherein the plurality of levels is eight levels.

13. A system for image processing comprising the server apparatus of any one of claims 1 to 12.

14. A method of configuring or constructing a neural multi-scale exposure fusion (MEF) algorithm for use in processing a plurality of differently exposed images to produce a composite image, the method comprising:

predefining a plurality of levels of pyramids of the neural augmented multi-scale exposure fusion function;

generating a plurality of first pyramids;

learning one or more weight maps of the plurality of first pyramids via an unsupervised method;

constructing or computing a plurality of hybrid pyramids based on on at least one edge-preserving smoothing function using the plurality of first pyramids as input; and

obtaining the composite image by collapsing the Laplacian pyramid.

15. The method of claim 14, further comprises generating a plurality of second pyramids based on the differently exposed images.

16. The method of claim 15, wherein the plurality of first pyramids are Gaussian pyramids and the plurality of second pyramids are Laplacian pyramids.

17. The method of claim 16, further comprises combining the Laplacian pyramids with the hybrid pyramids to form a fused Laplacian pyramid associated with the composite image.

18. The method of claim 17, wherein the fused Laplacian pyramid is collapsed to produce the composite image.

19. The method of any one of claims 14 to 18, wherein the edge-preserving smoothing function comprises edge-preserving smoothing (EPS) pyramids of the one or more weight maps, and content adaptive edge-preserving smoothing (CAS) pyramids of the one or more weight maps.

20. The method of any one of claims 14 to 19, wherein the unsupervised learning method comprises a Multi-Exposure Fusion Structural Similarity Index (MEF-SSIM) based loss function, the MEF-SSIM based loss function independent of one or more ground-truth images used to train the one or more weight maps.

21. A computer program element comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 14 to 20.

22. A non-transitory computer-readable medium comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 14 to 20.