US 20190075309A1

(19) **United States**

(12) **Patent Application Publication**

WANG et al.

(10) **Pub. No.: US 2019/0075309 A1**

(43) **Pub. Date: Mar. 7, 2019**

(54) **VIDEO COMPRESSED SENSING RECONSTRUCTION METHOD, SYSTEM, ELECTRONIC DEVICE, AND STORAGE MEDIUM**

(71) Applicant: **PING AN TECHNOLOGY (SHENZHEN) CO., LTD.**, Shenzhen (CN)

(72) Inventors: **Jianzong WANG**, Shenzhen (CN); **Jing XIAO**, Shenzhen (CN)

(73) Assignee: **PING AN TECHNOLOGY (SHENZHEN) CO., LTD.**, Shenzhen (CN)

(21) Appl. No.: **16/084,234**

(22) PCT Filed: **Jun. 30, 2017**

(86) PCT No.: **PCT/CN2017/091311**

§ 371 (c)(1),
(2) Date: **Sep. 12, 2018**

(30) **Foreign Application Priority Data**

Dec. 30, 2016 (CN) 201611260793.6

Publication Classification

(51) **Int. Cl.**

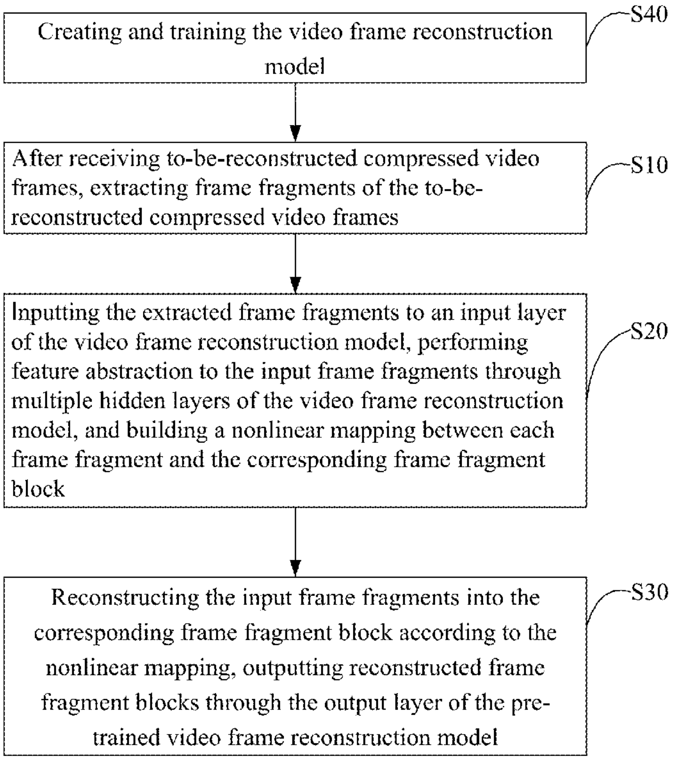
H04N 19/30	(2006.01)
H04N 19/14	(2006.01)
H04N 19/176	(2006.01)
H04N 19/89	(2006.01)
H04N 19/172	(2006.01)

(52) **U.S. Cl.**

CPC ***H04N 19/30*** (2014.11); ***H04N 19/14*** (2014.11); ***H04N 19/172*** (2014.11); ***H04N 19/89*** (2014.11); ***H04N 19/176*** (2014.11)

(57) **ABSTRACT**

The present disclosure provides a video compressed sensing reconstruction method, including: step B, after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the compressed video frames according to a predetermined extraction rule; step C, inputting the frame fragments into an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and a corresponding frame fragment block; and step D, reconstructing the input frame fragments to frame fragment blocks by the hidden layers according to the nonlinear mapping, and outputting the frame fragment blocks by an output layer of the video frame reconstruction model, and generating a reconstructed video based on the reconstructed frame fragment blocks. The present disclosure can render and reconstruct video frames quickly with a high quality.


Creating and training the video frame reconstruction model
S40
After receiving to-be-reconstructed compressed video frames, extracting frame fragments of the to-be-reconstructed compressed video frames
S10
Inputting the extracted frame fragments to an input layer of the video frame reconstruction model, performing feature abstraction to the input frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and the corresponding frame fragment block
S20
Reconstructing the input frame fragments into the corresponding frame fragment block according to the nonlinear mapping, outputting reconstructed frame fragment blocks through the output layer of the pre-trained video frame reconstruction model
S30

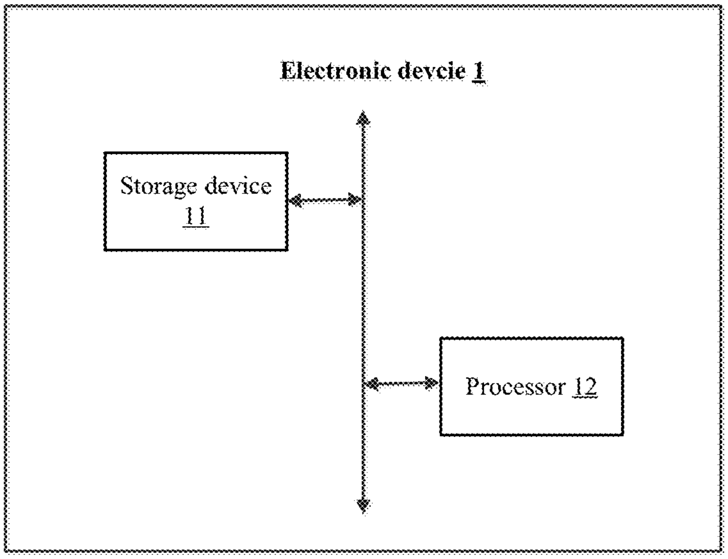
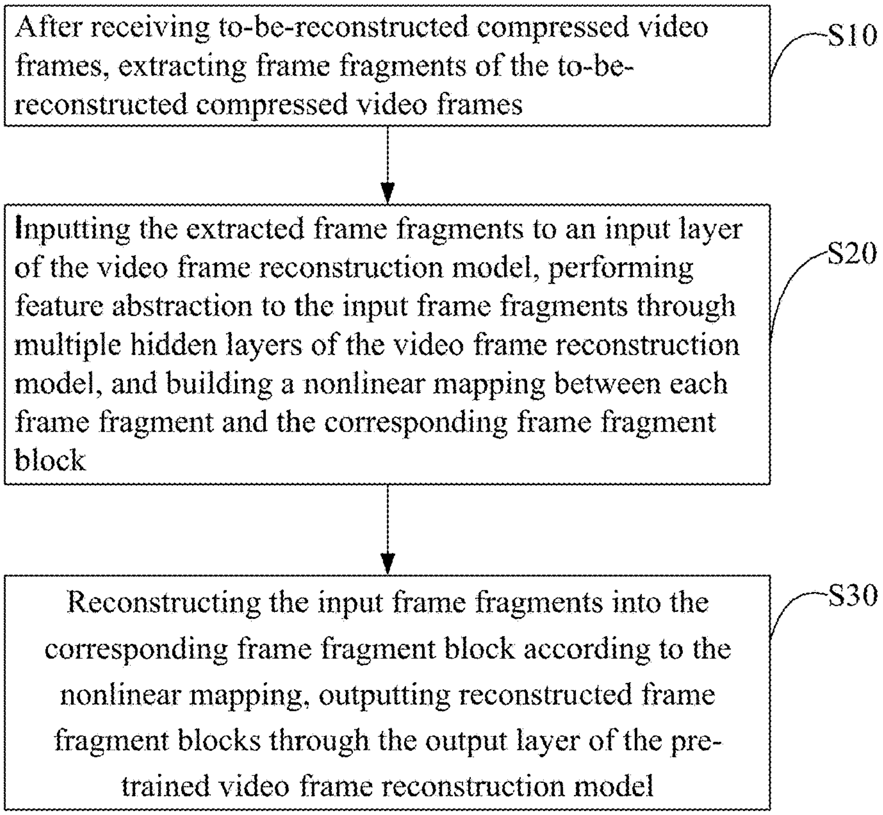


FIG. 1



After receiving to-be-reconstructed compressed video frames, extracting frame fragments of the to-be-reconstructed compressed video frames

Inputting the extracted frame fragments to an input layer of the video frame reconstruction model, performing feature abstraction to the input frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and the corresponding frame fragment block

Reconstructing the input frame fragments into the corresponding frame fragment block according to the nonlinear mapping, outputting reconstructed frame fragment blocks through the output layer of the pre-trained video frame reconstruction model

FIG. 2

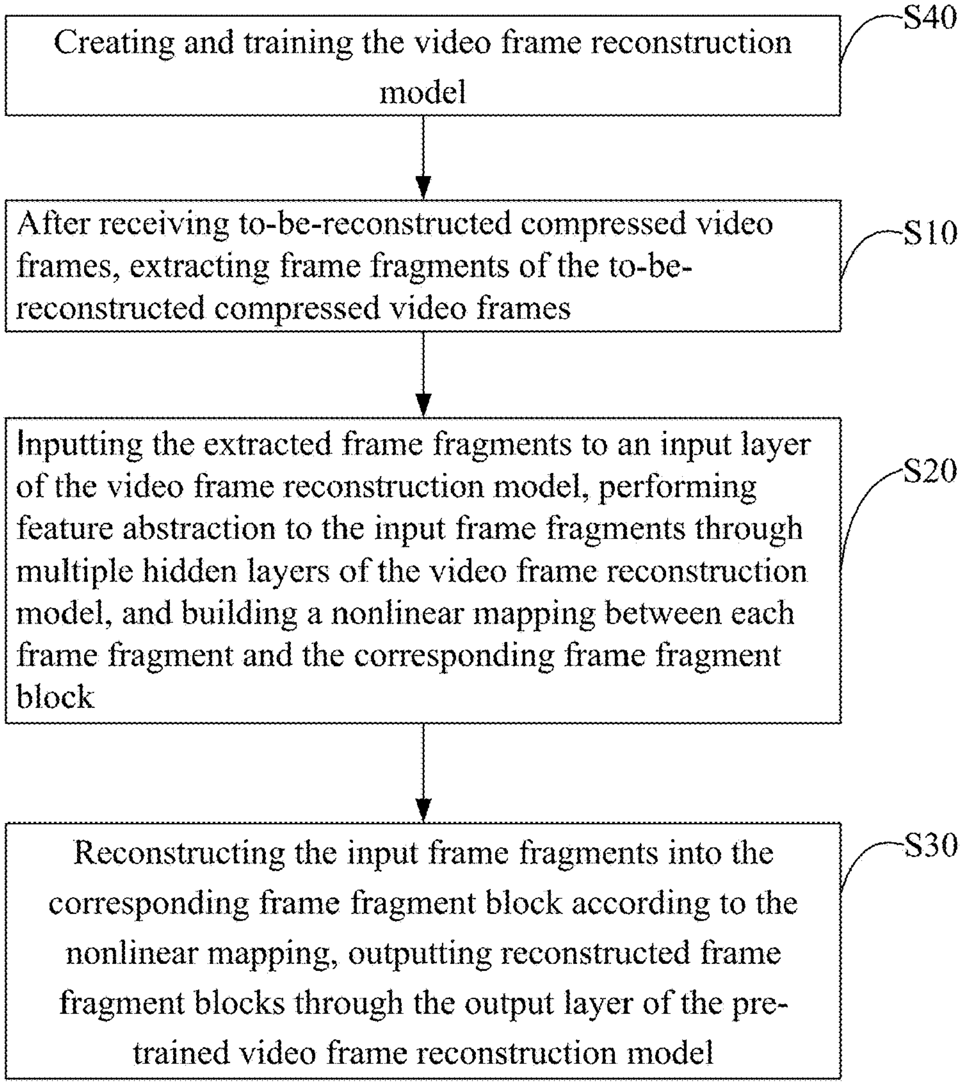

Creating and training the video frame reconstruction model
S40
After receiving to-be-reconstructed compressed video frames, extracting frame fragments of the to-be-reconstructed compressed video frames
S10
Inputting the extracted frame fragments to an input layer of the video frame reconstruction model, performing feature abstraction to the input frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and the corresponding frame fragment block
S20
Reconstructing the input frame fragments into the corresponding frame fragment block according to the nonlinear mapping, outputting reconstructed frame fragment blocks through the output layer of the pre-trained video frame reconstruction model
S30

FIG. 3

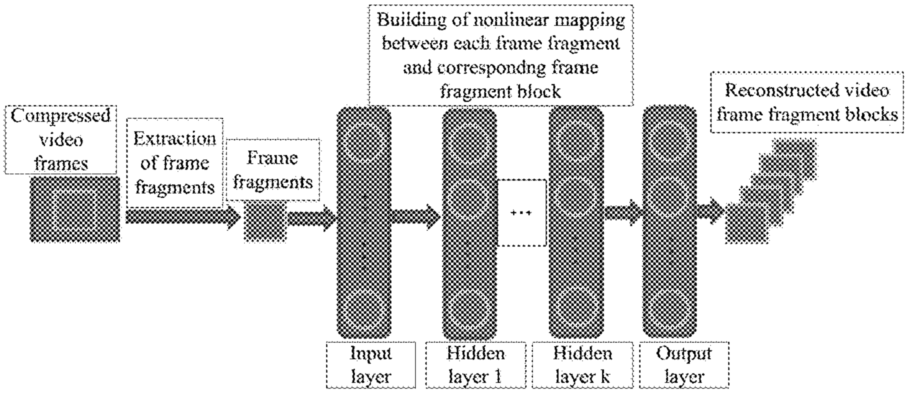

Building of nonlinear mapping between each frame fragment and correspondng frame fragment block
Compressed video frames
Extraction of frame fragments
Frame fragments
...
Reconstructed video frame fragment blocks
Input layer
Hidden layer 1
Hidden layer k
Output layer

FIG. 4

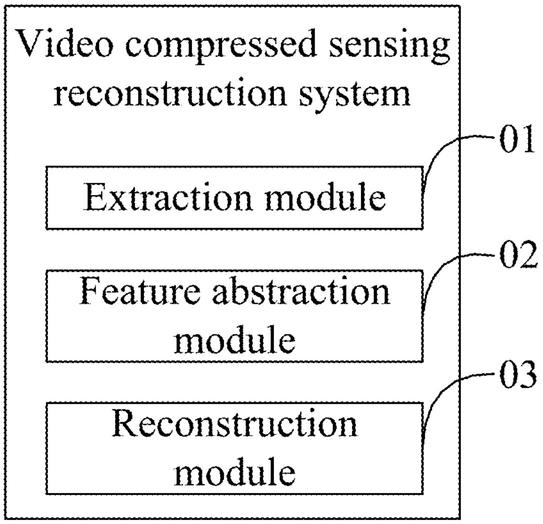

Video compressed sensing reconstruction system
Extraction module
01
Feature abstraction module
02
Reconstruction module
03

FIG. 5

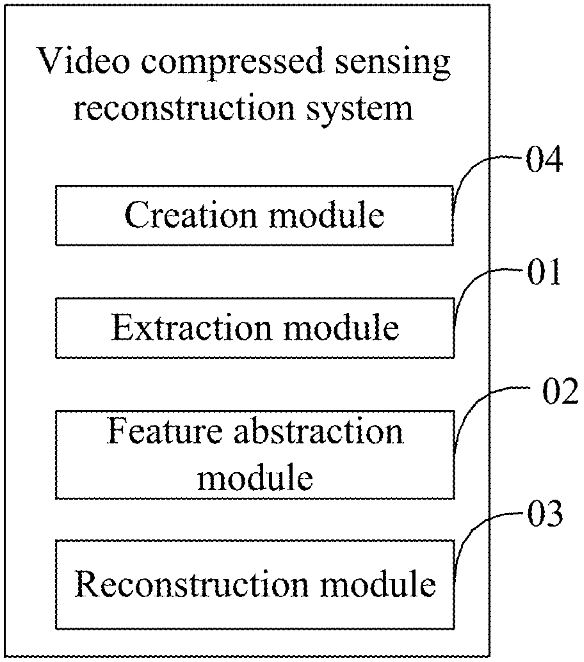

Video compressed sensing reconstruction system
Creation module
04
Extraction module
01
Feature abstraction module
02
Reconstruction module
03

FIG. 6

VIDEO COMPRESSED SENSING RECONSTRUCTION METHOD, SYSTEM, ELECTRONIC DEVICE, AND STORAGE MEDIUM

CROSS REFERENCE OF RELATED APPLICATIONS

[0001] The present application claims the benefit of Chinese Application No. 2016112607936, entitled "video compressed sensing reconstruction method and device" filed on Dec. 30, 2016, the entire content of which is incorporated herein in its entirety.

TECHNICAL FIELD

[0002] The present disclosure relates to computer technologies, and more particularly, to a video compressed sensing reconstruction method, system, device and storage medium.

BACKGROUND

[0003] At present, typical video compressed sensing algorithms based on time domain are very sensitive to computing complexity; especially when video frames are rendered and reconstructed, the computing speed is extremely slow. The situation cannot be obviously improved even graphics processing units are used to achieve parallel acceleration. Although there are algorithms capable of finishing sensing and reconstruction of video blocks, the reconstruction quality is relatively low. Thus, how to render and reconstruct video frames rapidly with a high quality has become a technical problem to be solved.

SUMMARY OF THE DISCLOSURE

[0004] The present disclosure provides a video compressed sensing reconstruction method, system, electronic device, and storage medium for rendering and reconstructing video frames rapidly with a high quality.

[0005] A first aspect of the present disclosure provides a video compressed sensing reconstruction method, including:

[0006] step B, after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the compressed video frames according to a predetermined extraction rule;

[0007] step C, inputting the frame fragments into an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and a corresponding frame fragment block; and

[0008] step D, reconstructing the input frame fragments to frame fragment blocks by the hidden layers according to the nonlinear mapping, and outputting the frame fragment blocks by an output layer of the video frame reconstruction model, and generating a reconstructed video based on the reconstructed frame fragment blocks.

[0009] A second aspect of the present disclosure provides a video compressed sensing reconstruction system, including:

[0010] an extraction module, configured for, after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the compressed video frames according to a predetermined extraction rule;

[0011] a feature abstraction module, configured for inputting the frame fragments into an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and a corresponding frame fragment block; and

[0012] a reconstruction module, configured for reconstructing the input frame fragments to frame fragment blocks by the hidden layers according to the nonlinear mapping, outputting the frame fragment blocks by an output layer of the video frame reconstruction model, and generating a reconstructed video based on the reconstructed frame fragment blocks.

[0013] A third aspect of the present disclosure provides an electronic device including a processor, a storage device and a video compressed sensing reconstruction system; the video compressed sensing reconstruction system is stored in the storage device and includes at least one computer-readable instruction which is capable of being executed by the processor to perform:

[0014] after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the compressed video frames according to a predetermined extraction rule;

[0015] inputting the frame fragments into an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and a corresponding frame fragment block; and

[0016] reconstructing the input frame fragments to frame fragment blocks by the multiple hidden layers according to the nonlinear mapping, outputting the frame fragment blocks by an output layer of the video frame reconstruction model, and generating a reconstructed video based on the reconstructed frame fragment blocks.

[0017] A fourth aspect of the present disclosure provides a computer-readable storage medium with at least one computer-readable instruction stored thereon, which can be executed by at least one processor to perform:

[0018] after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the compressed video frames according to a predetermined extraction rule;

[0019] inputting the frame fragments into an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and a corresponding frame fragment block; and

[0020] reconstructing the input frame fragments to frame fragment blocks by the multiple hidden layers according to the nonlinear mapping, outputting the frame fragment blocks by an output layer of the video frame reconstruction model, and generating a reconstructed video based on the reconstructed frame fragment blocks.

[0021] In the method, system, electronic device, and storage medium of the present disclosure, the frame fragments of the to-be-reconstructed compressed video frames are extracted according to the predetermined extraction rule; feature abstraction is performed to each frame fragment through the multiple hidden layers of the pre-trained video frame reconstruction model, and the nonlinear mapping between each frame fragment and the corresponding frame fragment block is built; and the input frame fragments are reconstructed to the frame fragment blocks according to the nonlinear mapping, and the reconstructed video blocks are output. Since the reconstruction of the video is carried out corresponding to the frame fragments extracted from the to-be-reconstructed compressed video frames rather than by processing the large compressed video frames directly, thus the computing complexity is reduced and the reconstruction speed of the video is improved. In addition, since the feature abstraction of the frame fragment is carried out through the multiple hidden layers of the pre-trained video frame reconstruction model, and the frame fragments are reconstructed to the frame fragment blocks for outputting, each detailed feature of the compressed video frames can be effectively extracted and thus the quality of the reconstruction of the video frames is improved.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] FIG. **1** is a schematic view showing an application environment of a video compressed sensing reconstruction method in accordance with an embodiment of the present disclosure;

[0023] FIG. **2** is a flow chart illustrating processes of a video compressed sensing reconstruction method in accordance with a first embodiment of the present disclosure;

[0024] FIG. **3** is a flowchart illustrating processes of a video compressed sensing reconstruction method in accordance with a second embodiment of the present disclosure;

[0025] FIG. **4** is a schematic view of a video frame reconstruction model applied in the video compressed sensing reconstruction method in accordance with an embodiment of the present disclosure;

[0026] FIG. **5** is a block diagram of a video compressed sensing reconstruction system in accordance with a first embodiment of the present disclosure; and

[0027] FIG. **6** is a block diagram of a video compressed sensing reconstruction system in accordance with a second embodiment of the present disclosure.

[0028] The realization of objectives, functional features and advantages of the present disclosure will be further described with reference to the accompanying drawings in combination with embodiments of the present disclosure.

PREFERRED EMBODIMENTS

[0029] For clearly understanding technical features, purpose, and effect of the present disclosure, embodiments are given in detail hereinafter with reference to the accompanying drawings.

[0030] FIG. **1** shows an application environment of a video compressed sensing reconstruction method in accordance with an embodiment of the present disclosure. An electronic device **1** can be one which is capable of performing numerical calculation and/or information processing automatically according to preset or pre-stored instructions. In some embodiments, the electronic device **1** can be a computer, a single network server, a server group having multiple network servers or a cloud server formed by a large number of host machines or network servers. Cloud computing is one type of distributed computing, referring to a super virtual computer formed by a group of discrete computers which are coupled together.

[0031] In the embodiment, the electronic device **1** includes a storage device **11** and a processor **12**. The processor **12** is configured for supporting the operation and running of the electronic device **1**, including one or more micro-processors or digital processors, etc. The storage device **11** is configured for storing various types of data or computer readable instructions. The storage device may include one or more non-volatile memory, such as ROM, EPROM, or flash memory. In an embodiment, the storage device **11** stores a video compressed sensing reconstruction program which includes at least one computer-readable instruction. The at least one computer-readable instruction can be executed by the processor **12** to perform a video compressed sensing reconstruction method of the present disclosure.

[0032] The present disclosure provides the video compressed sensing reconstruction method.

[0033] Referring to FIG. **2**, which is a flow chart illustrating processes of the video compressed sensing reconstruction method in accordance with a first embodiment of the present disclosure, the video compressed sensing reconstruction method includes steps as follows.

[0034] Step S**10**, after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the to-be-reconstructed compressed video frames according to a predetermined extraction rule.

[0035] In the embodiments, after the to-be-reconstructed compressed video frames are received, the compressed video frames are not rendered and reconstructed directly, and the frame fragments are extracted from the compressed video frames according to the predetermined extraction rule at first. The predetermined extraction rule can be extracting the frame fragments according to features like color, content, format, or area, etc., which is not limited herein.

[0036] In some embodiments, the predetermined extraction rule is: dividing the to-be-reconstructed compressed video frames into blocks such that the to-be-reconstructed compressed video frames are divided into several frame fragments. For example, various types of the compressed video frames, such as compressed video frames of JPEG or PNG type are divided into N*M frame fragments when the compressed video frames are divided into blocks, wherein N and M are positive integers. The frame fragments are equal to each other. In other embodiments, the compressed video frames can be divided into unequal frame fragments according to a preset ratio or randomly, which is not limited herein. Each frame fragment can have a regular shape like a square or a rectangle or irregular shapes, which is not limited herein.

[0037] Step S**20**, inputting the extracted frame fragments to an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the input frame fragments through multiple hidden layers of the pre-trained video frame reconstruction model, and building a nonlinear mapping between each frame fragment and the corresponding frame fragment block.

[0038] After the extraction of the frame fragments of the compressed video frames, the pre-trained video frame reconstruction model can be used to process the frame fragments correspondingly. The video frame reconstruction model can be established and trained each time when the video compressed sensing reconstruction is carried out or can be a pre-established and pre-trained model which is called each time when video compressed sensing reconstruction is carried out, which is not limited herein.

[0039] For example, in the embodiment, the pre-trained video frame reconstruction model includes the input layer, an output layer, and multiple hidden layers. After the frame fragments of the compressed video frames are extracted, the extracted frame fragments are input to the input layer of the pre-trained video frame reconstruction model; the hidden layers perform feature abstraction to the input frame fragments, build the nonlinear mapping between each frame fragment and the corresponding frame fragment block, and thus form a connection between each frame fragment and the reconstructed frame fragment block.

[0040] Step S**30**, reconstructing the input frame fragments into the corresponding frame fragment block according to the nonlinear mapping built by the hidden layers of the pre-trained video frame reconstruction model, outputting reconstructed frame fragment blocks through the output layer of the pre-trained video frame reconstruction model, and generating are constructed video based on the reconstructed frame fragment blocks.

[0041] The hidden layers of the pre-trained video frame reconstruction model reconstruct the input frame fragments into the final frame fragment blocks according to the nonlinear mapping, that is, the mapping between each frame fragment and the corresponding final reconstructed frame fragment block after feature abstraction of each frame is performed. The reconstructed video frame blocks are output through the output layer and the reconstructed video is generated based on the reconstructed frame fragment blocks. In some embodiments, the frame fragment blocks can be spliced or combined to generate the reconstructed video, thereby realizing the rendering and reconstruction of the compressed video frames.

[0042] In the method of the present disclosure, the frame fragments of the to-be-reconstructed compressed video frames are extracted according to the predetermined extraction rule; feature abstraction is performed to each frame fragment through the multiple hidden layers of the pre-trained video frame reconstruction model, and the nonlinear mapping between each frame fragment and the corresponding frame fragment block is built; and the input frame fragments are reconstructed to the frame fragment blocks according to the nonlinear mapping, and the reconstructed video blocks are output. Since the reconstruction of the video is carried out corresponding to the frame fragments extracted from the to-be-reconstructed compressed video frames rather than by processing the large compressed video frames directly, thus the computing complexity is reduced and the reconstruction speed of the video is improved. In addition, since the feature abstraction of the frame fragment is carried out through the multiple hidden layers of the pre-trained video frame reconstruction model, and the frame fragments are reconstructed to the frame fragment blocks for outputting, each detailed feature of the compressed video frames can be effectively extracted and thus the quality of the reconstruction of the video frames is improved.

[0043] Referring to FIG. **3**, a video compressed sensing reconstruction method in accordance with a second embodiment of the present disclosure is provided. The method of the second embodiment further includes following steps before step S**10** of the method of the first embodiment.

[0044] Step S**40**, creating and training the video frame reconstruction model which includes at least one input layer, at least one output layer, and multiple hidden layers.

[0045] In the embodiment, before the reconstruction of the video frames, the pre-trained video frame reconstruction model is created and trained. The pre-trained video frame reconstruction model includes at least one input layer, one output layer, and multiple hidden layers. Before the video frame reconstruction model is trained, training data and testing data are generated.

[0046] A preset number (e.g., 100) of videos in various types of natural scenes are obtained, and each video is converted into a grayscale space. A total amount of data of all the videos satisfies a preset value (e.g., 10K).

[0047] Each converted video is compressed through a measurement transformation matrix having a pre-defined size being $w_m \cdot * \cdot h_m \cdot * \cdot d_m$ (e.g., $w_m=8$, $h_m=8$, $d_m=16$). For example, w_b is a width of the video block having a preset number of videos, h_b is a height of the video block, d_b is a length of the video block (that is the number of the video frames). Each video block is $x_i \in w_b \times h_b \times d_b$, $i \in N$, wherein N is a positive integer no less than 1. The compressed video frame is $y_i \cdot \in \cdot w_b \cdot * \cdot h_b$, wherein $y_i = \varnothing_b x_i$, wherein $\varnothing_b$ is the measurement transformation matrix.

[0048] All the compressed videos are divided into a first data set and a second data set at a preset ratio such as X:Y (e.g., 7:3) wherein both X and Y are greater than 0. The number of the videos in first data set is greater than that of the videos in the second data set. The first data set is used as a training data set and the second data set is used as a testing data set.

[0049] In some embodiments, the training process of the video frame reconstruction model is as follows.

[0050] At the training of the video frame reconstruction model, a batch size of the input video frames is set to be 200, a total number of training times can be set to be $10*10^6$ times of iterations, and a size difference between every two video frames is specified within a range where an average value is 0 and a standard deviation is 1. At the beginning of the training process, neuron weight values of each hidden layer are initialized randomly, and the random value is selected from a uniform distribution within a range

$$\left(\frac{-1}{\sqrt{s}}, \frac{1}{\sqrt{s}}\right),$$

wherein the variants is the number of neurons of the previous hidden layer.

[0051] During the training of the video frame reconstruction model, each parameter the video frame reconstruction model is optimized using the stochastic gradient descent (SGD) algorithm. The SGD algorithm is applicable in optimization control processes in which many control variables and complex controlled systems are involved and accurate mathematical models cannot be established. In some embodiments, an initial learning rate can be set to be 0.001, and the learning rate changes to one tenth of the initial value after $3*10^6$ times of iterations. A momentum of the SGD algorithm can be set to 0.9, and a gradient can be cut during the descent of the random gradient. Supposed the objective function to be proposed is $E(x)=f(x)+r(x)$, wherein $f(x)$ is a loss function which can be any differentiable convex function for evaluating a training loss of the video frame reconstruction model, $r(x)$ is a normalized constraint factor for constraining the factor. According to a probability distribution of the parameters of the model, $r(x)$ typically includes a LI type of constraint (the model follows the Gauss distribution), and a L2 type of constraint (the model follows the Laplace distribution). The gradient can be kept within a certain range by cutting the weight updating gradient, such that a convergence of the model can be prevented from being affected by gradient explosion situation. The threshold value of the cutting of the gradient can be 10.

[0052] In other embodiments, as shown in FIG. **4**, which is a schematic view of the video frame reconstruction model used in the video compressed sensing method in accordance with an embodiment, the video frame reconstruction model includes an input layer, an output layer, and k hidden layers (k is a natural number greater than 1) each of which has the following formula:

$$h_k(y) \cdot = \cdot \theta(b_k+w_k y);$$

[0053] wherein $h_k(y) \in R^{L_k}$ is an activation value vector of the hidden layer k; L_k is the number of neurons of the hidden layer k, $\theta(*)$ that is $\theta(b_k+w_k y)$, is an activation function having a formula being $\theta(x) \cdot = \cdot \max(x, 0)$, $b_k \in R^{L_k}$ is a neuron bias vector of the hidden layer k, $w_k \in R^{L_{k-1} \times L_k}$ is a weight matrix, and $y \in R^{L_{k-1}}$ is an input vector of the hidden layer k.

[0054] The video frame reconstruction model $f(y_i;\omega)$ is obtained through training according to the activation value, the number of neurons, the activation function, the neuron bias value and the weight matrix, wherein ω is a parameter collection of the activation value, y_i the number of the neurons, the neuron bias vector, and the weight matrix, is the frame fragment input by the input layer, $f(y_i;\omega)$ is the nonlinear mapping between each frame fragment and the corresponding frame fragment block which is built by performing the feature abstraction to the frame fragments input by the input layer through the hidden layers.

[0055] Referring to FIG. **4**, the input layer receives the frame fragments; after the k hidden layers perform feature abstraction to the frame fragments, the frame fragment blocks enter into the output layer. Dimensions of the output layer are the same as the size of the video block finally reconstructed, both of which are $w_m \cdot * \cdot h_m \cdot * \cdot d_m$. In order to train the video frame reconstruction model, the weight values and bias of the model are continuously adjusted according to input parameters. Supposed that a set of all the parameters of the model is expressed as L(ω), an error back propagation (EP) algorithm is used to update the parameters, and the optimization function is mean squared error (MSE), a following formula can be obtained:

$$L(\omega) = \frac{1}{N} \Sigma_{i=1}^{N} \| f(y_i; \omega) - x_i \|_2^2 .$$

[0056] In some embodiments, dimensions of the input layer of the video frame reconstruction model may be set to be 8*8, dimensions of the output layer may be set to be 8*8*16, and the video frame reconstruction model may include 7 hidden layers having dimensions thereof set to be 128, 256, 384, 512, 512, 4096, and 2048, respectively.

[0057] The present disclosure further provides a video compressed sensing reconstruction system.

[0058] Referring to FIG. **5**, which is a block diagram of a video compressed sensing reconstruction system in accordance with an embodiment of the present disclosure, the system includes an extraction module **01**, a feature abstraction module **02**, and a reconstruction module **03**.

[0059] The extraction module **10** is configured for, after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the to-be-reconstructed compressed video frames according to a predetermined extraction rule.

[0060] In the embodiments, after the to-be-reconstructed compressed video frames are received, the compressed video frames are not rendered and reconstructed directly, and the frame fragments are extracted from the compressed video frames according to the predetermined extraction rule at first. The predetermined extraction rule can be extracting the frame fragments according to features like color, content, format, or area, etc., which is not limited herein.

[0061] In some embodiments, the predetermined extraction rule is: dividing the to-be-reconstructed compressed video frames into blocks such that the to-be-reconstructed compressed video frames are divided into several frame fragments. For example, various types of the compressed video frames, such as compressed video frames of JPEG or PNG type are divided into N*M frame fragments when the compressed video frames are divided into blocks, wherein N and M are positive integers. The frame fragments are equal to each other. In other embodiments, the compressed video frames can be divided into unequal frame fragments according to a preset ratio or randomly, which is not limited herein. Each frame fragment can have a regular shape like a square or a rectangle or irregular shapes, which is not limited herein.

[0062] The feature abstraction module **02** is configured for inputting the extracted frame fragments to an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the input frame fragments through multiple hidden layers of the pre-trained video frame reconstruction model, and building a nonlinear mapping between each frame fragment and the corresponding frame fragment block.

[0063] After the extraction of the frame fragments of the compressed video frames, the pre-trained video frame reconstruction model can be used to process the frame fragments correspondingly. The video frame reconstruction model can be established and trained each time when the video compressed sensing reconstruction is carried out or can be a pre-established and pre-trained model which is called each time when video compressed sensing reconstruction is carried out, which is not limited herein.

[0064] For example, in the embodiment, the pre-trained video frame reconstruction model includes the input layer, an output layer, and multiple hidden layers. After the frame fragments of the compressed video frames are extracted, the extracted frame fragments are input to the input layer of the pre-trained video frame reconstruction model; the hidden layers perform feature abstraction to the input frame fragments, build the nonlinear mapping between each frame fragment and the corresponding frame fragment block, and thus form a connection between each frame fragment and the reconstructed frame fragment block.

[0065] The reconstruction module **03** is configured for reconstructing the input frame fragments into the corresponding frame fragment block according to the nonlinear mapping built by the hidden layers of the pre-trained video frame reconstruction model, outputting reconstructed frame fragment blocks through the output layer of the pre-trained video frame reconstruction model, and generating are constructed video based on the reconstructed frame fragment blocks.

[0066] The hidden layers of the pre-trained video frame reconstruction model reconstruct the input frame fragments into the final frame fragment blocks according to the nonlinear mapping, that is, the mapping between each frame fragment and the corresponding final reconstructed frame fragment block after feature abstraction of each frame is performed. The reconstructed video frame blocks are output through the output layer and the reconstructed video is generated based on the reconstructed frame fragment blocks. In some embodiments, the frame fragment blocks can be spliced or combined to generate the reconstructed video, thereby realizing the rendering and reconstruction of the compressed video frames.

[0067] In the method of the present disclosure, the frame fragments of the to-be-reconstructed compressed video frames are extracted according to the predetermined extraction rule; feature abstraction is performed to each frame fragment through the multiple hidden layers of the pre trained pre-trained video frame reconstruction model, and the nonlinear mapping between each frame fragment and the corresponding frame fragment block is built; and the input frame fragments are reconstructed to the frame fragment blocks according to the nonlinear mapping, and the reconstructed video blocks are output. Since the reconstruction of the video is carried out corresponding to the frame fragments extracted from the to-be-reconstructed compressed video frames rather than by processing the large compressed video frames directly, thus the computing complexity is reduced and the reconstruction speed of the video is improved. In addition, since the feature abstraction of the frame fragment is carried out through the multiple hidden layers of the pre-trained video frame reconstruction model, and the frame fragments are reconstructed to the frame fragment blocks for outputting, each detailed feature of the compressed video frames can be effectively extracted and thus the quality of the reconstruction of the video frames is improved.

[0068] Referring to FIG. **6**, a video compressed sensing reconstruction system in accordance with a second embodiment of the present disclosure is provided. Based on the system of the first embodiment, the system of the second embodiment further includes a creation module **04**.

[0069] The creation module **04** is configured for creating and training the video frame reconstruction model which includes at least one input layer, at least one output layer, and multiple hidden layers.

[0070] In the embodiment, before the reconstruction of the video frames, the video frame reconstruction model is created and trained. The pre-trained video frame reconstruction model includes at least one input layer, one output layer, and multiple hidden layers. The creation module **04** includes a generation unit for generating training data and testing data. The generation unit obtains a preset number (e.g., 100) of videos in various types of natural scenes, and coverts and each video into a grayscale space. A total amount of data of all the videos satisfies a preset value (e.g., 10K).

[0071] The generation unit compresses each converted video through a measurement transformation matrix having a pre-defined size being $w_m \cdot * \cdot h_m \cdot * \cdot d_m$ (e.g., $w_m=8$, $h_m=8$, $d_m=16$). For example, w_b is a width of the video block having a preset number of videos, h_b is a height of the video block, d_b is a length of the video block (that is the number of the video frames). Each video block is $x_i \in w_b \times h_b \times d_b$, $i \in \cdot N$, wherein N is a positive integer no less than 1. The compressed video frame is $y_i \cdot \in \cdot w_b \cdot * \cdot h_b$, wherein $y_i = \O_b x_i$, wherein $\O_b$ is the measurement transformation matrix.

[0072] The generation unit divides all the compressed videos into a first data set and a second data set at a preset ratio such as X:Y (e.g., 7:3) wherein both X and Y are greater than 0. The number of the videos in first data set is greater than that of the videos in the second data set. The first data set is used as a training data set and the second data set is used as a testing data set.

[0073] In some embodiments, the training process of the video frame reconstruction model is as follows.

[0074] At the training of the video frame reconstruction model, a batch size of the input video frames is set to be 200, a total number of training times can be set to be $10*10^6$ times of iterations, and a size difference between every two video frames is specified within a range where an average value is 0 and a standard deviation is 1. At the beginning of the training process, neuron weight values of each hidden layer are initialized randomly, and the random value is selected from a uniform distribution within a range

$$\left(\frac{-1}{\sqrt{s}}, \frac{1}{\sqrt{s}}\right),$$

wherein the variants is the number of neurons of the previous hidden layer.

[0075] During the training of the video frame reconstruction model, each parameter the video frame reconstruction model is optimized using the stochastic gradient descent (SGD) algorithm. The SGD algorithm is applicable in optimization control processes in which many control variables and complex controlled systems are involved and accurate mathematical models cannot be established. In some embodiments, an initial learning rate can be set to be 0.001, and the learning rate changes to one tenth of the initial value after $3*10^6$ times of iterations. A momentum of the SGD algorithm can be set to 0.9, and a gradient can be cut during the descent of the random gradient. Supposed the objective function to be proposed is E(x)=f(x)+r(x), wherein f(x) is a loss function which can be any differentiable convex function for evaluating a training loss of the video frame reconstruction model, r(x) is a normalized constraint factor for constraining the factor. According to a probability distribution of the parameters of the model, r(x) typically includes a LI type of constraint (the model follows the Gauss distribution), and a L2 type of constraint (the model follows the Laplace distribution). The gradient can be kept within a certain range by cutting the weight updating gradient, such that a convergence of the model can be prevented from being affected by gradient explosion situation. The threshold value of the cutting of the gradient can be 10.

[0076] In other embodiments, the video frame reconstruction model includes an input layer, an output layer, and k hidden layers (k is a natural number greater than 1) each of which has the following formula:

$$h_k(y) \cdot = \cdot \theta(b_k + w_k y);$$

[0077] wherein $h_k(y) \in R^{L_k}$ is an activation value vector of the hidden layer k; L_k is the number of neurons of the hidden layer k, $\theta(*)$, that is $\theta(b_k + w_k y)$, is an activation function having a formula being $\theta(x) \cdot = \max(x, 0)$, $b_k \in R^{L_k}$ is a neuron bias vector of the hidden layer k, $w_k \in R^{L_{k-1} \times L_k}$ is a weight matrix, and $y \in R^{L_{k-1}}$ is an input vector of the hidden layer.

[0078] The video frame reconstruction model $f(y_i;\omega)$ is obtained through training according to the activation value, the number of neurons, the activation function, the neuron bias value and the weight matrix, wherein ω to is a parameter collection of the activation value, the number of the neurons, the neuron bias vector, and the weight matrix, y_i is the frame fragment input by the input layer, $f(y_i;\omega)$ is the nonlinear mapping between each frame fragment and the corresponding frame fragment block which is built by performing the feature abstraction to the frame fragments input by the input layer through the hidden layers.

[0079] Referring to FIG. **4**, the input layer receives the frame fragments, after the k hidden layers perform feature abstraction to the frame fragments, the frame fragment blocks enter into the output layer. Dimensions of the output layer are the same as the size of the video block finally reconstructed, both of which are $w_m\cdot*\cdot h_m\cdot*\cdot d_m$. In order to train the video frame reconstruction model, the weight values and bias of the model are continuously adjusted according to input parameters. Supposed that a set of all the parameters of the model is expressed as L(ω), an error back propagation (EP) algorithm is used to update the parameters, and the optimization function is mean squared error (MSE), a following formula can be obtained:

$$L(\omega) = \frac{1}{N}\Sigma_{i=1}^{N} \| f(y_i;\omega) - x_i \|_2^2 .$$

[0080] In some embodiments, dimensions of the input layer of the video frame reconstruction model may be set to be 8*8, dimensions of the output layer may be set to be 8*8*16, and the video frame reconstruction model may include 7 hidden layers having dimensions thereof set to be 128, 256, 384, 512, 512, 4096, and 2048, respectively.

[0081] In hardware implementation, the above extraction module **01**, the feature abstraction module **02**, the reconstruction module **03**, and the creation module **04** can be embedded in or independent from the electronic device as hardware, or can be stored in a storage device of the electronic device as software such that a processor of the electronic device can execute the above modules to perform corresponding operations. It is understood that the above processor can be a central processing unit, a micro-processor, or single chip, etc.

[0082] The present disclosure further provides a computer-readable storage medium on which a video compressed sensing reconstruction system is stored. The video compressed sensing reconstruction system can be executed by at least one processor such that the at least one processor can perform steps of the video compressed sensing reconstruction method of the above embodiments which includes step S**10**, S**20**, and S**30**, etc., which is not given in detail any more herein.

[0083] It should be noted that the term "comprising", "including" or any other variants thereof are intended to cover a non-exclusive inclusion, such that the process, method, product or device including a number of elements not only include these elements, but also other elements not explicitly listed, or but also inherent elements for the process, method, product or device. Unless otherwise restricted, the elements defined by the statement "comprise a . . . " does not exclude other elements included in the process, method, product or device including the said elements.

[0084] Through the foregoing description of the embodiments, it is clear to a person skilled in the art that the present invention may be implemented by software plus necessary universal hardware, and definitely may also be implemented by hardware, but in many cases, the software implementation is preferred. Based on such understanding, the essence of the technical solutions of the present disclosure, or part that makes contributions to the prior art, or part of the technical solution may be embodied in the form of a software product. The computer software product is stored in a storage medium (ROM, RAM, disk, disc), including several instructions such that any terminal device (which can be a mobile phone, a computer, an air conditioner, or a network device) can execute the methods of the above embodiments.

[0085] The contents described above are only preferred embodiments of the present disclosure, but the scope of the present disclosure is not limited to the embodiments. Any ordinarily skilled in the art would make any modifications or replacements to the embodiments in the scope of the present disclosure, and these modifications or replacements should be included in the scope of the present disclosure. Thus, the scope of the present disclosure should be subjected to the claims.

What is claimed is:

1. A video compressed sensing reconstruction method, comprising:

step B, after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the compressed video frames according to a predetermined extraction rule;

step C, inputting the frame fragments into an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and a corresponding frame fragment block; and

step D, reconstructing the input frame fragments to frame fragment blocks by the hidden layers according to the nonlinear mapping, and outputting the frame fragment blocks by an output layer of the video frame reconstruction model, and generating a reconstructed video based on the reconstructed frame fragment blocks.

2. The video compressed sensing reconstruction method of claim **1**, further comprising a following step before step B:

step A, creating and training the video frame reconstruction model which comprises at least one input layer, at least one output layer, and multiple hidden layers.

3. The video compressed sensing reconstruction method of claim **1**, wherein the video frame reconstruction model comprises one input layer, one output layer, and k hidden layers wherein k is a natural number being greater than 1, and each of the hidden layers comprises:

$$h_k(y)\cdot=\cdot\theta(b_k+w_k y);$$

wherein $h_k(y)\in R^{L_k}$ is an activation value vector of the hidden layer, L_k is the number of neurons of the hidden layer k, $\theta(b_k+w_k y)$ is an activation function, $b_k\in R^{L_k}$ is a neuron bias vector of the hidden layer, $w_k\in R^{L_{k-1}\times L_k}$ is a weight matrix, $y\in R^{L_{k-1}}$ is an input vector of the hidden layer; the video frame reconstruction model $f(y_i;\omega)$ is obtained through training according to the activation value, the number of neurons, the activation function, the neuron bias value, and the weight matrix, wherein $\omega\cdot$ is a parameter collection of the activation value the number of the neurons the neuron bias vector, and the weight matrix, y_i is the frame fragment input by the input layer, $f(y_i;\omega)\cdot$ is the nonlinear mapping between each frame fragment and the corresponding frame fragment block which is built by performing the feature abstraction to the frame fragments input by the input layer by the hidden layers.

4. The video compressed sensing reconstruction method of claim **2**, wherein the video frame reconstruction model comprises one input layer, one output layer, and k hidden layers wherein k is a natural number being greater than 1, and each of the hidden layers comprises:

$$h_k(y)\cdot=\cdot\theta(b_k+w_k y);$$

wherein $h_k(y)\in R^{L_k}$ is an activation value vector of the hidden layer, L_k is the number of neurons of the hidden layer k, $\theta(b_k+w_k y)$ is an activation function, $b_k\in R^{L_k}$ is a neuron bias vector of the hidden layer, $w_k\in R^{L_{k-1}\times L_k}$ is a weight matrix, $y\in R^{L_{k-1}}$ is an input vector of the hidden layer; the video frame reconstruction model $f(y_i;\omega)$ is obtained through training according to the activation value, the number of neurons, the activation function, the neuron bias value, and the weight matrix, wherein $\omega\cdot$ is a parameter collection of the activation value the number of the neurons the neuron bias vector, and the weight matrix, y_i is the frame fragment input by the input layer, $f(y_i;\omega)\cdot$ is the nonlinear mapping between each frame fragment and the corresponding frame fragment block which is built by performing the feature abstraction to the frame fragments input by the input layer by the hidden layers.

5. The video compressed sensing reconstruction method of claim **1**, wherein the predetermined extraction rule comprises:

dividing the to-be-reconstructed compressed video frames into blocks, and thus dividing the to-be-reconstructed compressed video frames into frame fragments.

6. The video compressed sensing reconstruction method of claim **2**, wherein the predetermined extraction rule comprises:

dividing the to-be-reconstructed compressed video frames into blocks, and thus dividing the to-be-reconstructed compressed video frames into frame fragments.

7. The video compressed sensing reconstruction method of claim **2**, wherein the step A further comprises a step of generating training data and testing data which comprises:

obtaining a preset number of videos of different types of natural scenes and converting each video into a gray-scale space;

compressing each converted video through a preset measurement transformation matrix; and

dividing the compressed videos into a first data set and a second data set at a preset ratio, wherein the first data set is used as a training data set and the second data set is used as a testing data set.

8-14. (canceled)

15. An electronic device, comprising a processor, a storage device and a video compressed sensing reconstruction system; the video compressed sensing reconstruction system being stored in the storage device and comprising at least one computer-readable instruction which is capable of being executed by the processor to perform:

after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the compressed video frames according to a predetermined extraction rule;

inputting the frame fragments into an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and a corresponding frame fragment block; and

reconstructing the input frame fragments to frame fragment blocks by the multiple hidden layers according to the nonlinear mapping, outputting the frame fragment blocks by an output layer of the video frame reconstruction model, and generating a reconstructed video based on the reconstructed frame fragment blocks.

16. The electronic device of claim **15**, wherein the at least one computer-readable instruction is executed by the processor to perform:

creating and training the video frame reconstruction model which comprises at least one input layer, at least one output layer, and multiple hidden layers.

17. The electronic device of claim **15**, wherein the video frame reconstruction model comprises one input layer, one outputting layer, and k hidden layers wherein k is a natural number being greater than 1, and each of the hidden layers comprises:

$$h_k(y)\cdot=\cdot\theta(b_k+w_k y)$$

wherein the video frame reconstruction model comprises one input layer, one output layer, and k hidden layers wherein k is a natural number being greater than 1, and each of the hidden layers comprises:

$$h_k(y)\cdot=\cdot\theta(b_k+w_k y)$$

wherein $h_k(y)\in R^{L_k}$ is an activation value vector of the hidden layer, L_k is the number of neurons of the hidden layer k, $\theta(b_k+w_k y)$ is an activation function, $b_k\in R^{L_k}$ is a neuron bias vector of the hidden layer, $w_k\in R^{L_{k-1}\times L_k}$ is a weight matrix, $y\in R^{L_{k-1}}$ is an input vector of the hidden layer; the video frame reconstruction model $f(y_i;\omega)$ is obtained through training according to the activation value, the number of neurons, the activation function, the neuron bias value, and the weight matrix, wherein $\omega\cdot$ is a parameter collection of the activation value the number of the neurons the neuron bias vector, and the weight matrix, y_i is the frame fragment input by the input layer, $f(y_i;\omega)\cdot$ is the nonlinear mapping between each frame fragment and the corresponding frame fragment block which is built by performing the feature abstraction to the frame fragments input by the input layer by the hidden layers.

18. The electronic device of claim **15**, wherein the predetermined extraction rule comprises:

dividing the to-be-reconstructed compressed video frames into blocks, and thus dividing the to-be-reconstructed compressed video frames into frame fragments.

19. The electronic device of claim **16**, wherein the at least one computer-readable instruction is executed by the processor to perform:

obtaining a preset number of videos of different types of natural scenes and converting each video into a gray-scale space;

compressing each converted video through a preset measurement transformation matrix; and dividing the compressed videos into a first data set and a second data set at a preset ratio, wherein the first data set is used as a training data set and the second data set is used as a testing data set.

20. A computer-readable storage medium with at least one computer-readable instruction stored thereon, which can be executed by at least one processor to perform:

after receiving to-be-reconstructed compressed video frames, extracting frame fragments of the compressed video frames according to a predetermined extraction rule;

inputting the frame fragments into an input layer of a pre-trained video frame reconstruction model, performing feature abstraction to the frame fragments through multiple hidden layers of the video frame reconstruction model, and building a nonlinear mapping between each frame fragment and a corresponding frame fragment block; and

reconstructing the input frame fragments to frame fragment blocks by the multiple hidden layers according to the nonlinear mapping, outputting the frame fragment blocks by an output layer of the video frame reconstruction model, and generating a reconstructed video based on the reconstructed frame fragment blocks.

21. The computer-readable storage medium of claim **20**, wherein the at least one computer-readable instruction is further executed by the processor to perform a following step before step B:

step A, creating and training the video frame reconstruction model which comprises at least one input layer, at least one output layer, and multiple hidden layers.

22. The computer-readable storage medium of claim **20**, wherein the video frame reconstruction model comprises one input layer, one output layer, and k hidden layers wherein k is a natural number being greater than 1, and each of the hidden layers comprises:

$$h_k(y) \cdot = \cdot \theta(b_k + w_k y);$$

wherein $h_k(y) \in R^{L_k}$ is an activation value vector of the hidden layer, L_k is the number of neurons of the hidden layer k, $\theta(b_k+w_k y)$ is an activation function, $b_k \in R^{L_k}$ is a neuron bias vector of the hidden layer, $w_k \in R^{L_{k-1} \times L_k}$ is a weight matrix, $y \in R^{L_{k-1}}$ is an input vector of the hidden layer; the video frame reconstruction model $f(y_i;\omega)$ is obtained through training according to the activation value, the number of neurons, the activation function, the neuron bias value, and the weight matrix, wherein $\omega\cdot$ is a parameter collection of the activation value the number of the neurons the neuron bias vector, and the weight matrix, y_i is the frame fragment input by the input layer, $f(y_i;\omega)\cdot$ is the nonlinear mapping between each frame fragment and the corresponding frame fragment block which is built by performing the feature abstraction to the frame fragments input by the input layer by the hidden layers.

23. The computer-readable storage medium of claim **21**, wherein the video frame reconstruction model comprises one input layer, one output layer, and k hidden layers wherein k is a natural number being greater than 1, and each of the hidden layers comprises:

$$h_k(y) \cdot = \cdot \theta(b_k + w_k y);$$

wherein $h_k(y) \in R^{L_k}$ is an activation value vector of the hidden layer, L_k is the number of neurons of the hidden layer k, $\theta(b_k+w_k y)$ is an activation function, $b_k \in R^{L_k}$ is a neuron bias vector of the hidden layer, $w_k \in R^{L_{k-1} \times L_k}$ is a weight matrix, $y \in R^{L_{k-1}}$ is an input vector of the hidden layer; the video frame reconstruction model $f(y_i;\omega)$ is obtained through training according to the activation value, the number of neurons, the activation function, the neuron bias value, and the weight matrix, wherein $\omega\cdot$ is a parameter collection of the activation value the number of the neurons the neuron bias vector, and the weight matrix, y_i is the frame fragment input by the input layer, $f(y_i;\omega)\cdot$ is the nonlinear mapping between each frame fragment and the corresponding frame fragment block which is built by performing the feature abstraction to the frame fragments input by the input layer by the hidden layers.

24. The computer-readable storage medium of claim **20**, wherein the predetermined extraction rule comprises:

dividing the to-be-reconstructed compressed video frames into blocks, and thus dividing the to-be-reconstructed compressed video frames into frame fragments.

25. The computer-readable storage medium of claim **21**, wherein the step A further comprises a step of generating training data and testing data which comprises:

obtaining a preset number of videos of different types of natural scenes and converting each video into a grayscale space;

compressing each converted video through a preset measurement transformation matrix; and

dividing the compressed videos into a first data set and a second data set at a preset ratio, wherein the first data set is used as a training data set and the second data set is used as a testing data set.

* * * * *